Geometric Programming for Communication Systems

Geometric Programming for Communication Systems

Mung Chiang

Electrical Engineering Department
Princeton University, Princeton
New Jersey 08544, USA

chiangm@princeton.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1 781 871 0245
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

Printed on acid-free paper

ISBN: 1-933019-09-3; ISSNs: Paper version 1567-2190; Electronic version 1567-2328

Contents

1

Introduction

1.1 Geometric Programming and Applications

Geometric Programming (GP) is a class of nonlinear optimization with many useful theoretical and computational properties. Although GP in standard form is apparently a non-convex optimization problem, it can be readily turned into a convex optimization problem, hence a local optimum is also a global optimum, the duality gap is zero under mild conditions,[1] and a global optimum can be computed very efficiently. 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 computational acceleration, distributed algorithms, and physical interpretations.

GP substantially broadens the scope of Linear Programming (LP) applications, and is naturally suited to model several types of important nonlinear systems in science and engineering. Since its inception

[1] Consider the Lagrange dual problem of a given optimization problem. Duality gap is the difference between the optimized primal objective value and the optimized dual objective value.

in 1960s,[2] GP has found applications in mechanical and civil engineering, chemical engineering, probability and statistics, finance and economics, control theory, circuit design, information theory, coding and signal processing, wireless networking, etc. For areas not related to communication systems, a very small sample of some of the GP application papers include [1, 24, 29, 38, 43, 44, 53, 57, 64, 65, 58, 92, 93, 104, 107, 112, 123, 125, 128]. Detailed discussion of GP can be found in the following books, book chapters, and survey articles: [52, 133, 10, 6, 51, 103, 54, 20]. Most of the applications in the 1960s and 1970s were in mechanical, civil, and chemical engineering. After a relatively quiet period in GP research in the 1980s and early to mid-1990s, GP has generated renewed interest since the late 1990s.

Over the last five years, GP has been applied to study a variety of problems in the analysis and design of communication systems, across many 'layers' in the layered architecture, from information theory and queuing theory to signal processing and network protocols. We also start to appreciate *why*, in addition to *how*, GP can be applied to a surprisingly wide range of problems in communication systems. These applications have in turn spurred new research activities on the theory and algorithms of GP, especially generalizations of GP formulations and distributed algorithms to solve GP in a network. This is a systematic survey of the applications of GP to the study of communication systems. It collects in one place various published results in this area, which are currently scattered in several books and many research papers, as well as a couple of unpublished results.

Although GP theory is already well-developed and very efficient GP algorithms are currently available through user-friendly software packages (e.g., MOSEK [129]), researchers interested in using GP still need to acquire the non-trivial capability of modelling or approximating engineering problems as GP. Therefore, in addition to the focus on the application aspects in the context of communication systems, this survey also provides a rather in-depth tutorial on the theory, algorithms, and modeling methods of GP.

[2] Appendix A briefly describes the history of GP.

1.2 Nonlinear Optimization of Communication Systems

LP and other classical optimization techniques have found important applications in communication systems for several decades (e.g., as surveyed in [15, 56]). Recently, there have been many research activities that utilize the power of recent developments in nonlinear convex optimization to tackle a much wider scope of problems in the analysis and design of communication systems.

These research activities are driven by both new demands in the study of communications and networking, and new tools emerging from optimization theory. In particular, a major breakthrough in optimization over the last two decades has been the development of powerful theoretical tools, as well as highly efficient computational algorithms like the interior-point methods (e.g., [12, 16, 17, 21, 97, 98, 111]), for nonlinear convex optimization, i.e., minimizing a convex function subject to upper bound inequality constraints on other convex functions and affine equality constraints:

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) \\ \mathsf{subject\ to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{c} \\ \mathsf{variables} & \mathbf{x} \in \mathbf{R}^n. \end{array} \tag{1.1}$$

The constant parameters are $\mathbf{A} \in \mathbf{R}^{l \times n}$ and $\mathbf{c} \in \mathbf{R}^l$. The objective function f_0 to be minimized and m constraint functions $\{f_i\}$ are convex functions.

From basic results in convex analysis [109], it is well known that for a convex optimization problem, a local minimum is also a global minimum. The Lagrange duality theory is also well developed for convex optimization. For example, the duality gap is zero under constraint qualification conditions, such as Slater's condition [21] that requires the existence of a strictly feasible solution to nonlinear inequality constraints. When put in an appropriate form with the right data structure, a convex optimization problem is also easy to solve numerically by efficient algorithms, such as the primal-dual interior-point methods [21, 97], which has worst-case polynomial-time complexity for a large class of functions and scales gracefully with problem size in practice.

Special cases of convex optimization include convex Quadratic Programming (QP), Second Order Cone Programming (SOCP), and Semidefinite Programming (SDP), as well as seemingly non-convex optimization problems that can be readily transformed into convex problems, such as GP. Some of these are covered in recent books on convex optimization, e.g., [12, 16, 17, 21, 97, 98]. While SDP and its special cases of SOCP and convex QP are now well-known in many engineering disciplines, GP is not yet as widely appreciated. This survey aims at enhancing the awareness of the tools available from GP in the communications research community, so as to further strengthen GP's appreciation–application cycle, where more applications (and the associated theoretical, algorithmic, and software developments) are found by researchers as more people start to appreciate the capabilities of GP in modeling, analyzing, and designing communication systems.

There are three distinctive characteristics in the nonlinear optimization framework for the study of communication systems:

- First, the watershed between efficiently solvable optimization problems and intractable ones is being recognized as 'convexity', instead of 'linearity' as was previously believed.[3] This has opened up opportunities on many nonlinear problems in communications and networking based on more accurate or robust modeling of channels and complex interdependency in networks. Inherently nonlinear problems in information theory may also be tackled.
- Second, the nonlinear optimization framework integrates various protocol layers into a coherent structure, providing a unified view on many disparate problems, ranging from classical Shannon theory on channel capacity and rate distortion [33] to Internet engineering such as inter-operability between TCP Vegas and TCP Reno congestion control [119].
- Third, some of these theoretical insights are being put into practice through field trials and industry adoption. Recent

[3] In some cases, global solutions and systematic relaxation techniques for non-convex optimization have also matured [101, 106].

examples include optimization-theoretic improvements of TCP congestion control [71] and DSL broadband access [118].

The phrase "nonlinear optimization of communication systems" in fact carries three different meanings. In the most straightforward way, an analysis or design problem in a communication system may be formulated as either minimizing a cost or maximizing a utility function over a set of variables confined within a constraint set. In a more subtle and recent approach, a given network protocol may be interpreted as a distributed algorithm solving an implicitly defined global optimization problem. In yet another approach, the underlying theory of a network control method or a communication strategy may be generalized using nonlinear optimization techniques, thus extending the scope of applicability of the theory. In Section 3, we will see that GP applications cover all three categories.

1.3 Overview

There are three main sections in this survey. Section 2 is a tutorial of GP: its basic formulations, convexity and duality properties, various extensions that significantly broaden the scope of applicability of the basic formulations, as well as numerical methods, robust solutions, and distributed algorithms for GP. Although this section does not cover any application topic, it is essential for modeling communication system problems in terms of GP and its generalizations.[4]

Section 3 is the core of this survey, presenting many applications of GP in the analysis and design of communication systems: the information theoretic problems of channel capacity, rate distortion, and error exponent in Subsection 3.1, construction of channel codes, relaxation of source coding problems, and digital signal processing algorithms for physical layer transceiver design in Subsection 3.2, network resource allocation algorithms such as power control in wireless networks in Subsection 3.3, network congestion control protocols in TCP Vegas and its cross-layer extensions in Subsection 3.4, and performance optimization of simple queuing systems in Subsection 3.5.

[4] For another very recent GP tutorial, readers are referred to a recent survey of GP for circuit design problems [20].

These applications generally fall into three categories: analysis (e.g., GP is used to characterize and bound information theoretic limits), forward engineering (e.g., GP is used to control transmit powers in wireless networks), and reverse engineering (e.g., GP is used to model congestion control or Highly Optimized Tolerance systems).

Then Section 4 explains why, rather than just how, GP can be applied to such a variety of problems in communication systems. As shown in Subsection 4.1, for problems based on stochastic models, GP is often applicable because large deviation bounds can be computed by GPs. As shown in Subsection 4.2, for problems based on deterministic models, reasons for applicability of GP is less well understood but may be due to GP's connections with proportional allocation, general market equilibrium, and generalized coding problems.

In the area of GP applications for communication systems, there are three most interesting directions of future research in author's view: distributed algorithms and heuristics for solving GP in a network, a systematic theory of using a nested family of GP relaxations for non-convex, generalized polynomial optimization, and the connections of GP with the theories of large deviation and general market equilibrium. These issues are discussed throughout the survey.

Some subsections in these three sections present unpublished results while most subsections summarize known results. In particular, Subsection 2.1 is partially based on [10, 21, 30, 52, 132], Subsection 2.2 on [6, 7, 10, 20, 51, 52, 103, 133], Subsection 2.3 on [21, 60, 67, 78, 37], Subsection 3.1 on [30, 33, 42, 82, 84, 120, 121, 122], Subsection 3.2 on [25, 30, 69, 75, 91], Subsection 3.3 on [37, 34, 35, 72, 73], Subsection 3.4 on [31, 88], Subsection 3.5 on [36, 68, 76], Subsection 4.1 on [30, 42, 45, 52, 108], and Subsection 4.2 on [28, 49, 70].

A brief historical account of the development of GP is provided in Appendix A and selected proofs are provided in Appendix B.

1.4 Notation

We will use the following notation. Vectors and matrices are denoted in boldface. Given two column vectors $\mathbf{x}$ and $\mathbf{y}$ of length n, we express the sum $\sum_{i=1}^{n} x_i y_i$ as an inner product $\mathbf{x}^T\mathbf{y}$. Componentwise inequalities

on a vector $\mathbf{x}$ with n entries are expressed using the $\succeq$ symbol: $\mathbf{x} \succeq 0$ denotes $x_i \geq 0, i = 1, 2, \ldots, n$. A column vector with all entries being 1 is denoted as $\mathbf{1}$. We use $\mathbf{R}_+^n$ and $\mathbf{R}_{++}^n$ to denote the non-negative and strictly positive quadrant of n-dimensional Euclidean space, respectively, and $\mathcal{Z}_+$ to denote the set of non-negative integers.

Sometimes a symbol has different meanings in different sections, because the same symbol is widely accepted as the standard notation representing different quantities in more than one field. For example, $\mathbf{P}$ denotes channel transition matrix in Subsection 3.1.1 on channel capacity, and denotes transmit power vector in Subsections 3.3.1 and 3.4.2 on wireless network power control. Such notational reuse should not cause any confusion since consistency is maintained within any single subsection.

All constrained optimization problems are written in this survey following a common format: objective function, constraints, and optimization variables. Constant parameters are also explicitly stated after the problem statement in cases where confusion may arise.

2

Geometric Programming

2.1 Formulations

2.1.1 Basic formulations and convexity property

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 a logarithmic change of variable. 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.

We first define a monomial as a function $f : \mathbf{R}^n_{++} \to \mathbf{R}$:[1]

$$f(\mathbf{x}) = dx_1^{a^{(1)}} x_2^{a^{(2)}} \cdots x_n^{a^{(n)}}$$

[1] Since the domain of monomials is the strictly positive quadrant of $\mathbf{R}^n$, when a GP is written in terms of monomials, it is implicitly assumed that the optimal variables cannot be zero. In theory, there is a loss of generality in this assumption for some applications. Numerically, this assumption may not introduce any difficulty since the interior-point method solves a GP through a feasible path inside the constraint set.

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

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

where $d_k \geq 0,\ k = 1, 2, \ldots, K$, and $a_k^{(j)} \in \mathbf{R},\ j = 1, 2, \ldots, n, k = 1, 2, \ldots, K$. The key features about a posynomial, which will be explained and utilized many places throughout this survey, are its positivity and convexity (in log domain).

For example, $2x_1^{-\pi}x_2^{0.5} + 3x_1x_3^{100}$ 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 a geometric program in standard form:[2]

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

where $f_i,\ i = 0, 1, \ldots, m$, are posynomials:

$$f_i(\mathbf{x}) = \sum_{k=1}^{K_i} d_{ik} x_1^{a_{ik}^{(1)}} x_2^{a_{ik}^{(2)}} \cdots x_n^{a_{ik}^{(n)}},$$

and $h_l,\ l = 1, 2, \ldots, M$ are monomials:

$$h_l(\mathbf{x}) = d_l x_1^{a_l^{(1)}} x_2^{a_l^{(2)}} \cdots x_n^{a_l^{(n)}}.$$

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

[2] Another name that is often used is 'posynomial form'.

Given a GP in standard form, we can form a matrix $\mathbf{A}$ where each row consists of the exponential constants associated with each monomial term that appears in the objective and constraints, and a vector $\mathbf{d}$ consisting of all the multiplicative constants. Each GP can be uniquely represented by the following data structure: $\mathbf{A}$, $\mathbf{d}$, and an identification of which rows in $\mathbf{A}$ and $\mathbf{d}$ belong to the objective function and which to each of the constraint functions.

As a small example, consider the following GP in standard form, with variables (x, y, z):[3]

$$\begin{array}{ll} \mathsf{minimize} & xy + xz \\ \mathsf{subject\ to} & \frac{0.8\sqrt{yz}}{x^2} \leq 1 \\ & \frac{0.5}{\sqrt{xy}} \leq 1 \\ & \frac{1}{x} \leq 1 \\ \mathsf{variables} & x, y, z. \end{array} \tag{2.2}$$

The constant parameters of this GP are:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ -2 & 1/2 & 1/2 \\ -1/2 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

and

$$\mathbf{d} = [1,\ 1,\ 0.8,\ 0.5,\ 1]^T$$

where the first two rows in $\mathbf{A}$ and $\mathbf{d}$ correspond to the two monomial terms in the objective function, and the last three rows each corresponds to a constraint.

GP in standard form is not a convex optimization problem, because posynomials are not convex functions. However, with a logarithmic change of all the variables and multiplicative constants: $y_i = \log x_i$,

[3] We will show in Subsection 3.1.1 that this GP is in fact computing a channel capacity with an input cost constraint.

$b_{ik} = \log d_{ik}, b_l = \log d_l$, we can turn it into the following problem:

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

where $\mathbf{a}_{ik} = [a_{ik}^{(1)}, a_{ik}^{(2)}, \ldots, a_{lk}^{(n)}]^T$, which is obviously equivalent[4] to the following GP in convex form:

$$\begin{array}{ll} \text{minimize} & p_0(\mathbf{y}) = \log \sum_{k=1}^{K_0} \exp(\mathbf{a}_{0k}^T \mathbf{y} + b_{0k}) \\ \text{subject to} & p_i(\mathbf{y}) = \log \sum_{k=1}^{K_i} \exp(\mathbf{a}_{ik}^T \mathbf{y} + b_{ik}) \leq 0, \\ & \qquad i = 1, 2, \ldots, m \\ & q_l(\mathbf{y}) = \mathbf{a}_l^T \mathbf{y} + b_l = 0, \quad l = 1, 2, \ldots, M \\ \text{variable} & \mathbf{y}. \end{array} \tag{2.3}$$

For the small example (2.2) of GP in standard form, the following problem is its convex form in $\tilde{x} = \log x, \tilde{y} = \log y, \tilde{z} = \log z$:

$$\begin{array}{ll} \text{minimize} & \log\left(\exp(\tilde{x} + \tilde{y}) + \exp(\tilde{x} + \tilde{z})\right) \\ \text{subject to} & 0.5\tilde{y} + 0.5\tilde{z} - 2\tilde{x} + \log 0.8 \leq 0 \\ & 0.5\tilde{x} + \tilde{y} + \log 0.5 \leq 0 \\ & -\tilde{x} \leq 0 \\ \text{variables} & \tilde{x}, \tilde{y}, \tilde{z}. \end{array}$$

To show that (2.3) is indeed a convex optimization problem, we need to show that the objective and inequality constraint functions are convex in $\mathbf{y}$. This convexity property can be readily verified through a positive-definiteness test of the Hessian. A more illuminating verification uses a duality argument.

Lemma 2.1. The *log-sum-exp function* $f(\mathbf{x}) = \log \sum_{i=1}^{n} e^{x_i}$ is convex in $\mathbf{x}$.

---

[4] Equivalence relationship between two optimization problems is used in a loose way throughout this survey. 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.

Proof. Consider the following log-sum inequality [42, 40] (readily proved by the convexity of $f(t) = t \log t, t \geq 0$):

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{2.4}$$

where $\mathbf{a}, \mathbf{b} \succeq 0$.[5]

Recall that, given a function $f : \mathbf{R}^n \to \mathbf{R}$, the function $f^\star : \mathbf{R}^n \to \mathbf{R}$, defined as

$$f^\star(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbf{dom}\, f} (\mathbf{y}^T \mathbf{x} - f(\mathbf{x})), \tag{2.5}$$

is called the conjugate function of f. Since $f^\star$ is the pointwise supremum of a family of affine functions of $\mathbf{y}$, it is always a convex function.[6]

Let $\hat{b}_i = \log b_i$ and $\sum_{i=1}^{n} a_i = 1$ in the log-sum inequality (2.4). We obtain

$$\log \left(\sum_{i=1}^{n} e^{\hat{b}_i} \right) \geq \mathbf{a}^T \hat{\mathbf{b}} - \sum_{i=1}^{n} a_i \log a_i,$$

with equality if and only if $a_i = \frac{e^{\hat{b}_i}}{\sum_j e^{\hat{b}_j}}$. This by definition shows that the log-sum-exp function is the conjugate function of negative entropy. Since all conjugate functions are convex, the log-sum-exp function is convex. □

The composition of a convex function with an affine function is a convex function, thus GP in convex form is indeed a convex optimization: minimizing a convex function subject to upper bound inequality constraints on convex functions and affine equality constraints.[7]

[5] This inequality also readily shows the convexity of the Kullback-Leibler divergence, or relative entropy, between two distributions $\mathbf{p}, \mathbf{q}$: $D(\mathbf{p} \| \mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ in $(\mathbf{p}, \mathbf{q})$ [40], which in turn shows that channel capacity and rate distortion problems, to be discussed in Subsection 3.1, are convex optimization problems.

[6] As a simple fact to be used in Subsection 4.1.1, it is easy to verify that, if $f^\star(\mathbf{y})$ is the conjugate of $f(\mathbf{x})$, then for a given $T > 0$, the perspective function $Tf^\star(\frac{\mathbf{y}}{T})$ is the conjugate of the scaled function $Tf(\mathbf{x})$.

[7] Sum-exp functions are also convex. In GP convex form, we further take the log of the sum-exp functions in the objective and constraints, which turns monomials into affine functions and also improves numerical stability of solution algorithms.

Convexity of the log-sum-exp function can also be verified from the following geometric inequality: the arithmetic mean is greater than or equal to the geometric mean [52]. It is for this reason that the name of geometric programming was used to describe the class of nonlinear optimization problems in the form of (2.1) or (2.3).

It is interesting to notice that, in the special case where all the posynomials in a GP in standard form are simply monomials, then this GP in convex form reduces to a LP. Hence GP can be viewed an an extension of LP.[8]

Examining GP in convex form, we appreciate why, in the definition of standard form GP, equality constraints can only be imposed on monomials. If there were posynomial equality constraints (or, equivalent, lower bound inequality constraints on posynomials), we would have obtained, after the logarithmic change of variable, a non-convex optimization problem, because the constraint set would not be convex even after the transformation. We also appreciate why, in the definition of posynomial, the exponential constants can be any real numbers (they appear only in an affine transformation of $\mathbf{y}$), but the multiplicative constants must be positive numbers (they need to be logarithmically transformed).

Note that although posynomial seems to be a non-convex function, it becomes a convex function after the log transformation, as shown in an example in Figure 2.1. Compared to the (constrained or unconstrained) minimization of a polynomial, the minimization of a posynomial in GP relaxes the integer constraint on the exponential constants but imposes a positivity constraint on the multiplicative constants and variables. There is a sharp contrast between these two problems: polynomial minimization is NP-hard, but GP can be turned into convex optimization with provably polynomial-time algorithms for a global optimum.

In an extension of GP called Signomial Programming, which will be discussed in Subsection 2.2.5, the restriction of non-negative multiplicative constants is removed, resulting in a general class of nonlinear

[8] This extension is different from the extension of LP to (convex) QP, which can be further extended to SOCP and SDP. GP is not a special case of SDP.

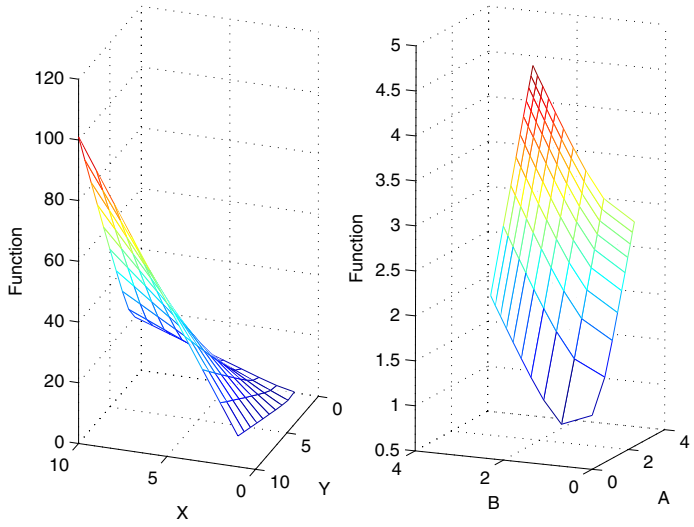

Fig. 2.1 A bi-variate posynomial before (left graph) and after (right graph) the log transformation. A non-convex function is turned into a convex one.

and truly non-convex problems that is simultaneously a generalization of GP and polynomial minimization over the positive quadrant, as summarized in the comparison Table 2.1.

	GP	*PMoP*	*SP*
c	$\mathbf{R}_+$	$\mathbf{R}$	$\mathbf{R}$
$a^{(j)}$	$\mathbf{R}$	$\mathcal{Z}_+$	$\mathbf{R}$
x_j	$\mathbf{R}_{++}$	$\mathbf{R}_{++}$	$\mathbf{R}_{++}$

Table 2.1 Comparison of GP, constrained polynomial minimization over the positive quadrant (PMoP), and Signomial Programming (SP). All three types of problems minimize a sum of monomials subject to upper bound inequality constraints on sums of monomials, but have different definitions of monomial: $c\prod_j x_j^{a^{(j)}}$. GP is known to be polynomial-time solvable, but PMoP and SP are not.

The objective function of Signomial Programming can be formulated as minimizing a ratio between two posynomials, which is not a posynomial since posynomials are closed under positive multiplication and addition but not division. As shown in Figure 2.2, a ratio between

two posynomials is a non-convex function both before and after the log transformation. Although it does not seem likely that Signomial Programming can be turned into a convex optimization problem, there are heuristics to solve it through a sequence of GP relaxations. Such methods current lack and would benefit significantly from a theoretical foundation similar to the sum-of-squares method [101, 102], which uses a nested family of SDP relaxations to solve constrained polynomial minimization problems.

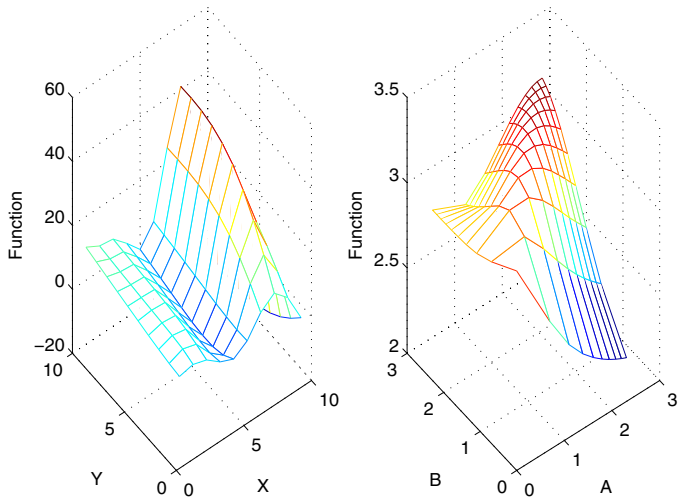

Fig. 2.2 Ratio between two bi-variate posynomials before (left graph) and after (right graph) the log transformation. It is a non-convex function in both cases.

2.1.2 Lagrange duality

The last subsection shows that a GP is a nonlinear, seemingly non-convex 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. 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.

Following the standard procedure of deriving the Lagrange dual problem [21], it is readily verified that for the following GP with m posynomial constraints,

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

the Lagrange dual problem is

$$\begin{array}{ll} \mathsf{maximize} & \mathbf{b}_0^T \boldsymbol{\nu}_0 - \sum_{j=1}^{K_0} \nu_{0j} \log \nu_{0j} \\ & + \sum_{i=1}^{m} \left(\mathbf{b}_i^T \boldsymbol{\nu}_i - \sum_{j=1}^{K_i} \nu_{ij} \log \frac{\nu_{ij}}{\mathbf{1}^T \boldsymbol{\nu}_i} \right) \\ \mathsf{subject\ to} & \boldsymbol{\nu}_i \succeq 0, \quad i = 0, \ldots, m \\ & \mathbf{1}^T \boldsymbol{\nu}_0 = 1 \\ & \sum_{i=0}^{m} \mathbf{A}_i^T \boldsymbol{\nu}_i = 0 \\ \mathsf{variables} & \boldsymbol{\nu}_i, \quad i = 0, 1, \ldots, m. \end{array} \tag{2.6}$$

The length of $\boldsymbol{\nu}_i$ is K_i, i.e., the number of monomial terms in the ith posynomial, $i = 0, 1, \ldots, m$. 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 $\mathbf{A}_i, \quad i = 1, 2, \ldots, m$, 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 $\mathbf{b}_i, \ i = 1, 2, \ldots, m$.

For the example GP (2.2) in the last subsection, its Lagrange dual problem is the following linearly constrained concave maximization:

$$\begin{array}{ll} \mathsf{maximize} & \nu_{01} + \nu_{02} - \nu_{01} \log \nu_{01} - \nu_{02} \log \nu_{02} \\ & + 0.8\nu_1 + 0.5\nu_2 + \nu_3 - \nu_1 \log \nu_1 - \nu_2 \log \nu_2 - \nu_3 \log \nu_3 \\ \mathsf{subject\ to} & \nu_{0j} \geq 0, \quad j = 1, 2 \\ & \nu_i \geq 0, \quad i = 1, 2, 3 \\ & \nu_{01} + \nu_{02} = 1 \\ & \mathbf{A}_0 \boldsymbol{\nu}_0 + \mathbf{A}_1 \nu_1 + \mathbf{A}_2 \nu_2 + \mathbf{A}_3 \nu_3 = 0 \\ \mathsf{variables} & \nu_{01}, \nu_{02}, \nu_1, \nu_2, \nu_3 \end{array}$$

where $\mathbf{A}_0 = [1, \ 1, \ 0; \ 1, \ 0, \ 1], \mathbf{A}_1 = [-2, \ 1/2, \ 1/2], \mathbf{A}_2 = [-1/2, \ -1, \ 0], \mathbf{A}_3 = [-1, \ 0, \ 0]$.

A special case is unconstrained GP:

$$\mathsf{minimize}_{\mathbf{y}} \log \sum_{i=1}^{N} \exp(\mathbf{a}_i^T \mathbf{y} + b_i).$$

From (2.6), the Lagrange dual problem of unconstrained GP reduces to

$$\begin{array}{ll} \mathsf{maximize} & \mathbf{b}^T \boldsymbol{\nu} - \sum_{i=1}^{N} \nu_i \log \nu_i \\ \mathsf{subject\ to} & \mathbf{1}^T \boldsymbol{\nu} = 1 \\ & \boldsymbol{\nu} \succeq 0 \\ & \mathbf{A}^T \boldsymbol{\nu} = 0 \\ \mathsf{variables} & \boldsymbol{\nu}. \end{array} \tag{2.7}$$

The first two constraints imply that the dual variable $\boldsymbol{\nu}$ must be a probability distribution. The dual problem (2.7) is a linearly constrained maximization of the entropy of distribution $\boldsymbol{\nu}$ plus a linear term $\mathbf{b}^T \boldsymbol{\nu}$.

The Lagrange dual problem (2.6) of a constrained GP can be interpreted as follows. Dual variable vector $\boldsymbol{\nu}_0$ is normalized but other dual variable vectors $\boldsymbol{\nu}_i,\ i = 1, 2, \ldots, m$, are not, and the objective function is a sum of linear terms and 'generalized' entropies of $\boldsymbol{\nu}_i$ (which are not normalized except when $i = 0$), to be maximized under a linear equality constraint where the weights are the exponential constants in the posynomials. The relationship between GP and free energy optimization in statistical physics will be discussed in Subsection 4.1.1.

By weak duality, any feasible solution of the dual GP, which can be easily computed by finding a solution to a system of linear inequalities, lower bounds the primal GP's optimal value. By strong duality, which holds for any GP in convex form that has a strictly feasible solution, the duality gap between a GP and its dual is zero. The optimal dual variables, as often is the case for convex optimization problems, provide very useful information about the sensitivity of the optimal solution to data perturbation and the tightness of a constraint at optimality.

2.1.3 Feasibility and sensitivity analysis

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

$$f_i(\mathbf{x}) \leq 1,\ \ i = 1, \ldots, m,\ \ h_l(\mathbf{x}) = 1,\ \ l = 1, \ldots, M, \tag{2.8}$$

is called a GP feasibility problem. Solving a 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 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 $\mathbf{x} \in \mathbf{R}^n$ [52, 20]:

$$\begin{array}{ll} \mathsf{minimize} & s \\ \mathsf{subject\ to} & f_i(\mathbf{x}) \leq s, \quad i = 1, \ldots, m \\ & s \geq 1 \\ \mathsf{variables} & \mathbf{x}, s. \end{array} \tag{2.9}$$

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 (2.9) and obtain the optimal $(s^*, \mathbf{x}^*)$. If $s^* = 1$, then the set of posynomial constraints $f_i(\mathbf{x}) \leq 1$ is feasible, and the associated $\mathbf{x}^*$ is a feasible solution to the original feasibility problem (2.9). 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 1th inequality constraint (with $u_1 > 0$) or tighten it (with $u_i < 0$), and shift the lth equality constraint (with $v_l \in \mathbf{R}$):

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) \\ \mathsf{subject\ to} & f_i(\mathbf{x}) \leq e^{u_i}, \quad i = 1, \ldots, m \\ & h_l(\mathbf{x}) = e^{v_l}, \quad l = 1, \ldots, M \\ \mathsf{variables} & \mathbf{x}. \end{array} \tag{2.10}$$

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:

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

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 (2.10): $S_i = -\lambda_i$ and $T_l = -\nu_l$ where $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are the Lagrange multipliers of the inequality and equality constraints in the convex form of (2.10), respectively.

There are also systematic procedures [52, 46, 48, 80] to obtain the optimizer $\mathbf{x}^*$ of the perturbed GP (2.10), without solving the perturbed problem from scratch, based on the constant parameter perturbations $(\mathbf{u}, \mathbf{v})$, and the exponent constant matrix $\mathbf{A}$, and multiplicative constant vector $\mathbf{d}$.

2.2 Extensions

The scope of GP formulations can be substantially expanded beyond the basic formulation in Section 2.1. We summarize these extensions in five groups in this section:

- Simple transformations by term rearrangements and partial change of variable.
- Generalized GP that allows compositions of posynomials with other functions.
- Extended GP based on other geometric inequalities.
- GP formulations based on monomial and posynomial approximations of nonlinear functions.
- Signomial Programming that allows posynomial equality constraints.

It is important to note that, unlike the first four groups of extensions, Signomial Programming cannot be transformed into convex

optimization problems. The third and fourth groups of extensions enlarge the scope of GP formulations so much that many convex optimization problems can fit into the GP framework. In contrast, the first and second groups of extensions transform a problem that already 'looks like' a GP into an equivalent GP in standard form. In the next two sections on GP applications in communication systems, we will primarily use the transformations in the first and second groups of extensions.

Unless specified otherwise, f denotes posynomials and g, h denote monomials in this subsection. When it is clear that $\mathbf{x}$ are the variables, we omit the variables field in the data structure of the representation of an optimization problem.

2.2.1 Simple transformations

It is trivial to realize that the following problems are GPs.

Extension 1: Maximize a monomial subject to posynomial upper bound inequality constraints:

$$\begin{array}{ll} \text{maximize} & h_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 1, \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 1, \quad j = 1, \ldots, M. \end{array}$$

This is still a GP because maximizing a monomial is equivalent to minimizing its reciprocal, which is another monomial (thus a posynomial):

$$\begin{array}{ll} \text{minimize} & \frac{1}{h_0(\mathbf{x})} \\ \text{subject to} & f_i(\mathbf{x}) \leq 1, \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 1, \quad j = 1, \ldots, M. \end{array}$$

Extension 2: The right hand side of poysnomial inequality and monomial equality constraints can be monomials instead of 1:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq g_i(\mathbf{x}), \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = g_j(\mathbf{x}), \quad j = 1, \ldots, M. \end{array}$$

This is still a GP because by dividing the right hand side monomial on both sides of the constraints, we obtain upper bound inequality

constraints on the ratio between a posynomial and a monomial (which is another posynomial) and equality constraints on the ratio between two monomials (which is another monomial):

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) \\ \mathsf{subject\ to} & \frac{f_i(\mathbf{x})}{g_i(\mathbf{x})} \leq 1, \quad i = 1, \ldots, m \\ & \frac{h_j(\mathbf{x})}{g_j(\mathbf{x})} = 1, \quad j = 1, \ldots, M. \end{array}$$

In general, we define an inverted posynomial as the ratio between a monomial and a posynomial. Lower bounding an inverted posynomial is allowed in a GP since it is equivalent to upper bounding a posynomial.

Extension 3: Positive sums and products of posynomials are also posynomials. For example,

$$(xy^{1/2} + z)(x^{-1/2} + yz) + xy = x^{1/2}y^{1/2} + xy^{2/3}z + x^{-1/2}z + yz^2 + xy.$$

Therefore, the objective function and inequality constraint functions in a GP can be any positive sums and products of posynomials.

Suppose the variables of an optimization problem can be separated into two sets, and no term in the problem involves variables from more than one set. If the problem is convex in one set of log-transformed variables and a convex optimization in the other set of variables, we can use a log change of variables only for the first set and obtain a convex optimization (although not a convex form GP). An example is the following:

Extension 4: An optimization problem is called a **Mixed Linear Geometric Programming (MLGP)** if with a log change of $\mathbf{x}$ variables, the following problem can be turned into a convex optimization:

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) + \mathbf{a}_0^T\mathbf{y} \\ \mathsf{subject\ to} & f_i(\mathbf{x}) + \mathbf{a}_i^T\mathbf{y} + d_i \leq 1, \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 1, \quad j = 1, \ldots, M \\ \mathsf{variables} & \mathbf{x}, \mathbf{y}. \end{array}$$

Extension 5: Consider the following unconstrained minimization problem:

$$\mathsf{minimize} \quad \sum_{i=1}^{m} \exp(f_i(\mathbf{x})),$$

where f_i are posynomials with non-negative exponents. Introducing auxiliary variables t_i, we can transform the above problem into the following equivalent problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} \exp(t_i) \\ \text{subject to} & f_i(\mathbf{x}) \leq t_i, \quad i = 1, \ldots, m \\ \text{variables} & \mathbf{x}, \mathbf{t}. \end{array}$$

At optimality, the inequality constraints will be tight. Now apply a log change of variable from $\mathbf{x}$ to $\mathbf{y}$, the above problem can be turned into a convex optimization in $(\mathbf{y}, \mathbf{t})$. This method of utilizing monotonicity of posynomials with non-negative exponents and introducing auxiliary variables is a common one that will also be used in the next subsection.

There are other simple transformations that only require basic arithmetics, such as the following:

Extension 6: Maximizing a sum of log of monomials:

$$\sum_{i=1}^{m} \log(h_i(\mathbf{x}))$$

is equivalent to a GP of minimizing the following monomial:

$$\prod_{i=1}^{m} (h_i^{-1}(\mathbf{x})).$$

2.2.2 Generalized GP

Generalized GP refers to minimizing a generalized posynomial subject to upper bound inequality constraints on generalized posynomials. A generalized posynomial is a composition of the following three functions, each of which generalizes the posynomial function.[9]

Extension 7: Consider composing posynomials $\{f_{ij}(\mathbf{x})\}$ with a posynomial with non-negative exponents $\{a_{ij}\}$:

$$F_1(\mathbf{x}) = \sum_i d_j \prod_j f_{ij}^{a_{ij}}(\mathbf{x}).$$

[9] The term 'Generalized GP' is also used by some authors to denote Signomial Programming, which will be discussed in Subsection 2.2.5. In this survey, Generalized GP should not be confused with Signomial Programming.

Introducing auxiliary variables t_{ij}, we see that minimizing $F_1(\mathbf{x})$ is equivalent to the following GP:

$$\begin{array}{ll} \text{minimize} & \sum_i d_j \prod_j t_{ij}^{a_{ij}} \\ \text{subject to} & f_{ij}(\mathbf{x}) \leq t_{ij}, \quad \forall i,j \\ \text{variables} & \mathbf{x}, \{t_{ij}\}. \end{array}$$

This equivalence is due to the monotonicity of posynomials with non-negative exponents. Similarly, an upper bound inequality constraint on $F_1(\mathbf{x})$ can be turned into the following posynomial constraints in variables $(\mathbf{x}, \{t_{ij}\})$: $\sum_i d_j \prod_j t_{ij}^{a_{ij}} \leq 1, \quad f_{ij}(\mathbf{x}) \leq t_{ij}, \quad \forall i,j$. Composing posynomials with a posynomial with non-negative exponents produces a generalized posynomial F_1. Posynomial objective and constraints in a GP can be substituted with generalized posynomials in the form of F_1 while maintaining the GP form of the problem.

Extension 8: The maximum of a finite number of posynomials is also a generalized posynomial, because

$$\text{minimize} \;\; F_2(\mathbf{x}) = \max_i \{f_i(\mathbf{x})\}$$

is equivalent to the following GP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_i(\mathbf{x}) \leq t, \quad \forall i \\ \text{variables} & \mathbf{x}, t. \end{array}$$

Extension 9: The following function is also a generalized posynomial:

$$F_3(\mathbf{x}) = \frac{f_1(\mathbf{x})}{h(\mathbf{x}) - f_2(\mathbf{x})}$$

where f_1 and f_2 are posynomials and h is a monomial.

This is because minimizing $F_3(\mathbf{x})$ is equivalent to the following GP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \frac{f_2(\mathbf{x})}{h(\mathbf{x})} + \frac{f_1(\mathbf{x})}{t h(\mathbf{x})} \leq 1 \\ \text{variables} & \mathbf{x}, t. \end{array}$$

Any combination of composition of F_1, F_2, F_3 is also a generalized posynomial. Minimizing a generalized posynomial subject to upper

bound inequality constraints on generalized posynomials is a Generalized GP, which can be turned into a standard form GP by introducing auxiliary variables as shown above.

As an example, the following problem in $\mathbf{x}$ is a generalized GP:

$$\begin{array}{ll} \text{minimize} & \max\{(x_1+x_2^{-1})^{0.5}, x_1x_3\} + (x_2+x_3^{-2.9})^{1.5} \\ \text{subject to} & \frac{(x_2x_3+x_2/x_1)^{\pi}}{x_1x_2-\max\{x_1^2x_3^3, x_1+x_3\}} \leq 10 \\ \text{variables} & x_1, x_2, x_3, \end{array}$$

which is equivalent to the following GP:

$$\begin{array}{ll} \text{minimize} & t_1 + t_2^{1.5} \\ \text{subject to} & 0.1(t_4^{\pi}+t_5)x_1^{-1}x_2^{-1} \leq 1 \\ & t_3^{0.5}t_1^{-1} \leq 1 \\ & x_1x_3t_1^{-1} \leq 1 \\ & (x_1+x_2^{-1})t_3^{-1} \leq 1 \\ & (x_2+x_3^{-2.9})t_2^{-1} \leq 1 \\ & (x_2x_3+x_2x_1^{-1})t_4^{-1} \leq 1 \\ & x_1^2x_3^2t_5^{-1} \leq 1 \\ & (x_1+x_3)t_5^{-1} \leq 1 \\ \text{variables} & x_1, x_2, x_3, t_1, t_2, t_3, t_4, t_5. \end{array}$$

2.2.3 Extended GP

The theory of GP, including the convexity and duality properties, can be developed from a basic geometric inequality: the arithmetic mean is greater than or equal to the geometric mean. GP can be extended by other geometric inequalities. This subsection provides a brief introduction to such Extended GP. Further details concerning Extended GP and its refined duality theory can be found in e.g. [52, 6].

The following inequality

$$\mathbf{x}^T\mathbf{y} \leq \lambda(\mathbf{y})G(\mathbf{x}) - F(\mathbf{y}), \tag{2.11}$$

for $\mathbf{x}$ in an open convex set and $\mathbf{y}$ in a cone, is called a geometric inequality if $\lambda(\mathbf{y})$ is non-negative, $G(\mathbf{x})$ is differentiable, and for every $\mathbf{x}$ there is a $\mathbf{y}$ for which the inequality becomes equality.

Different geometric inequalities lead to different classes of Extended GP. The following choice of functions leads to the basic version of convex form GP:

$$\begin{aligned} G(\mathbf{x}) &= \log \sum_{i=1}^{n} \exp(x_i) \\ \lambda(\mathbf{y}) &= \mathbf{1}^T \mathbf{y} \\ F(\mathbf{y}) &= -\sum_{i=1}^{n} y_i \log y_i - \lambda(\mathbf{y}) \log \lambda(\mathbf{y}). \end{aligned}$$

If $\mathbf{y}$ is a probability distribution, i.e., $\mathbf{y} \succeq 0, \mathbf{1}^T\mathbf{y} = 1$, the above geometric inequality reduces to the conjugacy relationship between log-sum-exp function and negative entropy. Indeed, it can be shown that $F(\mathbf{y})$ is simply the conjugate function of $G(\mathbf{x})$ restricted to the set of $\mathbf{y}$ such that there is an $\mathbf{x}$ for which $\nabla G(\mathbf{x}) = \mathbf{y}$.

In general, geometric inequality and conjugate function are related as follows [6]. By scaling the defining equation (2.5) of conjugacy relationship by $\lambda > 0$, we have:

$$\mathbf{x}^T\mathbf{y} \leq \lambda G(\mathbf{x}) + \lambda G^\star(\mathbf{y}/\lambda)$$

where $G^\star(\mathbf{y})$ is the conjugate function of $G(\mathbf{x})$. Now generalize the constant λ to a homogeneous positive function $\lambda(\mathbf{y})$ such that $\lambda(\nabla G(\mathbf{x})) = 1$, we have

$$\mathbf{x}^T\mathbf{y} \leq \lambda(\mathbf{y})G(\mathbf{x}) + \lambda(\mathbf{y})G^\star(\mathbf{y}/\lambda(\mathbf{y})).$$

Letting $F(\mathbf{y}) = \lambda(\mathbf{y})G^\star(\mathbf{y}/\lambda(\mathbf{y}))$ recovers the geometric inequality (2.11) from this generalization of the conjugacy relationship.

It can be readily verified [52] that $G(\mathbf{x})$ in a geometric inequality is always a convex function. Once a geometric inequality is obtained, minimizing an objective function in the form of $G(\mathbf{x})$ subject to inequality constraints on other functions in the form of $G(\mathbf{x})$ is a convex optimization, called the Extended GP derived from this geometric inequality. In this sense, Extended GP covers a very wide range of convex optimization problems. As long as the objective and constraint functions can be obtained from some geometric inequality, the associated convex optimization problem is an Extended GP.

Not only there is always a convex function accompanying a geometric inequality, geometric inequalities can also be constructed from convex functions. Suppose a function $g(\mathbf{x})$ is positive, convex, differentiable, and homogeneous of degree $p > 1$. Restrict its domain to be an open half space with a boundary containing the origin. Let $h(\mathbf{y})$ be the conjugate function of $g(\mathbf{x})$ and let q be such that $1/p + 1/q = 1$. Then the following geometric inequality, which is an extension of Holder's inequality, can be constructed [52]:

$$\mathbf{y}^T\mathbf{x} \leq (pg(\mathbf{x}))^{1/p}(qh(\mathbf{y}))^{1/q}.$$

In particular, given a differentiable (except possibly at the origin) norm, there is a positive, homogeneous function $\lambda(\mathbf{y})$ such that

$$\mathbf{y}^T\mathbf{x} \leq \lambda(\mathbf{y})\|\mathbf{x}\|,$$

which is a geometric inequality when $\mathbf{x}$ is confined to an open half space whose boundary contains the origin. This leads to the following:

Extension 10: Minimizing a differentiable norm over an open half space whose boundary contains the origin is an Extended GP.

As a more specific example of Extended GP, consider the following geometric inequality:

$$\begin{aligned} G(\mathbf{x}) &= \sum_{i=1}^{N-1} \exp(x_i + \log c_i) + x_N + \log c_N \\ \lambda(\mathbf{y}) &= y_N \\ F(\mathbf{y}) &= \log y_N \left(\sum_{i=1}^{N-1} y_i\right) - \sum_{i=1}^{N-1} y_i(\log y_i - 1) + \sum_{i=1}^{N} y_i \log c_i \end{aligned}$$

where $\mathbf{c} \succeq 0$ is a constant vector. This geometric inequality leads to the following [52]:

Extension 11: Minimizing the sum of a posynomial and the log of a monomial, subject to inequality posynomial constraints, is an Extended GP.

2.2.4 Approximation and fitting

Sometimes an objective or constraint function in an optimization problem is not a posynomial, and we would like to use a monomial or a

posynomial to approximate the given function so that optimizing or constraining this function can be accommodated in a GP. Intuitively, a function $f(\mathbf{x})$ can be accurately approximated by a posynomial if the function $F(\mathbf{y}) = \log f(\exp(\mathbf{y}))$ can be accurately approximated by a convex function.

Consider the simple case of monomial approximation [6, 20]. We are given a nonlinear function $f : \mathbf{R}^n \to \mathbf{R}$ that is differentiable at $\mathbf{x}_0 \succ 0$ and $f(\mathbf{x}_0) > 0$. We would like to approximate $f(\mathbf{x})$ with a monomial $\hat{f}(\mathbf{x}) = c\prod_{i=1}^n x_i^{a_i}$. We use the following log transformation: $y_i = \log x_i, g(\mathbf{y}) = \log f(\mathbf{y}), \hat{g}(\mathbf{y}) = \log \hat{f}(\mathbf{y}) = \log c + \mathbf{a}^T\mathbf{y}$. Equating the first order Taylor expansion of g at $\mathbf{y}_0$ with $\hat{g}(\mathbf{y})$, we obtain:

$$g(\mathbf{y}_0) + \nabla g(\mathbf{y}_0)^T(\mathbf{y} - \mathbf{y}_0) = \log c + \mathbf{a}^T\mathbf{y},$$

which implies that

$$\mathbf{a} = \nabla g(\mathbf{y}_0),$$

i.e.,

$$a_i = \frac{x_i}{f(\mathbf{x})}\frac{\partial f}{\partial x_i}\bigg|_{\mathbf{x}=\mathbf{x}_0}, \quad \forall i,$$

and that

$$c = \exp(g(\mathbf{y}_0) - \nabla g(\mathbf{y}_0)^T\mathbf{y}_0) = f(\mathbf{x})\prod_{i=1}^n x_i^{a_i}\bigg|_{\mathbf{x}=\mathbf{x}_0}.$$

Once $\mathbf{a}$ and c are computed, a monomial approximation $\hat{f}(\mathbf{x})$ to the original nonlinear function $f(\mathbf{x})$ is obtained.

Sometimes we are given a set of empirical data points, which we would like to curve-fit using monomials, posynomials, or generalized posynomials. Posynomials and generalized posynomials offer much flexibility in fitting empirical data, which can then be used in modeling the problem in a GP formulation. Methods for monomial and posynomial data fitting are explained in [20].

A limiting argument can also be used to allow a GP to approximate a nonlinear optimization problem.

Extension 12: Consider an unconstrained minimization of

$$f(\mathbf{x}) + \exp(g(\mathbf{x})).$$

Since $e^z = \lim_{\phi\to\infty}\left(1+\frac{z}{\phi}\right)^{\phi}$, the above minimization can be approximated for a large, fixed ϕ through the following GP, where t is an auxiliary variable:

$$\begin{array}{ll} \mathsf{minimize} & f(\mathbf{x}) + t^{\phi} \\ \mathsf{subject\ to} & t^{-1} + t^{-1}\phi^{-1}g(\mathbf{x}) \leq 1 \\ \mathsf{variables} & \mathbf{x}, t. \end{array}$$

Nonlinear programs that involve both posynomials and exponentials (or logarithms) of posynomials are called Transcendental GP. They are in general not convex optimization problems and the Lagrange dual problems may not be linearly constrained [6].

2.2.5 Signomial Programming

All of the GP extensions discussed so far (except Transcendental GP) can be converted into convex optimization problems. This subsection focuses on Signomial Programming (SP), which is an extension of GP that in general cannot be turned into convex problems.

In standard form GP, only upper bound inequality constraints are allowed on posynomials. Sometimes a posynomial represents a quality of service that needs to be lower bounded. Equality constraints on posynomials are also common in network modeling. In particular, flow conservation equality constraints are linear equality constraints. It is a limitation in GP modeling that lower bound inequalities (or equalities) on posynomials are not allowed in standard form GP.

This issue can be tackled by extending GP to SP: minimizing a signomial subject to upper bound inequality constraints on signomials, where a signomial is a sum of monomials, possibly with negative multiplicative coefficients:

$$s(\mathbf{x}) = \sum_{i=1}^{N} c_i \prod_{j=1}^{n} x_j^{a_i^{(j)}}$$

where $\mathbf{c} \in \mathbf{R}^N,\ a_i^{(j)} \in \mathbf{R},\ \forall i, j,\ \mathbf{x} \in \mathbf{R}^n_{++}$.

As shown in Table 2.1, SP covers a wide range of constrained, generalized polynomial minimization problems. Standard form GP is clearly

a special case of SP. Problems with posynomial equality constraints is a special case of problems with both upper and lower bound inequality constraints on posynomials, which is in turn equivalent to SP. Polynomial minimization with positive variables, which is NP-hard, is also a special case of SP where the exponents in signomials are non-negative integers.

The Lagrange dual problem of a SP also has the desired feature of being linearly constrained as in the dual of GP. However, in sharp contrast to GP, SP in general cannot be turned into convex optimization problems or be polynomial-time solved for global optimality, and the duality gap is non-zero.

There are at least four major approaches to solve, or approximately solve, a SP.

Approach 1: Branch and bound. The first approach is a standard branch and bound technique for general non-convex optimization, which does not utilize the special structure of SP and will not be discussed here.

Approach 2: Relaxations that are provably tight. The second approach is based on relaxations that do not incur any loss of generality at the optimal solution for some special cases of SP. For example [20], consider the following SP that is almost a standard form GP except an equality constraint on a posynomial $\tilde{f}(\mathbf{x})$:

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) \\ \mathsf{subject\ to} & f_i(\mathbf{x}) \leq 1, \;\; i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 1, \;\; j = 1, \ldots, M \\ & \tilde{f}(\mathbf{x}) = 1. \end{array} \tag{2.12}$$

We form a relaxation of the above problem by replacing the equality constraint on $\tilde{f}$ with an upper bound inequality constraint:

$$\begin{array}{ll} \mathsf{minimize} & f_0(\mathbf{x}) \\ \mathsf{subject\ to} & f_i(\mathbf{x}) \leq 1, \;\; i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 1, \;\; j = 1, \ldots, M \\ & \tilde{f}(\mathbf{x}) \leq 1. \end{array}$$

This is now a standard form GP in $\mathbf{x}$, which can be efficiently solved for a global optimizer $\tilde{\mathbf{x}}^*$. If the following three conditions are satisfied: there is a variable x_k such that it does not appear in any of the

monomials, f_i are decreasing functions of x_k for $i = 0, 1, \ldots, m$, and $\tilde{f}$ is a strictly increasing function of x_k, then we can increase the kth component of $\tilde{\mathbf{x}}^*$ until the $\tilde{f}(\mathbf{x}) \leq 1$ constraint becomes tight. The monotonicity assumptions clearly show that the resulting $\mathbf{x}$ is an optimizer $\mathbf{x}^*$ of the original problem (2.12). This approximation technique can be generalized to the case of multiple posynomial equality constraints [20]. It does not always apply to a general SP, but when it is applicable, it generates a globally optimal solution in polynomial time by using a GP solver.

Approach 3: Reversed GP. The third and fourth approaches work for any SP but may take exponential time to compute an optimal solution [7, 51]. The third approach converts a SP into a Reversed GP, and then apply a monomial approximation iteratively [10]. Reversed GP refers to minimizing a posynomial subject to both upper and lower bound inequality constraints.

Consider the following SP, written in a form where the monomial terms with negative multiplicative coefficients $\{f_{i2}(\mathbf{x})\}$ are separated from those monomial terms with positive multiplicative coefficients $\{f_{i1}(\mathbf{x})\}$, $i = 0, 1, \ldots, m$:

$$\begin{array}{ll} \mathsf{minimize} & f_{01}(\mathbf{x}) - f_{02}(\mathbf{x}) \\ \mathsf{subject\ to} & f_{i1}(\mathbf{x}) - f_{i2}(\mathbf{x}) \leq 1, \quad i = 1, \ldots, m. \end{array}$$

We first need to convert the signomial objective function into the form required by Reversed GP. If the optimal objective function value is positive, we introduce an auxiliary variable t and turn the objective to the minimization of t, with an additional constraint:

$$f_{01}(\mathbf{x}) - f_{02}(\mathbf{x}) \leq t,$$

which may be written as

$$f_{01}(\mathbf{x}) \leq s \leq f_{02}(\mathbf{x}) + t$$

where $s \geq 0$ is another auxiliary variable. The above inequalities can be written as two posynomial inequalities:

$$s^{-1} f_{02}(\mathbf{x}) + s^{-1} t \geq 1 \tag{2.13}$$

$$s^{-1} f_{01}(\mathbf{x}) \leq 1. \tag{2.14}$$

If instead the optimal objective function value is negative, we can introduce auxiliary variables s and t and turn the objective into minimizing t, subject to two additional posynomial constraints:

$$s^{-1}f_{01}(\mathbf{x}) + s^{-1}t \leq 1 \tag{2.15}$$

$$s^{-1}f_{02}(\mathbf{x}) \geq 1. \tag{2.16}$$

We do not know which of the above two cases should be used a priori. We can use one of them, and if the resulting optimal value is 0, we should shift to the other case.

We then transform the SP constraint $f_{i1}(\mathbf{x}) - f_{i2}(\mathbf{x}) \leq 1$ into the form required by Reversed GP, introducing auxiliary variables u_i:

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

which can be written as two posynomial constraints:

$$u_i^{-1}f_{i1}(\mathbf{x}) \leq 1$$
$$u_i^{-1}(f_{i2}(\mathbf{x}) + 1) \geq 1.$$

After rewriting a SP as a Reversed GP, we use the following approach for the lower bound inequality constraints on a posynomial $f_i(\mathbf{x})$. Since $f_i(\mathbf{x}) \geq 1$ is equivalent to $1/f_i(\mathbf{x}) \leq 1$, if we can approximate the posynomial $f_i(\mathbf{x})$ with a monomial, then a lower bound on $f_i(\mathbf{x})$ becomes an upper bound on a monomial, which is allowed in standard form GP.[10] This monomial approximation can be computed using the technique in Subsection 2.2.4. A simpler approximation is based on the geometric inequality that lead to the development of GP: the arithmetic mean is greater than or equal to the 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

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

[10] Alternatively, we can approximate a posynomial with an inverted posynomial, e.g., through the arithmetic mean harmonic mean inequality.

Let $\{u_i(\mathbf{x})\}$ be the monomial terms in a posynomial $f(\mathbf{x}) = \sum_i u_i(\mathbf{x})$. A lower bound inequality on posynomial $f(\mathbf{x})$ can now be approximated by an upper bound inequality on the following monomial:

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

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

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

which obviously satisfies the condition that $\boldsymbol{\alpha} \succ 0$ and $\mathbf{1}^T\boldsymbol{\alpha} = 1$. Given an $\boldsymbol{\alpha}$ for each lower bound posynomial inequality, a standard form GP can be obtained based on the above geometric mean approximation of a Reversed GP.

Notice that what is important is to have a monomial approximation of a posynomial. However, the geometric mean approximation and the above choice of $\boldsymbol{\alpha}$ may not lead to the best approximation in the sense of minimizing the approximation error or facilitating the computation of a global optimizer of SP.

An iterative procedure can now be used to solve the geometric mean approximation of a Reversed GP. Start with any feasible $\mathbf{x}^k$ and compute $\boldsymbol{\alpha}(\mathbf{x}^k)$. Then solve the resulting standard form GP to obtain $\mathbf{x}^{k+1}$. If it is feasible in the original constraints of Reversed GP and makes the approximation tight for (2.13,2.14,2.15,2.16), then stop. Otherwise compute $\boldsymbol{\alpha}(\mathbf{x}^{k+1})$ and repeat the iterations of solving a GP based on the geometric mean approximation using $\boldsymbol{\alpha}(\mathbf{x}^{k+1})$.

Approach 4: Complementary GP. The fourth approach to solve SP is similar to the third approach. It first converts a SP into a Complementary GP, which allows upper bound constraints on the ratio between two posynomials, and then applies a monomial approximation iteratively [10, 51]. This is called the condensation technique, which is an instance of the cutting-plane method for nonlinear programming.

The conversion from a SP into 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

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

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 again be carried out: given a feasible $\mathbf{x}^k$, from which a monomial approximations using $\boldsymbol{\alpha}(\mathbf{x}^k)$ 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, but may also converge to a local optimum.

Another choice is to make the monomial approximation for both the denominator posynomial $1 + f_{i2}(\mathbf{x})$ and numerator posynomial $f_{i2}(\mathbf{x})$. That turns all the constraints into monomials, and after the log transformation, approximate SP as a LP. This is called the double condensation method, and a similar iterative procedure can be carried out as in the last paragraph. A key difference from the (single) condensation method is that this LP approximation always generate solutions that are infeasible in the original SP. Therefore 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. The solution $\mathbf{x}^*$ at which all constraints in the original SP are satisfied is an optimum of the SP.

This iterative procedure of condensation uses a sequence of GP relaxations for a wide class of nonlinear non-convex optimization problems. Compared to the recently developed algebraic sum-of-squares method using SDP relaxations [102, 101, 106] for constrained polynomial minimization (which becomes a special case of SP when the variables are restricted to the positive quadrant), this GP relaxation method still lacks a theoretical foundation that guarantees its performance.

We conclude the quick tutorial on GP extensions in this subsection with the following comment. As evidenced through the exam-

ples in Subsection 2.2.3 and this subsection, GP, with all these extensions, in fact covers a surprisingly wide range of nonlinear (convex or non-convex) problems that may not even resemble a GP in standard form (2.1) or in convex form (2.3). Indeed, it is known [103] that any optimization where the feasible set is the intersection of the domain of objective and constraint functions and a cone can be turned into a (most generalized version of) GP. Different cones lead to different forms of GP in this most generalized sense, and different algebraic descriptions of the cones lead to different separability structures of the resulting GP.

2.3 Algorithms

2.3.1 Numerical methods for GP

During the 1960s and 1970s, a variety of numerical methods were proposed for GP, ranging from the original one by Duffin, Peterson, and Zener to ellipsoid methods. Some of them are based on primal GP while others on dual GP, some start with standard form while others convex form. Lists of representative GP problems for comparison of numerical methods and performance evaluation of some solution packages were published as well [6, 47].

A measure of how difficult is a GP was proposed in [52]. The degree of difficulty of a GP is the difference between the total number of monomial terms in the objective and constraints, and one plus the number of variables. When the degree of difficulty is zero, solving GP is equivalent to solving a system of linear equations. Modern numerical methods seem to perform well independent of the degree of difficulty.

There are at least two major approaches to solve a GP using modern convex optimization techniques. One is the interior-point method as in [97], and the other is an infeasible algorithm as in [78]. User-friendly softwares for GP are available on the Internet, such as the MOSEK package [129].

The standard barrier-based interior-point method for convex optimization can be applied to GP in a straightforward way, with a worst-case polynomial-time complexity and very efficient performance that

scales gracefully with problem size in practice. Consider the following GP in convex form with m inequality constraints:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) = \log \sum_{k=1}^{K_0} \exp(\mathbf{a}_{0k}^T \mathbf{x} + b_{0k}) \\ \text{subject to} & f_i(\mathbf{x}) = \log \sum_{k=1}^{K_i} \exp(\mathbf{a}_{ik}^T \mathbf{x} + b_{ik}) \leq 0, \quad i = 1, \ldots, m \\ \text{variables} & \mathbf{x}. \end{array}$$

The basic idea of the barrier-method is to solve a sequence of unconstrained problems that absorb the constraints into a new objective function, which is a weighted sum of the original objective function and a barrier function ϕ of the constraints. As the weight t on the original objective function becomes larger, the unconstrained problem becomes a tighter approximation of the original problem.

Barrier-method algorithm for GP [21]:
Given a strictly feasible point $\mathbf{x}$, which can be obtained either by verifying a given $\mathbf{x}$ to be strictly feasible or by solving a feasibility GP problem, and $t := t^{(0)} > 0, \mu > 1$, and error tolerance $\epsilon > 0$.
Repeat

(1) Centering step: compute $\mathbf{x}^*(t)$ by minimizing $tf_0(\mathbf{x}) + \phi(\mathbf{x})$ starting at $\mathbf{x}$. This is an unconstrained, smooth, convex minimization that can be readily carried out by a variety of iterative methods, such as gradient descent method or Newton's method.
(2) Update: $\mathbf{x} := \mathbf{x}^*(t)$.
(3) Stopping criterion: Quit if $\frac{m}{t} \leq \epsilon$.
(4) Increase t: $t := \mu t$.

If a log barrier function ϕ is used in item (1), we have

$$\phi(\mathbf{x}) = -\sum_{i=1}^{m} \log \left(-\log \sum_{k=1}^{K_i} \exp(\mathbf{a}_{ik}^T \mathbf{x} + b_{ik}) \right).$$

A concise discussion on computational complexity and parameter choices for the above algorithm can be found in [21].

We now turn to a more complicated infeasible algorithm following Kortanek, Xu, and Ye [78], which solves both the primal and dual GP simultaneously, starting with convex form, and is reported

to produce very competitive numerical efficiency for a wide range of GPs. The basic technique is to apply Newton's method to the perturbed Karush–Kuhn–Tucker (KKT) system with the help of predictor-corrector, coupled with effective techniques for choosing iterate directions and step lengths. Special structures of the Hessian of convex form GP is utilized in sparse matrix factorizations to accelerate the computation.

The infeasible primal-dual method generates subfeasible solutions whose primal and dual objective function values converge to the respective primal and dual optimal values. It is applied to the dual GP and its Wolfe dual problem. From Subsection 2.1.2, we know that the dual of a GP with n variables, $(m-1)$ monomial terms, and p inequality posynomial constraints, can be written as the following linearly constrained convex optimization:

$$\begin{array}{ll} \mathsf{minimize} & f(\mathbf{x}) \\ \mathsf{subject\ to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \succeq 0 \\ \mathsf{variables} & \mathbf{x} \end{array}$$

where constant matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$ and constant vector $\mathbf{b} \in \mathbf{R}^m$.[11] The classical Wolfe dual problem to the above dual GP can be written as:

$$\begin{array}{ll} \mathsf{maximize} & \mathbf{b}^T\mathbf{y} - \mathbf{x}^T\nabla f(\mathbf{x}) + f(\mathbf{x}) \\ \mathsf{subject\ to} & \mathbf{A}^T\mathbf{y} - \nabla f(\mathbf{x}) + \mathbf{z} = 0 \end{array}$$

where $\mathbf{x} \in \mathbf{R}^n_{++}$ are the primal variable, and $(\mathbf{y} \in \mathbf{R}^m, \mathbf{z} \in \mathbf{R}^n_+)$ are the dual variables.

The primal feasibility, dual feasibility, and complementarity feasibility residuals are respectively defined by:

$$\begin{aligned} \mathbf{r}_P(\mathbf{x}) &= \mathbf{b} - \mathbf{A}\mathbf{x} \\ \mathbf{r}_D(\mathbf{x}, \mathbf{y}, \mathbf{z}) &= \nabla f(\mathbf{x}) - \mathbf{A}^T\mathbf{y} - \mathbf{z} \\ \mu(\mathbf{x}, \mathbf{z}) &= \frac{\mathbf{x}^T\mathbf{z}}{n}. \end{aligned}$$

[11] Note that $\mathbf{A}$ in this subsection does not denote the exponent constant matrix used elsewhere in the survey. It represents the linear constraints in dual GP. The constant vector $\mathbf{b}$ is simply an $n+1$-dimensional vector with the first entry being 1 and the rest 0.

The duality gap is $\mathbf{x}^T \nabla f(\mathbf{x}) - \mathbf{b}^T \mathbf{y}$. Let $\mathbf{X} = \mathbf{diag}(\mathbf{x})$ and $\mathbf{Z} = \mathbf{diag}(\mathbf{z})$.

Infeasible primal-dual algorithm for GP and dual GP [78]: Given the dual objective function $f(\mathbf{x})$ and constants $(\mathbf{A}, \mathbf{b})$. Initialize $\mathbf{y}^0, \mathbf{x}^0 \succeq 0, \mathbf{z}^0 \succeq 0$.

(1) Compute the Hessian $\mathbf{H} = \nabla^2 f(\mathbf{x})$.
(2) Call **Prediction Step Routine** to compute $\gamma \in [0, 1]$ and $\eta \in [0, 1]$.
(3) Solve the following system of linear equation for $(\boldsymbol{\delta}_y, \boldsymbol{\delta}_x, \boldsymbol{\delta}_z)$:

$$\begin{pmatrix} 0 & \mathbf{A} & 0 \\ \mathbf{A}^T & -\mathbf{H} & \mathbf{I} \\ 0 & \mathbf{Z} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_y \\ \boldsymbol{\delta}_x \\ \boldsymbol{\delta}_z \end{pmatrix} = \begin{pmatrix} \eta \mathbf{r}_P \\ \eta \mathbf{r}_D \\ \gamma\mu \mathbf{1} - \mathbf{Xz} \end{pmatrix}. \quad (2.17)$$

(4) Call **Step Length Routine** to compute α.
(5) Update primal-dual solution:

$$\begin{aligned} \mathbf{y} &= \mathbf{y} + \alpha \boldsymbol{\delta}_y \\ \mathbf{x} &= \mathbf{x} + \alpha \boldsymbol{\delta}_x \\ \mathbf{z} &= \mathbf{z} + \alpha \boldsymbol{\delta}_z. \end{aligned}$$

(6) Call **Dual Slack Reset Routine** to reset $\mathbf{z}$.
(7) Call **Stopping Criterion Routine** to determine if the iterations can be stopped.

Prediction Step Routine

- Let $\eta = 1$ and $\gamma = 0$. Solve (2.17).
- Compute $\alpha = \min\left(\max_j \{-x_j/\delta_{x,j} : \delta_{x,j} < 0\}, (\max_j \{-z_j/\delta_{z,j} : \delta_{z,j} < 0\}\right)$.
- Compute $\gamma = \frac{1}{2\mu n}(\mathbf{x} + \alpha\boldsymbol{\delta}_x)^T(\mathbf{z} + \alpha\boldsymbol{\delta}_z)$ and $\eta = 1 - \gamma$.

Step Length Routine

- Compute $\alpha_R = \min\left(\max_j \{-x_j/\delta_{x,j} : \delta_{x,j} < 0\}, (\max_j \{-z_j/\delta_{z,j} : \delta_{z,j} < 0\}\right)$.

- Choose α_N such that $(\mathbf{x}(\alpha_N), \mathbf{y}(\alpha_N), \mathbf{z}(\alpha_N))$ (updated according to Step (5) in the algorithm) satisfy:

$$\begin{aligned} \|\mathbf{Xz}\| &\geq \sigma\mu \\ \|\mathbf{Xz} - \mu\mathbf{1}\| &\leq \beta\mu \end{aligned}$$

  where $\sigma \in (0,1)$ and $\beta > 0$ are constant parameters.
- Choose α_C such that $(\mathbf{x}(\alpha_C), \mathbf{y}(\alpha_C), \mathbf{z}(\alpha_C))$ (updated according to Step (5) in the algorithm) satisfy:

$$\begin{aligned} \theta_P \mathbf{x}(\alpha_C)^T \mathbf{z}(\alpha_C) &\geq \|\mathbf{r}_P\| \\ \theta_D \mathbf{x}(\alpha_C)^T \mathbf{z}(\alpha_C) &\geq \|\mathbf{r}_D\| \\ \mathbf{x}(\alpha_C)^T \mathbf{z}(\alpha_C) &\leq \theta_C \mathbf{x}^T \mathbf{z} \end{aligned}$$

  where $\theta_P, \theta_D, \theta_C > 0$ are constant parameters.
- Choose the step length as:

$$\alpha = \min\{\theta_1 \alpha_R, \alpha_N, \alpha_C\}$$

  where $\theta_1 \in (0,1)$ is a safety factor.

Dual Slack Reset Routine

- Let $\boldsymbol{\sigma} = \nabla f(\mathbf{x}) - \mathbf{A}^T\mathbf{y}$.
- For each i, if $\sigma_i \geq 0$, then

$$z_i = \begin{cases} \sigma_i, & \sigma_i \in (z_i/\theta_2, z_i\theta_2) \\ z_i/\theta_2, & \sigma_i \leq z_i/\theta_2 \\ z_i\theta_2, & \sigma_i \geq z_i\theta_2 \end{cases}$$

  where $\theta_2 > 0$ is a constant parameter.

Stopping Criterion Routine

If the following inequalities are satisfied, then stop the algorithm, otherwise, return to Step (1).

$$\begin{aligned} \frac{\|\mathbf{r}_P\|_1}{1 + \|\mathbf{x}\|_1} &< \epsilon_P \\ \frac{\|\mathbf{r}_D\|_1}{1 + \|\mathbf{z}\|_1} &< \epsilon_D \\ \frac{\mathbf{x}^T\mathbf{z}}{1 + \|\mathbf{x}\|_1 + \|\mathbf{z}\|_1} &< \epsilon_C \end{aligned}$$

where $\epsilon_P, \epsilon_D, \epsilon_C$ are constant parameters.

The following values for constant parameters are recommended in [78]:

- $\mathbf{x}^0 = \mathbf{1}, \mathbf{y}^0 = 0, \mathbf{z}^0 = \mathbf{1}$ as initial points.
- $\sigma = 10^{-8}, \beta = 10^3, \theta_P = 100\frac{\|\mathbf{r}_P^0\|}{\mathbf{x}^{0T}\mathbf{z}^0}, \theta_D = 10^4\frac{\|\mathbf{r}_P^0\|}{\mathbf{x}^{0T}\mathbf{z}^0}, \theta_C = 2$ and $\theta_1 = 0.9995$ in Step Length Routine.
- $\theta_2 = 100$ in Dual Slack Reset Routine.
- $\epsilon_P = 10^{-8}, \epsilon_D = 10^{-8}, \epsilon_C = 10^{-12}$ in Stopping Criterion Routine.

The most computationally intensive step in the algorithm is Step 3 that solves a KKT system of linear equations, which can be simplified as:

$$\begin{pmatrix} -\mathbf{X}^{-1}\mathbf{Z} - \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_x \\ \boldsymbol{\delta}_y \end{pmatrix} = \begin{pmatrix} \eta\mathbf{r}_D - \mathbf{X}^{-1}(\gamma\mu\mathbf{1} - \mathbf{X}\mathbf{z}) \\ \eta\mathbf{r}_P \end{pmatrix}.$$

Solving this linear system of equations can be accomplished by computing matrix $\mathbf{K} = \mathbf{A}(\mathbf{X}^{-1}\mathbf{Z} + \mathbf{H})\mathbf{A}^T$. Because the Hessian $\mathbf{H}$ of dual GP is a block diagonal matrix with sparse blocks, and $\mathbf{X}^{-1}\mathbf{Z} + \mathbf{H}$ has the same block diagonal structure with blocks $\bar{\mathbf{H}}_k$, matrix $\mathbf{K}$ can be written as

$$\mathbf{K} = \sum_{j=0}^{p} \mathbf{A}_j \bar{\mathbf{H}}_j \mathbf{A}_j^T.$$

This decomposition greatly simplifies the numerical solution in Step (3) of the algorithm.

As reported in [78], this infeasible primal-dual algorithm is tested on 19 typical GP problems, including 3 that are generally viewed as the most difficult, and the computational time is orders of magnitude faster than the earlier methods.

2.3.2 Numerical methods for robust GP

In the area of robust optimization [11, 13, 63, 62], we solve an optimization problem by taking into account possible perturbations of the problem parameters. A series of results on robust conic optimization, especially robust SDP and robust SOCP, have been obtained over the last decade in addition to the classical robust LP results.

Recall that the constant parameters of a GP can be written as a set of exponent constant matrices $\{\mathbf{A}_i\}$ and a set of multiplicative constant vectors $\{\mathbf{b}_i\}$, $i = 0, 1, \ldots, m$, where $\mathbf{A}_0$ and $\mathbf{b}_0$ correspond to the objective posynomial, and the other $\mathbf{A}_i \in \mathbf{R}^{K_i \times n}$ and $\mathbf{b}_i \in \mathbf{R}^{K_i}$ correspond to the m inequality constraints. In the worst-case robust GP formulation, each log-sum-exp function in the original GP need to be replaced by the supremum of the set of log-sum-exp functions whose parameters $\tilde{\mathbf{A}}_i$ and $\tilde{\mathbf{b}}_i$ belong to the image of a set $\mathcal{U}$ in $\mathbf{R}^L$ under an affine mapping:

$$(\tilde{\mathbf{A}}_i, \tilde{\mathbf{b}}_i) = \left(\mathbf{A}_i^0 + \sum_{j=1}^{L} u_j \mathbf{A}_i^j, \mathbf{b}_i^0 + \sum_{j=1}^{L} u_j \mathbf{b}_i^j \right), \quad \forall \mathbf{u} \in \mathcal{U}$$

where $\mathbf{A}_i^j, \mathbf{b}_i^j,\ j = 0, 1, \ldots, L$ are given matrices and vectors describing the uncertainty.

In some cases, a robust GP can be formulated as a GP [11]. A method to approximately solve a general robust GP has recently been proposed in [67]. This numerical method for robust GP is based on the idea of approximating a posynomial with a piecewise linear function.

First a general robust GP is reduced to a two-term robust GP, where each posynomial has only two monomial terms. This reduction can be conducted in a numerically efficient way and decreases the computational load of approximating a general multi-variate function. After the log transformation, a two-term posynomial is turned into a two-term log-sum-exp function:

$$\log(\exp(y_1) + \exp(y_2)),$$

which is then approximated by a r-term piecewise linear convex function:

$$\max_{i=1,2,\ldots,r} \{\mathbf{c}_i^T \mathbf{y} + d_i\}$$

where $\mathbf{c}_i, d_i$ are constants. Replacing the log-sum-exp functions with the best piecewise linear approximations, a robust GP is turned into a robust LP formulation.

When the set of uncertainty is polyhedral:

$$\mathcal{U} = \{\mathbf{u} \in \mathbf{R}^L | \mathbf{D}\mathbf{u} \preceq \mathbf{d}\}$$

where $(\mathbf{D}, \mathbf{d})$ describe the finite number of hyperplanes that define the uncertainly set, or ellipsoidal:

$$\mathcal{U} = \{\bar{\mathbf{u}} + \mathbf{D}\rho | \rho \in \mathbf{R}^L, \|\rho\|_2 \leq 1\}$$

where $\mathbf{D}$ describes the possible variations of GP parameters, the resulting robust LP becomes another LP or a Second Order Cone Program, respectively. Thus efficiently solvable as a standard convex optimization problem.

As the piecewise linear approximation becomes tighter, the approximate robust GP (in the form of a robust LP) approaches the exact formulation of the original robust GP. However, to maintain a given level of approximation accuracy, the size of robust LP that approximates the robust GP grows exponentially with the number of monomial terms.

2.3.3 Distributed algorithms

We present a systematic theory of distributed algorithms for GP, which is particularly useful for networking applications. While efficient and robust algorithms have been extensively studied for centralized solution of GP, distributed solutions for GP have not been fully explored before. This subsection shows how special structures of GP can be utilized for distributed computation of a globally optimal solution.

Based on the sparsity pattern of the exponent matrix $\mathbf{A}$, it is sometimes natural to decompose a GP into several small decoupled GPs. This is a straightforward application of the standard decomposition method for convex optimization with decoupled constraints (e.g., [18, 103]), and will not be further discussed here.

Special cases. We first present three special cases of GP where simple distributed algorithms have been found, by dual decomposition, or by linear system evolution using the Perron–Frobenius theory of positive matrix, or by message-passing-based iteration of gradient algorithms. These special cases include linearly constrained maximization of a monomial, unconstrained minimization of a product of posynomials, and feasibility problem with a special structure of the exponent constant matrix $\mathbf{A}$ and multiplicative constant vector $\mathbf{d}$. These cases are motivated by networking problems to be covered in Section 3.

Case 1. Consider the following linearly constrained monomial maximization:

$$\begin{array}{ll} \mathsf{maximize} & \prod_j x_j^{a_j} \\ \mathsf{subject\ to} & \sum_j R_{ij} x_j^{b_j} \leq 1, \ \ i = 1, \ldots, m \\ \mathsf{variables} & \mathbf{x} \end{array}$$

where $\{R_{ij}\}, \{a_j\}$ are non-negative constants, and $\{b_j\}$ are real constants. It is important to note that in this case the exponent constants $\{b_j\}$ do not depend on the constraint index i. A distributed algorithm for this class of GP is described in Subsection 3.4.1 in the application of TCP Vegas congestion control.

Case 2. Consider the following GP feasibility problem: find an $\mathbf{x}$ such that the following posynomial inequalities are satisfied:

$$\sum_{j \neq i} A_{ij} x_j x_i^{-1} \leq \rho, \ \ i = 1, \ldots, m$$

where A_{ij} are positive constants. This is a model of a wireless power control problem. Define a matrix $\mathbf{A}$ where the off-diagonal entries are A_{ij} and diagonal entries zero. It is known [60] that if $1/\rho$ is smaller than the Perron–Frobenius eigenvalue $\lambda_{max}(\mathbf{A})$ of $\mathbf{A}$, the feasibility problem has a solution, and the following simple, iterative update of $\mathbf{x}$ converges geometrically to a feasible solution over iterations indexed by t:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t).$$

This update can be carried out distributively. Each x_i is updated to make the ith posynomial constraint be satisfied with equality, assuming that the other $\{x_j, \ j \neq i\}$ do not change.

Case 3. Consider the following unconstrained minimization of the composition of a monomial and a posynomial of $\mathbf{x}$:

$$\mathsf{minimize} \prod_i \left(\sum_{j \neq i} A_{ij} x_j x_i^{-1} \right)^{a_i}$$

where A_{ij}, a_i are positive constants. In an application to jointly optimal congestion and power control in Subsection 3.4.2, a gradient method to solve this GP will be turned into a distributed algorithm with the

help of message passing among the network elements each controlling an x_i.

General approach. Based on recent results in [37], we show that, by allowing message passing, a dual decomposition technique can be used to distributively solve any standard form GP. Here, each term or each constraint equation has a corresponding interpretation of a network element (end user or intermediate nodes). This method can be used to distributively solve all the GP problems formulated in Section 3 for various applications in information theory, coding, signal processing, and networking, especially for power control problems. It can also be used to decouple any convex optimization with additive objective and constraint functions:

$$\begin{array}{ll} \mathsf{minimize} & \sum_s f_s(\mathbf{x}) \\ \mathsf{subject\ to} & \sum_j h_{ij}(\mathbf{x}) \leq 0, \quad \forall i \end{array}$$

where f_s and h_{ij} are all convex functions.

Objective and constraint functions in GP are additive, with coupling of variables across the monomial terms. Had there been no coupling, the additive structure leads to an easy parallel computation. The key approach to tackle the coupling problem is to introduce auxiliary variables and additional equality constraints, thus transferring the coupling in the objective function to coupling in the constraints, which can be decoupled by dual decomposition and solved by introducing 'consistency pricing'. Updates of 'consistency price' can be conducted via local communication channels among the variables that are coupled with each other. This method is illustrated through an unconstrained GP, and extensions to problems with constraints are straightforward.

Suppose we have the following unconstrained standard form GP in $\mathbf{x} \succ 0$:

$$\mathsf{minimize} \quad \sum_i f_i(x_i, \{x_j\}_{j \in I(i)}) \tag{2.18}$$

where x_i denotes the local variable of the ith user, $I(i)$ denotes the set of coupled variables from other users, and $\{f_i\}$ are posynomials. Making a change of variable $y_i = \log x_i, \forall i$, in the original problem, we obtain the following convex optimization problem in $\mathbf{y}$:

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

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

$$\begin{array}{ll} \mathsf{minimize} & \sum_i f_i(e^{y_i}, \{e^{y_{ij}}\}_{j\in I(i)}) \\ \mathsf{subject\ to} & y_{ij} = y_j, \ \ \forall j \in I(i), \forall i \\ \mathsf{variables} & \{y_i\}, \{y_{ij}\}. \end{array} \tag{2.19}$$

Each ith user controls the local variables $(y_i, \{y_{ij}\}_{j\in I(i)})$. Next, the Lagrangian of (2.19) 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 each partial Lagrangian term is

$$L_i(y_i, \{y_{ij}\}, \{\gamma_{ij}\}) = f_i(e^{y_i}, \{e^{y_{ij}}\}_{j\in I(i)}) + \Big(\sum_{j:i\in I(j)} \gamma_{ji}\Big) y_i - \sum_{j\in I(i)} \gamma_{ij} y_{ij}. \tag{2.20}$$

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 (2.18) is constrained, the additional constraints can be included in the minimization of each L_i.

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

$$\mathsf{maximize}_{\{\gamma_{ij}\}} \quad g(\{\gamma_{ij}\}) \tag{2.21}$$

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 (2.19) is convex with zero duality gap. Hence the Lagrange dual problem indeed solves the original standard GP problem. A simple way to solve the maximization in (2.21) is with the following update for the consistency prices:

$$\gamma_{ij}(t+1) = \gamma_{ij}(t) + \alpha(t)(y_j(t) - y_{ij}(t)). \tag{2.22}$$

Appropriate choice of the stepsize $\alpha(t) > 0$ leads to convergence of the dual algorithm [16].

Summarizing, we have the following

Distributed Algorithm for Unconstrained GP [37]:

The ith user does the following:

(1) Minimize L_i in (2.20) 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).

(2) Update the local consistency prices $\{\gamma_{ij}, j \in I(i)\}$ with (2.22).

The amount of message passing overhead in the above distributed algorithm can be substantially reduced using the structures of the coupling variables. The general approach of distributed GP solution and specific overhead reduction techniques will be illustrated in Subsection 3.3.1 through distributed GP-based power control.

3

Applications in Communication Systems

3.1 Information Theory

Materials in this subsection are in part based on recent publications [30, 33, 82, 84], and some problems have been studied in the 1980s and 1990s [120, 121, 122].

We are concerned with two information theoretic limits on data transmission and compression in this section, focusing on the single-user, discrete memoryless system models: channel capacity as the maximum transmission rate so that the decoding error probability vanishes as the codeword becomes long, and rate distortion function as the minimum rate required to describe a source so that the decoder's average distortion is no larger than a threshold distortion value.

We will show that the Lagrange dual problems of channel capacity and rate distortion can be simplified into GPs.[1] The structures of these GPs allow us to efficiently generate upper bounds on channel capacity and lower bounds on rate distortion by solving

[1] The Lagrange dual problem of a given primal problem may be represented in several different but equivalent ways. It is important to reduce it the most illuminating and useful form.

systems of linear inequalities (Subsections 3.1.1 and 3.1.2), to characterize Shannon duality between transmission and compression (Subsection 3.1.3), and to interpret channel capacity and rate distortion problems as free energy optimization in statistical physics (Subsection 4.1.1).

3.1.1 Channel capacity

First consider the problem of data transmission over a discrete memoryless channel with input $X \in \mathcal{X} = \{1, 2, \ldots, N\}$, output $Y \in \mathcal{Y} = \{1, 2, \ldots, M\}$, and channel law $P_{ij} = \mathbf{Prob}\{Y = j | X = i\}$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, M$. The channel law forms a channel matrix $\mathbf{P} \in \mathbf{R}^{N \times M}$, where the (i, j) entry of $\mathbf{P}$ is $P_{ij} \geq 0$ with $\mathbf{P1} = \mathbf{1}$. A distribution $\mathbf{p} \in \mathbf{R}^{N \times 1}$ on the input, together with a given channel matrix $\mathbf{P}$, induces a distribution $\mathbf{q} \in \mathbf{R}^{M \times 1}$ on the output by $\mathbf{q} = \mathbf{P}^T\mathbf{p}$, and a joint distribution $\mathbf{Q} \in \mathbf{R}^{N \times M}$ on the input output pair by $Q_{ij} = p_i P_{ij}$. We also associate with each input alphabet symbol i an input cost $s_i \geq 0$, forming a cost vector $\mathbf{s} \in \mathbf{R}^{N \times 1}$.

It is a key result in information theory (e.g., [40, 94]) that the capacity $C(S)$ of a discrete memoryless channel, under the input cost constraint $\mathbf{E_p}[\mathbf{s}] = \mathbf{p}^T\mathbf{s} \leq S$ for a given total cost $S \geq 0$, is

$$C(S) = \max_{\mathbf{p}: \mathbf{p}^T\mathbf{s} \leq S} I(X; Y) \tag{3.1}$$

where the mutual information I between input X and output Y is defined as

$$I(X; Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} Q_{ij} \log \frac{Q_{ij}}{p_i q_j} = H(Y) - H(Y|X) = -\sum_{j=1}^{M} q_j \log q_j - \mathbf{p}^T\mathbf{r}$$

where $\mathbf{r} \in \mathbf{R}^{N \times 1}$ and $r_i = -\sum_{j=1}^{M} P_{ij} \log P_{ij}$ is the conditional entropy of Y given $X = i$.

Therefore we view channel capacity as the optimal objective value of the following maximization problem, referred to as the channel capacity problem with input cost:[2]

$$\begin{array}{ll} \mathsf{maximize} & -\mathbf{p}^T\mathbf{r} - \sum_{j=1}^M q_j \log q_j \\ \mathsf{subject\ to} & \mathbf{P}^T\mathbf{p} = \mathbf{q} \\ & \mathbf{p}^T\mathbf{s} \leq S \\ & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}, \mathbf{q}. \end{array} \tag{3.2}$$

The constant parameters are $\mathbf{P}$, the channel matrix, and $r_i = -\sum_{j=1}^M P_{ij} \log P_{ij},\ \forall i$. In the special case of no input cost constraint, the channel capacity problem becomes:

$$\begin{array}{ll} \mathsf{maximize} & -\mathbf{p}^T\mathbf{r} - \sum_{j=1}^M q_j \log q_j \\ \mathsf{subject\ to} & \mathbf{P}^T\mathbf{p} = \mathbf{q} \\ & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}, \mathbf{q}. \end{array} \tag{3.3}$$

If we substituted $\mathbf{q} = \mathbf{P}^T\mathbf{p}$ in the objective function of (3.2), we would later find that the Lagrange dual problem can only be implicitly expressed through the solution of a system of linear equations. Keeping two sets of optimization variables, $\mathbf{p}$ and $\mathbf{q}$, and introducing the equality constraint $\mathbf{P}^T\mathbf{p} = \mathbf{q}$ in the primal problem is a key step to derive an explicit and simple Lagrange dual problem of the channel capacity problem.

Theorem 3.1. The Lagrange dual of the channel capacity problem with input cost (3.2) is the following GP (in convex form):

$$\begin{array}{ll} \mathsf{minimize} & \log \sum_{j=1}^M \exp(\alpha_j + \gamma S) \\ \mathsf{subject\ to} & \mathbf{P}\boldsymbol{\alpha} + \gamma\mathbf{s} \succeq -\mathbf{r} \\ & \gamma \geq 0 \\ \mathsf{variables} & \boldsymbol{\alpha}, \gamma \end{array} \tag{3.4}$$

The constant parameters are $\mathbf{P}, \mathbf{s}$ and S.

[2] Showing that channel capacity can be obtained as a maximized mutual information requires achievability and converse proofs, which are not discussed in this subsection.

An equivalent version of the Lagrange dual problem is the following GP (in standard form):

$$\begin{array}{ll} \mathsf{minimize} & w^S \sum_{j=1}^M z_j \\ \mathsf{subject\ to} & w^{s_i} \prod_{j=1}^M z_j^{P_{ij}} \geq e^{-H(\mathbf{P}^{(i)})}, \quad i = 1, 2, \ldots, N \\ & w \geq 1, \quad z_j \geq 0, \quad j = 1, 2, \ldots, M \\ \mathsf{variables} & \mathbf{z}, w \end{array} \tag{3.5}$$

where $\mathbf{P}^{(i)}$ is the ith row of $\mathbf{P}$.

Lagrange duality between problems (3.2) and (3.4) means the following:

- *Weak duality.* Any feasible $(\boldsymbol{\alpha}, \gamma)$ of the Lagrange dual problem (3.4) produces an upper bound on channel capacity with input cost: $\log \sum_{j=1}^M \exp(\alpha_j + \gamma S) \geq C(S)$.
- *Strong duality.* The optimal value of the Lagrange dual problem (3.4) is $C(S)$.

The proof can be found in Appendix B.1.

Corollary 3.1. The Lagrange dual of the channel capacity problem without input cost (3.3) is the following GP (in convex form)

$$\begin{array}{ll} \mathsf{minimize} & \log \sum_{j=1}^M e^{\alpha_j} \\ \mathsf{subject\ to} & \mathbf{P}\boldsymbol{\alpha} \succeq -\mathbf{r} \\ \mathsf{variables} & \boldsymbol{\alpha}. \end{array} \tag{3.6}$$

The constant parameters are $\mathbf{P}$.

An equivalent version of the Lagrange dual problem is the following GP (in standard form):

$$\begin{array}{ll} \mathsf{minimize} & \sum_{j=1}^M z_j \\ \mathsf{subject\ to} & \prod_{j=1}^M z_j^{P_{ij}} \geq e^{-H(\mathbf{P}^{(i)})}, \quad i = 1, 2, \ldots, N \\ & z_j \geq 0, \quad j = 1, 2, \ldots, M \\ \mathsf{variables} & \mathbf{z}. \end{array} \tag{3.7}$$

The constant parameters are $\mathbf{P}$, and $\mathbf{P}^{(i)}$ is the ith row of $\mathbf{P}$.

Lagrange duality between problems (3.3) and (3.6) means the following:

- *Weak duality.* $\log\left(\sum_{j=1}^{M} e^{\alpha_j}\right) \geq C$, for all $\boldsymbol{\alpha}$ that satisfy $\mathbf{P}\boldsymbol{\alpha} + \mathbf{r} \succeq 0$.
- *Strong duality.* $\log\left(\sum_{j=1}^{M} e^{\alpha_j^*}\right) = C$, where $\boldsymbol{\alpha}^*$ are the optimal dual variables.

We can also prove the weak duality result in Corollary 3.1 on channel capacity upper bound in a simple way without using the machinery of Lagrange duality. This short proof is shown in Appendix B.2.

Note that the Lagrange dual (3.7) of the channel capacity problem is a simple GP with a linear objective function and only monomial inequality constraints. Also, dual problem (3.5) is a generalized version of dual problem (3.7), weighing the objective function by w^S and each constraint by w^{s_i}, where w is the Lagrange multiplier associated with the input cost constraint. If the costs for all alphabet symbols are 0, we can analytically minimize the objective function over w by simply letting $w = 0$, indeed recovering the dual problem (3.7) for channels without the input cost constraint.

We can interpret the Lagrange dual problem (3.6) as follows. Let $\Lambda : \{1, \ldots, M\} \to \mathbf{R}$ be a real-valued function on the output space, with $\Lambda(j) = \alpha_j$. We can think of the variables $\boldsymbol{\alpha}$ as parameterizing all real-valued functions on the output space, so the dual problem is one over all real-valued functions on the output space. Since $(\mathbf{P}\boldsymbol{\alpha})_i = \sum_{j=1}^{M} \alpha_j P_{ij} = \mathbf{E}(\Lambda | X = i)$, the inequality constraint in the dual states that for each i, $\mathbf{E}(\Lambda | X = i)$ exceeds $-r_i = -H(Y|X = i)$. Since $\max_j \alpha_j \leq \log\left(\sum_{j=1}^{M} e^{\alpha_j}\right) \leq \max_j \alpha_j + \log M$, the objective function in the dual is a smooth approximation of the maximum function. Thus, the Lagrange dual problem asks us to consider all real-valued functions Λ on the output space, for which $\mathbf{E}(\Lambda | X = i)$ exceeds $-H(Y|X = i)$ for each i. Among all such Λ, we are to find the one that minimizes a smoothed approximation of the maximum value of Λ.

Suppose we have solved the GP dual[3] (3.4) of channel capacity. By strong duality, we obtain $C(S)$. We can also recover the optimal

[3] GP dual refers to the Lagrange dual problem (of some primal problem) that is a GP, not to be confused with the Lagrange dual problem of GP.

primal variables, i.e., the capacity achieving input distribution, from the optimal dual variables. For example, we can recover a least-norm capacity-achieving input distribution for a channel as follows. First, the optimal output distribution $\mathbf{q}^*$ can be recovered from the optimal dual variable $\boldsymbol{\alpha}^*$:

$$q_j^* = \exp\left(\alpha_j^* - C\right), \quad j = 1, 2, \ldots, M \tag{3.8}$$

where $C = \log \sum_{j=1}^{M} e^{\alpha_j^*}$, and the optimal input distribution $\mathbf{p}^*$ is a vector that satisfies the linear equations:

$$\begin{aligned} -\mathbf{p}^T\mathbf{r} &= C + e^{-C}\left(\sum_{j=1}^{M} \alpha_j^* e^{\alpha_j^*} - C\sum_{j=1}^{M} e^{\alpha_j^*}\right), \\ \mathbf{P}^T\mathbf{p} &= \mathbf{q}^*, \\ \mathbf{1}^T\mathbf{p} &= 1. \end{aligned}$$

The primal and dual problems of $C(S)$ can be simultaneously and efficiently solved through a primal-dual interior point algorithm [97, 21], which scales smoothly for different channels and alphabet sizes and provides an alternative to the classical Blahut–Arimoto algorithm [3, 19]. Due to the structure and sparsity of the exponent constant matrix $\mathbf{A}$ of the GP dual for channel capacity, standard GP algorithms like the ones in Subsection 2.3.1 can be further accelerated.

Complementary slackness between (3.2) and (3.4) states that $p_i^* = 0$ if $r_i + (\mathbf{P}\boldsymbol{\alpha}^*)_i + \gamma^* s_i > 0$ in the Lagrange dual of channel capacity. Therefore, from the optimal dual variables $(\boldsymbol{\alpha}^*, \gamma^*)$, we immediately obtain the support of the capacity achieving input distribution as the following set:

$$\{i | r_i + (\mathbf{P}\boldsymbol{\alpha}^*)_i + \gamma^* s_i = 0\}.$$

From the primal and dual problems of channel capacity, we obtain the following optimality conditions. If there are $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ satisfying the following KKT conditions [21] for a given $\mathbf{P}$:

$$\begin{aligned} &\boldsymbol{\lambda} \succeq 0, \\ &\mathbf{r} + \mathbf{P}\boldsymbol{\alpha} \succeq 0, \end{aligned}$$

$$\frac{e^{\alpha_j}}{\sum_{j'=1}^{M} e^{\alpha_{j'}}} + \sum_{i=1}^{N} \lambda_i P_{ij} = 0, \quad j = 1, 2, \ldots, M,$$
$$\lambda_i (r_i + \sum_{j=1}^{M} P_{ij} \alpha_j) = 0, \quad i = 1, 2, \ldots, N,$$

then the resulting $\log \sum_{j=1}^{M} e^{\alpha_j}$ is the channel capacity C.

Because the inequality constraints in the dual problem (3.4) are affine, it is easy to find a dual feasible $\boldsymbol{\alpha}$, and the resulting value of the dual objective function provides an easily-derived upper bound on channel capacity. Based on the sparsity pattern of a given channel matrix, tight analytic bounds may also be obtained from an appropriate selection of dual variables.

As a simple example, it is easy to verify that $\alpha_j = \log \max_i P_{ij}, \ \forall j$, satisfy the dual constraints and give the following

---

Corollary 3.2. Channel capacity is upper bounded in terms of a maximum likelihood receiver selecting $\text{argmax}_i P_{ij}$ for each output alphabet symbol j:

$$C \leq \log \sum_{j=1}^{M} \max_i P_{ij}, \tag{3.9}$$

which is tight if and only if the optimal output distribution q^* is

$$q_j^* = \frac{\max_i P_{ij}}{\sum_{k=1}^{M} \max_i P_{ik}}, \quad j = 1, 2, \ldots, M.$$

When there is an input cost constraint $\mathbf{p}^T \mathbf{s} \leq S$, the above upper bound becomes

$$C(S) \leq \log \sum_{j=1}^{M} \max_i (e^{-s_i} P_{ij}) + S \tag{3.10}$$

where each maximum likelihood decision is modified by the cost vector $\mathbf{s}$.

Of course, it is trivial to find a primal feasible point satisfying the linear inequality constraints of the primal problem (3.3), which gives a lower bound on channel capacity. This pair of bounds provides an estimate of $C(S)$. Therefore, given a dual feasible variable, by generating

the corresponding primal feasible variable, both an upper and a lower bound on channel capacity are obtained. The worst case difference between the true value of channel capacity and the estimate based on either bound is the gap between these two bounds. This is a computationally easy method to generate an estimate of channel capacity with bounded error. For example, for channel capacity without input cost, find any dual feasible $\boldsymbol{\alpha}$, from which we generate

$$q_j = \frac{e^{\alpha_j}}{\sum_{k=1}^{M} e^{\alpha_k}}.$$

If there is a $\mathbf{p}$ such that $\mathbf{P}^T\mathbf{p} = \mathbf{q}$, then the estimated channel capacity $\tilde{C} = \log \sum_{j=1}^{M} e^{\alpha_j}$ can only be Γ away from the true capacity C, where

$$\Gamma = \mathbf{p}^T\mathbf{r} + \frac{\sum_{j=1}^{M} \alpha_j e^{\alpha_j}}{\sum_{j=1}^{M} e^{\alpha_j}}.$$

There is a minmax Kullback–Leibler divergence (minmaxKL) characterization of discrete memoryless channel capacity with input cost in [42]:

$$C(S) = \min_{\mathbf{q}} \min_{\gamma \geq 0} \max_{i} \left[D(\mathbf{P}^{(i)} \| \mathbf{q}) + \gamma(S - s_i) \right] \tag{3.11}$$

where the minimization over $\mathbf{q}$ is over all possible output distributions.

This characterization of C obviously leads to the following known class of upper bounds on channel capacity: for any output distribution $\mathbf{q}$,

$$C \leq \max_{i} \sum_{j=1}^{M} P_{ij} \log \frac{P_{ij}}{q_j}, \tag{3.12}$$

which is shown in [42, 61], and has recently been used for simulating finite state channels in [124] and bounding the capacity of non-coherent multiple antenna fading channels in [83].

Since the Lagrange dual (3.4) and minmaxKL (3.11) characterizations both give $C(S)$, they must be equivalent. This equivalence can also be established directly. Let the dual variables $z_j = e^{\alpha_j} = \beta q_j$ where $\beta > 0$ and $\mathbf{q}$ is any distribution. Then the dual constraints become $\sum_{j=1}^{M} P_{ij} \log \frac{P_{ij}}{q_j} - \gamma s_i \leq \log \beta, \;\; i = 1, 2, \ldots, N$. Since the case of $C(S) = 0$ is trivial, assume $C(S) > 0$. By complementary slackness, if at optimality all the dual constraints are satisfied with strict inequalities, then

the optimal Lagrange multipliers (readily seen to be the optimal input distribution) of this GP must all be zero, contradicting our assumption that $C(S) > 0$. Therefore, $\max_i \left[\sum_{j=1}^M P_{ij} \log \frac{P_{ij}}{q_j^*} - \gamma^* s_i\right] = \log \beta^*$. By the strong duality part of Theorem 3.1,

$$\begin{aligned} C(S) &= \log \sum_{j=1}^M \beta^* q_j^* + \gamma^* S = \log \beta^* + \gamma^* S \\ &= \max_i \left(\sum_{j=1}^M P_{ij} \log \frac{P_{ij}}{q_j^*} - \gamma^* s_i \right) + \gamma^* S. \end{aligned}$$

Since at optimality, $\mathbf{q}^*$ must correspond to the output distribution induced by an optimal input distribution, restricting the minimization of dual variables $\mathbf{z}$ to a scaled version of an output distribution incurs no loss of generality. Thus the minmaxKL characterization (3.11) is recovered.

The above argument shows that the GP Lagrange dual (3.4) generates a broader class of upper bounds on $C(S)$, including the class of bounds from (3.12) as a special case. Specifically, the following bounds, readily extended from (3.12) and parameterized by output distributions $\mathbf{q}$ and $\gamma \geq 0$:

$$C(S) \leq \max_i \left[\sum_{j=1}^M P_{ij} \log \frac{P_{ij}}{q_j} - \gamma s_i \right] + \gamma S,$$

can be obtained from the GP dual by restricting the dual variables $(\mathbf{z}, \gamma)$ to be such that $\max_i \left[\sum_{j=1}^M P_{ij} \log \frac{P_{ij}}{z_j} - \gamma s_i\right] = 0$, and by restricting $\mathbf{z}$ to be a scaled output distribution.

For a memoryless channel with continuous alphabets, where the channel is a family of conditional distributions $P(y|x)$ and input cost constraint is $\int p(x)s(x)dx \leq S$, a derivation similar to the discrete case shows that the Lagrange dual of the channel capacity problem is the following continuous analog of GP:

$$\begin{array}{ll} \text{minimize} & \log \int z(y)dy + \gamma S \\ \text{subject to} & \int P(y|x) \log \frac{z(y)}{P(y|x)} dy + \gamma s(x) \geq 0, \quad \forall x \\ & z(y) \geq 0, \forall y, \quad \gamma \geq 0 \\ \text{variables} & z(y), \gamma. \end{array} \tag{3.13}$$

Although, in general, the KKT optimality condition can be complicated in the infinite dimensional case, weak duality and the Lagrange dual problem (3.13) readily lead to the following class of bounds [83, 33]: for any distribution $q(y)$ and $\gamma \geq 0$,

$$C(S) \leq \max_x \left[\int P(y|x) \log \frac{P(y|x)}{q(y)} dy - \gamma s(x) \right] + \gamma S.$$

3.1.2 Rate distortion

Consider the following information theoretic problem of data compression. Assume a source that produces a sequence of i.i.d. random variables $X_1, X_2, \ldots, X_n \sim \mathbf{p}$, where the state space of X_i is a discrete source alphabet $\mathcal{X}$ with N alphabet symbols and $\mathbf{p} \in \mathbf{R}^{N \times 1}$ is the source distribution. The encoder describes the source sequence X^n by an index $f_n(x^n) \in \{1, 2, \ldots, 2^{nR}\}$, where x^n is a realization of X^n. The decoder reconstructs X^n by an estimate $\hat{X}^n = g_n(f_n(X^n))$ in a finite reconstruction alphabet $\hat{\mathcal{X}}$. Given a bounded distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbf{R}_+$, the distortion $d(x^n, \hat{x}^n)$ between sequences x^n and $\hat{x}^n$ is the average distortion of these two n letter blocks.

It is another key result in information theory that the rate distortion function $R(D)$ of a discrete source can be evaluated as the minimum mutual information $I(X; \hat{X})$ between the source and the reconstruction under the distortion constraint:

$$R(D) = \min_{\mathbf{P}: \mathbf{E}[d(X,\hat{X})] \leq D} I(X; \hat{X}) \tag{3.14}$$

where $P_{ij} = \mathbf{Prob}\{\hat{X} = j | X = i\},\ i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, M$ are the reconstruction probabilities.

Writing out the minimization problem (3.14) explicitly, we have the following rate distortion problem:

$$\begin{array}{ll} \mathsf{minimize} & \sum_{i=1}^N \sum_{j=1}^M p_i P_{ij} \log \frac{P_{ij}}{\sum_k P_{kj} p_k} \\ \mathsf{subject\ to} & \sum_{i=1}^N \sum_{j=1}^M p_i P_{ij} d_{ij} \leq D \\ & \sum_{j=1}^M P_{ij} = 1, \quad i = 1, 2, \ldots, N \\ & P_{ij} \geq 0, \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, M \\ \mathsf{variables} & \mathbf{P}. \end{array} \tag{3.15}$$

The constant parameters are the source distribution $\mathbf{p}$, the distortion measures $d_{ij} = d(X = i, \hat{X} = j)$, and the distortion constraint D.

Theorem 3.2. The Lagrange dual of the rate distortion problem (3.15) is the following GP (in convex form):

$$\begin{array}{ll} \mathsf{maximize} & \mathbf{p}^T\boldsymbol{\alpha} - \gamma D \\ \mathsf{subject\ to} & \log \sum_{i=1}^N \exp(\log p_i + \alpha_i - \gamma d_{ij}) \leq 0, \\ & j = 1, 2, \ldots, M \ \gamma \geq 0 \\ \mathsf{variables} & \boldsymbol{\alpha}, \gamma. \end{array} \tag{3.16}$$

The constant parameters are $\mathbf{p}, d_{ij}$ and D.

An equivalent version of the Lagrange dual problem is the following GP (in standard form):

$$\begin{array}{ll} \mathsf{maximize} & w^{-D} \prod_{i=1}^N z_i^{p_i} \\ \mathsf{subject\ to} & \sum_{i=1}^N p_i z_i w^{-d_{ij}} \leq 1, \quad j = 1, 2, \ldots, M \\ & w \geq 1, \quad z_i \geq 0, \quad i = 1, 2, \ldots, N \\ \mathsf{variables} & \mathbf{z}, w. \end{array} \tag{3.17}$$

Lagrange duality between problems (3.14) and (3.16) means the following

- *Weak duality.* Any feasible $(\boldsymbol{\alpha}, \gamma)$ of the Lagrange dual problem (3.16) produce a lower bound on the rate distortion function:
$$\mathbf{p}^T\boldsymbol{\alpha} - \gamma D \leq R(D).$$

- *Strong duality.* The optimal value of the Lagrange dual problem (3.16) is $R(D)$.

In [14] Berger proved an equivalent formulation as (3.16). The proof in Appendix B.3 is simpler by directly using the Lagrange duality argument.

The Lagrange dual (3.17) of the rate distortion problem (3.15) is a simple GP: maximizing a monomial over posynomial constraints, in the form of maximizing a geometric mean $\prod_{i=1}^N z_i^{p_i}$ weighted by w^{-D}, under constraints on arithmetic means $\sum_{i=1}^N p_i z_i$ weighted by $w^{-d_{ij}}$.

A smaller shadow price w would reduce the objective value but also loosen each constraint, allowing larger dual variables z_i and possibly a higher objective value.

Similar to the case for channel capacity, we can now efficiently lower bound the rate distortion function from the dual problem in the GP form. In particular, due to the structure of the constraints in (3.17), for any given w, finding a dual feasible $\mathbf{z}$ reduces to the easy task of solving a system of linear inequalities. For example, with the Hamming distortion measure, it is easy to verify that $\alpha_i = \log\left(\frac{1-D}{p_i}\right)$, $\forall i$, and $\gamma = \log\left(\frac{(1-D)(N-1)}{D}\right)$ satisfy the Lagrange dual constraints in (3.16), and give the following lower bound:

$$R(D) \geq H(X) - H_0(D) - D\log(N-1) \tag{3.18}$$

where $H_0(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function.

Now consider guessing a random variable X based on another random variable $\hat{X}$. If we replace D by the probability of estimation error P_e and use the fact that $R(D) = \min I(X;\hat{X}) \leq H(X) - H(X|\hat{X})$, then the lower bound (3.18) recovers Fano's inequality:

$$H(X|\hat{X}) \leq H_0(P_e) + P_e\log(N-1) \tag{3.19}$$

that helps prove the converse theorem for channel capacity [40].

The problem of rate distortion with state information [39, 130] also has a Lagrange dual problem in the form of GP [33], and the Lagrange duality bounding technique leads to a generalization of Fano's inequality. The Lagrange dual problems for channel capacity and rate distortion with generalized information measures were also derived in [120, 121].

3.1.3 Shannon duality

Lagrange duality is often loosely stated as follows. Given an optimization problem called the primal problem, the objective and constraint functions in the dual problem can be obtained from those in the primal problem by some simple mappings of signs, variables, constant parameters and mathematical operations. This 'duality by mapping' that

turns one optimization problem into its Lagrange dual is also found in other duality relationships [90], such as that between controllability and observability. However, as can be easily verified, 'duality by mapping' does not hold between the primal problems of channel capacity (3.2) and rate distortion (3.15). It turns out that their Lagrange dual problems exhibit a precise 'duality by mapping'. Due to strong duality, this induces a 'duality by mapping' between the primal problems through the GP duals, as shown in Figure 3.1. Note that Lagrange duality is different from Shannon duality. Indeed, while channel capacity and rate distortion are 'somewhat dual' as commented by Shannon [114], their Lagrange dual problems are both GPs.

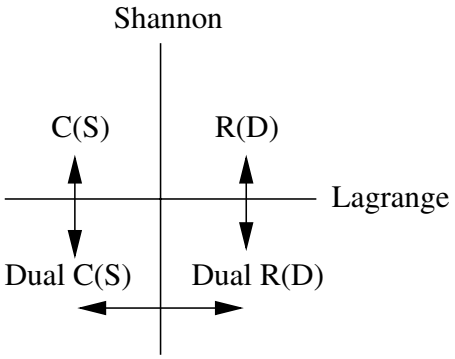

Fig. 3.1 Shannon duality characterized through the Lagrange dual problems of channel capacity and rate distortion.

We first summarize two versions of the Lagrange dual problems for $C(S)$ and $R(D)$ in Table 3.1, where the GP dual problems in standard form better illustrate the 'duality by mapping' relationships. The objective functions, constraint functions, variables, and constant parameters in the Lagrange dual problems of $C(S)$ and $R(D)$ can be obtained from one another through the following simple mappings:

Shannon Duality Correspondence

Channel capacity $C(S)$		Rate distortion $R(D)$
monomial	$\leftrightarrow$	posynomial
posynomial	$\leftrightarrow$	monomial

	Lagrange Dual of Channel Capacity	*Lagrange Dual of Rate Distortion*
Convex Form	minimize $\log \sum_{j=1}^{M} e^{\alpha_j + \gamma S}$ subject to $\sum_{j=1}^{M} P_{ij}(\alpha_j - \log P_{ij}) + \gamma s_i \geq 0$, $\gamma \geq 0$, variables: α_j, γ constants: channel P_{ij}, cost s_i, S	maximize $p\alpha - \gamma D$ subject to $\log \sum_{i=1}^{N} e^{\log p_i + \alpha_i - \gamma d_{ij}} \leq 0$ $\gamma \geq 0$ variables: α_i, γ constants: source p_i, distortion d_{ij}, D
Standard Form	minimize $w^S \sum_{j=1}^{M} z_j$ subject to $\prod_{j=1}^{M} e^{H(P^{(i)})} w^{s_i} z_j^{P_{ij}} \geq 1$, $w \geq 1, \quad z_j \geq 0$ variables: z_j, w constants: channel P_{ij}, cost s_i, S	maximize $w^{-D} \prod_{i=1}^{N} z_i^{p_i}$ subject to $\sum_i p_i z_i w^{-d_{ij}} \leq 1$, $w \geq 1, \quad z_i \geq 0$ variables: z_i, w constants: source p_i, distortion d_{ij}, D

Table 3.1 Lagrange dual problems of channel capacity with input cost and rate distortion.

$$
\begin{array}{rcl}
\text{minimization} & \leftrightarrow & \text{maximization} \\
\geq \text{constraints} & \leftrightarrow & \leq \text{constraints} \\
j(\text{receiver side index}) & \leftrightarrow & i(\text{sender side index}) \\
i(\text{sender side index}) & \leftrightarrow & j(\text{receiver side index}) \\
w^S & \leftrightarrow & w^{-D} \\
w^{s_i} & \leftrightarrow & w^{-d_{ij}} \\
z_j & \leftrightarrow & z_i^{p_i} \\
z_j^{P_{ij}} & \leftrightarrow & z_i \\
H(\mathbf{P}^{(i)}) & \leftrightarrow & -\log \dfrac{1}{p_i}
\end{array}
$$

Lagrange duality gives a detailed analysis of the structures in Shannon duality:

- It resolves the apparent asymmetry between maximizing over a vector $\mathbf{p}$ in the channel capacity problem and minimizing over a matrix $\mathbf{P}$ in the rate distortion problem. In the Lagrange dual of $C(S)$, there are as many optimization variables as output alphabet symbols, plus a shadow price for the cost constraint. In the Lagrange dual of $R(D)$, there are as many optimization variables as input alphabet symbols, plus a shadow price for the distortion constraint.
- It answers the following question: since a vector $\mathbf{p}$ (the source distribution) is given in the rate distortion problem, and a matrix $\mathbf{P}$ (the channel matrix) is given in the channel capacity problem, what is the proper analog of p_i (the ith entry in $\mathbf{p}$) in the channel capacity problem? The last pair in the Shannon duality correspondence shows that the proper analog of $\log \frac{1}{p_i}$ in rate distortion is $H(\mathbf{P}^{(i)})$ in channel capacity: $\log \frac{1}{p_i}$ is the number of bits to describe an alphabet symbol with probability p_i in the Shannon code for lossless compression, and $H(\mathbf{P}^{(i)})$ is the number of bits needed to describe without loss a source whose distribution follows the ith row of channel matrix. This correspondence can be interpreted in the context of universal source coding, where each row $\mathbf{P}^{(i)}$

of the channel matrix represents a possible distribution of a source.

- It confirms Shannon's remark in [114] on introducing input costs to enhance duality. From the Lagrange dual problems in standard form GP, it is easy to see that input costs $\mathbf{s}$ and cost constraint S in the $C(S)$ dual problem are complementary to distortion measures d_{ij} and distortion constraint D in the $R(D)$ dual problem.
- The dual variable $w \geq 1$ can be interpreted as the shadow price associated with the input cost constraint S and with the reconstruction distortion constraint D, respectively. From local sensitivity analysis [21], the optimal $-w^*$ tells us approximately how much increase in capacity $C(S)$ or reduction in rate $R(D)$ would result if the cost or distortion constraint can be relaxed by a small amount. From global sensitivity analysis [21], if w^* is large, then tightening the cost or distortion constraint will greatly decrease capacity (in the channel capacity problem) or increase rate (in the rate distortion problem). If w^* is small, then loosening the cost or distortion constraint will not significantly increase capacity or decrease rate.

In addition to the above Lagrange duality based characterization, Shannon duality has also been characterized for single-user and multiple-user models through functional duality and operational duality.[4]

3.1.4 Error exponent

Channel capacity is the x-intercept of an exponential decay curve of the decoding error probability: when the transmission rate R is below capacity C, the average decoding error probability $\bar{P}_e^{(N)}(R)$ decreases exponentially as the codebook length N tends to infinity, and

[4] The functional duality [105] and Lagrange duality approach together imply that solving a GP in the form of (3.4) induces a set of problem parameters for another GP in the form of (3.16), whose optimal value equals that of the first GP.

the corresponding exponent is called the reliability function $E(R) = \lim_{N\to\infty} -\frac{1}{N}\log \bar{P}_e^{(N)}(R)$.

Reliability function is usually approximated by the following random coding upper bound on error probability [61]:

$$\bar{P}_e^{(N)}(R) \leq \exp(-NE_r(R)). \tag{3.20}$$

The random coding exponent $E_r(R)$ is the maximized value of the objective function in the following optimization problem:

$$\begin{array}{ll} \mathsf{maximize} & E_0(\rho, \mathbf{p}) - \rho R \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ & \rho \in [0,1] \\ \mathsf{variables} & \mathbf{p}, \rho \end{array} \tag{3.21}$$

where

$$E_0(\rho, \mathbf{p}) = -\log \sum_j \left(\sum_i p_i \left(P_{ij}\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{3.22}$$

The constant parameters are the channel $\mathbf{P}$ and a given rate R.

To upper bound the decoding error probability, we can solve problem (3.21), by first maximizing over $\mathbf{p}$ and then ρ. Properties of such optimization have been analytically studied, e.g., in [61]. Here we show that the problem of maximizing over $\mathbf{p}$ for a given ρ is a Reversed GP, and indeed a special one that can be turned into a convex optimization and has an unconstrained optimization as its Lagrange dual problem, thus leading to efficient upper bounds on the rate achievable for a finite block length and a desired error probability.

First, we show that maximizing $E_0(\rho, \mathbf{p})$ over $\mathbf{p}$ for a given ρ can be turned into a Reversed GP in convex form: minimizing a log-sum-exp objective function with equality constraints on other log-sum-exp functions. Let $A_{ij} = P_{ij}^{1/(1+\rho)}$. Maximizing E_0 over $\mathbf{p}$ for a given ρ can be written as:

$$\begin{array}{ll} \mathsf{minimize} & \log \sum_j \left(\sum_i p_i A_{ij}\right)^{1+\rho} \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}. \end{array} \tag{3.23}$$

Introducing the following variables:

$$\begin{aligned} r_j &= \log \sum_i p_i A_{ji}, \quad \forall j, \\ t_i &= \log p_i, \quad \forall i, \end{aligned}$$

we can rewrite (3.23) as the following problem:

$$\begin{array}{ll} \mathsf{minimize} & \log \sum_j \exp(r_j(1+\rho)) \\ \mathsf{subject\ to} & \log \sum_i \exp(t_i) A_{ij} = r_j, \quad \forall j \\ & \log \sum_i \exp(t_i) = 0 \\ \mathsf{variables} & \mathbf{r}, \mathbf{t}, \end{array}$$

which is equivalent to the following Reversed GP in variables $\mathbf{r}, \mathbf{t}$, where the constraints are written as equalities on log-sum-exp functions:

$$\begin{array}{ll} \mathsf{minimize} & \log \sum_j \exp(r_j(1+\rho)) \\ \mathsf{subject\ to} & \log \sum_i \exp(t_i - r_j + \log A_{ij}) = 0 \quad \forall j \\ & \log \sum_i \exp(t_i) = 0 \\ \mathsf{variables} & \mathbf{r}, \mathbf{t}. \end{array}$$

While in general Reversed GP cannot be turned into convex optimization problems and the duality gap is strictly positive, in this case the error exponent problem (3.23) can be transformed into a convex problem through a different change of variable, as shown in Appendix B.4 that proves the following theorem:

Theorem 3.3. The Lagrange dual problem of (3.23) is the following unconstrained concave maximization over $\boldsymbol{\alpha}$:

$$\mathsf{maximize} \quad \left[\theta(\rho) \sum_j \alpha_j^{(1+\rho)/\rho} - \max_i \left\{ \sum_j A_{ij} \alpha_i \right\} \right]. \tag{3.24}$$

where $\theta(\rho) = \frac{\rho(-1)^{1/\rho}}{(1+\rho)^{1+1/\rho}}$ and $A_{ij} = P_{ij}^{1/(1+\rho)}$.

By weak duality, the dual problem (3.24) gives the following bound parameterized by $\boldsymbol{\alpha}$:

$$E_0(\rho) \leq \max_i \left\{ \sum_j A_{ij} \alpha_i \right\} - \theta(\rho) \sum_j \alpha_j^{(1+\rho)/\rho}$$

By strong duality, the optimized value of (3.24) equals $-E_0(\rho)$.

Equivalent forms of the above dual characterization of the random coding error exponent have been obtained in [42], as recently pointed out by [82].

A converse proof of channel capacity was given [4] by providing a lower bound on error probability through optimizing E_0 over a different range of ρ:

$$\bar{P}_e^{(N)}(R) \leq \exp(-N\hat{E}_r(R))$$

where

$$\hat{E}_r(R) = \max_{\rho\in[-1,0]} \left[-\rho R + \min_{\mathbf{p}}(E_0(\rho, \mathbf{p}))\right].$$

The Lagrange dual problem of minimizing $E_0(\rho, \mathbf{p})$ over $\mathbf{p}$ for a given $\rho \in [-1, 0]$ has also been derived into a form similar to (3.24) [84].

Finally, we consider the problem of maximizing the rate R under a given constraint $\bar{P}_{e,N}$ on the decoding error probability for codewords with blocklength N. This problem appears in the 'achievability' part of Shannon's channel capacity theorem [113]. Again we restrict to the relaxed case where a $\rho \in [0, 1]$ is given, and the optimization variables are $(R, \mathbf{p})$:

$$\begin{array}{ll} \mathsf{maximize} & R \\ \mathsf{subject\ to} & E_r(R) \leq \bar{P}_{e,N} \\ & R \geq 0 \\ & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}, R. \end{array} \tag{3.25}$$

Substituting the definition of $E_r(R)$ into the above problem, we can rewrite (3.25) as:

$$\begin{array}{ll} \mathsf{maximize} & R \\ \mathsf{subject\ to} & E_0(\mathbf{p}) - \rho R \geq -\frac{\log \bar{P}_{e,N}}{N} \\ & R \geq 0 \\ & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}, R, \end{array}$$

which is in turn equivalent to

$$\begin{array}{ll} \mathsf{maximize} & \frac{1}{\rho}\left[E_0(\mathbf{p}) + \frac{\log \bar{P}_{e,N}}{N}\right] \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p}, \end{array}$$

since at optimality $R = \frac{1}{\rho}\left[E_0(\mathbf{p}) + \frac{\log P_{e,N}}{N}\right]$. Up to a constant scaling and shift, problems (3.25) and (3.23) are equivalent. Therefore, Theorem 3.3 leads to the following

---

Corollary 3.3. The maximum achievable rate R with codeword block-length N under a decoding error probability $\bar{P}_{e,N}$ is upper bounded by

$$R \leq \frac{1}{\rho}\left[\max_i \left\{\sum_j A_{ij}\alpha_i\right\} - \theta(\rho)\sum_j \alpha_j^{(1+\rho)/\rho} + \frac{\log \bar{P}_{e,N}}{N}\right] \tag{3.26}$$

where $\rho \in [0, 1]$.

3.2 Coding and Signal Processing

Materials in this subsection are in part based on [25, 30, 69, 75, 91].

3.2.1 Channel coding: group code for gaussian channel

Consider an additive Gaussian channel:

$$\mathbf{y} = \mathbf{x} + \mathbf{z}$$

where $\mathbf{x} \in \mathbf{R}^n$ is the transmitted signal, $\mathbf{z}$ is additive Gaussian noise with zero mean and variance N, and $\mathbf{y}$ is the received signal. GP, or more precisely, SP, can be used to solve the initial vector problem in the theory of group codes for Gaussian channels.

Let $\mathcal{G}$ be a finite group of real orthogonal $n \times n$ matrices and $\mathbf{x}$ a unit vector in $\mathbf{R}^n$. The group code generated by $\mathcal{G}$ and $\mathbf{x}$ is $\mathcal{G}\mathbf{x} = \{\mathbf{G}\mathbf{x} | \mathbf{G} \in \mathcal{G}\}$. The optimal initial vector $\mathbf{x}$ maximizes the minimum distance d of the group code generated by $\mathcal{G}$:

$$\mathbf{x} = \operatorname*{argmax}_{\mathbf{z}\in\mathbf{R}^n, \|\mathbf{z}\|=1} \min_{\mathbf{G}\in\mathcal{G}, \mathbf{G}\neq id} \{d(\mathbf{z}, \mathbf{G}\mathbf{z})\}.$$

We will focus on codes generated by primitive permutation groups.

This problem was first formulated in [116]. Using a feasible directions algorithm, [74] presents a table of the optimal codes for well-known primitive permutation groups of degree less than or equal to 12. A GP-based algorithm in [75] finds the optimal minimum distances for all the named groups of degree less than or equal to 20 listed in [115].

For a unit vector $\mathbf{z}$, we have $d^2(\mathbf{Gz}, \mathbf{z}) = 2 - 2\mathbf{z}^T\mathbf{Gz}$. Therefore, maximizing the minimum distance is equivalent to minimizing the maximum inner product between $\mathbf{Gz}$ and $\mathbf{z}$. It has been shown [74] that a necessary condition for $\mathbf{z}$ to be an optimal initial vector for a permutation group of degree n is $\mathbf{1}^T\mathbf{z} = 0$. Together with the condition that $\mathbf{z}$ is a unit vector, the following optimization formulation of the initial vector problem is formed:

$$\begin{array}{ll} \mathsf{minimize} & \max_{\mathbf{G}\in\mathcal{S}} \mathbf{z}^T\mathbf{Gz} \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{z} = 0 \\ & \mathbf{z}^T\mathbf{z} = 1 \\ \mathsf{variables} & \mathbf{z} \end{array}$$

where $\mathcal{S}$ denotes the set of primitive permutation groups. Introducing an auxiliary variable t, this minmax problem is equivalent to the following optimization problem:

$$\begin{array}{ll} \mathsf{minimize} & t \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{z} = 0 \\ & \mathbf{z}^T\mathbf{z} = 1 \\ & \mathbf{z}^T\mathbf{Gz} \le t, \ \ \forall \mathbf{G} \in \mathcal{S} \\ \mathsf{variables} & \mathbf{z}, t. \end{array} \tag{3.27}$$

Now we apply another standard method of GP modeling: turn $\mathbf{z}$ into two vectors of variables $\mathbf{z}^+$ and $\mathbf{z}^-$, consisting of the non-negative and negative entries of $\mathbf{z}$, respectively. Then problem (3.27) is equivalent to the following optimization problem:

$$\begin{array}{ll} \mathsf{minimize} & t \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{z}^+ - \mathbf{1}^T\mathbf{z}^- = 0 \\ & \mathbf{z}^{+T}\mathbf{z}^+ + \mathbf{z}^{-T}\mathbf{z}^- - 2\mathbf{z}^+\mathbf{z}^- = 1 \\ & \mathbf{z}^{+T}\mathbf{Gz}^+ + \mathbf{z}^{-T}\mathbf{Gz}^- - \mathbf{z}^{+T}\mathbf{Gz}^- - \mathbf{z}^{-T}\mathbf{Gz}^+ \le t, \\ & \quad \forall \mathbf{G} \in \mathcal{S} \\ & \mathbf{z}^+, \mathbf{z}^- \succeq 0, \ \ t > 0 \\ \mathsf{variables} & \mathbf{z}^+, \mathbf{z}^-, t. \end{array} \tag{3.28}$$

While this is not a GP due to the posynomial equality constraints, it is a SP, and can be solved using the double condensation method in Subsection 2.2.5 through an iterative procedure [75]:

(1) Find a feasible point of problem (3.28).
(2) Use the double condensation method to convert (3.28) into an LP and compute its optimal solution.
(3) If the optimal solution is not feasible in (3.28), recondense the most violated constraint and solve the new problem. Repeat this until the resulting optimal solution is feasible in (3.28).
(4) If the change in t over the last iteration is smaller than a stopping criterion threshold ϵ, stop the iteration. Otherwise repeat from Step 2.

This method significantly improves the earlier numerical techniques. The order of the largest permutation group for which the initial vector problem is solved increases from 7920 to 322560, which corresponds to the largest group found in the table in [115]. A complete list of optimal permutation group codes' properties can be found in [75].

3.2.2 Source coding: relaxation of lossless code

Lossless source coding is a data compression problem and a special case of the rate distortion problem treated in Subsection 3.1.2. A source code $\mathcal{C}$ for a random variable X is a mapping from the range of X to the set of finite length strings of symbols from a W-ary alphabet. When $W = 2$, $\mathcal{C}$ is a binary source code. Let $\mathcal{C}(i)$ denote the codeword corresponding to $X = i$ and l_i denote the length of $\mathcal{C}(i)$, $i = 1, 2, \ldots, N$. We are primarily interested in the class of source codes called the prefix code, which means no codeword is a prefix of any other codeword, and, consequently, the codes can be instantaneously decoded. Kraft inequality states that for any prefix code $\sum_{i=1}^{N} W^{-l_i} \leq 1$. Conversely, for a set of codeword lengths satisfying the Kraft inequality, there exits a prefix code with these codeword lengths [40].

In lossless source coding using prefix codes over W-ary alphabet, we would like to minimize the the expected codeword length $L = \sum_i p_i l_i$ subject to the Kraft inequality. This is an integer optimization problem:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i p_i l_i \\ \mathsf{subject\ to} & \sum_i W^{-l_i} \leq 1 \\ & \mathbf{l} \in \mathcal{Z}_+^N \\ \mathsf{variables} & \mathbf{l}. \end{array} \tag{3.29}$$

The constant parameters are $\mathbf{p}$ and W. We show that relaxed versions of lossless source coding problems without the integer constraints are GPs. Let $z_i = W^{-l_i}$, we rewrite the problem of expected codeword length minimization as the following GP:

$$\begin{array}{ll} \mathsf{minimize} & \prod_i z_i^{-p_i} \\ \mathsf{subject\ to} & \mathbf{1}^T \mathbf{z} \leq 1 \\ & \mathbf{z} \succeq 0 \\ \mathsf{variables} & \mathbf{z}. \end{array} \tag{3.30}$$

The constant parameters are $\mathbf{p}$.

Interestingly, the lossless source coding problem (3.30) is indeed a special case of the GP dual of the lossy data compression problem (3.17), with $D = 0, d = 0$ and $\mathbf{p} = \mathbf{1}$. Problem (3.30) is a simple relative entropy minimization with an analytic solution. Indeed, it is readily shown [40] that letting

$$l_i = -\log_W p_i, \ \ \forall i,$$

solves problem (3.29), and the minimized L^* is $H(\mathbf{p})$. However, since l_i must be an integer, we let $l_i = \lceil -\log p_i \rceil$ as in the Shannon–Fano code [40], and the resulting L is close to the optimum without the integer constraint:

$$H(\mathbf{p}) \leq L \leq H(\mathbf{p}) + 1.$$

In general, we have the following source coding problem:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i f(l_i) \\ \mathsf{subject\ to} & \sum_i W^{-l_i} \leq 1 \\ & g(\mathbf{l}) \leq 0 \\ & \mathbf{l} \succeq 0 \\ \mathsf{variables} & \mathbf{l}. \end{array} \tag{3.31}$$

where the objective function to be minimized is the sum of some convex function f of l_i over i, g in the constraint can be any convex function of $\mathbf{l}$, and the other constraints are Kraft's inequality and nonnegativity constraint.

Among all the possible choices of f, Huffman coding is known to be optimal only when f is a linear or exponential function [69]. For the linear case when $f(l_i) = p_i l_i$, we have shown that the source coding problem is equivalent to the following GP where a monomial is minimized:

$$\begin{array}{ll} \mathsf{minimize} & \prod_i z_i^{-p_i} \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{z} \leq 1 \\ & \mathbf{z} \succeq 0 \\ \mathsf{variables} & \mathbf{z}. \end{array}$$

Recall that the objective function of a GP can be a posynomial rather than just a monomial as above. We may ask which special case of problem (3.31) would the following GP corresponds to:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i p_i z_i^{-\beta} \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{z} \leq 1 \\ & \mathbf{z} \succeq 0 \\ \mathsf{variables} & \mathbf{z}. \end{array}$$

It is easy to verify, by a logarithmic change of variable $l_i = -\log_W z_i$, that the above GP is equivalent to the following problem, where f in (3.31) is an exponential function of the optimization variables $\mathbf{l}$:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i p_i b^{l_i} \\ \mathsf{subject\ to} & \sum_i W^{-l_i} \leq 1 \\ & \mathbf{l} \succeq 0 \\ \mathsf{variables} & \mathbf{l}. \end{array} \tag{3.32}$$

where $b = W^\beta$.

Similar to the case with a linear f, where the optimal value of the objective function is $H(\mathbf{p})$, the optimal value of the objective function

for the case with an exponential f is also known in closed form [25] as $b^{H_\gamma(\mathbf{p})}$, where $\gamma = \frac{1}{1+\beta}$ and H_γ is the Renyi entropy of order γ:[5]

$$H_\gamma(\mathbf{p}) = \frac{1}{\gamma - 1} \log \sum_i p_i^\gamma. \tag{3.33}$$

This result can again be readily derived through Lagrange duality. Since $W \geq 2$, the constraint $\mathbf{l} \succeq 0$ in (3.32) is redundant. Now we find the Lagrange dual problem of (3.32), by first forming the Lagrangian:

$$L(\mathbf{l}, \lambda) = \sum_i p_i b^{l_i} + \lambda \left(\sum_i W^{-l_i} - 1 \right).$$

Differentiating $L(\mathbf{l}, \lambda)$ with respect to l_i and equating with zero gives

$$l_i^* = \frac{1}{1+\beta} \log_W \left(\frac{\lambda}{\beta p_i} \right), \quad \forall i.$$

Therefore, the Lagrange dual of the source coding problem with exponential penalty is the following unconstrained maximization with a scalar variable $\lambda \geq 0$:

$$\text{maximize} \quad \sum_i \left(p_i (\beta p_i)^{-\frac{\beta}{1+\beta}} + (\beta p_i)^{\frac{1}{1+\beta}} \right) \lambda^{\frac{\beta}{1+\beta}} - \lambda,$$

which can be analytically solved to give the optimal value $b^{H_\gamma(p)}, \gamma = \frac{1}{1+\beta}$. By strong duality, this is also the optimal value of the primal problem (3.32).

3.2.3 Signal processing: multi-access transmitter design

GP can be used to optimize a special case of linear transceiver design in m-user multiple access communication systems. For notational simplicity, consider a two-user multiple access channel (MAC) as shown in Figure 3.2. The n-user case is a straightforward extension.

The received signal $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{H}_1 \mathbf{F}_1 \mathbf{x}_1 + \mathbf{H}_2 \mathbf{F}_2 \mathbf{x}_2 + \rho \mathbf{n} \tag{3.34}$$

where $\mathbf{x}_i$, $i = 1, 2$, are the message signals, linearly preprocessed by $\mathbf{F}_i$, transmitted through channels $\mathbf{H}_i$, and corrupted by additive

[5] The Renyi entropy $H_\gamma(\mathbf{p})$ is a generalization of the Shannon entropy $H(\mathbf{p})$ with $\lim_{\gamma \to 1} H_\gamma(\mathbf{p}) = H(\mathbf{p})$.

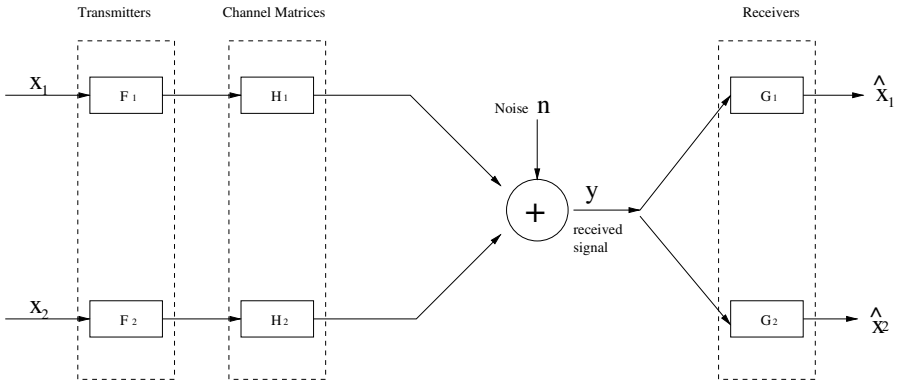

Fig. 3.2 A two-user multiple access communication system.

noise $\mathbf{n}$ with noise variance $\rho > 0$. The received signal $\mathbf{y}$ will be linearly equalized by $\mathbf{G}_i$: the decoded signal is given by $\hat{\mathbf{x}}_i = \text{sign}(\mathbf{G}_i\mathbf{y})$, $i = 1, 2$.

The problem is to design the transceivers for both users: $\mathbf{F}_1, \mathbf{F}_2, \mathbf{G}_1, \mathbf{G}_2$, for given channel matrices $\mathbf{H}_1, \mathbf{H}_2$. In [91], a special case of this multi-access joint transmitter and receiver design problem is examined for zero-forcing equalizers and MMSE transmitters. This case is applicable, for example, to multiple adjacent CDMA cells where each base station is constrained to use a decorrelating detector for its own cell.

Let $\mathbf{e}_i$ be the error vector for user i, $i = 1, 2$. Consider the error vector and Mean Squared Error (MSE) for a zero-forcing equalizer for user 1. The error vector is

$$\mathbf{e}_1 = \mathbf{G}_1\mathbf{y} - \mathbf{x}_1 = (\mathbf{G}_1\mathbf{H}_1\mathbf{F}_1 - \mathbf{I})\mathbf{x}_1 + \mathbf{G}_1\mathbf{H}_2\mathbf{F}_2\mathbf{x}_2 + \rho\mathbf{G}_1\mathbf{n},$$

the zero-forcing equalizer is

$$\mathbf{G}_1 = (\mathbf{H}_1\mathbf{F}_1)^{-1},$$

and the resulting MSE is

$$\text{MSE} = \mathbf{Tr}\left((\mathbf{G}_1\mathbf{H}_2\mathbf{F}_2)(\mathbf{G}_1\mathbf{H}_2\mathbf{F}_2)^T\right) + \rho^2\,\mathbf{Tr}\left(\mathbf{G}_1\mathbf{G}_1^T\right).$$

Introducing new matrix variables $\mathbf{U}_1 = \mathbf{F}_1\mathbf{F}_1^T$ and $\mathbf{V}_1 = \mathbf{G}_1^T\mathbf{G}_1$, the MSE can be written as

$$\text{MSE} = \mathbf{Tr}(\mathbf{V}_1\mathbf{H}_2\mathbf{U}_2\mathbf{H}_2^T) + \rho^2\,\mathbf{Tr}(\mathbf{V}_1).$$

The zero-forcing condition becomes

$$\mathbf{V}_1^{-1} = \mathbf{H}_1\mathbf{U}_1\mathbf{H}_1^T.$$

Adding the power constraint $\mathbf{Tr}(\mathbf{U}_i) \leq P_i,\ i = 1, 2$, we have the following formulation of zero-forcing MMSE transceiver design problem:

$$\begin{array}{ll} \mathsf{minimize} & \mathbf{Tr}(\mathbf{V}_1\mathbf{H}_2\mathbf{U}_2\mathbf{H}_2^T) + \rho^2\,\mathbf{Tr}(\mathbf{V}_1) \\ & + \mathbf{Tr}(\mathbf{V}_2\mathbf{H}_1\mathbf{U}_1\mathbf{H}_1^T) + \rho^2\,\mathbf{Tr}(\mathbf{V}_2) \\ \mathsf{subject\ to} & \mathbf{V}_i^{-1} = \mathbf{H}_i\mathbf{U}_i\mathbf{H}_i^T,\ \ i = 1, 2 \\ & \mathbf{Tr}(\mathbf{U}_i) \leq P_i,\ \ i = 1, 2 \\ & \mathbf{V}_i \succeq 0,\ \ \mathbf{U}_i \succeq 0,\ \ i = 1, 2 \\ \mathsf{variables} & \mathbf{V}_1, \mathbf{V}_2, \mathbf{U}_1, \mathbf{U}_2. \end{array} \tag{3.35}$$

An alternating optimization method can be applied to (3.35), where $\mathbf{U}_1, \mathbf{V}_1$ are fixed and $\mathbf{U}_2, \mathbf{V}_2$ optimized in one iteration, and in the next iteration, $\mathbf{U}_2, \mathbf{V}_2$ are fixed and $\mathbf{U}_1, \mathbf{V}_1$ optimized. In each iteration, (3.35) is reduced to a SDP.

Suppose the channel matrices $\mathbf{H}_i$ are diagonal with n independent tones indexed by j. If $\mathbf{V}_1, \mathbf{U}_1$ are given positive definite diagonal matrices, then the alternating optimization procedure can be shown to lead to diagonal $\mathbf{V}_i, \mathbf{U}_i$ for both users in all iterations. It is conjectured [91] that the optimal transceiver matrices are indeed diagonal. In this case, the non-convex optimization (3.35) reduces to the following problem:

$$\begin{array}{ll} \mathsf{minimize} & \sum_{j=1}^n \left(|h_2(j)|^2 u_2(j) v_1(j) + |h_1(j)|^2 u_1(j) v_2(j)\right) \\ & + \rho^2 \sum_{j=1}^n (v_1(j) + v_2(j)) \\ \mathsf{subject\ to} & \sum_{j=1}^n u_i(j) \leq P_i,\ \ i = 1, 2 \\ & |h_i(j)|^2 u_i(j) \geq v_i^{-1}(j),\ \ i = 1, 2,\ \ j = 1, \ldots, n \\ & \mathbf{v}_i \succeq 0,\ \ i = 1, 2 \\ \mathsf{variables} & \mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_1, \mathbf{u}_2, \end{array}$$

which, by eliminating variables $\mathbf{u}_1, \mathbf{u}_2$, is in turn equivalent to the following GP in variables $\mathbf{v}_1$ and $\mathbf{v}_2$, which are the diagonal vectors of $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively:

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{n} \left(v_1^{-1}(j) v_2(j) + v_1(j) v_2^{-1}(j) \right) \\ & + \rho^2 \sum_{j=1}^{n} (v_1(j) + v_2(j)) \\ \text{subject to} & \sum_{j=1}^{n} h_i^{-2}(j) v_i^{-1}(j) \leq P_i, \quad i = 1, 2 \\ & \mathbf{v}_i \succeq 0, \quad i = 1, 2 \\ \text{variables} & \mathbf{v}_1, \mathbf{v}_2 \end{array} \tag{3.36}$$

where j are the independent tones in the channels, and $v_i(j)$ denotes the jth component of vector $\mathbf{v}_i,\ i = 1, 2$.

After solving GP (3.36) to obtain the (squared) optimal receivers $\mathbf{v}_i$, we can recover the (squared) optimal transmitters: $u_i^*(j) = |h_i(j)|^{-2} (v_i^*(j))^{-1},\ i = 1, 2,\ j = 1, \ldots, n$.

3.3 Network Resource Allocation

Materials in this subsection are in part based on [37, 34, 35, 72, 73].

GP in standard form can be used to efficiently optimize network resource allocations for nonlinear objectives under nonlinear Quality of Service (QoS) constraints. The key idea is that resources are often allocated proportional to some parameters, and when resource allocations is 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. As this subsection will illustrate through examples in wireless power control and admission control, this idea can be further extended to provide a GP-based method of resource allocation. Specific examples in wireless power control are first presented in Subsection 3.3.1 before the general method is discussed in Subsection 3.3.2.

3.3.1 Wireless network power control

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 subsection, 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 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 subsection shows that, when SIR is much larger than 0dB, 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. When SIR is comparable to or below 0dB, the power control problems are *truly* non-convex with no efficient and global solution methods. In this case, we present a heuristic based on SP condensation that empirically performs well in computing 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 four numerical examples in this subsection.

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 SIR for the receiver on logical link i is:

$$\mathrm{SIR}_i = \frac{P_i G_{ii} F_{ii}}{\sum_{j \neq i}^{N} P_j G_{ij} F_{ij} + n_i} \tag{3.37}$$

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

$$R_i = \frac{1}{T}\log_2(K\mathrm{SIR}_i) \tag{3.38}$$

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. This approximation is the watershed between convexity and non-convexity in power control problems, and later this subsection we will discuss how to solve truly non-convex power control problems when SIR is not high. For notational simplicity in the rest of this subsection, 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 [73]

$$P_{o,i} = 1 - \prod_{j \neq i} \frac{1}{1 + \frac{\text{SIR}_{th}G_{ij}P_j}{G_{ii}P_i}},$$

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 $\mathbf{P}$:

$$\prod_{j \neq i} \left(1 + \frac{\text{SIR}_{th}G_{ij}P_j}{G_{ii}P_i} \right) \leq \frac{1}{1 - P_{o,i,max}}. \tag{3.39}$$

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

Proposition 3.1. The following nonlinear problem of maximizing the SIR of a particular user i^* is a GP:

$$\begin{array}{ll} \mathsf{maximize} & \mathrm{SIR}_{i^*}(\mathbf{P}) \\ \mathsf{subject\ to} & \mathrm{SIR}_i(\mathbf{P}) \geq \mathrm{SIR}_{i,min}, \ \ \forall i \\ & P_{i1}G_{i1} = P_{i2}G_{i2} \\ & 0 \leq P_i \leq P_{i,max}, \ \ \forall i \\ \mathsf{variables} & \mathbf{P}. \end{array}$$

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:

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

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 $(\mathbf{P}, t)$.

Power control example 1. A simple system comprised of five users is employed 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 $P_{max} = 0.5\mathrm{mW}$. The noise power is $0.5\mu\mathrm{W}$ 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 3.3.

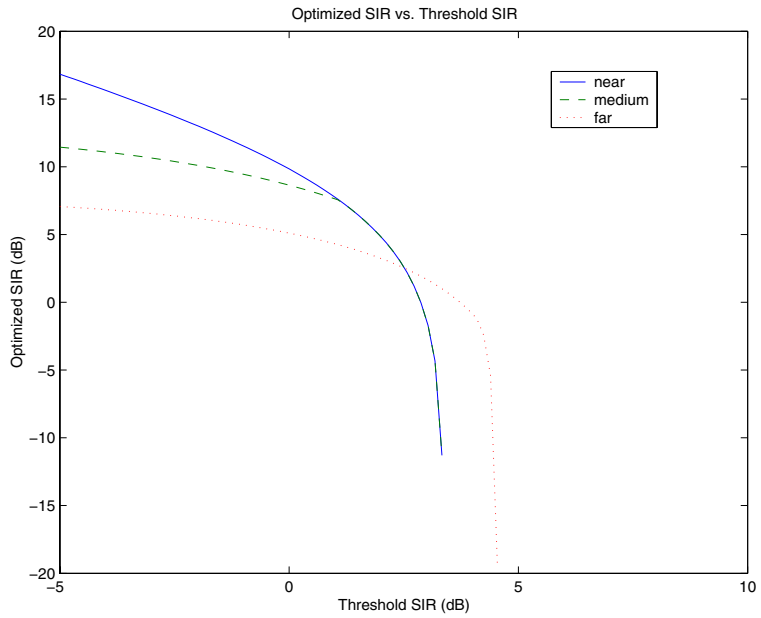

Fig. 3.3 Constrained optimization of power control in a cellular network (Example 1).

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 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.

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 3.2. The following nonlinear problem of maximizing system throughput is a GP:

$$\begin{array}{ll} \mathsf{maximize} & R_{system}(\mathbf{P}) \\ \mathsf{subject\ to} & R_i(\mathbf{P}) \geq R_{i,min}, \ \ \forall i \\ & P_{o,i}(\mathbf{P}) \leq P_{o,i,max}, \ \ \forall i \\ & 0 \leq P_i \leq P_{i,max}, \ \ \forall i \\ \mathsf{variables} & \mathbf{P}. \end{array} \tag{3.40}$$

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 (3.38,3.39).

There are several obvious variations of problem (3.40) 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 (3.40) plus the additional constraints $\mathrm{ISR}_i \leq t_i$

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 be used for admission control based on GP feasibility study, and for determining which QoS constraint is a performance bottleneck, i.e., meet tightly at the optimal power allocation, based on GP sensitivity analysis.

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 will be explained in Subsection 3.4.2, and we focus on the second approach in this subsection, by extending problem formulations, such as (3.40), to the multihop case as in the following example.

Power control example 2. Consider a simple four node multihop network shown in Figure 3.4. There are two connections $A \rightarrow B \rightarrow D$ and $A \rightarrow C \rightarrow D$. Nodes A and D, as well as B and C, are separated by a distance of 20m. Power fall-off factor is -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 (3.40), 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.

For this simple network, 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 $A \rightarrow B \rightarrow D$, while U_3 requires 10kbps sent from $A \rightarrow B$. 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

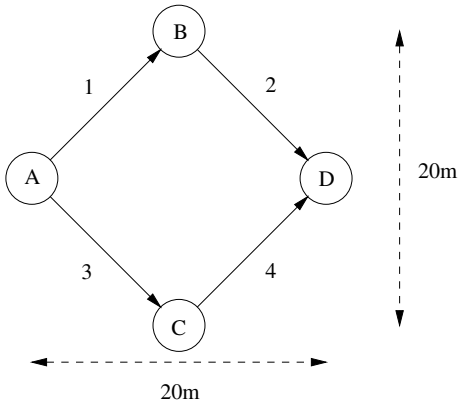

Fig. 3.4 A small wireless multihop network (Example 2).

no feasible power allocation solution, so she is not admitted to the system.

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. 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

$$\bar{D}_i = \frac{1}{\Gamma R_i(\mathbf{P}) - \Lambda_i}. \tag{3.41}$$

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 3.3. 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:

$$\begin{array}{ll} \text{maximize} & R_{system}(\mathbf{P}) \\ \text{subject to} & \bar{D}_i(\mathbf{P}) \leq \bar{D}_{i,max}, \ \ \forall i \\ & P_{BO,i}(\mathbf{P}) \leq P_{BO,i,max}, \ \ \forall i \\ & \text{Same constraints as in problem (3.40)} \\ \text{variables} & \mathbf{P}. \end{array} \tag{3.42}$$

Power control 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 3.5. 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.

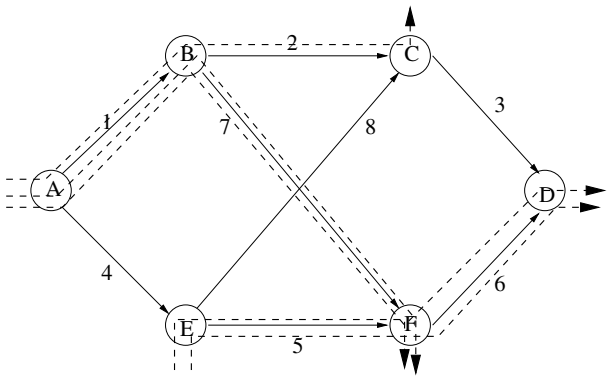

Fig. 3.5 Topology and flows in a multihop wireless network (Example 3).

Figure 3.6 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 a 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.

SP formulations and solutions for the 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 the rest of this subsection.

The first limitation is the assumption that SIR is much larger than 0dB, which can be removed by the condensation method for SP as introduced in Subsection 2.2.5. When SIR is not much larger than 0dB,

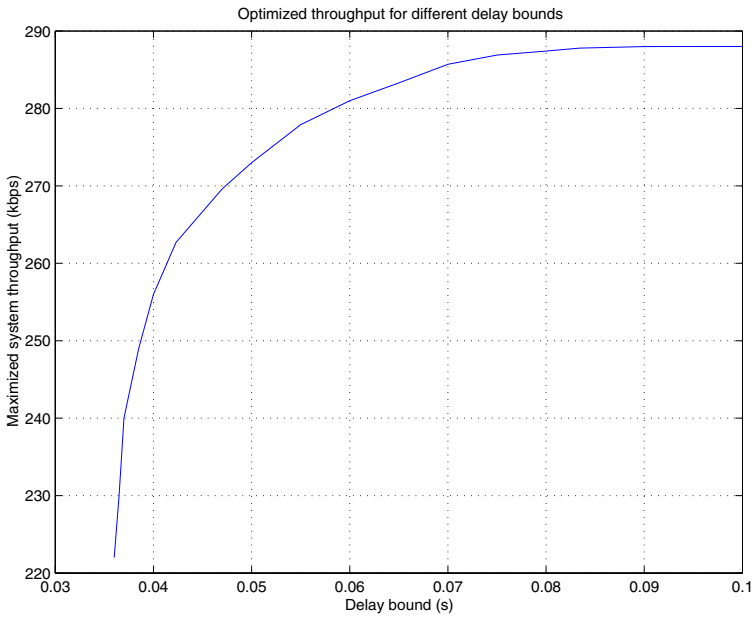

Fig. 3.6 Optimal tradeoff between maximized system throughput and average delay constraint (Example 3).

the approximation of $\log(1+\text{SIR})$ as $\log \text{SIR}$ does not hold. Unlike SIR, $1+\text{SIR}$ 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 non-convex problem that is intrinsically intractable.

Figure 3.7 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 non-convex power control problems that cannot be turned into convex formulations through a series of GPs.

GP-based power control problems in the medium to low SIR regimes become SPs, which can be solved by the single or double condensation method. We focus on the single condensation method here.

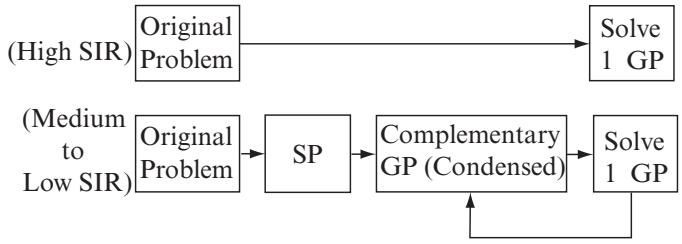

Fig. 3.7 GP-based power control for general SIR regime.

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 3.2, which is explicitly written out as:

$$\begin{array}{ll} \text{minimize} & \prod_{i=1}^{N} \frac{1}{1+\text{SIR}_i} \\ \text{subject to} & (2^{TR_{i,min}}-1)\frac{1}{\text{SIR}_i} \leq 1, \quad i=1,\ldots,N \\ & (\text{SIR}_{th})^{N-1}(1-P_{o,i,max})\prod_{j\neq i}^{N}\frac{G_{ij}P_j}{G_{ii}P_i} \leq 1, \\ & i=1,\ldots,N \\ & P_i(P_{i,max})^{-1} \leq 1, \quad i=1,\ldots,N \\ \text{variables} & \mathbf{P}. \end{array} \tag{3.43}$$

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 an algorithm consisting of the following steps:

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

(1): Evaluate the denominator posynomial of the (3.43) objective function with the given $\mathbf{P}$.

(2): Compute for each term i in this posynomial,

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

(3): Condense the denominator posynomial of the (3.43) objective function into a monomial with weights α_i (Subsection 2.2.5).

(4): Solve the resulting GP e.g., using an interior-point method.

(5): Go to STEP 1 using $\mathbf{P}$ from step (4).

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

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.

Power control 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 $\text{SIR}_{th} = -10$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 (3.43), we pick a random initial feasible power vector $\mathbf{P}$ uniformly between 0 and $\mathbf{P}_{max}$. We compare 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). 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.

We have thus far discussed a power control problem (3.43) where the objective function needs to be condensed. The method is also applicable if some constraint functions are signomials and need to be condensed [37].

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 (3.43) is solved, a slightly modified SP is formulated and solved:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \prod_{i=1}^{N} \frac{1}{1+\text{SIR}_i} \leq t \\ & t \leq \frac{t_0}{\alpha} \\ & \text{Same set of constraints as problem (3.43)} \\ \text{variables} & \mathbf{P}, t. \end{array} \tag{3.44}$$

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 (3.44) 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 a 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 (3.43), then (3.44) would be infeasible, and the iteration of SPs stops.

The above heuristic is applied to the rare instances of power control 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 (3.44).

Distributed algorithms. As an application of the distributed algorithm for GP, we maximize the total system throughput in the high SIR regime, assuming the same $P_{\max}$ at each transmitter. If we directly applied the distributed approach described in Subsection 2.3.3, the resulting algorithm 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 the 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 (after the log change of variables to $\{\tilde{P}_{ij}^R, \tilde{P}_j\}$):

$$\begin{array}{ll} \text{minimize} & \sum_i \log\left(G_{ii}^{-1}\exp(-\tilde{P}_i)\left(\sum_{j\neq i}\exp(\tilde{P}_{ij}^R)+n_i\right)\right) \\ \text{subject to} & \tilde{P}_{ij}^R = \tilde{G}_{ij} + \tilde{P}_j, \quad \forall i,j \\ & \text{Constraints local to each user} \\ \text{variables} & \{\tilde{P}_{ij}^R\}, \{\tilde{P}_j\}. \end{array} \tag{3.45}$$

The partial Lagrangian is

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

from which the dual variable update is found as

$$\begin{aligned} \gamma_{ij}(t+1) &= \gamma_{ij}(t)+\alpha(t)\left(\tilde{P}_{ij}^R-\left(\tilde{G}_{ij}+\tilde{P}_j\right)\right) \\ &= \gamma_{ij}(t)+\alpha(t)\left(\tilde{P}_{ij}^R-\log G_{ij}P_j\right), \end{aligned} \tag{3.46}$$

with stepsizes $\alpha(t)$.

Each user has to minimize the following partial Lagrangian over $\left(\tilde{P}_i,\left\{\tilde{P}_{ij}^R\right\}_j\right)$ for given $\{\gamma_{ij}\}_j$, subject to its own local constraints:

$$\begin{aligned} L_i\left(\tilde{P}_i,\left\{\tilde{P}_{ij}^R\right\}_j,\{\gamma_{ij}\}_j\right) &= \log\left(G_{ii}^{-1}\exp(-\tilde{P}_i)\left(\sum_{j\neq i}\exp(\tilde{P}_{ij}^R)+n_i\right)\right) \\ &\quad +\sum_{j\neq i}\gamma_{ij}\tilde{P}_{ij}^R-\left(\sum_{j\neq i}\gamma_{ji}\right)\tilde{P}_i. \end{aligned}$$

Two practical observations are in order:

- For the minimization of the above local Lagrangian term, 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 (3.46), 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.

To conclude this subsection, we outline several research issues that remain to be further explored for GP-based power control: 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.

3.3.2 General structure

The applications of standard form GP to power control problems can be generalized to other constrained optimization of resource allocation, and the associated problems of feasibility study, sensitivity analysis, admission control, and admission pricing, as long as the QoS metrics can be simplified and reduced to the form of inverted posynomials.

Consider n users indexed by i sharing a common pool of communication resource X, such as bandwidth or buffer. The amount of resource allocated to connection i is denoted by x_i. Consider the following Generalized Proportional Allocation (GPA) form, where the total resource X is allocated to connection i in proportion to some p_i and normalized by a sum of parameters $\sum_j \gamma_{ji} + \alpha\nu_i$:

$$x_i = \frac{p_i}{\sum_j \gamma_{ji} + \alpha\nu_i} X, \quad \forall i \tag{3.47}$$

where the allocation parameters $p_i, \nu_i, \gamma_{ji} \geq 0$ belong to a fixed range of values for each user i (different ranges for different QoS classes), and $\alpha \geq 0$ is a given weight. Usually, there is a tradeoff among the users in resource allocation, i.e., γ_{ji} are increasing functions of p_j.

This form of resource allocation appears in many applications, with power control being an important special case. A possible explanation of the source from which this form arises can be found in Subsection 4.2.1. As a simple example of (3.47), recall the Generalized Processor

Sharing scheme [100] where an egress link with a total rate of R is shared among multiple connections each receiving rate R_i:

$$R_i = \frac{\phi_i}{\sum_j \phi_j} R, \ \ \forall i$$

where $\{\phi_i\}$ are the parameters that can be directly varied in the system design to produce a desirable $\mathbf{R}$.

As another example, [87] shows that rate control on a single link through the congestion control protocol of TCP Reno produces the following rate allocation proportional to the inverse of round trip delay D_i:

$$R_i = \frac{\frac{1}{D_i}}{\sum_j \frac{1}{D_j}} R, \ \ \forall i.$$

These allocation parameters $\{p_i, \nu_i, \gamma_{ji}\}$ can be optimized, within certain ranges, to maximize the resource received by a particular user i^* in the highest QoS class, subject to constraints that lower bound the resources received by each of the other users. Alternatively, for maxmin fairness, allocation parameters can be optimized to maximize the resource received by the user with the minimum received resource. GP can be used to solve them globally and efficiently.

$$\begin{array}{ll} \mathsf{maximize} & x_{i^*} \quad (\text{or maximize } \min_i x_i) \\ \mathsf{subject\ to} & x_i \geq x_{i,min}, \ \ \forall i \\ & \text{variables } (p_i, \gamma_{ji}, \nu_i) \geq \text{lower bounds} \\ & \text{variables } (p_i, \gamma_{ji}, \nu_i) \leq \text{upper bounds.} \end{array} \tag{3.48}$$

Theorem 3.4. The optimization problem (3.48) for resource allocation in the GPA form (3.47) is a GP.

The proof is very simple as shown in Appendix B.5, where the key observation is that the GPA form is an inverted posynomial. The general method of distributed solution for GP in Subsection 2.3.3 can also be applied to such network resource allocation problems.

As shown in the power control examples in the last subsection, the scope of GP method of efficient resource allocation can be extended

to accommodate several types of nonlinear functions of the GPA form (which are in turn nonlinear functions of the underlying design variables) in the objective and constraints.

Optimization (3.48) may not have any feasible solution. Indeed, if the QoS constraints $x_i \geq x_{i,min}, \forall i$, are too strict, there may exist no resource allocation that simultaneously meets all the constraints. Due to the GP form of the problem, feasibility of resource allocation in the GPA form can also be efficiently determined, and used for admission control and pricing in a network: a new user is admitted into the system only if the resulted new problem (3.48) is still feasible, and the user is charged in proportion to the resulting reduction in the objective value of (3.48).

Furthermore, bottlenecks of resource allocation constraints are readily detected, so that if additional resources becomes available, we know where to allocate them to alleviate the bottlenecks of resource demands. Associate a Lagrange dual variable $\sigma_i \geq 0$ for each resource demand constraint $x_i \geq x_{i,min}$. By complementary slackness, if an optimal dual variable $\sigma_1^* > 0$, then we know that the QoS requirement constraint for user 1 is tight at optimality, i.e., $x_1^* = x_{1,min}$.

If the objective is instead to maximize the total resources obtained by a group of users, then the objective cannot be turned into minimization of a posynomial, but a ratio between two posynomials. In that case, the problem becomes a SP and the condensation or double condensation method needs to be applied to solve the problem.

As another example in addition to the power control examples, we show a simple application of the GP method of resource allocation for Connection Admission Control (CAC). Consider the ingress of a switch or a network as shown in Figure 3.8. There are K connections trying to get admitted into the system. They first pass through a traffic shaping mechanism, such as leaky buckets, to conform the connections to their respective provisioned rates λ_i specified in the QoS service level agreement. Due to limitation in the available resource, the CAC controller has to enforce admission control among the contending connections.

Consider the following simple CAC algorithm where the CAC controller has an exponential service time with rate μ. If the first service

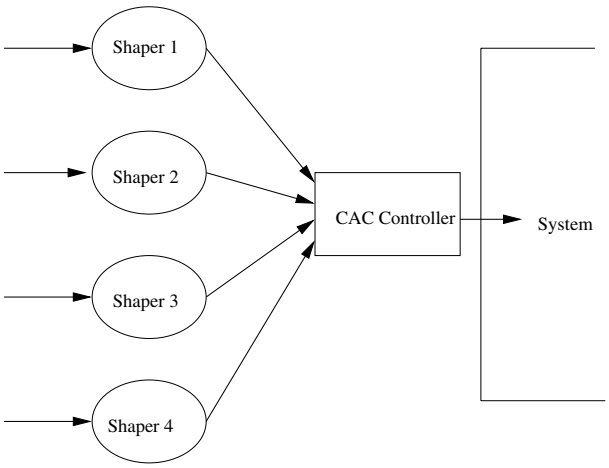

Fig. 3.8 Flows contending for admission into the system through traffic shapers and a connection admission controller.

time of the CAC controller occurs before any packet from the contending connections arrive at the controller, no connection will be admitted to the system. However, if packets from some connections arrive before the first service time of the CAC controller, then the connection whose packet arrives first will be admitted and the other connections will not be admitted.

Lemma 3.1. The total rate of admission R_a and the rate of admission for each connection R_i (normalized by the maximum total rate) are of the following GPA forms:

$$\begin{aligned} R_a &= \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{K} \lambda_k + \mu}, \\ R_i &= \frac{\lambda_i}{\sum_{k=1}^{K} \lambda_k + \mu}, \quad \forall i. \end{aligned}$$

Intuitively, the relative magnitudes of μ and $\sum_{i=1}^{K} \lambda_i$ determine the admission rate. The relative magnitudes among λ_i determine the allocation among the connections. Using GP we show how the parameters μ and $\boldsymbol{\lambda}$ can be dynamically optimized to provide a flexible control of

both the total admission rate and a constrained rate allocation among the contending connections.

Proposition 3.4. The following nonlinear problem of maximizing the admission rate for a particular connection i^*, subject to the total admission rate constraint $R_{a,max}$, the QoS constraints of guaranteed rate $R_{i,min}$ for each connection i, and the range constraints on variables $\boldsymbol{\lambda}$ and μ, is a GP:

$$\begin{array}{ll} \mathsf{maximize} & R_{i^*}(\boldsymbol{\lambda}, \mu) \\ \mathsf{subject\ to} & R_a(\boldsymbol{\lambda}, \mu) \leq R_{a,max} \\ & R_i(\boldsymbol{\lambda}, \mu) \geq R_{i,min}, \quad \forall i \\ & \lambda_{i,max} \geq \lambda_i \geq \lambda_{i,min}, \quad \forall i \\ & \mu_{max} \geq \mu \geq \mu_{min} \\ \mathsf{variables} & \boldsymbol{\lambda}, \mu. \end{array} \tag{3.49}$$

Although the first constraint in problem (3.49) is an upper bound instead of the lower bounds on inverted posynomials as in (3.48), it is still equivalent to a posynomial upper bound $(1 - R_{a,max}) \left(\sum_{j=1}^{K} \lambda_j \mu^{-1} + 1 \right) \leq 1$ due to the specific GPA structure in this case.

Proposition 3.5. The following nonlinear problem of maximizing the admission rate for the worst-case connection is a GP:

$$\begin{array}{ll} \mathsf{maximize} & \min_i R_i(\boldsymbol{\lambda}, \mu) \\ \mathsf{subject\ to} & \text{Same constraints as in problem (3.49)} \\ \mathsf{variables} & \boldsymbol{\lambda}, \mu. \end{array} \tag{3.50}$$

Note that that the parameters $\lambda_{i,min}, \lambda_{i,min}, \mu_{max}$ and μ_{max} determine the ranges over which λ_i and μ can vary. Larger the $\lambda_{i,max}$, higher the rate connection i could be allowed to receive under the constrained optimization. The parameters $\lambda_{i,min}$ and $\lambda_{i,max}$ can be found through a lookup table that maps the QoS class of connection i to the range of λ_i allowed. The rate $R_{i,min}$ guaranteed for each connection can be read from the traffic descriptor of the connection.

As a numerical example, we consider a scenario where there are five connections contending to get admitted into the system. Problem parameter $R_{a,max}$ is determined based on the congestion condition of the network, and is periodically updated. The connection characteristics and minimum admission requirements are shown in Table 3.2. The connection admission controller varies $\mu : 0 \leq \mu \leq 1$ and the traffic shapers vary $\boldsymbol{\lambda} : \lambda_{i,min} \leq \lambda_i \leq \lambda_{i,max}$ to control the total system admission rate and each individual connection's admission rate by solving a GP.

Connection	$\lambda_{i,min}$	$\lambda_{i,max}$	$R_{i,min}$
1	0.21875	0.37500	0.15625
2	0.18750	0.31250	0.09375
3	0.25000	0.40625	0.15625
4	0.28125	0.43750	0.21875
5	0.09375	0.18750	0.06250

Table 3.2 Arrival traffic bounds and minimum rate requirements.

Figures 3.9 and 3.10 present simulation results illustrating how $(\boldsymbol{\lambda}, \mu)$ are dynamically optimized. The data points for each time instance in the graphs are obtained by solving a corresponding GP. In Figure 3.9, $(\boldsymbol{\lambda}, \mu)$ are chosen such that connection 1 admission rate is the largest possible while ensuring that the minimum admission rate requirements $R_{i,min}$ are met for all other connections, and that the total admission rate does not exceed the maximum rate $R_{a,max}$ allowed by the system. Connection 1 is always favored over the other connections whenever possible under the QoS constraints. For instance, although connection 4 has a higher minimum admission rate requirement than connection 1, connection 1 is admitted more often.

In Figure 3.10, $(\boldsymbol{\lambda}, \mu)$ are chosen to maximize the minimum admission rate among all connections, using the same set of $R_{a,max}$ constraint values as in Figure 3.9. With this objective, if there were no minimum admission rate requirements, all connections would have been admitted equally. However, because different connections have different characteristics and requirements, admission rates will vary. Connections

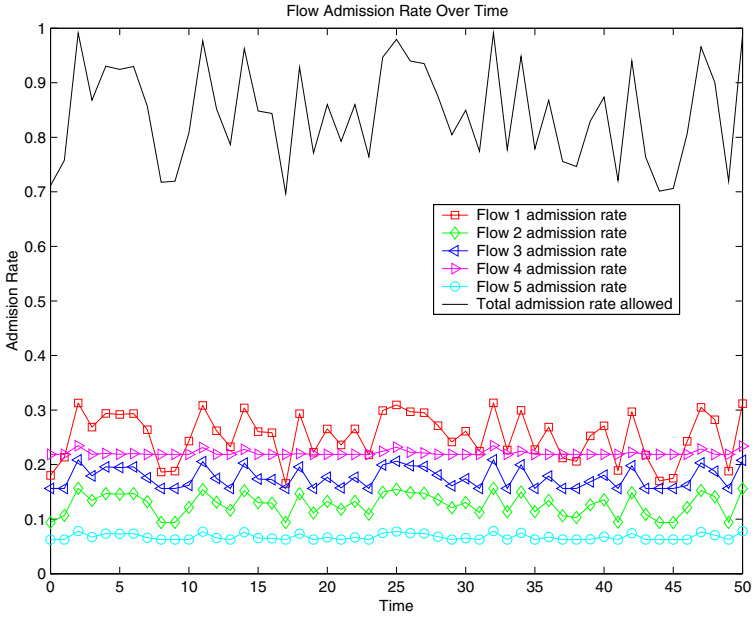

Fig. 3.9 Optimized CAC: Maximize the admission rate for connection 1.

that have relatively small minimum admission rate requirements (e.g., connections 2 and 5) are usually admitted at rates higher than requested. Intuitively, in Figure 3.10 all connections are treated as equally as possible, resulting in a narrower band of admission rate curves.

3.4 Network Congestion Control

Materials in this subsection are in part based on [31, 88].

3.4.1 TCP Vegas congestion control

Transmission Control Protocol (TCP) is one of the two widely-used transport layer protocols in the Internet. A main function performed by TCP is network congestion control and end-to-end rate allocation. Roughly speaking, there are two phases of TCP congestion control:

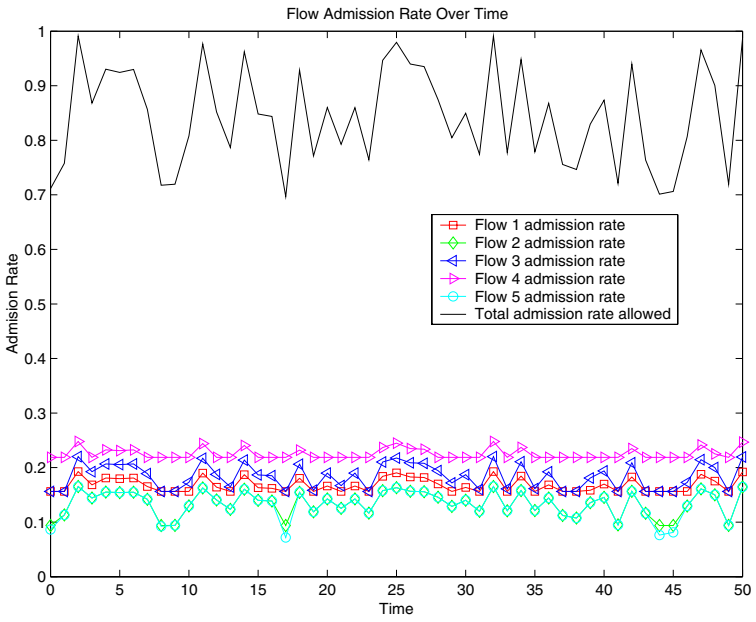

Fig. 3.10 Optimized CAC: Maximize the minimum admission rate among all connections.

slow start and congestion avoidance. Long-lived flows usually spend most of the time in congestion avoidance. Similar to recent work on utility maximization models of TCP, we assume a deterministic flow model for the average, equilibrium behavior of the congestion avoidance phase. TCP uses sliding windows to adjust the allowed transmission rate at each source based on implicit or explicit feedback of the congestion signals generated by Active Queue Management (AQM) at the routers. Among the variants of TCP, such as Tahoe, Reno, Vegas, and FAST, some use loss as the congestion signal and others use delay. Delay-based congestion signals, like those in TCP Vegas, have more desirable properties on convergence, stability, and fairness [87].

The basic rate allocation mechanism of TCP Vegas is as follows. Let d_s be the propagation delay along the path originating from source s, and D_s be the propagation plus congestion-induced queuing delay. Obviously $d_s = D_s$ when there is no congestion on all the links used

by source s. The window size w_s is updated at each source s according to whether the difference between the expected rate $\frac{w_s}{d_s}$ and the actual rate $\frac{w_s}{D_s}$, where D_s is estimated by the timing of ACK packets, is smaller than a parameter α_s:

$$w_s(t+1) = \begin{cases} w_s(t) + \frac{1}{D_s(t)} & \text{if } \frac{w_s(t)}{d_s} - \frac{w_s(t)}{D_s(t)} < \alpha_s \\ w_s(t) - \frac{1}{D_s(t)} & \text{if } \frac{w_s(t)}{d_s} - \frac{w_s(t)}{D_s(t)} > \alpha_s \\ w_s(t) & \text{else.} \end{cases}$$

The end-to-end throughput for each path is the allowed source rate x_s, which is proportional to the window size: $x_s(t) = \frac{w_s(t)}{D_s(t)}$.

Since the seminal work by Kelly *et al.* [77] that analyzes network rate allocation as a distributed solution of utility maximization, TCP congestion control mechanisms have been shown as approximated distributed algorithms solving appropriately formulated utility maximization problems (e.g., [79, 81, 85, 86, 87, 88, 89, 95, 96, 117]). A central approach in this series of work is to interpret source rates as *primal variables*, link congestion measures as *dual variables*, and a TCP–AQM protocol as a distributed algorithm over the Internet implicitly to solve the following network utility maximization.

Consider a wired communication network with L links, each with a fixed capacity of c_l bps,[6] and S sources, each transmitting at a source rate of x_s bps. Each source emits one flow, using a fixed set $L(s)$ of connected links in its path, and has an increasing, strictly concave, and twice differentiable utility function $U_s(x_s)$. Network utility maximization is the problem of maximizing the total utility $\sum_s U_s(x_s)$ over the source rates $\mathbf{x}$, subject to linear flow constraints $\sum_{s:l\in L(s)} x_s \le c_l$ for all links l:

$$\begin{array}{ll} \mathsf{maximize} & \sum_s U_s(x_s) \\ \mathsf{subject\ to} & \sum_{s:l\in L(s)} x_s \le c_l, \quad \forall l \\ & \mathbf{x} \succeq 0 \\ \mathsf{variables} & \mathbf{x}. \end{array} \tag{3.51}$$

Different TCP-AQM protocols solve for different strictly concave utility functions using different types of congestion signals. For example, TCP Vegas is shown [88] to be implicitly solving (3.51) for weighted

[6] Capacity in terms of attainable throughput with a given code and modulation, not the information theoretic limit of channel capacity.

logarithmic utility functions: $U_s(x_s) = \alpha_s d_s \log x_s$, using queuing delays as congestion signals. Although TCP and AQM protocols were designed and implemented without regard to utility maximization, now they can be *reverse-engineered* to determine the underlying utility functions and to rigorously characterize many important equilibrium and dynamic properties.

In particular, TCP Vegas has been reverse-engineered as a distributed algorithm implicitly solving the following logarithmic network utility maximization problem:

$$\begin{array}{ll} \mathsf{maximize} & \sum_s \alpha_s d_s \log x_s \\ \mathsf{subject\ to} & \sum_{s:l\in L(s)} x_s \le c_l, \quad \forall l \\ & \mathbf{x} \succeq 0 \\ \mathsf{variables} & \mathbf{x}. \end{array} \tag{3.52}$$

This Vegas problem is readily seen to be a GP as in Extension 6 in Subsection 2.2.1. As discussed in Subsection 2.3.3 and the next subsection, this GP can be distributively solved by a Lagrangian decomposition of the coupling (flow conservation) constraint. TCP Vegas congestion control solves this GP using queuing delay as the dual variables, i.e., the link congestion prices.

It is also known [77] that the weighted log utility functions implicitly maximized by TCP Vegas lead to a weighted proportionally fair allocation of link capacities, i.e., given the optimizer $\mathbf{x}^*$ to (3.52), the following inequality holds for any $\mathbf{x}$ that satisfies the constraint of (3.52):

$$\sum_s \alpha_s d_s \frac{x_s - x_s^*}{x_s^*} \le 0.$$

3.4.2 Jointly optimal congestion and power control

Consider a wireless network with multihop transmission and interference-limited link rates. Congestion control mechanisms, such as those in TCP, regulate the allowed source rates so that the total traffic load on any link does not exceed the available capacity. At the same time, the attainable data rates on wireless links depend on the interference levels, which in turn depend on the power control policy.

This subsection describes a distributed algorithm for *jointly* optimal end-to-end congestion control and per-link power control. The algorithm utilizes the coupling between the transport and physical layers to increase end-to-end throughput and to enhance energy efficiency in a wireless multihop network.

This presents a step towards understanding 'layering' as 'optimization decomposition', where the overall communication network is modeled by a generalized utility maximization problem, each layer corresponds to a decomposed subproblem, and the interfaces among layers are quantified as functions of primal or dual variables coordinating the subproblems. In the case of the transport and physical layers, link congestion prices turn out to be the optimal 'layering prices'.

Network utility maximization problems are linearly constrained by link capacities that are assumed to be fixed quantities. However, network resources can sometimes be allocated to change link capacities. This formulation of network utility maximization with 'elastic' link capacities [31, 99, 131] leads to a new approach of congestion avoidance in wireless multihop networks. The current approach of congestion control in the Internet is to avoid the development of a bottleneck link by reducing the allowed transmission rates from all the sources using this link. Intuitively, an alternative approach is to build, in real time, a larger transmission 'pipe' and 'drain' the queued packets faster on a bottleneck link. Indeed, a smart power control algorithm would allocate just the 'right' amount of power to the 'right' nodes to alleviate the bottlenecks, which may then induce an increase in end-to-end TCP throughput. But there are two major difficulties in making this idea work: defining which link constitutes a 'bottleneck' *a priori* is difficult, and changing the transmit power on one link also affects the data rates available on other links. Due to interference in wireless networks, increasing the capacity on one link reduces those on other links. We need to find an algorithm that distributively and adaptively detects the 'bottleneck' links and optimally 'shuffles' them around in the network.

Consider a wireless multihop network with N nodes and an established logical topology and routing, where some nodes are sources of

transmission and some nodes act as 'voluntary' relay nodes. A sequence of connected links $l \in L(s)$ forms a route originating from source s. Let x_s be the transmission rate of source s, and c_l be the capacity on logical link l. Revisiting the utility maximization formulation (3.51), for which TCP congestion control solves, we observe that in an interference-limited wireless network, data rates attainable on wireless links are not fixed numbers $\mathbf{c}$ as in (3.51), and instead can be written, in the high SIR regime, as a global and nonlinear function of the transmit power vector $\mathbf{P}$ and channel conditions: $c_l(\mathbf{P}) = \frac{1}{T}\log(K\text{SIR}_l(\mathbf{P}))$.

With the above model, we have specified the following network utility maximization with 'elastic' link capacities and logarithmic utilities:

$$\begin{array}{ll} \text{maximize} & \sum_s \alpha_s d_s \log x_s \\ \text{subject to} & \sum_{s:l\in L(s)} x_s \le c_l(\mathbf{P}), \ \ \forall l \\ & \mathbf{x}, \mathbf{P} \succeq 0 \\ \text{variables} & \mathbf{x}, \mathbf{P}. \end{array} \tag{3.53}$$

The key difference from the standard utility maximization (3.51) is that each link capacity c_l is now a function of the new optimization variables: the transmit powers $\mathbf{P}$. The design space is enlarged from $\mathbf{x}$ to both $\mathbf{x}$ and $\mathbf{P}$, which are clearly coupled in (3.53). Linear flow constraints on $\mathbf{x}$ become nonlinear constraints on $(\mathbf{x}, \mathbf{P})$. In practice, problem (3.53) is also constrained by the maximum and minimum transmit powers allowed at each transmitter on link l: $P_{l,min} \le P_l \le P_{l,max}$, $\forall l$.

In the context of wireless ad hoc networks, new distributed algorithms are needed to solve (3.53). Thus the major challenges are the two global dependencies:

- Source rates $\mathbf{x}$ and link capacities $\mathbf{c}$ are globally coupled across the network, as reflected in the range of summation $\{s : l \in L(S)\}$ in the constraints in (3.53).
- Each link capacity $c_l(\mathbf{P})$ is a global function of all the interfering powers.

The following distributive algorithm converges to the joint and global optimum of (3.53).

Jointly Optimal Congestion-control and Power-control (JOCP) Algorithm [31]:
During each time slot t, the following four updates are carried out simultaneously until convergence:

(1) At each intermediate node, a weighted queuing delay λ_l is implicitly updated,[7] where $\gamma > 0$ is a constant weight:

$$\lambda_l(t+1) = \left[\lambda_l(t) + \frac{\gamma}{c_l(t)}\left(\sum_{s:l\in L(s)} x_s(t) - c_l(t)\right)\right]^+. \quad (3.54)$$

(2) At each source, total delay D_s is measured and used to update the TCP window size w_s. Consequently, the source rate x_s is updated:

$$w_s(t+1) = \begin{cases} w_s(t) + \frac{1}{D_s(t)} & \text{if } \frac{w_s(t)}{d_s} - \frac{w_s(t)}{D_s(t)} < \alpha_s \\ w_s(t) - \frac{1}{D_s(t)} & \text{if } \frac{w_s(t)}{d_s} - \frac{w_s(t)}{D_s(t)} > \alpha_s \\ w_s(t) & \text{else.} \end{cases} \quad (3.55)$$

$$x_s(t+1) = \frac{w_s(t+1)}{D_s(t)}.$$

(3) Each transmitter j calculates a message $m_j(t) \in \mathbf{R}_+$ based on locally measurable quantities, and passes the message to all other transmitters by a flooding protocol:

$$m_j(t) = \frac{\lambda_j(t)\mathrm{SIR}_j(t)}{P_j(t)G_{jj}}.$$

(4) Each transmitter updates its power based on locally measurable quantities and the received messages, where $\kappa > 0$ is a constant weight:

$$P_l(t+1) = P_l(t) + \frac{\kappa\lambda_l(t)}{P_l(t)} - \kappa\sum_{j\neq l} G_{lj}m_j(t). \quad (3.56)$$

[7] This is using an average model for deterministic fluids. The difference between the total ingress flow intensity and the egress link capacity, divided by the egress link capacity, gives the average time that a packet needs to wait before being sent out on the egress link.

With the maximum and minimum transmit power constraint $(P_{l,max}, P_{l,min})$ on each transmitter, the updated power is projected onto the interval $[P_{l,max}, P_{l,min}]$.

We first present some intuitive arguments on this algorithm before proving the convergence theorem and discussing the practical implementation issues. Item (2) is simply the TCP Vegas window update [23]. Item (1) is a modified version of queuing delay price update [88] (and the original update [23] is an approximation of item (1)). Items (3) and (4) describe a new power control using message passing [32]. Taking in the current values of $\frac{\lambda_j(t)\mathrm{SIR}_j(t)}{P_j(t)G_{jj}}$ as the messages from other transmitters indexed by j, the transmitter on link l adjusts its power level in the next time slot in two ways: first increase power directly proportional to the current price (e.g., queuing delay in TCP Vegas) and inversely proportional to the current power level, then decrease power by a weighted sum of the messages from all the other transmitters, where the weights are the path losses G_{lj}. Intuitively, if the local queuing delay is high, transmit power should increase, with more moderate increase when the current power level is already high. If queuing delays on other links are high, transmit power should decrease in order to reduce interference on those links.

Note that to compute m_j, the values of queuing delay λ_j, signal-interference-ratio SIR_j, and received power level P_jG_{jj} can be directly measured by node j locally. This algorithm only uses the resulting message m_j but not the individual values of $\lambda_j, \mathrm{SIR}_j, P_j$ and G_{jj}. Each message is simply a real number.

The known source algorithm (3.55) and queue algorithm (3.54) of TCP-AQM, together with the new power control algorithm (3.56), form a set of distributed, joint congestion control and resource allocation in wireless multihop networks. As the transmit powers change, SIR and thus data rate also change on each link, which in turn change the congestion control dynamics. At the same time, congestion control dynamics change the dual variables $\boldsymbol{\lambda}(t)$, which in turn change the transmit powers. Figure 3.11 shows this nonlinear coupling of 'supply' (regulated by power control) and 'demand' (regulated by congestion control), through the same shadow prices $\boldsymbol{\lambda}$ that

are currently used by TCP to regulate distributed demand. Now $\boldsymbol{\lambda}$ serves the second function of cross-layer coordination in the JOCP Algorithm.

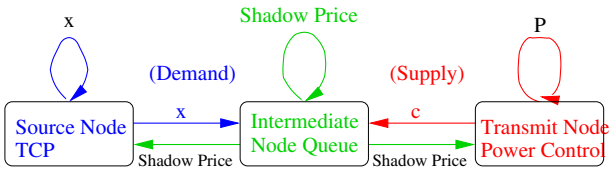

Fig. 3.11 Nonlinearly coupled dynamics of joint congestion and power control.

Notice that there is no need to change the existing TCP congestion control and queue management algorithms. All that is needed to achieve the joint and global optimum of (3.53) is to utilize the values of weighted queuing delay in designing power control algorithm in the physical layer.[8]

The advantage of such a joint control can be captured in a small illustrative example, where the logical topology and routes for four multi-hop connections are shown in Figure 3.12. Sources at each of the four flows use TCP Vegas window updates with α_s ranging from 3 to 5. The path losses G_{ij} are determined by the relative physical distances d_{ij}, which we vary in different experiments, by $G_{ij} = d_{ij}^{-4}$. The target BER is 10^{-3} on each logical link.

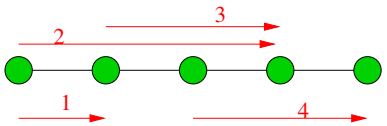

Fig. 3.12 The logical topology and routes in an illustrative example.

Transmit powers, as regulated by the proposed distributed power control, and source rates, as regulated through TCP Vegas window

[8] This approach is complementary to some recent suggestions in the Internet community to pass physical layer information for a better control of routing and congestion in upper layers.

update, are shown in Figure 3.13. The initial conditions of the graphs are based on the equilibrium states of TCP Vegas with fixed power levels of 2.5mW. With power control, the transmit powers $\mathbf{P}$ distributively adapt to induce a 'smart' capacity $\mathbf{c}$ and queuing delay $\boldsymbol{\lambda}$ configuration in the network, which in turn lead to increases in end-to-end throughput as indicated by the rise in all the allowed source rates. Notice that some link capacities actually decrease while the capacities on the bottleneck links rise to maximize the total network utility. This is achieved through a distributive adaptation of power, which lowers the power levels that cause most interference on the links that are becoming a bottleneck in the dynamic demand-supply balancing process. Confirming our intuition, such a 'smart' allocation of power tends to reduce the spread of queuing delays, thus preventing any link from becoming a bottleneck. Queuing delays on the four links do not become the same though, due to the asymmetry in traffic load on the links and different weights in the logarithmic utility objective functions. The end-to-end throughput per watt of power transmitted, i.e., the throughput-power ratio, is 82% higher with power control.

We can associate a Lagrange multiplier λ_l for each of the constraints $\sum_{s:l\in L(s)} x_s \leq c_l(\mathbf{P})$. To find the stationary points of the Lagrangian, we need to solve the following Lagrangian maximization problems: $I_{system}(\mathbf{x}, \mathbf{P}, \boldsymbol{\lambda}) = (\sum_s \alpha_s d_s \log x_s - \sum_l \lambda_l \sum_{s:l\in L(s)} x_s) + (\sum_l \lambda_l c_l(\mathbf{P}))$. By linearity of the differentiation operator, this can be decomposed into two separate maximization problems:

$$\begin{aligned} \mathsf{maximize}_{\mathbf{x}\succeq 0} \quad & \sum_s \alpha_s d_s \log x_s - \sum_s \sum_{l\in L(s)} \lambda_l x_s \\ \mathsf{maximize}_{\mathbf{P}\succeq 0} \quad & \sum_l \lambda_l c_l(\mathbf{P}). \end{aligned}$$

Both maximization problems are readily verified to be GPs, one in $\mathbf{x}$ and another in $\mathbf{P}$. The first maximization is already implicitly solved by the TCP Vegas congestion control mechanism. But we still need to solve the second maximization, using the Lagrange multipliers $\boldsymbol{\lambda}$ as the shadow prices to allocate exactly the right power to each transmitter. Since the data rate on each wireless link is a global function of all the

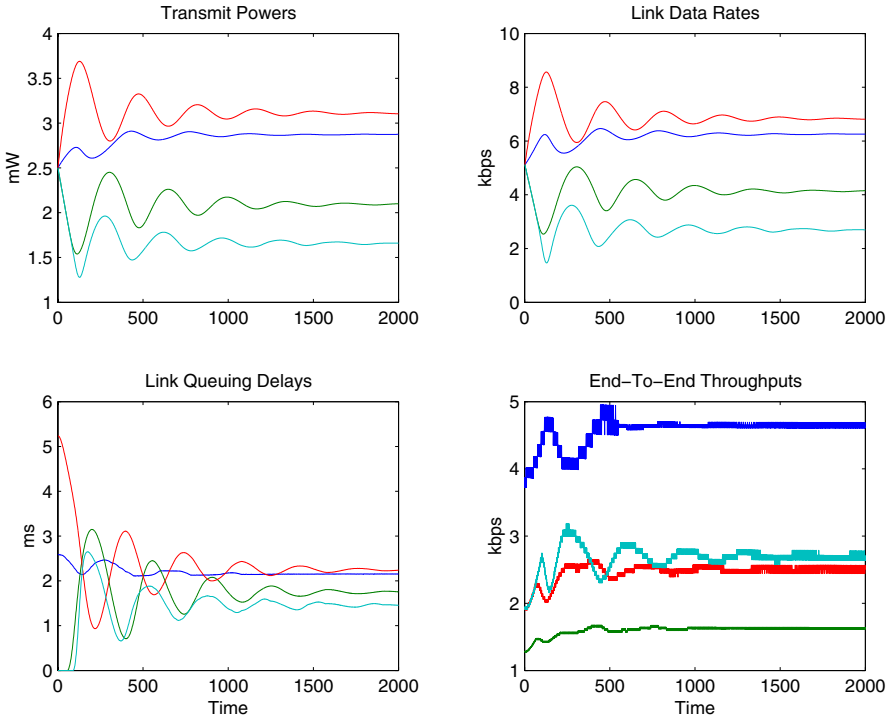

Fig. 3.13 A typical numerical example of joint TCP Vegas congestion control and power control. The top left graph shows the primal variables $\mathbf{P}$. The lower left graph shows the dual variables $\boldsymbol{\lambda}$. The lower right graph shows the primal variables $\mathbf{x}$, i.e., the end-to-end throughput. In order of their y-axis values after convergence, the curves in the top left, top right, and bottom left graphs are indexed by the third, first, second, and fourth links in Figure 2. The curves in the bottom right graph are indexed by flows 1, 4, 3, 2.

transmit powers, the power control problem cannot be nicely decoupled into local problems for each link as in [131]. More message passing in needed.

Convergence of the nonlinearly coupled system, formed by the JOCP Algorithm and illustrated in Figure 3.11, is guaranteed under two mild assumptions. First, P_l are within a range between $P_{l,min} > 0$ and $P_{l,max} < \infty$ for each link l. Second, when link prices are high enough, source rates can be made very small: for any $\epsilon > 0$, there exists a λ_{max} such that if $\lambda_l > \lambda_{max}$, then $x_s(\boldsymbol{\lambda}) < \epsilon$ for all sources s that use link l. Appendix B.6 proves the following

Theorem 3.5. For small enough constants γ and κ, the distributed JOCP Algorithm (3.54,3.55,3.56) converges to the global optimum of the joint congestion control and power control problem (3.53).

Extensions to other TCP variants and multi-commodity flow routing are discussed in [31]. Other properties of this distributed algorithm of GP are shown: the convergence is geometric, can be maintained under finite asynchronism, is robust to estimation error of path loss and to packet loss due to channel fading, and can be accelerated by passing slightly more information among the nodes. Simplified heuristics for partial messaging are also possible.

3.5 Queuing Theory

Materials in this subsection are in part based on [36, 68, 76].

Queuing systems form a fundamental part for different types of communication systems, such as computer multiprocessor networks and communications data networks. Queuing systems are also an integral part of various network elements, such as the input and output buffers of a packet switch. We often would like to optimize some performance metrics of queuing systems, for example, buffer occupancy, overall delay, jittering, workload, and probabilities of certain states. However, optimizing the performance of even simple queues like the $M/M/m/m$ queue is in general a difficult problem because of the nonlinearity of the performance metrics as functions of the arrival and service rates.

We show how convexity properties of queuing systems, in the form of posynomials (that appear as product form distributions) and log-sum-exp functions (that appear in effective bandwidth formulations), can be used to turn some of these problems into polynomial time solvable ones. We provide a suite of GP formulations to efficiently and globally optimize the performance of queuing systems under QoS constraints, first for single Markovian queues, then for blocking probability minimization and service rate allocation through the effective bandwidth approach, and finally for closed networks of queues.

First consider a simple example of minimizing the service load of an $M/M/1$ queue, over the arrival rate λ and service rate μ, with

constraints on average queuing delay W, total delay D, and queue occupancy Q. The following proposition can be readily verified through basic formulas of $M/M/1$ queues [15]:

Proposition 3.6. The following nonlinear optimization of service load minimization is a GP:

$$\begin{array}{ll} \text{minimize} & \frac{\mu}{\lambda} \\ \text{subject to} & W \le W_{max} \\ & D \le D_{max} \\ & Q \le Q_{max} \\ & \lambda \ge \lambda_{min} \\ & \mu \le \mu_{max} \\ \text{variables} & \lambda, \mu. \end{array} \tag{3.57}$$

The constant parameters are the performance upper bounds W_{max}, D_{max} and Q_{max}, and practical constraints on the maximum service rate μ_{max} of the queue that cannot be exceeded, and the minimum incoming traffic rate λ_{min} that must be supported.

The above formulation can be extended to a Markovian queuing system with N queues sharing a pool of service rate bounded by μ_{max} (for example, connected to a common egress link). The arrival rate to be supported for each individual queue i is bounded by $\lambda_{i,min}$. There are delay and queue occupancy bounds $W_{i,max}$, $D_{i,max}$ and $Q_{i,max}$ for each queue i. The objective now becomes minimizing a weighted sum of the service loads for all the queues, with constant weights $\boldsymbol{\alpha}$:

Corollary 3.4. The following nonlinear minimization of a weighted sum of service loads is a GP:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} \alpha_i \frac{\mu_i}{\lambda_i} \\ \text{subject to} & W_i \le W_{i,max}, \ \ \forall i \\ & D_i \le D_{i,max}, \ \ \forall i \\ & Q_i \le Q_{i,max}, \ \ \forall i \\ & \lambda_i \ge \lambda_{i,min}, \ \ \forall i \\ & \sum_{i=1}^{N} \mu_i \le \mu_{max} \\ \text{variables} & \boldsymbol{\lambda}, \boldsymbol{\mu}. \end{array} \tag{3.58}$$

We now optimize specific queue occupancy probabilities of an $M/M/m/m$ queue. The steady state probability of state k is given by

$$p_k = \frac{(\frac{\lambda}{\mu})^k \frac{1}{k!}}{\sum_{i=0}^{m} (\frac{\lambda}{\mu})^i \frac{1}{i!}}.$$

In many applications of queuing systems to network design, we would like to maximize the probability of a particular desirable state, without making the probabilities of other states too small. For example, we may want to design a telephone call service center so as to maximize the probability that a particular number of telephone lines (e.g., 90%) are in use at any given time.

Proposition 3.7. The following nonlinear optimization of $M/M/m/m$ queues is a GP:

$$\begin{array}{ll} \mathsf{maximize} & p_k(\lambda, \mu) \\ \mathsf{subject\ to} & p_j(\lambda, \mu) \geq C_j, \quad \forall j \\ & \lambda \geq \lambda_{min} \\ & \mu \leq \mu_{max} \\ \mathsf{variables} & \lambda, \mu. \end{array} \tag{3.59}$$

The constant parameters are λ_{min}, μ_{max} and $C_j, j = 1, 2, \ldots, m$.

One approach to study the buffer overflow probability is through the blocking probability of an $M/M/1/B$ queue with a fixed buffer of size B:

$$p_B = \frac{(\frac{\lambda}{\mu})^B \frac{1}{B!}}{\sum_{i=0}^{B} (\frac{\lambda}{\mu})^i \frac{1}{i!}}.$$

Minimizing p_B is equivalent to maximizing a posynomial of λ and μ, which is in turn equivalent to maximizing a convex function. Therefore, minimizing blocking probability cannot be turned into a GP. One possible heuristic is to use GP to maximize the probability of some state $k, k < B$, subject to lower bounds on p_j for all other $j < B$. Since $p_B = 1 - \sum_{i=0}^{B-1} p_i$, this heuristic essentially minimizes the blocking probability.

An alternative way to characterize buffer overflow is through the large deviation approach, where the blocking probability is guaranteed statistically: for a connection X with a prescribed service rate R in the queue, we would like to ensure that the probability of overflow (receiving more than R bps from X) over a time scale of t is exponentially small:

$$\mathbf{Prob}\left\{\sum_{i=1}^{t} X(i) \geq R\right\} \leq \exp(-sR) \tag{3.60}$$

where $s \geq 0$ is the under-subscription factor. Smaller s implies more aggressive statistical multiplexing of multiple connections to one queue. This number R is called the effective bandwidth EB of X (as first proposed in [68], used in many papers since, and nicely reviewed in [76]).

Using the Chernoff bound, the effective bandwidth is given by

$$\mathrm{EB}(X) = \frac{1}{st} \log \mathbf{E}\left[\exp(sX)\right]. \tag{3.61}$$

In practice, the expectation is replaced by the empirical average of traffic data over a time period of $\tilde{t}$ that is much larger than the time scale factor t:

$$\mathrm{EB}(X) = \frac{1}{st} \log \left(\frac{t}{\tilde{t}} \sum_{i=1}^{\frac{\tilde{t}}{t}} \exp(sX(i)) \right)$$

where $X(i)$ is the number of bits sent by connection X during the ith time slot.

In traffic engineering, we want to either minimize the assigned service rate $\mathrm{EB}(X)$ subject to constraints that lower bound the traffic intensity $\{X(i)\}$ to be supported (i.e., exponentially small probability of overflow or blocking), or maximize the traffic intensity subject to constraints upper bounding the service rate that can be assigned to X. Both problems can be formulated as GPs, and we focus on the first formulation here. Constraints on the minimal level of traffic intensity to be supported by $\mathrm{EB}(X)$ are indexed by j and induced by the stochasticity of other connections sharing the queue buffer.

Proposition 3.8. The following nonlinear problem of constrained buffer allocation is a GP:

$$\begin{array}{ll} \text{minimize} & \mathrm{EB}(X) \\ \text{subject to} & \sum_i P_{ij} X(i) \geq X_{min,j}, \ \ \forall j \\ \text{variables} & X(i), \ \ \forall i. \end{array} \tag{3.62}$$

The constant parameters are P_{ij} and $X_{min,j}$.

A numerical example is summarized as follows. With $s = 0.5, t =$ 5ms, we impose a set of 10 different constraints to specify the type of an arrival curve a queue should be able to support without blocking. The GP solution returns the minimized effective bandwidth as $\mathrm{EB}^*(X) =$ 1.7627Mbps, and the envelope of supportable arrival curves is shown in Figure 3.14. Connections with arrival curves below this envelope will not cause buffer overflow or queue blocking with a probabilistic guarantee as in (3.60).

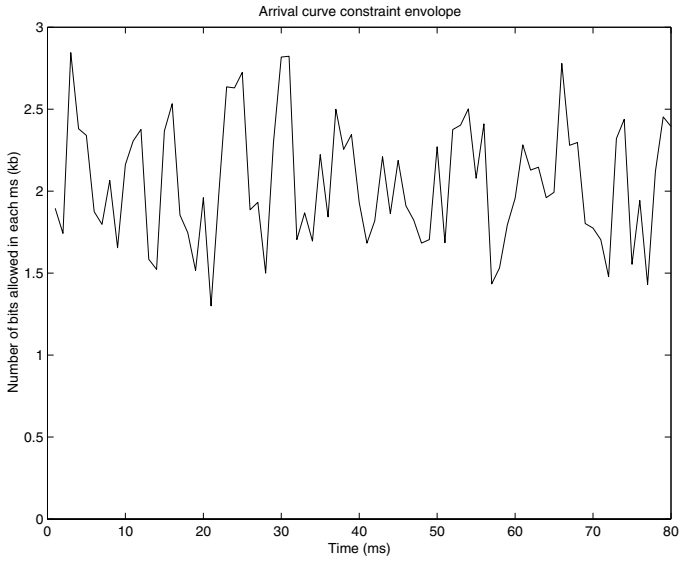

Fig. 3.14 The envelope of arrival curves supportable by $\mathrm{EB}^*(X) = 1.7627$.

In some queuing problems, a fixed number of customers or tasks circulate indefinitely in a closed network of queues. For example, some

computer system models assume that at any given time a fixed number of programs occupy the resource. Such problems can be modeled by a closed queuing network consisting of K nodes, where each node k consists of m_k identical exponential servers, each with an average service rate μ_k. There are always exactly N customers in the system. Once served at node k, a customer goes to node j with probability p_{kj}. Then for each node k, the average arrival rate to the node, λ_k, is given by $\lambda_k = \sum_{j=1}^{K} p_{kj}\lambda_j$.

The steady state probability that there are n_k customers in node k, for $k = 1, 2, \ldots, K$, is given by a closed network Jackson's theorem:

$$\mathbf{Prob}(n_1, n_2, \ldots, n_K) = \frac{1}{G(K)} \prod_{k=1}^{K} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^{n_k}}{\beta_k(n_k)},$$

where

$$\beta_k(n_k) = \begin{cases} n_k! & n_k \leq m_k \\ m_k! m_k^{n_k - m_k} & n_k > m_k, \end{cases}$$

and the normalization constant $G(K)$ is given by

$$G(K) = \sum_{\mathbf{s}} \prod_{k=1}^{K} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^{n_k}}{\beta_k(n_k)}$$

where the summation over $\mathbf{s}$ is taken over all state vectors $\mathbf{s} = (n_1, n_2, \ldots, n_K)$ satisfying $\sum_{k=1}^{K} n_k = N$.

Proposition 3.9. The following nonlinear problem of maximizing the probability of state $(n_1 = n_1^*, \ldots, n_K = n_K^*)$ with $\sum_{k=1}^{K} n_k = N$, subject to constraints on other states, is a GP:

$$\begin{array}{ll} \mathsf{maximize} & \mathbf{Prob}(n_1 = n_1^*, \ldots, n_K = n_K^*) \\ \mathsf{subject\ to} & \mathbf{Prob}(n_1, \ldots, n_K) \geq \text{Constant} \\ & \mu_{k,min} \leq \mu_k \leq \mu_{k,max}, \quad \forall k \\ & \sum_{k=1}^{K} m_k \mu_k \leq \mu_{total} \\ \mathsf{variables} & \boldsymbol{\mu} \end{array} \tag{3.63}$$

where there is a constraint of the first type for each steady state probability $\mathbf{Prob}(n_1, \ldots, n_K)$. The constant parameters are $\mu_{k,min}, \mu_{k,max}, m_k,\ k = 1, 2, \ldots, K$, and μ_{total}.

The above optimization problem can be viewed as a problem of resource (i.e., service capacity $\boldsymbol{\mu}$) allocation in a closed queuing network. The goal is to maximize the probability that the system is in a particular state subject to constraints on other states and the limited system resource μ_{total}.

4

Why Is Geometric Programming Useful for Communication Systems

The last section summarizes GP's applications to information theory and queuing theory, coding and signal processing, network resource allocation and protocols. This section initiates an exploration of the plausible reasons why GP is useful to such a variety of topics in the analysis and design of communication systems.

4.1 Stochastic Models

Materials in this subsection are in part based on [30, 42, 45, 52, 108].

Many problems in information theory, queuing theory, and coding are based on stochastic models of communication systems, where the probability of an undesirable event is to be bounded or minimized. For example, given a family of conditional distributions describing a channel, we would like the probability of decoding error to vanish exponentially as the codeword length goes to infinity. Or given a queuing discipline and arrival and departure statistics, we would like the probability of buffer overflow to vanish exponentially as the buffer size increases.

Large deviation principles govern such exponential behavior in stochastic systems. It is well known that convex analysis is closely

related to large deviation principles [45, 127]. In Subsection 4.1.2 we will sharpen this general connection by showing that several important large deviation bounds can be computed by GP. It is this connection between GP and large deviation characterizations that is the underlying reason for GP's applicability to communication system problems based on stochastic models.

This discussion is first motivated in Subsection 4.1.1 by examining the relationship between GP and classical statistical physics. Such relationships were recognized since the 1960s [52] and is not surprising as one can easily realize that the log-sum-exp function, which can be the objective or constraint functions of a convex form GP, is the log partition function of the Boltzmann distribution. A posynomial can be transformed into not just a convex function, but also one in the 'right' convexity structure with interesting interpretations from statistical physics.

4.1.1 Interpretations from statistical physics

Consider a system governed by classical statistical physics with n states at temperature T, where each state i has energy e_i and probability p_i of occurring. Given an energy vector $\mathbf{e}$ and a probability vector $\mathbf{p}$, the average energy is $U(\mathbf{p}, \mathbf{e}) = \mathbf{p}^T\mathbf{e}$ and the entropy is $H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i$. The Gibbs free energy is defined as a weighted difference between the two:

$$G(\mathbf{p}, \mathbf{e}) = U(\mathbf{p}, \mathbf{e}) - TH(\mathbf{p}) = \mathbf{p}^T\mathbf{e} + T\sum_{i=1}^{n} p_i \log p_i.$$

Solving the problem of Gibbs free energy minimization:

$$\begin{array}{ll} \mathsf{minimize} & \mathbf{p}^T\mathbf{e} + T\sum_{i=1}^{n} p_i \log p_i \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \mathsf{variables} & \mathbf{p} \end{array} \tag{4.1}$$

with constant parameters $\mathbf{e}$, is important in statistical physics with several interpretations, e.g., (4.1) strikes a balance between energy minimization and entropy maximization.

Following the argument in the discussion of the conjugacy relationship between log-sum-exp and negative entropy in Subsection 2.1.1, it is easy to see that the Boltzmann distribution, denoted by $\tilde{\mathbf{b}}$, minimizes $G(\mathbf{p}, \mathbf{e})$ over $\mathbf{p}$ for a given energy vector $\mathbf{e}$, where $\tilde{b}_i$ is proportional to $\exp(-\frac{e_i}{T})$. The proportionality constant needed for normalization is called the partition function $Z(\mathbf{e}) = \sum_{i=1}^{n} \exp(\frac{-e_i}{T})$. The Gibbs free energy $G(\mathbf{p}, \mathbf{e})$ induced by the Boltzmann distribution $\mathbf{p} = \tilde{\mathbf{b}}$ is called the Helmholtz free energy $F(\mathbf{e})$, which is the negative logarithm of $Z(\mathbf{e})$ scaled by T:[1]

$$F(\mathbf{e}) = G(\tilde{\mathbf{b}}, \mathbf{e}) = -T \log \sum_{i=1}^{n} \exp(-\frac{e_i}{T}).$$

Due to the convexity of the Gibbs free energy in $\mathbf{p}$ and the concavity of the Helmholtz free energy in $\mathbf{e}$, $\max_{\mathbf{e}} \min_{\mathbf{p}} G(\mathbf{p}, \mathbf{e}) = \min_{\mathbf{p}} \max_{\mathbf{e}} G(\mathbf{p}, \mathbf{e})$. Therefore, maximizing the Helmholtz free energy is equivalent to finding the minimum Gibbs free energy for the worst-case energy vector.

The Gibbs free energy can also be generalized as follows [108]. Consider a multiple phase chemical system with K phases and J types of substances. Let n_{jk} be the number of atomic weights of substance j that are in phase k, and let e_{jk} be the energy of substance j in phase k, $j = 1, 2, \ldots, J,\ k = 1, 2, \ldots, K$. The multiphase equilibrium problem is to minimize the following generalized Gibbs free energy with unit temperature over $\{n_{jk}\}$:

$$\sum_{j,k} n_{jk} e_{jk} + \sum_{j,k} n_{jk} \log \left(\frac{n_{jk}}{\sum_{j'} n_{j'k}} \right). \tag{4.2}$$

Now suppose the distribution on the states is not the Boltzmann distribution $\tilde{\mathbf{b}}$, but some general distribution $\mathbf{q}$. In this case, we will get a corresponding value for the Gibbs free energy $G(\mathbf{q}, \mathbf{e})$, and the

[1] With a logarithmic transformation of variables: $\tilde{e}_i = -\log p_i,\ \forall i$, it is easy to see that the Renyi entropy (3.33) is a log-sum-exp function. If the order of the Renyi entropy is $1/T$, the Renyi entropy is a scaled Helmholtz free energy:

$$H_\gamma(\mathbf{p}) = \frac{1}{T-1} F(\tilde{\mathbf{e}}).$$

difference between this value and the Helmholtz free energy, normalized by the temperature T, is given by:

$$\frac{1}{T}\left(G(\mathbf{q},\mathbf{e})-G(\tilde{\mathbf{b}},\mathbf{e})\right)=\frac{\mathbf{q}^T\mathbf{e}}{T}-H(\mathbf{q})-\frac{F(\mathbf{e})}{T}=\sum_{i=1}^{n}q_i\log\frac{q_i}{\tilde{b}_i}=D(\mathbf{q}\|\tilde{\mathbf{b}}). \tag{4.3}$$

Therefore, another way to derive the Boltzmann distribution is through minimizing the difference between a general Gibbs free energy and the Helmholtz free energy, expressed as a KL divergence (relative entropy), over the probability simplex. Minimizing KL divergence $D(\mathbf{q}\|\tilde{\mathbf{b}})$ over $\mathbf{q}$ is precisely the dual objective function of an unconstrained GP. Indeed, we can rewrite the dual objective in (2.7) as a KL divergence minimization: minimize$_{\boldsymbol{\nu}} D(\boldsymbol{\nu}\|\mathbf{d})$, or in the exponentiated form: maximize$_{\boldsymbol{\nu}} \prod_i \left(\frac{d_i}{\nu_i}\right)^{\nu_i}$ where $b_i=\log d_i,\ \forall i$.

In general, the KL divergence $D(\mathbf{q}_1\|\mathbf{q}_2)$ between $\mathbf{q}_1$ and $\mathbf{q}_2$ is the Gibbs free energy $G(\mathbf{q}_1,\mathbf{e})$ where the energy vector is the negative log likelihood of $\mathbf{q}_2$: $e_i=-\log q_{2,i},\ \forall i$. And the dual objective of an unconstrained GP (2.7) over $\boldsymbol{\nu}$ is equivalent to minimizing the Gibbs free energy $G(\boldsymbol{\nu},-\mathbf{b})$ at unit temperature.

In summary, GP in convex form is equivalent to a constrained Helmholtz free energy maximization problem, the dual problem of GP is equivalent to a linearly-constrained generalized Gibbs free energy minimization problem, and the dual problem of unconstrained GP is equivalent to the Gibbs free energy minimization.

Now recall some of the GP applications to information theory. Shannon [113] considered communication as a problem of reproducing an i.i.d. stochastic source (with N alphabet symbols) at the destination. If each alphabet symbol in the source has probability p_i of appearing, and the string is n symbols long, then for large n, the probability of a typical string is approximately $\prod_{i=1}^{N} p_i^{np_i}=e^{-nH(\mathbf{p})}$, and to the first order in the exponent, the number of typical sequences is $K=e^{nH(\mathbf{p})}=\prod_{i=1}^{N}p_i^{-np_i},\ \forall i$. It turns out that K is also the exponentiated objective function of the dual problem of an unconstrained GP: if we let the constants $x_i=n,\forall i$, and the variables $z_i=np_i$, then $K=\prod_{i=1}^{N}\left(\frac{x_i}{z_i}\right)^{z_i}$. Therefore, maximizing the number of typical sequences is Lagrange dual to this unconstrained GP.

For the primal problem of rate distortion (3.14), [14] shows that minimizing the Lagrangian of rate distortion is a Gibbs free energy minimization problem. Furthermore, as can be readily verified, the Lagrange dual problem (3.16) of $R(D)$ is minimizing an average energy under the Helmholtz free energy constraints, where the energy of state i in the jth constraint is $e_{ij} = \gamma d_{ij} - \alpha_i - \log p_i$.

For channel capacity problems in Subsection 3.1.1, we note that the primal problem of channel capacity without input cost is a generalized Gibbs free energy minimization, where each state i has energy r_i, temperature is unity, average energy $U = \mathbf{p}^T\mathbf{r}$ is on the input distribution, but entropy is on the output distribution $\mathbf{q}$ induced by the input distribution and the channel. Minimizing the Lagrangian $\mathbf{p}^T(\mathbf{r}+\mathbf{s}) + \sum_{j=1}^{M} q_j \log q_j$ of channel capacity with input cost is still a Gibbs free energy minimization problem, with the energy for each state i increased by the input cost s_i. The Lagrange dual problem (3.6) of $C(S)$ is a Helmholtz free energy maximization problem under average energy constraints: energy for each state is $-\alpha_i$, the objective is to maximize the Helmholtz free energy $F(-\boldsymbol{\alpha})$, and the average energy constraints are $\sum_{j=1}^{M} P_{ij}(-\alpha_j) \le r_i, i = 1, 2, \ldots, N$.

Turning to lossless source coding problems in Subsection 3.2.2, the Huffman code is known to be optimal for the following family of problems with exponential codeword length penalty parameterized by $\beta \ge 0$ [25]:

$$\text{minimize} \quad \left[\frac{1}{\beta}\log\sum_i p_i 2^{\beta l_i}\right]. \tag{4.4}$$

As $\beta \to \infty$, this problem reduces to a minmax redundancy problem [50]. The Huffman code gives the maximizing energy vector for all tilted Helmholtz free energy minimization problems of the form (4.4).

4.1.2 Large deviation bounds

Large deviation principles [45] characterize the limiting behavior of a family of probability measures $\{\mu_\epsilon\}$ on a probability space $\mathcal{S}$ and the associated Borel field $\mathcal{B}$ in terms of rate functions. A family of probability measures $\{\mu_\epsilon\}$ satisfies the large deviation principle with a

rate function I if, for all $\Gamma \in \mathcal{B}$ such that the closure of Γ is its interior, we have:

$$\lim_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) = \inf_{\mathbf{z} \in \Gamma} I(\mathbf{z}). \tag{4.5}$$

Large deviation principles have been used in bandwidth allocation and admission control algorithms (e.g., [127]) for communication systems, and their relationship with the method of types is also well understood [41, 42].

Continuing with the discussion on statistical physics, we would like to bound the probability that the n-sample average energy $\bar{e}$ exceeds a given threshold τ, for $\tau > U(\tilde{\mathbf{b}}, \mathbf{e})$. This probability is upper bounded by

$$\mathbf{Prob}\{\bar{e} \geq \tau\} \leq \exp(-n(\lambda\tau - \Lambda(\lambda))) \tag{4.6}$$

where $\Lambda(\lambda), \lambda \geq 0$, is the log moment generating function:

$$\Lambda(\lambda) = \log \sum_i \tilde{b}_i \exp(\lambda e_i).$$

We can define a tilted Boltzmann distribution $\tilde{b}(\lambda)$ parameterized by λ:

$$\tilde{b}_i(\lambda) = \frac{1}{Z(\lambda, \mathbf{e})} \tilde{b}_i \exp(\lambda e_i), \quad \forall i$$

where $Z(\lambda, \mathbf{e}) = \exp(\Lambda(\lambda))$ is the partition function.

Optimizing the bound (4.6) over $\lambda \geq 0$, we obtain the Chernoff exponent $E_c(\tau)$:[2]

$$E_c(\tau) = \max_{\lambda \geq 0} [\lambda\tau - \Lambda(\lambda)], \tag{4.7}$$

and the Chernoff bound:

$$\mathbf{Prob}\{\bar{e} \geq \tau\} \leq \exp(-nE_c(\tau)).$$

The optimization problem (4.7) is a dual problem of GP: KL divergence minimization under a linear constraint. Specifically, we can form the Lagrangian of the problem of minimize$_{\mathbf{q}:\mathbf{q}^T\mathbf{e} \geq \tau} D(\mathbf{q}||\tilde{\mathbf{b}})$ as:

$$L(\mathbf{q}, \lambda) = D(\mathbf{q}||\tilde{\mathbf{b}}) + \lambda(\tau - \mathbf{q}^T\mathbf{e}),$$

[2] This function $E_c(\tau)$ can be viewed as the conjugate function of $\Lambda(\lambda)$.

and then minimize L over $\mathbf{q}$ to obtain the optimal $\mathbf{q}^*$ as a tilted Boltzmann distribution with a tilted inverse temperature $1-\lambda$. This gives the Lagrange dual function

$$g(\lambda) = \min_{\mathbf{q}} L(\mathbf{q}, \lambda) = \lambda\tau + F(\lambda, \mathbf{e})$$

where $F(\lambda, \mathbf{e})$ is a tilted Helmholtz free energy.[3] Since $F(\lambda, \mathbf{e}) = -\Lambda(\lambda)$, it follows that the dual problem is $\max_{\lambda\geq 0}[\lambda\tau - \Lambda(\lambda)]$.

Therefore, by strong duality for GP in convex form, we have

$$\max_{\lambda\geq 0}[\lambda\tau - \Lambda(\lambda)] = \min_{\mathbf{q}:\mathbf{q}^T\mathbf{e}\geq\tau} D(\mathbf{q}||\tilde{\mathbf{b}}). \tag{4.8}$$

Note that $\mathbf{Prob}\{\bar{e} \geq \tau\}$ is also lower bounded by the probability of an n-tuple of type $\mathbf{q}$:

$$\mathbf{Prob}\{\bar{\mathbf{e}} \geq \tau\} \geq \min_{\mathbf{q}:\mathbf{q}^T\mathbf{e}\geq\tau} \mathbf{Prob}\{\mathbf{q}|\tilde{\mathbf{b}}\},$$

and recall the following simple result from the method of types. Let $\mathbf{q}$ be an arbitrary distribution on a discrete alphabet $\{1, 2, \ldots, M\}$. As $n \to \infty$, the probability of all sequences of type $\mathbf{q}$[4] under a given distribution $\mathbf{p}$ on the source is known to be:

$$\mathbf{Prob}\{\mathbf{q}|\mathbf{p}\} = \exp(-nD(\mathbf{q}||\mathbf{p})). \tag{4.9}$$

The Chernoff bound (4.6) can now be seen to be exponentially tight, because (4.6) and (4.9) imply that

$$\min_{\mathbf{q}:\mathbf{e}^T\mathbf{q}\geq\tau} D(\mathbf{q}||\tilde{\mathbf{b}}) \geq \lim_{n\to\infty} -\frac{1}{n}\log \mathbf{Prob}\{\bar{\mathbf{e}} \geq \tau\} \geq \max_{\lambda\geq 0}[\lambda\tau - \Lambda(\lambda)],$$

which, together with (4.8), implies that

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbf{Prob}\{\bar{e} \geq \tau\} = E_c(\tau).$$

[3] A tilted Helmholtz free energy $F\left(\lambda, \frac{\mathbf{e}}{1-\lambda}\right)$ also covers Gallager's function $E_0(\rho)$ [61]:

$$E_0(\rho) = -\log \sum_i \left(\sum_j x_j \left(P_{ij}\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

in the random coding error exponent as a special case, by substituting $\lambda = \frac{\rho}{1+\rho}$.

[4] An n-tuple vector $\mathbf{x}$ is called of type $\mathbf{q}$ if the number of appearances of x_i is equal to nq_i, $\forall i$.

Next, consider random variables $X_1, X_2, \ldots, X_n$, with a finite alphabet size $|\mathcal{X}| = N$ and i.i.d. $\sim \mathbf{q}$. Let Γ be a set of probability measures such that it is the closure of its interior. Applying the method of types bound as above, the following Sanov's theorem can be readily proved [40]:

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbf{q}(\Gamma) = \min_{\mathbf{p}\in\Gamma} D(\mathbf{p}||\mathbf{q}).$$

If the set Γ can be represented through linear constraints on $\mathbf{p}$, the asymptotics $\min_{\mathbf{p}\in\Gamma} D(\mathbf{p}||\mathbf{q})$ is a dual problem of GP (in the form of Gibbs free energy minimization). For some special Γ, analytic closed form solutions can also be obtained. For example, let $\Gamma = \{\mathbf{p} : \sum_i p_i g(x_i) \geq \alpha\}$ where g is a given deterministic function and α a given scalar. It can be verified that the $\mathbf{p}^*$ that minimizes $D(\mathbf{p}||\mathbf{q})$ over $\mathbf{p} \in \Gamma$ is the tilted Boltzmann distribution:

$$p_i^* = \tilde{b}_i(\lambda) = \frac{1}{Z(\lambda, \mathbf{x})} q_i \exp(\lambda g(x_i)), \quad \forall i$$

where $Z(\lambda, \mathbf{x})$ is the partition function, and λ is the Lagrange multiplier that makes $\mathbf{p}^*$ satisfy the inequality in the definition of Γ.

Now consider n i.i.d. random variables $\{Y_j\}$ with a probability measure $\boldsymbol{\mu} \in \mathbf{R}_+^m$. Let $X_j = g(Y_j)$ for a deterministic scalar-valued function g, and consider the empirical mean $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$. By applying Sanov's theorem, the following Cramer's theorem is obtained for finite subsets of $\mathbf{R}$ [45]:

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\mathop{\mathbf{Prob}}_{\boldsymbol{\mu}}\{\bar{X}_n \in \Gamma\}\right) = -\inf_{z\in\Gamma} I(z)$$

where the rate function is

$$I(z) = \sup_{\lambda\in\mathbf{R}} [\lambda z - \Lambda(\lambda)],$$

$\Lambda(\lambda)$ is the log moment generating function:

$$\Lambda(\lambda) = \log \sum_{i=1}^m \mu_i \exp(\lambda g(y_i)),$$

and $\{y_i\}$ are the values taken by each random variable Y_j.

Cramer's theorem can be interpreted as giving an exponentially tight bound $I(\mathbf{z})$ on the deviation of the empirical mean of i.i.d. random variables through the following GP (in the form of minimizing a KL divergence, or the conjugate of a tilted Helmholtz free energy):

$$I(z) = \min_{\boldsymbol{\nu}:\mathbf{g}^T\boldsymbol{\nu}=z} D(\boldsymbol{\nu}||\boldsymbol{\mu}),$$

where $\mathbf{g} = [g(y_1), \ldots, g(y_m)]$.

This connection can also be illustrated by rewriting the exponential bound as

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log\left(\underset{\mu}{\mathbf{Prob}}\{\bar{X}_n \in \Gamma\}\right) &= -\inf_{z\in\Gamma}\left[\sup_{\lambda\in\mathbf{R}} [\lambda z - \Lambda(\lambda)]\right] \\ &= -\sup_{\lambda\in\mathbf{R}}\left[\inf_{z\in\Gamma}[\lambda z] - \Lambda(\lambda)\right] \\ &= \inf_{z\in\Gamma,\lambda\in\mathbf{R}} \left[\Lambda(\lambda) - \lambda z\right],\end{aligned}$$

where the last expression shows the symmetry in minimizing over both z and λ through a GP.

The approach of using exponential asymptotics usually relies on the assumption that the underlying random variables are i.i.d. However, this assumption can be relaxed to a Markovity assumption. This is useful for problems in communication systems where the events or traffic may not be i.i.d. but can be adequately modeled as systems with finite memory, since a random process with finite memory can be turned into a Markov process by exponentially increasing the state space size.

Let $\{Y_l\}$ be a finite state Markov chain with m states and an irreducible transition matrix $\mathbf{A}$, $X_l = g(Y_l)$ for a deterministic d-valued function g, and $\bar{X}_n = \frac{1}{n}\sum_{l=1}^n X_l$. For every $\mathbf{z} \in \mathbf{R}^d$, define

$$I(\mathbf{z}) = \sup_{\boldsymbol{\lambda}\in\mathbf{R}^d}\left[\boldsymbol{\lambda}^T\mathbf{z} - \log\rho(\mathbf{A}_{\boldsymbol{\lambda}})\right] \tag{4.10}$$

where $\rho(\mathbf{A}_{\boldsymbol{\lambda}})$ is the Perron–Frobenius eigenvalue of matrix $\mathbf{A}_{\boldsymbol{\lambda}}$ with the (i,j)th entry being $A_{ij}\exp(\boldsymbol{\lambda}^T g(y_j))$. It is known [45] that $\bar{X}_n$ satisfies the large deviation principle with rate function I.

It is apparent that $\log \rho(\mathbf{A}_\lambda)$ in the Markov case plays the role of $\Lambda(\lambda)$ in the i.i.d. case. More precisely, the following proposition [30] can be proved as in Appendix B.7:

Proposition 4.1. The rate function $I(\mathbf{z})$ in (4.10) can be evaluated at any given $\mathbf{z}$ as the optimized value of the following GP:

$$\begin{array}{ll} \text{minimize} & \rho \prod_{k=1}^{d} \tilde{\lambda}_k^{-z_k} \\ \text{subject to} & \sum_{j=1}^{m} (\rho\nu_i)^{-1} (A_{ij}) \nu_j \prod_{k=1}^{d} \tilde{\lambda}_k^{g_k(y_j)} \leq 1, \quad \forall i \\ \text{variables} & \tilde{\boldsymbol{\lambda}}, \boldsymbol{\nu}, \rho \end{array} \tag{4.11}$$

where $\tilde{\lambda}_k = e^{\lambda_k}$. The constant parameters include the arguments of the rate function $\mathbf{z}$, the transition matrix $\mathbf{A}$, and the values of the deterministic functions $\{g_k(y_j)\}$, $j = 1, \ldots, m$, $k = 1, \ldots, d$.

Additional constraints of the form $f(\tilde{\boldsymbol{\lambda}}) \leq 1$ can be added to problem (4.11), where f is a posynomial function to be specified depending on the application of this large deviation principle. For example, if f is a monomial, then $f(\tilde{\boldsymbol{\lambda}}) = 1$ is equivalent to a linear constraint on $\boldsymbol{\lambda}$.

4.2 Deterministic Models

Materials in this subsection are in part based on [28, 49, 30, 70].

Unlike in the case of problems based on stochastic models, there is not a single most convincing explanation of why GP has been found to be useful for problems in communication systems based on deterministic models. Three plausible explanations are provided in this subsection: GP is useful for deterministic models when network resources are allocated to competing users in a proportional way, or when there is a general market equilibrium with linear utilities, or when the system is designed according to the Highly Optimized Tolerance theory.

4.2.1 Proportional allocation

In Subsection 3.3.2 we have seen that GP in standard form can be used to solve network resource allocation problems when the objective function to be maximized, and the constraint functions to be lower bounded, can be written as inverted posynomials in the Generalized Proportional

Allocation (GPA) form (3.47). GPA form allocation includes various examples as special cases, including power control in the high SIR regime and some admission control, rate control, and queuing system optimization problems. It turns out that the GPA form of allocation can be interpreted as the solution to an implicit optimization of an information theoretic quantity.

Relative Entropy Minimization (REM) is a convex optimization problem in the following form:[5]

$$\begin{array}{ll} \text{minimize} & D(\mathbf{p}\|\mathbf{x}) + \alpha(\boldsymbol{\nu}^T\mathbf{x}) \\ \text{subject to} & \mathbf{A}\mathbf{x} \preceq \mathbf{v} \\ & \mathbf{x} \succeq 0 \\ \text{variables} & \mathbf{x}. \end{array} \tag{4.12}$$

The constant parameters are $\mathbf{p}, \mathbf{A}, \mathbf{v}, \boldsymbol{\nu}, \alpha \succeq 0$. Both $\mathbf{x}$ and $\mathbf{p}$ are non-negative vectors that may not be normalized.

As can be easily verified, REM can also be written as an Extended GP (Extension 11 in Subsection 2.2.3) of minimizing a posynomial and the log of a monomial in variables $\mathbf{x}$:

$$\begin{array}{ll} \text{minimize} & \alpha(\boldsymbol{\nu}^T\mathbf{x}) + \log \prod_i x_i^{-p_i} \\ \text{subject to} & \mathbf{A}\mathbf{x} \preceq \mathbf{v} \\ & \mathbf{x} \succeq 0 \\ \text{variables} & \mathbf{x}. \end{array} \tag{4.13}$$

Proposition 4.2. The solution of a REM problem (4.12) is in the GPA form (3.47).

This proposition can be readily proved as in Appendix B.8. For every REM problem (4.12), there corresponds a proportional allocation in the parameterized form of (3.47). However, there can be many REM problems whose solutions are in the form of a given proportional allocation.

[5] The network utility model in Kelly [77] is a special case of REM where $\alpha = 0$ and $\mathbf{A}$ is a $0-1$ matrix denoting the routing decisions.

4.2.2 General market equilibrium

Many network resource allocation methods, especially the network utility maximization formulation of congestion control, are closely related with the economics theory of general market equilibrium. It has recently been shown that in the case of linear utility, the Arrow–Debreu market equilibrium can be obtained by a Mixed Linear Geometric Program as introduced in Subsection 2.2.1, with implications to the set of equilibria and to the complexity of computing a market equilibrium.

In 1874, Walras [126] introduced the following problem. Consider a model of market where every person has an initial endowment of divisible goods. If there exits a price vector, which assigns every good a price, such that it is possible for every person to sell the initial endowment and buy an optimal bundle of goods with the entire revenue, then such a price vector is called a general market equilibrium. A special case was independently proposed by Fischer [22] in 1891, where there are two kinds of people, producers who have initial endowments of goods and want to earn money, and consumers who have money and want to maximize utilities for goods. A market equilibrium is set of prices assigned to the goods so that when every consumer buys an optimal bundle then the market clears, i.e., all the money is spent and all the goods are sold.

Arrow and Debreu [5] in 1954 showed that, when the utility functions are concave, a general market equilibrium exists. Since then, many researchers have studied numerical methods for computing a market equilibrium in polynomial time. For the case of linear utilities in the Fischer model, Eisenberg and Gale [55] provided a convex optimization formulation to obtain a market equilibrium, which was solved by an ellipsoid method together with diophantine approximation. Alternatively, the convex optimization problem can be solved by interior-point methods and the result be extended to nonlinear concave utilities in the Fischer model. For the more general Walras model, there was no polynomial time algorithm and many approximation heuristics were developed for various special cases. A recent paper [70] provides the first polynomial time algorithm for linear utilities in the Walras model. The key idea is to show the equivalence between the set of general

market equilibria and the feasible set of a nonconvex optimization problem, which turns out to be a MLGP and can be turned into a convex problem. We follow the development in [70] in the rest of this subsection.

There are n people and m goods. We index a person by i and a good by j. Without loss of generality, assume that each person is endowed with one unit of good, and has a utility function that is assumed to be linear in the Walras model. Person i has a following linear utility function:

$$\sum_j u_{ij} x_{ij}$$

where u_{ij} is a constant parameter and x_{ij} is the amount of good j consumed by person i. The Arrow–Debreu Theorem [5] states that there exists a price vector $\mathbf{p} \neq 0$, called the general market equilibrium, such that buying and selling can be done to clear the market.

Without loss of generality, we assume that everyone likes something, i.e., for every i, there is a j such that $u_{ij} > 0$, and that there is an equilibrium where every price p_i is non-zero. The first assumption is easily seen to incur no loss of generality. The case for the second assumption is more involved, and a proof that there exists an equilibrium where no price is zero can be found in [70].

Consider the following nonconvex feasibility problem:

$$\begin{array}{ll} \mathsf{maximize} & \text{No Objectives} \\ \mathsf{subject\ to} & \sum_i x_{ij} = 1, \ \ \forall j \\ & \frac{p_i}{p_j} \leq \frac{\sum_k u_{ik} x_{ik}}{u_{ij}}, \ \ \forall i, j \\ & x_{ij} \geq 0, \ \ \forall i, j \\ & p_i > 0, \ \ \forall i \\ \mathsf{variables} & \{x_{ij}\}, \mathbf{p}. \end{array} \tag{4.14}$$

The constraints of $\sum_i x_{ij} = 1$ and $x_{ij} \geq 0$ state that $\{x_{ij}\}$ is a feasible assignment of goods to people. The constraint $p_i > 0$ comes from the assumption of strictly positive prices. The most interesting constraint is the second one, and as shown in Appendix B.9, the following theorem can be proved [70].

Theorem 4.1. The set of general market equilibria is the same as the set of feasible solutions of problem (4.14).

It is readily recognized that (4.14) is a MLGP feasibility problem, and thus can be transformed into a convex optimization problem by taking the logarithm of prices $\mathbf{p}$. Let $\tilde{p}_i = \log p_i, \ \forall i$. Then the feasibility problem (4.14) can be written as the following convex feasibility problem:

$$\begin{array}{ll} \mathsf{maximize} & \text{No Objectives} \\ \mathsf{subject\ to} & \sum_i x_{ij} = 1, \ \ \forall j \\ & \exp(\tilde{p}_i - \tilde{p}_j) \leq \frac{\sum_k u_{ik} x_{ik}}{u_{ij}}, \ \ \forall i, j \\ & x_{ij} \geq 0, \ \ \forall i, j \\ \mathsf{variables} & \{x_{ij}\}, \{\tilde{p}_i\} \end{array} \tag{4.15}$$

This result immediately implies the following:

Corollary 4.1. The set of general market equilibria's assignments of goods to people is convex. The set of general market equilibria's log prices $\tilde{\mathbf{p}}$ is convex.

Theorem 4.1 also leads to the conclusion that either the ellipsoid or interior-point method provides a polynomial time algorithm to compute a general market equilibrium for linear utilities in the general Walras model. Extensions to nonlinear concave utilities are also presented in [70].

4.2.3 Generalized source coding

In [49], a generalized source coding problem is proposed as an example of a class of optimization problems called the probability-loss-resource (PLR) problem, which is equivalent to the family of source coding problems with either linear or exponential penalty function in Subsection 3.2.2, and thus a special case of GP.

PLR problem refers to the following minimization of expected loss $\mathbf{E_p}[\mathbf{l}]$ under a given probability distribution $\mathbf{p}$, over a global and linear constraint $R \geq 0$ on resources $\mathbf{r}$:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i p_i l_i \\ \mathsf{subject\ to} & l_i = f_\beta(r_i), \ \ \forall i \\ & \sum_i r_i \leq R \\ & \mathbf{r} \succeq 0 \\ \mathsf{variables} & \mathbf{r}, \mathbf{l} \end{array} \tag{4.16}$$

where the resource vs. loss function f_β is parameterized by a scalar parameter $\beta \geq 0$:

$$f_\beta(r_i) = \begin{cases} -c \log(r_i), & \beta = 0 \\ \frac{c}{\beta}\left(r_i^{-\beta} - 1\right), & \beta > 0, \end{cases} \tag{4.17}$$

so that $f_\beta(1) = 0$, and the marginal loss per unit resource decreases proportional to $r_i^{-(\beta+1)}$ for $\beta \geq 0$ with proportionality constant c. Problem (4.16) can be solved analytically just as in the case of source coding with linear or exponential penalty functions.

It is shown in [49] that the empirical data on the distributions of file sizes on the Internet and of sizes of forest fires match the optimal solutions of PLR problems very well, with $\beta = 1$ and $\beta = 2$ respectively. Explanations of this remarkable fit of PLR problem with data from apparently unrelated fields are provided in [49], in the context of the Highly Optimized Tolerance (HOT) framework that explains phenomena in complex systems. HOT provides an alternative, with significantly different implications, to the Self Organized Criticality (SOC) framework [2, 9] in statistical physics to explain power law distributions. In SOC theory, complexity is emphasized as emerging between order and disorder at a phase transition in an interconnection of components and otherwise largely random. The details associated with the initiation of events would be a statistically inconsequential factor and large events would be the result of random internal fluctuations. In contrast, HOT systems arise when deliberate robust design aims for a specific level of tolerance to uncertainty, which is traded-off against the cost of the compensating resources. Optimization of this trade-off, possibly through solving the PLR problem, leads to high performance

and high throughput, ubiquitous power law distributions of event sizes, and potentially high sensitivities to design flaws and unanticipated perturbations. It is argued in [49, 27] that many mechanisms in complex systems are implicitly solving a PLR problem with different $\beta, c, \mathbf{p}, R$ constants.

It is readily seen that when $\beta = 0$, the PLR problem is equivalent to the following GP:

$$\begin{array}{ll} \mathsf{minimize} & \prod_i r_i^{-cp_i} \\ \mathsf{subject\ to} & \sum_i r_i \leq R \\ \mathsf{variables} & \mathbf{r}, \end{array}$$

and when $\beta > 0$, the PLR problem is equivalent to another GP:

$$\begin{array}{ll} \mathsf{minimize} & \sum_i \frac{cp_i}{\beta} r_i^{-\beta} \\ \mathsf{subject\ to} & \sum_i r_i \leq R \\ \mathsf{variables} & \mathbf{r}. \end{array}$$

Therefore, according to the HOT theory, the class of GP with single-variable monomial terms in the objective and linear constraints provide a prototype problem for which processes in complex systems are implicitly solving.

A

History of Geometric Programming

In 1961, Zener published a seminal paper [132] in the *Proceedings of the National Academy of Sciences*, and observed that some engineering design problems can be formulated as optimization of 'generalized polynomials', and that if the number of terms exceed the number of variables by one, the optimal design can be found by solving a system of linear equations. In 1967, Duffin, Peterson, and Zener published the book *Geometric Programming: Theory and Applications* [52] that started the field of GP as a branch of nonlinear optimization.

Several important developments of GP took place in the late 1960s and 1970s. GP was tied with convex optimization and Lagrange duality, and was extended to include more general formulations beyond posynomials. Several numerical algorithms to solve GP were proposed and tested (e.g., in [47]). And researchers in mechanical engineering, civil engineering, and chemical engineering found successful applications of GP to their problems.

There are several books on nonlinear optimization that have a section on GP, e.g., [8] in 1973, [57] in 1997, [59] in 1999, and [21] in 2004. In addition, there have also been at least five books devoted to GP. Other than a very recent one [26], the other four were published

during the late 1960s and 1970s: the pioneering book by Duffin, Peterson, and Zener [52] in 1967, two follow-up books [133, 10] in 1971 and 1976, respectively, and a book of collected papers [6] in 1980. The one in 1980 summarized major developments in GP since its initiation in 1967, and contained a comprehensive list of over 120 papers published on the theory, algorithms, and applications of GP up to 1980. There were also three *SIAM Review* papers surveying the status of GP at the time of their publications: [51] in 1970, [103] in 1976, and [54] in 1980.

However, as researchers felt that most of the theoretical, algorithmic and application aspects of GP had been exhausted by the early 1980s, the period of 1980–1998 was relatively quiet. Over the last few years, however, GP started to receive renewed attention from the operations research community, for example, in a special issue of *Annals of Operations Research* in 2000 and a special session in the INFORMS annual meeting in 2001. In particular, we now have very efficient algorithms to solve GP, either by general purpose convex optimization solvers [97, 21], or by more specialized methods [78]. Approximation methods for robust GP and distributed algorithms for GP have also appeared recently.

New and surprising applications of GP have also been found recently by the electrical engineering and computer science communities, including some that are very different from GP's traditional applications. Two areas of applications have been particularly prominent. One is digital and analog circuit design, since the mid-1980s and especially since the late 1990s [44, 66]. Another is GP applications in communication systems, the subject of this text, since the mid-1990s and especially over the last five years. Insights on why GP is useful for communication systems have also been obtained.

B

Some Proofs

B.1 Proof of Theorem 3.1

Proof. In order to find the Lagrange dual of problem (3.2), we first form the Lagrangian L as

$$L(\mathbf{p}, \mathbf{q}, \boldsymbol{\nu}, \mu, \boldsymbol{\lambda}, \gamma) = -\mathbf{p}^T\mathbf{r} - \sum_j q_j \log q_j + (\mathbf{q} - \mathbf{P}^T\mathbf{p})^T\boldsymbol{\nu} + \mu(1 - \mathbf{1}^T\mathbf{p}) + \mathbf{p}^T\boldsymbol{\lambda} + \gamma(S - \mathbf{p}^T\mathbf{s}) \quad \text{(B.1)}$$

with Lagrange multiplier vector $\boldsymbol{\nu} \in \mathbf{R}^{M\times 1}$, Lagrange multiplier $\mu, \gamma \in \mathbf{R}$, and Lagrange multiplier vector $\boldsymbol{\lambda} \in \mathbf{R}^{N\times 1}$. Since $\boldsymbol{\lambda}$ and γ correspond to the inequality constraints, we have $\boldsymbol{\lambda} \succeq 0$ and $\gamma \geq 0$.

We then find the Lagrange dual function $g(\boldsymbol{\nu}, \mu, \boldsymbol{\lambda}, \gamma) = \sup_{\mathbf{p},\mathbf{q}} L(\mathbf{p}, \mathbf{q}, \boldsymbol{\nu}, \mu, \boldsymbol{\lambda}, \gamma)$ by finding the $\mathbf{p}$ and $\mathbf{q}$ that maximize L, which is a concave function of $(\mathbf{p}, \mathbf{q})$. First, note that L is a linear function of $\mathbf{p}$, thus bounded from above only when it is identically zero. As a result, $g(\boldsymbol{\nu}, \mu, \boldsymbol{\lambda}, \gamma) = \infty$ unless $\mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} - \boldsymbol{\lambda} + \gamma\mathbf{s} = 0$, which is equivalent to $\mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} + \gamma\mathbf{s} \succeq 0$ since $\boldsymbol{\lambda} \succeq 0$.

Assuming $\mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} + \gamma\mathbf{s} \succeq 0$, the Lagrangian becomes $-\sum_j (q_j \log q_j - \nu_j q_j) + \mu + \gamma S$, which we must now maximize over $\mathbf{q}$. To find the maximum of $-q_j \log q_j + \nu_j q_j$ over q_j, we set the derivative with

respect to q_j equal to zero: $\log q_j + 1 - \nu_j = 0$. Thus, $q_j = \exp(\nu_j - 1)$ is the maximizer, with the associated maximum value

$$-q_j \log q_j + \nu_j q_j = -e^{\nu_j - 1}(\nu_j - 1) + \nu_j e^{\nu_j - 1} = e^{\nu_j - 1}.$$

Therefore, the Lagrange dual function is

$$g(\nu, \mu, \gamma) = \begin{cases} \sum_j \exp(\nu_j - 1) + \mu + \gamma S, & \mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} + \gamma\mathbf{s} \succeq 0 \\ \infty, & \text{otherwise.} \end{cases} \tag{B.2}$$

By making the constraint $\mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} + \gamma\mathbf{s} \succeq 0$ explicit, we obtain the Lagrange dual problem:

$$\begin{array}{ll} \textsf{minimize} & \sum_j \exp(\nu_j - 1) + \mu + \gamma S \\ \textsf{subject to} & \mathbf{r} + \mathbf{P}\boldsymbol{\nu} + \mu\mathbf{1} + \gamma\mathbf{s} \succeq 0, \quad \gamma \geq 0 \\ \textsf{variables} & \boldsymbol{\nu}, \mu, \gamma. \end{array}$$

The constant parameters are $\mathbf{P}$.

Letting $\boldsymbol{\alpha} = \boldsymbol{\nu} + \mu\mathbf{1}$, and using the fact $\mathbf{P}\mathbf{1} = \mathbf{1}$, we rewrite the dual problem as

$$\begin{array}{ll} \textsf{minimize} & \exp(-1-\mu) \sum_j e^{\alpha_j} + \mu + \gamma S \\ \textsf{subject to} & \mathbf{r} + \mathbf{P}\boldsymbol{\alpha} + \gamma\mathbf{s} \succeq 0, \quad \gamma \geq 0 \\ \textsf{variables} & \boldsymbol{\alpha}, \mu, \gamma. \end{array}$$

Since the dual variable μ, which is the Lagrange multiplier corresponding to the primal constraint $\mathbf{1}^T\mathbf{p} = 1$, is unconstrained in the dual problem, we can minimize the dual objective function over μ analytically, and obtain the minimizing $\mu = \log \sum_j e^{\alpha_j} - 1$. The resulting dual objective function is $\log \sum_j e^{\alpha_j} + \gamma S$. The Lagrange dual problem is simplified to the following GP in convex form:

$$\begin{array}{ll} \textsf{minimize} & \log \sum_j e^{\alpha_j} + \gamma S \\ \textsf{subject to} & \mathbf{P}\boldsymbol{\alpha} + \gamma\mathbf{s} \succeq -\mathbf{r}, \quad \gamma \geq 0 \\ \textsf{variables} & \boldsymbol{\alpha}, \gamma. \end{array}$$

We can turn this GP into standard form, through an exponential change of the variables $z_j = e^{\alpha_j}$ and the dual objective function:

$$\begin{array}{ll} \textsf{minimize} & w^S \sum_j z_j \\ \textsf{subject to} & w^{s_i} \prod_j z_j^{P_{ij}} \geq e^{-H(\mathbf{P}^{(i)})}, \quad i = 1, 2, \ldots, N \\ & z_j \geq 0, \quad j = 1, 2, \ldots, M, \quad w \geq 1 \\ \textsf{variables} & \mathbf{z}, w. \end{array}$$

The constant parameters are entries of the channel matrix $\mathbf{P}$, and $\mathbf{P}^{(i)}$ is the ith row of $\mathbf{P}$.

The weak duality part of the proposition follows directly from a standard fact in Lagrange duality theory [21]: the Lagrange dual function is always an upper bound on the primal maximization problem.

It is well-known that the objective function to be maximized in the primal problem (3.2) is concave in $(\mathbf{p}, \mathbf{q})$ and the constraint functions are affine. The strong duality part of the proposition holds because the primal problem (3.1) is a convex optimization satisfying Slater's condition [21]. □

B.2 Proof of Corollary 3.1

Proof. We are given $\boldsymbol{\alpha} \in \mathbf{R}^{M\times 1} : \sum_j P_{ij}\alpha_j \geq \sum_j P_{ij} \log P_{ij}, i = 1, 2, \ldots, N$, and $\mathbf{p} \in \mathbf{R}^{N\times 1}, \mathbf{q} \in \mathbf{R}^{M\times 1} : \mathbf{p} \succeq 0, \mathbf{1}^T\mathbf{p} = 1, \mathbf{P}^T\mathbf{p} = \mathbf{q}$. Through the second derivative test, $f(t) = t \log t, t \geq 0$ is readily verified to be convex, i.e., $\sum_j \theta_j f(t_j) \geq f(\sum_j \theta_j t_j)$ with $\boldsymbol{\theta} \succeq 0, \mathbf{1}^T\boldsymbol{\theta} = 1$. Letting $t_j = \frac{q_j}{e^{\alpha_j}}$ and $\theta_j = \frac{e^{\alpha_j}}{\sum_k e^{\alpha_k}}$ and using $\mathbf{1}^T\mathbf{q} = \mathbf{1}^T\mathbf{P}^T\mathbf{p} = \mathbf{1}^T\mathbf{p} = 1$ gives $\log \sum_j e^{\alpha_j} \geq \sum_j \alpha_j q_j - \sum_j q_j \log q_j$. Since $\sum_j \alpha_j q_j = \sum_j \alpha_j \sum_i p_i P_{ij} = \sum_i p_i \sum_j P_{ij}\alpha_j \geq \sum_i p_i \sum_j P_{ij} \log P_{ij}$, we have $\log \sum_j e^{\alpha_j} \geq \sum_{i,j} p_i P_{ij} \log P_{ij} - \sum_j q_j \log q_j = I(X;Y)$, i.e., any feasible dual objective value is an upper bound on channel capacity. This proves the weak duality part of Corollary 3.1. □

B.3 Proof of Theorem 3.2

Proof. In order to find the Lagrange dual of problem (3.15), we first form the Lagrangian:

$$L(\mathbf{P}, \boldsymbol{\mu}, \gamma, \boldsymbol{\Lambda}) = \sum_{i,j} p_i P_{ij} \log \frac{P_{ij}}{\sum_k P_{kj} p_k} + \sum_i \mu_i \sum_j P_{ij} - \sum_i \mu_i + \gamma \sum_{i,j} p_i P_{ij} d_{ij} - \gamma D - \sum_{i,j} \lambda_{ij} P_{ij} \quad \text{(B.3)}$$

with Lagrange multiplier vector $\boldsymbol{\mu} \in \mathbf{R}^{N\times 1}$, Lagrange multiplier $\gamma \in \mathbf{R}$ and Lagrange multiplier matrix $\boldsymbol{\Lambda} \in \mathbf{R}^{M\times N}$, with (i, j) entry of $\boldsymbol{\Lambda}$

denoted as λ_{ij}. Since γ and $\mathbf{\Lambda}$ correspond to the inequality constraints, we have $\gamma \geq 0$ and $\lambda_{ij} \geq 0,\ i = 1, 2, \ldots, N,\ j = 1, 2 \ldots, M$.

We then find the Lagrange dual function $g(\boldsymbol{\mu}, \gamma, \mathbf{\Lambda}) = \inf_{\mathbf{P}} L(\mathbf{P}, \boldsymbol{\mu}, \gamma, \mathbf{\Lambda})$ by finding the $\mathbf{P}$ that minimizes L, which is a convex function of P_{ij}. We let the derivatives of L with respect to P_{ij} be equal to 0:

$$p_i \left[\log \left(\frac{P_{ij}}{z_i \sum_k P_{kj} p_k} \right) + \gamma d_{ij} - \frac{\lambda_{ij}}{p_i} \right] = 0$$

where $z_i = \exp(-\frac{\mu_i}{p_i})$. This gives the following condition on the minimizer $\mathbf{P}$ of L:

$$P_{ij} = z_i q_j \exp \left(\frac{\lambda_{ij}}{p_i} - \gamma d_{ij} \right) \tag{B.4}$$

where $q_j = \sum_k P_{kj} p_k$. Now multiply both sides of (B.4) by p_i, sum over i, and cancel q_j on both sides, we obtain the following condition: $\sum_i z_i p_i \exp \left(\frac{\lambda_{ij}}{p_i} - \gamma d_{ij} \right) = 1,\ \ j = 1, 2, \ldots, M$, which, by the definition of z_i and the condition $\lambda_{ij} \geq 0$, is equivalent to

$$\sum_i p_i \exp \left(-\frac{\mu_i}{p_i} - \gamma d_{ij} \right) \leq 1,\ \ j = 1, 2, \ldots, M. \tag{B.5}$$

Substituting the minimizer (B.4) and the condition (B.5) into L (B.3), we obtain the Lagrange dual function

$$g(\boldsymbol{\mu}, \gamma) = \begin{cases} -\sum_i \mu_i - \gamma D, & \sum_i p_i \exp \left(-\frac{\mu_i}{p_i} - \gamma d_{ij} \right) \leq 1 \\ -\infty, & \text{otherwise.} \end{cases} \tag{B.6}$$

By making the constraints explicit, we obtain the Lagrange dual problem:

$$\begin{array}{ll} \mathsf{maximize} & -\sum_i \mu_i - \gamma D \\ \mathsf{subject\ to} & \sum_i p_i \exp \left(-\frac{\mu_i}{p_i} - \gamma d_{ij} \right) \leq 1,\ \ j = 1, 2, \ldots, M \\ & \gamma \geq 0 \\ \mathsf{variables} & \boldsymbol{\mu}, \gamma. \end{array}$$

The constant parameters are $\mathbf{p}, d_{ij}$ and D.

Now we change the dual variables from $\boldsymbol{\mu}$ to $\boldsymbol{\alpha}$: $\alpha_i = -\frac{\mu_i}{p_i}$, and rewrite the dual problem as

$$\begin{array}{ll} \text{maximize} & \sum_i p_i\alpha_i - \gamma D \\ \text{subject to} & \log\sum_i \exp(\log p_i + \alpha_i - \gamma d_{ij}) \leq 0, \quad j = 1, 2, \ldots, M \\ & \gamma \geq 0 \\ \text{variables} & \boldsymbol{\alpha}, \gamma. \end{array}$$

In order to bring the dual problem (3.16) to the standard form of GP, we use an exponential change of the variable $w = e^{\gamma}, z_i = e^{\alpha_i}$ to rewrite the dual problem as

$$\begin{array}{ll} \text{maximize} & w^{-D}\prod_i z_i^{p_i} \\ \text{subject to} & \sum_i p_i z_i w^{-d_{ij}} \leq 1, \quad j = 1, 2, \ldots, M \\ & w \geq 1, \quad z_i \geq 0, \quad i = 1, 2, \ldots, N \\ \text{variables} & \mathbf{z}, w. \end{array}$$

The constant parameters are $\mathbf{p}, d_{ij}$ and D.

The weak duality part of the proposition follows directly from a standard fact in Lagrange duality theory [21]: the Lagrange dual function is always a lower bound on the primal minimization problem.

It is well-known that the objective function in the primal problem (3.15) is convex in P_{ij}, and the constraints are affine. The strong duality part of the proposition holds because the primal problem (3.1) is a convex optimization satisfying Slater's condition [21]. □

B.4 Proof of Proposition 3.3

Proof. Let $A_{ij} = P_{ij}^{1/(1+\rho)}$. Problem (3.23) can be written as:

$$\begin{array}{ll} \text{minimize} & \sum_j w_j^{1+\rho} \\ \text{subject to} & \mathbf{A}^T\mathbf{p} = \mathbf{w} \\ & \mathbf{1}^T\mathbf{p} = 1 \\ & \mathbf{p} \succeq 0 \\ \text{variables} & \mathbf{w}, \mathbf{p}. \end{array}$$

The Lagrangian can be written as:

$$L(\mathbf{p}, \mathbf{w}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}) = \sum_j w_j^{1+\rho} + \boldsymbol{\alpha}^T(\mathbf{w} - \mathbf{A}^T\mathbf{p}) + \beta(1 - \mathbf{1}^T\mathbf{p}) - \boldsymbol{\lambda}^T\mathbf{p}.$$

The Lagrangian is convex in $\mathbf{w}$ and can be minimized over $\mathbf{w}$. The minimizer is

$$w_j^* = [-\alpha_j/(1+\rho)]^{1/\rho}, \quad \forall j.$$

Since the Lagrangian is affine in $\mathbf{p}$, its minimized value is finite if and only if the coefficient of $\mathbf{p}$ is zero:

$$-\mathbf{A}^T\boldsymbol{\alpha} - \beta\mathbf{1} - \boldsymbol{\lambda} = 0.$$

Let $\theta(\rho)$ be a constant that only depends on ρ:

$$\theta(\rho) = \frac{\rho(-1)^{1/\rho}}{(1+\rho)^{1+1/\rho}}.$$

Substituting $\mathbf{w}^*$ into the Lagrangian, we can write the Lagrange dual problem as:

$$\begin{array}{ll} \text{maximize} & \theta(\rho)\sum_j \alpha_j^{(1+\rho)/\rho} + \beta \\ \text{subject to} & -\mathbf{A}^T\boldsymbol{\alpha} - \beta\mathbf{1} - \boldsymbol{\lambda} = 0 \\ & \boldsymbol{\lambda} \succeq 0 \\ \text{variables} & \boldsymbol{\alpha}, \boldsymbol{\lambda}, \beta, \end{array}$$

which, by removing the slack variable $\boldsymbol{\lambda}$, is equivalent to

$$\begin{array}{ll} \text{maximize} & \theta(\rho)\sum_j \alpha_j^{(1+\rho)/\rho} - \beta \\ \text{subject to} & \mathbf{A}^T\boldsymbol{\alpha} \preceq \beta\mathbf{1} \\ \text{variables} & \boldsymbol{\alpha}, \beta. \end{array}$$

Since at optimality, $\beta^*(\boldsymbol{\alpha}) = \max_i\{\sum_j A_{ij}\alpha_i\}$, the Lagrange dual problem can be simplified to an unconstrained problem over $\boldsymbol{\alpha}$:

$$\text{maximize} \quad \left[\theta(\rho)\sum_j \alpha_j^{(1+\rho)/\rho} - \max_i\left\{\sum_j A_{ij}\alpha_i\right\}\right].$$

□

B.5 Proof of Theorem 3.4

Proof. The claim is readily verified if the objective in (3.48) is to maximize x_{i^*}. In this case, omitting the monomial constraints in the form of range constraints on the variables, we can rewrite (3.48) as

$$\begin{array}{ll} \text{minimize} & \frac{1}{x_{i^*}} \\ \text{subject to} & \frac{1}{x_i} \le \frac{1}{x_{i,min}} \quad i = 1, 2, \ldots, N \\ \text{variables} & \mathbf{x}. \end{array} \tag{B.7}$$

By the structure of GPA forms, x_i are inverted posynomials of the variables $\mathbf{p}, \boldsymbol{\nu}$ and γ_{ij}, thus (B.7) is minimizing a posynomial subject to upper bound constraints on other posynomials. Therefore, (3.48) is equivalent to a GP.

To prove the claim for the maxmin fairness case, we can use the following technique to convert the problem of maximizing (over variables $\mathbf{z}$) the minimum of $g_j(\mathbf{z})$ to be maximizing over $(\mathbf{z}, t)$ (where t is an auxiliary variable) such that $g_j(\mathbf{z}) \geq t, \forall j$. Specifically, for the maxmin fair optimization, the following problem

$$\begin{array}{ll} \mathsf{maximize} & \min_{j=1,2,\ldots,M} g_j(\mathbf{z}) \\ \mathsf{subject\ to} & f_i(\mathbf{z}) \geq 1, \ \ i = 1, 2, \ldots, N \\ \mathsf{variables} & \mathbf{z} \end{array} \tag{B.8}$$

where g_j, f_i are inverted posynomials, is easily verified to be equivalent to the following problem:

$$\begin{array}{ll} \mathsf{maximize} & t \\ \mathsf{subject\ to} & g_j(\mathbf{z}) \geq t, \ \ j = 1, 2, \ldots, M \\ & f_i(\mathbf{z}) \geq 1, \ \ i = 1, 2, \ldots, N \\ \mathsf{variables} & \mathbf{z}, t. \end{array} \tag{B.9}$$

Now we rewrite the optimization (B.9) as

$$\begin{array}{ll} \mathsf{minimize} & t^{-1} \\ \mathsf{subject\ to} & \frac{t}{g_j(\mathbf{z})} \leq 1, \ \ j = 1, 2, \ldots, M \\ & \frac{1}{f_i(\mathbf{z})} \leq 1, \ \ i = 1, 2, \ldots, N \\ \mathsf{variables} & \mathbf{z}. \end{array}$$

The objective function is a monomial, and the inequality constraints are posynomials of $(\mathbf{z}, t)$. Therefore, this is a GP in standard form. □

B.6 Proof of Theorem 3.5

Proof. We first associate a Lagrange multiplier λ_l for each of the constraints $\sum_{s:l \in L(s)} x_s \leq c_l(\mathbf{P})$. Using the KKT optimality conditions for convex optimization [16, 21], solving problem (3.53) is equivalent to

satisfying the complementary slackness condition and finding the stationary points of the Lagrangian.

Complementary slackness condition states that at optimality, the product of the dual variable and the associated primal constraint must be zero. This condition is satisfied since the equilibrium queuing delay must be zero if the total equilibrium ingress rate at a router is strictly smaller than the egress link capacity.

We now find the stationary points of the Lagrangian: $I_{system}(\mathbf{x}, \mathbf{P}, \boldsymbol{\lambda}) = (\sum_s \alpha_s d_s \log x_s - \sum_l \lambda_l \sum_{s:l\in L(s)} x_s) + (\sum_l \lambda_l c_l(\mathbf{P}))$. By linearity of the differentiation operator, this can be decomposed into two separate maximization problems:

$$\begin{aligned} &\mathsf{maximize}_{\mathbf{x}\succeq 0} \quad \sum_s \alpha_s d_s \log x_s - \sum_s \sum_{l\in L(s)} \lambda_l x_s, \\ &\mathsf{maximize}_{\mathbf{P}\succeq 0} \quad I_{power}(\mathbf{P}, \boldsymbol{\lambda}) = \sum_l \lambda_l c_l(\mathbf{P}). \end{aligned}$$

The first maximization is already implicitly solved by the congestion control mechanism. But we still need to solve the second maximization.

We first establish that, if the algorithm converges, the convergence is indeed toward the global optimum. As in Subsection 3.4.2, this can be established by showing that the second partial Lagrangian maximization problem is a GP. We can also directly verify that the partial Lagrangian to be maximized $I_{power}(\mathbf{P}) = \sum_l \lambda_l \log(\mathrm{SIR}_l(\mathbf{P}))$ is a strictly concave function of a logarithmically transformed power vector. Let $\tilde{P}_l = \log P_l, \ \ \forall l$, we have $I_{power}(\tilde{\mathbf{P}}) =$

$$\begin{aligned} &\sum_l \lambda_l \log \frac{G_{ll} e^{\tilde{P}_l}}{\sum_k G_{lk} e^{\tilde{P}_k} + n_l} \\ &= \sum_l \lambda_l \left[\log(G_{ll} e^{\tilde{P}_l}) - \log \left(\sum_k G_{lk} e^{\tilde{P}_k} + n_l \right) \right] \\ &= \sum_l \lambda_l \left[\log(G_{ll} e^{\tilde{P}_l}) - \log \left(\sum_k \exp(\tilde{P}_k + \log G_{lk}) + n_l \right) \right]. \end{aligned}$$

The first term in the square bracket is linear in $\tilde{\mathbf{P}}$, and the second term is concave in $\tilde{\mathbf{P}}$ as directly verified below.

Taking the derivative of $I_{power}(\tilde{\mathbf{P}})$ with respect to $\tilde{P}_l$, we have

$$\begin{aligned}\nabla_l I_{power}(\tilde{\mathbf{P}}) &= \lambda_l - \sum_{j\neq l} \frac{\lambda_j G_{jl} e^{\tilde{P}_l}}{\sum_{k\neq j} G_{jk} e^{\tilde{P}_k} + n_j} \\ &= \lambda_l - P_l \sum_{j\neq l} \frac{\lambda_j G_{jl}}{\sum_{k\neq j} G_{jk} P_k + n_j}.\end{aligned}$$

Taking derivatives again, for each of the nonlinear $-\lambda_l \log\left(\sum_k \exp(\tilde{P}_k + \log G_{lk}) + n_l\right)$ terms in $I_{power}(\tilde{\mathbf{P}})$, we obtain the Hessian:

$$\mathbf{H}^l = \frac{-\lambda_l}{(\sum_k z_{lk} + n_l)^2} \left(\left(\sum_k z_{lk} + n_l \right) \mathbf{diag}(\mathbf{z}_l) - \mathbf{z}_l \mathbf{z}_l^T \right)$$

where $z_{lk} = \exp(\tilde{P}_k + \log G_{lk})$ and $\mathbf{z}_l$ is a column vector $[z_{l1}, z_{l2}, \ldots, z_{lN}]^T$.

Matrix $\mathbf{H}^l$ is indeed negative definite: for all vectors $\mathbf{v}$,

$$\mathbf{v}^T \mathbf{H}^l \mathbf{v} = \frac{-\lambda_l \left((\sum_k z_{lk} + n_l) (\sum_k v_k^2 z_{lk}) - (\sum_k v_k z_{lk})^2 \right)}{(\sum_k z_{lk} + n_l)^2} < 0. \quad \text{(B.10)}$$

This is because of the Cauchy–Schwarz inequality: $(\mathbf{a}^T\mathbf{a})(\mathbf{b}^T\mathbf{b}) \geq (\mathbf{a}^T\mathbf{b})^2$ where $a_k = v_k\sqrt{z_{lk}}$ and $b_k = \sqrt{z_{lk}}$ and the fact that $n_l > 0$. Therefore, $I_{power}(\tilde{P})$ is a strictly concave function of $\tilde{\mathbf{P}}$, and its Hessian is a negative definite block diagonal matrix $\mathbf{diag}(\mathbf{H}^1, \mathbf{H}^2, \ldots, \mathbf{H}^L)$.[1]

Coming back to the $\mathbf{P}$ solution space instead of $\tilde{\mathbf{P}}$, it is easy to verify that the derivative of $I_{power}(\mathbf{P})$ with respect to P_l is

$$\nabla_l I_{power}(\mathbf{P}) = \frac{\lambda_l}{P_l} - \sum_{j\neq l} \frac{\lambda_j G_{jl}}{\sum_{k\neq j} G_{jk} P_k + n_j}.$$

Therefore, the logarithmic change of variables simply scales each entry of the gradient by P_l: $\nabla_l I_{power}(\mathbf{P}) = \frac{1}{P_l} \nabla_l I_{power}(\tilde{\mathbf{P}})$. Power update can be conducted in either $\mathbf{P}$ or $\tilde{\mathbf{P}}$ domain.

[1] Interestingly, some of the propositions about JOCP in [31] depend on the invertibility of $\mathbf{H}$, which are provided for by the nonzero noise terms.

We now use the gradient method [16], with a constant step size κ, to maximize $I_{power}(\mathbf{P})$:

$$\begin{aligned} P_l(t+1) &= P_l(t) + \kappa \nabla_l I_{power}(\mathbf{P}) \\ &= P_l(t) + \kappa \left(\frac{\lambda_l(t)}{P_l(t)} - \sum_{j \neq l} \frac{\lambda_j(t) G_{jl}}{\sum_{k \neq j} G_{jk} P_k(t) + n_j} \right). \end{aligned}$$

Simplifying the equation and using the definition of SIR, we can write the gradient steps as the following distributed power control algorithm with message passing:

$$P_l(t+1) = P_l(t) + \frac{\kappa \lambda_l(t)}{P_l(t)} - \kappa \sum_{j \neq l} G_{lj} m_j(t)$$

where $m_j(t)$ are messages passed from node j:

$$m_j(t) = \frac{\lambda_j(t) \mathrm{SIR}_j(t)}{P_j(t) G_{jj}}.$$

These are exactly items (3) and (4) in the JOCP Algorithm.

It is known [16] that when the step size along the gradient direction is optimized, the gradient-based iterations converge. Such an optimization of step size κ in (3.56) would require global coordination in a wireless ad hoc network, and is undesirable or infeasible. However, in general gradient-based iterations with a constant step size may not converge.

By the descent lemma [16], convergence of the gradient-based optimization of a function $f(\mathbf{x})$, with a constant step size κ, is guaranteed if $f(\mathbf{x})$ has the Lipschitz continuity property: $\|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\| \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|$ for some $L > 0$, and the step size is small enough: $\epsilon \leq \kappa \leq \frac{2-\epsilon}{L}$ for some $\epsilon > 0$. It is known that $f(\mathbf{x})$ has the Lipschitz continuity property if it has a Hessian bounded in l_2 norm.

The Hessian $\mathbf{H}$ of $\sum_l \lambda_l c_l(\mathbf{P})$ can be verified to be

$$H_{ll} = \sum_{j \neq l} \lambda_j \left(\frac{G_{jl}}{\sum_{k \neq j} G_{jk} P_k + n_j} \right)^2 - \frac{\lambda_l}{P_l^2}, \tag{B.11}$$

$$H_{li} = \sum_{j \neq l,i} \frac{\lambda_j G_{jl} G_{ji}}{\left(\sum_{k \neq j} G_{jk} P_k + n_j \right)^2}, \quad i \neq l. \tag{B.12}$$

It is known that the second assumption for Theorem 3.5 leads to the conclusion that $\boldsymbol{\lambda}$ are upper bounded, which, together with the first assumption for Theorem 3.5, shows that $\|\mathbf{H}\|_2$ is upper bounded. The upper bound can be estimated by the following inequality:

$$\|\mathbf{H}\|_2 \leq \sqrt{\|\mathbf{H}\|_1 \|\mathbf{H}\|_\infty}$$

where $\|\mathbf{H}\|_1$ is the maximum column-sum matrix norm of $\mathbf{H}$, and $\|\mathbf{H}\|_\infty$ is the maximum row-sum matrix norm.

Therefore, the power control part (3.56) converges for a small enough step size κ:

$$\epsilon \leq \kappa \leq \frac{2-\epsilon}{L'}$$

where

$$\begin{aligned}(L')^2 = \max_i \Bigg(& \sum_l \sum_{j\neq l,i} \frac{\lambda_j G_{jl} G_{ji}}{\left(\sum_{k\neq j} G_{jk} P_k + n_j\right)^2} \\ & + \left| \sum_{j\neq l} \lambda_j \left(\frac{G_{jl}}{\sum_{k\neq j} G_{jk} P_k + n_j} \right)^2 - \frac{\lambda_l}{P_l^2} \right| \Bigg) \\ & \times \max_l \Bigg(\sum_i \sum_{j\neq l,i} \frac{\lambda_j G_{jl} G_{ji}}{\left(\sum_{k\neq j} G_{jk} P_k + n_j\right)^2} \\ & + \left| \sum_{j\neq l} \lambda_j \left(\frac{G_{jl}}{\sum_{k\neq j} G_{jk} P_k + n_j} \right)^2 - \frac{\lambda_l}{P_l^2} \right| \Bigg)\end{aligned}$$

and ϵ can be any small positive number $\leq \frac{2}{1+L'}$.

It is known [88] that TCP Vegas converges for a small enough step size $0 < \gamma \leq \frac{2\alpha_{min} d_{min} c_{min}}{L_{max} S_{max} x_{max}^2}$, where α_{min} and d_{min} are the smallest TCP source parameters α_s and d_s among the sources, respectively, x_{max} is the largest possible source rates, c_{min} is the smallest link data rate, L_{max} is the largest number of links any path has, and S_{max} is the largest number of sources sharing a link.

Convergence proof for TCP Vegas in [88] assumes that $c_{min} \neq 0$. Since SIR_l is lower bounded by $\frac{P_{l,min} G_{ll}}{\sum_{j\neq l} P_{j,max} G_{lj} + n_l}$, each c_l is lower

bounded by a strictly positive number. (In fact, the formulation in (3.53) assumes high SIR in the first place.) Consequently, TCP Vegas (3.55,3.54) also converges. By the convergence result of the simultaneous gradient method to the saddle point of minmax problems [17, 110] (in this case, minimizing the Lagrangian over dual variables and maximizing it over the primal variables to the saddle point of the Lagrangian, which is the optimal $(\mathbf{x}^*, \mathbf{P}^*)$), the JOCP Algorithm converges.

Since c_l can be turned into a concave function in $\tilde{\mathbf{P}}$, each constraint $\sum_{s:l\in L(s)} x_s - c_l(\mathbf{P}) \leq 0$ in (3.53) is an upper bound constraint on a convex function in $(\mathbf{x}, \tilde{\mathbf{P}})$. So problem (3.53) can be turned into maximizing a strictly concave objective function over a convex constraint set. The established convergence is thus indeed toward a unique global optimum. □

B.7 Proof of Proposition 4.1

Proof. The Perron–Frobenius eigenvalue $\rho(\mathbf{B})$ of a positive $n\times n$ matrix $\mathbf{B}$ can be characterized as:

$$\rho(\mathbf{B}) = \min\{\lambda | \mathbf{B}\mathbf{v} \preceq \lambda\mathbf{v} \text{ for some } \mathbf{v} \succ 0\}.$$

Therefore, computation of Perron–Frobenius eigenvalues can be conducted by the following GP:

$$\begin{array}{ll} \text{minimize} & \rho \\ \text{subject to} & \sum_{j=1}^n \frac{B_{ij}v_j}{\rho v_i} \leq 1, \quad i = 1, \ldots, n \\ \text{variables} & \mathbf{v}, \rho. \end{array} \tag{B.13}$$

Substituting B_{ij} as $A_{ij}\exp(\boldsymbol{\lambda}^T g(y_j))$, we see that the constraint in problem (B.13) becomes the constraint in problem (4.11).

The objective of maximizing $\boldsymbol{\lambda}^T\mathbf{z} - \log\rho(\mathbf{A}_{\boldsymbol{\lambda}})$ is equivalent to minimizing $\log\rho - \log\prod_{k=1}^d \tilde{\lambda}_k^{z_k}$, or minimizing $\rho\prod_{k=1}^d \tilde{\lambda}_k^{-z_k}$, where $\tilde{\lambda}_k = \exp(\lambda_k)$. Therefore, the rate function $I(\mathbf{z})$ of Markov chain large deviation bounds can be computed by GP (4.11). □

B.8 Proof of Proposition 4.2

Proof. We need to show that the Lagrange dual problem of (4.12) is

$$\begin{array}{ll} \mathsf{maximize} & \sum_i p_i \log \beta_i - \mathbf{v}^T \boldsymbol{\lambda} \\ \mathsf{subject\ to} & \mathbf{A}^T \boldsymbol{\lambda} + \alpha \boldsymbol{\nu} = \boldsymbol{\beta} \\ & \boldsymbol{\lambda} \succeq 0 \\ \mathsf{variables} & \boldsymbol{\lambda}, \boldsymbol{\beta}. \end{array} \tag{B.14}$$

The constant parameters are $\mathbf{A}, \mathbf{v}, \mathbf{p}, \boldsymbol{\nu}$ and α. Furthermore, the optimal primal variables $\mathbf{x}^*$ can be obtained from the optimal dual variables $\boldsymbol{\lambda}^*$ as

$$x_i^* = \frac{p_i}{\sum_j \lambda_j^* A_{ji} + \alpha \nu_i}, \quad \forall i.$$

Indeed, we can form the Lagrangian of the primal problem (4.12), ignoring the constant term of $\sum_i p_i \log p_i$:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}) = -\sum_i p_i \log x_i + \alpha(\boldsymbol{\nu}^T \mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{v}) - \boldsymbol{\sigma}^T \mathbf{x}$$

where $\boldsymbol{\lambda}, \boldsymbol{\sigma} \succeq 0$ are the Lagrange multiplier vectors. Let the derivative of $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma})$ with respect to x_i be equal to 0, we obtain

$$x_i = \frac{p_i}{\sum_j \lambda_j A_{ji} + \alpha \nu_i - \sigma_i}.$$

Substitute this $\mathbf{x}$ into the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma})$, we obtain the Lagrange dual function $g(\boldsymbol{\lambda}, \boldsymbol{\sigma})$:

$$\begin{array}{c} -\sum_i p_i \log \frac{p_i}{\sum_j \lambda_i A_{ji} + \alpha \nu_i - \sigma_i} + \sum_j \lambda_j \sum_i A_{ji} \frac{p_i}{\sum_j \lambda_j A_{ji} + \alpha \nu_i - \sigma_i} - \boldsymbol{\lambda}^T \mathbf{v} \\ -\sum_i \sigma_i \frac{p_i}{\sum_j \lambda_i A_{ji} + \alpha \nu_i - \sigma_i}, \end{array}$$

which can be simplified to

$$g(\boldsymbol{\lambda}, \boldsymbol{\sigma}) = \sum_i p_i \log \left(\sum_j \lambda_j A_{ji} + \alpha \nu_i - \sigma_i \right) - \mathbf{v}^T \boldsymbol{\lambda} + \sum_i p_i - \sum_i p_i \log p_i.$$

Therefore, the Lagrange dual problem can be stated as

$$\begin{array}{ll} \mathsf{maximize} & \sum_i p_i \log \left(\sum_j \lambda_j A_{ji} + \alpha \nu_i - \sigma_i \right) - \mathbf{v}^T \boldsymbol{\lambda} \\ \mathsf{subject\ to} & \boldsymbol{\sigma}, \boldsymbol{\lambda} \succeq 0 \\ \mathsf{variables} & \boldsymbol{\sigma}, \boldsymbol{\lambda}. \end{array}$$

Since the objective function is a non-increasing function of $\boldsymbol{\sigma} \succeq 0$, we let $\boldsymbol{\sigma} = 0$, and simplify the Lagrange dual problem to

$$\begin{array}{ll} \mathsf{maximize} & \sum_i p_i \log\left(\sum_j \lambda_j A_{ji} + \alpha\nu_i\right) - \mathbf{v}^T\boldsymbol{\lambda} \\ \mathsf{subject\ to} & \boldsymbol{\lambda} \succeq 0 \\ \mathsf{variables} & \boldsymbol{\lambda}. \end{array}$$

Now letting $\boldsymbol{\beta} = \mathbf{A}^T\boldsymbol{\lambda} + \alpha\boldsymbol{\nu}$ proves the claim. □

B.9 Proof of Theorem 4.1

Proof. It is easy to see that a market equilibrium must satisfy the constraints. To see that any feasible solution must also be a market equilibrium, multiply the second constraint by $x_{ij}p_j$ and sum over j, which gives:

$$\sum_j u_{ij}x_{ij} \leq \sum_k u_{ik}x_{ik}\frac{\sum_j x_{ij}p_j}{p_i}, \quad \forall i.$$

By assumptions, $\sum_k u_{ik}x_{ik} \neq 0$ and $p_i \neq 0$. Therefore, we have

$$p_i \leq \sum_j x_{ij}p_j, \quad \forall i.$$

Sum over i and change the order of summation over i and j, we obtain:

$$\sum_i p_i \leq \sum_j p_j \sum_i x_{ij},$$

which should be an equality. Therefore, either the inequality constraint is tight $\frac{p_i}{p_j} = \frac{\sum_k u_{ik}x_{ik}}{u_{ij}}$ or the factor $x_{ij}p_j$ with which we multiply the inequality is 0. Since $p_j \neq 0$, either $x_{ij} = 0$, i.e., good j is not assigned to person i, or $\frac{p_i}{p_j} = \frac{\sum_k u_{ik}x_{ik}}{u_{ij}}$. This is the definition of a general market equilibrium. □

Acknowledgements

I would like to acknowledge, first and foremost, Stephen Boyd, who introduced convex optimization to me, and my other co-authors on GP-related publications and preprints: David Julian, Daniel O'Neill, and Arak Sutivong at Stanford University, and Ying Li, Daniel Palomar, Chee Wei Tan, and Sergio Verdú at Princeton University.

I am also grateful to illuminating discussions related to GP with Toby Berger, Thomas Cover, and David Forney Jr. on information theoretic and statistical physics topics, with Steven Low on networking topics, and with John Doyle, Seung Jean Kim, Tom Z. Q. Luo, Asuman Ozdaglar, Pablo Parrilo, Lin Xiao, Yinyu Ye, and Wei Yu on optimization topics.

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

References

[1] M. Abou-El-Ata, H. Bergany, and M. El-Wakeel, "Probabilistic multi-item inventory model with varying order cost under two restrictions: A geometric programming approach," *International Journal of Production Economics*, vol. 83, no. 3, pp. 223–231, 2003.

[2] R. Albert and A. L. Barabasi, "Statistical mechanics of complex networks," *Review of Modern Physics*, vol. 47, 2002.

[3] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 14–20, Jan. 1972.

[4] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 19, pp. 357–359, May 1973.

[5] K. Arrow and G. Debreu, "Existence of an equilibrium for a competitive economy," *Econometrica*, vol. 22, pp. 265–290, 1954.

[6] M. Avriel, *Advances in Geometric Programming*. Plenum Press, 1980.

[7] M. Avriel, R. Dembo, and U. Passy, "Solution of generalized geometric programs," *International Journal of Numerical Methods in Engineering*, vol. 9, pp. 149–168, 1975.

[8] M. Avriel, M. J. Rijckaert, and D. J. Wilde, *Optimization and Design*. Prentice Hall, 1973.

[9] P. Bak, *How Nature Works: The Science of Self Organized Criticality*. Copernicus, 1996.

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

[11] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.
[12] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.
[13] A. Ben-Tal, A. Nemirovski, and C. Ross, "Robust solutions of uncertain quadratic and conic-quadratic problems," *SIAM Journal of Optimization*, vol. 13, no. 2, pp. 535–560, 2002.
[14] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice Hall, 1971.
[15] D. Bertsekas and R. G. Gallager, *Data Networks*. Prentice Hall, 1991.
[16] D. P. Bertsekas, *Nonlinear Programming, 2nd Ed.* Athena Scientific, 1999.
[17] D. P. Bertsekas, E. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.
[18] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*. Prentice Hall, 1989.
[19] R. E. Blahut, "Computation of channel capacity and rate distortion function," *IEEE Transactions on Information Theory*, vol. 18, pp. 450–473, 1972.
[20] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Stanford University EE Technical Report*, 2004.
[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[22] W. C. Brainard and H. E. Scarf, "How to compute equilibrium prices in 1891," *Cowles Foundation Discussion Paper*, 2000.
[23] L. S. Brakmo and L. L. Peterson, "TCP Vegas: End-to-end congestion avoidance on a global Internet," *IEEE Journal of Selected Areas in Communications*, vol. 13, pp. 1465–1480, October 1995.
[24] D. Bricker, K. Kortanek, and L. Xui, "Maximum likelihood estimates with order restrictions on probabilities and odd ratios: A geometric programming approach," *Journal of Applied Mathematics and Decision Sciences*, vol. 1, no. 1, pp. 53–65, 1997.
[25] L. Campbell, "A coding theorem and Renyi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.
[26] B. Y. Cao, *Fuzzy Geometric Programming*. Kluwar Academic Publisher, October 2002.
[27] J. M. Carlson and J. Doyle, "Complexity and robustness," *Proceedings of National Academy of Sciences*, vol. 99, pp. 2538–2545, February 2002.
[28] J. M. Carlson and J. C. Doyle, "Highly Optimized Tolerance: Robustness and design in complex systems," *Physics Review Letters*, vol. 84, no. 11, pp. 2529–2532, 2000.
[29] T. Y. Chen, "Structural optimization using single-term posynomial geometric programming," *Computers and Structures*, vol. 45, pp. 911–918, 1992.
[30] M. Chiang, *Solving Nonlinear Problems in Communication Systems Using Dualities and Geometric Programming*. PhD thesis, Stanford University, 2003.
[31] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Joingly optimal congestion control and power control," *IEEE Journal of Selected Areas in Communications*, vol. 23, no. 1, pp. 104–116, 2005.

[32] M. Chiang and N. Bambos, "Distributed network control through sum product algorithms on graphs," in *Proceedings of IEEE Infocom*, 2002.

[33] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Transactions on Information Theory*, vol. 50, pp. 245–258, Feb. 2004.

[34] M. Chiang, D. O'Neill, D. Julian, and S. Boyd, "Resource allocation for QoS provisioning in wireless ad hoc networks.," in *Proceedings of IEEE Infocom*, 2001.

[35] M. Chiang and A. Sutivong, "Efficient nonlinear optimization of resource allocation," in *Proceedings of IEEE Infocom*, 2003.

[36] M. Chiang, A. Sutivong, and S. Boyd, "Efficient nonlinear optimizations of queuing systems," in *Proceedings of IEEE Infocom*, 2002.

[37] M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian, ch. Geometric programming for wireless network power control, *Resource Allocation in Next Generation Wireless Networks*. Nova Science Publisher, 2005.

[38] C. Chu and D. Wong, "VLSI circuit performance optimization by geometric programming," *Annals of Operations Research*, vol. 105, pp. 37–60, 2001.

[39] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with state information," *IEEE Transactions on Information Theory*, vol. 48, pp. 1629–1638, June 2002.

[40] T. M. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 1991.

[41] I. Csiszar, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[42] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[43] W. Daems, G. Geilen, and W. Sansen, "Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits," *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems*, vol. 22, no. 5, pp. 517–534, 2003.

[44] J. Dawson, S. Boyd, M. Hershenson, and T. Lee, "Optimal allocation of local feedback in multistage amplifiers via geometric programming," *IEEE Transactions on Circuits and Systems I*, vol. 48, no. 1, pp. 1–11, 2001.

[45] A. Dembo and D. Zeitouni, *Large Deviations: Techniques and Applications*. Springer Verlag, 1998.

[46] R. Dembo, "Sensitivity analysis in geometric programming," *Journal of Optimization Theory and Applications*, vol. 37, pp. 1–21, 1982.

[47] R. S. Dembo, "A set of geometric programming test problems," *Mathematical Programming*, vol. 10, no. 192-213, 1976.

[48] J. Dinkel, M. Kochenberger, and S. Wong, "Sensitivity analysis procedures for geometric programs: Computational aspects," *ACM Transactions on Mathematical Software*, vol. 4, no. 1, pp. 1–14, 1978.

[49] J. C. Doyle and J. M. Carlson, "Power laws, Highly Optimized Tolerance and generalized source coding," *Physics Review Letters*, vol. 84, no. 24, pp. 5656–5659, 2000.

[50] M. Drmota and W. Szpankowski, "The precise minimax redundancy," in *Proceedings of IEEE International Symposium on Information Theory*, 2002.

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

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

[53] A. Dutta and D. V. Rama, "An optimization model of communications satellite planning," *IEEE Transactions on Communications*, vol. 40, no. 9, pp. 1463–1473, 1992.

[54] J. Ecker, "Geometric programming: Methods, computations and applications," *SIAM Review*, vol. 22, no. 3, pp. 338–362, 1980.

[55] E. Eisenberg and D. Gale, "Consensus of subjective probabilities: The pari-mutual method," *Annals of Mathematical Statistics*, vol. 30, pp. 165–168, 1959.

[56] A. Ephremides and S. Verdú, "Control and optimization methods in communication network problems," *IEEE Transactions on Automatic Control*, vol. 9, pp. 930–942, 1989.

[57] S. C. Fang, J. R. Rajasekera, and H. Tsao, *Entropy Optimization and Mathematical Programming*. Kluwer Academic Publishers, 1997.

[58] F. Feigin and U. Passy, "The geometric programming dual to the extinction probability problem in simple branch processes," *The Annals of Probability*, vol. 9, no. 3, pp. 498–503, 1981.

[59] C. A. Floudas, *Deterministic Global Optimization: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, 1999.

[60] G. J. Foschini and Z. Miljanic, "A simple distributed autonomour power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, 1993.

[61] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[62] L. E. Ghaoui and H. Lebret, "Robust solutions to least square problems with uncertain data," *SIAM Journal of Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, 1997.

[63] L. E. Ghaoui and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 33–52, 1998.

[64] H. Greenberg, "Mathematical programming models for environmental quality control," *Operations Research*, vol. 43, no. 4, pp. 578–622, 1995.

[65] P. Hajela, "Geometric programming strategies for large scale structural synthesis," *AIAA Journal*, vol. 24, no. 7, pp. 1173–1178, 1986.

[66] M. Hersehnson, S. Boyd, and T. H. Lee, "Optimal design of a CMOS Op-Amp via geometric programming," *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems*, vol. 20, no. 1, pp. 1–21, 2001.

[67] K. L. Hsiung, S. J. Kim, and S. Boyd, "Robust geometric programming via piecewise linear approximiation," *Mathematical Programming*, 2005.

[68] J. Y. Hui, "Resource allocation for broadband networks," *IEEE Journal of Selected Areas in Communications*, pp. 1598–1608, 1988.

[69] P. A. Humblet, "Generalization of Huffman coding to minimize the probability of buffer overflow," *IEEE Transactions on Information Theory*, vol. 27, pp. 230–232, 1981.

[70] K. Jain, "A polynomial time algorithm for computing the Arrow-Debreu market equilibrium for linear utilities," in *Proceedings of IEEE Foundation of Computer Science*, 2004.

[71] C. Jin, D. X. Wei, and S. H. Low, "TCP FAST: motivation, architecture, algorithms, performance," in *Proceedings of IEEE Infocom*, 2004.

[72] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proceedings of IEEE Infocom*, 2002.

[73] S. Kandukuri and S. Boyd, "Optimal power control in interference limited fading wireless channels with outage probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, Jan. 2002.

[74] J. Karlof, "Permutation codes for the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 35, pp. 726–732, July 1989.

[75] J. K. Karlof and Y. O. Chang, "Optimal permutation codes for the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 356–358, 1997.

[76] F. P. Kelly, "Notes on effective bandwidth," *Stochastic Networks: Theory and Applications*, pp. 141–168, 1996.

[77] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of Operations Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

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

[79] S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses and ECN marks," *IEEE/ACM Transactions on Networking*, pp. 689–702, Oct. 2003.

[80] J. Kyparsis, "Sensitivity analysis in geometric programming: Theory and computation," *Annals of Operations Research*, vol. 27, pp. 39–64, 1990.

[81] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Transactions on Networking*, vol. 10, no. 2, pp. 272–286, 2002.

[82] A. Lapidoth and N. Miliou, "Duality bounds on the cut-off rate with applications to Ricean fading," *Submitted to IEEE Transactions on Information Theory*, March 2005.

[83] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[84] Y. Li, M. Chiang, and S. Verdu, "Lagrange duality of random coding error exponents," Preprint, Princeton University, 2005.

[85] S. H. Low, "A duality model of TCP and queue management algorithms," *IEEE/ACM Transactions on Networking*, vol. 11, no. 4, pp. 525–536, 2003.

[86] S. H. Low and D. E. Lapsley, "Optimization flow control, I: basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[87] S. H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Control Systems Magazine*, vol. 22, pp. 28–43, February 2002.
[88] S. H. Low, L. Peterson, and L. Wang, "Understanding Vegas: A duality model," *Journal of ACM*, vol. 49, no. 2, pp. 207–235, 2002.
[89] S. H. Low and R. Srikant, "A mathematical framework for designing a low-loss, low-delay internet," *Networks and Spatial Economics, special issue on "Crossovers between transportation planning and telecommunications",* E. Altman and L. Wynter, 2003.
[90] D. Luenberger, "A double look at duality," *IEEE Transactions Automatic Control*, 1992.
[91] T. Luo, "Optimal zero forcing transceiver design for multiaccess communications," McMaster University ECE Technical Report, 2001.
[92] C. Maranas and C. Foudas, "Global optimization in generalized geometric programming," *Computers and Chemical Engineering*, vol. 21, no. 4, pp. 351–369, 1997.
[93] M. Mazumdar and T. R. Jefferson, "Maximum likelihood estimates for multinomial probabilities via geometric programming," *Biometrika*, vol. 70, no. 1, pp. 257–261, 1983.
[94] R. McEliece, *Information Theory and Coding*. Wiley, 1976.
[95] J. Mo, R. La, V. Anantharam, and J. Walrand, "Analysis and comparison of TCP Reno and Vegas," in *Proceedings of IEEE Infocom*, 1999.
[96] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.
[97] Y. Nesterov and A. Nemirovsky, *Interior Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
[98] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer Verlag, 1999.
[99] D. O'Neill, "Adaptive congestion control for wireless networks using TCP," in *Proceedings of IEEE International Conference on Communications*, 2003.
[100] A. K. Parekh and R. Gallager, "A Generalized processor sharing approach to flow control in integrated services networks: The single node case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344–357, 1993.
[101] P. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, Caltech, 2000.
[102] P. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical Programming Series B*, vol. 96, no. 2, pp. 293–320, 2003.
[103] E. L. Perterson, "Geometric programming: A survey," *SIAM Review*, vol. 18, pp. 1–51, 1976.
[104] E. Peterson, "Investigation of path following algorithms for signomial geometric programming problems," *Annals of Operations Research*, vol. 105, pp. 15–19, 2001.
[105] S. Pradhan, J., and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information caes," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1181–1203, 2003.
[106] S. Prajna, A. Papachristodoulou, and P. A. Parrilo, "Introducing SOSTOOLS: A general purpose sum of squares programming solver," in *Proceedings of IEEE Conference on Decision and Control*, 2002.

[107] J. Rajasekera and M. Yamada, "Estimating the firm value distribution function by entropy optimization and geometric programming," *Annals of Operations Research*, vol. 105, pp. 61–75, 2001.
[108] F. Reif, *Fundamentals of Statistical and Thermal Physics*. McGraw Hill, 1965.
[109] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
[110] R. T. Rockafellar, ch. Saddle-points and convex analysis, *Differential Games and Related Topics*. North-Holland, 1971.
[111] R. T. Rockafellar, "Lagrange multipliers and optimality," *SIAM Review*, vol. 35, pp. 183–283, 1993.
[112] S. Sapatnekar, V. Rao, P. Vaidya, and S. Kang, "An exact solution to the transistor sizing problem for CMOS circuits using convex optimization," *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems*, vol. 12, no. 11, pp. 1621–1634, 1993.
[113] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, pp. 379–423, 623–656, 1948.
[114] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, 1959.
[115] C. Sims, "Computational methods for permutation groups," in *Computational Problems in Abstract Algebra*, 1970.
[116] D. Slepian, "Group codes for the Gaussian channel," *Bell Systems Technical Journals*, vol. 17, pp. 575–602, 1968.
[117] R. Srikant, *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.
[118] T. Starr, M. Sorbara, J. Cioffi, and P. Silverman, *DSL Advances*. Prentice Hall, 2003.
[119] A. Tang, J. Wang, S. H. Low, and M. Chiang, "Equilibrium of heterogeneous congestion control protocols," in *Proceedings of IEEE Infocom*, 2005.
[120] M. Teboulle and A. Ben-Tal, "Rate distortion theory with generalized infomration measure via convex programming duality," *IEEE Transactions on Information Theory*, vol. 32, pp. 630–641, 1986.
[121] M. Teboulle and A. Ben-Tal, "Extension of some results for capacity using a generalized information measure," *Applied Mathematics and Optimization*, vol. 17, pp. 121–132, 1988.
[122] T. Terlaky, E. Klafszky, and Mayer, "A geometric programming approach to the channel capacity problem," *Engineering Mathematics*, vol. 19, pp. 115–130, 1992.
[123] J. F. Tsai, H. L. Li, and N. Z. Hu, "Global optimization for signomial discrete programming problems in engineering design," *Engineering Optimization*, vol. 34, no. 6, pp. 613–622, 2002.
[124] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proceedings of Information Theory on Workshop*, 2001.
[125] T. Wall, D. Greening, and R. Woolsey, "Solving complex chemical equilibria using a geometric programming based technique," *Operations Research*, vol. 34, no. 3, pp. 345–355, 1986.

[126] L. Walras, *Elements of Pure Economics, Or The Theory of Social Wealth*. Lausanne, 1874.

[127] A. Weiss, "An introduction to large deviations for communication networks," *IEEE Journal of Selected Areas in Communications*, vol. 13, no. 6, pp. 938–952, 1995.

[128] D. Wong, "Maximum likelihood, entropy maximization, and the geometric programming approaches to the calibration of trip distribution models," *Transportation Research Part B: Methodological*, vol. 15, no. 5, pp. 329–343, 1981.

[129] www.mosek.com, *MOSEK Optimization Toolbox*. 2002.

[130] A. D. Wyner and J. Ziv, "The rate distortion function for source coding with side information at the decode," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, 1976.

[131] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation for wireless networks," *IEEE Transactions of Communications*, vol. 52, no. 7, pp. 1136–1144, July 2004.

[132] C. Zener, "A mathematical aid in optimizing engineering design," in *Proceedings of National Academy of Sciences*, pp. 537–539, 1961.

[133] C. Zener, *Engineering Design By Geometric Programming*. Wiley, 1971.

Foundations and Trends® in Communications and Information Theory

Volume 2 Issue 1/2, 2005

Editorial Board

Editor-in-Chief: Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544,USA
verdu@princeton.edu

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2005, Volume 2, 4 issues. ISSN paper version 1567-2190 (USD 200 N. America; EUR 200 Outside N. America). ISSN online version 1567-2328 (USD 250 N. America; EUR 250 Outside N. America). Also available as a combined paper and online subscription (USD 300 N. America; EUR 300 Outside N. America).

Foundations and Trends® in
Communications and Information Theory
Vol. 2, No 1/2 (2005) 1–153

now
the essence of knowledge

Geometric Programming for Communication Systems

Mung Chiang

Princeton University, Princeton, New Jersey 08544, USA,
chiangm@princeton.edu

Abstract

Geometric Programming (GP) is a class of nonlinear optimization with many useful theoretical and computational properties. Over the last few years, GP has been used to solve a variety of problems in the analysis and design of communication systems in several 'layers' in the communication network architecture, including information theory problems, signal processing algorithms, basic queuing system optimization, many network resource allocation problems such as power control and congestion control, and cross-layer design. We also start to understand why, in addition to how, GP can be applied to a surprisingly wide range of problems in communication systems. These applications have in turn spurred new research activities on GP, especially generalizations of GP formulations and development of distributed algorithms to solve GP in a network. This text provides both an in-depth tutorial on the theory, algorithms, and modeling methods of GP, and a comprehensive survey on the applications of GP to the study of communication systems.