Geometric Programming for Communication Systems

Mung Chiang
Electrical Engineering Department, Princeton University

Tutorial Presentation
August 2005

Outline

Part I: Geometric Programming

- What's GP? Why GP?
- GP and free energy optimization
- History of GP

Part II: GP Applications in Communication Systems

- Information theory and coding
- Wireless network power control
- Internet protocol TCP congestion control

Collaborators: S. Boyd, D. Julian, Y. Li, D. O'Neill, D. Palomar, A. Sutivong, C. W. Tan

References

- Overview: M. Chiang, "Geometric Programming for Communication Systems," *Foundations and Trends in Communications and Information Theory,* vol. 2, no. 1, pp. 1-156, Aug. 2005.

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

- Power control and distributed algorithm: M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian "Geometric programming for power control" *IEEE Trans. Wireless Communications*, 2006.

- Network protocol and TCP congestion control: M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control," *IEEE J. Sel. Areas Comm.,* vol. 23, no. 1, pp. 104-116, Jan. 2005.

Part I.A

GP: Formulations and Duality

Monomials and Posynomials

Monomial is a function $f : \mathbf{R}^n_+ \to \mathbf{R}$:

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

Multiplicative constant $d \geq 0$

Exponential constants $a^{(j)} \in \mathbf{R}, j = 1, 2, \dots, n$

Posynomial: A sum of monomials:

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

where $d_k \geq 0,\ k = 1, 2, \dots, K$, and $a_k^{(j)} \in \mathbf{R},\ j = 1, 2, \dots, n, k = 1, 2, \dots, K$

Example: $\sqrt{2} x^{-0.5} y^{\pi} z$ is a monomial, $x - y$ is not a posynomial

GP

- GP standard form in variables x:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 1, \;\; i = 1, 2, \ldots, m, \\ & h_l(x) = 1, \;\; l = 1, 2, \ldots, M \end{array}$$

where $f_i, i = 0, 1, \ldots, m$ are posynomials and $h_l, l = 1, 2, \ldots, M$ are monomials

Log transformation: $y_j = \log x_j, b_{ik} = \log d_{ik}, b_l = \log d_l$

- GP convex form in variables y:

$$\begin{array}{ll} \text{minimize} & p_0(y) = \log \sum_{k=1}^{K_0} \exp(a_{0k}^T y + b_{0k}) \\ \text{subject to} & p_i(y) = \log \sum_{k=1}^{K_i} \exp(a_{ik}^T y + b_{ik}) \le 0, \;\; i = 1, 2, \ldots, m, \\ & q_l(y) = a_l^T y + b_l = 0, \;\; l = 1, 2, \ldots, M \end{array}$$

In convex form, GP with only monomials reduces to LP

Example

In fact a channel capacity problem:

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

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

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

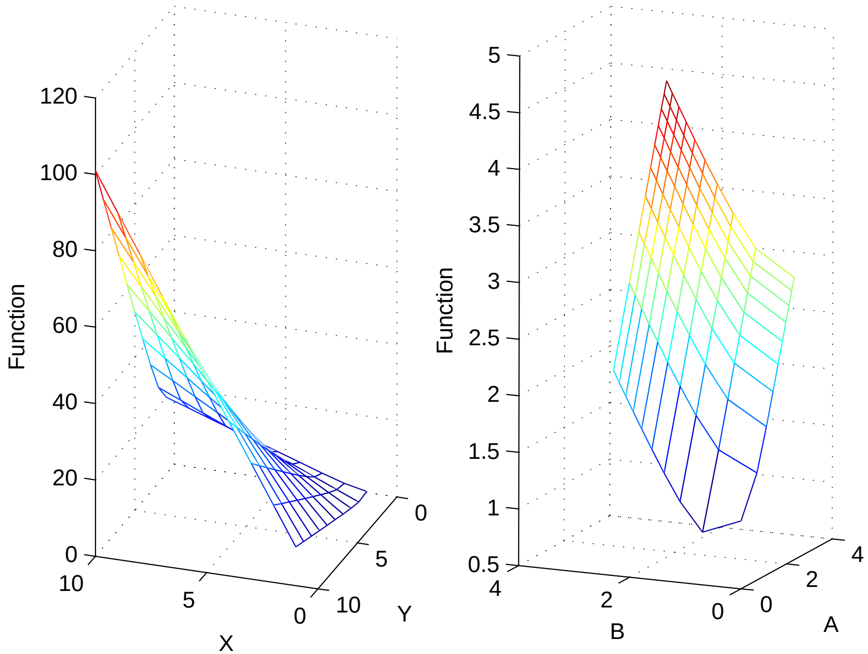
Convex form GP:

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

Pseudo-Nonconvexity

A bi-variate posynomial before (left graph) and after (right graph) the log transformation

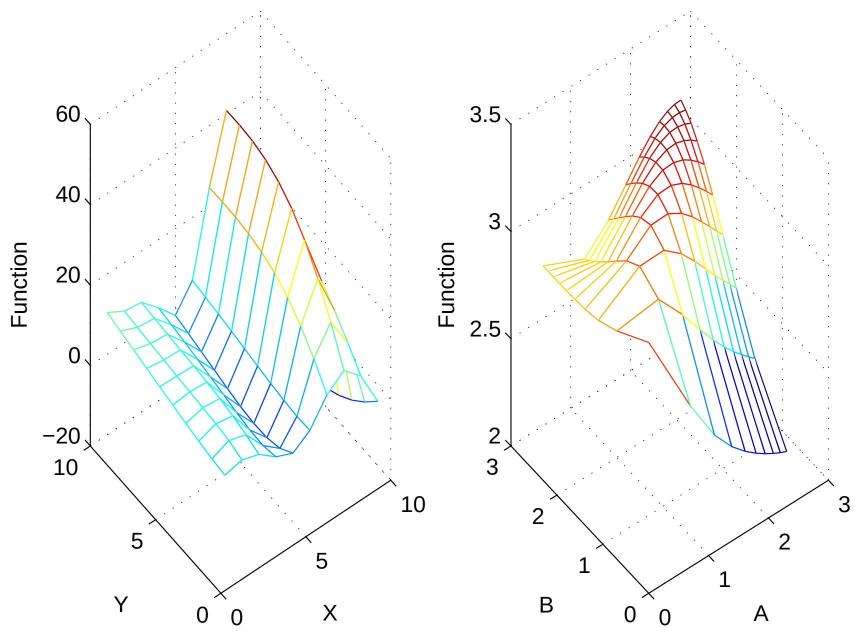
A non-convex function is turned into a convex one



Nonconvexity

A bi-variate signomial (ratio between two posynomials) before (left graph) and after (right graph) the log transformation

A non-convex function remains a non-convex one



GP, SP, PMoP

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 polynomial-time solvable, but PMoP and SP are not

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

Dual GP

Primal problem: Unconstrained GP in variables y

$$\text{minimize} \quad \log \sum_{i=1}^{N} \exp(a_i^T y + b_i).$$

Lagrange dual problem in variables ν:

$$\begin{array}{ll} \text{maximize} & b^T \nu - \sum_{i=1}^{N} \nu_i \log \nu_i \\ \text{subject to} & \mathbf{1}^T \nu = 1, \\ & \nu \succeq 0, \\ & A^T \nu = 0 \end{array}$$

Dual GP

Primal problem: General GP in variables y

$$
\begin{array}{ll}
\text{minimize} & \log \sum_{j=1}^{k_0} \exp(a_{0j}^T y + b_{0j}) \\
\text{subject to} & \log \sum_{j=1}^{k_i} \exp(a_{ij}^T y + b_{ij}) \leq 0, \quad i = 1, \ldots, m,
\end{array}
$$

Lagrange dual problem in variables $\nu_i, \ i = 0, 1, \ldots, m$:

$$
\begin{array}{ll}
\text{maximize} & b_0^T \nu_0 - \sum_{j=1}^{k_0} \nu_{0j} \log \nu_{0j} + \sum_{i=1}^{m} \left(b_i^T \nu_i - \sum_{j=1}^{k_i} \nu_{ij} \log \frac{\nu_{ij}}{\mathbf{1}^T \nu_i} \right) \\
\text{subject to} & \nu_i \succeq 0, \quad i = 0, \ldots, m, \\
 & \mathbf{1}^T \nu_0 = 1, \\
 & \sum_{i=0}^{m} A_i^T \nu_i = 0
\end{array}
$$

A_0 is the matrix of the exponential constants in the objective function, and $A_i, \ i = 1, 2, \ldots, m$ are the matrices of the exponential constants in the constraint functions

Example

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

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

GP Extensions

- 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 (covers a wide range of conic convex optimization)
- GP formulations based on monomial and posynomial approximations of nonlinear functions (approximates a wide range of nonconvex optimization)

- Signomial Programming that allows posynomial equality constraints

Generalized GP

Rule 1: Composing posynomials $\{f_{ij}(\mathbf{x})\}$ with a posynomial with non-negative exponents $\{a_{ij}\}$ is a generalized posynomial

Rule 2: The maximum of a finite number of posynomials is also a generalized posynomial

Rule 3: f_1 and f_2 are posynomials and h is a monomial :

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

Example:

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

Solving GP

- Level 1: local optimum is global optimum
- Level 2: polynomial time to compute global optimum
- Level 3: efficient practical algorithm (e.g., primal-dual interior-point method)
- Level 4: free software (e.g., MOSEK)
- Level 5: robust solution (Hsiung, Kim, Boyd 2005)
- Level 6: distributed solution (Tan, Palomar, Chiang 2005)

Distributed Algorithm for GP

Example: Unconstrained GP in standard form:

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

Making a change of variable $y_i = \log x_i, \forall i$:

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

Introducing auxiliary variables $\{y_{ij}\}$ for the coupled arguments, and additional equality consistency constraints:

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

Forming the Lagrangian:

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

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

Minimization of the Lagrangian with respect to primal variables $(\{y_i\},\{y_{ij}\})$ can be done distributively by each user

Master dual problem has to be solved:

$$\text{maximize}_{\{\gamma_{ij}\}} \qquad g(\{\gamma_{ij}\})$$

where

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

Gradient update for the consistency prices:

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

Why GP

- Nonlinear nonconvex problem, can be turned into nonlinear convex problem
- Linearly constrained dual problem

- Theoretical structures: global optimality, zero duality gap, KKT condition, sensitivity analysis
- Numerical efficiency: interior-point, robust, distributed algorithms
- Surprisingly wide range of applications

Part I.B

GP and Statistical Physics

Free Energy Optimization

Given an energy vector $\mathbf{e}$ and a probability vector $\mathbf{p}$

- Average energy: $U(\mathbf{p}, \mathbf{e}) = \mathbf{p}^T \mathbf{e}$
- Entropy: $H(\mathbf{p}) = -\sum_{i=1}^n p_i \log p_i$
- Gibbs free energy:

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

Gibbs free energy minimization:

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

Solution: Boltzmann distribution $\tilde{\mathbf{b}}$

Helmholtz free energy:

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

Helmholtz free energy maximization:

$$\max_{\mathbf{e}} \min_{\mathbf{p}} G(\mathbf{p}, \mathbf{e}) = \min_{\mathbf{p}} \max_{\mathbf{e}} G(\mathbf{p}, \mathbf{e})$$

Generalization:

Multiple phase chemical system with K phases and J types of substances

n_{jk}: number of atomic weights of substance j in phase k

e_{jk}: energy of substance j in phase k, $j = 1, 2, \ldots, J$, $k = 1, 2, \ldots, K$

Multiphase equilibrium problem: minimize the generalized Gibbs free energy with unit temperature over $\{n_{jk}\}$:

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

Free Energy and GP

- GP in convex form is equivalent to a constrained Helmholtz free energy maximization problem

- Dual problem of GP is equivalent to a linearly-constrained generalized Gibbs free energy minimization problem

- Dual problem of unconstrained GP is equivalent to the Gibbs free energy minimization

Large Deviation Bounds

Probability of an undesirable event is to be bounded or minimized:

- Given a family of conditional distributions describing a channel, probability of decoding error to vanish exponentially as the codeword length goes to infinity

- Given a queuing discipline and arrival and departure statistics, probability of buffer overflow to vanish exponentially as the buffer size increases.

Large deviation principles govern such exponential behavior in stochastic systems

Can be obtained by GP:

- IID case

- Markov case

Part I.C

History of GP

History of GP: Theory

1961: Zener

1967: Duffin, Peterson, Zener (Geometric Programming: Theory and Applications)

1967 - 1980: many generalizations, structures of convexity and duality

1971: Zener (Engineering Design by Geometric Programming)

1976: Beightler and Philips (Applied Geometric Programming)

1980: Avriel (Advances in Geometric Programming)

1970, 1976, 1980: 3 SIAM Review papers

History of GP: Algorithms

1960s-1970s: Classical method: dual-based, cutting plane, ...

1996: Primal-dual interior-point method (Kortanek, Xu, Ye)

2005: Robust GP

1993-2005: Distributed algorithm for some GP

History of GP: Applications

1960s - 1970s: Mechanical/civil engineering: structure design

1960s - 1970s: Chemical engineering: statistical mechanical equilibrium

1960s - 1970s: Economics: growth modelling

1960s - 1970s: Limited applications in optimal control and network flow

Modern applications in EE and CS:

Late 1990s: Analog circuit design (Hershensen, Boyd, Lee)

2000 - 2005: A variety of problems in communication systems

GP for Communication Systems

1. Information Theory and Coding:

- Channel capacity and rate distortion
- Channel coding
- Large deviation bounds

2. Network Resource Allocation:

- Wireless network power control
- Network control protocol analysis
- Rate allocation and admission control
- Proportional allocation, market equilibrium theory

3. Signal Processing Algorithms
4. Queuing System Optimization

Where Are We Now?

Since mid 1990s, for GP we have:

- Very efficient, quite robust, sometimes distributed algorithms
- Surprising new applications in Electrical Engineering and Computer Science
- Understand not just 'how', but also 'why' it is useful

Appreciation-Application cycle:

Compared to other convex optimization e.g., SDP and applications, sill not many people are aware of new advances in GP

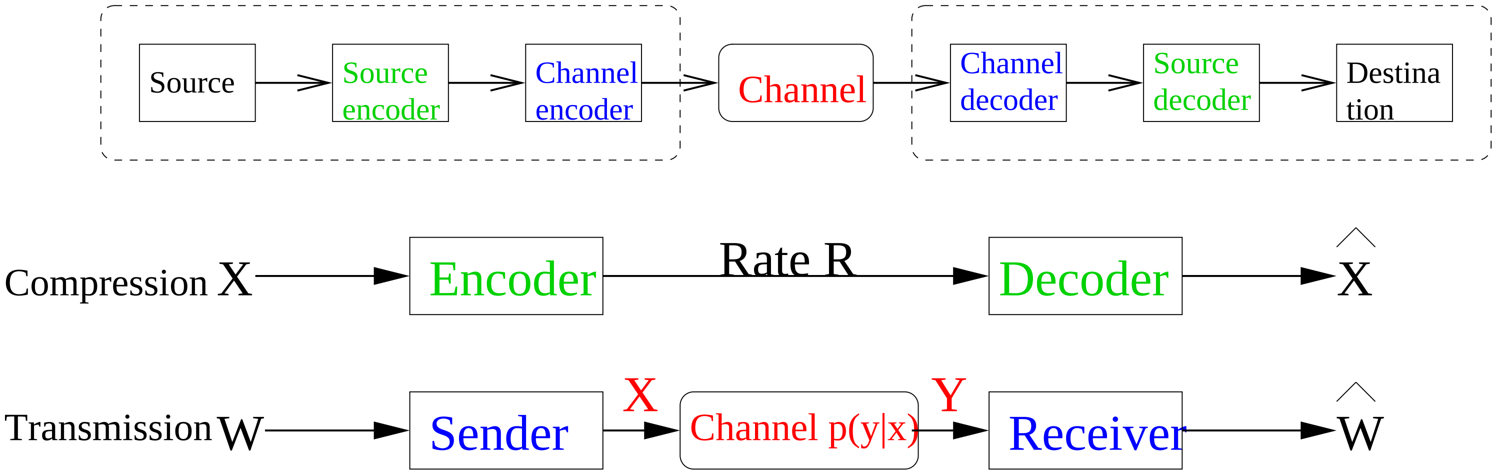
Part II.A

GP and Information Theory

Basic Information Theory

Fundamental limits of data transmission and compression

Rate distortion and channel capacity (Shannon 1948, 1959):



- What's the minimum rate needed for a small distortion?
- Can reliable transmission be done: decoding error probability $\to 0$?

Channel Capacity

Given channel $P_{ij} = \mathbf{Prob}\{Y = j | X = i\}, \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, M$

A distribution $\mathbf{p} \in \mathbf{R}^{1\times N}$ on the input, together with a given channel matrix $\mathbf{P}$, induces a distribution $\mathbf{q} \in \mathbf{R}^{1\times M}$ on the output by $\mathbf{q} = \mathbf{p}\mathbf{P}$.

Associate with each input alphabet symbol i an input cost $s_i \geq 0$

Capacity $C(S)$ under input cost constraint:

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

Mutual information between input X and output Y:

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

$r_i = -\sum_{j=1}^{M} P_{ij} \log P_{ij}$: conditional entropy of Y given $X = i$.

GP Solves Channel Capacity

Lagrange dual of the channel capacity problem is a GP:

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

Optimization variables: $\boldsymbol{\alpha}$ and γ. Constant parameters: $\mathbf{P}, \mathbf{s}$ and S.

GP in standard form:

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

Optimization variables: $\mathbf{z}$ and w.

Some of the Implications

- Weak duality. Any feasible $(\boldsymbol{\alpha}, \gamma)$ of the Lagrange dual problem produce 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 is $C(S)$.
- Also can recover the optimal primal variables, i.e., the capacity achieving input distribution, from the optimal dual variables.
- By complementary slackness, from the optimal dual variables $(\boldsymbol{\alpha}^*, \gamma^*)$, we immediately obtain the support of the capacity achieving input distribution:

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

Upper Bound Generation

Inequality constraints in the dual problem are affine $\Rightarrow$ Easy to find a dual feasible $\boldsymbol{\alpha}$ and upper bound on $C(S)$

Example: a maximum likelihood receiver selecting $\mathrm{argmax}_i\, P_{ij}$ for each output symbol j, and

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

which is tight if and only if the optimal output distribution $\mathbf{q}^*$ is

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

With an input cost constraint $\mathbf{ps} \leq S$, the above upper bound becomes

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

where each maximum likelihood decision is modified by the costs

Error Exponent

Average decoding error probability $\bar{P}_e^{(N)}(R)$ decreases exponentially as the codebook length N tends to infinity:

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

Random coding exponent $E_r(R)$ is the maximized value of:

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

where

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

Maximizing E_0 over $\mathbf{p}$ for a given ρ:

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

Lagrange dual problem: unconstrained concave maximization over $\boldsymbol{\alpha}$

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

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

Corollary: Maximum achievable rate R with finite codeword blocklength N under a decoding error probability $\bar{P}_{e,N}$ upper bounded by

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

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

Rate Distortion Problem

• A source produces a sequence of i.i.d. random variables $X_1, X_2, \ldots, X_n \sim \mathbf{p}$

• An encoder describes the source sequence X^n by an index $f_n(x^n) \in \{1, 2, \ldots, 2^{nR}\}$

• A 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

Rate distortion function $R(D)$ gives the minimum rate needed to describe the source so that distortion is smaller than D:

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

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

GP Solves Rate Distortion

Lagrange dual of the rate distortion problem is a GP:

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

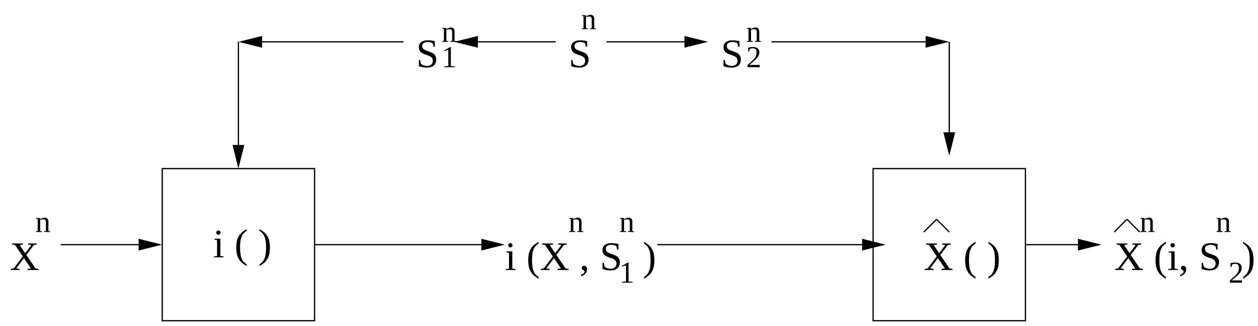
Optimization variables: $\boldsymbol{\alpha}$ and γ. Constant parameters: $\mathbf{p}, d_{ij}$ and D.

GP in standard form:

$$\begin{array}{ll} \text{maximize} & w^{-D} \prod_{i=1}^{N} z_i^{p_i} \\ \text{subject to} & \sum_{i=1}^{N} p_i z_i w^{-d_{ij}} \leq 1, \;\; j = 1, 2, \ldots, M, \\ & w \geq 1, \quad z_i \geq 0, \;\; i = 1, 2, \ldots, N \end{array}$$

Optimization variables: $\mathbf{z}$ and w.

Rate Distortion with State Information



- Correlated random variables (X, S_1, S_2) i.i.d. $\sim p(x, s_1, s_2)$ with finite alphabet sets

S_1 at sender, S_2 at receiver

Reconstruct $\hat{X}$ with distortion less than D

- Rate distortion with state information is known

Lower Bound Generation

Lagrange dual problem is another GP

Dual feasible points:

$\mu_{il} = \left(\sum_{k'} q_{k'} Q_{k'il}\right) \log \frac{1-D}{\max_k Q_{kil}}$, $\gamma = \log\left(\frac{(1-D)(N-1)}{D}\right)$

(N: size of source alphabet set)

$$R_{S_1,S_2}(D) \geq -H_0(D) - D\log(N-1) +$$
$$\sum_{i,l} \mathbf{Prob}\{X=i, S_1=l\}(-\max_k \log \mathbf{Prob}\{X=i, S_1=l|S_2=k\})$$

(H_0: binary entropy function)

Source Coding Problem

A source code $\mathcal{C}$ for a random variable X: a mapping from the range of X to the set of finite length strings of symbols from a W-ary alphabet

$\mathcal{C}(i)$: codeword corresponding to $X = i$

l_i: length of $\mathcal{C}(i)$, $i = 1, 2, \ldots, N$

Prefix code: no codeword is a prefix of any other codeword

Integer optimization problem:

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

Let $z_i = W^{-l_i}$, relaxed codeword length minimization is GP:

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

Exponential penalty:

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

is another GP:

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

Free Energy Interpretation

Maximizing the number of typical sequences is Lagrange dual to an unconstrained GP

Minimizing the Lagrangian of rate distortion is a Gibbs free energy minimization problem

Lagrange dual problem of $C(S)$ in GP convex form: Physical interpretation:

- Each output alphabet is a state
- Each dual variable is energy
- Dual objective: maximize Helmholtz free energy
- Dual constraints: average energy constraints

Shannon Duality Through Lagrange Duality

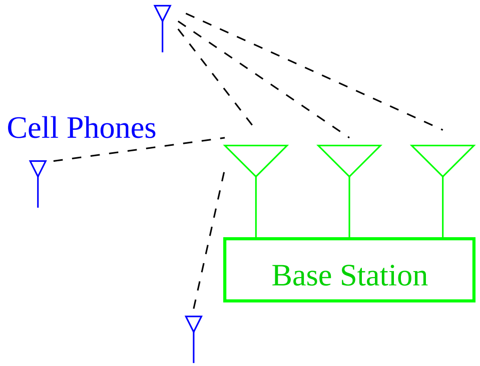
Channel capacity $C(S)$		Rate distortion $R(D)$
monomial (posynomial)	$\leftrightarrow$	posynomial (monomial)
minimization	$\leftrightarrow$	maximization
$\geq$ constraints	$\leftrightarrow$	$\leq$ constraints
j(receiver side index)	$\leftrightarrow$	i(sender side index)
i(sender side index)	$\leftrightarrow$	j(receiver side index)
$M+1$ variables	$\leftrightarrow$	$N+1$ variables
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
$H(\mathbf{P}^{(i)})$	$\leftrightarrow$	$-\log \frac{1}{p_i}$

Part II.B

GP and Network Resource Allocation

Wireless Network Power Control

Wireless CDMA cellular or multi-hop networks:



Competing users all want:

- O1: High data rate
- O2: Low queuing delay
- O3: Low packet drop probability due to channel outage

Optimize over transmit powers $\mathbf{P}$ such that:

- O1, O2 or O3 optimized for 'premium' QoS class (or for maxmin fairness)
- Minimal QoS requirements on O1, O2 and O3 met for all users

A Sample of Power Control Problems

2 classes of traffic traverse a multihop wireless network:

maximize	Total System Throughput
subject to	Data $\text{Rate}_1 \geq$ Rate Requirement_1
	Data $\text{Rate}_2 \geq$ Rate Requirement_2
	Queuing $\text{Delay}_1 \leq$ Delay Requirement_1
	Queuing $\text{Delay}_2 \leq$ Delay Requirement_2
	Drop $\text{Prob}_1 \leq$ Drop Requirement_1
	Drop $\text{Prob}_2 \leq$ Drop Requirement_2
variables	Powers

Wireless Channel Models

Signal Interference Ratio:

$$\mathsf{SIR}_i(\mathbf{P}) = \frac{P_i G_{ii}}{\sum_{j \neq i}^{N} P_j G_{ij} + n_i}.$$

Attainable data rate at high SIR:

$$c_i(\mathbf{P}) = \frac{1}{T} \log_2(K \mathsf{SIR}_i(\mathbf{P})).$$

Outage probability on a wireless link:

$$P_{o,i}(\mathbf{P}) = \mathbf{Prob}\{\mathsf{SIR}_i(\mathbf{P}) \leq \mathsf{SIR}_{th}\}$$

Average (Markovian) queuing delay with Poisson(Λ_i) arrival:

$$\bar{D}_i(\mathbf{P}) = \frac{1}{c_i(\mathbf{P}) - \Lambda_i}$$

Wireless Network Power Control

This suite of nonlinear nonconvex power control problems can be solved by GP (in standard form)

- Global optimality obtained efficiently
- For many combination of objectives and constraints
- Multi-rate, Multi-class, Multi-hop
- Feasibility $\Rightarrow$ Admission control
- Reduction in objective $\Rightarrow$ Admission pricing

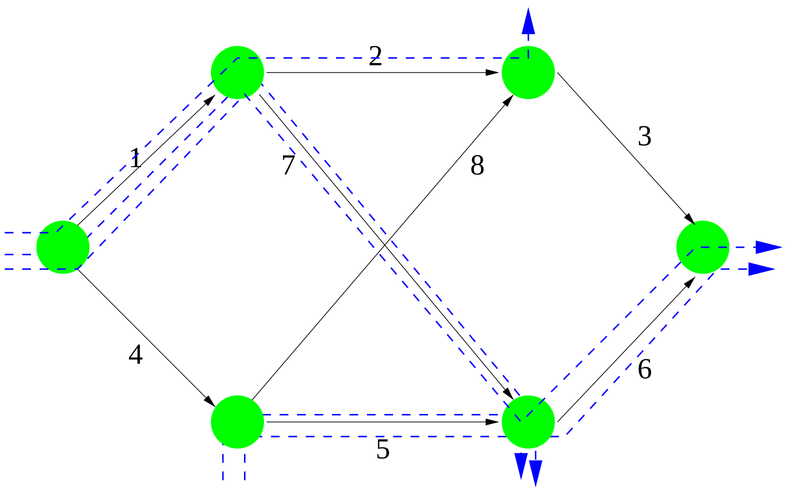
Earlier power control methods can only deal with

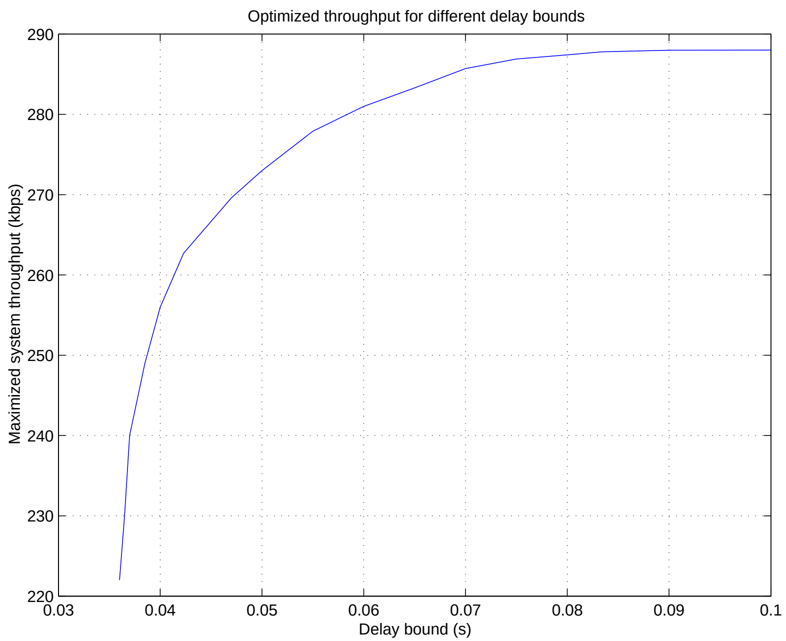
- Single link
- Single class constraints on data rates
- Linear objectives (sum of powers)

Numerical Example: Optimal Throughput-Delay Tradeoff

- 6 nodes, 8 links, 5 multi-hop flows, Direct Sequence CDMA
- max. power 1mW, target BER 10^{-3}, path loss $=$ distance^{-4}

Maximized throughput of network increases as delay bound relaxes





Heuristics: Delay bound violation or system throughput reduction

Low SIR Case

- SIR($\mathbf{P}$) is an inverted posynomial
- $(1 + \text{SIR}(\mathbf{P}))$ is a ratio of two posynomials

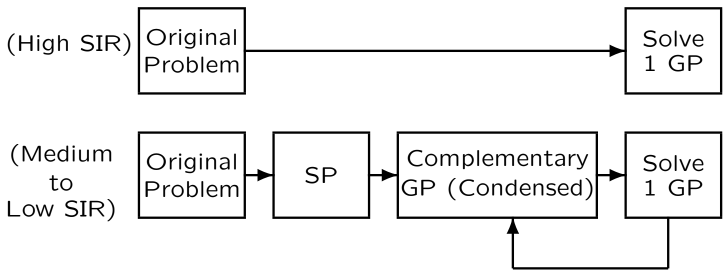
QoS constrained system throughput maximization:

$$
\begin{array}{rrcl}
\text{maximize} & R_{system}(\mathbf{P}) & & \\
\text{subject to} & R_i(\mathbf{P}) & \geq & R_{i,min}, \forall i, \\
 & P_{o,i}(\mathbf{P}) & \leq & P_{o,i,max}, \forall i, \\
 & P_i & \leq & P_{i,max}, \forall i,
\end{array}
$$

which is explicitly written out as:

$$
\begin{array}{rrcl}
\text{minimize} & \displaystyle\prod_{i=1}^{N} \frac{1}{1+\mathsf{SIR}_i} & & \\
\text{subject to} & \displaystyle(2^{TR_{i,min}}-1)\frac{1}{\mathsf{SIR}_i} & \leq & 1, \forall i, \\
& \displaystyle(\mathsf{SIR}_{th})^{N-1}(1-P_{o,i,max})\prod_{i\neq k}^{N}\frac{G_{ik}P_k}{G_{ii}P_i} & \leq & 1, \forall i, \\
& P_i(P_{i,max})^{-1} & \leq & 1, \forall i,
\end{array}
$$

Posynomial constraints but signomial objective function

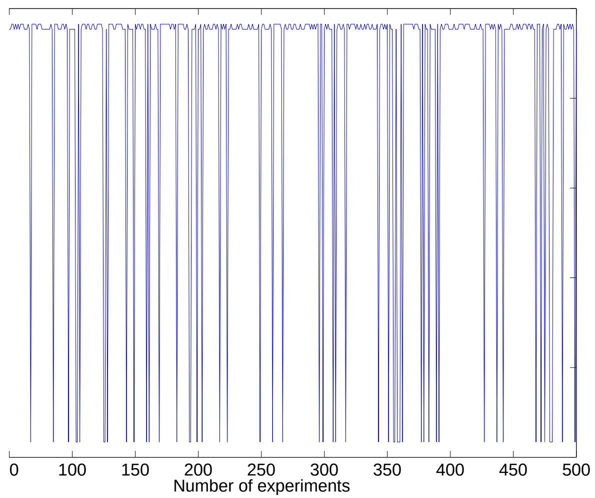


Numerical Example

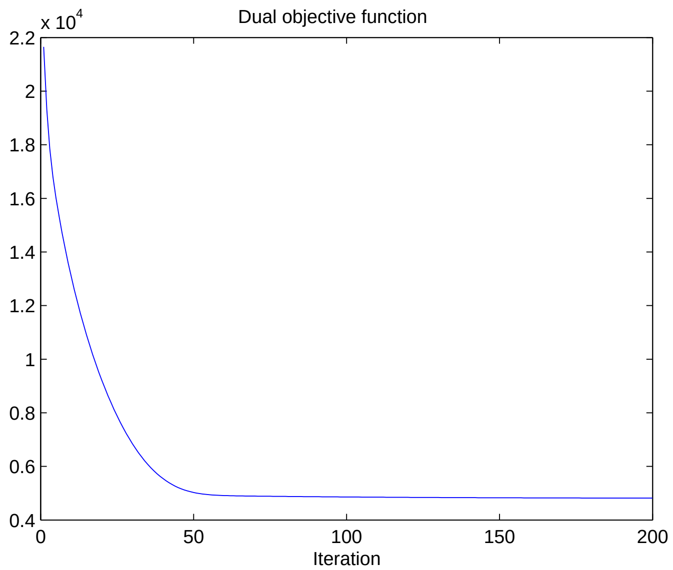
Obtained the globally optimal power allocation in 96% of trials

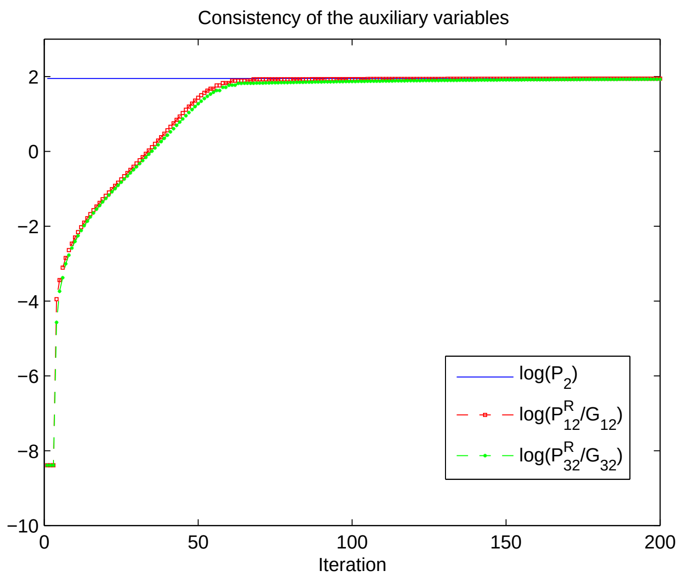
Achieved 100% success rate after solving one more SP

Efficient way to solve this NP-hard problem



Distributed GP





Summary

- Theory, algorithm, and modeling techniques for GP
- Extensions to distributed algorithm and truly nonconvex formulations
- Wide range of applications to communication systems and networks
- Start to know why it is useful, e.g., connection with large deviation theory