Geometric Programming

Chee Wei Tan

Convex Optimization and its Applications to Computer Science

Outline

- Geometric Programming in Standard and Convex Forms
- Convexity of LogSumExp
- Generalized GP
- Dual GP
- Examples
- Signomial Programming

Monomials and Posynomials

Monomial as a function $f : \mathbf{R}^n_+ \rightarrow \mathbf{R}$:

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

where the multiplicative constant $d \geq 0$ and the exponential constants $a^{(j)} \in \mathbf{R}, j = 1, 2, \dots, n$

Sum of monomials is called a posynomial:

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

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

Standard GP and Convex GP

- GP standard form with variables x:

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

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 with variables y:

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

In convex form, GP with only monomials reduces to LP

Convexity of LogSumExp

Log sum inequality (readily proved by the convexity of $f(t) = t\log t, t \geq 0$):

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

where $a_i, b_i \in \mathbf{R}_+, i = 1, 2, \ldots, n$

Let $\hat{b}_i = \log b_i$ and $\sum_{i=1}^{n} a_i = 1$:

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

So LogSumExp is the conjugate function of negative entropy

Since all conjugate functions are convex, LogSumExp is convex

GP and Convexity

- The following problem can be turned into an equivalent standard GP:

$$
\begin{array}{ll}
\text{maximize} & x/y \\
\text{subject to} & 2 \leq x \leq 3 \\
& x^2 + 3y/z \leq \sqrt{y} \\
& x/y = z^2
\end{array}
$$

$$
\begin{array}{ll}
\text{minimize} & x^{-1}y \\
\text{subject to} & 2x^{-1} \leq 1, \quad (1/3)x \leq 1 \\
& x^2 y^{-1/2} + 3y^{1/2} z^{-1} \leq 1 \\
& xy^{-1}z^{-2} = 1
\end{array}
$$

- Let p, q be posynomials and r monomial

$$
\begin{array}{ll}
\text{minimize} & p(x)/(r(x) - q(x)) \\
\text{subject to} & r(x) > q(x)
\end{array}
$$

which is equivalent to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & p(x) \leq t(r(x) - q(x)) \\ & (q(x)/r(x)) < 1 \end{array}$$

which is in turn equivalent to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & (p(x)/t + q(x))/r(x) \leq 1 \\ & (q(x)/r(x)) < 1 \end{array}$$

- Generalized posynomials: f is a generalized posynomial if it can be formed using addition, multiplication, positive power, and maximum, starting from posynomials. Composition of posynomials.

- Generalized GP: minimize generalized posynomials over upper bound inequality constraints on other generalized posynomials
  Generalized GP can be turned into equivalent standard GP

Generalized GP

- Rule 1: Composing posynomials $\{f_{ij}(\mathbf{x})\}$ with a posynomial with nonnegative 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:

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

Unconstrained GP

minimize

$$f(x) = \log\left(\sum_{i=1}^{m} \exp(a_i^T x + b_i)\right)$$

Optimality condition has no analytic solution:

$$\nabla f(x^*) = \frac{1}{\sum_{j=1}^{m} \exp(a_j^T x^* + b_j)} \sum_{i=1}^{m} \exp(a_i^T x^* + b_i) a_i = 0$$

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 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:

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

where variables are $\nu_i, \ \ i = 0, 1, \ldots, m$

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: DMC Capacity Problem

Discrete Memoryless Channel (DMC): $x \in \mathbf{R}^n$ is distribution of input; $y \in \mathbf{R}^m$ is distribution of output;

$P \in \mathbf{R}^{m \times n}$ gives conditional probabilities: $y = Px$

Primal Channel Capacity Problem:

$$\begin{array}{ll} \text{maximize} & -c^T x - \sum_{i=1}^m y_i \log y_i \\ \text{subject to} & x \geq 0,\ \mathbf{1}^T x = 1,\ y = Px, \end{array}$$

where $c_j = -\sum_{i=1}^m p_{ij} \log p_{ij}$

Dual Channel Capacity Problem is a simple GP:

$$\begin{array}{ll} \text{minimize} & \log \sum_{i=1}^m e^{u_i} \\ \text{subject to} & c + P^T u \geq 0, \end{array}$$

Example: Machine Proving of Math Inequalities

The Arithmetic-Geometric Mean Inequality: Let $\boldsymbol{\alpha} \in R_+^N$ be a given positive probability vector, i.e., $\boldsymbol{\alpha}^T \mathbf{1} = 1$. Then, for any N positive real numbers $v_1, \ldots, v_N$, we have $\displaystyle\sum_{i=1}^N \alpha_i v_i \geq \prod_{i=1}^N v_i^{\alpha_i}$.

Solve a standard GP problem:

$$\text{minimize} \quad \sum_{i=1}^N \alpha_i \left(v_i^{1-\alpha_i} \prod_{j \neq i}^N v_j^{-\alpha_j} \right)$$

whose optimal value is 1 and achieved by $v_1 = v_2 = \cdots = v_N$.

Example: Machine Proving of Math Inequalities

The 26th Vojtěch Jarník International Mathematical Competition, Ostrava 2016: Let a, b, c be positive real numbers with $a + b + c = 1$. Prove that $\left(\frac{1}{a} + \frac{1}{bc}\right)\left(\frac{1}{b} + \frac{1}{ac}\right)\left(\frac{1}{c} + \frac{1}{ab}\right) \geq 1728$.

Solve a standard GP relaxation that is tight at optimality[1]:

$$
\begin{array}{ll}
\text{minimize} & pqr \\
\text{subject to} & p^{-1}a^{-1} + p^{-1}b^{-1}c^{-1} \leq 1, \\
& q^{-1}b^{-1} + q^{-1}a^{-1}c^{-1} \leq 1, \\
& r^{-1}c^{-1} + r^{-1}a^{-1}b^{-1} \leq 1, \\
& a + b + c \leq 1,
\end{array}
$$

whose optimal value is 1728 and achieved by $a = b = c = 1/3$.

[1] https://docs.mosek.com/modeling-cookbook/expo.html

Example: Machine Proving of Math Inequalities

Let a, b, c be positive real numbers with $abc = 1$. Prove that

$$\frac{1}{a^3(b+c)} + \frac{1}{b^3(a+c)} + \frac{1}{c^3(a+b)} \geq \frac{3}{2}.$$

In mathematics, the Nesbitt's inequality[2] states that for positive real numbers a, b and c,

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geq \frac{3}{2},$$

with equality if and only if $a = b = c$. Can you use generalized GP to validate or prove these mathematical inequalities?

[2] Nesbitt, A.M., Problem 15114, Educational Times, 2, 37–38, 1903. the Nesbitt's inequality is a special case of the famous Shapiro's inequality proposed by H. Shapiro in 1954.

Signomial Programming (SP)

- A signomial is a linear combination of monomials of some positive variables $x_1, \ldots, x_n$. Signomials are more general than posynomials (which are signomials with all positive coefficients).

- Consider the following Signomial Programming (SP) problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m, \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p. \end{array} \tag{1}$$

where $f_0(\mathbf{x})$ is convex, and $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$, $h_1(\mathbf{x}), \ldots, h_p(\mathbf{x})$ are all signomials.

Signomial Programming (SP)

- Some SP's are already convex, while some others can be transformed to GPs, and therefore solved efficiently.

- An example of a SP that can be efficiently solved: Each inequality constraint signomial $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$ has exactly one term with a negative coefficient: $f_i = p_i - q_i$ where p_i is posynomial, and q_i is monomial. Each equality constraint signomial $h_1(\mathbf{x}), \ldots, h_p(\mathbf{x})$ has exactly one term with a positive coefficient and one term with a negative coefficient: $h_i = r_i - s_i$ where r_i and s_i are monomials.

- An example of a SP that is convex: $h_i(\mathbf{x})$, $\forall\, i$ are linear and $f_i(\mathbf{x})$ is a sum with each summand being one of two forms: $c_i \prod_{j=1} x_j^{\alpha_j}$ where either $c_i > 0$ and $\alpha_j \leq 0$ or else $c_i < 0$ and $\alpha_j > 0$ with $\mathbf{1}^T \boldsymbol{\alpha} \leq 1$ for $i = 1, \ldots, m$.

Successive Convex Approximation Method

- Since SP is nonconvex in general, directly solving (1) can be difficult. Let us solve it by an iterative series of approximations $\tilde{f}_i(\mathbf{x}) \approx f_i(\mathbf{x}), \forall \mathbf{x}$, each of which can be optimally solved easily.

- If the approximations satisfy the following three properties, then the solutions of an iterative series of approximations converge to a point satisfying the Karush-Kuhn-Tucker (KKT) conditions of the original problem.

1. $f_i(\mathbf{x}) \leq \tilde{f}_i(\mathbf{x})$ for all $\mathbf{x}$,
2. $f_i(\mathbf{x}_0) = \tilde{f}_i(\mathbf{x}_0)$, where x_0 is the optimal solution of the approximated problem in the previous iteration,
3. $\nabla f_i(\mathbf{x}_0) = \nabla \tilde{f}_i(\mathbf{x}_0)$.

Successive Convex Approximation Method

- Condition (1) guarantees that the approximation $\tilde{f}_i(\mathbf{x})$ is tightening the constraints in Problem (1), and any solution of the approximated problem will be a feasible point of the original problem in Problem (1).

- Condition (2) guarantees that the solution of each approximated problem will decrease the cost function: $f_0\left(\mathbf{x}^{(k)}\right) \leq f_0\left(\mathbf{x}^{(k-1)}\right)$, where $x^{(k)}$ is the solution to the k-th approximated problem.

- Condition (3) guarantees that the KKT conditions of the original Problem (1) will be satisfied after the series of approximations converges.

Successive Convex Approximation Method

Algorithm Successive approximation to a nonconvex problem

Input Method to approximate $f_i(\mathbf{x})$ with $\tilde{f}_i(\mathbf{x}), i = 1, \ldots, m$, around some point of interest $\mathbf{x}_0$.

Output A vector that satisfies the KKT conditions of the original problem.

0) Choose an initial feasible point $\mathbf{x}^{(0)}$ and set $k = 1$.

1) Form the k-th approximated problem of (1) based on approximating $f_i(\mathbf{x})$ with $\tilde{f}_i(\mathbf{x})$ around the previous point $\mathbf{x}^{(k-1)}$.

2) Solve the k-th approximated problem to obtain $\mathbf{x}^{(k)}$.

3) Increment k and go to step 2 until convergence to a stationary point.

Single Condensation Method by GP

- An inequality with a signomial can be rewritten into a ratio of posynomials upper bounded by one: $f(\mathbf{x})/g(\mathbf{x}) \leq 1$. Next, approximate the denominator of this ratio, $g(\mathbf{x})$, with a monomial $\tilde{g}(\mathbf{x})$, but leave the numerator $f(\mathbf{x})$ as a posynomial.

- *Lemma 1*: Let $g(\mathbf{x}) = \sum_i u_i(\mathbf{x})$ be a posynomial. Then

$$g(\mathbf{x}) \geq \tilde{g}(\mathbf{x}) = \prod_i \left(\frac{u_i(\mathbf{x})}{\alpha_i} \right)^{\alpha_i}.$$

  If, in addition, $\alpha_i = u_i(\mathbf{x}_0)/g(\mathbf{x}_0), \forall i$, for any fixed positive $\mathbf{x}_0$, then $\tilde{g}(\mathbf{x}_0) = g(\mathbf{x}_0)$, and $\tilde{g}(\mathbf{x}_0)$ is the best local monomial approximation to $g(\mathbf{x}_0)$ near $\mathbf{x}_0$ in the sense of first order Taylor approximation.

Single Condensation Method by GP

Proof:

- Arithmetic-geometric mean inequality: $\sum_i \alpha_i v_i \geq \prod_i v_i^{\alpha_i}$, 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 $\sum_i u_i \geq \prod_i (u_i/\alpha_i)^{\alpha_i}$.

- The inequality becomes an equality if we let $\alpha_i = u_i / \sum_i u_i, \forall i$, which satisfies the condition that $\boldsymbol{\alpha} \succeq 0$ and $\mathbf{1}^T \boldsymbol{\alpha} = 1$.

- The best local monomial approximation $\tilde{g}(\mathbf{x}_0)$ of $g(\mathbf{x}_0)$ near $\mathbf{x}_0$ can be easily verified.

Single Condensation Method by GP

Proposition 1:

The approximation of a ratio of posynomials $f(\mathbf{x})/g(\mathbf{x})$ with $f(\mathbf{x})/\tilde{g}(\mathbf{x})$, where $\tilde{g}(\mathbf{x})$ is the monomial approximation of $g(\mathbf{x})$ using the Arithmetic-Geometric Mean Inequality approximation in Lemma 1, satisfies the three conditions for the convergence of the successive approximation method.

Proof:

Conditions (1) and (2) are clearly satisfied since $g(\mathbf{x}) \geq \tilde{g}(\mathbf{x})$ and $\tilde{g}(\mathbf{x}_0) = g(\mathbf{x}_0)$ (Lemma 1). Condition (3) is easily verified by taking derivatives of $g(\mathbf{x})$ and $\tilde{g}(\mathbf{x})$.

Single Condensation Method by GP

- Suppose we want to minimize $f_0(\mathbf{x})$ subject to an equality constraint on a ratio of posynomials: $f(\mathbf{x})/g(\mathbf{x}) = 1$.

- We can rewrite this as minimizing $f_0(\mathbf{x}) + \phi t$ subject to $f(\mathbf{x})/g(\mathbf{x}) \leq 1$ and $f(\mathbf{x})/g(\mathbf{x}) \geq 1 - t$ where ϕ is a sufficiently large number to guarantee that the optimum solution will have $t \approx 0$. [3]

- The second constraint can be rewritten as $g(\mathbf{x})/f(\mathbf{x} + tg(\mathbf{x})) \leq 1$ where we can apply the single condensation method to the denominators of both constraints.

[3] One choice is to use $\phi = 1 + k$ at the k-th iteration.

Double Condensation Method by GP

- Another choice of approximation to $f(\mathbf{x})/g(\mathbf{x}) \leq 1$ is to make a double monomial approximation for both the denominator and numerator.
- We can still use the arithmetic-geometric mean approximation of Lemma 1 as a monomial approximation for the denominator $g(\mathbf{x})$.
- But, Lemma 1 cannot be used as a monomial approximation for the numerator $f(\mathbf{x})$.
- To satisfy the three conditions for the convergence of the successive approximation method, a monomial approximation for the numerator $f(\mathbf{x})$ should satisfy $f(\mathbf{x}) \leq \tilde{f}(\mathbf{x})$.

GP Properties and Software

- 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
- Surprisingly wide range of applications

Freely available software: Stanford CVX and GGPLAB

`https://web.stanford.edu/~boyd/ggplab`

Summary

- Nonlinearity in Posynomial, Generalized Posynomial, LogSumExp
- Standard GP, Convex GP, Dual GP, Generalized GP (GGP), SP
- A variety of problems: Structural design in mechanical engineering, Growth modeling in economics (1960s-1970s), Analog and digital circuit design (late 1990s), Communication system problems (early 2000s) and many other applications in practice and analysis

Reading assignment: Sections 4.5, 5.7 of textbook.

- S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A Tutorial on Geometric Programming," Optimization and Engineering, 8(1):67-127, 2007
- M. Chiang, "Geometric programming for communication systems," Foundations and Trends in Communications and Information Theory, 2005