Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 2, Lecture 6: Geometric programming

February 15, 2019 *University of Illinois at Urbana-Champaign*

1 Unconstrained geometric programs

Unconstrained programs are a generalization of the final example problem from the previous lecture, where we wanted to minimize $f(x,y) = 2xy + \frac{y}{x^2} + \frac{3x}{y}$ for $x, y > 0$.

Define a *posynomial term* in variables $t_1, t_2, \dots, t_m$ to be a function of the form

$$Ct_1^{\alpha_1} t_2^{\alpha_2} \cdots t_m^{\alpha_m}$$

where $\alpha_1, \alpha_2, \dots, \alpha_m$ are real powers (not necessarily integers, and not necessarily positive), and $C > 0$ is a positive real constant.

A *posynomial* is just a sum of such posynomial terms. These are supposed to be analogous to polynomials: we generalize polynomials by allowing arbitrary powers, but add the restriction that all coefficients must be positive (hence the prefix "pos-").

An unconstrained geometric program (GP) is the problem of minimizing a posynomial over all positive real inputs. Formally, the problem is

$$\begin{aligned} \underset{\mathbf{t}\in\mathbb{R}^m}{\text{minimize}} \quad & g(t) = \text{Term}_1(\mathbf{t}) + \text{Term}_2(\mathbf{t}) + \cdots + \text{Term}_n(\mathbf{t}) \\ \text{subject to} \quad & t_1, t_2, \dots, t_m > 0 \end{aligned}$$

where for $1 \le i \le n$, $\text{Term}_i(\mathbf{t})$ is a posynomial term

$$\text{Term}_i(\mathbf{t}) = C_i t_1^{\alpha_{i1}} t_2^{\alpha_{i2}} \cdots t_m^{\alpha_{im}}.$$

For example, when minimizing $f(x,y) = 2xy + \frac{y}{x^2} + \frac{3x}{y}$, $\text{Term}_1(x,y) = 2x^1y^1$, $\text{Term}_2(x,y) = x^{-2}y^1$, and $\text{Term}_3(x,y) = 3x^1y^{-1}$.

These are called geometric programs because we plan to use the AM-GM inequality to solve them. We will choose weights $\delta_1, \delta_2, \dots, \delta_n > 0$ with $\delta_1 + \delta_2 + \cdots + \delta_n = 1$ and then use the inequality

$$\begin{aligned} \text{Term}_1(\mathbf{t}) + \text{Term}_2(\mathbf{t}) + \cdots + \text{Term}_n(\mathbf{t}) &= \delta_1 \cdot \frac{\text{Term}_1(\mathbf{t})}{\delta_1} + \delta_2 \cdot \frac{\text{Term}_2(\mathbf{t})}{\delta_2} + \cdots + \delta_n \cdot \frac{\text{Term}_n(\mathbf{t})}{\delta_n} \\ &\ge \left(\frac{\text{Term}_1(\mathbf{t})}{\delta_1}\right)^{\delta_1} \left(\frac{\text{Term}_2(\mathbf{t})}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{\text{Term}_n(\mathbf{t})}{\delta_n}\right)^{\delta_n} \end{aligned}$$

to obtain a lower bound on $g(\mathbf{t})$.

[1] This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-spring-2019.html

2 The dual of the unconstrained GP

Not all weights $\delta_1, \delta_2, \ldots, \delta_n$ will work, of course. In the case of an unconstrained GP, the only time that this inequality actually gives us a lower bound on $g(\mathbf{t})$ is when the powers of $t_1, t_2, \ldots, t_n$ cancel and we get a constant on the right-hand side of the AM-GM inequality.

What do we need for that to happen? Consider a general upper bound on $f(x,y) = 2xy + \frac{y}{x^2} + \frac{3x}{y}$: we have

$$f(x,y) \geq \left(\frac{2xy}{\delta_1}\right)^{\delta_1} \left(\frac{y/x^2}{\delta_2}\right)^{\delta_2} \left(\frac{3x/y}{\delta_3}\right)^{\delta_3}$$

where x appears to the power $\delta_1 - 2\delta_2 + \delta_3$ and y appears to the power $\delta_1 + \delta_2 - \delta_3$. To get a constant lower bound, we need to choose values of δ_1, δ_2, and δ_3 satisfying

$$\begin{aligned}
\delta_1 - 2\delta_2 + \delta_3 &= 0, && \text{(power of } x\text{)}\\
\delta_1 + \delta_2 - \delta_3 &= 0, && \text{(power of } y\text{)}\\
\delta_1 + \delta_2 + \delta_3 &= 1, && \text{(AM-GM requirement)}\\
\delta_1, \delta_2, \delta_3 &> 0. && \text{(AM-GM requirement)}
\end{aligned}$$

In general, the lower bound contains the variable t_i raised to the power $\delta_1\alpha_{1i} + \delta_2\alpha_{2i} + \cdots + \delta_n\alpha_{ni}$ and we get $m+1$ equations: one from the AM-GM requirement and m from each of the variables $t_1, \ldots, t_m$.

In the example problem, there is only one solution to the system of equations in $\boldsymbol{\delta}$. But in general, there could be many solutions. In such a case, it makes sense to pick the best lower bound.

Assuming that the powers of $t_1, t_2, \ldots, t_m$ all cancel, we can simplify the lower bound:

$$g(\mathbf{t}) \geq \left(\frac{\text{Term}_1(\mathbf{t})}{\delta_1}\right)^{\delta_1} \left(\frac{\text{Term}_2(\mathbf{t})}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{\text{Term}_n(\mathbf{t})}{\delta_n}\right)^{\delta_n} = \left(\frac{C_1}{\delta_1}\right)^{\delta_1} \left(\frac{C_2}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{C_n}{\delta_n}\right)^{\delta_n}.$$

So the problem of picking the best lower bound we can itself becomes an optimization problem, called the *dual geometric program*:

$$\begin{aligned}
\underset{\boldsymbol{\delta} \in \mathbb{R}^n}{\text{maximize}} \quad & v(\boldsymbol{\delta}) = \left(\frac{C_1}{\delta_1}\right)^{\delta_1} \left(\frac{C_2}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{C_n}{\delta_n}\right)^{\delta_n} \\
\text{subject to} \quad & \delta_1\alpha_{11} + \delta_2\alpha_{21} + \cdots + \delta_n\alpha_{n1} = 0, && \text{(power of } t_1\text{)}\\
& \delta_1\alpha_{12} + \delta_2\alpha_{22} + \cdots + \delta_n\alpha_{n2} = 0, && \text{(power of } t_2\text{)}\\
& \cdots \\
& \delta_1\alpha_{1m} + \delta_2\alpha_{2m} + \cdots + \delta_n\alpha_{nm} = 0, && \text{(power of } t_m\text{)}\\
& \delta_1 + \delta_2 + \cdots + \delta_n = 1, && \text{(AM-GM requirement)}\\
& \delta_1, \delta_2, \ldots, \delta_n > 0. && \text{(AM-GM requirement)}
\end{aligned}$$

Instead of finding the optimal solution $\mathbf{t}^*$ to the original problem, which involves lots of messy calculations with derivatives, we can try to learn something about $\mathbf{t}^*$ by finding the optimal solution $\boldsymbol{\delta}^*$ to the dual problem.

3 Relationship between primal and dual

To go from the dual back to the primal, we can try to use the equality condition of the AM-GM inequality. For any vector $\boldsymbol{\delta}$ that solves the dual geometric program, *if* there is going to be any $\mathbf{t}$ with $g(\mathbf{t}) = v(\delta)$, then the inequality we used must be an equality: we must have

$$\delta_1 \cdot \frac{\mathrm{Term}_1(\mathbf{t})}{\delta_1} + \delta_2 \cdot \frac{\mathrm{Term}_2(\mathbf{t})}{\delta_2} + \cdots + \delta_n \cdot \frac{\mathrm{Term}_n(\mathbf{t})}{\delta_n} = \left(\frac{\mathrm{Term}_1(\mathbf{t})}{\delta_1}\right)^{\delta_1} \left(\frac{\mathrm{Term}_2(\mathbf{t})}{\delta_2}\right)^{\delta_2} \cdots \left(\frac{\mathrm{Term}_n(\mathbf{t})}{\delta_n}\right)^{\delta_n}$$

and we know from the AM-GM inequality that this is only possible if

$$\frac{\mathrm{Term}_1(\mathbf{t})}{\delta_1} = \frac{\mathrm{Term}_2(\mathbf{t})}{\delta_2} = \cdots = \frac{\mathrm{Term}_n(\mathbf{t})}{\delta_n}.$$

Moreover, these must all be equal to their mean $v(\boldsymbol{\delta})$.

But it's possible that this set of equations for $\mathbf{t}$ has no solution. In fact, this will happen a lot of the time.

Theorem 3.1. *Given a dual solution $\boldsymbol{\delta}$, if the equations*

$$\frac{\mathrm{Term}_1(\mathbf{t})}{\delta_1} = \frac{\mathrm{Term}_2(\mathbf{t})}{\delta_2} = \cdots = \frac{\mathrm{Term}_n(\mathbf{t})}{\delta_n} = v(\boldsymbol{\delta})$$

have a solution $\mathbf{t}$ with $t_1, t_2, \dots, t_m > 0$, then actually $\mathbf{t}$ is the primal optimal solution and $\boldsymbol{\delta}$ is the dual optimal solution, with $g(\mathbf{t}) = v(\boldsymbol{\delta})$.

Proof. If we can solve this to get a solution $\mathbf{t}$ with $t_1, t_2, \dots, t_m > 0$, then we have $g(\mathbf{t}) = v(\boldsymbol{\delta})$ by the equality condition of the AM-GM inequality.

That is what guarantees optimality. Since $v(\boldsymbol{\delta})$ is a lower bound on the primal problem, then for any other primal solution $\mathbf{t}'$, we have $g(\mathbf{t}') \geq v(\boldsymbol{\delta})$. So $g(\mathbf{t}') \geq g(\mathbf{t})$, which makes $\mathbf{t}$ optimal.

Similarly, any other dual solution $\boldsymbol{\delta}'$ is a lower bound, so in particular it satisfies $g(\mathbf{t}) \geq v(\boldsymbol{\delta}')$. Therefore $v(\boldsymbol{\delta}) \geq v(\boldsymbol{\delta}')$, which makes $\boldsymbol{\delta}$ optimal. □

This gives us a strategy for trying to solve the geometric program: construct the dual problem, find the optimal solution $\boldsymbol{\delta}^*$, and try to solve the equation above. If it works, then the solution $\mathbf{t}^*$ will be the optimal primal solution.

Will it work? Well, at least it will work if an optimal primal solution exists to begin with:

Theorem 3.2. *If $\mathbf{t}^*$ is an optimal primal solution, then*

$$\boldsymbol{\delta}^* = \left(\frac{\mathrm{Term}_1(\mathbf{t}^*)}{g(\mathbf{t}^*)}, \frac{\mathrm{Term}_2(\mathbf{t}^*)}{g(\mathbf{t}^*)}, \dots, \frac{\mathrm{Term}_n(\mathbf{t}^*)}{g(\mathbf{t}^*)}\right)$$

is an optimal dual solution and $g(\mathbf{t}^*) = v(\boldsymbol{\delta}^*)$.

Proof. If $\mathbf{t}^*$ is an optimal primal solution, then it is a critical point, so $\nabla g(\mathbf{t}^*) = \mathbf{0}$.

The gradient ∇g can be expressed in terms of g. We have

$$\frac{\partial}{\partial t_j}\,\mathrm{Term}_i(\mathbf{t}) = C_i t_1^{\alpha_{i1}} t_2^{\alpha_{i2}} \cdots (\alpha_{ij} t_j^{\alpha_{ij}-1}) \cdots t_m^{\alpha_{im}} = \frac{\alpha_{ij}}{t_j}\,\mathrm{Term}_i(\mathbf{t}).$$

So

$$\frac{\partial}{\partial t_j} g(\mathbf{t}) = \frac{\alpha_{1j}}{t_j}\,\mathrm{Term}_1(\mathbf{t}) + \frac{\alpha_{2j}}{t_j}\,\mathrm{Term}_2(\mathbf{t}) + \cdots + \frac{\alpha_{nj}}{t_j}\,\mathrm{Term}_n(\mathbf{t}).$$

So if $\frac{\partial}{\partial t_j} g(\mathbf{t}^*) = 0$, then we have

$$\alpha_{1j}\,\mathrm{Term}_1(\mathbf{t}^*) + \alpha_{2j}\,\mathrm{Term}_2(\mathbf{t}^*) + \cdots + \alpha_{nj}\,\mathrm{Term}_n(\mathbf{t}^*) = t_j \cdot 0 = 0$$

which means that $(\mathrm{Term}_1(\mathbf{t}^*), \mathrm{Term}_2(\mathbf{t}^*), \ldots, \mathrm{Term}_n(\mathbf{t}^*))$ satisfies the "power of t_j" constraint of the dual GP.

That doesn't make $(\mathrm{Term}_1(\mathbf{t}^*), \mathrm{Term}_2(\mathbf{t}^*), \ldots, \mathrm{Term}_n(\mathbf{t}^*))$ a dual solution: it probably doesn't satisfy the constraint $\delta_1 + \delta_2 + \cdots + \delta_n = 1$. But if we divide all the components of this vector by their sum $g(\mathbf{t}^*)$, we force the result to satisfy that constraint as well.

So the point $\boldsymbol{\delta}^*$ given by $\delta_i = \frac{\mathrm{Term}_i(\mathbf{t}^*)}{g(\mathbf{t}^*)}$ is a feasible dual solution.

Also, we know that

$$\frac{\mathrm{Term}_1(\mathbf{t})}{\delta_1^*} = \frac{\mathrm{Term}_2(\mathbf{t}^*)}{\delta_2^*} = \frac{\mathrm{Term}_n(\mathbf{t}^*)}{\delta_n^*} = g(\mathbf{t}^*)$$

because of how we defined $\boldsymbol{\delta}^*$. So we're in the equality condition of the AM-GM inequality, which means that $g(\mathbf{t}^*) = v(\boldsymbol{\delta}^*)$ and $\boldsymbol{\delta}^*$ is an optimal solution. □

You should notice that the equations in Theorem 3.1 and Theorem 3.2 are inverses of each other: if you use Theorem 3.2 to find $\boldsymbol{\delta}^*$ from $\mathbf{t}^*$, and then use Theorem 3.1 to find a $\mathbf{t}$ from $\boldsymbol{\delta}^*$, we'll get $\mathbf{t} = \mathbf{t}^*$.