Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 6, Lecture 5: Equality Constraints

April 10, 2019 *University of Illinois at Urbana-Champaign*

1 Equality constraints in the KKT theorem

Suppose that we want to solve an optimization problem such as

$$(P) \qquad \begin{cases} \underset{\mathbf{x}\in S}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & g(\mathbf{x}) = 0. \end{cases}$$

(To simplify matters, for now there is only one constraint, and it is an equality constraint.)

The KKT theorem generally deals with inequality constraints. Can we make it work here?

We can encode a single equality constraint as two inequalities, and rewrite the problem as

$$(P') \qquad \begin{cases} \underset{\mathbf{x}\in S}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & g(\mathbf{x}) \le 0, \\ & -g(\mathbf{x}) \le 0. \end{cases}$$

But because these two constraints are identical except for the sign, some further simplifications occur. Writing down the Lagrangian of P' gives us

$$L(\mathbf{x}, \lambda_1, \lambda_2) = f(\mathbf{x}) + \lambda_1 g(\mathbf{x}) + \lambda_2(-g(\mathbf{x})) = f(\mathbf{x}) + (\lambda_1 - \lambda_2)g(\mathbf{x}).$$

So let's define a new variable μ equal to $\lambda_1 - \lambda_2$. Even though we are required to have $\lambda_1, \lambda_2 \ge 0$, their difference isn't constrained to be nonnegative: we can achieve any desired value of μ by setting $\lambda_1 = \max\{\lambda_2, 0\}$ and $\lambda_2 = \max\{-\lambda_2, 0\}$. (This is only one of many ways to get μ.)

Complementary slackness becomes a redundant condition: since we have $g(\mathbf{x}) = 0$ for a feasible $\mathbf{x}$ anyway, we don't learn anything from the equation $\mu g(\mathbf{x}) = 0$.

So the effect of the equality constraint is that the corresponding dual variable is allowed to be any real number. This generalizes. We can have a mix of equality and inequality constraints: a program of the form

$$(P) \qquad \begin{cases} \underset{\mathbf{x}\in S}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \le 0, \quad i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, \ell. \end{cases}$$

Then the saddle point KKT theorem tells us the following. Suppose that $\mathbf{x}^* \in S$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ with $\boldsymbol{\lambda}^* \ge \mathbf{0}$, and $\boldsymbol{\mu}^* \in \mathbb{R}^\ell$ satisfy:

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

1. $L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ for all $\mathbf{x}^* \in \mathbb{R}^n$
2. $L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \ge L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu})$ for all $\boldsymbol{\lambda} \in R^m$ and $\boldsymbol{\mu} \in \mathbb{R}^\ell$ with $\boldsymbol{\lambda} \ge \mathbf{0}$:
3. For $i = 1, \dots, m$, $\lambda_i^* g_i(\mathbf{x}^*) = 0$, and for $j = 1, \dots, \ell$, $h_i(\mathbf{x}^*) = 0$.

Then $\mathbf{x}^*$ is an optimal solution of P. Here, $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is the Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \mu_j h_j(\mathbf{x}).$$

The gradient form of the KKT theorem is changed similarly. Recall that in the gradient form of the KKT, we are already required to check that $\mathbf{x}$ is feasible ourselves: we don't get it for free. So there's no form of complementary slackness at all for $h_1, h_2, \dots, h_\ell$: the only constraint on the μ-variables comes from the equation $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$.

It doesn't quite make sense to think of $\boldsymbol{\mu}$ as a sensitivity vector anymore. However, there is still a more general form of the Slater condition:

Theorem 1.1. *Suppose that P is convex: S is a convex set, $f, g_1, g_2, \dots, g_m$ are convex functions and $h_1, h_2, \dots, h_\ell$ are* linear functions *(possibly with a constant term). We say that P is superconsistent if there is a point $\mathbf{x} \in S$ with $g_i(\mathbf{x}) < 0$ for $i = 1, \dots, m$.*

If P is superconsistent and $\mathbf{x}^*$ is an optimal solution of P, then there are $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ that, together with $\mathbf{x}^*$, satisfy the conditions of the saddle point KKT theorem.

Intuitively, this is our definition of a convex program because that we want both h_i and $-h_i$ to be convex functions. This only happens if $h_1, h_2, \dots, h_\ell$ are all linear. In that case, the feasible region of P is a convex set, despite the equality constraints.

The reason that this slightly-more-general notion of superconsistency still works is that, when we impose some linear equality constraints, we are essentially optimizing over an affine subspace of $\mathbb{R}^n$, which looks like $\mathbb{R}^k$ for some $k \le n$. Working in this subspace, we can apply the usual form of the KKT theorem and the Slater condition.

2 Equality constraints in the penalty method

Compared with this, the penalty method is simpler to work with. In fact, the penalty method only gets less complicated when our constraints are equality constraints. To enforce a constraint $h(\mathbf{x}) = 0$ in the penalty method, we simply add the penalty term $k \cdot h(\mathbf{x})^2$, with no weird $^+$ operator.

In theory, we could prove that this works (and that all of our results continue to apply when we do this) by writing $h(\mathbf{x}) = 0$ as $h(\mathbf{x}) \le 0$ and $-h(\mathbf{x}) \le 0$, and noticing that the penalty terms for these two constraints simplify to the constraint above. But it just makes sense that this penalty method does what we want it to do.

3 Equality constraints in geometric programming

Finally, let's consider what happens when we get creative with the constraints of a geometric program.

A general geometric program looks like

$$(GP) \qquad \begin{cases} \underset{\mathbf{t}\in\mathbb{R}^m}{\text{minimize}} & g_0(\mathbf{t}) \\ \text{subject to} & g_i(\mathbf{t}) \le 1, \quad 1 \le i \le m, \\ & \mathbf{t} > \mathbf{0} \end{cases}$$

where $g_0, g_1, \ldots, g_m$ are posynomials.

There is some flexibility here. For any $C > 0$, we can deal with the constraint $g(\mathbf{t}) \le C$, by rewriting it as $C^{-1}g(\mathbf{t}) \le 1$.

But this cannot be done with a negative C, because the coefficients in a posynomial cannot be positive. Similarly, we cannot add the constraint $g(\mathbf{t}) \ge 1$ to a geometric program: there is no way to put that constraint in standard form. (We can't rewrite it as $-g(\mathbf{t}) \le -1$: this is not a posynomial constraint!) Equality constraints are equally impossible.

There is one exception. Suppose that $g(\mathbf{t})$ is a posynomial with only one term. (A posymonomial??) Then we *can* write down the constraint $g(\mathbf{t}) \ge 1$, by taking the reciprocal of both sides and writing down $\frac{1}{g(\mathbf{t})} \le 1$. And together, the two constraints $\frac{g(\mathbf{t})}{C} \le 1$ and $\frac{C}{g(\mathbf{t})} \le 1$ encode an equality constraint $g(\mathbf{t}) = C$.

Let's look at what this actually does to the dual. Consider the geometric program

$$\begin{aligned} \underset{x,y\in\mathbb{R}}{\text{minimize}} \quad & 3x + y \\ \text{subject to} \quad & x^2 y = 12. \end{aligned}$$

We first put this in standard form as:

$$\begin{aligned} \underset{x,y\in\mathbb{R}}{\text{minimize}} \quad & 3x + y \\ \text{subject to} \quad & \frac{1}{12}x^2 y \le 1, \\ & 12x^{-2}y^{-1} \le 1. \end{aligned}$$

Writing down the dual, we get

$$\begin{aligned} \underset{\delta_1,\delta_2,\delta_3,\delta_4}{\text{minimize}} \quad & \left(\frac{3}{\delta_1}\right)^{\delta_1} \left(\frac{1}{\delta_2}\right)^{\delta_2} \left(\frac{1/12}{\delta_3}\right)^{\delta_3} \left(\frac{12}{\delta_4}\right)^{\delta_4} \delta_3^{\delta_3}\delta_4^{\delta_4} \\ \text{subject to} \quad & \delta_1 + 2\delta_3 - 2\delta_4 = 0, \\ & \delta_2 + \delta_3 - \delta_4 = 0, \\ & \delta_1 + \delta_2 = 1 \end{aligned}$$

with some rather special positivity constraints: we get $\delta_1, \delta_2 > 0$ as usual, but for δ_3 we have "$\delta_3 > 0$ or $\delta_3 = 0$" and for δ_4 we have "$\delta_4 > 0$ or $\delta_4 = 0$", which can just be written as $\delta_3, \delta_4 \ge 0$.

Next, things simplify a lot. We can rewrite this dual geometric program as

$$\begin{aligned}
\underset{\delta_1,\delta_2,\delta_3,\delta_4}{\text{minimize}} \quad & \left(\frac{3}{\delta_1}\right)^{\delta_1}\left(\frac{1}{\delta_2}\right)^{\delta_2}\left(\frac{1}{12}\right)^{\delta_3-\delta_4} \\
\text{subject to} \quad & \delta_1 + 2(\delta_3 - \delta_4) = 0, \\
& \delta_2 + (\delta_3 - \delta_4) = 0, \\
& \delta_1 + \delta_2 = 1, \\
& \delta_1, \delta_2 > 0, \delta_3, \delta_4 \geq 0.
\end{aligned}$$

Every time we see δ_3 or δ_4 in this program, we are always looking at the difference $\delta_3 - \delta_4$. Even though δ_3 and δ_4 are nonnegative, their difference can be anything. So by making the substitution $\eta = \delta_3 - \delta_4$, we get a simpler dual

$$\begin{aligned}
\underset{\delta_1,\delta_2,\eta}{\text{minimize}} \quad & \left(\frac{3}{\delta_1}\right)^{\delta_1}\left(\frac{1}{\delta_2}\right)^{\delta_2}\left(\frac{1}{12}\right)^{\eta} \\
\text{subject to} \quad & \delta_1 + 2\eta = 0, \\
& \delta_2 + \eta = 0, \\
& \delta_1 + \delta_2 = 1, \\
& \delta_1, \delta_2 > 0.
\end{aligned}$$

(The value of η doesn't do anything for us when we are trying to go from the dual back to the primal, apart from being part of the dual objective function. But the constraint that $x^2 y = 12$ will be very useful.)

There is one more trick that can make our geometric programs more flexible. Suppose we have the optimization problem

$$\begin{aligned}
\underset{x,y\in\mathbb{R}}{\text{minimize}} \quad & \sqrt{x^2+y^2} + xy \\
\text{subject to} \quad & \frac{1}{x} + \frac{1}{y} \leq 1, \\
& x, y > 0.
\end{aligned}$$

What is this nonsense? $\sqrt{x^2+y^2}$ doesn't even look like a posynomial.

We can make this work by adding a third variable z, and replace $\sqrt{x^2+y^2}$ by z. To relate this to the original problem, we ask that $z \geq \sqrt{x^2+y^2}$ (and since we're minimizing, z will actually be equal to this lower bound), which we can encode as a posynomial constraint:

$$\begin{aligned}
\underset{x,y,z\in\mathbb{R}}{\text{minimize}} \quad & z + xy \\
\text{subject to} \quad & \frac{1}{x} + \frac{1}{y} \leq 1, \\
& \frac{x^2}{z^2} + \frac{y^2}{z^2} \leq 1, \\
& x, y, z > 0.
\end{aligned}$$