Math 484: Nonlinear Programming[1] Mikhail Lavrov

# Chapter 5, Lecture 5: KKT Theorem, Saddle Point Form

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

## 1 Bounding the value function

Last time, we looked at what happens when we relax or tighten the constraints on a convex program: changing

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

to

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

The value function $MP(\mathbf{z})$ gives us (more or less) the optimal value of this perturbed variant of $P$, in terms of $\mathbf{z}$.

Intuitively (we'll make this precise later) knowing that $MP$ is a convex function tells us that it cannot decrease too quickly. So if we make $P$ an unconstrained program, allowing $\mathbf{x}$ to violate the constraint $g(\mathbf{x}) \le \mathbf{0}$, then a small violation in the constraint can only yield a small improvement. If we modify the objective function $f(\mathbf{x})$ to take that into account, we can forget about the constraints at all, and then everything is easy.

That's the plan. To make it work, we need two technical lemmas:

**Lemma 1.1** (Sensitivity vector lemma)**.** *Let $P$ be a convex program with a point $\mathbf{x}^* \in S$ such that $\mathbf{g}(\mathbf{x}^*) < \mathbf{0}$. (This is called the Slater condition.)*

*Then there is some $\boldsymbol{\lambda} \in \mathbb{R}^m$, with $\boldsymbol{\lambda} \ge \mathbf{0}$, such that*

$$MP(\mathbf{z}) \ge MP(\mathbf{0}) - \boldsymbol{\lambda} \cdot \mathbf{z}.$$

A vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ with $\boldsymbol{\lambda} \ge \mathbf{0}$ satisfying this inequality is called a *sensitivity vector*:[2] it measures the sensitivity of $P$ to changes in the constraints.

**Lemma 1.2** ("Moving the goalposts" lemma)**.** *Let $P$ be any nonlinear program. For any $\mathbf{x}^{(0)} \in S$,*

*(a)* $\mathbf{x}^{(0)}$ *is feasible for* $P(\mathbf{g}(\mathbf{x}^{(0)}))$*, and*

*(b)* $MP(\mathbf{g}(\mathbf{x}^{(0)})) \le f(\mathbf{x}^{(0)})$.

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

[2] Later, we'll call $\boldsymbol{\lambda}$ a "KKT multiplier" by analogy with Lagrange multipliers, or a "dual solution". Just giving you a heads up that this thing goes by multiple names.

The first lemma, when it applies, relates the original problem $P$ to the perturbed problems $P(\mathbf{z})$, which justifies thinking about them to begin with.

The second lemma says that any element of $S$, even if it's not feasible for $P$, still gives a bound on $MP(\mathbf{z})$ for some $\mathbf{z}$, and therefore (by the first lemma) it still gives a bound on $MP(\mathbf{0})$, the minimum value of $P$.

### 1.1 Proving the sensitivity vector lemma

*Proof.* The condition that there is some $\mathbf{x}^* \in S$ such that $\mathbf{g}(\mathbf{x}^*) < \mathbf{0}$ is called the Slater condition. It tells us that $\mathbf{0}$ is an interior point of the domain of $MP$: we can make the constraints of $P$ a little tighter, and $\mathbf{x}^*$ will stay feasible.

If $\mathbf{0}$ is an interior point, then (as we showed two lectures ago) $MP$ has a subgradient at $\mathbf{0}$, which is exactly the inequality

$$MP(\mathbf{z}) \geq MP(\mathbf{0}) - \boldsymbol{\lambda} \cdot \mathbf{z}.$$

Usually, we'd write that with a $+$; here, we write it with a $-$, because then we can show that $\boldsymbol{\lambda} \geq 0$.

To see this, take $\mathbf{z} = \mathbf{e}^{(i)}$. Then the subgradient inequality becomes

$$MP(\mathbf{e}^{(i)}) \geq MP(\mathbf{0}) - \lambda_i$$

but on the other hand we have $MP(\mathbf{0}) \geq MP(\mathbf{e}^{(i)})$: changing the $i^{\text{th}}$ constraint from $g_i(\mathbf{x}) \leq 0$ to $g_i(\mathbf{x}) \leq 1$ can't make the minimum value larger, only smaller. So $MP(\mathbf{0}) \geq MP(\mathbf{0}) - \lambda_i$, whcih means $\lambda_i \geq 0$. □

### 1.2 Proving the "Moving the goalposts" lemma

*Proof.* Part (a) is saying "$\mathbf{x}$ becomes feasible if we change the constraints to make $\mathbf{x}$ feasible". The program $P(\mathbf{g}(\mathbf{x}^{(0)}))$ is

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

and setting $\mathbf{x}$ to $\mathbf{x}^{(0)}$ makes the constraint be $\mathbf{g}(\mathbf{x}^{(0)}) \leq \mathbf{g}(\mathbf{x}^{(0)})$, which is true.

We defined $MP(\mathbf{g}(\mathbf{x}^{(0)}))$ to be the greatest lower bound on $f(\mathbf{x})$ for all $\mathbf{x}$ which are feasible for $P(\mathbf{g}(\mathbf{x}^{(0)}))$. One such $\mathbf{x}$ is $\mathbf{x}^{(0)}$ itself, so $MP(\mathbf{g}(\mathbf{x}^{(0)}))$ is a lower bound on $f(\mathbf{x}^{(0)})$, which is part (b) of the lemma. □

## 2 The Karush–Kuhn–Tucker theorem, saddle point form

For $\mathbf{x} \in S$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$, we define the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda})$ by

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda} \cdot \mathbf{g}(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}).$$

We will state the Karush–Kuhn–Tucker theorem slightly differently from the way it is stated in the textbook. The textbook states it for all convex programs satisfying the Slater condition: convex programs with a point satisfying all constraints strictly (with a $<$ in place of a $\leq$).

However, we only need this hypothesis to apply Lemma 1.1 and get a sensitivity vector: a vector $\boldsymbol{\lambda} \in \mathbb{R}^m$, with $\boldsymbol{\lambda} \geq \mathbf{0}$, such that

$$MP(\mathbf{z}) \geq MP(\mathbf{0}) - \boldsymbol{\lambda} \cdot \mathbf{z},$$

So we'll skip that step and just ask directly: do we have such a $\boldsymbol{\lambda}$?

**Theorem 2.1** (Karush–Kuhn–Tucker theorem, saddle point form)**.** *Let $P$ be any nonlinear program. Suppose that $\mathbf{x}^* \in S$ and $\boldsymbol{\lambda}^* \geq 0$. Then $\mathbf{x}^*$ is an optimal solution of $P$ and $\boldsymbol{\lambda}^*$ is a sensitivity vector for $P$ if and only if:*

1. *$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*)$ for all $\mathbf{x} \in S$. (Minimality of $\mathbf{x}^*$)*
2. *$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \geq L(\mathbf{x}^*, \boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \geq \mathbf{0}$. (Maximality of $\boldsymbol{\lambda}^*$)*
3. *$\lambda_i^* g_i(\mathbf{x}^*) = 0$ for $i = 1, 2, \ldots, m$. (Complementary slackness)*

## 2.1 Proving that these conditions are necessary

Let $\mathbf{x}^*$ be an optimal solution of $P$ and let $\boldsymbol{\lambda}^*$ be a sensitivity vector.

Take any $\mathbf{x} \in S$. By Lemma 1.2, $MP(\mathbf{g}(\mathbf{x})) \leq f(\mathbf{x})$. Using the sensitivity vector inequality, we get

$$f(\mathbf{x}) \geq MP(\mathbf{g}(\mathbf{x})) \geq MP(\mathbf{0}) - \boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}).$$

But $MP(\mathbf{0})$ is just $f(\mathbf{x}^*)$, so we can rewrite this as

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) + \boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\lambda}^*).$$

This is how the Lagrangian enters the picture. In particular, setting $\mathbf{x}^* = \mathbf{x}$, we get

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^*) + \boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}^*),$$

so $\boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}^*) \geq 0$.

But we have $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$ (since $\mathbf{x}^*$ is feasible for $P$) and $\boldsymbol{\lambda}^* \geq 0$ (by assumption). So the dot product $\boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}^*)$ is a sum of nonpositive products $g_i(\mathbf{x}^*) \leq \lambda_i^*$. In particular, it's always $\leq 0$, and the only way it can also be $\geq 0$ is if each of these products is 0.

This gives us the complementary slackness condition, and says that $\boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}^*) = 0$.

Going back to the inequality $f(\mathbf{x}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*)$: now that we know that $\boldsymbol{\lambda}^* \cdot g(\mathbf{x}^*) = 0$, we can add it to the left-hand side, and get $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*)$. This proves condition 1.

Finally, we have $\mathbf{g}(\mathbf{x}^*) \cdot \boldsymbol{\lambda} \leq 0$ for any $\boldsymbol{\lambda} \geq 0$, since (once again) this is a sum of nonpositive products. Since $\boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}^*) = 0$, we have

$$\mathbf{g}(\mathbf{x}^*) \cdot \boldsymbol{\lambda} \leq \mathbf{g}(\mathbf{x}^*) \cdot \boldsymbol{\lambda}^*$$

and adding $f(\mathbf{x}^*)$ to both sides gives $L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, proving condition 2.

### 2.2 Proving that these conditions are sufficient

First, we use the three conditions to prove that $\mathbf{x}^*$ is feasible: that $g_i(\mathbf{x}^*) \le 0$ for $i = 1, 2, \ldots, m$. To do this, take $\boldsymbol{\lambda} = \boldsymbol{\lambda}^* + \mathbf{e}^{(i)}$: we get

$$L(\mathbf{x}^*, \boldsymbol{\lambda}) = f(\mathbf{x}^*) + \mathbf{g}(\mathbf{x}^*) \cdot (\boldsymbol{\lambda}^* + \mathbf{e}^{(i)}) = f(\mathbf{x}^*) + \mathbf{g}\mathbf{x}^* \cdot \boldsymbol{\lambda}^* + g_i(\mathbf{x}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*) + g_i(\mathbf{x}^*).$$

But condition 2 tells us that $L(\mathbf{x}^*, \boldsymbol{\lambda}) \le L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, so we must have $g_i(\mathbf{x}^*) \le 0$.

Next, we prove that $\mathbf{x}^*$ is optimal. Take any feasible $\mathbf{x}$; that is, any $\mathbf{x} \in S$ with $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$; then

$$\begin{aligned} f(\mathbf{x}) &\ge f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\lambda}^*) && \text{(since each term in the sum is } \le 0) \\ &\ge L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) && \text{(by condition 1)} \\ &= f(\mathbf{x}^*) && \text{(by condition 3, since } \lambda_i^* g_i(\mathbf{x}^*) = 0 \text{ for all } i) \end{aligned}$$

This shows that $f(\mathbf{x}^*) \le f(\mathbf{x})$ for any feasible $\mathbf{x}$, making it optimal.

Finally, we prove that $\boldsymbol{\lambda}^*$ is a sensitivity vector. To do this, we must show that for all $\mathbf{z}$ in the domain of MP,

$$MP(\mathbf{z}) \ge MP(\mathbf{0}) - \boldsymbol{\lambda}^* \cdot \mathbf{z}.$$

Pick a vector $\mathbf{x}$ feasible for $P(\mathbf{z})$: that is, $\mathbf{x} \in S$, and $\mathbf{g}(\mathbf{x}) \le \mathbf{z}$. Then $\boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}) \le \boldsymbol{\lambda}^* \cdot \mathbf{z}$.

Condition 2 tells us that $f(\mathbf{x}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*)$, so

$$f(\mathbf{x}^*) \le f(\mathbf{x}) + \boldsymbol{\lambda}^* \cdot \mathbf{g}(\mathbf{x}) \le \boldsymbol{\lambda}^* \cdot \mathbf{z}$$

which we can rearrange to get

$$f(\mathbf{x}) \ge f(\mathbf{x}^*) - \boldsymbol{\lambda}^* \cdot \mathbf{z} = MP(\mathbf{0}) - \boldsymbol{\lambda}^* \cdot \mathbf{z}.$$

That is, $MP(\mathbf{0}) - \boldsymbol{\lambda}^* \cdot \mathbf{z}$ is a lower bound on $f(\mathbf{x})$ for any $\mathbf{x}$ feasible for $P(\mathbf{z})$. In the meantime, $MP(\mathbf{z})$ is by definition the greatest lower bound on $f(\mathbf{z})$ for any $\mathbf{x}$ feasible for $P(\mathbf{z})$. Therefore

$$MP(\mathbf{z}) \ge MP(\mathbf{0}) - \boldsymbol{\lambda}^* \cdot \mathbf{z}$$

which shows that $\boldsymbol{\lambda}^*$ is a sensitivity vector and completes the proof.