Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 7, Lecture 1: The KKT Theorem and Local Minimizers

April 29, 2019 *University of Illinois at Urbana-Champaign*

1 From the KKT conditions to local minimizers

We return to the KKT theorem, where we solve

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

by finding $\mathbf{x} \in S$ and $\boldsymbol{\lambda} \ge \mathbf{0}$ satisfying the gradient conditions

1. $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$. That is, $\nabla f(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}) = \mathbf{0}$.
2. For each i, $g_i(\mathbf{x}) \le 0$.
3. For each i, either $\lambda_i = 0$ or $g_i(\mathbf{x}) = 0$.

Suppose that, with no assumptions on convexity, we get a solution $(\mathbf{x}, \boldsymbol{\lambda})$. What can we say about it?

What we can say is something weak that *almost*, but not quite, implies "$\mathbf{x}$ is a local minimizer".

We know $\mathbf{x}$ is feasible for P. Call a direction $\mathbf{u}$ a *feasible direction from* $\mathbf{x}$ if $\mathbf{x} + t\mathbf{u}$ is feasible for P, for at least some sufficiently small $t > 0$.

If $\mathbf{u}$ is a feasible direction, then we'd better have $\nabla_{\mathbf{u}} g_i(\mathbf{x}) \le 0$, or $\nabla g_i(\mathbf{x}) \cdot \mathbf{u} \le 0$, whenever $g_i(\mathbf{x}) = 0$. (Whenever g_i is an *active constraint*.) Here, we're just saying that when $g_i(\mathbf{x})$ has already reached 0, it can't be increasing any farther in the direction $\mathbf{u}$.

Because of complementary slackness, we know that $\lambda_i \nabla g_i(\mathbf{x}) \cdot u \le 0$ no matter if $g_i(\mathbf{x}) = 0$ or not. For an active constraint, the inequality holds even before we multiply by λ_i. For an inactive constraint, $\lambda_i = 0$, so we just multiplied by 0.

In other words, for any feasible direction $\mathbf{u}$, we have

$$\left(\sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x})\right) \cdot \mathbf{u} = \sum_{i=1}^m (\lambda_i \nabla g_i(\mathbf{x}) \cdot \mathbf{u}) \le 0$$

and therefore, by the first of the KKT conditions, $\nabla f(\mathbf{x}) \cdot \mathbf{u} \ge 0$. Moreover, $\nabla f(\mathbf{x}) \cdot \mathbf{u} = 0$ only if $\nabla g_i(\mathbf{x}) \cdot \mathbf{u} = 0$ for all constraints with $\lambda_i > 0$: that is, if $\mathbf{u}$ is moving in a direction tangent to those constraints.

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

If this inequality were strict, we could kind of conclude that $\mathbf{x}$ is a local minimizer of P, because the function f would be increasing in any feasible direction.[2] As it is, we don't really know how f is changing in the directions where this is equal to 0: the directions tangent to the active constraints.

The point $\mathbf{x}$ is still *doing something special* in those directions, because in those cases we have $\nabla f(\mathbf{x}) \cdot \mathbf{u} = 0$, so it's not just an arbitrary point on the boundary that has this property. But $\mathbf{x}$ might not be doing anything more special than "acting like a critical point".

2 Note on feasible directions

Talking about "feasible directions" is not actually the best way to understand local minimizers of P. Suppose that our constraint is $x^2 + y^2 \leq 1$ and we're at the point $(1, 0)$. Then the direction $\mathbf{u} = (0, 1)$ is a direction tangent to the constraint, but it's not a "feasible direction", and so we can't understand it by looking along the line in the direction of $\mathbf{u}$.

To understand local minimizers properly, we have to relax our idea of going in a direction $\mathbf{u}$ along the path $h(t) = \mathbf{x} + t\mathbf{u}$, which doesn't work in this case. Instead, we allow ourselves to follow an arbitrary path $h(t)$ with $h(0) = \mathbf{x}$ and $\mathbf{h}'(0) = \mathbf{u}$. (For example, $h(t) = (\cos t, \sin t)$ is such a path in the example above.)

To explain why this the right idea, here is a result about such paths.

Lemma 2.1 (Lemma we're not going to prove)**.** *At a point $\mathbf{x}$ satisfying $\mathbf{g}(\mathbf{x}) \leq 0$, let $I \subseteq \{1, \dots, m\}$ be the set of all i such that $g_i(\mathbf{x}) = 0$.*

Suppose that the gradients $\{\nabla g_i(\mathbf{x}) : i \in I\}$ ***are linearly independent****.*

Then for every direction $\mathbf{u}$ such that $\nabla g_i(\mathbf{x}) \cdot \mathbf{u} = 0$ for all $i \in I$, there is a path $h(t)$ with

- $h(0) = \mathbf{x}$. *(We start at $\mathbf{x}$.)*
- $\mathbf{h}'(0) = \mathbf{u}$. *(We go in the direction $\mathbf{u}$, but maybe not in a straight line.)*
- $\mathbf{g}(h(t)) \leq 0$ *for all sufficiently small t. ($h(t)$ stays feasible as we go.)*

This is a consequence of the implicit function theorem: the result that lets us go from an implicit description $\{(x, y) : x^2 + y^2 = 1\}$ to an explicit description $\{(\cos t, \sin t) : 0 \leq t < 2\pi\}$ at least as long as we're sufficiently close to a point.

We call a feasible point $\mathbf{x}$ *regular* if it satisfies this hypothesis: the gradients $\{\nabla g_i(\mathbf{x}) : i \in I\}$ are linearly independent.

3 From local minimizers to the KKT conditions

What we really want is a result that says "if $\mathbf{x}$ is a local minimizer, then it satisfies the gradient KKT conditions". Then, by solving the equations, we'd find all local minimizers; assuming a global minimizer exists, we'd find that too.

[2] Even then, we'd be leaving out a few details: there might be points arbitrarily close to $\mathbf{x}$ that are not in any feasible direction from $\mathbf{x}$.

Unfortunately, this is not quite true without some additional assumptions. The assumptions we need to make vary, but here is a common one.

Theorem 3.1. *Suppose* $\mathbf{x}$ *is a local minimizer of* P *and a regular point. Then there is a* $\boldsymbol{\lambda} \geq \mathbf{0}$ *such that* $(\mathbf{x}, \boldsymbol{\lambda})$ *satisfy the gradient KKT conditions.*

Proof. As before, let $I = \{i : g_i(\mathbf{x}) = 0\}$.

We want to express $\nabla f(\mathbf{x})$ as a linear combination of the vectors $\{\nabla g_i(\mathbf{x}) : i \in I\}$: that's what conditions 1 and 3 of the gradient KKT theorem promise us. (Condition 1 says $\nabla f(\mathbf{x})$ is a linear combination of all the gradients; condition 3 says that the gradients $\nabla g_i(\mathbf{x})$ with $i \notin I$ have a coefficient of 0.)

Suppose this is impossible. Then by a fact from linear algebra, there is some direction $\mathbf{u}$ that distinguishes $\nabla f(\mathbf{x})$ from these gradients: we have

$$\nabla g_i(\mathbf{x}) \cdot \mathbf{u} = 0$$

for all $i \in I$, but $\nabla f(\mathbf{x}) \cdot \mathbf{u} \neq 0$. We can replace $\mathbf{u}$ by $-\mathbf{u}$, if necessary, to get $\nabla f(\mathbf{x}) \cdot \mathbf{u} < 0$.

By Lemma 2.1, there is a path $h(t)$ that starts at $\mathbf{x}$, goes in the direction $\mathbf{u}$, and stays in the feasible region of P. But now,

$$\frac{d}{dt} f(h(t))\bigg|_{t=0} = \nabla f(h(0)) \cdot \mathbf{h}'(0) = \nabla f(\mathbf{x}) \cdot \mathbf{u} < 0.$$

So by going along the path $h(t)$, the value of f decreases. This means that $\mathbf{x}$ is not a local minimizer of P after all!

So if $\mathbf{x}$ is a local minimizer, then there is some $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}) = \mathbf{0}$$

and $\lambda_i = 0$ whenever $g_i(\mathbf{x}) < 0$. This proves conditions 1 and 3 of the gradient KKT theorem, and condition 2 was one of our assumptions to begin with.

A final thing we should check is that $\boldsymbol{\lambda} \geq \mathbf{0}$. (This is something that identifies $\mathbf{x}$ as a local minimizer specifically, rather than a local maximizer, for instance.)

To do this, pick any $j \in I$, and let $\mathbf{u}$ be a direction such that $\nabla g_j(\mathbf{x}) \cdot \mathbf{u} = -1$ while $\nabla g_i(\mathbf{x}) \cdot \mathbf{u} = 0$ for all other $i \in I$.

By Lemma 2.1, applied to the set of constraints ignoring g_j, there is a path $h(t)$ from $\mathbf{x}$ in the direction of $\mathbf{u}$ that stays feasible. Well, maybe it violates g_j, because we left that out. But no: the condition $\nabla g_j(\mathbf{x}) \cdot \mathbf{u} = -1$ says that g_j is increasing as we head in the direction of $\mathbf{u}$, so that it is also still satisfied, at least for small t.

We have

$$0 = \nabla f(\mathbf{x}) \cdot \mathbf{u} + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}) \cdot \mathbf{u} = \nabla f(\mathbf{x}) \cdot \mathbf{u} + \lambda_j \cdot -1 + \sum_{i \neq j} \lambda_i \cdot 0$$

and therefore, after everything cancels, $\nabla f(\mathbf{x}) \cdot \mathbf{u} = \lambda_j$.

Well, we'd better have $\lambda_j \geq 0$, because f shouldn't be decreasing in the direction $\mathbf{u}$ (otherwise, once again, tha path h would prove that $\mathbf{x}$ isn't a local minimizer). Because j could have been any element of I, and because we have $\lambda_i = 0$ for $i \notin I$, we conclude $\boldsymbol{\lambda} \geq 0$ and we are done. □

The condition that $\mathbf{x}$ should be a regular point is—well, it's not the strongest condition we could ask for, but not satisfying it can definitely cause problems. Generally speaking, the same kinds of nonlinear programs that violate Slater's condition (when they're convex) are also the ones where we get non-regular points.

Consider, for example, the constraints $x^2 + y^2 \leq 1$ and $x \geq 1$.

Together, these constraints only have the feasible point $(1, 0)$. The gradients of $g_1(x, y) = x^2 + y^2 - 1$ and $g_2(x, y) = 1 - x$ are $(2, 0)$ and $(-1, 0)$ at this point. So they're not linearly independent.

The vector $\mathbf{u} = (0, 1)$ is orthogonal to both of these gradients. However, there's no path from $(1, 0)$ going in the direction of $\mathbf{u}$ that stays feasible: there's no path from $(1, 0)$ that goes in any direction that stays feasible! So our proof falls flat, and in fact the point $(1, 0)$ is not found by the gradient KKT conditions here.

It's not just isolated points like this one that cause us trouble. If we consider the same set of constraints in $\mathbb{R}^3$, then the entire line $(1, 0, z)$ is feasible. However, that line still does not allow us to take any path going in the direction $\mathbf{u} = (0, 1, 0)$, so the proof doesn't work.

The intuition is that in this case, the KKT conditions require the point $(1, 0, 0)$ to be a local minimizer in the direction $(0, 1, 0)$ before they can find it, because that direction is orthogonal to all the active constraints. But *actually being a local minimizer* does not require that, because going in this direction is impossible while remaining feasible. So this is a case where the KKT conditions are stronger than necessary, and there are some points they don't find.