Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 1, Lecture 3: Critical points in $\mathbb{R}^n$

January 23, 2019 *University of Illinois at Urbana-Champaign*

1 Restrictions to a line

Thinking about an n-dimensional function $f : \mathbb{R}^n \to \mathbb{R}$ is hard. We understand one-variable functions; functions of several variables are harder to work with.

There is a useful trick for fixing this which we'll use throughout this class in many ways. Given a point $\mathbf{x} \in \mathbb{R}^n$, and a direction $\mathbf{u} \in \mathbb{R}^n$ (with $\mathbf{u} \neq \mathbf{0}$), the line through $\mathbf{x}$ in the direction of $\mathbf{u}$ is parametrized as $\{\mathbf{x} + t\mathbf{u} : t \in \mathbb{R}\}$. As you plug in different values of t, you get different points along this line. We define the *restriction of f to the line through $\mathbf{x}$ in the direction of $\mathbf{u}$* to be the function

$$\phi_{\mathbf{u}}(t) = f(\mathbf{x} + t\mathbf{u}).$$

This is a one-variable function which describes what f does along a single line through $\mathbf{x}$.

Today, we'll use such restrictions to understand minimizers of f. The global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ is what you'd expect based on our one-dimensional definition: a point $\mathbf{x}^* \in \mathbb{R}^n$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Lemma 1.1. *A point $\mathbf{x}^* \in \mathbb{R}^n$ is a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ if and only if, for every direction $\mathbf{u} \in \mathbb{R}^n$, $t = 0$ is the global minimizer of $\phi_{\mathbf{u}}(t) = f(\mathbf{x}^* + t\mathbf{u})$.*

Proof. Suppose $\mathbf{x}^*$ is a global minimizer of f. Then, for every $\mathbf{u}$ and for every t, $\phi_{\mathbf{u}}(0) = f(\mathbf{x}^*) \leq f(\mathbf{x}^* + t\mathbf{u}) \leq \phi_{\mathbf{u}}(t)$, making 0 a global minimizer of $\phi_{\mathbf{u}}$.

On the other hand, suppose that $t = 0$ is a global minimizer of $\phi_{\mathbf{u}}$ for any $\mathbf{u}$. If you pick any $\mathbf{x} \in \mathbb{R}^n$, you can set $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$. Then we have $\phi_{\mathbf{u}}(0) = f(\mathbf{x}^*)$ and $\phi_{\mathbf{u}}(1) = f(\mathbf{x})$. Since $\phi_{\mathbf{u}}(0) \leq \phi_{\mathbf{u}}(1)$, we must have $f(\mathbf{x}^*) \leq f(\mathbf{x})$, and because this works for any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^*$ is a global minimizer of f. □

The plan is to understand global minimizers in $\mathbb{R}^n$ by using restrictions to a line to turn them into global minimizers in $\mathbb{R}$, which we already know about.

2 The first-derivative test in $\mathbb{R}^n$

If $\phi_{\mathbf{u}}(t) = f(\mathbf{x} + t\mathbf{u})$, what is the derivative of $\phi_{\mathbf{u}}(t)$?

Here, we can apply the chain rule for multi-variable functions. For any $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$, we have

$$\frac{d}{dt} f(\mathbf{g}(t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{g}(t)) \frac{d}{dt} g_i(t).$$

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

When we look at $\phi_{\mathbf{u}}(t) = f(\mathbf{x} + t\mathbf{u})$, the inside function $\mathbf{g}(t)$ is just the linear function $\mathbf{x} + t\mathbf{u}$. Its i^{th} component is $x_i + tu_i$, whose derivative is just u_i. So we have

$$\phi'_{\mathbf{u}}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x} + t\mathbf{u})u_i.$$

To make sense of this, define $\nabla f(\mathbf{x})$ to be the vector of all of f's partial derivatives: $\nabla f(\mathbf{x})^{\mathsf{T}} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix}$. Then the sum becomes a dot product:

$$\phi'_{\mathbf{u}}(t) = \nabla f(\mathbf{x} + t\mathbf{u}) \cdot \mathbf{u}.$$

Fine print: the multivariate chain rule only works under some assumptions. It's enough if we assume that the partial derivatives $\frac{\partial f}{\partial x_i}$ are all continuous (at $\mathbf{x} + t\mathbf{u}$, in this case). We'll sum this up as "∇f is continuous".

Theorem 2.1. *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, if ∇f is continuous and $\mathbf{x}^*$ is a global minimizer of f, then $\nabla f(\mathbf{x}^*) = \mathbf{0}$. (When $\nabla f(\mathbf{x}^*) = \mathbf{0}$, we call $\mathbf{x}^*$ a critical point of f.)*

Proof. If $\mathbf{x}^*$ is a global minimizer of f, then $t = 0$ is a global minimizer of $\phi_{\mathbf{u}}(t) = f(\mathbf{x}^* + t\mathbf{u})$ for each direction $\mathbf{u}$. We already know that for the one-variable function $\phi_{\mathbf{u}}(t)$, this means that $\phi'_{\mathbf{u}}(0) = 0$.

Our application of the chain rule has shown that $\phi'_{\mathbf{u}}(0) = \nabla f(\mathbf{x}^*) \cdot \mathbf{u}$. (This is known as a *directional derivative* of f.) So we must have $\nabla f(\mathbf{x}^*) \cdot \mathbf{u} = 0$ for any $\mathbf{u}$.

In particular, $\nabla f(\mathbf{x}^*) \cdot \nabla f(\mathbf{x}^*) = \|\nabla f(\mathbf{x}^*)\|^2 = 0$, which means $\nabla f(\mathbf{x}^*) = \mathbf{0}$. □

3 The second-derivative test in $\mathbb{R}^n$

Now let's do the same thing for second derivatives, and see what happens.

For the second derivative of $\phi_{\mathbf{u}}(t)$, we have to apply the chain rule twice. Starting from the expression

$$\phi'_{\mathbf{u}}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x} + t\mathbf{u})u_i$$

we take another derivative and get

$$\phi''_{\mathbf{u}}(t) = \sum_{i=1}^{n} u_i \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x} + t\mathbf{u})u_j \right).$$

This is not a dot product but a matrix product. The second derivatives of f can be put in a matrix, which we call the Hessian matrix of f and write Hf. That is,

$$Hf(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{bmatrix}.$$

With this notation, we have $\phi''_{\mathbf{u}}(t) = \mathbf{u}^{\mathsf{T}} Hf(\mathbf{x} + t\mathbf{u})\mathbf{u}$.

For this to apply, we assume that all second partial derivatives of f are continuous, which we sum up as "Hf is continuous".

Theorem 3.1. *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, if Hf is continuous, $\nabla f(\mathbf{x}^*) = \mathbf{0}$, and $Hf(\mathbf{x})$, for any $\mathbf{x} \in \mathbb{R}^n$, has the property that*

$$\mathbf{u}^{\mathsf{T}} Hf(\mathbf{x})\mathbf{u} \geq 0 \qquad \text{for all } \mathbf{u} \in \mathbb{R}^n,$$

then $\mathbf{x}^*$ is a global minimizer of f.

Proof. Consider any direction $\mathbf{u}$ and let $\phi_{\mathbf{u}}(t) = f(\mathbf{x}^* + t\mathbf{u})$.

Since $\nabla f(\mathbf{x}^*) = \mathbf{0}$, we have $\phi'_{\mathbf{u}}(0) = 0$. Our very generous hypothesis about f tells us that $\phi''_{\mathbf{u}}(t) = \mathbf{u}^{\mathsf{T}} Hf(\mathbf{x}^* + t\mathbf{u})\mathbf{u} \geq 0$ for all t.

This makes 0 a global minimizer of $\phi_{\mathbf{u}}$, for any $\mathbf{u}$. Therefore $\mathbf{x}^*$ is a global minimizer of f. □

4 Minimizing over other sets

Suppose f is not defined on all of $\mathbb{R}^n$: we have a subset $D \subseteq \mathbb{R}^n$, and a function $f : D \to \mathbb{R}$. (This means that restrictions of f to a line will also have a different domain: $\phi_{\mathbf{u}}(t) = f(\mathbf{x} + t\mathbf{u})$ is only defined on the set $\{t : \mathbf{x} + t\mathbf{u} \in D\}$.)

Does everything we've done still work?

The answer is yes, with two assumptions. First, we want to consider points $\mathbf{x}^*$ that aren't on the boundary of D. If $\mathbf{x}^*$ is on the boundary, then there are directions $\mathbf{u}$ in which we don't expect $\phi_{\mathbf{u}}(t) = f(\mathbf{x}^* + t\mathbf{u})$ to increase as t goes up, because the point $\mathbf{x}^* + t\mathbf{u}$ immediately leaves D. So that's no good.

Second, we want $\mathbf{x}^*$ to be able to "see" every point $\mathbf{x} \in D$ along a straight line, so that we can compare them via some restriction $\phi_{\mathbf{u}}(t)$. We'll have more to say about which sets D this happens for when we get to chapter 2. For now, just remember that this could be a problem.

One kind of set D for which both conditions are met is the "ball of radius r around $\mathbf{x}^*$": the set

$$\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}^*\| < r\}$$

of points that are distance less than r away from $\mathbf{x}^*$. We denote it by $B(\mathbf{x}^*, r)$.

We say that $\mathbf{x}^*$ is a *local minimizer* of $f : \mathbb{R}^n \to \mathbb{R}$ if there is some radius $r > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ whenever $\|\mathbf{x}^* - \mathbf{x}\| < r$. In other words, $\mathbf{x}^*$ is a global minimizer of f but on the domain $B(\mathbf{x}^*, r)$ only.

By applying our results after replacing the domain of f with $B(\mathbf{x}^*, r)$, we immediately get the following theorems:

Theorem 4.1. *Suppose $f : D \to \mathbb{R}$ has continuous ∇f and $\mathbf{x}^*$ is not on the boundary of D.*

If $\mathbf{x}^*$ is a local minimizer of f, then $\mathbf{x}^*$ is a critical point of f: $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Theorem 4.2. *Suppose $f : D \to \mathbb{R}$ has continuous Hf and $\mathbf{x}^*$ is a critical point of f.*

If we can pick an $r > 0$ such that whenever $\|\mathbf{x} - \mathbf{x}^*\| < r$, $Hf(\mathbf{x})$ has the property

$$\mathbf{u}^\mathsf{T} Hf(\mathbf{x})\mathbf{u} \geq 0 \qquad \text{for all } \mathbf{u} \in \mathbb{R}^n,$$

then $\mathbf{x}^*$ is a local minimizer of f.

Looking back at this theorem we've shown just now, one thing is still mysterious. What exactly is this property of $Hf(\mathbf{x})$ that we're asking for? How do we test for it? This will be our next topic.

Later, we will refine this theorem. For functions $f : \mathbb{R} \to \mathbb{R}$, the second derivative test could only report the results "x^* is a local minimizer", "x^* is a local maximizer", and "the test is inconclusive". But in higher dimensions, we will often be able to identify that a point $\mathbf{x}^*$ is definitely *not* a local minimizer or maximizer. (Sometimes, however, the test will still be inconclusive.)