Math 484: Nonlinear Programming[1] — Mikhail Lavrov

Chapter 1, Lecture 1: 1-Dimensional Optimization

January 14, 2019 — *University of Illinois at Urbana-Champaign*

1 Useful information about the class

See the course webpage (in the footnote at the bottom of this page) for information about when things happen, how grading works, and so forth.

- The first homework assignment is due Friday, January 25th in class. The first exam will be in the evening on Wednesday, February 6th. There's some sort of deadline for registering for the 4-credit version of this class, if you want to do that.
- Homework grades and solutions will be posted on Moodle.
- Notes for each lecture will be posted on the course webpage. (But consider taking your own notes, as well.)

2 About optimization

"Nonlinear programming" suggests telling a computer to do nonlinear stuff. But that's not what the word "programming" means in the title of this class. In this context, "programming" means "optimization": finding a point in a set that maximizes the value of a function.

In general, we might be looking at problems of the following form:

$$\begin{aligned} \underset{x_1,x_2,\dots,x_n\in\mathbb{R}}{\text{minimize}} \quad & f(x_1,x_2,\dots,x_n) \\ \text{subject to} \quad & g_1(x_1,x_2,\dots,x_n) \le 0, \\ & \dots, \\ & g_m(x_1,x_2,\dots,x_n) \le 0. \end{aligned}$$

Without any assumptions on the functions f and $g_1, \dots, g_m$, this is basically impossible to do.

So instead, we:

- look at special cases which we do know how to solve,
- reduce complicated cases to simpler ones, and
- classify problems into cases which can or cannot be handled.

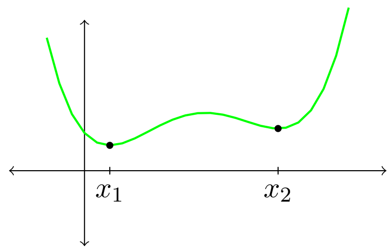
The beginning of this class (the material from Chapter 1) will consider the case of unconstrained optimization: minimizing or maximizing a function $f : \mathbb{R}^n \to \mathbb{R}$ with no further conditions on where the optimal solution can be.

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

Today, we will consider an even more specific case: optimizing one-dimensional functions $f : \mathbb{R} \to \mathbb{R}$. This should mostly be a review of some facts from single-variable calculus. Our goal is to learn some terminology, and to prove these facts in the one-dimensional setting so that in the next lecture, we can try to extend their proofs to the general case.

3 Single-variable calculus

Consider the function $f : \mathbb{R} \to \mathbb{R}$ graphed below:

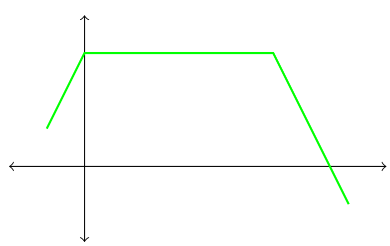


A point x^* is a *global minimizer* if $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}$. For this function, x_1 is a global minimizer.

A point x^* is a *local minimizer* if there's some positive radius $r > 0$ such that $f(x^*) \leq f(x)$, not necessarily for all $x \in \mathbb{R}$, but at least for all x such that $|x - x^*| < r$: all points close to x^*. A local minimizer might not be the best possible solution, but it's the best possible among nearby points. In this example, both x_1 and x_2 are local minimizers.

In fact, both of the examples so far are *strict* minimizers of their type. A point x^* is a *strict global minimizer* if $f(x^*) < f(x)$ for all $x \in \mathbb{R}$, and a *strict local minimizer* if there's some positive radius $r > 0$ such that $f(x^*) < f(x)$ for all x with $|x - x^*| < r$.

For the function $f(x) = \sin x$, the values $x = \dots, -\frac{5\pi}{2}, -\frac{\pi}{2}, \frac{3\pi}{2}, \frac{7\pi}{2}, \dots$ are global minimizers that are not strict. All of these x-values achieve the same minimum value of $f(x)$. Or consider the function below:



All the points along the horizontal line segment, except for its endpoints, are local minimizers: close to those points, you can't do any better. They're not strict local minimizers, because there are lots of points nearby that are tied for best.

We can also make the same definitions for maximizers (local and global, strict and non-strict). Usually, the theory for maximizing functions turns out to be the same as for minimizing them, and for convenience, we'll just assume that we're always minimizing.

Also, I make a notational promise to you: whenever I write x^* (or something similar), the $*$ is supposed to indicate that we're considering an optimal value of some sort: usually, a local or global minimizer. Or at least we're hoping that it is one of these.

Now we'll begin with the promised review of basic facts from calculus.

3.1 The first derivative test

Theorem 3.1 (1.1.4 in the textbook)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function and x^* is a local minimizer of f, then $f'(x^*) = 0$.*

Proof. Knowing that x^* is a local minimizer tells us that we can choose a radius $r > 0$ such that $f(x^*) \le f(x)$ for all x in the range $(x^* - r, x^* + r)$.

We want to show $f'(x^*) = 0$, so let's recall the definition of a derivative:

$$f'(x^*) = \lim_{h \to 0} \frac{f(x^* + h) - f(x^*)}{h}.$$

Since $h \to 0$ in the limit, it's okay to only consider h with $0 < |h| < r$. There are two cases:

- When $0 < h < r$, we have $f(x^* + h) \ge f(x^*)$ by the definition of a local minimizer, so the fraction $\frac{f(x^*+h)-f(x^*)}{h}$ divides a nonnegative number by a positive number. Therefore $\frac{f(x^*+h)-f(x^*)}{h} \ge 0$. As a result, the limit $f'(x^*)$ should also be ≥ 0.
- When $-r < h < 0$, we also have $f(x^* + h) \ge f(x^*)$ by the definition of a local minimizer, so the fraction $\frac{f(x^*+h)-f(x^*)}{h}$ divides a nonnegative number by a *negative* number. Therefore $\frac{f(x^*+h)-f(x^*)}{h} \le 0$. As a result, the limit $f'(x^*)$ should also be ≤ 0.

Since $f'(x^*) \le 0$ and $f'(x^*) \ge 0$, we must have $f'(x^*) = 0$, completing the proof. □

This motivates the definition of a *critical point*: a point x^* is a critical point of f if $f'(x^*)$ is 0.

To convince yourself that you fully understand this proof, here are two variants to think about:

1. (Easy) It's also true that if x^* is a local maximizer of a differentiable function f, then $f'(x^*)$ must be 0. How does the proof change for local maximizers?
2. (Harder) In the textbook, the theorem actually says something slightly different: assuming that f is only defined on an interval $[a, b]$, it shows that a local minimizer is either a critical point, or else one of a or b. Why do endpoints matter for the proof? (And what should the definition of a local minimizer even be in this case, so that it can apply to a or b?)

3.2 The second derivative test

Once again, the textbook states this in a slightly stronger form, and you should compare the statement in the textbook to this one.

Theorem 3.2 (1.1.5 in the textbook)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is a function with a continuous second derivative and x^* is a critical point of f, then:*

(a) If we know that $f''(x) \ge 0$ for all $x \in \mathbb{R}$, then x^* is a global minimizer on $\mathbb{R}$.

(a') If we only know there is an interval $[a, b]$ containing x^* such that $f''(x) \geq 0$ for all $x \in [a, b]$, then x^* is a global minimizer on $[a, b]$.

(b) If, instead, we only know that $f''(x^*) > 0$, then x^* is a local minimizer.

Proof. The key fact we need to prove this theorem is Taylor's formula. Given the values of f and its derivatives at x^*, we can approximate f's value at another point x as follows:

- Linearly: $f(x) = f(x^*) + f'(\xi)(x - x^*)$ for some ξ between x and x^*.
- Quadratically: $f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(\xi)\frac{(x-x^*)^2}{2!}$ for some ξ between x and x^*.
- Cubically: $f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(x^*)\frac{(x-x^*)^2}{2!} + f'''(\xi)\frac{(x-x^*)^3}{3!}$ for some ξ between x and x^*.
- And so forth, for as long as the derivatives exist.

Here, we will only use the second of the approximations:

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + f''(\xi)\frac{(x - x^*)^2}{2!}$$

for some ξ between x and x^*.

Since x^* is a critical point, $f'(x^*) = 0$, and the second term $f'(x^*)(x - x^*)$ disappears. If, as in part (a) of the theorem, we assume that f'' is nonnegative everywhere, then the third term is also nonnegative (since $(x - x^*)^2$ can never be negative). In other words, for any $x \in \mathbb{R}$,

$$f(x) = f(x^*) + (\text{zero}) + (\text{something nonnegative}),$$

which means that $f(x^*) \leq f(x)$: x^* is a global minimizer.

Moreover, if x and x^* are both in some interval $[a, b]$, then ξ will also be in that interval. So if we know that f'' is nonnegative on $[a, b]$, then the same argument shows that $f(x^*) \leq f(x)$ for all $x \in [a, b]$.

Finally, if all we know is that $f''(x^*) > 0$, then—by continuity of f''—there is an interval $[x^* - r, x^* + r]$ around x^* where f'' is still nonnegative. So x^* is a global minimizer of f on that interval. However, since we have no control over r, all this tells us is that x^* is a local minimizer. □