Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 5, Lecture 4: Convex (and Nonlinear) Programming

March 13, 2019 *University of Illinois at Urbana-Champaign*

1 A general nonlinear program

Our optimization problems or "nonlinear programs" are going to have the form

$$(P) \qquad \begin{cases} \underset{\mathbf{x}\in S}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & g_1(\mathbf{x}) \le 0 \\ & \dots \\ & g_m(\mathbf{x}) \le 0 \end{cases}$$

where S is a subset of $\mathbb{R}^n$ and $f, g_1, g_2, \dots, g_m$ are functions $S \to \mathbb{R}$. As shorthand, we write

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

thinking of $\mathbf{g}$ as a function $S \to \mathbb{R}^m$, and the $\le \mathbf{0}$ comparison checking if every component of $\mathbf{g}(\mathbf{x})$ is nonpositive.

This setup means that there's two places where constraints on $\mathbf{x}$ come from: $\mathbf{x}$ must satisfy $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$, and $\mathbf{x}$ must be an element of S. The second type of constraint is a bit of a cheat: we can't say much about arbitrary sets S, so the constraint "$\mathbf{x} \in S$" will remain mysterious to us forever. We will be able to say more if as many of the constraints on $\mathbf{x}$ as possible get written as $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$.

(Sometimes, we are forced to "hide" a constraint in S, because the functions f and $g_1, g_2, \dots, g_m$ are not all defined on all of $\mathbb{R}^m$, or don't have the properties we want on all of $\mathbb{R}^m$. In that case, we let S be the correct domain of those functions.)

We call

$$F = \{\mathbf{x} \in S : \mathbf{g}(\mathbf{x}) \le \mathbf{0}\}$$

the *feasible region* of P. A point in the feasible region is called the feasible solution.

We call

$$MP = \inf\{f(\mathbf{x}) : \mathbf{x} \in F\}$$

the *value* of P. Here, inf denotes the infimum or greatest lower bound. This is, more or less, the value of $f(\mathbf{x})$ at the optimal solution to P, with two caveats:

- If P is unbounded, and $f(\mathbf{x})$ can be made arbitrarily negative, then $MP = -\infty$. If P is inconsistent, and F is empty, then $MP = +\infty$.

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

- If P has solutions getting arbitrarily close to a value z, but are always bigger than z, then $MP = z$ as well. (We saw this happen with geometric programs.)

So far we have been talking about the fully general case of nonlinear programming. The nonlinear program P is a convex program if:

- the set S is a convex set, and
- the functions $f, g_1, \dots, g_m$ are convex functions.

In the case of a convex program, the feasible region F will also be convex. It is a good exercise, which I may or may not assign for homework, to check this by using the definition of convex sets.

For now, we will assume that we're dealing with a convex program. We'll see later which of our results require convexity and which ones don't.

2 Perturbations of the constraints

To better understand P, we generalize it to the perturbation $P(\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^m$:

$$(P(\mathbf{z})) \qquad \begin{cases} \underset{\mathbf{x} \in S}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & g_1(\mathbf{x}) \le z_1 \\ & \dots \\ & g_m(\mathbf{x}) \le z_m. \end{cases}$$

One reason to look at $P(\mathbf{z})$ is sensitivity analysis: relaxing the constraints to see what happens tells us how which constraints matter and which ones don't. Another reason is that we want to convert P to an unconstrained optimization problem somehow, and $P(\mathbf{z})$ helps with that because every point $\mathbf{x} \in S$ is feasible for $P(\mathbf{z})$ for *some* $\mathbf{z}$.

For $P(\mathbf{z})$, we define $F(\mathbf{z}) = \{\mathbf{x} : \mathbf{x} \in S, \mathbf{g}(\mathbf{x}) \le \mathbf{z}\}$, and

$$MP(\mathbf{z}) = \inf\{f(\mathbf{x}) : \mathbf{x} \in F(\mathbf{z})\}.$$

This $MP(\mathbf{z})$ is called the *value function*: you tell it how you're changing the constraints, and it tells you how that affects the value.

Sometimes $MP(\mathbf{z})$ is $\pm\infty$. Usually, we restrict the domain of MP to those $\mathbf{z}$ for which $F(\mathbf{z}) \neq \varnothing$, and so $MP(\mathbf{z}) < \infty$. This still allows us to end up with $MP(\mathbf{z}) = -\infty$; we are sad about that, but we're not going to do anything about it for now.

Theorem 2.1. *Let P be a convex program. Then the domain of MP (the set of $\mathbf{z}$ for which $MP(\mathbf{z})$ is feasible) is a convex set, and $MP(\mathbf{z})$ is a convex function on that domain.*

Proof. Suppose that $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}$ are in the domain of MP.

Because we take the inf, there is not necessarily an $\mathbf{x}^{(i)} \in F(\mathbf{z}^{(i)})$ for which $f(\mathbf{x}^{(i)}) = MP(\mathbf{z}^{(i)})$. (Maybe we can only get arbitrarily close.) However, for any $\epsilon > 0$, we can find $\mathbf{x}^{(1)} \in F(\mathbf{z}^{(1)})$ and

$\mathbf{x}^{(2)} \in F(\mathbf{z}^{(2)})$ such that

$$f(\mathbf{x}^{(1)}) \leq MP(\mathbf{z}^{(1)}) + \epsilon \quad \text{and} \quad f(\mathbf{x}^{(2)}) \leq MP(\mathbf{z}^{(2)}) + \epsilon.$$

Then for $0 \leq t \leq 1$:

1. We first show that $t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}$ is in the domain of MP, proving that it's a convex set. To do this, we'll show that $t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}$ is an element of $F(t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)})$, so in particular $F(t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}) \neq \varnothing$.

   Since S is a convex set and $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in S$, we have $t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)} \in S$. For every g_i, $1 \leq i \leq m$, we have

$$\begin{aligned} g_i(t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}) &\leq tg_i(\mathbf{x}^{(1)}) + (1-t)g_i(\mathbf{x}^{(2)}) && \text{(since } g_i \text{ is convex)} \\ &\leq t{z_i}^{(1)} + (1-t){z_i}^{(2)} && \text{(since } g_1(\mathbf{x}^{(1)}) \leq {z_i}^{(1)} \text{ and } g_2(\mathbf{x}^{(2)}) \leq {z_i}^{(2)}) \\ &\leq [t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}]_i \end{aligned}$$

   and putting these together we get $\mathbf{g}(t\mathbf{x}^{(1)}+(1-t)\mathbf{x}^{(2)}) \leq t\mathbf{z}^{(1)}+(1-t)\mathbf{z}^{(2)}$, so $t\mathbf{x}^{(1)}+(1-t)\mathbf{x}^{(2)}$ is feasible for $t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}$.

2. Second, we show that MP is a convex function on its domain. By convexity of f, we have

$$f(t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}) \leq tf(\mathbf{x}^{(1)}) + (1-t)f(\mathbf{x}^{(2)}).$$

   Due to the way we chose $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, this tells us that

$$f(t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}) \leq tMP(\mathbf{z}^{(1)}) + (1-t)MP(\mathbf{z}^{(2)}) + \epsilon.$$

   Since $MP(\mathbf{z})$ is a lower bound on $f(\mathbf{x})$ for all $\mathbf{x} \in F(\mathbf{z})$, we also have

$$MP(t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}) \leq tMP(\mathbf{z}^{(1)}) + (1-t)MP(\mathbf{z}^{(2)}) + \epsilon.$$

   If this is true for all $\epsilon > 0$, then the difference $MP(t\mathbf{z}^{(1)} + (1-t)\mathbf{z}^{(2)}) - [tMP(\mathbf{z}^{(1)}) + (1-t)MP(\mathbf{z}^{(2)})]$ is less than any positive number, so it must be 0, and the inequality holds for $\epsilon = 0$ as well.

   Therefore we have checked the definition of convex functions for MP. □

Here is a simple example of the value function in action. Consider the convex program

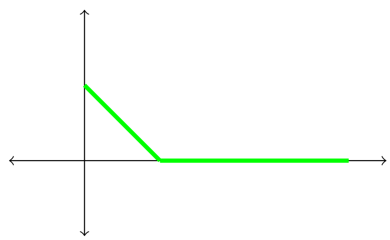
$$(P) \qquad \begin{cases} \underset{(x,y)\in\mathbb{R}^2}{\text{minimize}} & |x-1| \\ \text{subject to} & \sqrt{x^2+y^2} \leq 0. \end{cases}$$

Then $P(z)$ tells us to minimize $|x-1|$ subject to $\sqrt{x^2+y^2} \leq z$. Geometrically, we're looking for a point in the disk of radius z around $\mathbf{0}$ which is as close to the line $x = 1$ as possible.

There are three cases:

- For $z \geq 1$, the point $(1, 0)$ is feasible for $P(z)$, since $1^2 + 0^2 \leq z$. So we can make $|x - 1| = 0$; we can't do better than that, so $MP(z) = 0$.
- For $0 \leq z < 1$, the best we can do is set $x = z$ and $y = 0$. Then $MP(z) = |z - 1| = 1 - z$.
- For $z < 0$, there are no (x, y) such that $\sqrt{x^2 + y^2} \leq z$. So $MP(z) = +\infty$: we are outside the domain of MP.

Plotting $MP(z)$, we get:



In particular, you should notice that even though $MP(z)$ is convex, it is not a "nice" convex function: it has sharp corners when the behavior of $P(z)$ changes. This is not just because $|x - 1|$ has a corner: with more constraints, it can happen even when the functions we start with are perfectly nice. (Even with linear functions.) For example, see Example 5.2.9 in the textbook.

You should also notice that $MP(z)$ is a decreasing function: this is a universal feature. Increasing z relaxes the constraints, so the value of $P(z)$ can only get smaller. If P has multiple constraints, then $MP(\mathbf{z})$ is a decreasing function in each of its variables: increasing z_i relaxes the i^{th} constraint, so the value will either get smaller or (in case the i^{th} constraint didn't matter) stay the same.