Math 484: Nonlinear Programming[1] Mikhail Lavrov

Chapter 2, Lecture 4: Jensen's inequality

February 11, 2019 *University of Illinois at Urbana-Champaign*

1 Jensen's inequality

Jensen's inequality—one of the most useful inequalities that ever inequalitied—is the result below:

Theorem 1.1. *For any* $\lambda_1, \lambda_2, \dots, \lambda_k \ge 0$ *with* $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$*, if* $f : C \to \mathbb{R}$ *is convex and* $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)} \in C$*, then*

$$f(\lambda_1 \mathbf{x}^{(1)} + \lambda_2 \mathbf{x}^{(2)} + \cdots + \lambda_k \mathbf{x}^{(k)}) \le \lambda_1 f(\mathbf{x}^{(1)}) + \lambda_2 f(\mathbf{x}^{(2)}) + \cdots + \lambda_k f(\mathbf{x}^{(k)}).$$

This might seem very similar to the property of convex sets we proved in the previous lecture: that a convex combination of points in a convex set C is still an element of C. It is! It is so similar, in fact, that we can take a shortcut and get this theorem as a corollary of the theorem from the last lecture. (For the non-shortcut proof, which is essentially a rehash of the proof of the previous theorem, see your textbook.)

We'll need a definition first. Given a subset $C \subseteq \mathbb{R}^n$ and a function $f : C \to \mathbb{R}$, its *epigraph* is the set

$$\text{epi}(f) = \{(\mathbf{x}, y) \in C \times \mathbb{R} : y \ge f(\mathbf{x})\}.$$

The prefix "epi" means "above", so "epigraph" means "above the graph", and this is just what the epigraph is: it's the subset of $\mathbb{R}^{n+1}$ (one dimension higher, because we're graphing) above the graph of f.

The key relationship between convex functions and convex sets is that **the function f is a convex function if and only if its epigraph** epi(f) **is a convex set**. I will not prove this, but essentially the definition of a convex function checks the "hardest case" of convexity of epi(f). This is the case where we pick two points on the boundary of the epigraph, a.k.a. the graph of f itself.

Now, to prove the theorem.

Proof. For each of the points $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$, there is a corresponding point in $C \times \mathbb{R}$: the points $(\mathbf{x}^{(1)}, f(\mathbf{x}^{(1)}))$ through $(\mathbf{x}^{(k)}, f(\mathbf{x}^{(k)}))$. These are points on the graph of f,and therefore in epi(f).

Because epi(f) is a convex set, their convex combination with weights $\lambda_1, \dots, \lambda_k$ is still in epi(f). That is,

$$\lambda_1 \begin{bmatrix} \mathbf{x}^{(1)} \\ f(\mathbf{x}^{(1)}) \end{bmatrix} + \cdots + \lambda_k \begin{bmatrix} \mathbf{x}^{(k)} \\ f(\mathbf{x}^{(k)}) \end{bmatrix} \in \text{epi}(f).$$

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

That is,

$$\begin{bmatrix} \lambda_1 \mathbf{x}^{(1)} + \cdots + \lambda_k \mathbf{x}^{(k)} \\ \lambda_1 f(\mathbf{x}^{(1)}) + \cdots + \lambda_k f(\mathbf{x}^{(k)}) \end{bmatrix} \in \operatorname{epi}(f).$$

What does it mean for this point to be in $\operatorname{epi}(f)$? It means that its y-coordinate is above the value of f at its $\mathbf{x}$-coordinate. Therefore

$$f(\lambda_1 \mathbf{x}^{(1)} + \lambda_2 \mathbf{x}^{(2)} + \cdots + \lambda_k \mathbf{x}^{(k)}) \le \lambda_1 f(\mathbf{x}^{(1)}) + \lambda_2 f(\mathbf{x}^{(2)}) + \cdots + \lambda_k f(\mathbf{x}^{(k)}),$$

which is the inequality that we wanted. □

Jensen's inequality can be sharpened. If $f : C \to \mathbb{R}$ is strictly convex, and $\lambda_1, \lambda_2, \dots, \lambda_k > 0$, then the only way to get the equation

$$f(\lambda_1 \mathbf{x}^{(1)} + \lambda_2 \mathbf{x}^{(2)} + \cdots + \lambda_k \mathbf{x}^{(k)}) = \lambda_1 f(\mathbf{x}^{(1)}) + \lambda_2 f(\mathbf{x}^{(2)}) + \cdots + \lambda_k f(\mathbf{x}^{(k)})$$

is by setting $\mathbf{x}^{(1)} = \mathbf{x}^{(2)} = \cdots = \mathbf{x}^{(n)}$. This is hard to prove via the epigraph approach, but if you write an induction proof similar to the one we wrote for convex combinations, then it falls out of the definition.

2 Applications of Jensen's inequality

Jensen's inequality—even applied to simple, one-dimensional convex functions—is useful for solving optimization problems in one simple step.

Taking the weights $\lambda_1 = \cdots = \lambda_k = \frac{1}{k}$, Jensen's inequality says that

$$\frac{1}{k} f(x_1) + \cdots + \frac{1}{k} f(x_k) \ge f(\frac{1}{k} x_1 + \cdots + \frac{1}{k} x_k),$$

or

$$f(x_1) + \cdots + f(x_k) \ge k \cdot f\left(\frac{x_1 + \cdots + x_k}{k}\right).$$

In other words, if $x_1 + x_2 + \cdots + x_k$ is fixed and $f : \mathbb{R} \to \mathbb{R}$ is convex, then the sum $f(x_1) + f(x_2) + \cdots + f(x_k)$ is minimized by setting $x_1, \dots, x_k$ all equal to their average. (If f is strictly convex, then this is the unique minimizer.)

2.1 Classic calculus problem

Given 100 feet of fencing, what is the largest rectangular region we can enclose?

Let x_1 be the height and x_2 the width. We are given $2x_1 + 2x_2 = 100$, or $x_1 + x_2 = 50$.

We want to maximize $x_1 x_2$, which does not look like Jensen's inequality. But it's equivalent to minimize $-\log(x_1 x_2) = -\log(x_1) + -\log(x_2)$.

Since $f(x) = -\log x$ is convex, $f(x_1) + f(x_2)$ is minimized when we take $x_1 = x_2 = 25$, giving an area of $x_1 x_2 = 625$.

2.2 Standard combinatorics problem

The integers $1, 2, \ldots, 100$ are colored by 10 colors. At least how many pairs $\{a, b\} \subseteq \{1, 2, \ldots, 100\}$ have the same color?

Let $x_1, x_2, \ldots, x_{10}$ be the number of integers that get color $1, 2, \ldots, 10$. We are given $x_1 + x_2 + \cdots + x_{10} = 100$, since all integers get a color.

If color i has x_i integers, there are $\binom{x_i}{2} = \frac{x_i(x_i-1)}{2}$ pairs of integers that both have color i. So we are trying to minimize

$$\binom{x_1}{2} + \cdots + \binom{x_{10}}{2}.$$

Since $f(x) = \binom{x}{2}$ is a convex function, this is minimized when $x_1 = x_2 = \cdots = x_{10} = 10$. In this case, we have $\binom{10}{2} = 45$ pairs of the same color for each color, and 450 pairs total.

2.3 Indian Math Olympiad, 1995

As an unexpected bonus, Jensen's inequality is useful for high school math competitions.

The problem was this: prove that if $x_1, x_2, \ldots, x_n > 0$ and $x_1 + x_2 + \cdots + x_n = 1$, then

$$\frac{x_1}{\sqrt{1-x_1}} + \frac{x_2}{\sqrt{1-x_2}} + \cdots + \frac{x_n}{\sqrt{1-x_n}} \geq \sqrt{\frac{n}{n-1}}.$$

First, we check that $f(t) = \frac{t}{\sqrt{1-t}}$ is convex. It's easier to check $g(t) = f(1-t) = \frac{1-t}{\sqrt{t}}$ because $g(t) = \frac{1}{\sqrt{t}} - \frac{t}{\sqrt{t}} = t^{-1/2} + (-t^{1/2})$, and both terms are convex. Since $g(t)$ is convex, $f(t) = g(1-t)$ is convex by the second composition-of-convex-functions result we proved.

Now, by Jensen's inequality with weights $\lambda_1 = \cdots = \lambda_n = \frac{1}{n}$, we have

$$\frac{1}{n}\left(\frac{x_1}{\sqrt{1-x_1}} + \frac{x_2}{\sqrt{1-x_2}} + \cdots + \frac{x_n}{\sqrt{1-x_n}}\right) \geq f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = f\left(\frac{1}{n}\right) = \frac{1/n}{\sqrt{1-1/n}}$$

which simplifies to the inequality we wanted.

2.4 The AM-GM inequality

The first example we did can be generalized to a result called the AM-GM (Arithmetic Mean-Geometric Mean) inequality. It states the following:

Theorem 2.1 (AM-GM inequality)**.** *For any $x_1, x_2, \ldots, x_n \geq 0$,*

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}$$

with equality only if $x_1 = x_2 = \cdots = x_n$.

It also has a weighted form:

Theorem 2.2 (Weighted AM-GM inequality)**.** *For any* $x_1, x_2, \ldots, x_n \geq 0$ *and for any weights* $\delta_1, \delta_2, \ldots, \delta_n > 0$ *with* $\delta_1 + \delta_2 + \cdots + \delta_n = 1$,

$$\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n \geq x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}$$

with equality only if $x_1 = x_2 = \cdots = x_n$.

As before, let $f(t) = -\ln t$: this is a strictly convex function on $(0, \infty)$, since $f''(t) = \frac{1}{t^2} > 0$ for all t. Jensen's inequality says that

$$f(\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n) \leq \delta_1 f(x_1) + \delta_2 f(x_2) + \cdots + \delta_n f(x_n).$$

When $x_1, x_2, \ldots, x_n$ are not all equal, because f is strictly convex, we get a $>$ in this inequality. That's where the equality condition of AM-GM comes from.

Now let's try to simplify this inequality a bit. Once we replace f by its definition, we get

$$-\ln(\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n) \leq -\delta_1 \ln x_1 - \delta_2 \ln x_2 - \cdots - \delta_n \ln x_n$$

and we can negate both sides to reverse the inequality:

$$\ln(\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n) \geq \delta_1 \ln x_1 + \delta_2 \ln x_2 + \cdots + \delta_n \ln x_n.$$

Now get rid of the ln by applying e^x to both sides:

$$\begin{aligned} \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_n x_n &\geq e^{\delta_1 \ln x_1 + \delta_2 \ln x_2 + \cdots + \delta_n \ln x_n} \\ &= e^{\delta_1 \ln x_1} e^{\delta_2 \ln x_2} \cdots e^{\delta_n \ln x_n} \\ &= x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}. \end{aligned}$$

This gives us the weighted AM-GM inequality.

(A minor note: f is convex on $(0, \infty)$ and not even defined at 0, but we stated AM-GM for $x_1, x_2, \ldots, x_n \geq 0$. Is this a problem? It's easily fixed: when $x_i = 0$ for any i, then $x_1^{\delta_1} x_2^{\delta_2} \cdots x_n^{\delta_n}$ immediately becomes 0. On the other side, the arithmetic mean remains nonnegative, and it's strictly positive unless $x_1 = x_2 = \cdots = x_n = 0$. So we're still good.)