Math 484: Nonlinear Programming[1]	Mikhail Lavrov

Chapter 2, Lecture 3: Building convex functions

February 8, 2019 — *University of Illinois at Urbana-Champaign*

1 The plan

Today's goal is to figure out how we can tell if a function is convex.

So far, we know three ways, each with their own drawbacks:

1. The definition: a function $f : C \to \mathbb{R}$ is convex if and only if, for all $\mathbf{x}, \mathbf{y} \in C$ and $0 \le t \le 1$,
$$f(t\mathbf{x} + (1-t)\mathbf{y}) \le tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$
This is a useful property, and often useful in proofs, but for many actual functions, it's not clear how to prove this.

2. The tangent line test: for differential functions, f is convex if and only if, for all $\mathbf{x}, \mathbf{y} \in C$,
$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}).$$
This is more useful as a consequence of convexity, rather than as a way to prove that a function is convex.

3. The second-derivative test: assuming that $Hf(\mathbf{x})$ exists for all $\mathbf{x}$, f is convex if, for all $\mathbf{x} \in C$, $Hf(\mathbf{x}) \succeq 0$.

   This is often the way to go, but if f is a complicated function, lots of computation is involved.

Today, we take a different approach. The definition and the second-derivative test are easy to use for functions with simple definitions. To deal with more complicated functions, we look at how they are built out of simpler "building blocks". We will prove several results about how we can manipulate convex functions to get more complicated convex functions.

1.1 Strictly convex functions

But first, an aside for another definition.

Given a set $C \subseteq \mathbb{R}^n$ (convex, as always), a function $f : C \to \mathbb{R}$ is called *strictly convex* when, for all $\mathbf{x}, \mathbf{y} \in C$ with $\mathbf{x} \ne \mathbf{y}$ and $0 < t < 1$,
$$f(t\mathbf{x} + (1-t)\mathbf{y}) < tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

These have a slightly sharper version of most properties of convex functions. For example, if a function is strictly convex, then any local minimizer (and any critical point) is not just a global minimizer, but a strict global minimizer. So having strict convexity is often nice.

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

We'll need to watch out for it today so that we know what operations preserve strict convexity, not just convexity.

Warning: the second-derivative tests can show that a function is strictly convex. If $Hf(\mathbf{x}) \succ 0$ for all $\mathbf{x} \in C$, then f is strictly convex. But the implication doesn't go both ways. For example, $f(x) = x^4$ has $f''(0) = 0$, but is still strictly convex.

2 Building convex functions

2.1 The less-scary ways to combine functions

The simplest and most important operations that preserve convexity are addition and multiplication by a positive scalar.

Theorem 2.1 (Theorem 2.3.10 in the textbook)**.** *Suppose that $f_1, f_2, \dots, f_k$ are convex functions $C \to \mathbb{R}$ and $\alpha_1, \alpha_2, \dots, \alpha_k$ are positive*[2] *scalars. Then*

$$f(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i f_i(\mathbf{x}) = \alpha_1 f_1(\mathbf{x}) + \alpha_2 f_2(\mathbf{x}) + \cdots + \alpha_k f_k(\mathbf{x})$$

is convex. Moreover, if at least one f_i is strictly convex, then f is strictly convex.

Proof. It's enough to prove two simple cases of this theorem rather than deal with the arbitrary sum.

First, if f is (strictly) convex and $\alpha > 0$, then αf is (strictly) convex. This holds because we can just multiply both sides of the definition by α:

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \ge tf(\mathbf{x}) + (1-t)f(\mathbf{y}) \iff \alpha f(t\mathbf{x} + (1-t)\mathbf{y}) \ge t\alpha f(\mathbf{x}) + (1-t)\alpha f(\mathbf{y}).$$

Second, if f and g are convex, then their sum h defined by $h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ is convex. This is also just a matter of adding together the two inequalities:

$$\begin{aligned}
h(t\mathbf{x} + (1-t)\mathbf{y}) &= f(t\mathbf{x} + (1-t)\mathbf{y}) + g(t\mathbf{x} + (1-t)\mathbf{y}) && \text{(definition of } h\text{)} \\
&\le tf(\mathbf{x}) + (1-t)f(\mathbf{y}) + g(t\mathbf{x} + (1-t)\mathbf{y}) && (f \text{ is convex}) \\
&\le tf(\mathbf{x}) + (1-t)f(\mathbf{y}) + tg(\mathbf{x}) + (1-t)g(\mathbf{y}) && (g \text{ is convex}) \\
&= th(\mathbf{x}) + (1-t)h(\mathbf{y}). && \text{(definition of } h\text{)}
\end{aligned}$$

There are two inequalities here. If either f or g is strictly convex, then one inequality is strict; so the whole inequality becomes strict, and the sum $h = f + g$ is strictly convex.

To get the full theorem, we just build up the combination of k functions by induction. □

Here is another relatively simple result. It rarely comes up, but when it does, it's often the only tool we have.

[2] The textbook says "nonnegative", but if $\alpha_i = 0$ it's as though we didn't include f_i at all.

Theorem 2.2. *Suppose that $f_1, f_2, \ldots, f_k$ are (strictly) convex functions $C \to \mathbb{R}$. Then*

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})\}$$

is (strictly) convex.

Proof. As usual, take $\mathbf{x}, \mathbf{y} \in C$ and $t \in [0,1]$. Then $f(t\mathbf{x}+(1-t)\mathbf{y})$ must be equal to $f_i(t\mathbf{x}+(1-t)\mathbf{y})$ for some i, and we have

$$\begin{aligned} f(t\mathbf{x} + (1-t)\mathbf{y}) &= f_i(t\mathbf{x} + (1-t)\mathbf{y}) && \text{(we are at a point where } f = f_i\text{)} \\ &\le t f_i(\mathbf{x}) + (1-t) f_i(\mathbf{y}) && (f_i \text{ is convex}) \\ &\le t f(\mathbf{x}) + (1-t) f(\mathbf{y}) \end{aligned}$$

where the last inequality holds because for any $i = 1, 2, \ldots, k$ and any point $\mathbf{x} \in C$, $f(\mathbf{x}) \ge f_i(\mathbf{x})$ because $f(\mathbf{x})$ is a maximum of several values including $f_i(\mathbf{x})$.

If all the f_i are strictly convex and $0 < t < 1$, we get to use a strict inequality in this proof and so f is strictly convex. □

It's also important to mention that multiplying two convex functions does not guarantee convexity: for example, $f(x) = x^2 - 1$ is convex, but $f(x) \cdot f(x) = (x^2-1)^2$ is not. Also, the minimum of two convex functions isn't convex, even though min looks a lot like max.

2.2 Compositions of functions

The final way of combining functions we'll cover is composition. We ask: *when* is it true (it's certainly not always true) that the composition $g(f(\mathbf{x}))$ of two convex functions f and g is convex?

Theorem 2.3 (Also Theorem 2.3.10 in the textbook)**.** *Suppose $f : C \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is not only convex but* increasing*: when $x_1 \le x_2$, $g(x_1) \le g(x_2)$. Then $h(\mathbf{x}) = g(f(\mathbf{x}))$ is convex.*

(If f is strictly convex and g is strictly increasing—when $x_1 < x_2$, $g(x_1) < g(x_2)$—then h is strictly convex as well.)

Proof. The proof is short; the hard part is watching out for this rule in examples. We have (in the usual setup for a convexity proof):

$$\begin{aligned} f(t\mathbf{x}) + (1-t)\mathbf{y}) &\le t f(\mathbf{x}) + (1-t) f(\mathbf{y}), && (f \text{ is convex}) \\ g(f(t\mathbf{x}) + (1-t)\mathbf{y})) &\le g(t f(\mathbf{x}) + (1-t) f(\mathbf{y})) && (g \text{ is increasing}) \\ &\le t g(f(\mathbf{x})) + (1-t) g(f(\mathbf{y})). && (g \text{ is convex}) \end{aligned}$$

If f is strictly convex, then the first inequality is strict (it's $<$). If g is strictly increasing, then that strict inequality is preserved, so h is strictly convex as well. □

Another useful case, which is in the textbook as a post-chapter exercise:

Theorem 2.4. *If $f : \mathbb{R}^m \to \mathbb{R}^n$ has the form $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for a matrix A and a vector $\mathbf{b}$, and $g : C \to \mathbb{R}$ is convex, so is $h(\mathbf{x}) = g(f(\mathbf{x}))$ as a function $f^{-1}(C) \to \mathbb{R}$.*

Proof. Such a function f has the useful property that it's convex, and the definition of convex is always an equality: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $t \in [0,1]$ (actually, any t), we have

$$\begin{aligned} f(t\mathbf{x} + (1-t)\mathbf{y}) &= A(t\mathbf{x} + (1-t)\mathbf{y}) + \mathbf{b} \\ &= t(A\mathbf{x} + \mathbf{b}) + (1-t)(A\mathbf{y} + \mathbf{b}) \\ &= tf(\mathbf{x}) + (1-t)f(\mathbf{y}). \end{aligned}$$

So we have

$$\begin{aligned} h(t\mathbf{x} + (1-t)\mathbf{y}) &= g(f(t\mathbf{x} + (1-t)\mathbf{y})) && \text{(definition of } h\text{)} \\ &= g(tf(\mathbf{x}) + (1-t)f(\mathbf{y})) && \text{(what we proved above)} \\ &\le tg(f(\mathbf{x})) + (1-t)g(f(\mathbf{y})) && (g \text{ is convex}) \\ &\le th(\mathbf{x}) + (1-t)h(\mathbf{y}). && \text{(definition of } h\text{)} \end{aligned}$$

□

Noteworthy special case: if $g : \mathbb{R} \to \mathbb{R}$ is convex, so is $h(x) = g(ax + b)$. Also, by choosing the matrix A appropriately, we know that $h(\mathbf{x}) = g(x_i)$ is convex as a function $\mathbb{R}^n \to \mathbb{R}$.

3 Examples

3.1 Example 2.3.11.c in the textbook

To check that $f(x_1, x_2) = x_1^2 - 4x_1x_2 + 5x_2^2 - \ln x_1x_2$ is convex for $x_1, x_2 > 0$, write f as a sum

$$f(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + (-\ln x_1) + (-\ln x_2).$$

The first term is convex because $\begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \succeq 0$.

The second term and theird term are convex because $g(x) = -\ln x$ is convex on $(0, \infty)$ (and by our last result, plugging in just x_1 or just x_2 doesn't change this).

Finally, the sum of three convex functions is convex.

3.2 A function that will be useful for us soon

The function $f(x, y, z) = \left(\frac{x}{2}\right)^x \left(\frac{y}{3}\right)^y \left(\frac{z}{4}\right)^z$ is convex on $\{(x, y, z) : x, y, z > 0\}$.

To show this, write it as $e^{x \ln \frac{x}{2} + y \ln \frac{y}{3} + zln \frac{z}{4}}$. If ewe can show that the inside function is convex, we are done, because e^t is convex and increasing.

For any constant C, $g(x) = x \ln \frac{x}{C}$ has first derivative $\ln \frac{x}{C} + x \cdot \frac{1}{x/C} = \ln \frac{x}{C} + C$, and second derivative $\frac{C}{x}$. When $x > 0$, this is guaranteed to be positive, so g is convex. Applying this to the three parts of the exponent with $C = 2$, $C = 3$, and $C = 4$ concludes the example.