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 $x, y \in C$ and $0 \leq t \leq 1$,

   $$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

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

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

3. The epigraph test: $f$ is convex if and only if

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

   is a convex set.

   This is rarely useful outside proofs: testing if a set is convex is not any easier than testing if a function is convex.

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 $x, y \in C$ with $x \neq y$ and $0 < t < 1$,

$$f(tx + (1 - t)y) < tf(x) + (1 - t)f(y).$$

This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-fall-2018.html
These have a slightly sharper version of most properties of convex functions. For example, Jensen’s inequality holds strictly (that is, with a $<$) for strictly convex functions. And 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.

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

The second-derivative tests can show that a function is strictly convex: if $Hf(x) \succ 0$ for all $x \in C$, 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 $f_1, f_2, \ldots, f_k$ are convex functions $C \to \mathbb{R}$ and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive\(^2\) scalars. Then

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

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 $\alpha f$ is (strictly) convex. This holds because we can just multiply both sides of the definition by $\alpha$:

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \iff \alpha f(tx + (1-t)y) \geq t\alpha f(x) + (1-t)\alpha f(y).$$

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

$$h(tx + (1-t)y) = f(tx + (1-t)y) + g(tx + (1-t)y) \overset{\text{(definition of } h)}{=} tf(x) + (1-t)f(y) + g(tx + (1-t)y) \overset{f \text{ is convex}}{\leq} tf(x) + (1-t)f(y) + tg(x) + (1-t)g(y) \overset{g \text{ is convex}}{\leq} th(x) + (1-t)h(y).$$

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. \(\square\)

\(^2\)The textbook says “nonnegative”, but if $\alpha_i = 0$ it’s as though we didn’t include $f_i$ at all.
Here is another relatively simple result. It rarely comes up, but when it does, it’s often the only tool we have.

**Theorem 2.2.** Suppose that \( f_1, f_2, \ldots, f_k \) are (strictly) convex functions \( C \to \mathbb{R} \). Then
\[
f(x) = \max\{f_1(x), f_2(x), \ldots, f_k(x)\}
\]
is (strictly) convex.

**Proof.** A short argument is that \( \text{epi}(f) = \text{epi}(f_1) \cap \text{epi}(f_2) \cap \cdots \cap \text{epi}(f_k) \) and we know intersections of convex sets are convex. But we can also get there by working with the inequalities.

As usual, take \( x, y \in C \) and \( t \in [0, 1] \). Then \( f(tx + (1-t)y) \) must be equal to \( f_i(tx + (1-t)y) \) for some \( i \), and we have
\[
\begin{align*}
f(tx + (1-t)y) &= f_i(tx + (1-t)y) & \text{(we are at a point where } f = f_i) \\
&\leq tf_i(x) + (1-t)f_i(y) & \text{ (} f_i \text{ is convex)} \\
&\leq tf(x) + (1-t)f(y)
\end{align*}
\]
where the last inequality holds because for any \( i = 1, 2, \ldots, k \) and any point \( x \in C \), \( f(x) \geq f_i(x) \) because \( f(x) \) is a maximum of several values including \( f_i(x) \).

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 \( g(f(x)) \) 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 \( x_1 \leq x_2 \), \( g(x_1) \leq g(x_2) \). Then \( h(x) = g(f(x)) \) 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):
\[
\begin{align*}
f(tx) + (1-t)y &\leq tf(x) + (1-t)f(y), & \text{(} f \text{ is convex)} \\
g(f(tx) + (1-t)y) &\leq g(tf(x) + (1-t)f(y)) \\
&\leq tg(f(x)) + (1-t)g(f(y)). & \text{(} g \text{ is convex)}
\end{align*}
\]
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 \( f(x) = Ax + b \) for a matrix \( A \) and a vector \( b \), and \( g : C \to \mathbb{R} \) is convex, so is \( h(x) = g(f(x)) \) 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 \( x, y \in \mathbb{R}^m \) and \( t \in [0, 1] \) (actually, any \( t \)), we have

\[
\begin{align*}
f(tx + (1-t)y) &= A(tx + (1-t)y) + b \\
&= t(Ax + b) + (1-t)(Ay + b) \\
&= tf(x) + (1-t)f(y).
\end{align*}
\]

So we have

\[
\begin{align*}
h(tx + (1-t)y) &= g(f(tx + (1-t)y)) \quad \text{(definition of \( h \))} \\
&= g(tf(x) + (1-t)f(y)) \quad \text{(what we proved above)} \\
&\leq tg(f(x)) + (1-t)g(f(y)) \quad \text{(\( g \) is convex)} \\
&\leq th(x) + (1-t)h(y). \quad \text{(definition of \( h \))}
\end{align*}
\]

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 \( h(x) = g(x_i) \) 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

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

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

The second term and the third 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 Indian Math Olympiad, 1995

We probably won’t get to this in class, but it’s nice to know that what we’ve learned so far lets you do well on national 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

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

This is an application of Jensen’s inequality. 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

\[
\frac{1}{n} \left( \frac{x_1}{\sqrt{1 - x_1}} + \frac{x_2}{\sqrt{1 - x_2}} + \cdots + \frac{x_n}{\sqrt{1 - x_n}} \right) = \frac{1/n}{\sqrt{1 - 1/n}}
\]

which simplifies to the inequality we wanted.