1 Definitions of convex functions

Convex functions \( f : \mathbb{R}^n \to \mathbb{R} \) with \( Hf(x) \succeq 0 \) for all \( x \), or functions \( f : \mathbb{R} \to \mathbb{R} \) with \( f''(x) \geq 0 \) for all \( x \), are going to be our model of what we want convex functions to be. But we actually work with a slightly more general definition that doesn’t require us to say anything about derivatives.

Let \( C \subseteq \mathbb{R}^n \) be a convex set. A function \( f : C \to \mathbb{R} \) is convex on \( C \) if, for all \( x, y \in C \), the inequality holds that
\[
f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).
\]
(We ask for \( C \) to be convex so that \( tx + (1-t)y \) is guaranteed to stay in the domain of \( f \).)

This is easiest to visualize in one dimension:

The point \( tx + (1-t)y \) is somewhere on the line segment \( [x, y] \). The left-hand side of the definition, \( f(tx + (1-t)y) \), is just the value of the function at that point: the green curve in the diagram. The right-hand side of the definition, \( tf(x) + (1-t)f(y) \), is the dashed line segment: a straight line that meets \( f \) at \( x \) and \( y \).

So, geometrically, the definition says that secant lines of \( f \) always lie above the graph of \( f \).

Although the picture we drew is for a function \( \mathbb{R} \to \mathbb{R} \), nothing different happens in higher dimensions, because only points on the line segment \( [x, y] \) (and \( f \)'s values at those points) play a role in the inequality.

The nice thing about this definition is that:

- All the nice things we’ve said about functions with \( Hf(x) \succeq 0 \), and more, still hold for convex functions in general.
- But we don’t have to deal with derivatives to prove them.

For example, we can show the following result.

---

1This document comes from the Math 484 course webpage: https://faculty.math.illinois.edu/~mlavrov/courses/484-fall-2018.html
**Theorem 1.1.** If $C \subseteq \mathbb{R}^n$ is a convex set, $f : C \to \mathbb{R}$ is a convex function, and $x^* \in C$ is a local minimizer of $f$, then it is a global minimizer.

**Proof.** Suppose not: suppose there is a point $y \in C$ with $f(y) < f(x^*)$.

Then the secant line that meets $f$ at $x^*$ and $y$ is decreasing in slope in the direction from $x^*$ to $y$. In particular, if we move along the secant line from $x^*$ to $y$, we always stay below $f(x^*)$. But the graph of $f$ is below the secant line, so the graph of $f$ always stays below $f(x^*)$.

This means that if we move in the direction of $y$ even a tiny bit, then we get a point with smaller $f$-value, contradicting the assumption that $x^*$ was a local minimizer.

Okay, we can also do all that in algebra, which might be more convincing to some people. Assuming $x^*$ is a local minimizer, there is an open ball $B$ around $x^*$ such that $f(x^*) \leq f(y)$ for all $y \in B$. For any $x$, choose $t$ small enough that $tx + (1 - t)x^* \in B$. Then

$$f(x^*) \leq f(tx + (1 - t)x^*) \leq tf(x) + (1 - t)f(x^*)$$

by convexity of $f$, which can be rearranged to $tf(x^*) \leq tf(x)$, or $f(x^*) \leq f(x)$.

\[\Box\]

2 Jensen’s inequality

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

**Theorem 2.1.** For any $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ with $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$, if $f : C \to \mathbb{R}$ is convex and $x^{(1)}, \ldots, x^{(k)} \in C$, then

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

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

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

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 $\text{epi}(f)$ is a convex set. I will not prove this, but essentially the definition of a convex function checks the “hardest case” of convexity of $\text{epi}(f)$: the case where we pick two points on the boundary of the epigraph—the graph of $f$ itself.

Now, to prove the theorem.
Proof. For each of the points \( x^{(1)}, \ldots, x^{(k)} \), there is a corresponding point in \( C \times \mathbb{R} \): the points \((x^{(1)}, f(x^{(1)})\) through \((x^{(k)}, f(x^{(k)})\). 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 \( \lambda_1, \ldots, \lambda_k \) is still in epi(\( f \)). This convex combination is the point

\[
(x^*, y^*) = (\lambda_1 x^{(1)} + \cdots + \lambda_k x^{(k)}, \lambda_1 f(x^{(1)}) + \cdots + \lambda_k f(x^{(k)})).
\]

Its first \( n \) coordinates (its position in the graph) are the weighted average of \( x^{(1)}, \ldots, x^{(k)} \), and its \((n+1)\)th coordinate (its height in the graph) is the weighted average of their \( f \)-values.

What does it mean for this point to be in epi(\( f \))? It means that its \( y \)-coordinate is above the value of \( f \) at its \( x \)-coordinate: \( f(x^*) \leq y^* \).

Looking at what \( x^* \) and \( y^* \) are, this is precisely the inequality

\[
f(\lambda_1 x^{(1)} + \lambda_2 x^{(2)} + \cdots + \lambda_k x^{(k)}) \leq \lambda_1 f(x^{(1)}) + \lambda_2 f(x^{(2)}) + \cdots + \lambda_k f(x^{(k)})
\]

that we wanted. \( \square \)

3 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

\[
\frac{1}{k} f(x_1) + \cdots + \frac{1}{k} f(x_k) \geq f\left(\frac{1}{k} x_1 + \cdots + \frac{1}{k} x_k\right),
\]

or

\[
f(x_1) + \cdots + f(x_k) \geq k \cdot f\left(\frac{x_1 + \cdots + x_k}{k}\right).
\]

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 \( x_1, \ldots, x_k \) all equal to their average.

3.1 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 \( x_1 x_2 = 625 \).
3.2 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.