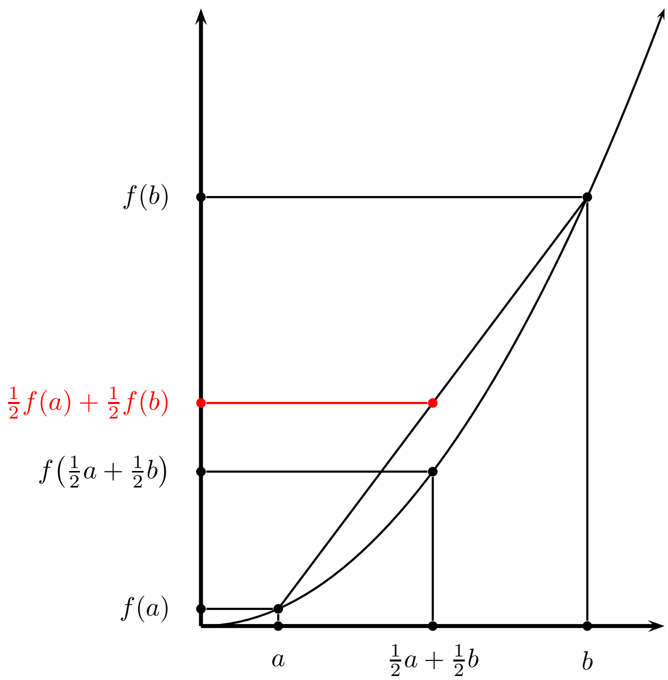
CS 624
Convex Functions

1 Definitions



We say that a function is *convex* iff it lies below any of its secants.

More precisely, suppose that $a < b$. The point halfway between a and b is of course $\frac{1}{2}a + \frac{1}{2}b$. The point on the line from $\bigl(a, f(a)\bigr)$ to $\bigl(b, f(b)\bigr)$ that is halfway in between those points has

$$\begin{aligned} x\text{-coordinate} &= \tfrac{1}{2}a + \tfrac{1}{2}b \\ y\text{-coordinate} &= \tfrac{1}{2}f(a) + \tfrac{1}{2}f(b) \end{aligned}$$

And so if f is a convex function, we must have

$$f\left(\tfrac{1}{2}a + \tfrac{1}{2}b\right) \leq \tfrac{1}{2}f(a) + \tfrac{1}{2}f(b)$$

Suppose we want to consider a different point between a and b. For instance, suppose we consider the point that is $\frac{1}{3}$ of the way from a to b. This point is

$$\tfrac{2}{3}a + \tfrac{1}{3}b$$

and because f is convex, we have

$$f\left(\tfrac{2}{3}a + \tfrac{1}{3}b\right) \le \tfrac{2}{3}f(a) + \tfrac{1}{3}f(b)$$

We can go even farther: suppose $0 \le p \le 1$ and we set $q = 1 - p$, so $p + q = 1$ and both p and q are between 0 and 1. Then the same reasoning shows that

$$f(pa + qb) \le pf(a) + qf(b)$$

2 Jensen's inequality

And in fact, we can go even farther. Suppose we have a set of numbers $\{p_1, p_2, \dots, p_n\}$ such that

- $0 \le p_i \le 1$ for all i, and
- $\sum_{i=1}^n p_i = 1$.

Then if $\{x_1, x_2, \dots, x_n\}$ are any real numbers, we have

$$f\left(\sum_{i=1}^n p_i x_i\right) \le \sum_{i=1}^n p_i f(x_i)$$

This can be proved by induction, among other ways. It is called *Jensen's inequality.* It is the single most important fact about convex functions, and it is incredibly useful.

Note that the numbers x_i are just any real numbers. And the indices i are not special either. The fact that they run over the set $\{1, 2, \dots, n\}$ is not significant—they could run over any finite set.

For instance, we could let S be a finite set, and let s run over the elements of S, and let $x(s)$ be a function on S (and, similarly, $p(s)$ would also be a function on S). Then Jensen's inequality would be written

$$f\left(\sum_{s\in S} p(s)x(s)\right) \le \sum_{s\in S} p(s)f\big(x(s)\big)$$

3 An application to probability theory

Suppose S is a "sample space" in probability theory. To keep things simple—which is all we will need in this course, really—let us assume that S is a finite set, and that each element $s \in S$ has an associated probability $p(s)$, such that for all $s \in S$, $0 \le p(s) \le 1$ and

$$\sum_{s \in S} p(s) = 1$$

Now suppose X is a random variable on S. This is just a fancy way of saying that X is a function on S. In other words, X associates to every element $s \in X$ a number $X(s)$ (with probability $p(s)$).

Again to make things simple, we will assume that X is real-valued. (One can also consider functions with values in the complex numbers or in higher-dimensional spaces, for instance, but we won't have any need for this here.)

By definition, the *expectation* of X (or the *expected value* of X) is

$$E(X) = \sum_{s \in S} p(s) X(s)$$

Now suppose that f is any convex function. Then of course $f(X)$ is also a random variable. (Clearly it's a function on S.) And by Jensen's inequality we have

$$\begin{aligned} f\bigl(E(X)\bigr) &= f\Bigl(\sum_{s \in S} p(s) X(s)\Bigr) \\ &\le \sum_{s \in S} p(s) f\bigl(X(s)\bigr) \\ &= E\bigl(f(X)\bigr) \end{aligned}$$

4 More about convex functions and applications of Jensen's inequality[1]

4.1 How can you tell if a function is convex?

One easy way is this: A function f (which we will assume is defined on an interval, which may be all of $\mathbf{R}$) is convex if it is twice differentiable (that is, if the second derivative $f''(x)$ exists for all x on that interval) and if $f''(x) \ge 0$ for all x on that interval[2].

4.1 Exercise *Prove this. [Hint: The condition implies that $f'(x)$ is increasing. Then you could use the mean value theorem.]*

This immediately shows that the functions $f(x) = x^p$ are convex for every $p \ge 1$, and also shows that the functions $g(x) = b^x$ are convex for every $b \ge 1$. (And *please* remember that the derivative $g'(x)$ is not entirely trivial...) And the function $h(x) = -\log x$ is also convex.

[1] You don't need to read this section for the course. It's just some more background that I hope some of you will find interesting and give you some more insight into how people look at these things.

[2] Note that this condition is sufficient for f to be convex, but is not necessary. For instance, the function $f(x) = |x|$ is convex, but is not twice differentiable—even its first derivative does not exist at $x = 0$.

4.2 Some famous applications of Jensen's inequality

Let us consider the convex function $f(x) = x^2$. We'll consider a set of n numbers $\{x_i : 1 \leq i \leq n\}$ and corresponding "weights" $\{p_i : 1 \leq i \leq n\}$ with each $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. (In the simplest case, all the weights are equal, and each p_i is just $\frac{1}{n}$. If you are seeing this for the first time, that's a good simplification to make. You might find it helpful to write out everything in this section for yourself with $\frac{1}{n}$ substituted for each p_i.) From Jensen's inequality then, we get

$$\Big(\sum_{i=1}^n p_i x_i\Big)^2 \leq \sum_{i=1}^n p_i x_i^2$$

Or equivalently,

$$\sum_{i=1}^n p_i x_i \leq \Big(\sum_{i=1}^n p_i x_i^2\Big)^{\frac{1}{2}}$$

The left-hand side of this inequality is a weighted arithmetic mean[3] of the numbers $\{x_i\}$. (The weights are just the numbers $\{p_i\}$.) The right-hand-side is the weighted *root-mean-square* of those numbers. The root-mean-square is a kind of average that is often used in physics and engineering to express an average error. So we see that Jensen's inequality shows that the arithmetic mean is bounded above by the root-mean-square.

For another example, let us consider the convex function $f(x) = -\log x$. Of course $\log x$ is only defined for $x > 0$, so we will assume that all the numbers $\{x_i\}$ are positive. From Jensen's inequality, we get

$$-\log\Big(\sum_{i=1}^n p_i x_i\Big) \leq \sum_{i=1}^n p_i(-\log x_i)$$

Or equivalently,

$$-\log\Big(\sum_{i=1}^n p_i x_i\Big) \leq -\sum_{i=1}^n p_i \log x_i$$

which is equivalent to

$$\sum_{i=1}^n p_i \log x_i \leq \log\Big(\sum_{i=1}^n p_i x_i\Big)$$

Now the left-hand side of this inequality is just

$$\sum_{i=1}^n \log\Big(x_i^{p_i}\Big) = \log \prod_{i=1}^n x_i^{p_i}$$

[3]often called simply the "weighted average"

So we have

$$\log \prod_{i=1}^{n} x_i^{p_i} \leq \log\Big(\sum_{i=1}^{n} p_i x_i\Big)$$

and applying the function $x \to e^x$ to both sides (which we can do since the exponential function is an increasing function, and so maintains the inequality), we get

$$\prod_{i=1}^{n} x_i^{p_i} \leq \sum_{i=1}^{n} p_i x_i$$

This is famous *Theorem of the Arithmetic and Geometric Means*. The expression on the right is just (as before) the weighted arithmetic mean of the numbers $\{x_i\}$, and the expression on the left is the weighted geometric mean of those numbers, and the theorem states that the geometric mean is always bounded above by the arithmetic mean.

In the simple case when $n = 2$ and $p_1 = p_2 = \frac{1}{2}$, this theorem simply becomes the statement

$$\sqrt{ab} \leq \frac{a+b}{2}$$

and this inequality is very easy to prove directly. But the general form is much harder. Sometimes you see it proved by induction. But it's amazing how simply it follows from Jensen's inequality.