EE364a Review Session 2: Convex Functions

April 14, 2009

In this review session, we will cover convex functions and some methods for determining if a function is convex. We will talk about some basic building blocks and then operations which preserve convexity. We'll also briefly discuss the conjugate function.

Announcements:

- TA office hours: Tuesdays 6:15–8:15pm, in Meyer 143 (location change); Wednesdays and Thursdays 4–8pm, in Packard 277.
- Download and install CVX from `http://www.stanford.edu/~boyd/cvx/`. Be sure to skim through the user guide (sections 2 and 3 will be particularly helpful for this week's homework).

Tools for determining convexity

A way to check the convexity of a function is to check if it satisfies one of the conditions of convexity:

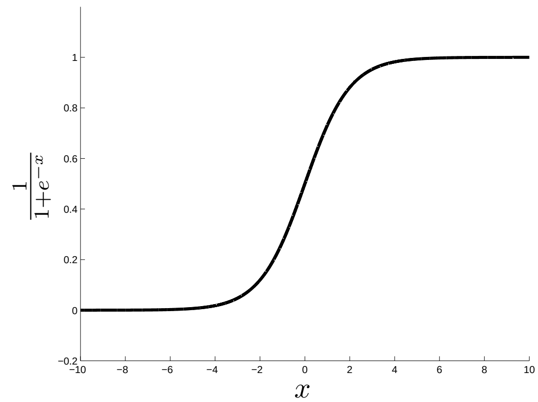
- Definition of convexity: If the function satisfies Jensen's inequality: $f(\theta x+(1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$ for $0 \leq \theta \leq 1$, then the function is convex. It is often sufficient to check for midpoint convexity—that is, by only considering $\theta = \frac{1}{2}$.
- First order condition: For a differentiable function $f(x)$, the first order condition $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ says that the function always lies above any tangent. If the first order condition holds, then the function is convex.
- Second order condition: If the Hessian is positive semidefinite (*i.e.*, $\nabla^2 f(x) \succeq 0$), then the function is convex.
- Epigraph: A function is convex if and only if its epigraph is a convex set.
- Restriction to a line: If the restriction of a function to any line within its domain is always convex, then the function is convex.
- Simple examples: negative log, norm, quadratic-over-linear, log-sum-exp, ...

Example: Consider the sigmoid function (over **R**)

$$f(x) = \frac{1}{1+e^{-x}}.$$

Is it convex? Concave? Quasiconvex? Quasiconcave? Log-convex? Log-concave?

Solution. For functions over **R**, it is often helpful to first plot the function.



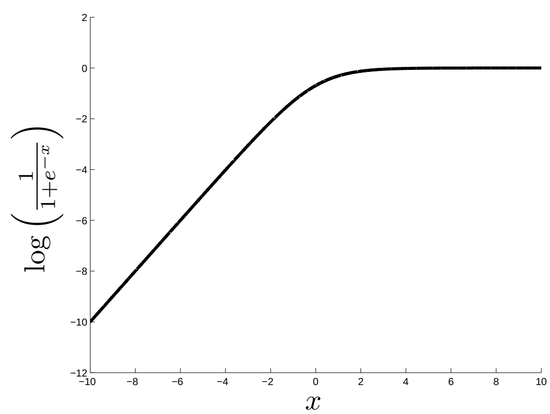
The function is obviously not convex. If we draw the chord between two points x_1 and x_2, where x_1 and x_2 are both greater than 0, it is easy to see that the chord lies below the function. Similarly, we can show that the function is not concave, by drawing the chord between two points which are both less than 0; this chord will lie above the function. Alternatively, we can compute the second derivative

$$f''(x) = -\frac{e^{-x}(1-e^{-x})}{(1+e^{-x})^3},$$

which is negative for positive x, and is positive for negative x.

The sigmoid function is quasiconvex, as its sublevel sets are convex. Additionally, the function is quasiconcave, as its superlevel sets are also convex.

The sigmoid function is not log-convex. This can be quickly verified by plotting $\log f(x)$, which is not convex.



However, the sigmoid function is log-concave. In fact, $\log f(x)$ is the negative of the log-sum-exp function of $z = (0, -x)$; *i.e.*, $\log f(x) = -\log(e^0 + e^{-x})$. Log-concavity of the sigmoid function follows from the fact that log-sum-exp is a convex function.

Operations that preserve convexity

The basic tests for checking the convexity of a function become very powerful tools when used in conjunction with convexity-preserving operations. By starting with some basic building block convex functions and applying convexity-preserving operations, the problem of determining convexity of a function becomes more tractable. Some operations which preserve convexity are the following:

- Nonnegative weighted sum
- Composition with an affine function
- Pointwise maximum and supremum (sup means least upper bound; see Appendix A.2.2)
- Minimization (over convex sets)
- Perspective
- Composition

Example: Is the following a convex function (in $x, y, z \in \mathbf{R}$)?

$$f(x,y,z) = 4 \cdot \max\left(1 + |x| - y, \frac{1}{\sqrt{z}}, 0\right) + 3 \cdot \frac{(x-z)^2}{y+1}$$

(with domain $y + 1 > 0$, $z > 0$)

Solution: The function can be separated into the two terms of the non-negative weighted sum. Consider the individual arguments inside the max term. $|x|$ is convex in x, and $1 - y$ is affine, so $1 + |x| - y$ is convex. Also, $\frac{1}{\sqrt{z}}$ is a negative-power function, so it is convex in z. The max term is convex, since each of its arguments is convex. The second term $\frac{(x-z)^2}{y+1}$ is the composition of the quadratic-over-linear function $\frac{s^2}{t}$ with the affine function that maps (x, y, z) to $(x - z, y + 1)$—so this term is convex. The function $f(x, y, z)$ is the nonnegative weighted sum of convex functions, and thus is convex.

Composition rules

Consider the composition of a function $g : \mathbf{R}^n \to \mathbf{R}^k$ with the function $h : \mathbf{R}^k \to \mathbf{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x)).$$

Given certain conditions on g_i and h, the composition rules can be used to show convexity (or concavity) of f. For example, f is convex if g_i are convex, h is convex, and $\tilde{h}$ is nondecreasing in each argument (where $\tilde{h}$ is the extended-value extension function of h; please read through §3.2.4 for a discussion of why it is necessary for the extended-value extension to have the monotonicity property). We can easily show that $f'' \geq 0$ for differentiable g and h, and $n = 1$:

$$f''(x) = g'(x)^T \underbrace{\nabla^2 h(g(x))}_{\succeq 0} g'(x) + \underbrace{\nabla h(g(x))}_{\succeq 0}^T \underbrace{g''(x)}_{\succeq 0}.$$

Of course, this composition rule also holds for non-differentiable functions g_i and h. We sketch a proof using Jensen's inequality. Consider two points x and y in $\mathbf{dom}\, f$. Then

$$\begin{aligned} f(\theta x + (1-\theta)y) &= h(g(\theta x + (1-\theta)y)) \\ &\leq h(\theta g(x) + (1-\theta)g(y)) \\ &\leq \theta h(g(x)) + (1-\theta)h(g(y)) \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

The first inequality holds because each of the g_i is convex, and h is nondecreasing in each argument. The second inequality is because h is convex.

Using composition rules, we can ascertain convexity for composition functions. This allows us to recognize an even larger class space of functions to be convex.

Conjugate function

The conjugate of a function $f(x)$ is defined to be

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x)).$$

For $f : \mathbf{R} \to \mathbf{R}$, the interpretation of the conjugate is that for a given y, $f^*(y)$ is the maximum gap between the function $f(x)$ and the line xy.

Exercise 3.36(a): Derive the conjugate of the *max function*

$$f(x) = \max_{i=1,\ldots,n} x_i \text{ on } \mathbf{R}^n.$$

Solution (partial): We give some arguments which can be used toward finding the conjugate of the max function, and only consider the case where $n = 2$. The details and more general arguments are left for the reader to work out.

The first step is to determine the domain of the conjugate function $f^*(y)$ (*i.e.*, the values of y for which $y^T x - f(x)$ is bounded above). First, we consider y which has some negative entries. For example, suppose $y = (-1, 0)$. If we consider $x = -te_1$, then $y^T x - \max x_i = t - 0 \to \infty$ as $t \to \infty$. This gives a clue that for the conjugate to be bounded above, y must have all positive entries; this will need to be proved.

From the preceding argument, we can restrict the domain of the conjugate function to y such that $y \succeq 0$. Now consider the following vector $y = (0.7, 0.7)$. If we choose $x = t\mathbf{1}$, then $y^T x - \max x_i = t(\mathbf{1}^T y) - t = 1.4t - t \to \infty$ as $t \to \infty$. Therefore, $y = (0.7, 0.7)$ is not in $\mathbf{dom}\, f^*$. In fact, by considering $x = t\mathbf{1}$ and $y \succeq 0$, it must be true that $\mathbf{1}^T y = 1$ for

$y^T x - \max x_i$ to be bounded above. (You will need to show why this is the case, and also show why any vector y such that $\mathbf{1}^T y < 1$ or $\mathbf{1}^T y > 1$ will lead to the conjugate becoming unbounded above.)

For y such that $y \succeq 0$ and $\mathbf{1}^T y = 1$, it can be shown that

$$\sup_{x \in \mathbf{dom}\, f} (y^T x - \max_{i=1,\ldots,n} x_i) = 0$$

(why?). Then

$$f^*(y) = \begin{cases} 0 & \text{if } y \succeq 0 \text{ and } \mathbf{1}^T y = 1 \\ \infty & \text{otherwise} \end{cases}$$