

Convexity Reference

This document is just a reference for various properties of convexity, some of which may be useful for Homework 2. It's meant to make your life easier by not having to cross-reference things from various different sources.

Convex set Recall that a set C is convex if for all pairs of points x and y in C , and for any $\lambda \in [0, 1]$, the point $\lambda x + (1 - \lambda)y$ is also in C .

Function extensions When talking about convex functions and optimization, it is often convenient to consider the *extension* of a function to all of \mathbb{R}^n , where the function's value is ∞ on points outside its domain. Thus a given function $f : A \rightarrow \mathbb{R}$, where $A \subseteq \mathbb{R}^n$, can be extended to $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ as

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \text{dom } f \\ \infty & x \notin \text{dom } f \end{cases}$$

where $\text{dom } f$ is the domain of f . In this document (and on the homework), we will assume that all functions have been thus extended. This allows us to simplify our notation. Note that minimizing \tilde{f} is identical to minimizing f . The advantage comes from allowing us to say, for example, simply $h = f + g$ rather than $h(x) = f(x) + g(x)$ if $x \in \text{dom } f \cap \text{dom } g$. We now define $\text{dom } \tilde{f}$ to be $\{x : \tilde{f}(x) < \infty\}$ (so that dom is not quite the traditional definition of domain).

Convex functions The *epigraph* is the region of a plot of f that lies above the function:

$$\text{epi } f = \{(x, t) \mid x \in \text{dom } f, t \geq f(x)\}.$$

We call a function convex if its epigraph is a convex set.

This condition is equivalent to the statement that chords (line segments connecting the points $(x, f(x))$ and $(y, f(y))$) lie above the function:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{for all } \lambda \in [0, 1]. \quad (1)$$

Make sure you understand why this is true.

One reason we care about convex functions is that any local minima must also be global minima (why?). Note that the minimum may not be unique, but if there are multiple minima they must form a convex set.

Calculus of convex functions All of the following functions are convex:

- (a) $\alpha f(x)$ for convex f and $\alpha > 0$.
- (b) $f(Ax + b)$ for convex f , matrix A , and vector b .
- (c) $f(x) + g(x)$ for convex f and g .
- (d) $\max(f(x), g(x))$ for convex f and g .
- (e) $g(f(x))$ for convex f with range $A \subseteq \mathbb{R}$ and $g : A \rightarrow \mathbb{R}$ both convex and non-decreasing.
- (f) $\sup_{z \in C} f(x, z)$ for $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ convex in x , C a convex subset of $\text{dom } f$.

Convexity Reference

First-order condition A function f differentiable on the interior of its domain $\text{int dom } f$ is convex iff

$$\text{dom}(f) \text{ is convex and } f(y) - f(x) \geq \nabla f(x)^T (y - x) \text{ for all } x, y \in \text{int dom}(f) \quad (2)$$

where $\nabla f(x)^T$ is the gradient as a row vector, $\left[\frac{\partial f}{\partial x_1}(x) \quad \dots \quad \frac{\partial f}{\partial x_n}(x) \right]$.

This is equivalent to the condition that tangent planes lie below (or on) the function; make sure you understand why this is true.

Second-order condition There is also a second-order condition for functions f twice-differentiable on the interior of their domains:

$$\text{dom}(f) \text{ is convex and } \nabla^2 f(x) \succeq 0 \text{ for all } x \in \text{int dom}(f) \quad (3)$$

where $A \succeq 0$ means the matrix A is positive semidefinite and $\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$ is

the Hessian at x .

Strict convexity A *strictly convex* function has tangent planes strictly less than the function (at all points other than where they are tangent). The equivalent of the conditions (1) (2) (3) are

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) \quad \text{for all } x \neq y \in \text{dom } f, \lambda \in (0, 1) \quad (4)$$

$$\text{dom}(f) \text{ is convex and } f(y) - f(x) > \nabla f(x)^T (y - x) \text{ for all } x \neq y \in \text{int dom}(f) \quad (5)$$

$$\text{dom}(f) \text{ is convex and } \nabla^2 f(x) \succ 0 \text{ for all } x \in \text{int dom } f \quad (6)$$

The notation $\nabla^2 f(x) \succ 0$ means that $\nabla^2 f(x)$ is (strictly) positive definite. Note that (6) is a sufficient condition but is not necessary. (Can you think of a counterexample?) The other conditions are both sufficient and necessary (when applicable).

A strictly convex function has a unique minimum.

Strong convexity An *m-strongly convex* function is lower-bounded by tangent parabolas with a fixed curvature coefficient $m > 0$. The conditions become:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}m\lambda(1 - \lambda)\|x - y\|_2^2 \quad \text{for all } x, y \in \text{dom } f, \lambda \in [0, 1] \quad (7)$$

$$\text{dom}(f) \text{ is convex and } f(y) - f(x) \geq \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|_2^2 \text{ for all } x, y \in \text{int dom}(f) \quad (8)$$

$$\text{dom}(f) \text{ is convex and } \nabla^2 f(x) \succeq mI \text{ for all } x \in \text{int dom } f \quad (9)$$

$A \succeq B$ means that $A - B \succeq 0$. $A \succeq mI$ is equivalent to the smallest eigenvalue of A being at least m (why?). Each of these conditions is both sufficient and necessary (when applicable).

Strong convexity is useful in proving that some optimization algorithms converge, as we saw in class.