10-725: Optimization **Fall 2012**

Lecture 3: September 4

Lecturer: Geoff Gordon *Scribes: Adona Iosif, Xiaowen Ding, Yue Yu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

In this lecture we look in more detail at *convexity*. We cover the definition and properties of convex sets and functions, and provide a toolkit of techniques to prove convexity. We will later use these techniques to show that particular optimization problems are convex, and can thus be efficiently solved using standard methods such as gradient and subgradient descent.

3.1 Convex Sets

Definition 3.1 *A set C is convex if the line segment between any two points in C lies in C, i.e. $\forall x_1, x_2 \in C, \forall \theta \in [0,1]$*

$$\theta x_1 + (1-\theta)x_2 \in C.$$

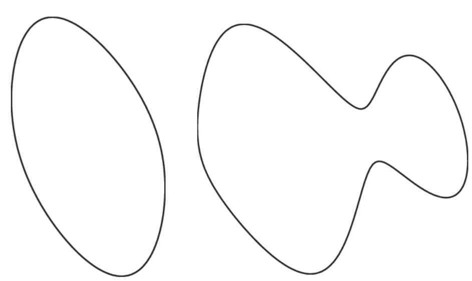

Figure 3.1: Example of a convex set (left) and a non-convex set (right).

Simple examples of convex sets are:

- The empty set $\emptyset$, the singleton set $\{x_0\}$, and the complete space $\mathbb{R}^n$;
- Lines $\{a^T x = b\}$, line segments, hyperplanes $\{A^T x = b\}$, and halfspaces $\{A^T x \leq b\}$;
- Euclidian balls $B(x_0, \epsilon) = \{x \mid ||x - x_0||_2 \leq \epsilon\}$.

We can generalize the definition of a convex set above from two points to any number of points n. A convex combination of points $x_1, x_2, ..., x_k \in C$ is any point of form $\theta_1 x_1 + \theta_2 x_2 + ... + \theta_k x_k$, where $\theta_i \geq 0, i = 1...k, \sum_{i=1}^{k} \theta_i = 1$. Then, a set C is convex iff any convex combination of points in C is in C.

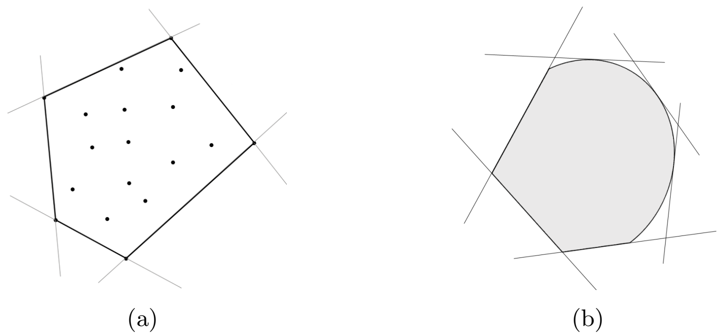

Figure 3.2: (a) Representation of a convex set as the convex hull of a set of points. (b) Representation of a convex set as the intersection of a (possibly infinite) number of halfspaces.

We can take this even further to *infinite countable sums*: C convex iff $\forall x_i \in C, \theta_i \geq 0, i = 1, 2.., \sum_{i=1}^{\infty} \theta_i = 1$:

$$\sum_{i=1}^{\infty} \theta_i x_i \in C$$

if the series converges.

Most generally, C is convex iff for any *random variable* X over C, $\mathbb{P}(X \in C) = 1$, its expectation is also in C:

$$\mathbb{E}(X) \in C.$$

3.1.1 Alternative Representations

In different contexts, different representations of a convex set may be natural or useful. In the following sections we introduce the *convex hull* and *intersection of halfspaces* representations, which can be used to show that a set is convex, or prove general properties about convex sets.

3.1.1.1 Convex Hull

Definition 3.2 *The convex hull of a set C is the set of all convex combinations of points in C:*

$$\boldsymbol{conv}(C) = \{\theta_1 x_1 + ... + \theta_k x_k | x_i \in C, \theta_i \geq 0, i = 1, ...k, \sum_{i=1}^{k} \theta_i = 1\}$$

The convex hull of a set C is the smallest convex set which includes C:

- **conv(C)** is convex
- $C \subseteq$ **conv(C)**
- $\forall C'$, C' convex, $C \subseteq C' \Rightarrow$ **conv(C)** $\subseteq C'$

Lemma 3.3 *Any closed convex set C can be written as the convex hull of a possibly infinite set of points X:*

$$C = hull(X)$$

Indeed, any closed convex set is the convex hull of itself. However, we may be able to find a set X of much smaller dimensionality than C, such that we still have $C = hull(X)$. (See Figure 3.2a)

3.1.1.2 Intersection of Halfspaces

Lemma 3.4 *Any closed convex set C can be written as the possibly infinite intersection of a set of halfplanes:*

$$C = \cap_i \{x | a_i x + b_i \leq 0\}$$

Indeed, any closed convex set is the intersection of all halfspaces that contain it:

$$C = \cap\{H | H halfspaces, C \subseteq H\}.$$

However, we may be able to find a much smaller set of halfspaces such that the representation still holds. (See Figure 3.2b)

3.1.2 Convexity preserving operations

A standard way to prove that a set (or later, a function) is convex is to build it up from simple sets for which convexity is known, by using convexity preserving operations. We present some of the basic operations below:

- **Intersection** If C, D are convex sets, then $C \cap D$ is also convex.
- **Affine transform** If C is a convex set, $C \subseteq \mathbb{R}^n$, $A \in \mathbb{R}^{mxn}$, $b \in \mathbb{R}^m$, then

  $$AC + b = \{Ax + b | x \in C\} \subseteq \mathbb{R}^m$$

  is also convex. In particular, the following operations also preserve convexity:

  - **Translation** $C + b$
  - **Scaling** αC
  - **Orthogonal projection** $T = \{x_1 \in \mathbb{R}^{n_1} | (x_1, x_2) \in C \text{ for some } x_2 \in \mathbb{R}^{n_2}\}$
  - **Set sum** $C_1 + C_2 = \{c_1 + c_2 | c_1 \in C_1, c_2 \in C_2\}$
    Indeed, the set sum $C_1 + C_2$ is the image of the carthesian product $C_1 \times C_2$ through the affine transform $f(x) = f(x_1, x_2) = x_1 + x_2 = (1, 1) \cdot x$.
- **Perspective transform** If $C \in \mathbb{R}^n \times \mathbb{R}_{++}$ is a convex set, then the perspective transform $P(C)$ with

  $$P(x) = P(x_1, x_2, ..., x_n, t) = (x_1/t, x_2/t, ..., x_n/t) \in \mathbb{R}^n$$

  is also convex.

We can prove convexity preservation under intersection and affine transforms trivially from the definition of a convex set. For perspective transforms we show that segments between any two points $x, y \in C$ are mapped to segments between the transforms of the points $P(x), P(y) \in P(C)$, and thus any line segment between two points in $P(C)$ is also in $P(C)$. For more details and examples see [Boyd2004], pg 35.

An example of operation that does *not* preserve convexity is set union.

3.1.3 Proving a set is convex

To recap, there are multiple different ways to prove that a set C is convex. Some of the most common ones we've seen are:

- Using the definition of a convex set
- Writing C as the convex hull of a set of points X, or the intersection of a set of halfspaces
- Building it up from convex sets using convexity preserving operations

3.1.4 Separating and supporting hyperplane theorems

An important idea that we will use later in the analysis of convex optimization problems is the use of hyperplanes or affine functions to separate convex sets that do not intersect. The two main results are the *separating hyperplane theorem* and the *supporting hyperplane theorem.*

3.1.4.1 Separating hyperplane theorem

The separating hyperplane theorem states that, for any two convex sets C and D which do not intersect, $C \cap D = \emptyset$, there exists a hyperplane such that C and D are on opposite sides of the hyperplane (See Figure 3.3). More rigurously:

Theorem 3.5 *Let C, D be convex sets such that $C \cap D = \emptyset$. Then there exist $a \neq 0$, b such that:*

$$\forall x \in C, a^T x \leq b \text{ and } \forall y \in D, a^T y \geq b.$$

The separating hyperplane theorem is geometrically intuitive. [Boyd2004], pg 45, goes through the proof for the special case of two sets C and D which are a finite distance $dist(C, D) = min_{c \in C, d \in D} ||c - d||_2$ apart, and for which there exist two points $c \in C, d \in D$ which actually achieve this minimum. In this case it can be shown that any hyperplane perpendicular on the segment $[c, d]$, which goes through the segment $[c, d]$, is a separating hyperplane. In particular, the hyperplane through $\frac{c+d}{2}$ is a separating hyperplane. The proof is a proof by contradiction, which essentially shows that, if one of the sets (say D) would intersect our proposed hyperplane, than we could find another d' closer to c than d.

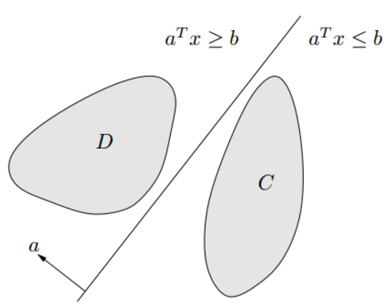

Figure 3.3: The hyperplane $x | a^T x = b$ separates the disjoint convex sets C and D. The affine function $a^T x - b$ is non-positive on C and non-negative on D.

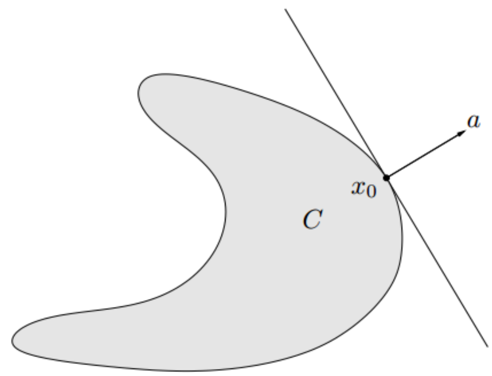

Figure 3.4: The hyperplane $x|a^Tx = a^tx_0$ supports C at x_0.

A related concept is *strict separation*, where the separating hyperplane does not intersect either C or D:

$$\forall x \in C, a^Tx < b \text{ and } \forall y \in D, a^Ty > b.$$

In general, not any two convex sets C and D have a strictly separating hyperplane. A sufficient (but not necessary) condition for this to hold is that:

- both C and D are closed
- at least one of the sets is compact

A subtle example where strict separation does not hold is inbetween the halfspace $y \leq 0$ and the set above the function $\frac{1}{x}$, $y \geq \frac{1}{x}$. The two set boundaries converge towards each other, but never intersect. Thus, we can find a separating hyperplane, $y = 0$, but not a strictly separating hyperplane. In this case, the two sets were closed, but neither of them was compact.

3.1.4.2 Supporting hyperplane theorem

A related result is the *supporting hyperplane theorem*. A supporting hyperplane of the set C at a boundary point $x_0 \in$ **bd C** = **cl C - int C** is any hyperplane $\{x|a^Tx = b, a \neq 0\}$ such that the entire convex set lies on one side of the hyperplane:

$$\forall x \in C, a^Tx \leq a^Tx_0.$$

Theorem 3.6 *For any convex set C and any boundary point $x_0 \in$ **bd C** there exists a supporting hyperplane for C at x_0.*

The proof of the theorem is trivial, and can be found in [Boyd2004], pg 50. The main idea behind it is to use the *separating hyerplane theorem* applied to **int C** and x_0, and show that the resulting separating hyperplane at x_0 is a supporting hyperplane for C (See image 3.4).

In Homework 1, we proved a partial converse to the supporting hyperplane theorem: If a set is closed, has non-empty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.

Next we look at convex functions.

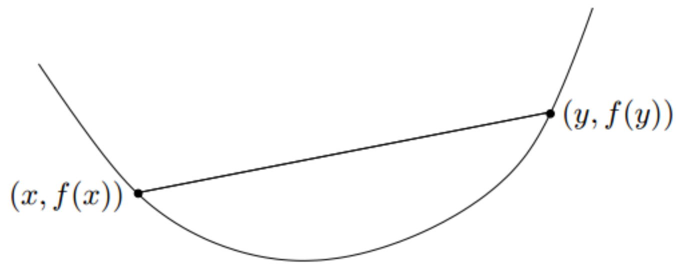

Figure 3.5: Graph of a convex function. The line segment betwen any two points on the graph lies above the graph.

3.2 Convex Functions

A *convex function* is a function defined on a convex domain such that, for any two points in the domain, the segment between the two points lies above the function curve between them (See figure 3.5). We will show below that this definition is closely connected to the concept of a convex set: a function f is convex if and only if its *epigraph*, the set of all points above the function graph, is a convex set. We restate these results more precisely:

Definition 3.7 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom(f) is a convex set and if $\forall x, y \in dom(f), \forall \theta \in [0, 1]$, we have:*

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set of points $epi(f) = \{(x, t)|x \in dom(f), t \geq f(x)\}$.

Lemma 3.8 *The function f is convex iff the set epi(f) is convex.*

3.2.1 Criteria for convexity

As with sets, there are multiple ways to characterize a convex function, each of which may by convenient or insightful in different contexts. Below we explain the most commonly used three criteriea.

3.2.1.1 Zeroth order

Theorem 3.9 *A function is convex iff it is convex when restricted to any line that intersects its domain. In other words, f is convex iff $\forall x \in dom(f), \forall v$, the function $g(t) = f(x + tv)$ is convex, when restricted to its domain $\{t|x + tv \in dom(f)\}$.*

This property is useful because it allows us to reduce the problem of checking the convexity of a multivariate function to checking the convexity of a uni-variate function, for which we can use much simpler criteria.

We can intuitively visualize why the property holds by imagining e.g. a 3D convex cup-shaped function, choosing any point (x_0, y_0) on the function graph, and taking a vertical slice through (x_0, y_0). The resulting plane intersects the domain of f on a line, $x_0 + tv$, and generates a new 2D function $g(t)$. Intuitively, we expect this slice function to be convex (in this case, a 2D cup), and indeed this is what the zeroth-order property states.

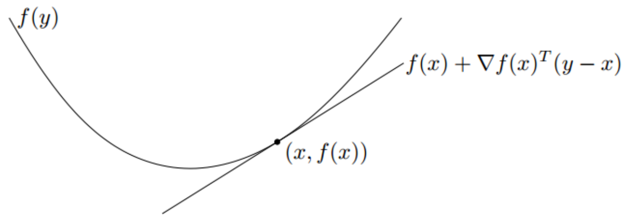

Figure 3.6: If f is convex and differentiable, then $f(x) + \nabla f(x)^T(y - s) \le f(y)$ for all $x, y \in dom(f)$

3.2.1.2 First order

Theorem 3.10 *Let f be a differentiable function. Then f is convex iff dom(f) is a convex set and:*

$$f(y) \ge f(x) + \nabla f(x)^T(y - x).$$

[Shamelessly reproduced from [Boyd2004], pg 69, since it explains it so well:]

The function $g(y) = f(x)+\nabla f(x)^T(y-s)$ is the first-order Taylor approximation of f near x. The inequality above states that for a convex function, the first-order Taylor approximation is in fact a *global underestimator* of the function (See Figure 3.6). Conversely, if the first-order Taylor approximation of a function is always a global underestimator, then the function is convex.

The inequality above shows that from *local information* about a convex function (*i.e.* its value and derivative at a point) we can derive *global information* (*i.e.* a global underestimator of it). This is perhaps the most important property of convex functions and convex optimization problems, and begins to explain why convex optimization problems are so much more easy to solve than non-convex optimizatino problems. As one simple example, the inequality above shows that if $\nabla f(x) = 0$ then for all $y \in dom(f)$, $f(y) \ge f(x)$, so x is a global minimizer of f. [end of reproduction]

[Boyd2004], pg 70, gives a neat proof of the first-order criterion by first showing it holds for the $n = 1$ case, and then using the zero-th order criterion for the case $n > 1$ to reduce the problem back to $n = 1$. An additional intuition for why Theorem 3.10 holds is that the term $g(x) = f(x) + \nabla f(x)^T(y - x)$ represents a hyperplane through x which has $\nabla f(x)$ slopes with respect to each axis of the input, and which *supports the epigraph of* f. We expect such a hyperplane exists because, since f is convex, so is $epi(f)$. Thus, the first-order criterion is just an analytical expression of the supporting hyperplane theorem for convex sets!

3.2.1.3 Second order

If f is twice differentiable, we have a simple second-order criterion, which is the most common way to show well behaved functions are convex:

Theorem 3.11 *Let f be a twice differentiable function on an open domain $dom(f)$. The f is convex iff $dom(f)$ is convex and its Hessian is positive semidefinite:*

$$H(x) = \nabla^2 f(x) \succeq 0.$$

[Remember that a matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefinite, $A \succeq 0$, iff $\forall x \in \mathbb{R}^n, x^T AX \ge 0$.]

For a function on $\mathbb{R}$, this reduces to the simple condition $f''(x) \ge 0$, which means that the first derivative $f'(x)$ is non-decreasing.

3.2.2 Convexity preserving operations

Just as for convex sets, we consider standard operations which perserve function convexity. These are useful tools for proving a complicated function is convex by showing it is constructed from simpler convex functions using legal operations. If f, f_i are convex functions, then the following functions are also convex:

- **Non-negative weighted sum** $\forall \omega_1, ...\omega_k \geq 0, \sum_{i=1}^{k} \omega_i f_i$ is convex
- **Composition with affine mapping** $f(Ax + b)$ is convex
- **Composition with monotone convex** $g(f(x))$ is convex for f convex, g convex and non-decreasing.

  We can see this easily in the case $f : \mathbb{R} \to \mathbb{R}$:

  $$h''(x) = [g(f(x))]'' = [g'(f(x))f'(x)]' = g''(f(x))[f'(x)]^2 + g'(f(x))f''(x) \geq 0$$

  If f, g are convex than $g''(f(x))$ and $f''(x)$ are both ≥ 0. Further, if g is non-decreasing, $g'(f(x))$ is also ≥ 0, making all the terms in $h''(x) \geq 0$.

  In this form we can also easily see other combinations of properties of f and g which lead to known properties of h. For example, if f is concave, g convex and non-increasing, then h is again convex.

- **Pointwise maximum** $f(x) = \max_i f_i(x)$ is convex

  This follows directly from the fact that the epigraph of a max/sup of functions is the intersection of the epigraphs of the functions, and that set convexity is preserved under intersection (See Figure 3.7).

- **Minimum over a convex set** $g(x) = inf_{y \in C} f(x, y)$ is convex

  We can also get an intuition for this result by looking at it in terms of epigraphs. Indeed, assuming that all infimums are actually achieved, for any x, $g(x) = f(x, y)$ for some $y \in C$. Thus, (x, t) is in the epigraph of g iff (x, y, t) is in the epigraph of f for some $y \in C$. In other words:

  $$\textbf{epi } g = \{(x, t) | (x, y, t) \in \textbf{epi} f, y \in C\}$$

  and the epigraphf of g is the projection of the epigraph of f on one of its components. Since we've shown that projection on components preserves set convexity, we must have **epi** g convex, and thus the function g is also convex.

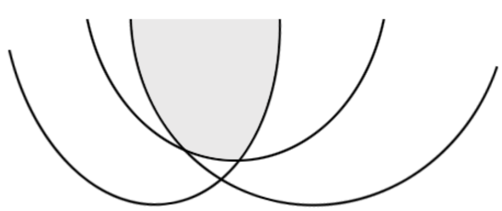

Figure 3.7: The point-wise maximum of convex functions is convex.

3.2.3 Proving a function is convex

Already in the previous section, we've used three different techniques (!) for proving function convexity: the definition of a convex function, for non-negative weighted sums and composition with affine mapping; the 2^{nd} order criteria, for composition with monotone convex; and the connection with convex sets, for maximum and minimum. Let's review all the different ways one can prove a function is convex:

- Using the definition of a convex function
- Showing that the function's epigraph is a convex set
- Using one of the $0^{th}, 1^{st}$ or 2^{nd} order criteria:
  - 0^{th} order: $g(t) = f(x + tv)$ is a convex function of t for all x, v
  - 1^{th} order: $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ for all x, y
  - 2^{th} order: $\nabla^2 f(y) \geq 0$ for all x
- Building it up from convex functions using convexity preserving operations

And that's all, folks!

References

[Boyd2004] S. Boyd and L. Vandenberghe, "Convex optimization" *Cambridge University Press*, 2004