Lecture 4: Convexity

Lecturer: Barnabás Póczos   *Scribes: Jessica Chemali, David Fouhey, Yuxiong Wang*

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

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

The goals of this lecture are to:

1. give definitions that are important to convexity as well as examples of convex sets and basic properties;

2. define convex functions and their properties, as well as some examples.

4.1   Basic Definitions

We begin by formalizing a few mathematical objects that we will use throughout the lecture:

Definition 4.1 *A line passing through x_1 and $x_2 \in \mathbb{R}^n$ forms the set $\{x \in \mathbb{R}^n | x = \theta x_1 + (1-\theta)x_2, \theta \in \mathbb{R}\}$.*

Definition 4.2 *The line segment passing through x_1 and x_2 is defined similarly, but with θ confined to $[0,1]$, or $\{x \in \mathbb{R}^n | x = \theta x_1 + (1-\theta)x_2, \theta \in [0,1]\}$.*
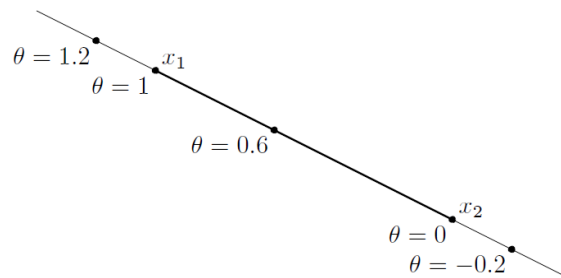

Figure 4.1: Line and line segment.

Definition 4.3 *A Euclidean ball with radius r centered at x_0, $B(x_o, r)$, is: $\{x \in \mathbb{R}^n | \|x - x_0\|_2 \le r\}$*

Definition 4.4 *An L_p ball is the equivalent, but for the L_p distance, or $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ $\{x \in \mathbb{R}^n | \|x - x_0\|_p \le r\}$*

Definition 4.5 *A half-space $HS(a, b)$ is a set of points on one side of a hyperplane: $\{x \in \mathbb{R}^n | a^T x \le b\}$.*
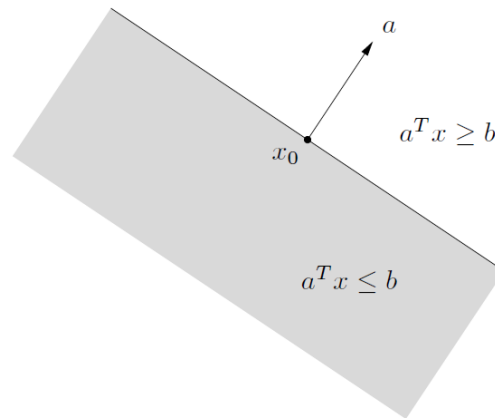
Figure 4.2: Halfspace.

Definition 4.6 *A hyperplane $HS(a, b)$ is defined similarly, but with equality, or $\{x \in \mathbb{R}^n | a^T x = b\}$* .
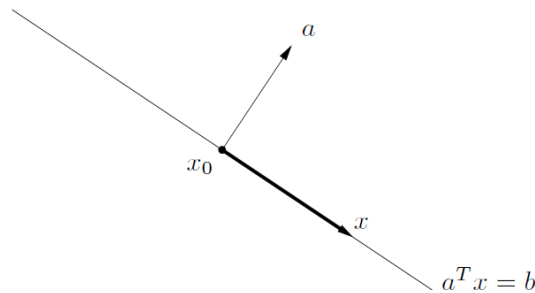

Figure 4.3: Hyperplane.

Definition 4.7 *A cone is any set such that if $x \in C$ and $\theta \geq 0$ then $\theta x \in C$.*

We can immediately use these to give a definition of an affine set. Specifically, a set C is affine if for any points in it, the line through them is also contained in C. For example, the affine hull of a circle in $\mathbb{R}^2$ is $\mathbb{R}^2$ and the affine hull of a line in $\mathbb{R}^2$ is $\mathbb{R}$. Written more formally, for $x_1, x_2 \in C$ and $\theta \in C$, $\theta x_1 + (1-\theta)x_2 \in C$. This definition also leads to the definition of the affine hull of a set, which is the smallest affine set containing C. The affine hull can be written out explicitly as:

$$\text{Aff}[C] = \left\{ \theta_1 x_1 + \cdots + \theta_k x_k | x_i \in C, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

Similarly, a conic hull is

$$\text{Cone}[C] = \{x | x = \theta_1 x_1 + \ldots + \theta_k x_k, \theta_i \geq 0, x_i \in C\}$$
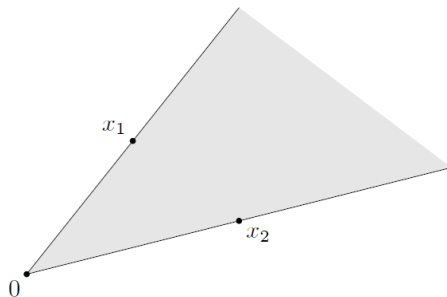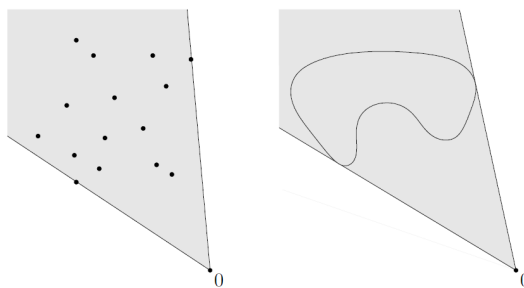
Figure 4.4: Cone.

Figure 4.5: Conic Hull.

Before we proceed with the remaining definitions, we also need to give a few topological definitions. Given a set C,

Definition 4.8 *We say that x is on the boundary of C, or ∂C, if for some small enough $\epsilon > 0$, the epsilon-radius ball centered at x covers both inside and outside the set, or $B(x, \epsilon) \cap C \neq \emptyset$ and $B(x, \epsilon) \cap C^c \neq \emptyset$.*

Definition 4.9 *We say that x is in the interior of C if $\exists\, \epsilon > 0$ s.t. $B(x, \epsilon) \subset C$.*

Definition 4.10 *We say that the relative interior of C, rel int C is the same as the interior, but with the ball restricted to the affine hull of C. Specifically, x is in rel int C if $B(x, \epsilon) \cap Aff[C] \subset C$. This definition is more natural in some ways. Consider a line segment in a 2D-space. No points in the space are in the interior the line segment since any epsilon ball contains points off the line; however, if we restrict our consideration to the affine hull of the line segment (the equivalent line), some points are.*

We can then use set operations to take these and generate more definitions and properties of C:

1. The closure of C, cl C, is $C \cup \partial C$.

2. The relative boundary of C, rel∂C, is cl C\rel int C.

3. C is closed if $\partial C \subset C$

4. C is open if $\partial C \cap C = \emptyset$

5. C is compact if it is closed and bounded in $\mathbb{R}^n$.

4.2   Convex sets

4.2.1   Definitions

Definition 4.11 *A set $C \subset R^n$ is convex if for any two points in C, the line segment joining them is contained in C. Formally, it is convex if and only if for all $x_1, x_2 \in C$ and $\theta \in [0, 1]$, $\theta x_1 + (1 - \theta)x_2 \in C$.*
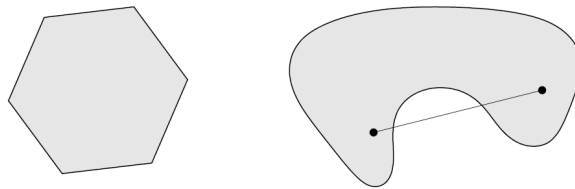

Figure 4.6: Convex sets.

Definition 4.12 *A convex set is strictly convex if for any two points in the set in general position, the line segment less the endpoints is contained in int C. Formally, if for all $x_1, x_2 \in C$ with $x_1 \neq x_2$ and $\theta \in (0, 1)$ $\theta x_1 + (1 - \theta)x_2 \in int\ C$.*

Definition 4.13 *The convex hull of a set C is the set of all convex combinations of the points in C*

$$conv[C] = \{\theta_1 x_1 + \cdots + \theta_k x_k | x_i \in C, \theta_i \geq 0\ \forall i = 1, \ldots, k, \sum_{i=1}^{k} \theta_i = 1, k \in \mathbb{Z}^+\}$$

Here, we are constraining θ_i in both ways that we did before: $\theta_i \geq 0$ as in the conic hull, and $\sum_{i=1}^{k} \theta_i = 1$ as in the affine hull. Just as in the affine hull case, the $conv[C]$ is the smallest convex set that also contains C. Note that: $conv[C]$ is convex and $C \subset conv[C]$.
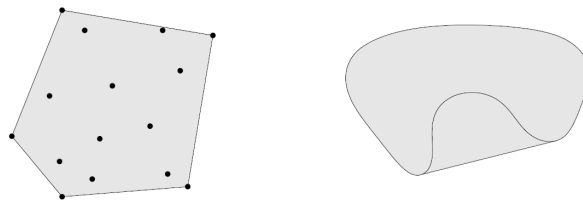

Figure 4.7: Convex hull.

4.2.2 Examples

Example 4.14 *Many of the objects defined earlier are convex sets: lines, line segments, hyperplanes and half spaces. Additionally, the empty set $\emptyset$ and singleton sets $\{x\}$ are convex, as are complete spaces $\mathbb{R}^d$. L_p Balls are convex for $p \geq 1$, but are not for $0 < p < 1$.*

Example 4.15 *A more complex example of a convex set is a polyhedron. This, at least in this course, is a solution set to a finite set number of linear equalities and inequalities (i.e., $\{x | Ax \leq b, Cx = d\}$. If the solution set is bounded, we call it a polytope. As we will see later, this is easy to show as convex, as it is a an intersection of halfspaces (from $Ax \leq b$) and hyperplanes (from $Cx = d$), each of which are convex.*

Example 4.16 *One can also construct convex sets via the convex combination of points. The definition of the convex hull gives it for a finite set of points. One can also define it for an infinite sum: given $x_1, x_2, \ldots \in C$ and $\theta_1, \theta_2, \ldots \geq 0$ such that $\sum_{i=1}^{\infty} \theta_i = 1$, then $\sum_{i=1}^{\infty} \theta_i x_i \in C$ so long as the series converges. Similarly, one can use a density function p, and $\int_C p(x)xdx \in C$ if the integral exists.*

Example 4.17 *We can extend this to more complex domains, such as matrices. For instance, the set of $n \times n$ positive semi-definite (PSD) matrices form a convex cone which we denote S_+^n. Recall that a matrix A is PSD if and only if $A = A^T$ and $x^T A x \geq 0$ for any $x \in \mathbb{R}^n$; equivalently a symmetric matrix is PSD if all its eigenvalues are positive. This may also be written as $A \geq 0$, which suggests a partial ordering defined by the PSD property, namely, $M \geq N$ if and only if $M - N$ is PSD. We can prove convexity by linearity: given PSD matrices A and B, $x^T[\theta A + (1 - \theta)B]x] = \theta x^T A x + (1 - \theta)x^T B x \geq 0$ as it is a convex sum of non-negative terms (since A and B are PSD).*

4.2.3 Representations

We can represent a convex set in two equivalent ways.

Definition 4.18 *The primal representation represents a convex set C using its convex hull: a convex combination of its points. The number of points might be finite (like in the case of a polyhedron) or infinite (like the case of a circle).*

Definition 4.19 *The dual representation represents a closed set C as the intersection of all the closed halfspaces containing it.*

4.2.4 Convexity Preserving Operations

One easy way to show that a set is convex is to construct it from convex sets via convexity preserving operations. Here are a few. Given convex sets $C, D \subset \mathbb{R}^n$, $b \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$, $\alpha \in \mathbb{R}$, the following sets are convex:

1. Translation: $C + b = \{x + b : x \in C\}$

2. Scaling: $\alpha C = \{\alpha x : x \in C\}$

3. Intersection: $C \cap D$; this can also be extended to an infinite number of sets.
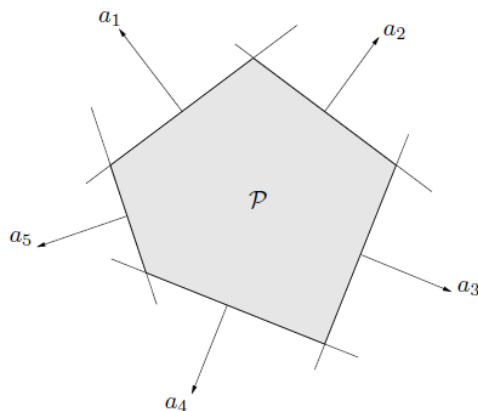
Figure 4.8: Dual representation as Intersection of hyperplanes.

4. Affine: $AC + b = \{Ax + b | x \in C\} \subset \mathbb{R}^m$

5. Set sum: $C + D = \{x + y | x \in C, y \in D\}$.

6. Direct sum: $C \times D = \{(x, y) \in \mathbb{R}^{n+m}, x \in C, y \in D\}$

A more involved operation is perspective projection: suppose $C \subset \mathbb{R}^n \times \mathbb{R}_{++}$ is convex, then $P(C)$ is also convex, where P is perspective projection, or:

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

This can be extended further to linear-fractional functions.

Theorem: image of linear-fractional functions. Let $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$, $d \in \mathbb{R}$. Then, define f as:

$$f(x) = \frac{Ax + b}{c^T x + d}, \qquad \text{dom } f = \{x | c^T x + d > 0\}$$

Then $f(X)$ is also convex.

Example 4.20 *Let $C = \{p_{ij} | u = u_i \in [0, 1], v = v_j \in [0, 1], i \in [1, m], j \in [1, n]\}$. The set C represents the joint probability distribution of discrete variables u and v. Let*

$$D = \{f_{ij} | i \in [1, m], j \in [1, n]\}, where$$

$$f_{ij} = Pr(u = u_i | v = v_j) = \frac{p_{ij}}{\sum_i p_{ij}}$$

By the linear-fractionals theorem above: if C is convex, then D is also convex.

4.2.5 Separating hyperplanes

Theorem: Let C and D be two convex sets such that $C \cap D = \emptyset$. Then $\exists \, a^T \in \mathbb{R}^n$, $b \in \mathbb{R}$, s.t

$$\forall x \in C \; a^T x \leq b,$$
$$\forall x \in D \; a^T x \geq b.$$

Strict inequalities do not always hold even when C and D do not intersect. For example, when points of both C and D are points of the separating hyperplane.
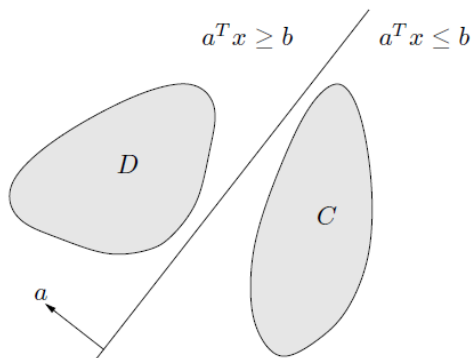

Figure 4.9: Separating hyperplane.

Definition 4.21 *Strong Separation: Let C and D be two disjoint convex sets. We say that they are stongly separated if they can be shifted by a small amount and stay separated by a hyperplane $H(a, b)$. Formally:*

$$\exists \ \epsilon > 0 \ \ s.t. \ a^T(C + B(0, \epsilon)) > b \ \ and$$
$$a^T(D + B(0, \epsilon)) < b.$$

Definition 4.22 *Proper Separation: C and D are properly separated, if they are separated by hyperplane $H(a, b)$ and it is not the case that $C \subset H(a, b)$ and $D \subset H(a, b)$.*

Definition 4.23 *Strict Separation: C and D are strictly separated by hyperplane $H(a,b)$ if $a^T x < b$ and $a^T y > b$ $\forall x \in C$ and $y \in D$.*

Theorem 4.24 *Strong Separation Theorem: If C and D are non-empty convex sets in $\mathbb{R}^n$, cl $C \cap$ cl $D = \emptyset$, and one of them is bounded, then there is a hyperplane that separates them strongly.*

Example 4.25 *Why do we need boundedness? We can construct one example to show that without the boundedness property of at least one of the sets, there's no strong separation. Let C be the epigraph of $f(x) = \frac{1}{x}$ and D be the set of all points below the line $y = 0$ in the 2D euclidean space. Then because $f(x)$ tends to 0, the line $y = 0$ does not strongly separate the two sets.*

Theorem 4.26 *Another version of the theorem is that C and D are strongly separated iff the distance between any two points $x \in C$ and $y \in D$ is greater than 0. In the example above the distance between C and D decreases to 0 as x tends to infinity.*

Theorem 4.27 *Supporting plane theorem: For any point x_0 at the boundary of a convex set, $\exists$ a hyperplane that lies entirely on one side of the set.*

There is a very useful partial converse for this theorem: If a set C is closed and has a non-empty interior, and $\exists$ a supporting hyperplane for all its boundary points, then C is convex.
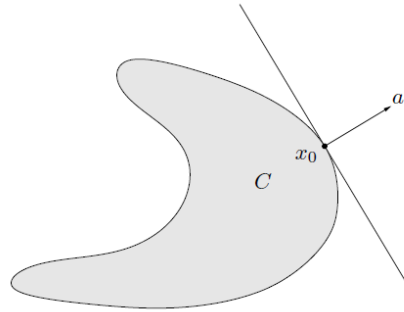
Figure 4.10: Supporting hyperplane.

4.2.6   How to prove that a set is convex

- Use the definition.

- Represent it as a convex hull.

- Represent it as an intersection of halfspaces.

- Use the supporting hyperplane converse theorem.

- Construct C from convex sets using convexity-preserving operations.

4.3   Convex Function

4.3.1   Definitions

Definition 4.28 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\operatorname{dom} f$ is a convex set and if for all $x, y \in \operatorname{dom} f$, and θ with $0 \le \theta \le 1$, we have*

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

Definition 4.29 *A function f is strictly convex if whenever $x \ne y$, and $0 < \theta < 1$, strict inequality holds, that is, we have*

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

Example 4.30 x^4 *is strictly convex.*

Definition 4.31 f *is concave if $-f$ is convex.*

The geometric interpretation of convex functions is shown in Fig. **??**. The chord from x to y (*i.e.*, line segment) between any two points on the graph lies above the graph of f.
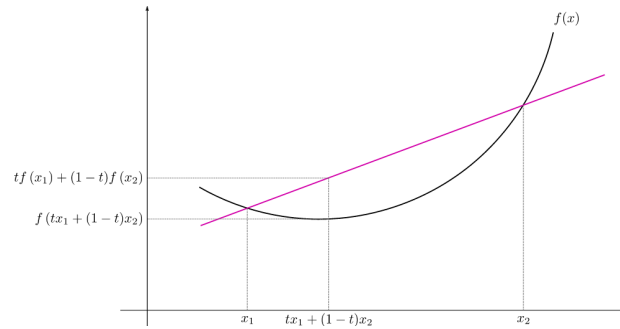
Figure 4.11: Graph of a convex function

4.3.2 Strongly Convex Function

Definition 4.32 *A differentiable function f is called $m-$strongly convex if $m > 0$ and*

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq m \|x - y\|_2^2, \forall x, y \in \mathrm{dom} f$$

An equivalent condition is

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2} \|y - x\|_2^2, \forall x, y \in \mathrm{dom} f$$

It is not necessary for a function to be differentiable. We could have the definition without gradient.

Definition 4.33 *A function f is called $m-$strongly convex if $m > 0$ and for $0 \leq t \leq 1$*

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \frac{1}{2}mt(1 - t) \|x - y\|_2^2, \forall x, y \in \mathrm{dom} f$$

If the function is twice continuously differentiable, we could have the definition with Hessian matrix.

Definition 4.34 *f is called $m-$stronly convex if $m > 0$ and*

$$\nabla^2 f(x) \geq mI, \forall x, y \in \mathrm{dom} f$$

A strongly convex function is also strictly convex, but not vice-versa.

Example 4.35 *Convex functions*

- $|x|^p, p \geq 1, \forall x \in \mathbb{R}.$
- *Max function:* $f(x) = max(x_1, ..., x_n), \forall x \in \mathbb{R}^n.$
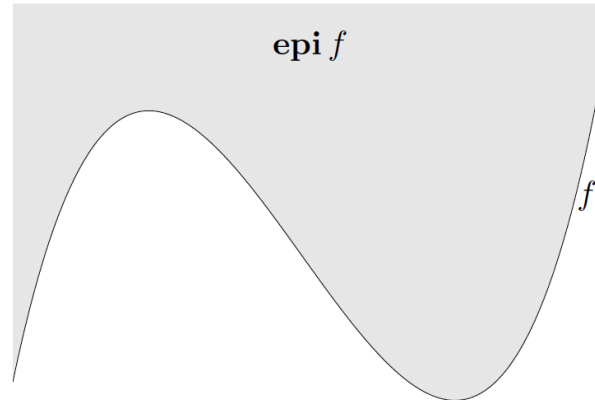- *Norms: Every norm on $\mathbb{R}^n$ is convex.*

Figure 4.12: Epigraph of a function f, shown shaded. The lower boundary, shown darker, is the graph of f.

Example 4.36 *Concave functions*

- *Geometric mean:* $f(x) = (\prod_{i=1}^{n} x_i)^{1/n} is concave, \forall x \in R_{++}^n$.
- *Log determinant:* $\log\det(X) is concave, \forall X \in S_{++}^n$.

4.3.3 Extended-value Extension

It is often convenient to extend a convex function to all of $\mathbb{R}^n$ by defining its value to be ∞ outside its domain.

Definition 4.37 $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \bigcup \{\infty\}$ *is extended-value extension of* f:

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

The extension $\tilde{f}$ is defined on all $\mathbb{R}^n$, and takes values in $\mathbb{R} \bigcup \{\infty\}$. This does not change its convexity

Theorem 4.38 f *is convex*

$\Leftrightarrow \tilde{f}$ *is convex*

$\Leftrightarrow \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta)\tilde{f}(y), 0 \leq \theta \leq 1$
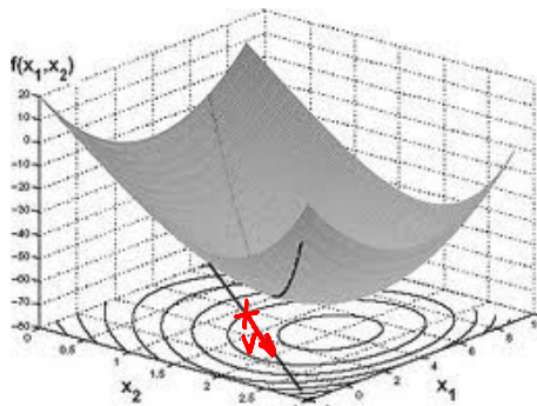
4.3.4 Epigraph

Definition 4.39 *The epigraph of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is defined as*

$\mathrm{epi} f = \{(x, t) | x \in \mathrm{dom} f, f(x) \leq t\}$

The definition is illustrated in Fig. **??**. The link between convex sets and convex functions is via the epigraph:

Theorem 4.40 f *is convex* $\Leftrightarrow \mathrm{epi} f$ *is a convex set*

Graph courtesy of Prof. Robert Freund
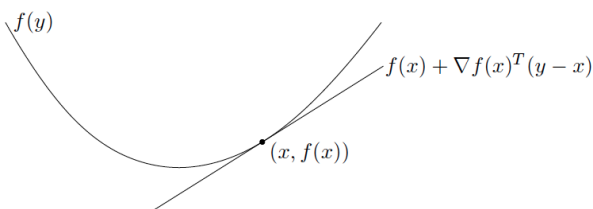
Figure 4.13: 0th order property.

Figure 4.14: 1st order property: If f is convex and differentiable, then $f(y) \geq f(x) + \nabla f(x)^T (y - x)$.

4.3.5 Convex Function Properties

4.3.5.1 0th Order Characterization

A function is convex if and only if when restricted to any line that intersects its domain is convex. That is:

$f : \mathbb{R}^n \to \mathbb{R}$ is convex

$\Leftrightarrow g(x) = f(x + tv)$ is convex, $\mathrm{dom}g = \{t | x + tv \in \mathrm{dom}f\}, \forall x \in \mathrm{dom}f, \forall v \in \mathbb{R}^n,$

This is useful, because we only need to check the convexity of 1D functions, as Fig. **??** shows.

4.3.5.2 1st Order Characterization

Let f be a differentiable function, $\mathrm{dom}f$ is open and convex, then we have

f is convex $\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y - x)$

It is illustrated as Fig. **??** shows. The inequality states that for a convex function, the first-order Taylor approximation is a global underestimator of the function. Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.

The inequality also shows that if $\nabla f(x) = 0$, then for $\forall y \in \text{dom} f$, $f(y) \geq f(x)$, and x is a global minimizer of the function.

4.3.5.3   2nd Order Characterization

Let f be twice differentiable, $\text{dom} f$ is open, then we have

f is convex $\Leftrightarrow \nabla^2 f(x) \geq 0, \forall x \in \text{dom} f$

If $\nabla^2 f(x) > 0, \forall x \in \text{dom} f$, f is strictly convex. The converse is not true.

For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^4$ is strictly convex but has zero second derivative at $x = 0$

4.3.6   Jensen's Inequality

The basic inequality, $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$, is sometimes called Jensen's inequality. Now we extended it to convex combinations of more than two points. If x is a random variable such that $x \in \text{dom} f$ with probability one, and f is convex, then we have

$$f(\mathrm{E}x) \leq \mathrm{E}f(x),$$

provided the expectations exist.

4.3.7   Convexity-preserving function operations

1. Nonnegative Weighted Sum:

   If f_1, f_2 are convex, $\omega_i \geq 0 \Rightarrow h(x) = \omega_1 f_1(x) + \omega_2 f_2(x)$ is convex.

2. Pointwise Max/Sup:

   If f, g are convex $\Rightarrow m(x) = \max\{f(x), g(x)\}$ is convex.

3. Extension of Pointwise Max/Sup:

   If $f(x, y)$ is convex in x for each $y \Rightarrow g(x) = \sup_{y \in C} f(x, y)$ is convex in x, provided $g(x) > -\infty$ for some x.

4. Affine Map:

   If $f : \mathbb{R}^n \to \mathbb{R}$ is convex $\Rightarrow g(x) = f(Ax + b)$ is convex, where $A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$.

5. Composition:

   If f, g are convex, and g is non-decreasing $\Rightarrow h(x) = g(f(x))$ is convex.

6. Perspective Map:

   If $f(x)$ is convex $\Rightarrow g(x, t) = tf(x/t)$ is convex.

4.3.8   How to prove a function is convex

- Use definition directly.

- Prove that epigraph is convex vis set methods.

- 0th, 1st, 2nd order convexity properties.

- Construct f from simpler convex functions using convexity-preserving operations.

4.4   Summary

4.4.1   Convex Sets

- Representation: convex hull, intersect hyperplanes.

- Supporting, separating hyperplanes.

- Operations that preserve convexity.

4.4.2   Convex Functions

- Epigraph.

- 0 orders, 1st order, 2nd order conditions.

- Operations that preserve convexity.