# Convexity

We'll assume throughout, without always saying so, that we're in the finite-dimensional Euclidean vector space $\mathbb{R}^n$, although sometimes, for statements that hold in any vector space, we'll say explicitly that we're in a vector space $V$. Theorems and remarks that are accompanied by the symbol ♣ are given without proof but are easy to prove. They provide good exercises in beginning to work with convexity and related concepts.

**Definition:** A set S in a vector space $V$ is **convex** if for any two points $x$ and $y$ in $S$, and any $\lambda$ in the unit interval $[0,1]$, the point $(1-\lambda)x + \lambda y$ is in $S$.

**Theorem:**♣ The intersection of any collection of convex sets is convex — *i.e.*, if for each $\alpha$ in some set $A$ the set $S_\alpha$ is convex, then the set $\bigcap_{\alpha \in A} S_\alpha$ is convex.

**Theorem:**♣ If $X_1, X_2, \ldots, X_m$ are convex sets, then $\sum_1^m X_i$ is convex.

**Definition:** Let $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ and let $b$ be a real number. The set of solutions of the linear equation $p_1 x_1 + \cdots + p_n x_n = b$ is a **hyperplane** in $\mathbb{R}^n$, and is denoted $H(\mathbf{p}, b)$ — *i.e.*, $H(\mathbf{p}, b) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{p} \cdot \mathbf{x} = b\}$.

**Remark:**♣ For any non-zero real number $\lambda$, we have $H(\lambda\mathbf{p}, \lambda b) = H(\mathbf{p}, b)$. Therefore any hyperplane can be represented by a $\mathbf{p}$ whose Euclidean norm $\|\mathbf{p}\|$ satisfies $\|\mathbf{p}\| = 1$: Given a hyperplane $H(\mathbf{p}, b)$ with $\|\mathbf{p}\| \neq 1$, let $\lambda = 1/\|\mathbf{p}\|$ and let $\mathbf{p}' = \lambda\mathbf{p}$ and $b' = \lambda b$; then $\|\mathbf{p}'\| = 1$ and $\mathbf{p}' \cdot \mathbf{x} = b'$ if and only if $\mathbf{p} \cdot \mathbf{x} = b$, so that $H(\mathbf{p}', b') = H(\mathbf{p}, b)$.

**Definition:** A **closed half-space** is a set of the form $\{x \in \mathbb{R}^n \mid \mathbf{p} \cdot \mathbf{x} \leqq b\}$ for some $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. An **open half-space** is a set of the form $\{x \in \mathbb{R}^n \mid \mathbf{p} \cdot \mathbf{x} < b\}$ for some $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Informally, a closed (resp. open) half-space is the set of all points on one side of a hyperplane, including (resp. not including) the hyperplane itself. Thus, a hyperplane $H$ in $\mathbb{R}^n$ divides $\mathbb{R}^n$ into two half-spaces which have in common either no points (if at least one of the half-spaces is open) or all the points in $H$ (if both half-spaces are closed).

**Remark:**♣ Every half-space (whether open or closed) is a convex set.

**Definition:** A real-valued function $f : X \to \mathbb{R}$ on a convex set $X$ is **concave** if

$$\forall \mathbf{x}, \mathbf{y} \in X, \lambda \in (0,1) : f\big((1-\lambda)\mathbf{x} + \lambda\mathbf{y}\big) \geqq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}\},$$

and $f$ is **convex** if

$$\forall \mathbf{x}, \mathbf{y} \in X, \lambda \in (0,1) : f\big((1-\lambda)\mathbf{x} + \lambda\mathbf{y}\big) \leqq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}\}.$$

**Definition:** A real-valued function $f : X \to \mathbb{R}$ on a convex set $X$ is **quasiconcave** if

$$\forall \mathbf{x}, \mathbf{y} \in X, \lambda \in (0,1) : f\big((1-\lambda)\mathbf{x} + \lambda \mathbf{y}\big) \geqq \min\{f(\mathbf{x}), f(\mathbf{y})\},$$

and $f$ is **quasiconvex** if and only if

$$\forall \mathbf{x}, \mathbf{y} \in X, \lambda \in (0,1) : f\big((1-\lambda)\mathbf{x} + \lambda \mathbf{y}\big) \leqq \max\{f(\mathbf{x}), f(\mathbf{y})\}.$$

In other words, $f$ is quasiconcave if and only if the value of $f$ at any convex combination of two points in $X$ is at least as large as its value at one of the two points; and $f$ is quasiconvex if the value of $f$ at any convex combination of two points is at least as small as it is at one of the two points.

**Definition:** Let $f : X \to \mathbb{R}$ be a real-valued function on a set $X \subseteq \mathbb{R}^n$, and let $a \in \mathbb{R}$. The **$f$-upper-contour set** and the **$f$-lower-contour set** for $a$ are the sets

$$U_f(a) = \{\mathbf{x} \in X \mid f(\mathbf{x}) \geqq a\} \qquad \text{and} \qquad L_f(a) = \{\mathbf{x} \in X \mid f(\mathbf{x}) \leqq a\}.$$

Let $\mathbf{x} \in X$; in a slight abuse of notation, we say the **$f$-upper-contour set** and the **$f$-lower-contour set** of $\mathbf{x}$ are the sets $U_f(f(\mathbf{x}))$ and $L_f(f(\mathbf{x}))$ — *i.e.*,

$$U_f(\mathbf{x}) = \{\mathbf{z} \in X \mid f(\mathbf{z}) \geqq f(\mathbf{x})\} \qquad \text{and} \qquad L_f(\mathbf{x}) = \{\mathbf{z} \in X \mid f(\mathbf{z}) \leqq f(\mathbf{x})\}.$$

The **strict upper- and lower-contour sets** are defined by replacing the weak inequalities with strict inequalities, and are denoted by $U_f^\circ(\cdot)$ and $L_f^\circ(\cdot)$. When we need to make the distinction, the sets defined by weak inequalities are called the **weak upper- and lower-contour sets**.

**Theorem:**$^\clubsuit$ A function $f : X \to \mathbb{R}$ is quasiconcave if and only if all of its upper-contour sets are convex — *i.e.*, if and only if $\forall \mathbf{x} \in X : U_f(\mathbf{x})$ is convex. And $f$ is quasiconvex if and only if all of its lower-contour sets are convex — $\forall \mathbf{x} \in X : L_f(\mathbf{x})$ is convex.

**Definition:** A function $f : X \to \mathbb{R}$ is **strictly** concave/convex/quasiconcave/quasiconvex if the inequality in the relevant definition above is strict whenever $\mathbf{x}$ and $\mathbf{y}$ are distinct points.

**Remark:**$^\clubsuit$ A function $f$ is concave if and only if $-f$ is convex. The same is true for each variant of concave and convex functions. We can therefore work only with concave functions (and variants), since every definition and result can be directly translated into a statement that replaces concave with convex and $f$ with $-f$.

**Remark:**$^\clubsuit$ If a function $f$ is concave then it is quasiconcave; if $f$ is strictly concave then it is strictly quasiconcave.

**Exercise:** Let $f : \mathbb{R}^2_{++} \to \mathbb{R}$ be the function defined by $f(\mathbf{x}) = x_1 x_2$. Verify that $f$ is neither concave nor convex, but is strictly quasiconcave.

**Remark:** If $f : X \to \mathbb{R}$ is a concave function and $f(\overline{x}) < f(\overline{\overline{x}})$, then $f(x(\lambda)) > f(\overline{x})$ for all $\lambda \in (0,1)$, where $x(\lambda) = (1-\lambda)\overline{x} + \lambda\overline{\overline{x}}$.

**Proof:**

$$f(x(\lambda) \geqq (1-\lambda)f(\overline{x}) + \lambda f(\overline{\overline{x}}) > (1-\lambda)f(\overline{x}) + \lambda f(\overline{x}) = f(\overline{x}). \quad \blacksquare$$

Suppose that $f(\overline{x}) < f(\overline{\overline{x}})$, as in the preceding remark, and that $b$ is a value between $f(\overline{x})$ and $f(\overline{\overline{x}})$, as depicted in Figure 1. The Intermediate Value Theorem for continuous real functions tells us that if $X \subseteq \mathbb{R}$ and if $f$ is continuous, there is a value of $x$ between $\overline{x}$ and $\overline{\overline{x}}$ for which $f(x) = b$. The following theorem is a kind of intermediate value theorem for concave functions: it tells us that if $f$ is concave (but not necessarily continuous, and more important, the domain $X$ need not be only one-dimensional), then there is a value of $\lambda$, say $\lambda^*$, such that for all larger values of $\lambda$, we have $f(x(\lambda)) > b$. The intuition for the theorem is clear in Figure 2 if we remember that the graph of a concave function must everywhere be on or above the line segment joining the points $(\overline{x}, f(\overline{x}))$ and $(\overline{\overline{x}}, f(\overline{\overline{x}}))$.

**Theorem:** If $f : X \to \mathbb{R}$ is a concave function and $f(\overline{x}) < b < f(\overline{\overline{x}})$, then $f(x(\lambda)) > b$ for every $\lambda > \lambda^*$, where

$$\lambda^* = \frac{b - f(\overline{x})}{f(\overline{\overline{x}}) - f(\overline{x})} \qquad \text{and} \qquad x(\lambda) = (1-\lambda)\overline{x} + \lambda\overline{\overline{x}}.$$

**Proof:** We first show that $f(x(\lambda^*)) \geqq b$, as follows:

$$\begin{aligned}
f(x(\lambda^*)) &\geqq (1-\lambda)f(\overline{x}) + \lambda f(\overline{\overline{x}}), \quad \text{by concavity of } f \\
&= \frac{f(\overline{\overline{x}}) - b}{f(\overline{\overline{x}}) - f(\overline{x})} f(\overline{x}) + \frac{b - f(\overline{x})}{f(\overline{\overline{x}}) - f(\overline{x})} f(\overline{\overline{x}}) \\
&= \frac{1}{f(\overline{\overline{x}}) - f(\overline{x})} \left[ f(\overline{\overline{x}})f(\overline{x}) - bf(\overline{x}) + bf(\overline{\overline{x}}) - f(\overline{\overline{x}})f(\overline{x}) \right] \\
&= \frac{f(\overline{\overline{x}}) - f(\overline{x})}{f(\overline{\overline{x}}) - f(\overline{x})} b \; = \; b.
\end{aligned}$$

Now there are three cases to consider: (i) If $f(x(\lambda^*)) < f(\overline{\overline{x}})$, then the conclusion, $f(x(\lambda)) > b$ for every $\lambda > \lambda^*$, follows from the preceding remark, where the roles of $\overline{x}$ and $\overline{\overline{x}}$ in the remark are played here by $x(\lambda^*)$ and $\overline{\overline{x}}$, respectively. (ii) If $f(x(\lambda^*)) = f(\overline{\overline{x}})$, then the conclusion is obvious. (iii) if $f(x(\lambda^*)) > f(\overline{\overline{x}})$, then the conclusion follows from the remark, where the roles of $\overline{x}$ and $\overline{\overline{x}}$ in the remark are played here by $\overline{\overline{x}}$ and $x(\lambda^*)$, respectively. $\blacksquare$

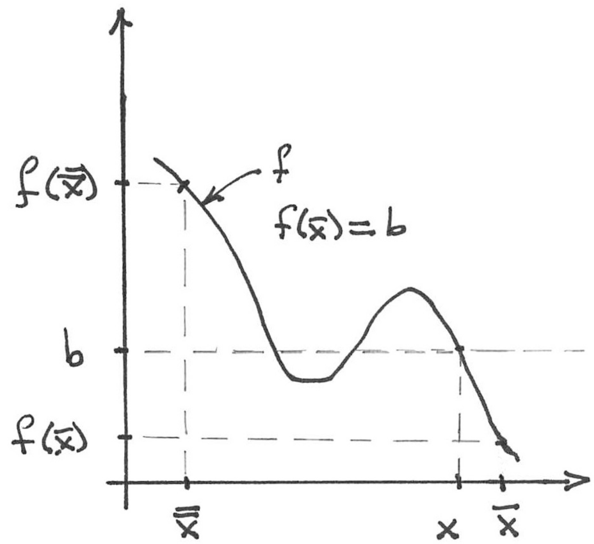


Figure 1

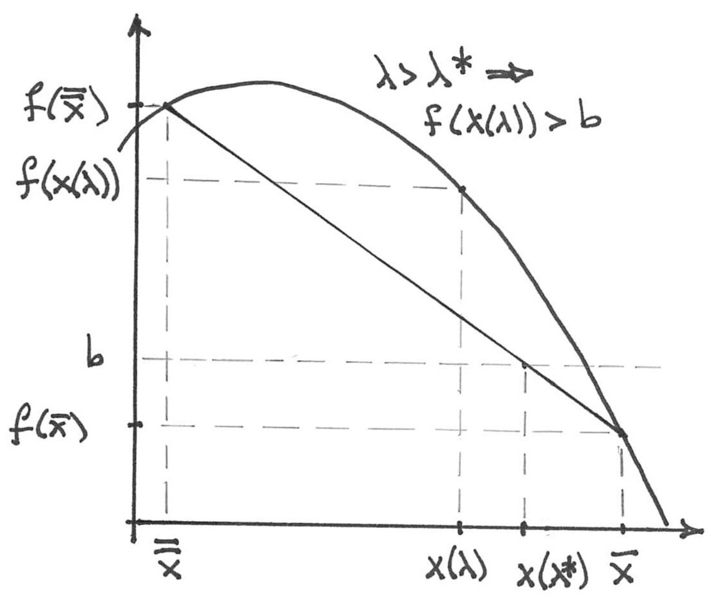


Figure 2

## Preview: Convex Optimization

**Definition:** The hyperplane $H(\mathbf{p}, b)$ **separates** sets $X$ and $Y$ in $\mathbb{R}^n$ if for all $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, we have $\mathbf{p} \cdot \mathbf{x} \leqq b \leqq \mathbf{p} \cdot \mathbf{y}$.

The following theorem is a classical theorem of convex analysis. The theorem seems obviously true from an intuitive, geometric perspective. The theorem and its proof confirm this intuition. The proof will be given later in the course.

**Minkowski's Theorem:** If $X$ and $Y$ are nonempty disjoint convex sets in $\mathbb{R}^n$, then there is a hyperplane that separates them — *i.e.*, there exist a $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ and a $b \in \mathbb{R}$ such that for all $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, we have $\mathbf{p} \cdot \mathbf{x} \leqq b \leqq \mathbf{p} \cdot \mathbf{y}$.

**Theorem:** Let $f : X \to \mathbb{R}$ be a continuous quasiconcave function on a convex domain $X \subseteq \mathbb{R}^n$; and let $S$ be a convex subset of $X$. Let $\overline{\mathbf{x}}$ be an element of $S$ at which $f$ does not attain a local maximum on $X$. Then $\overline{\mathbf{x}}$ maximizes $f$ on $S$ if and only if there is a $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ such that

$$\text{(a)} \;\; \overline{\mathbf{x}} \text{ maximizes } f(\mathbf{x}) \text{ s.t. } \mathbf{p} \cdot \mathbf{x} \leqq \mathbf{p} \cdot \overline{\mathbf{x}} \quad \text{and} \quad \text{(b)} \;\; \overline{\mathbf{x}} \text{ maximizes } \mathbf{p} \cdot \mathbf{x} \text{ s.t. } \mathbf{x} \in S.$$

**The Beginning of a Proof:** It's easy to see that (a) and (b) together imply that $\overline{\mathbf{x}}$ maximizes $f$ on $S$: if $\mathbf{x} \in S$, then $\mathbf{p} \cdot \mathbf{x} \leqq \mathbf{p} \cdot \overline{\mathbf{x}}$ according to (b); and therefore $f(\mathbf{x}) \leqq f(\overline{\mathbf{x}})$ according to (a).

To prove the converse, we assume that $\overline{\mathbf{x}}$ maximizes $f$ on $S$, and we will show that there is a $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ that satisfies (a) and (b). Let $U = \{\mathbf{x} \in X \mid f(\mathbf{x}) > f(\overline{\mathbf{x}})\}$, the strict $f$-upper-contour set of $\overline{\mathbf{x}}$. $U$ is nonempty and convex, and is disjoint from $S$; and since $S$ is also nonempty and convex, the Minkowski Theorem guarantees the existence of a $\mathbf{p} \neq \mathbf{0} \in \mathbb{R}^n$ and a real number $b > 0$ such that the hyperplane $H(\mathbf{p}, b)$ separates the two sets — *i.e.*,

$$(*) \quad \forall \mathbf{x} \in S : \mathbf{p} \cdot \mathbf{x} \leqq b \qquad \text{and} \qquad (**) \quad \forall \mathbf{x} \in U : b \leqq \mathbf{p} \cdot \mathbf{x}.$$

The remainder of the proof consists of first showing that $(*)$ and $(**)$ together imply that $b = \mathbf{p} \cdot \overline{\mathbf{x}}$, and then using this fact to establish the conclusions (a) and (b) of the theorem. This is the place where we use the continuity of $f$ — note that we so far haven't made use of continuity. We defer this part of the proof to later in the course. $\blacksquare$

Here are several observations about this convex optimization theorem:

(1) If $f$ is strictly quasiconcave, or if $S$ is (informally speaking, for the time being) a "strictly convex" set, then a maximizer $\overline{\mathbf{x}}$ of $f$ on $S$ will be unique. If *both* of these "strict" conditions are satisfied, then $\overline{\mathbf{x}}$ will be (a) a unique maximizer of $f$ subject to $\mathbf{p} \cdot \mathbf{x} \leqq \mathbf{p} \cdot \overline{\mathbf{x}}$ and (b) a unique maximizer of $\mathbf{p} \cdot \mathbf{x}$ subject to $\mathbf{x} \in S$.

(2) The theorem is an *existence theorem*: one of its conclusion says that *there exists* a vector $\mathbf{p}$ that satisfies (a) and (b). (Minkowski's Theorem is also an existence theorem.) The $\mathbf{p}$ that exists is often interpreted as a list of prices of various goods, and the theorem is then interpreted as a theorem about "separation of decisions" or "decentralizing decision-making." For example, suppose $S$ is a set of feasible output combinations (a production-possibilities set); $f$ is a utility or objective function; and we wish to find an $\overline{\mathbf{x}}$ that maximizes $f$ over $S$. The theorem says that there is some list $\mathbf{p}$ of prices such that a person who knows the function $f$ and who knows the prices and the budget amount $b$, but who knows nothing about the set $S$, could choose $\mathbf{x}$ so as to maximize $f$ subject to the restriction that the value of $\mathbf{x}$ (at prices $\mathbf{p}$) not exceed the budget $b$; and another person, knowing only the set $S$ and the prices $\mathbf{p}$, but knowing nothing about $f$, could choose $\mathbf{x}$ to maximize the value of $\mathbf{x}$ among all the $\mathbf{x} \in S$; and the (unknown) $\overline{\mathbf{x}}$ would be a solution for each of the two decision-makers. And it would be the *unique* solution of each person's problem if $f$ is strictly quasiconcave or $S$ is "strictly convex." Alternatively, *one* person who wishes to choose a best $\mathbf{x}$ in $S$ can separate his decision-making into two distinct decision problems, a consumption decision that doesn't involve $S$ and a production decision that doesn't involve $f$.

Of course, we may not know the separating price-list $\mathbf{p}$ any more than we know the maximizing plan $\overline{\mathbf{x}}$ that we're trying to find. A potential way to find the correct $\mathbf{p}$ and $\overline{\mathbf{x}}$ might be to iteratively try various price-lists $\mathbf{p}$. As long as the two decision-makers choose different plans, we adjust the price-list until (we hope) the price-list converges to a $\mathbf{p}$ at which they choose the same plan. The theorem guarantees that if there is a maximizing $\overline{\mathbf{x}}$ then there will be such a separating $\mathbf{p}$, and that when both decision-makers choose the same plan we will have the "correct" separating $\mathbf{p}$ and therefore the plan they both choose will be the maximizing plan. There's an obvious similarity here to a market process, where the prices adjust to excess demand or supply until reaching an "equilibrium" at which demand equals supply. This similarity is not accidental; we'll see this again in Econ 501B.

(3) The theorem makes no mention of differentiability of $f$ or of differentiability of any functions that might define the set $S$. So the theorem ensures the existence of a separating $\mathbf{p}$ in a broad range of situations where differentiability is not present.

(4) The constraint set, or feasible set, $S$ is often defined by a set of inequality constraints, such as

$$g_i(\mathbf{x}) \leqq c_i \quad (i = 1, \ldots, m),$$

where each function $g_i(\cdot)$ is quasiconvex and where *all* of the $m$ inequalities have to be satisfied. Then $S$ is the intersection of $m$ convex sets

$$S_i = \{\mathbf{x} \,|\, g_i(\mathbf{x}) \leqq c_i\}, \quad (i = 1, \ldots, m),$$

as in Figure 3, and therefore $S$ itself is convex, as required in the theorem. In particular, the constraints could be linear, as in Figure 4.

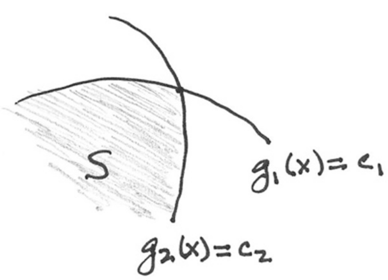


Figure 3

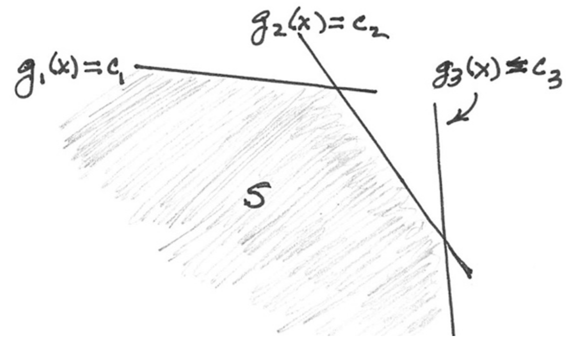


Figure 4