Chapter 9

Convex sets

9.1 Affine sets

Definition 40. *(Geometric). A set $C \subset \mathbb{R}^n$ is affine if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C, the line passing through $\mathbf{x}_1$ and $\mathbf{x}_2$ lies entirely in C.*

(Algebraic). A set $C \subset \mathbb{R}^n$ is affine if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C and any real scalar θ, the point $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$.

The algebraic version simply expresses the geometry of the line passing through $\mathbf{x}_1$ and $\mathbf{x}_2$ as the parameterization $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$.

Definition 41. *Given a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ of weights such that $\sum_{i=1}^{k} \theta_i = 1$, the $\boldsymbol{\theta}$-affine combination of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$ is the point $\theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k$.*

An affine combination of points is, therefore, a constrained linear combination where the weights are constrained to sum to 1.

Proposition 7. *A set $C \subset \mathbb{R}^n$ is affine if and only if it contains all affine combinations of its points*

Proof. 'if' part: The reverse direction is easy. If C contains all affine combinations of its points, then it contains all affine combinations of any two of its points. This verifies that C is affine.

'only if' part: Choose k (k arbitrary) points in C, and a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ of weights such that $\sum_{i=1}^{k} \theta_i = 1$. Then we want to show that the $\boldsymbol{\theta}$-affine combination $\mathbf{x} = \theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k$ is a point in C. We show this by mathematical induction on k.

For the base step $k = 2$, the affine combination $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2$ is a point in C because of the premise in the forward direction that C is affine.

For the inductive step, suppose for any $k-1$ points in C, their affine combination lies in C.

Then for any k arbitrarily chosen points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in C, their $\boldsymbol{\theta}$-affine combination can be written as

$$\begin{aligned} \theta_1 \mathbf{x}_1 + \ldots + \theta_k \mathbf{x}_k &= (1-\theta_k) \underbrace{\left(\frac{\theta_1}{1-\theta_k} \mathbf{x}_1 + \ldots + \frac{\theta_{k-1}}{1-\theta_k} \mathbf{x}_{k-1} \right)}_{\in C \text{ by inductive hypothesis}} + \theta_k \mathbf{x}_k \\ &\in C \qquad\qquad (C \text{ is an affine set}) \end{aligned}$$

This proves the forward direction and hence the proposition. Q.E.D.

Definition 42. *The affine hull* $\mathbf{aff}(C)$ *of a set* $C \subset \mathbb{R}^n$ *is the set of all affine combination of points in* C*. Formally,*

$$\mathbf{aff}(C) := \{\mathbf{x} \in \mathbb{R}^n : \textit{ there exist points } \mathbf{x}_1, , \mathbf{x}_k \textit{ in } C \textit{ and weights } \theta_1, \ldots, \theta_k \textit{ such that } \sum_{i=1}^{k} \theta_i = 1 \textit{ and } \mathbf{x} = \theta_1 \mathbf{x}_1 + \ldots + \theta_k \mathbf{x}_k\}$$

The affine hull $\mathbf{aff}(C)$ of any set $C \subset \mathbb{R}^n$ is an affine set itself. To see this, take two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{aff}(C)$. Since these points are in $\mathbf{aff}(C)$, $\mathbf{x}$ is some $\boldsymbol{\theta}$-affine combination of k points in C, and $\mathbf{y}$ is some $\boldsymbol{\beta}$-affine combination of n points in C. But this means that for $\alpha \in \mathbb{R}$ we can write

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} = \sum_{i=1}^{k} \alpha \theta_i \mathbf{x}_i + \sum_{j=1}^{n} (1-\alpha) \beta_j \mathbf{y}_i$$

$$\text{where} \quad \sum_{i=1}^{k} \alpha \theta_i + \sum_{j=1}^{n} (1-\alpha)\beta_j = \alpha (\sum_{i=1}^{k} \theta_i) + (1-\alpha)(\sum_{j=1}^{n} \beta_j) = 1$$

In other words, $\alpha \mathbf{x} + (1-\alpha)\mathbf{y}$ is some affine combination of points in C. As a result, it lies in $\mathbf{aff}(C)$. This shows $\mathbf{aff}(C)$ is affine.

Proposition 7 and the definition of an affine hull imply that the affine hull of any affine set C is the set C itself. In other words, if the underlying set C is affine, then the operation of taking an affine hull does not do anything - it adds no new points to the set.

The next result brings out the key idea about affine sets - they are simply translates of subspaces. Subspaces, we already know, have to go through the origin in the ambient space. If we use any vector of the subspace to do a parallel translation of that subspace, we get an affine set.

Proposition 8. *(Affine sets are translated subspaces). If* $C \subset \mathbb{R}^n$ *is an affine set and* $\mathbf{x}_0$ *is a point in* C*, then the set*

$$V = C - \{\mathbf{x}_0\} = \{\mathbf{x} - \mathbf{x}_0 : \mathbf{x} \in C\}$$

is a subspace in $\mathbb{R}^n$*.*

Proof. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two points in V and θ_1 and θ_2 be real scalars. Then there exist points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C such that $\mathbf{v}_1 = \mathbf{x}_1 - \mathbf{x}_0$ and $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{x}_0$. Therefore

$$\begin{aligned}\theta_1\mathbf{v}_1 + \theta_2\mathbf{v}_2 &= \theta_1(\mathbf{x}_1 - \mathbf{x}_0) + \theta_2(\mathbf{x}_2 - \mathbf{x}_0)\\ &= \theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 - (\theta_1 + \theta_2)\mathbf{x}_0\\ &= \underbrace{\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + (1 - \theta_1 - \theta_2)\mathbf{x}_0}_{\in C \text{ as an affine combination of points in } C} - \mathbf{x}_0\\ &\in V\end{aligned}$$

This shows V is a subspace. Q.E.D.

Proposition 9. *(Arbitrary intersection of affine sets is an affine set). If $\{C_i\}_{i\in I}$ is a family of affine sets in $\mathbb{R}^n$, then $C = \cap_{i\in I}$ is an affine set.*

Proof. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two points in C and θ_1 and θ_2 be real scalars such that $\theta_1 + \theta_2 = 1$. Then since both the points are in every C_i, and every C_i is an affine set, the line $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2$ is also in every C_i. This implies that the line $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2$ is also in C. C is therefore an affine set. Q.E.D.

9.2 Hyperplanes

The trivial affine sets of $\mathbb{R}^n$ are a point set, which is zero dimensional; and $\mathbb{R}^n$ itself, which is full dimensional. Between these, there are affine sets of various dimensions. A hyperplane in $\mathbb{R}^n$ is an $n-1$ dimensional affine set. Hyperplanes are largest affine sets in $\mathbb{R}^n$ (other than of course $\mathbb{R}^n$ itself), and they are a natural generalization of planes in 3D to higher dimensions. This explains the prefix 'hyper'. More precisely, we have

Definition 43. *Given a nonzero vector $\mathbf{a} \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$, a hyperplane $H(\mathbf{a}, b)$ in $\mathbb{R}^n$ is the set*

$$H(\mathbf{a}, b) := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x}\rangle = b\} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1x_1 + \ldots + a_nx_n = b\}$$

(**Exercise**). Verify that a hyperplane $H(\mathbf{a}, b)$ is an affine set.

From the definition, we can read out some simple instances of hyperplanes.

• hyperplanes in $\mathbb{R}^2$ are sets of the form $\{(x_1, x_2) \in \mathbb{R}^2 : a_1x_1 + a_2x_2 = b\}$. These are simply lines in the plane.
• hyperplanes in $\mathbb{R}^3$ are sets of the form $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : a_1x_1 + a_2x_2 + a_3x_3 = b\}$. These are simply planes in 3D.

The definition also expresses two ways to think about a hyperplane.

Analytically, a hyperplane in $\mathbb{R}^n$ is the solution set of a linear equation in n variables.

The fact that a hyperplane is an affine set gives us the simple conclusion that the solution set of a linear equation is an affine set. Generalizing this observation, the solution set of a linear equation system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix, can be analytically viewed as the intersection of solution sets of m linear equations $\langle \mathbf{a}_i, \mathbf{x} \rangle = b_i$, where $\mathbf{a}_i$ is the i-th row of A (viewed as a column vector). By Proposition 9, we obtain the conclusion that

The solution set of a linear equation system $A\mathbf{x} = \mathbf{b}$ is an affine set.

To develop a geometric view of the hyperplane, take a point $\mathbf{x}_0$ in the hyperplane so that $\langle \mathbf{a}, \mathbf{x}_0 \rangle = b$. Then for any point $\mathbf{x}$ in the hyperplane, the vector $\mathbf{x} - \mathbf{x}_0$ lies in the hyperplane and we have

$$H(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle = 0\} \tag{9.1}$$

(9.1) expresses the fundamental orthogonality property that the vector $\mathbf{a}$ is orthogonal to any vector that lies in the hyperplane. Figure 9.1 illustrates this property.

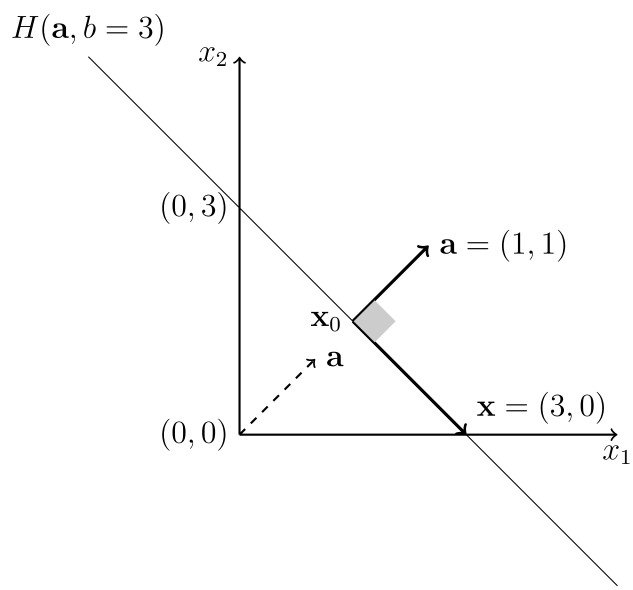

Figure 9.1: Hyperplane geometry diagram: In $\mathbb{R}^2$, the line $x_1 + x_2 = 3$ is the hyperplane $H(\mathbf{a}, b)$ with $\mathbf{a} = (1, 1)$ and $b = 3$. The vector $\mathbf{x}_0$ is the point $(1.5, 1.5)$. This means the vector $\mathbf{x} - \mathbf{x}_0 = (1.5, -1.5)$. The diagram displays the orthogonality property - the normal vector $\mathbf{a}$ is orthogonal to the vector $\mathbf{x} - \mathbf{x}_0$ that lies in the hyperplane. Algebraically, $\langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle = 0$.

From (9.1), we have

$\mathbf{x} \in H(\mathbf{a}, b)$ if and only if $\langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle = 0$ if and only if $\mathbf{x} - \mathbf{x}_0 \in \mathbf{a}^\perp$ if and only if $\mathbf{x} \in \mathbf{x}_0 + \mathbf{a}^\perp$

where $\mathbf{a}^\perp = \{\mathbf{y} : \langle \mathbf{a}, \mathbf{y} \rangle = 0\}$ is the orthogonal complement of vector $\mathbf{a}$. The orthogonal complement $\mathbf{a}^\perp$ is the set of all vectors that are orthogonal to $\mathbf{a}$. From basic linear algebra, we know that $\mathbf{a}^\perp$ is a subspace. We can therefore say that

The geometric view of the hyperplane is that it is some translate of the orthogonal complement $\mathbf{a}^{\perp}$.

(**Halfspaces**). A hyperplane $H(\mathbf{a}, b)$ in $\mathbb{R}^n$ has two closed sets associated with it: the closed positive halfspace $H^+(\mathbf{a}, b)$ and the closed negative halfspace $H^-(\mathbf{a}, b)$. These sets are defined as

$$H^+(\mathbf{a}, b) := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \geq b\} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1 x_1 + \ldots + a_n x_n \geq b\}$$
$$H^-(\mathbf{a}, b) := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \leq b\} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1 x_1 + \ldots + a_n x_n \leq b\}$$

Analytically, a closed halfspace is the solution set of a linear inequality. Figure 9.2 illustrates these halfspaces and the key observation that the positive halfspace extends in the direction of the normal vector.

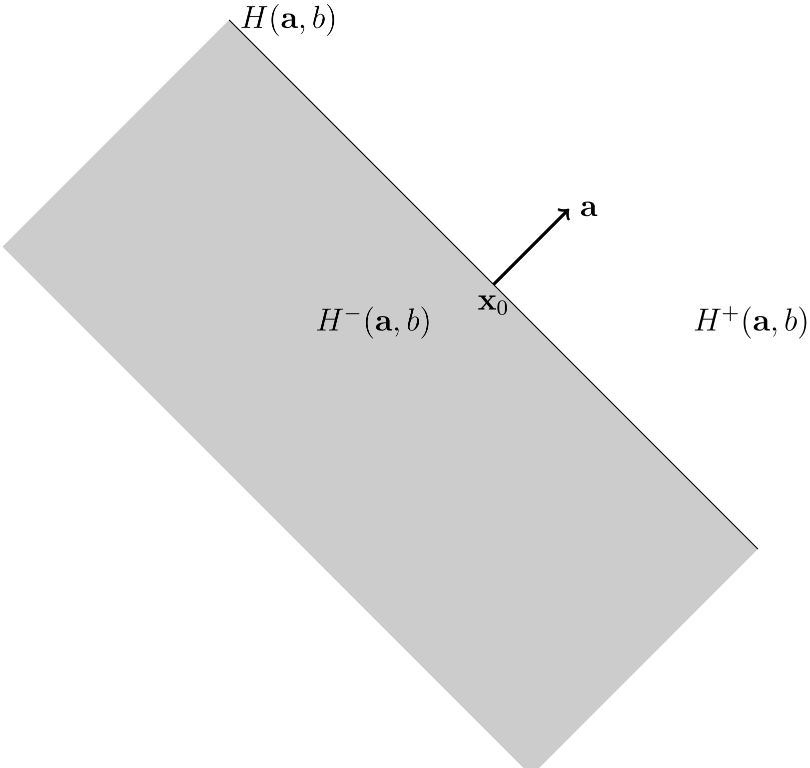

Figure 9.2: Halfspace orientation diagram: The hyperplane $H(\mathbf{a}, b)$ determines two halfspaces. The positive closed halfspace $H^+(\mathbf{a}, b)$ is the region $\{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \geq b\}$; and extends in the direction of the normal vector $\mathbf{a}$. The negative closed halfspace $H^-(\mathbf{a}, b)$ is the region $\{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}$; and extends in the direction of $-\mathbf{a}$.

The halfspaces associated with a hyperplane $H(\mathbf{a}, b)$ can also be described in terms of an angular property. The closed negative halfspace $H^-(\mathbf{a}, b)$ consists of all those points $\mathbf{x}$ such that the normal vector $\mathbf{a}$ makes an obtuse angle with the vector $\mathbf{x} - \mathbf{x}_0$, which is a vector in the negative halfspace. A similar angular property holds for the closed positive halfspace. Figure 9.3 illustrates the halfspaces in terms of this angular property.

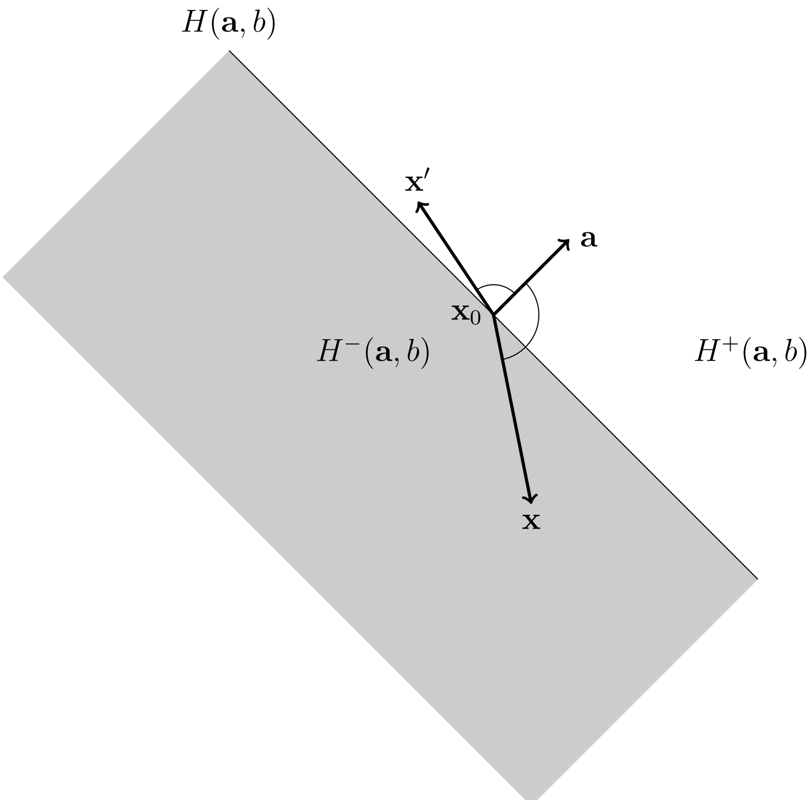

Figure 9.3: Halfspace angular property diagram: If $\mathbf{x}_0 \in H(\mathbf{a}, b)$ and $\mathbf{x} \in H^-(\mathbf{a}, b)$, then this means $\langle \mathbf{a}, \mathbf{x}_0 \rangle = b$ and $\langle \mathbf{a}, \mathbf{x} \rangle \leq b$ implying $\langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle \leq 0$. In other words, the normal vector $\mathbf{a}$ makes an obtuse angle with the vector $\mathbf{x} - \mathbf{x}_0$. Similarly, if $\mathbf{x}' \in H^+(\mathbf{a}, b)$, then $\mathbf{a}$ makes an acute angle with the vector $\mathbf{x}' - \mathbf{x}_0$.

9.3 Convex sets

Definition 44. *(Geometric). A set $C \subset \mathbb{R}^n$ is convex if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C, the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ lies entirely in C.*

(Algebraic). A set $C \subset \mathbb{R}^n$ is convex if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C and any scalar $\theta \in [0, 1]$, the point $\theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in C$.

The algebraic version simply expresses the geometry of the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ as the parameterization $\theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$. Figure 9.4 shows some convex sets in the plane. In order to appreciate the definition, Figure 9.5 shows some nonconvex sets in the plane as well.

Definition 45. *Given a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ of weights such that for every $i = 1, \ldots, k$, $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$, the $\boldsymbol{\theta}$-convex combination of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$ is the point* in the negative halfspace. A similar angular property holds for the closed positive halfspace. Figure 9.3 illustrates the halfspaces in terms of this angular property.

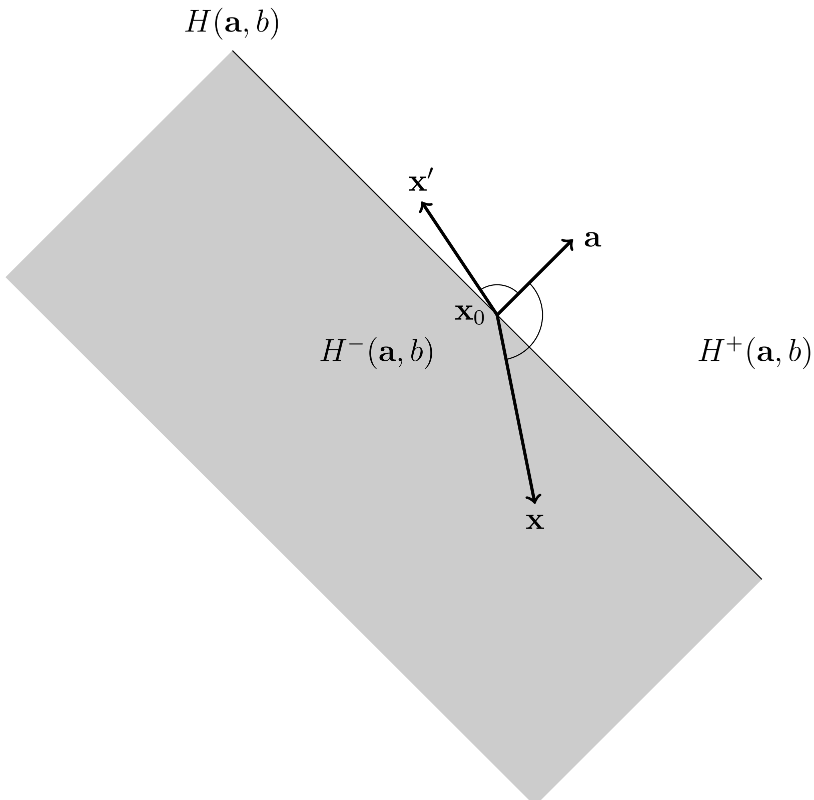

Figure 9.3: Halfspace angular property diagram: If $\mathbf{x}_0 \in H(\mathbf{a}, b)$ and $\mathbf{x} \in H^-(\mathbf{a}, b)$, then this means $\langle \mathbf{a}, \mathbf{x}_0 \rangle = b$ and $\langle \mathbf{a}, \mathbf{x} \rangle \leq b$ implying $\langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle \leq 0$. In other words, the normal vector $\mathbf{a}$ makes an obtuse angle with the vector $\mathbf{x} - \mathbf{x}_0$. Similarly, if $\mathbf{x}' \in H^+(\mathbf{a}, b)$, then $\mathbf{a}$ makes an acute angle with the vector $\mathbf{x}' - \mathbf{x}_0$.

9.3 Convex sets

Definition 44. *(Geometric). A set $C \subset \mathbb{R}^n$ is convex if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C, the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ lies entirely in C.*

(Algebraic). A set $C \subset \mathbb{R}^n$ is convex if given any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C and any scalar $\theta \in [0, 1]$, the point $\theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in C$.

The algebraic version simply expresses the geometry of the line segment joining $\mathbf{x}_1$ and $\mathbf{x}_2$ as the parameterization $\theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$. Figure 9.4 shows some convex sets in the plane. In order to appreciate the definition, Figure 9.5 shows some nonconvex sets in the plane as well.

Definition 45. *Given a vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ of weights such that for every $i = 1, \dots, k$, $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$, the $\boldsymbol{\theta}$-convex combination of points $\mathbf{x}_1, \dots, \mathbf{x}_k$ in $\mathbb{R}^n$ is the point*

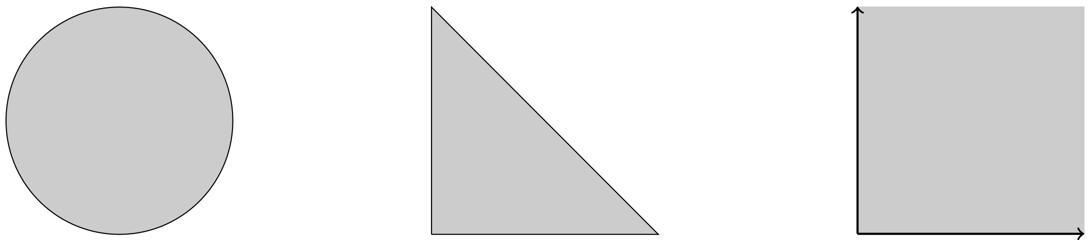

Figure 9.4: Some convex sets in the plane: a circular disk, a triangular disk, nonnegative quadrant of cartesian plane

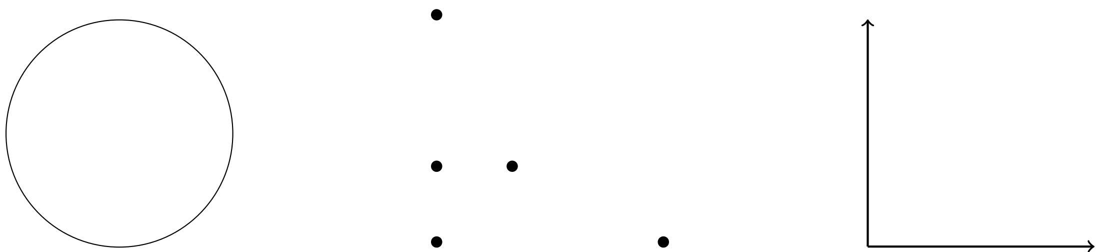

Figure 9.5: Some non-convex sets in the plane: a circle, a finite set of points, nonnegative axes of cartesian plane

$\theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k$.

A convex combination of points is, therefore, a constrained affine combination where the weights are constrained to be nonnegative.

Proposition 10. *A set $C \subset \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its points*

Proof. 'if' part: The reverse direction is easy. If C contains all convex combinations of its points, then it contains all convex combinations of any two of its points. This verifies that C is convex.

'only if' part: Choose k (k arbitrary) points in C, and a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ of weights such that for every $i = 1, \ldots, k$, $\theta_i \geq 0$ and $\sum_{i=1}^{k} \theta_i = 1$. Then we want to show that the $\boldsymbol{\theta}$-convex combination $\mathbf{x} = \theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k$ is a point in C. We show this by mathematical induction on k.

For the base step $k = 2$, the convex combination $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2$ is a point in C because of the premise in the forward direction that C is convex.

For the inductive step, suppose for any $k-1$ points in C, their convex combination lies in C. Then for any k arbitrarily chosen points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in C, their $\boldsymbol{\theta}$-convex combination can be written as

$$\theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k = (1-\theta_k)\underbrace{\left(\frac{\theta_1}{1-\theta_k}\mathbf{x}_1 + \ldots + \frac{\theta_{k-1}}{1-\theta_k}\mathbf{x}_{k-1}\right)}_{\in C \text{ by inductive hypothesis}} + \theta_k\mathbf{x}_k$$

$$\in C \qquad\qquad (C \text{ is a convex set})$$

This proves the forward direction and hence the proposition. Q.E.D.

Definition 46. *The convex hull $\mathbf{co}(C)$ of a set $C \subset \mathbb{R}^n$ is the set of all convex combination of points in C. Formally,*

$$\mathbf{co}(C) := \{\mathbf{x} \in \mathbb{R}^n : \text{ there exist points } \mathbf{x}_1, , \mathbf{x}_k \text{ in } C \text{ and weights } \theta_1, \ldots, \theta_k \text{ such that}$$
$$\text{for every } i = 1, \ldots, k, \quad \theta_i \geq 0, \quad \sum_{i=1}^{k} \theta_i = 1 \text{ and } \mathbf{x} = \theta_1 \mathbf{x}_1 + \ldots + \theta_k \mathbf{x}_k\}$$

The convex hull of the nonconvex sets shown in Figure 9.5 are the corresponding convex sets shown in 9.4.

The convex hull $\mathbf{co}(C)$ of any set $C \subset \mathbb{R}^n$ is a convex set itself. To see this, take two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{co}(C)$. Since these points are in $\mathbf{co}(C)$, $\mathbf{x}$ is some $\boldsymbol{\theta}$-convex combination of k points in C, and $\mathbf{y}$ is some $\boldsymbol{\beta}$-convex combination of n points in C. But this means that for $\alpha \in [0, 1]$ we can write

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} = \sum_{i=1}^{k} \alpha \theta_i \mathbf{x}_i + \sum_{j=1}^{n} (1-\alpha)\beta_j \mathbf{y}_i$$
$$\text{where} \quad \sum_{i=1}^{k} \alpha \theta_i + \sum_{j=1}^{n} (1-\alpha)\beta_j = \alpha(\sum_{i=1}^{k} \theta_i) + (1-\alpha)(\sum_{j=1}^{n} \beta_j) = 1$$

In other words, $\alpha \mathbf{x} + (1-\alpha)\mathbf{y}$ is some convex combination of points in C. As a result, it lies in $\mathbf{co}(C)$. This shows $\mathbf{co}(C)$ is convex.

Proposition 10 and the definition of a convex hull imply that that the convex hull of any convex set C is the set C itself. In other words, if the underlying set C is convex, then the operation of taking a convex hull does not do anything - it adds no new points to the set.

The forward direction of Proposition 10 generalizes as follows: Let us imagine the convex set $C \subset \mathbb{R}^n$ as the sample space for the random vector $\mathbf{x}$ which is distributed on C according to the probability density $p(\bullet)$. Then

$$\mathbb{E}(\mathbf{x}) = \int_C \mathbf{x} f(\mathbf{x}) \in C$$

9.4 Polyhedra

Definition 47. *(Geometric). A closed polyhedron is an intersection of finitely many closed halfspaces. Formally, given hyperplanes $H(\mathbf{a}_1, b_1), \ldots, H(\mathbf{a}_m, b_m)$ where for every $i = 1, \ldots, m$, $\mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$, define a polyhedron to be the set*

$$P := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i \text{ for every } i = 1, \ldots, m\} = H^-(\mathbf{a}_1, b_1) \cap \ldots \cap H^-(\mathbf{a}_m, b_m)$$

(Analytic). A closed polyhedron is the solution set of a linear inequality system. This is clear by taking an analytic look at the geometric definition above. To write it more compactly, construct an $m \times n$ matrix A by having $\mathbf{a}_i \in \mathbb{R}^n$ as its i-th row for every $i = 1, \ldots, m$, and let $\mathbf{b} = (b_1, \ldots, b_m) \in \mathbb{R}^m$. Then a closed polyhedron is the set

$$P := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$$

A closed polyhedron is a (closed) convex set. This is because it is constructed up from closed halfspaces, which are convex sets, using the convexity preserving operation of intersecting finitely many of them.

Example 25. Consider the geometric region in the plane described by the set $P = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq 1; x_1 \geq 0\}$. The set P, as the solution set of a linear inequality system, is a polyhedron. It can be geometrically expressed as $H^-((1,1),2) \cap H^-((-1,0),0)$. Figure 9.6 shows the polyhedron P as extending indefinitely in the downward direction.

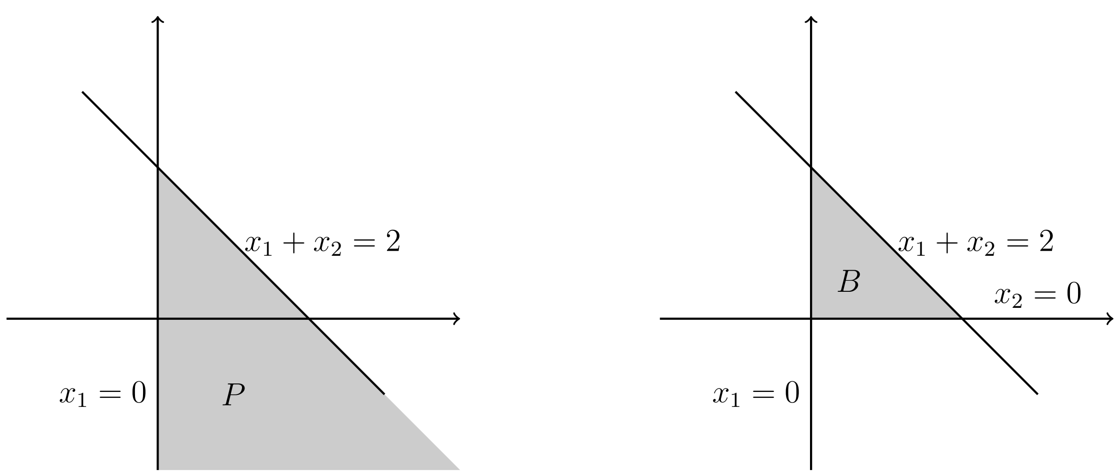

Figure 9.6: The sets P and B are both polyhedrons. The set B is a polytope as well.

Definition 48. *(Polytope). A set $P \subset \mathbb{R}^n$ is called a polytope if it is the convex hull of finitely many points in $\mathbb{R}^n$.*

Note that a polytope is also a closed polyhedron, but one which is bounded.

Example 26. Consider the geometric region in the plane described by the set $B = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq 1; x_1 \geq 0; x_2 \geq 0\}$. The set B, as the solution set of a linear inequality system, is again a polyhedron. It can also be geometrically expressed as $H^-((1,1),2) \cap H^-((-1,0),0) \cap H^-((0,-1),0)$. Figure 9.6 shows the polyhedron B as a bounded set. Note also that B is the convex hull of three points: $(0,0)$, $(2,0)$ and $(0,2)$. The bounded polyhedron B is, therefore, a polytope as well.

A special kind of polytope, which routinely arises in applications which involve optimizing over probability distributions, is known as the probability simplex.

Example 27. (Probability Simplex). A probability simplex in $\mathbb{R}^n$ is the convex hull of the standard basis vectors of $\mathbb{R}^n$. Formally, if $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denote the standard basis vectors of $\mathbb{R}^n$, then the probability simplex in is the set

$$S := \mathbf{co}(\{\mathbf{e}_1, \ldots, \mathbf{e}_n\})$$

By its definition, a probability simplex is a polytope. It can also be expressed as the set

$$S := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \text{ for every } i = 1, \ldots, n, \quad x_i \geq 0; \text{ and } x_1 + \ldots + x_n = 1\}$$

This way of writing the probability simplex reveals the reason behind its name - any point of this set can be thought of as a discrete probability distribution over its vertices - the points corresponding to the standard basis vectors. Figure 9.7 shows the probability simplexes in $\mathbb{R}^2$ and $\mathbb{R}^3$.

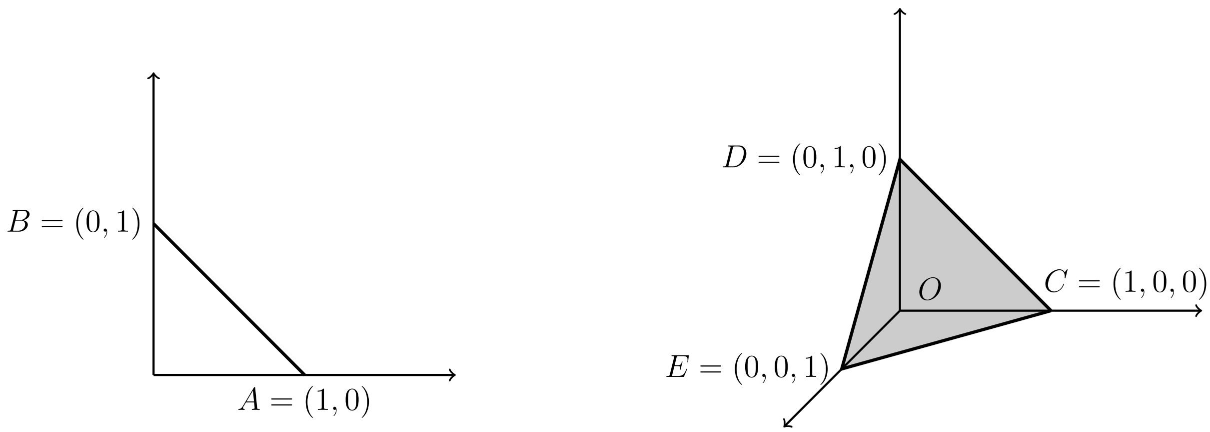

Figure 9.7: Left Panel: A probability simplex in $\mathbb{R}^2$ is the line segment AB. Right Panel: A probability simplex in $\mathbb{R}^3$ is the triangular face CDE of the tetrahedron $OCDE$.

9.5 Some convex sets in Economics

1. *Feasible payoff set in a bilateral trade/bargaining problem.*

Suppose X is the underlying set of physical alternatives in some euclidean space, and $\mathbf{d} \in X$ be a distinguished status quo point. For player/trader $i = 1, 2$, let $u_i : X \mapsto \mathbb{R}$ be her utility function. In a bilateral trade problem, the set X is the set of commodity allocations while $\mathbf{d}$ is interpreted as the no-trade allocation; in a bilateral bargaining problem, the set X is the set of contractual outcomes while $\mathbf{d}$ is interpreted as the disagreement outcome.

Under the assumption that X is convex and u_1 and u_2 are concave, the feasible payoff set defined by

$$F = \{(v_1, v_2) \in \mathbb{R}^2 : \text{ there is an } \mathbf{x} \in X \text{ such that } u_1(\mathbf{d}) \leq v_1 \leq u_1(\mathbf{x}) \text{ and } u_2(\mathbf{d}) \leq v_2 \leq u_2(\mathbf{x})\}$$

is convex. To show this, take two points (v_1, v_2) and (w_1, w_2) in F, and a scalar $\theta \in [0, 1]$. Then we know that

$$\text{there is an } \mathbf{x} \in X \text{ such that } u_1(\mathbf{d}) \leq v_1 \leq u_1(\mathbf{x}) \text{ and } u_2(\mathbf{d}) \leq v_2 \leq u_2(\mathbf{x}) \tag{9.2}$$

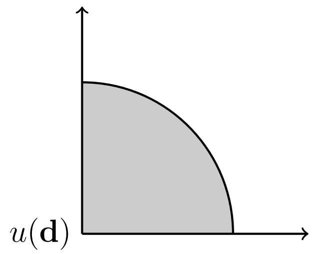

Figure 9.8: An example of a feasible payoff set in a bilateral trade/bargaining problem.

$$\text{there is a } \mathbf{y} \in X \text{ such that } u_1(\mathbf{d}) \le w_1 \le u_1(\mathbf{y}) \text{ and } u_2(\mathbf{d}) \le w_2 \le u_2(\mathbf{y}) \tag{9.3}$$

Multiply the inequalities in (9.2) by θ, those in (9.3) by $1-\theta$, add up, and use concavity of u_1 and u_2 to write

$$\begin{aligned} u_1(\mathbf{d}) &\le \theta v_1 + (1-\theta)w_1 \le \theta u_1(\mathbf{x}) + (1-\theta)u_2(\mathbf{y}) \le u_1(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \qquad \text{and} \\ u_2(\mathbf{d}) &\le \theta v_2 + (1-\theta)w_2 \le \theta u_2(\mathbf{x}) + (1-\theta)u_2(\mathbf{y}) \le u_2(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \end{aligned} \tag{9.4}$$

Since X is convex, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in X$. (9.4) then implies that $\theta(v_1, v_2) + (1-\theta)(w_1, w_2) \in F$. So F is convex.

2. A player's feasible set of mixed strategies in a finite strategic game. Consider, for instance, the Prisoners' Dilemma game shown in Table 9.1. For any player, the set of pure strategies is the finite set $S = \{C, D\}$ while the set of mixed strategies is

$$\Delta(S) = \{(p_1, p_2) \in [0,1] \times [0,1] : p_1 + p_2 = 1\}$$

where p_1 is the probability with which she plays C. Notice that $\Delta(S)$ is a probability simplex in $\mathbb{R}^2$.

		Player 2	
		C	D
Player 1	C	$2,2$	$0,3$
	D	$3,0$	$1,1$

Table 9.1: Prisoners' Dilemma

3. Cooperative payoff set in a finite strategic game.

The cooperative payoff set in a strategic game is the set of payoffs achievable by means of correlated strategies. A correlated strategy in a strategic game is simply a distribution over the set of pure strategy profiles in the game. For instance, in the Prisoners' Dilemma game shown in Table 9.1, the set of correlated strategies in the game is

$$\Delta(A) = \{(p_1, p_2, p_3, p_4) \in [0,1] \times [0,1] \times [0,1] \times [0,1] : p_1 + p_2 + p_3 + p_4 = 1\}$$

		Player 2 C	Player 2 D
Player 1	C	p_1	p_4
	D	p_3	p_2

Table 9.2: Correlated strategy in Prisoners' Dilemma

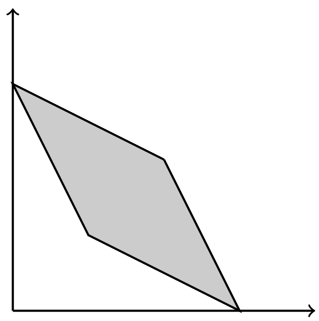

Figure 9.9: Cooperative payoff set of Prisoners' Dilemma.

where the probability weight on each pure strategy profile can be read off conveniently from Table 9.2.

The set of payoffs achievable by means of correlated strategies is then given by

$$\{p_1(2,2)+p_2(1,1)+p_3(3,0)+p_4(0,3) : (p_1,p_2,p_3,p_4) \in \Delta(A)\} = \mathbf{co}(\{(2,2),(1,1),(3,0),(0,3)\})$$

and is shown in Figure 9.9.

9.6 Convexity preserving operations on convex sets

Given a basic library of convex sets, we can create new convex sets using some simple operations that preserve convexity.

Rule 1. Arbitrary intersection of convex sets is convex

Let $\{C_i\}_{i\in I}$ be a family of convex sets where I is an arbitrary index set that can be finite or continuous. Then Rule 1 says that their intersection $C := \cap_{i\in I} C_i$ is a convex set.

The validity of this rule is easy to establish by simply testing the set C against the definition of a convex set. This is left as an exercise.

Example 1. A closed polyhedron is defined as the solution of a linear inequality system, and is given by $P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$ where A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.

For $i = 1, \ldots, n$, let $\mathbf{a}_i \in \mathbb{R}^n$ be the rows of the matrix A (viewed as column vectors). Then the closed polyhedron can be alternatively expressed as

$$P = \{\mathbf{x} \in \mathbb{R}^n : \text{for every } i = 1, \ldots, n, \quad \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i\}$$

Viewed this way, P is the intersection of finitely many closed halfspaces, all of which are convex sets. Hence, P is also a convex set.

Rule 2. Cartesian product of convex sets is convex

For $i = 1, \ldots, k$, let C_i be convex sets in $\mathbb{R}^{n_i}$. Then Rule 2 says that their cartesian product $C := C_1 \times \ldots \times C_k$ is a convex set.

The validity of this rule is easy to establish by simply testing the set C against the definition of a convex set. This is left as an exercise.

Rule 3. Image of a convex set under an affine function is convex. Similarly, pre-image of a convex set under an affine function is convex.

We know that an affine function f from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be written as $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, where A is an $m \times n$ matrix, $\mathbf{b}$ is some fixed vector in $\mathbb{R}^m$ and $\mathbf{x}$ is a general domain point in $\mathbb{R}^n$. Suppose then that C is a convex set in $\mathbb{R}^n$. Then the image $f(C) := \{f(\mathbf{x}) : \mathbf{x} \in S\}$ is a convex set in $\mathbb{R}^m$. To check this, simply verify the definition. Take two points $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$ in the image set $f(C)$, and a scalar $\theta \in [0, 1]$.

$$\begin{aligned} \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) &= \theta(A\mathbf{x}_1 + b) + (1-\theta)(A\mathbf{x}_2 + \mathbf{b}) \\ &= A(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + \mathbf{b} \\ &= f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \end{aligned}$$

Since C is convex, the point $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C$; so $f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \in f(C)$. Hence $f(C)$ is convex.

The conclusion that pre-image of a convex set under an affine function is convex can be established on similar lines. Some applications of Rule 3 are:

Rule 3a. Scaling and translations preserve the convexity of a set.

If C is a convex set in $\mathbb{R}^n$, then for any real scalar k, the scaled set $kC := \{k\mathbf{x} : \mathbf{x} \in C\}$ is a convex set. This is an immediate application of Rule 3 because kC is the image of the convex set C under the linear map $f(\mathbf{x}) = k\mathbf{x}$.

Similarly, for any vector $\mathbf{a}$ the translated set $C + \mathbf{a} := \{\mathbf{x} + \mathbf{a} : \mathbf{x} \in C\}$ is a convex set. This again is an application of Rule 3 because $C + \mathbf{a}$ is the image of the convex set C under the linear map $f(\mathbf{x}) = \mathbf{x} + \mathbf{a}$.

Rule 3b. Slice of a convex set is convex. Shadow of a convex set is convex.

Let C be a convex set in $\mathbb{R}^n \times \mathbb{R}^m$. Let $p : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ be the projection mapping defined by $p(\mathbf{x}, \mathbf{y}) = \mathbf{x}$. Then the slice of C along $\mathbf{y}$ is defined as the set

$$C(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}, \mathbf{y}) \in C\}$$

Applying Rule 3, $C(\mathbf{y})$ is convex as the image of the convex set C under the restricted linear map $p(\bullet, \mathbf{y})$.

The shadow or the projection of C onto $\mathbb{R}^n$ is defined as the set

$$C_1 = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}, \mathbf{y}) \in C \text{ for some } \mathbf{y} \in C\}$$

Applying Rule 3, C_1 is convex as the image of the convex set C under the linear map $p(\bullet, \bullet)$.

Rule 3c. The generalized Minkowski sum of two convex sets is convex.

Let C_1 and C_2 be two convex sets in $\mathbb{R}^n$. Then for given real scalars k_1 and k_2, their generalized Minkowski sum is defined as

$$k_1 C_1 + k_2 C_2 := \{k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2 : \mathbf{x}_1 \in C_1 \text{ and } \mathbf{x}_2 \in C_2\}$$

Rule 3c asserts that $k_1 C_1 + k_2 C_2$ is a convex set. This is true by Rule 3 because $k_1 C_1 + k_2 C_2$ is the image of the convex set $C_1 \times C_2$ (by Rule 2) under the linear map $f(\mathbf{x}_1, \mathbf{x}_2) = k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2$.

Chapter 10

Convex Functions

Definition 49. *(On restricted domain). Suppose $C \subset \mathbb{R}^n$ is a nonempty convex set. A function $f : C \mapsto \mathbb{R}$ is convex on C if for any two points $\mathbf{x}$ and $\mathbf{y}$ in C, and for any scalar $\theta \in [0, 1]$, the following inequality holds*

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \tag{10.1}$$

The defining inequality for a convex function has a geometric content which is depicted in Figure 10.1 - the line segment between the two points $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ on the graph $\mathbf{gr}(f)$ of the function f is always above that part of $\mathbf{gr}(f)$ whose points have their first n coordinates as some convex combination of $\mathbf{x}$ and $\mathbf{y}$.

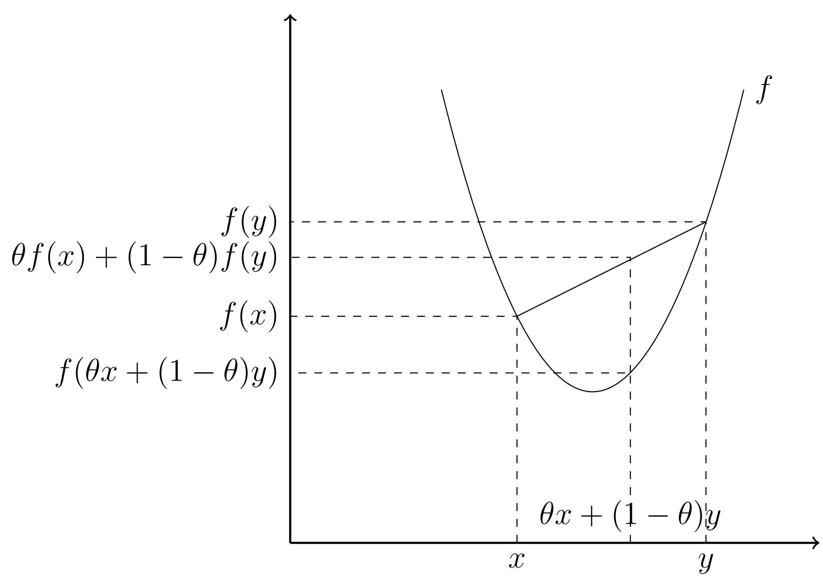

Figure 10.1: Geometric content of Definition 1 illustrated by a univariate convex function f.

The convexity requirement on C in Definition 49 is needed for the point $\theta\mathbf{x} + (1-\theta)\mathbf{y}$ that appears on the left side of (10.1) to be a domain point of f. Even so, it is convenient sometimes to have a definition of a convex function on the unrestricted domain $\mathbb{R}^n$. This can be done if we let f take the value ∞ at some points. We now define convex functions with unrestricted domain.

Definition 50. *(On unrestricted domain). A function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$, not identically ∞, is convex if for any two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, and for any scalar $\theta \in [0, 1]$, the following inequality holds (when considered in $\mathbb{R} \cup \{\infty\}$)*

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}) \tag{10.2}$$

The (effective) domain of a function f which is convex in the sense of Definition 50 is the set

$$\mathbf{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < \infty\} \tag{10.3}$$

Example 28. Consider the univariate function $f(x) = -\log x$, which is a convex function on its domain $\mathbb{R}_{++}$. We might want to make its domain unrestricted in a way that still keeps the function convex. To do this, extend the function f to $\hat{f} : \mathbb{R} \mapsto \mathbb{R} \cup \{\infty\}$ defined as

$$\hat{f}(x) = \begin{cases} -\log x & \text{if } x > 0 \\ \infty & \text{if } x \leq 0 \end{cases}$$

To see that $\hat{f}$ is convex on $\mathbb{R}$, take two points x and y on the real line. Check (10.2) for various possible cases - (a) $x > 0$ and $y > 0$, (b) $x \leq 0$ and $y > 0$, (c) $x \leq 0$ and $y \leq 0$.

It is important to realize that a function f which is convex in the sense of Definition 49 is also convex in the sense of Definition 50 and vice versa. To see this, suppose f is convex in the sense of Definition 49. To make it convex in the sense of Definition 50, simply extend f to entire $\mathbb{R}^n$ by letting $f(\mathbf{x}) = \infty$ when $\mathbf{x} \notin C$. We need to verify (10.2) for this extended function. There are two cases - either $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C$ or $\theta\mathbf{x} + (1-\theta)\mathbf{y} \notin C$. In the former case, (10.2) follows from (10.1) whenever $\mathbf{x}$ and $\mathbf{y}$ are both in C, otherwise it follows because the left side of (10.2) is finite while the right side is ∞. In the latter case, due to convexity of C, it must be that one of $\mathbf{x}$ and $\mathbf{y}$ are not in C. Consequently, (10.2) follows because both its left and right sides are ∞. Suppose now that f is convex in the sense of Definition 50. To make it convex in the sense of Definition 49, choose $C = \mathbf{dom}(f)$.

10.1 Epigraphs and sublevel sets

In the study of multivariate functions, the notion of graph $\mathbf{gr}(f)$ of a function and that of level sets $\mathbf{L}_t(f)$ (for any real t) of a function are two fundamental ideas. These sets associated with $f : \mathbb{R}^n \mapsto \mathbb{R}$ are defined as

$$\begin{aligned} \mathbf{gr}(f) &= \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : t = f(\mathbf{x})\} \\ \mathbf{L}_t(f) &= \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = t\} \end{aligned}$$

For convex functions, however, certain other related sets are much more important. These are first defined for general multivariate functions below.

Definition 51. *(Epigraph). The epigraph of a function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$, not identically ∞, is the nonempty set*

$$\mathbf{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : t \geq f(\mathbf{x})\} \tag{10.4}$$

Definition 52. *(Sublevel sets). The sublevel set of a function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$, not identically ∞, at level $t \in \mathbb{R}$ is the (possibly empty) set*

$$\mathbf{S}_t(f) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq t\} \tag{10.5}$$

(10.4) and (10.5) imply the following relation between these sets

$$(\mathbf{x}, t) \in \mathbf{epi}(f) \quad \text{if and only if} \quad \mathbf{x} \in \mathbf{S}_t(f) \tag{10.6}$$

(10.6) suggests that slicing the epigraph $\mathbf{epi}(f)$ along t (and projecting it onto $\mathbb{R}^n$) will give the t-sublevel set $\mathbf{S}_t(f)$.

For general multivariate functions, there is hardly any structure we can ascribe to these sets. For convex functions though, these sets turn out to be convex ! In fact, the convexity of epigraph is a complete characterization of convexity of the associated function. Theorem 34 and Theorem 35 establish these claims.

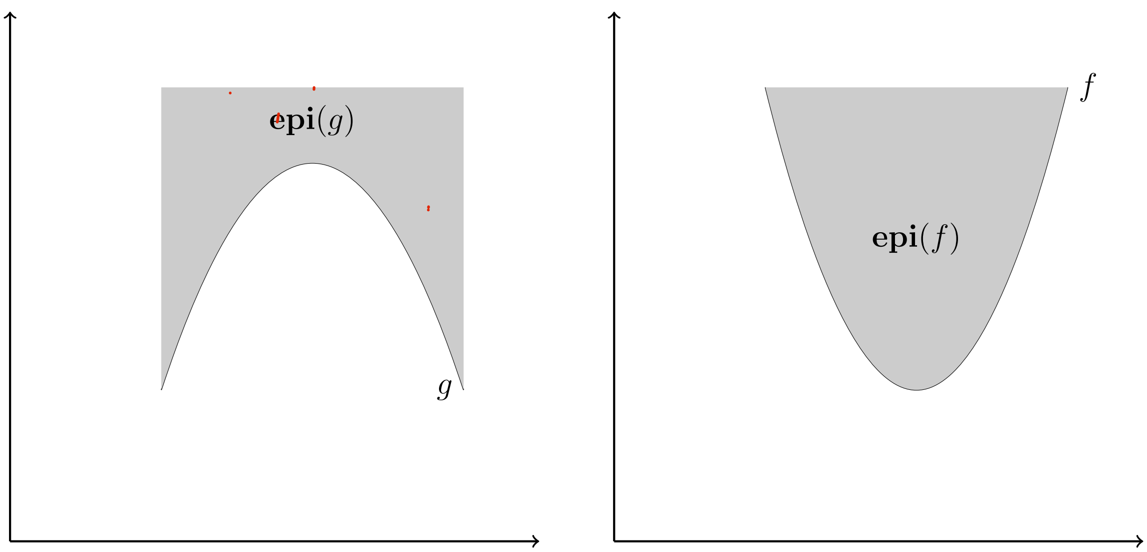

Figure 10.2: Left Panel: epigraph of a nonconvex (in fact, a concave) univariate function. Right Panel: epigraph of a univariate convex function f.

Theorem 34. *(Epigraph characterization of convex functions). Let the function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$ be not identically equal to ∞. Then f is a convex function if and only if $\mathbf{epi}(f)$ is a convex set.*

Proof. 'only if' part: Take two points $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$ in $\mathbf{epi}(f)$, and a scalar $\theta \in [0,1]$. Then the θ-convex combination of these two points is the point $\theta(\mathbf{x}_1, t_1) + (1-\theta)(\mathbf{x}_2, t_2) = (\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta t_1 + (1-\theta)t_2)$. This point lies in $\mathbf{epi}(f)$ because

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) \leq \theta t_1 + (1-\theta)t_2 \tag{10.7}$$

where, the first inequality in the chain is because f is convex, and the second because $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$ are points in $\mathbf{epi}(f)$. This proves the forward direction of the proposition.

'if' part: Take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and a scalar $\theta \in [0,1]$. Set $t_1 := f(\mathbf{x}_1)$ and $t_2 := f(\mathbf{x}_2)$. Then the points $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$ are in $\mathbf{epi}(f)$. Since $\mathbf{epi}(f)$ is a convex set, the point $\theta(\mathbf{x}_1, t_1) + (1-\theta)(\mathbf{x}_2, t_2) = (\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta t_1 + (1-\theta)t_2)$ is a point in $\mathbf{epi}(f)$. But this, by the definition of $\mathbf{epi}(f)$ means

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta t_1 + (1-\theta)t_2 = \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) \tag{10.8}$$

The ordering of expressions at the extreme ends of (10.8) implies that f is convex. This proves the reverse direction of the proposition. Q.E.D.

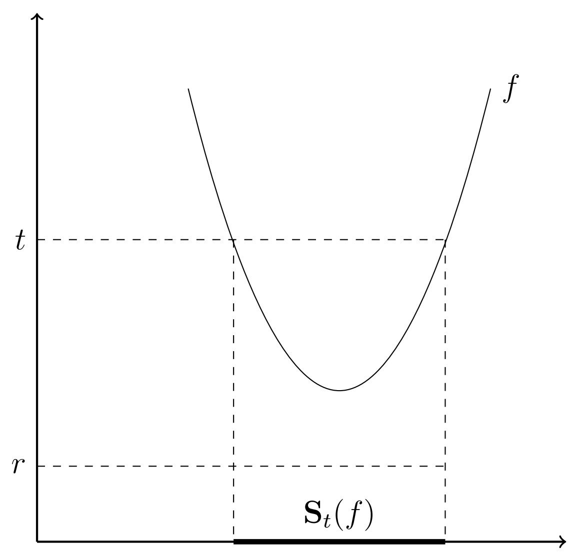

Figure 10.3: The t-sublevel set of a univariate convex function f is shown in dark as an interval, a convex set. The r-sublevel set, however, is an empty set as f always takes values above r. Nevertheless, the r-sublevel set is trivially convex.

Theorem 35. *(Sublevel sets of convex functions are convex). Let the function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$ be not identically equal to ∞. If f is a convex function, then for any $t \in \mathbb{R}$, the sublevel set $\mathbf{S}_t(f)$ is a convex set.*

Proof. Fix $t \in \mathbb{R}$. Take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbf{S}_t(f)$, and a scalar $\theta \in [0,1]$. Then

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) \leq \theta t + (1-\theta)t = t \tag{10.9}$$

where the first inequality is because f is convex, and the second because $\mathbf{x}_1$ and $\mathbf{x}_2$ are points in $\mathbf{S}_t(f)$. The ordering of expressions at the extreme ends of (10.9) implies that the point $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$ is in the sublevel set $\mathbf{S}_t(f)$. This proves it is a convex set. Q.E.D.

The epigraph characterization of convex functions immediately yields Jensen's inequality - a generalization of the defining inequality (10.1) for convex functions.

Theorem 36. *(Jensen's inequality). Let the function $f : \mathbb{R}^n \mapsto \mathbb{R}\cup\{\infty\}$ be not identically equal to ∞. If f is a convex function, then for any finite set of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in* ***dom***(f)*, and for scalars $\theta_1, \ldots, \theta_k$ such that for every $i = 1, \ldots, k$, $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$, we have*

$$f\Big(\sum_{i=1}^k \theta_i \mathbf{x}_i\Big) \leq \sum_{i=1}^k \theta_i f(\mathbf{x}_i) \tag{10.10}$$

Proof. The k points $(\mathbf{x}_1, f(\mathbf{x}_1)), \ldots, (\mathbf{x}_k, f(\mathbf{x}_k))$, being points in $\mathbf{gr}(f)$, are also in $\mathbf{epi}(f)$. Since f is convex, by Theorem 34, $\mathbf{epi}(f)$ is a convex set. Hence, the point

$$\theta_1(\mathbf{x}_1, f(\mathbf{x}_1)) + \ldots + \theta_k(\mathbf{x}_k, f(\mathbf{x}_k)) = \big(\theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k, \theta_1 f(\mathbf{x}_1) + \ldots + \theta_k f(\mathbf{x}_k)\big) \tag{10.11}$$

also lies in $\mathbf{epi}(f)$. But this implies that Jensen's inequality (10.10) holds. Q.E.D.

10.2 Differentiable convex functions

10.2.1 First order characterization

If a convex function from $C \subset \mathbb{R}^n$ to $\mathbb{R}$ is differentiable, then using the derivative information about f at any point $\mathbf{x} \in C$, we can obtain an estimate of f at any other point $\mathbf{y} \in C$. This affine estimate, we know, is read off on the tangent hyperplane to f at $\mathbf{x}$ and is given by

$$\text{AffineApproximation }(\mathbf{y}) = f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

The first order condition for convex functions says that regardless of the point $\mathbf{y}$, the affine estimate is always an underestimate of the actual function value $f(\mathbf{y})$.

$$\text{First order condition : } \quad \text{for every } \mathbf{y}, \quad f(\mathbf{y}) \geq \text{AffineApproximation }(\mathbf{y})$$

In other words, if we define

$$\text{ApproximationError }(\mathbf{y}) = f(\mathbf{y}) - \text{AffineApproximation }(y)$$

then the first order condition for convex functions can be recast in this language as

$$\text{First order condition : } \quad \text{for every } \mathbf{y}, \quad \text{ApproximationError }(\mathbf{y}) \geq 0$$

Figure 10.4 emphasizes these underestimates at two points y'' and y' for a univariate convex function by showing the extent of the positive approximation errors by downward arrows.

where the first inequality is because f is convex, and the second because $\mathbf{x}_1$ and $\mathbf{x}_2$ are points in $\mathbf{S}_t(f)$. The ordering of expressions at the extreme ends of (10.9) implies that the point $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$ is in the sublevel set $\mathbf{S}_t(f)$. This proves it is a convex set. Q.E.D.

The epigraph characterization of convex functions immediately yields Jensen's inequality - a generalization of the defining inequality (10.1) for convex functions.

Theorem 36. *(Jensen's inequality). Let the function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$ be not identically equal to ∞. If f is a convex function, then for any finite set of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in **dom**(f), and for scalars $\theta_1, \ldots, \theta_k$ such that for every $i = 1, \ldots, k$, $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$, we have*

$$f\Big(\sum_{i=1}^{k} \theta_i \mathbf{x}_i\Big) \leq \sum_{i=1}^{k} \theta_i f(\mathbf{x}_i) \tag{10.10}$$

Proof. The k points $(\mathbf{x}_1, f(\mathbf{x}_1)), \ldots, (\mathbf{x}_k, f(\mathbf{x}_k))$, being points in $\mathbf{gr}(f)$, are also in $\mathbf{epi}(f)$. Since f is convex, by Theorem 34, $\mathbf{epi}(f)$ is a convex set. Hence, the point

$$\theta_1(\mathbf{x}_1, f(\mathbf{x}_1)) + \ldots + \theta_k(\mathbf{x}_k, f(\mathbf{x}_k)) = \big(\theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k, \theta_1 f(\mathbf{x}_1) + \ldots + \theta_k f(\mathbf{x}_k)\big) \tag{10.11}$$

also lies in $\mathbf{epi}(f)$. But this implies that Jensen's inequality (10.10) holds. Q.E.D.

10.2 Differentiable convex functions

10.2.1 First order characterization

If a convex function from $C \subset \mathbb{R}^n$ to $\mathbb{R}$ is differentiable, then using the derivative information about f at any point $\mathbf{x} \in C$, we can obtain an estimate of f at any other point $\mathbf{y} \in C$. This affine estimate, we know, is read off on the tangent hyperplane to f at $\mathbf{x}$ and is given by

$$\text{AffineApproximation }(\mathbf{y}) = f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

The first order condition for convex functions says that regardless of the point $\mathbf{y}$, the affine estimate is always an underestimate of the actual function value $f(\mathbf{y})$.

$$\text{First order condition :} \quad \text{for every } \mathbf{y}, \quad f(\mathbf{y}) \geq \text{AffineApproximation }(\mathbf{y})$$

In other words, if we define

$$\text{ApproximationError }(\mathbf{y}) = f(\mathbf{y}) - \text{AffineApproximation }(y)$$

then the first order condition for convex functions can be recast in this language as

$$\text{First order condition :} \quad \text{for every } \mathbf{y}, \quad \text{ApproximationError }(\mathbf{y}) \geq 0$$

Figure 10.4 emphasizes these underestimates at two points y'' and y' for a univariate convex function by showing the extent of the positive approximation errors by downward arrows.

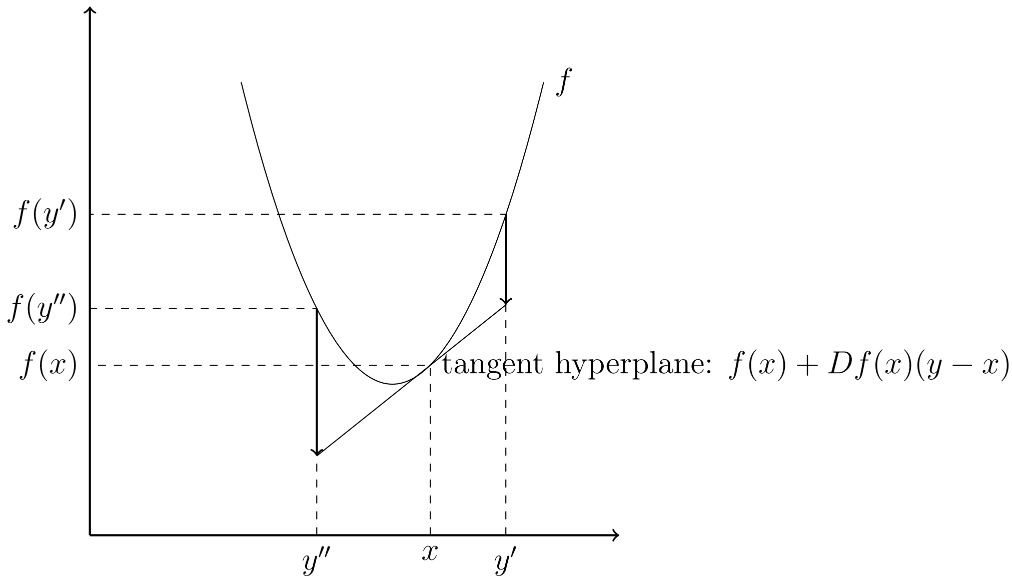

Figure 10.4: Geometric content of first order characterization of convex functions: for any point y, the affine approximation of $f(y)$ on the tangent hyperplane through any point x is an underestimate.

An important geometric implication of the first order conditions is that the graph of a differentiable convex function always lies weakly above the tangent hyperplane at any point on the graph.

Theorem 37. *(First order characterization of differentiable convex functions). Let C be a nonempty convex set in $\mathbb{R}^n$ and the function $f : C \mapsto \mathbb{R}$ be differentiable on C. Then f is a convex function on C if and only if*

$$\text{for any two points } \mathbf{x}, \mathbf{y} \in C, \quad f(\mathbf{y}) \geq f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{10.12}$$

Proof. 'only if' part: Take two points $\mathbf{x}, \mathbf{y} \in C$ and a scalar $\theta \in (0, 1)$. Since f is convex, we have the inequality

$$f\big(\theta\mathbf{y} + (1-\theta)\mathbf{x}\big) \leq \theta f(\mathbf{y}) + (1-\theta)f(\mathbf{x}) \tag{10.13}$$

This is equivalently rewritten as

$$f\big(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})\big) - f(\mathbf{x}) \leq \theta\big(f(\mathbf{y}) - f(\mathbf{x})\big) \tag{10.14}$$

Divide by θ and take limits as $\theta \to 0$ to get

$$Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) = D_{\mathbf{y}-\mathbf{x}}f(\mathbf{x}) := \lim_{\theta \to 0} \frac{f\big(\mathbf{x} + \theta(\mathbf{y} - \mathbf{x})\big) - f(\mathbf{x})}{\theta} \leq f(\mathbf{y}) - f(\mathbf{x}) \tag{10.15}$$

This proves the forward direction of the proposition.

'if' part: Take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in C, and a scalar $\theta \in [0,1]$. Since C is a convex set, the point $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$ is a point in C. In the first step, use (10.12) to estimate the function value at $\mathbf{x}_1$ using the derivative information at $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$ as

$$f(\mathbf{x}_1) \geq f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + Df(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big(\mathbf{x}_1 - (\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big) \tag{10.16}$$

In the second step, use (10.12) to estimate the function value at $\mathbf{x}_2$ using the derivative information at $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$ as

$$f(\mathbf{x}_2) \geq f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + Df(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big(\mathbf{x}_2 - (\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big) \tag{10.17}$$

Taking the convex combination θ (10.16) $+(1-\theta)$ (10.17) of these inequalities, and making use of the fact that the derivative $Df(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big(\bullet\big)$ is a linear map, gives

$$\theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) \geq f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + Df(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big(\mathbf{0}\big) = f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \tag{10.18}$$

The inequality between the expressions at the extreme ends of (10.18) is the defining inequality for convex functions, implying that f is convex. This proves the reverse direction of the proposition. Q.E.D.

10.2.2 Second order characterization

The second order conditions allow us to recognize multivariate convex functions that are twice differentiable by checking their Hessian matrix for positive semi-definiteness at every point. This can get unwieldy for many functions whose Hessian matrix varies across points in complicated ways. However, for functions whose Hessian matrix can be inspected without great difficulty for positive semi-definiteness, the second order conditions are quite handy.

In this section, we want to understand why the second order conditions hold. We first develop a simple result that helps reduce the task of checking the convexity of a multivariate function to that of checking the convexity of a univariate function. This result will be put to use in developing the important second order characterization of twice differentiable functions.

Suppose $C \subset \mathbb{R}^n$ is a convex set. Recall that given a point $\mathbf{x}_0 \in C$ and a direction $\mathbf{d} \in \mathbb{R}^n$, a line passing through $\mathbf{x}_0$ in the direction $\mathbf{d}$ is parameterized as $\mathbf{x}_0 + t\mathbf{d}$ where $t \in \mathbb{R}$. Let $T = \{t \in \mathbb{R} : \mathbf{x}_0 + t\mathbf{d} \in C\}$. Then T is a subset of real line.

Definition 53. *A univariate function $\phi : T \mapsto \mathbb{R}$ is a restriction of a multivariate function $f : C \mapsto \mathbb{R}$ on the line $\mathbf{x}_0 + t\mathbf{d}$ if $\phi(t) = f(\mathbf{x}_0 + t\mathbf{d})$.*

By rules of calculus, we have

$$\phi'(t) = Df(\mathbf{x}_0 + t\mathbf{d})\mathbf{d} \tag{10.19}$$
$$\phi''(t) = \mathbf{d}^T D^2 f(\mathbf{x}_0 + t\mathbf{d})\mathbf{d} \tag{10.20}$$

Lemma 1. *(Simplification Result). Let C be a nonempty convex set in $\mathbb{R}^n$. A multivariate function $f : C \mapsto \mathbb{R}$ is convex on C if and only if its univariate restriction $\phi : T \mapsto \mathbb{R}$ is convex on T.*

Proof. 'only if' part: Fix a line by fixing a point $\mathbf{x}_0 \in C$ and a direction $\mathbf{d} \in \mathbb{R}^n$. Take two points t_1 and t_2 in T and a scalar $\theta \in [0, 1]$. Then

$$\begin{aligned}
\phi(\theta t_1 + (1-\theta)t_2) &= f\big(\mathbf{x}_0 + (\theta t_1 + (1-\theta)t_2)\mathbf{d}\big) \\
&= f\big(\theta(\mathbf{x}_0 + t_1\mathbf{d}) + (1-\theta)(\mathbf{x}_0 + t_2\mathbf{d})\big) \\
&\leq \theta f(\mathbf{x}_0 + t_1\mathbf{d}) + (1-\theta)f(\mathbf{x}_0 + t_2\mathbf{d}) \qquad \text{(by convexity of } f\text{)} \\
&= \theta\phi(t_1) + (1-\theta)\phi(t_2)
\end{aligned}$$

This proves the convexity of ϕ, and thereby the forward direction of the proposition.

'if' part: Take two points $\mathbf{x}$ and $\mathbf{y}$ in C, and a scalar $\theta \in [0, 1]$. Then

$$\begin{aligned}
f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= f\big(\mathbf{y} + \theta(\mathbf{x} - \mathbf{y})\big) \\
&= \phi(\theta) \\
&= \phi(\theta 1 + (1-\theta)0) \\
&\leq \theta\phi(1) + (1-\theta)\phi(0) \qquad \text{(by convexity of } \phi\text{)} \\
&= \theta f\big(\mathbf{y} + 1.(\mathbf{x} - \mathbf{y})\big) + (1-\theta)f\big(\mathbf{y} + 0.(\mathbf{x} - \mathbf{y})\big) \\
&= \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})
\end{aligned}$$

where in the penultimate equality, in passing from ϕ to f, we have chosen the point as $\mathbf{y}$ and the direction as $\mathbf{x} - \mathbf{y}$. This proves the convexity of f, and thereby the reverse direction of the proposition. Q.E.D.

Theorem 38. *(Second order characterization of twice differentiable convex functions). Let C be a nonempty convex set in $\mathbb{R}^n$ and the function $f : C \mapsto \mathbb{R}$ be twice differentiable on C. Then f is a convex function on C if and only if $D^2 f(\mathbf{x})$ is positive semidefinite for every $\mathbf{x} \in C$.*

Proof. 'only if' part: Fix $\mathbf{x} \in C$. Fix a direction $\mathbf{d} \in \mathbb{R}^n$ as well, thereby fixing a line. Since f is twice differentiable and convex on C, Lemma 1 implies its restriction ϕ on $T = \{t \in \mathbb{R} : \mathbf{x} + t\mathbf{d} \in C\}$ is twice differentiable and convex as well. Applying the second derivative characterization of twice differentiable and convex univariate functions, we have

$$\mathbf{d}^T D^2 f(\mathbf{x} + t\mathbf{d})\mathbf{d} = \phi''(t) \geq 0 \qquad \text{for every } t \in T$$

Since $0 \in T$, we have $\mathbf{d}^T D^2 f(\mathbf{x})\mathbf{d} \geq 0$ for arbitrarily chosen $\mathbf{d}$. This implies $D^2 f(\mathbf{x})$ is positive semidefinite, thereby proving the forward direction.

'if' part: Fix a line by fixing a point $\mathbf{x}_0 \in C$ and a direction $\mathbf{d} \in \mathbb{R}^n$. Since $D^2 f(\mathbf{x})$ is positive semidefinite for every $\mathbf{x} \in C$, we have

$$\phi''(t) = \mathbf{d}^T D^2 f(\mathbf{x}_0 + t\mathbf{d})\mathbf{d} \geq 0 \qquad \text{for every } t \in T$$

Applying the second derivative characterization of twice differentiable and convex univariate functions, we deduce that ϕ is a convex univariate function on T. Lemma 1 then implies the multivariate function f is convex on C. This proves the reverse direction. Q.E.D.

10.3 Basic library of univariate convex/concave functions

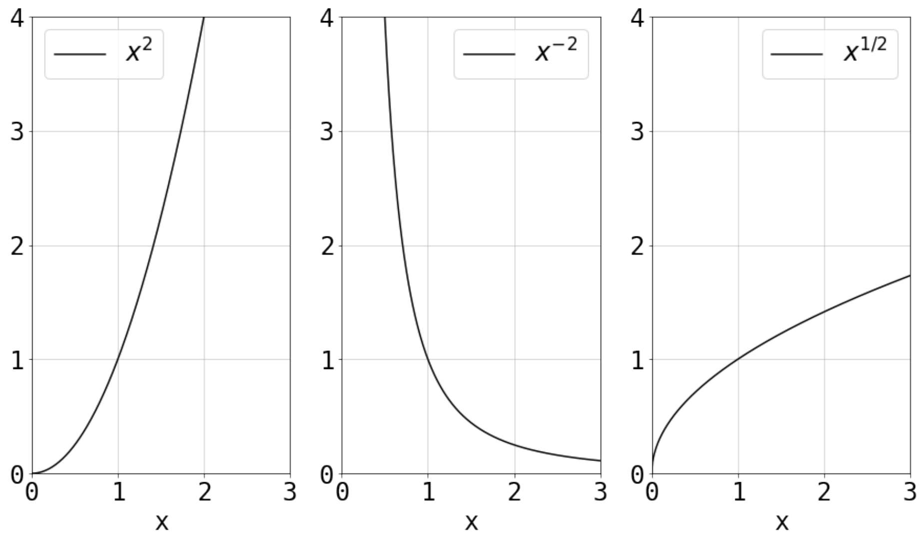

Figure 10.5: Power function x^p on $\mathbb{R}_+$ is convex when $p \geq 1$ or $p \leq 0$; and concave when $p \in [0, 1]$. The graphs here are drawn for $p = 2, -2, 1/2$. In CVXPY, this is the atomic function power(x, p). When $p = 1/2$, this is the same as the atomic function sqrt(x). Similarly, when $p = 2$, this is the same as the atomic function square(x).

'if' part: Fix a line by fixing a point $\mathbf{x}_0 \in C$ and a direction $\mathbf{d} \in \mathbb{R}^n$. Since $D^2 f(\mathbf{x})$ is positive semidefinite for every $\mathbf{x} \in C$, we have

$$\phi''(t) = \mathbf{d}^T D^2 f(\mathbf{x}_0 + t\mathbf{d})\mathbf{d} \geq 0 \qquad \text{for every } t \in T$$

Applying the second derivative characterization of twice differentiable and convex univariate functions, we deduce that ϕ is a convex univariate function on T. Lemma 1 then implies the multivariate function f is convex on C. This proves the reverse direction. Q.E.D.

10.3 Basic library of univariate convex/concave functions

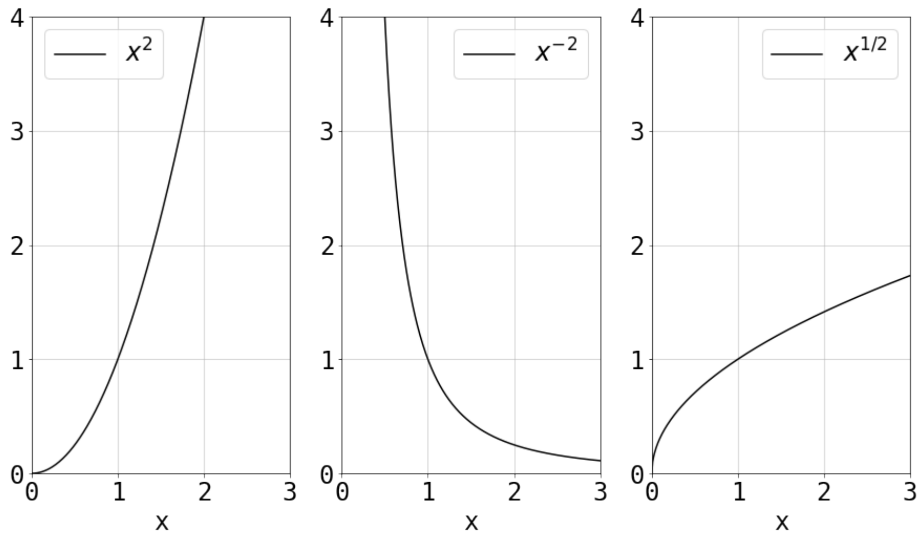

Figure 10.5: Power function x^p on $\mathbb{R}_+$ is convex when $p \geq 1$ or $p \leq 0$; and concave when $p \in [0, 1]$. The graphs here are drawn for $p = 2, -2, 1/2$. In CVXPY, this is the atomic function power(x, p). When $p = 1/2$, this is the same as the atomic function sqrt(x). Similarly, when $p = 2$, this is the same as the atomic function square(x).

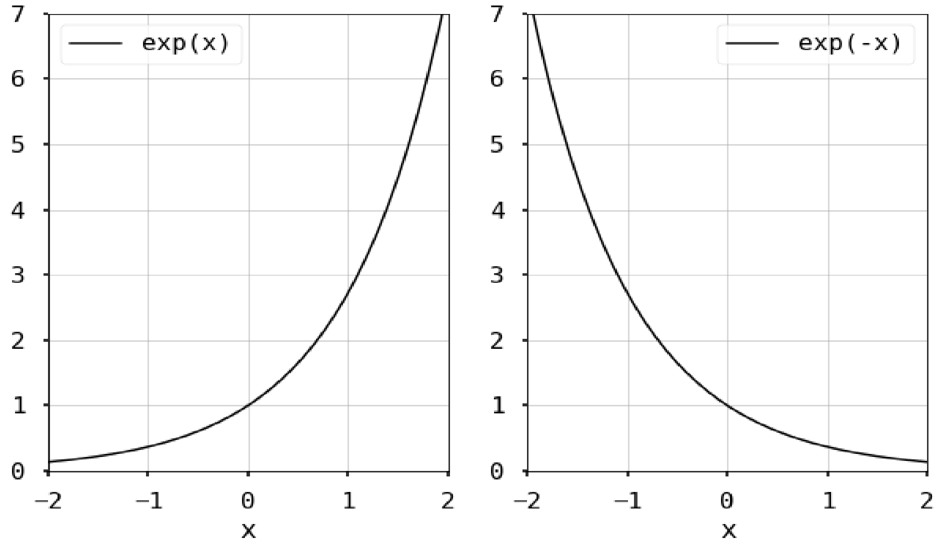

Figure 10.6: Exponential function $\exp(ax)$ is convex on $\mathbb{R}$. The graphs are drawn for $a = 1$ and $a = -1$. In CVXPY, this is the atomic function $\exp(x, p)$.

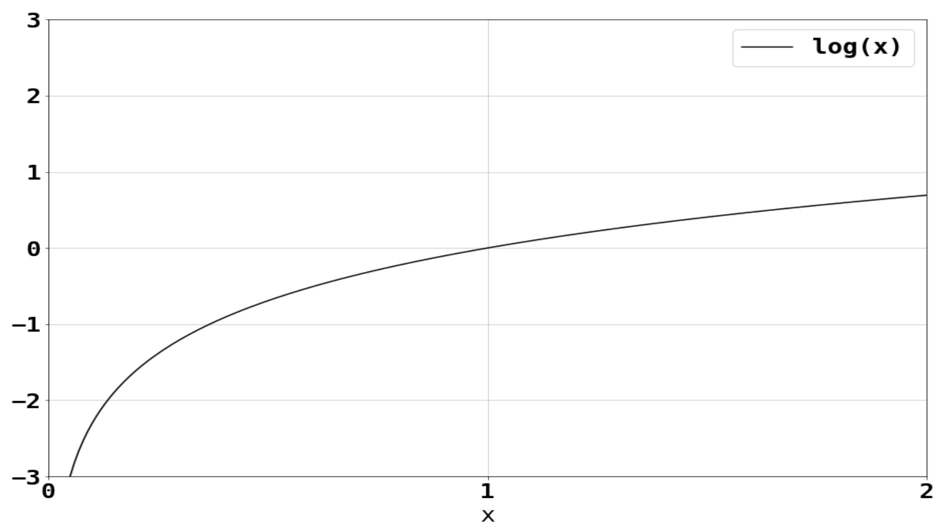

Figure 10.7: Logarithm function $\log(x)$ is concave on $\mathbb{R}_{++}$. In CVXPY, this is the atomic function $\log(x)$.

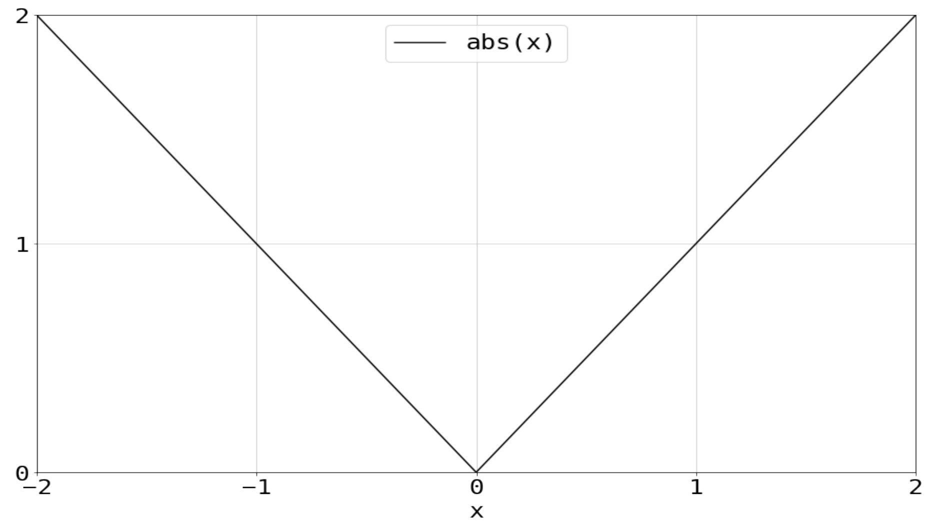

Figure 10.8: Absolute value function $\mathrm{abs}(x)$ is convex on $\mathbb{R}$. In CVXPY, this is the atomic function $\mathrm{abs}(x)$.

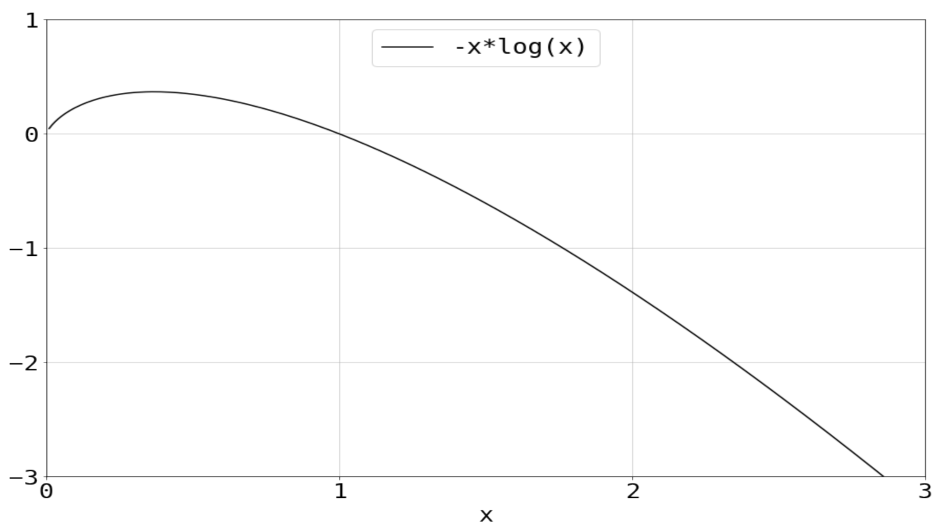

Figure 10.9: Entropy function $-x\log(x)$ is concave on $\mathbb{R}_{++}$. In CVXPY, this is the atomic function $\mathrm{entr}(x)$.

10.4 Basic library of multivariate convex/concave functions

1. *Linear functions.* A linear function on $\mathbb{R}^n$ is specified as

$$f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle = c_1 x_1 + \ldots + c_n x_n$$

for some given vector $\mathbf{c} \in \mathbb{R}^n$. By basic definitions, a linear function is both convex and concave. In CVXPY, this is written as the atomic function c @ x.

2. *Quadratic functions.* A quadratic function on $\mathbb{R}^n$ is specified as

$$f(\mathbf{x}) = \mathbf{x}^T P \mathbf{x}$$

for some given $n \times n$ matrix P. f is twice differentiable and $D^2 f(\mathbf{x}) = P$ regardless of $\mathbf{x}$. By the second order characterization of twice differentiable functions, f is convex on $\mathbb{R}^n$ if P is positive semidefinite; and concave on $\mathbb{R}^n$ if P is negative semidefinite. In CVXPY, this is written as the atomic function quad_form(x, P).

3. *Norm function.* A norm function $\|\bullet\| : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. To show this, simply verify the definition. Take two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and a scalar $\theta \in [0, 1]$. Then

$$\begin{aligned} \|\theta\mathbf{x} + (1-\theta)\mathbf{y}\| &\leq \|\theta\mathbf{x}\| + \|(1-\theta)\mathbf{y}\| && \text{(by triangle inequality of the norm)} \\ &\leq \theta \|\mathbf{x}\| + (1-\theta) \|\mathbf{y}\| && \text{(by homogeneity property of the norm)} \end{aligned}$$

This verifies that the norm is a convex function on $\mathbb{R}^n$. In CVXPY, the atomic functions for various norms are

L^2/Euclidean- norm :	norm(x, 2) to denote	$\sqrt{x_1^2 + \ldots + x_n^2}$
L^∞/max- norm :	norm(x, "inf") to denote	$\max(\lvert x_1\rvert, \ldots, \lvert x_n\rvert)$
L^1/absolute value- norm :	norm(x, 1) to denote	$\lvert x_1\rvert + \ldots + \lvert x_n\rvert$

4. *Max function.* The max function $f : \mathbb{R}^n \mapsto \mathbb{R}$ defined as

$$f(\mathbf{x}) = \max(x_1, \ldots x_n)$$

is convex on $\mathbb{R}^n$. Take two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and a scalar $\theta \in [0, 1]$. Then

$$\begin{aligned} f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \max_{i=1,\ldots,n} \theta x_i + (1-\theta) y_i \\ &\leq \theta \max_{i=1,\ldots,n} x_i + (1-\theta) \max_{i=1,\ldots,n} y_i \\ &= \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \end{aligned}$$

This verifies that the max function is a convex function on $\mathbb{R}^n$. In CVXPY, this is the atomic function max(x).

5. *Weighted geometric mean function.* Given a vector of positive weights $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n_{++}$, the weighted geometric mean function $f : \mathbb{R}^n_{++} \mapsto \mathbb{R}$ defined as

$$f(\mathbf{x}) = (x_1^{p_1} \ldots x_n^{p_n})^{1/p} \quad \text{where} \quad p := p_1 + \ldots + p_n$$

is concave on $\mathbb{R}^n_{++}$. We will show this using the weighted AM-GM inequality, which says that for a list of positive real numbers $(a_1, \ldots, a_n)$ and positive weights $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n_{++}$, where $p := p_1 + \ldots + p_n$, we have

$$\frac{p_1 a_1 + \ldots + p_n a_n}{p} \geq (a_1^{p_1} \ldots a_n^{p_n})^{1/p}$$

Take two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and a scalar $\theta \in [0,1]$. Apply the weighted AM-GM inequality to the list $\left(\frac{x_1}{\theta x_1+(1-\theta)y_1},\ldots,\frac{x_n}{\theta x_n+(1-\theta)y_n}\right)$ to get

$$\sum_{i=1}^{n}\left(\frac{p_i}{p}\right)\left(\frac{x_i}{\theta x_i+(1-\theta)y_i}\right) \geq \prod_{i=1}^{n}\left(\frac{x_i}{\theta x_i+(1-\theta)y_i}\right)^{p_i/p} = \frac{f(\mathbf{x})}{f(\theta\mathbf{x}+(1-\theta)\mathbf{y})} \tag{10.21}$$

Again, apply the weighted AM-GM inequality to the list $\left(\frac{y_1}{\theta x_1+(1-\theta)y_1},\ldots,\frac{y_n}{\theta x_n+(1-\theta)y_n}\right)$ to get

$$\sum_{i=1}^{n}\left(\frac{p_i}{p}\right)\left(\frac{y_i}{\theta x_i+(1-\theta)y_i}\right) \geq \prod_{i=1}^{n}\left(\frac{y_i}{\theta x_i+(1-\theta)y_i}\right)^{p_i/p} = \frac{f(\mathbf{y})}{f(\theta\mathbf{x}+(1-\theta)\mathbf{y})} \tag{10.22}$$

Now multiplying (10.21) by θ and (10.22) by $1-\theta$ and adding up, we get

$$1 = \sum_{i=1}^{n}\left(\frac{p_i}{p}\right) \geq \frac{\theta f(\mathbf{x})+(1-\theta)f(\mathbf{y})}{f(\theta\mathbf{x}+(1-\theta)\mathbf{y})} \tag{10.23}$$

This shows that f is concave. In CVXPY, this is the atomic function geo_mean(x, p).

10.5 Convex functions in Economics

1. *CARA utility function.* Consider the vNM utility function $u:\mathbb{R}_+ \mapsto \mathbb{R}$ defined on positive levels of wealth as

$$u(w) = -\exp(-aw)$$

where $a > 0$ is the Arrow-Pratt coefficient of absolute risk aversion for u. u is a concave univariate function on $\mathbb{R}_+$ because $\exp(-aw)$ is convex function of w.

2. *CRRA utility function.* Consider the vNM utility function $u:\mathbb{R}_+ \mapsto \mathbb{R}$ defined on positive levels of wealth as

$$u(w) = \frac{w^{1-r}}{1-r}$$

where $r(\neq 1) > 0$ is the Arrow-Pratt coefficient of relative risk aversion for u. When $r \in (0,1)$, u is concave as a positive scaling of the power function w^{1-r}, which is concave for such r. On the other hand, when $r > 1$, u is concave as a negative scaling of the power function w^{1-r}, which is convex for such r.

3. *Linear utility function.* Consider the vNM utility function $u:\mathbb{R}_+^2 \mapsto \mathbb{R}$ defined on nonnegative levels of consumption of two commodities as

$$u(x,y) = ax+by \quad \text{where} \quad a>0, b>0 \text{ are positive constants}$$

Being linear, u can of course be seen as a concave (or convex) function. In consumer theory, this utility function is used to model preferences over two commodities which are perfect substitutes in consumption. In welfare economics, this is used to model a utilitarian welfare function which equates social welfare to the aggregate sum of individual welfares but is neutral to the distribution of the aggregate welfare among individuals.

4. *Min function.* Consider the function $f : \mathbb{R}^2_+ \mapsto \mathbb{R}$ defined as

$$f(x, y) = \min(ax, by) \quad \text{where} \quad a > 0, b > 0 \text{ are positive constants}$$

Since $-f(x, y) = \max(-ax, -by)$ is a convex function, f is a concave function. In consumer theory, this utility function is used to model preferences over two commodities which are perfect complements in consumption. In production theory, this function is used to model a production technology that uses inputs in fixed proportions. In welfare economics, this is used to model a Rawlsian welfare function which equates social welfare to the welfare of the worst-off individual in society.

5. *Cobb-Douglas function.* Consider the function $f : \mathbb{R}^2_+ \mapsto \mathbb{R}$ defined as

$$f(x, y) = x^a y^b \quad \text{where} \quad a > 0, b > 0 \text{ are positive constants}$$

For $0 < a + b \leq 1$, f is a concave function. To see this,

$$\begin{aligned} \text{define } g : \mathbb{R}^2_+ \mapsto \mathbb{R}_+ \text{ by} \quad & g(x, y) := (x^a y^b)^{1/(a+b)} \\ \text{define } h : \mathbb{R}_+ \mapsto \mathbb{R}_+ \text{ by} \quad & h(t) := t^{a+b} \\ \text{it follows then that} \quad & f = hog \end{aligned}$$

Now for $0 < a + b \leq 1$, h is a nondecreasing concave function. g is a concave function as the weighted geometric mean over $\mathbb{R}^2_+$. Therefore, f is concave as a nondecreasing concave transformation of a concave function.

In consumer theory, this utility function is used to model preferences over two commodities which are imperfect complements in consumption. In production theory, this function is used to model a production technology that is either constant returns to scale $a + b = 1$ or decreasing returns to scale $a + b < 1$. In welfare economics and bargaining theory, a version of the Cobb-Douglas function, given by

$$N(u_1, u_2) = (u_1 - d_1)^a (u_2 - d_2)^b \quad \text{where} \quad a > 0, b > 0 \text{ are positive constants such that } a + b = 1$$

is knows as the generalized Nash product function for given disagreement point payoffs $\mathbf{d} = (d_1, d_2)$.

6. *Surplus in a bilateral trade problem.* Consider a bilateral trade problem involving quantities of a single commodity produced by a producer at cost $c(q)$ and demanded by a consumer with utility function $u(q)$. The trade surplus as a function of traded quantity q is given by

$$S(q) = u(q) - c(q)$$

Under the standard assumption that $u(.)$ is strictly increasing (consumer prefers more to less) and strictly concave (consumer has diminishing marginal utility) while $c(.)$ is strictly increasing (it costs more to produce more) and strictly convex (producer has increasing marginal costs), the trade surplus $S(.)$ is concave on $\mathbb{R}_+$ as the sum of two concave functions $u(.)$ and $-c(.)$.

7. *Monopolist's profit function.* Consider a monopoly market with demand function $Q(p)$ and linear costs cq with $c > 0$. The monopolist's profit as a function of the price it sets is given by

$$\pi(p) = (p - c)Q(p)$$

Under the assumptions that $Q(.)$ is twice differentiable, strictly decreasing and strictly concave, the profit function $\pi(.)$ is concave on $\mathbb{R}_+$. This is easily verified using the second order conditions for a univariate concave function:

$$\pi''(p) = (p - c)Q''(p) + 2Q'(p) < 0$$

8. *Cournot oligopolist's profit function.* Consider a symmetric oligopoly market (n firms) with inverse demand function $P(Q)$ and linear costs cq with $c > 0$. The oligopolist i's profit as a function of the quantity it produces is given by

$$\pi_i(q_i) = P(q_i + Q_{-i})q_i - cq_i$$

Under the assumptions that $P(.)$ is twice differentiable, strictly decreasing and strictly concave, the profit function $\pi_i(.)$ is concave on $\mathbb{R}_+$. This is easily verified using the second order conditions for a univariate concave function:

$$\pi_i''(q_i) = P''(q_i + Q_{-i})q_i + 2P'(q_i + Q_{-i}) < 0$$

9. *Risk-averse decision maker's utility function in a statistical decision problem.* Suppose a risk averse decision maker (DM) faces a statistical decision problem where he chooses an action $a \in \mathbb{R}$ which pays off a return that depends on the realization of a real-valued random variable X. If the DM is an vNM expected utility maximizer, than her expected utility

$$\mathbb{E}_X[u(a, X)]$$

is a concave function of her action a. This is argued as follows. Fix actions a_1 and a_2 and a scalar $\theta \in [0, 1]$. Then for every realization x of X, we have

$$\begin{aligned} u(\theta a_1 + (1-\theta)a_2, x) &\geq \theta u(a_1, x) + (1-\theta)u(a_2, x) && \text{(DM risk averse)} \\ \mathbb{E}_X[u(\theta a_1 + (1-\theta)a_2, X)] &\geq \mathbb{E}_X[\theta u(a_1, X) + (1-\theta)u(a_2, X)] && \text{(Expectation is a monotone operator)} \\ &= \theta\mathbb{E}_X[u(a_1, X)] + (1-\theta)\mathbb{E}_X[u(a_2, X)] && \text{(Expectation is a linear operator)} \end{aligned}$$

This shows that $\mathbb{E}_X[u(a, X)]$ is a concave function of her action a.

10.6 Convexity preserving operations on convex functions

Given a basic library of convex functions, we can create new convex functions using some simple operations that preserve convexity. In this chapter, we focus on a subset of these operations.

marginal costs), the trade surplus $S(.)$ is concave on $\mathbb{R}_+$ as the sum of two concave functions $u(.)$ and $-c(.)$.

7. *Monopolist's profit function.* Consider a monopoly market with demand function $Q(p)$ and linear costs cq with $c > 0$. The monopolist's profit as a function of the price it sets is given by

$$\pi(p) = (p - c)Q(p)$$

Under the assumptions that $Q(.)$ is twice differentiable, strictly decreasing and strictly concave, the profit function $\pi(.)$ is concave on $\mathbb{R}_+$. This is easily verified using the second order conditions for a univariate concave function:

$$\pi''(p) = (p - c)Q''(p) + 2Q'(p) < 0$$

8. *Cournot oligopolist's profit function.* Consider a symmetric oligopoly market (n firms) with inverse demand function $P(Q)$ and linear costs cq with $c > 0$. The oligopolist i's profit as a function of the quantity it produces is given by

$$\pi_i(q_i) = P(q_i + Q_{-i})q_i - cq_i$$

Under the assumptions that $P(.)$ is twice differentiable, strictly decreasing and strictly concave, the profit function $\pi_i(.)$ is concave on $\mathbb{R}_+$. This is easily verified using the second order conditions for a univariate concave function:

$$\pi_i''(q_i) = P''(q_i + Q_{-i})q_i + 2P'(q_i + Q_{-i}) < 0$$

9. *Risk-averse decision maker's utility function in a statistical decision problem.* Suppose a risk averse decision maker (DM) faces a statistical decision problem where he chooses an action $a \in \mathbb{R}$ which pays off a return that depends on the realization of a real-valued random variable X. If the DM is an vNM expected utility maximizer, than her expected utility

$$\mathbb{E}_X[u(a, X)]$$

is a concave function of her action a. This is argued as follows. Fix actions a_1 and a_2 and a scalar $\theta \in [0, 1]$. Then for every realization x of X, we have

$$\begin{aligned}
u(\theta a_1 + (1-\theta)a_2, x) &\geq \theta u(a_1, x) + (1-\theta)u(a_2, x) && \text{(DM risk averse)} \\
\mathbb{E}_X[u(\theta a_1 + (1-\theta)a_2, X)] &\geq \mathbb{E}_X[\theta u(a_1, X) + (1-\theta)u(a_2, X)] && \text{(Expectation is a monotone operator)} \\
&= \theta \mathbb{E}_X[u(a_1, X)] + (1-\theta)\mathbb{E}_X[u(a_2, X)] && \text{(Expectation is a linear operator)}
\end{aligned}$$

This shows that $\mathbb{E}_X[u(a, X)]$ is a concave function of her action a.

10.6 Convexity preserving operations on convex functions

Given a basic library of convex functions, we can create new convex functions using some simple operations that preserve convexity. In this chapter, we focus on a subset of these operations.

Rule A. Nonnegative scaling of a convex function is convex

Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function. Fix a scalar $\alpha \geq 0$. Then the scaled function $\alpha f : \mathbb{R}^n \mapsto \mathbb{R}$ is defined as $(\alpha f)(\mathbf{x}) := \alpha f(\mathbf{x})$. Let us use the definition of a convex function to verify that αf is convex. To this end, take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and a scalar $\theta \in [0, 1]$. Then we have

$$\begin{aligned}
(\alpha f)(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) &:= \alpha f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \\
&\leq \alpha\big(\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2)\big) && \text{(by convexity of } f\text{)} \\
&= \theta(\alpha f)(\mathbf{x}_1) + (1-\theta)(\alpha f)(\mathbf{x}_2) && \text{(by definition of } \alpha f\text{)}
\end{aligned}$$

This proves αf is convex.

Rule B. Addition of two convex functions is convex

Suppose f_1 and f_2 are both convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. Then the addition function $f_1 + f_2 : \mathbb{R}^n \mapsto \mathbb{R}$ is defined as $(f_1 + f_2)(\mathbf{x}) := f_1(\mathbf{x}) + f_2(\mathbf{x})$. Let us use the definition of a convex function to verify that $f_1 + f_2$ is convex. Take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and a scalar $\theta \in [0, 1]$. Then the verification follows the familiar template

$$\begin{aligned}
(f_1 + f_2)(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) &:= f_1(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + f_2(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \\
&\leq \big(\theta f_1(\mathbf{x}_1) + (1-\theta) f_1(\mathbf{x}_2)\big) + \big(\theta f_2(\mathbf{x}_1) + (1-\theta) f_2(\mathbf{x}_2)\big) \\
&= \theta(f_1 + f_2)(\mathbf{x}_1) + (1-\theta)(f_1 + f_2)(\mathbf{x}_2)
\end{aligned}$$

This proves $f_1 + f_2$ is convex.

Rule 1. A conical combination of convex functions is convex

Rule A and Rule B can be combined to give a general rule: if $f_1, \ldots, f_m$ are m real-valued convex functions on a common domain, then given weights $w_1 \geq 0$, …, $w_m \geq 0$, their conical sum function $f := w_1 f_1 + \ldots + w_m f_m$ is a convex function on the common domain.

Example 29. For a fixed parameter $a \in (0, 1)$, consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by

$$f(x_1, x_2) = a \exp(x_1) + (1-a) \exp(x_2)$$

One can argue that f is a convex function on $\mathbb{R}^2$ as follows

Step 1. From our basic library, we know that $\exp(x)$ is convex univariate function. It follows therefore (try to show this formally) that $(x_1, x_2) \mapsto \exp(x_1)$ and $(x_1, x_2) \mapsto \exp(x_2)$ are convex functions on $\mathbb{R}^2$.

Step 2. Invoke the conical combination rule to the two convex functions of Step 1 to conclude that f is a convex function on $\mathbb{R}^2$.

Rule 2. Convex transformation of an affine function is convex

We know that an affine function from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be written as $A\mathbf{x} + \mathbf{b}$, where A is an $m \times n$ matrix, $\mathbf{b}$ is some fixed vector in $\mathbb{R}^m$ and $\mathbf{x}$ is a general domain point in $\mathbb{R}^n$. Suppose

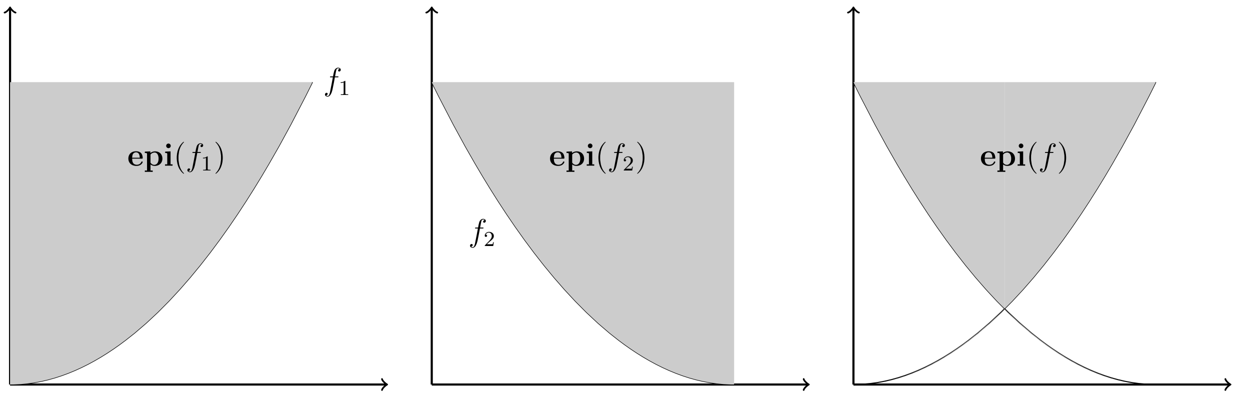

Figure 10.10: Illustration of Rule 3 for the case when both $\mathbf{dom}(f_1) \subset \mathbb{R}$ and $\mathbf{dom}(f_2) \subset \mathbb{R}$.

then that f is a real-valued convex function on the range of $A\mathbf{x}+\mathbf{b}$ in $\mathbb{R}^m$. Then the convex transformation $f(A\mathbf{x}+\mathbf{b})$ of the affine function $A\mathbf{x}+\mathbf{b}$ is a convex function from $\mathbb{R}^n$ to $\mathbb{R}$. To check this, simply verify the definition. Take two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and a scalar $\theta \in [0,1]$.

$$\begin{aligned} f\big(A(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) + \mathbf{b}\big) &= f\big(\theta(A\mathbf{x}_1 + b) + (1-\theta)(A\mathbf{x}_2 + \mathbf{b})\big) \\ &\leq \theta f(A\mathbf{x}_1 + \mathbf{b}) + (1-\theta) f(A\mathbf{x}_2 + \mathbf{b}) \quad \text{(by convexity of } f) \end{aligned}$$

Example 30. Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by

$$f(x_1, x_2) = \exp(2x_1 + 3x_2 + 5)$$

One can argue that f is a convex function on $\mathbb{R}^2$ as follows

Step 1. The mapping g defined by $g(x_1, x_2) = 2x_1 + 3x_2 + 5$ is an affine function on $\mathbb{R}^2$. The mapping h defined by $h(t) = \exp(t)$ is a convex univariate function.

Step 2. As f can be expressed as the composition hog, invoke Rule 2 to conclude that f is a convex function on $\mathbb{R}^2$.

Rule 3. Pointwise maximum of two convex functions is convex

Suppose f_1 and f_2 are both convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. Rule 3 says that their pointwise maximum function

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\} \text{ defined on the common domain } \mathbf{dom}(f_1) \cap \mathbf{dom}(f_2)$$

is a convex function.

To establish this rule, we use the epigraph characterization of convex functions: $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function if and only if $\mathbf{epi}(f) \subset \mathbb{R}^n \times \mathbb{R}$ is a convex set. Using this characterization, we know $\mathbf{epi}(f_1)$ and $\mathbf{epi}(f_2)$ are convex sets. The first step is to establish that $\mathbf{epi}(f) = \mathbf{epi}(f_1) \cap \mathbf{epi}(f_2)$. This follows from the following chain of equivalences:

$$(\mathbf{x}, t) \in \mathbf{epi}(f) \text{ iff } t \geq f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$$

$$\begin{aligned}
&\text{iff } t \geq f_1(\mathbf{x}) \text{ and } t \geq f_2(\mathbf{x}) \\
&\text{iff } (\mathbf{x}, t) \in \mathbf{epi}(f_1) \text{ and } (\mathbf{x}, t) \in \mathbf{epi}(f_2) \\
&\text{iff } (\mathbf{x}, t) \in \mathbf{epi}(f_1) \cap \mathbf{epi}(f_2)
\end{aligned}$$

The next step is to observe that $\mathbf{epi}(f)$, as the intersection of two convex sets, is itself a convex set. The epigraph characterization of convexity of a function then gives us the convexity of f. Figure 10.10 illustrates Rule 3 on a simple instance when the domains of f_1 and f_2 are subsets of the real line.

Generalized Rule 3. Pointwise supremum of an infinite set of convex functions is convex

Suppose we have an infinite family of functions $f(\mathbf{x}, \mathbf{y})$ indexed by a continuous variable $\mathbf{y} \in Y$ such that for every $\mathbf{y} \in Y$, $f(\mathbf{x}, \mathbf{y})$ is a convex function of $\mathbf{x}$. Then the function

$$g(\mathbf{x}) = \sup_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$$

is a convex function on its domain $\{\mathbf{x} : (\mathbf{x}, \mathbf{y}) \in \mathbf{dom} f$ for every $\mathbf{y} \in Y$ and $\sup_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y}) < \infty\}$. The argument is analogous to the two functions case. It follows because $\mathbf{epi}(g) = \cap_{\mathbf{y} \in Y} \mathbf{epi} f(\bullet, \mathbf{y})$; and, arbitrary intersection of convex sets is convex.

Example 31. Consider the pointwise maximum function $f : \mathbb{R}^n \mapsto \mathbb{R}$ of a finite number of affine functions $\{\langle \mathbf{a}_1, \mathbf{x} \rangle + \mathbf{b}_1, \ldots, \langle \mathbf{a}_L, \mathbf{x} \rangle + \mathbf{b}_L\}$ defined by

$$f(\mathbf{x}) = \max_i \, \langle \mathbf{a}_i, \mathbf{x} \rangle + \mathbf{b}_i$$

This is also a piecewise affine function. Then since affine functions are also convex, by invoking generalized Rule 3, we get the conclusion that f is a convex function on $\mathbb{R}^n$.

Rule 4. Nondecreasing convex transformation of a convex function is convex

Suppose f is a convex function from $\mathbb{R}^n$ to $\mathbb{R}$; and g is a nondecreasing convex function from $\mathbf{range}(f) \subset \mathbb{R}$ to $\mathbb{R}$. Then Rule 4 says that the composition $gof : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function.

To establish the validity of the rule, simply verify its conclusion against the definition of a convex function. So fix two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and a scalar $\theta \in [0, 1]$.

$$\begin{aligned}
(gof)(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) &:= g\big(f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2)\big) \\
&\leq g\big(\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2)\big) && (f \text{ convex and } g \text{ nondecreasing}) \\
&\leq \theta g(f(\mathbf{x}_1)) + (1-\theta) g(f(\mathbf{x}_2)) && (g \text{ convex}) \\
&= \theta (gof)(\mathbf{x}_1) + (1-\theta)(gof)(\mathbf{x}_2)
\end{aligned}$$

This proves that the composition map gof is convex.

Example 32. Consider the squared norm function $h : \mathbb{R}^n \mapsto \mathbb{R}$ defined by

$$h(\mathbf{x}) = \|\mathbf{x}\|^2$$

The argument that h is a convex function on $\mathbb{R}^n$ runs as follows

Step 1. The mapping f defined by $f(\mathbf{x}) = \|\mathbf{x}\|$ is a convex function on $\mathbb{R}^n$. The mapping $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ defined by $g(t) = t^2$ is a nondecreasing convex univariate function.

Step 2. As h can be expressed as the composition gof, invoke Rule 4 to conclude that h is a convex function on $\mathbb{R}^n$.

The nondecreasing condition on g is necessary. The next example shows a case with a decreasing g, where the composition map is actually concave.

Example 33. Consider the negative exponential function $h : \mathbb{R} \mapsto \mathbb{R}$ defined by

$$h(x) = -e^x$$

The function h can be viewed as a composition map gof where the mapping f defined by $f(x) = e^x$ is a convex function on $\mathbb{R}$ and the mapping $g : \mathbb{R} \mapsto \mathbb{R}$ defined by $g(t) = -t$ is a decreasing convex (in fact linear) univariate function. However, h is a concave function.

Rule 5. Partial minimization of a convex function preserves convexity

Suppose $f : (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ is a convex function in its arguments $(\mathbf{x}, \mathbf{p})$. Define $V : \mathbb{R}^m \mapsto \mathbb{R}$ by

$$V(\mathbf{p}) := \min_{\mathbf{x}\in C} f(\mathbf{x}, \mathbf{p}) \quad \text{where} \quad C \subset \mathbb{R}^n \text{ is a nonempty compact convex set}$$

Then Rule 5 says that V is a convex function in $\mathbf{p}$.

To establish this rule, simply test V against the definition of a convex function. Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be two points in $\mathbb{R}^m$ and $\theta \in [0, 1]$ be a scalar. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be the partial minimizers of f in C corresponding to $\mathbf{p}_1$ and $\mathbf{p}_2$ respectively. In other words, $V(\mathbf{p}_1) = f(\mathbf{x}_1, \mathbf{p}_1)$ and $V(\mathbf{p}_2) = f(\mathbf{x}_2, \mathbf{p}_2)$. It follows then

$$\begin{aligned}
\theta V(\mathbf{p}_1) + (1-\theta)V(\mathbf{p}_2) &= \theta f(\mathbf{x}_1, \mathbf{p}_1) + (1-\theta) f(\mathbf{x}_2, \mathbf{p}_2) && \text{(by definition of } \mathbf{x}_i\text{)} \\
&\geq f\big(\theta(\mathbf{x}_1, \mathbf{p}_1) + (1-\theta)(\mathbf{x}_2, \mathbf{p}_2)\big) && \text{(by convexity of } f\text{)} \\
&= f\big(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2\big) && (\mathbb{R}^n \times \mathbb{R}^m \text{ is a vector space}) \\
&\geq V\big(\theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2\big) && \text{(by definition of } V\big(\theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2\big)\text{)}
\end{aligned}$$

When $\mathbf{x}$ is interpreted as the vector of optimization/decision variables and $\mathbf{p}$ is interpreted as a vector of parameters in the context of a minimization problem, Rule 5 says that if the objective function f is jointly convex in the decision variables and the parameters, then the value function V is a convex function of parameters.