Convex Optimization for Machine Learning

with Mathematica Applications

Chapter 5

Convex Sets, Convex Functions and Sub-gradient

M. M. Hammad
Department of Mathematics
Faculty of Science
Damanhour University
Egypt

Chapter 5

Convex Sets, Convex Functions and Sub-gradient

5.1 Convex sets

Convexity, or convex analysis, is an area of mathematics where one studies questions related to two basic objects, namely convex sets and convex functions. Triangles, rectangles and "certain" polygons are examples of convex sets in the plane, and the quadratic function $f(x) = ax^2 + bx + c$ is convex provided that $a \geq 0$. Actually, the points in the plane on or above the graph of this quadratic function is another example of a convex set. But one may also consider convex sets in $\mathbb{R}^n$ for any n, and convex functions of several variables.

Convexity is the mathematical core of optimization, and it plays an important role in many other mathematical areas as statistics, approximation theory, differential equations and mathematical economics.

A basic optimization problem is to minimize a real-valued function f of n variables, say $f(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_n) \in A$ and A is the domain of f. A global minimum of f is a point $\mathbf{x}^*$ with $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in A$, where A is the domain of f. Often it is hard to find a global minimum so one settles with a local minimum point which satisfies $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x}$ in A that are sufficiently close to $\mathbf{x}^*$. There are several optimization algorithms that can locate a local minimum of f.

Unfortunately, the function value in a local minimum may be much larger than the global minimum value. This raises the question: Are there functions where a local minimum point is also a global minimum? The main answer to this question is:

If f is a convex function, and the domain A is a convex set, then a local minimum point is also a global minimum point! Thus, one can find the global minimum of convex functions whereas this may be hard (or even impossible) in other situations.

We start the study of convexity with sets. Geometric ideas play an underlying role in convex analysis, its extensions, and applications.

Definition: Given two elements $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$, define the interval/line segment

$$[\mathbf{a}, \mathbf{b}] \coloneqq \{\lambda\mathbf{a} + (1-\lambda)\mathbf{b} : \lambda \in [0,1]\}. \tag{5.1}$$

The parameter value $\lambda = 0$ corresponds to $\mathbf{b}$, and the parameter value $\lambda = 1$ corresponds to $\mathbf{a}$. Values of the parameter λ between 0 and 1 correspond to the (closed) line segment between $\mathbf{a}$ and $\mathbf{b}$. Note that if $\mathbf{a} = \mathbf{b}$, then this interval reduces to a singleton $[\mathbf{a}, \mathbf{b}] = \{\mathbf{a}\}$. Equivalently, we have

$$[\mathbf{a}, \mathbf{b}] \coloneqq \{\mathbf{b} + \lambda(\mathbf{a} - \mathbf{b}) : \lambda \in [0,1]\}. \tag{5.2}$$

This expression gives another interpretation: the line segment is the sum of the base point $\mathbf{b}$ (corresponding to $\lambda = 0$) and the direction $(\mathbf{a} - \mathbf{b})$ (which points from $\mathbf{b}$ to $\mathbf{a}$) scaled by the parameter λ. Thus, λ gives the fraction of the way from $\mathbf{b}$ to $\mathbf{a}$. As λ increases from 0 to 1, the point moves from $\mathbf{b}$ to $\mathbf{a}$. See figure 5.1.

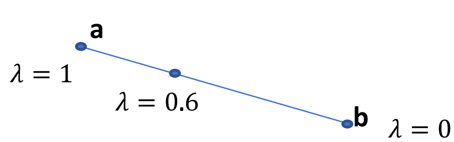

Figure 5.1.

Definition: a set C is convex if the line segment between any two points in C lies in C. See figure 5.2. Equivalently, Ω is convex if

$$\lambda\mathbf{a} + (1-\lambda)\mathbf{b} \in \Omega \tag{5.3}$$

for all $\mathbf{a}, \mathbf{b} \in \Omega$ and $\lambda \in [0,1]$.
Loosely speaking a convex set in $\mathbb{R}^2$ (or $\mathbb{R}^n$) is a set "with no holes".

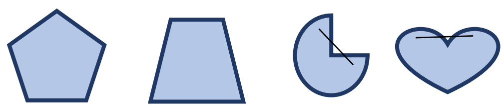

Figure 5.2.

Some simple examples

- The empty set $\emptyset$ is convex.
- any single point $\{x_0\}$ is convex.
- the whole space $\mathbb{R}^n$ is convex.
- A line segment is convex.

Hyperplanes, Half-Spaces, and Support Line/Hyperplane

A hyperplane is a set of the form

$$H = \{\mathbf{x}: \langle \mathbf{a}, \mathbf{x} \rangle \;=\; b\}, \tag{5.4}$$

where $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq 0$, and $b \in \mathbb{R}$. Analytically it is the solution set of a nontrivial linear equation among the components of $\mathbf{x}$. Geometrically, the hyperplane H can be interpreted as the set of points with a constant inner product to a given vector $\mathbf{a}$, or as a hyperplane with normal vector $\mathbf{a}$; the constant $b \in \mathbb{R}$ determines the offset of the hyperplane from the origin. This geometric interpretation can be understood by expressing the hyperplane in the form

$$\{\mathbf{x}: \langle \mathbf{a}, (\mathbf{x} - \mathbf{x}_0) \rangle = 0\}, \tag{5.5}$$

where $\mathbf{x}_0$ is any point in the hyperplane (i.e., any point that satisfies $\langle \mathbf{a}, \mathbf{x}_0 \rangle = b$).

A hyperplane divides $\mathbb{R}^n$ into two half-spaces. A (closed) half-space is a set of the form

$$\{\mathbf{x}: \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}, \tag{5.6}$$

where $\mathbf{a} \neq 0$, i.e., the solution set of one (nontrivial) linear inequality. Half-spaces are convex. This is illustrated in figure 5.3 and figure 5.4.

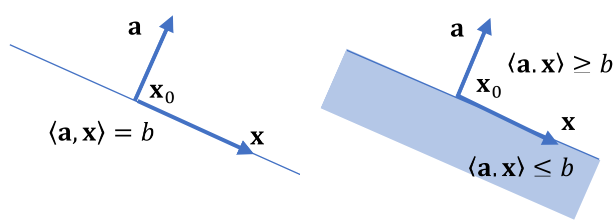

Figure 5.3.

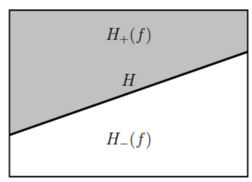

Figure 5.4. The two half-spaces determined by a hyperplane, H.

Theorem 5.1: The hyperplane is a convex set.

Proof:

Let $\langle \mathbf{a}|\mathbf{x}\rangle = b$ be a hyperplane. $\mathbf{x}_1$, and $\mathbf{x}_2$ are any points on the hyperplane. Then $\langle \mathbf{a}|\mathbf{x}_1\rangle = b$, $\langle \mathbf{a}|\mathbf{x}_2\rangle = b$. Therefore, for $0 \leq \lambda \leq 1$,

$$\begin{aligned}\langle \mathbf{a}|(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)\rangle &= \langle \mathbf{a}|\lambda\mathbf{x}_1\rangle + \langle \mathbf{a}|(1-\lambda)\mathbf{x}_2\rangle \\ &= \lambda\langle \mathbf{a}|\mathbf{x}_1\rangle + (1-\lambda)\langle \mathbf{a}|\mathbf{x}_2\rangle \\ &= \lambda b + (1-\lambda)b \\ &= b\end{aligned}$$

Hence, $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, for $0 \leq \lambda \leq 1$ lies in the hyperplane. So, the hyperplane is convex.

■

Definition: An important property of any convex set K in the plane is that at every point p on the boundary of K, there exists at least one line ℓ (or generally a $(n-1)$-dimensional hyperplane in higher dimensions) that passes through p such that K lies entirely in one of the closed half-planes (half-spaces) defined by ℓ (see figure. 5.5). Such a line is called a support line/hyperplane for K.

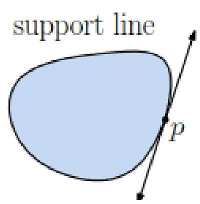

Figure 5.5.

Euclidean ball

Theorem 5.2: The Euclidean ball is a convex set.

Proof:

The (Euclidean) ball (or just ball) in $\mathbb{R}^n$ has the form

$$B(\mathbf{x}_c, r) = \{\mathbf{x}: \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\} = \{\mathbf{x}: \langle(\mathbf{x} - \mathbf{x}_c), (\mathbf{x} - \mathbf{x}_c)\rangle \leq r^2\}.$$

The vector $\mathbf{x}_c$ is the center of the ball and the scalar r is its radius; $B(\mathbf{x}_c, r)$ consists of all points within a distance r of the center $\mathbf{x}_c$. If $\|\mathbf{x}_1 - \mathbf{x}_c\|_2 \leq r$, $\|\mathbf{x}_2 - \mathbf{x}_c\|_2 \leq r$, and $0 \leq \theta \leq 1$, then

$$\begin{aligned}\|\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 - \mathbf{x}_c\|_2 &= \|\theta\mathbf{x}_1 - \theta\mathbf{x}_c + (1-\theta)\mathbf{x}_2 - \mathbf{x}_c + \theta\mathbf{x}_c\|_2 \\ &= \|\theta(\mathbf{x}_1 - \mathbf{x}_c) + (1-\theta)(\mathbf{x}_2 - \mathbf{x}_c)\|_2 \\ &\leq \theta\|(\mathbf{x}_1 - \mathbf{x}_c)\|_2 + (1-\theta)\|(\mathbf{x}_2 - \mathbf{x}_c)\|_2 \\ &\leq \theta r + (1-\theta)r \\ &\leq r\end{aligned}$$

■

Positive Semi-definite Matrices

Theorem 5.3: The set of positive semidefinite matrices are convex.

Proof:

Let $\mathbf{M}_1$ and $\mathbf{M}_2$ are two positive semidefinite matrices. So that, $\langle \mathbf{x}|\mathbf{M}_1|\mathbf{x}\rangle \geq 0$ and $\langle \mathbf{x}|\mathbf{M}_2|\mathbf{x}\rangle \geq 0$. If

$$\mathbf{M} = \theta\mathbf{M}_1 + (1-\theta)\mathbf{M}_2$$

we have

$$
\begin{aligned}
\langle \mathbf{x}|\mathbf{M}|\mathbf{x}\rangle &= \langle \mathbf{x}|(\theta\mathbf{M}_1 + (1-\theta)\mathbf{M}_2)|\mathbf{x}\rangle \\
&= \theta\langle \mathbf{x}|\mathbf{M}_1|\mathbf{x}\rangle + (1-\theta)\langle \mathbf{x}|\mathbf{M}_2|\mathbf{x}\rangle \\
&\geq 0
\end{aligned}
$$

Hence, the set of positive semidefinite matrices is convex.

■

Example 5.1

Show that $C = \{(x_1, x_2): 2x_1 + 3x_2 = 7\} \subset \mathbb{R}^2$ is a convex set.

Solution

Let $\mathbf{x}, \mathbf{y} \in C$, where $\mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2)$. The line segment joining $\mathbf{x}$ and $\mathbf{y}$ is the set

$$W = \{\mathbf{w}: \mathbf{w} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}, 0 \leq \lambda \leq 1\}$$

For some $0 \leq \lambda \leq 1$, let $\mathbf{w} = (w_1, w_2)$ be a point of set W, so that

$$w_1 = \lambda x_1 + (1-\lambda)y_1, \qquad w_2 = \lambda x_2 + (1-\lambda)y_2$$

Since $\mathbf{x}, \mathbf{y} \in C$, $2x_1 + 3x_2 = 7$ and $2y_1 + 3y_2 = 7$. But,

$$
\begin{aligned}
2w_1 + 3w_2 &= 2(\lambda x_1 + (1-\lambda)y_1) + 3(\lambda x_2 + (1-\lambda)y_2) \\
&= \lambda(2x_1 + 3x_2) + (1-\lambda)(2y_1 + 3y_2) \\
&= 7\lambda + 7(1-\lambda) \\
&= 7
\end{aligned}
$$

Therefore, $\mathbf{w} = (w_1, w_2) \in C$. Since $\mathbf{w}$ is any point of C, $\mathbf{x}$, $\mathbf{y} \in C \Rightarrow [\mathbf{x}: \mathbf{y}] \subset C$. Hence C is convex.

Example 5.2

Show that in $\mathbb{R}^3$, the closed ball $x_1^2 + x_2^2 + x_3^2 \leq 1$ is a convex set.

Solution

Let $S = \{(x_1, x_2, x_3): x_1^2 + x_2^2 + x_3^2 \leq 1\}$. Also, let $\mathbf{x}, \mathbf{y} \in S$, where

$$\mathbf{x} = (x_1, x_2, x_3), \qquad \mathbf{y} = (y_1, y_2, y_3).$$

Then, by giving condition, we have

$$x_1^2 + x_2^2 + x_3^2 \leq 1 \text{ and } y_1^2 + y_2^2 + y_3^2 \leq 1$$

Now, for some scalar λ, $0 \leq \lambda \leq 1$, we have

$$
\begin{aligned}
\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\|^2 &= (\lambda x_1 + (1-\lambda)y_1)^2 + (\lambda x_2 + (1-\lambda)y_2)^2 + (\lambda x_3 + (1-\lambda)y_3)^2 \\
&= \lambda^2(x_1^2 + x_2^2 + x_3^2) + (1-\lambda)^2(y_1^2 + y_2^2 + y_3^2) + 2\lambda(1-\lambda)(x_1y_1 + x_2y_2 + x_3y_3)
\end{aligned}
$$

By Schwartz's inequality,

$$x_1y_1 + x_2y_2 + x_3y_3 \leq \sqrt{x_1^2 + x_2^2 + x_3^2}\sqrt{y_1^2 + y_2^2 + y_3^2}$$

We have

$$\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\|^2 \leq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) = \big(\lambda + (1-\lambda)\big)^2 = 1$$

Therefore, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is a point in S. Thus, $\mathbf{x}$, $\mathbf{y} \in S \Rightarrow [\mathbf{x}: \mathbf{y}] \subset S$. Hence S is convex.

Definition: Given $\boldsymbol{\omega}_1, \ldots \boldsymbol{\omega}_m \in \mathbb{R}^n$, the element

$$\mathbf{x} = \lambda_1\boldsymbol{\omega}_1 + \cdots + \lambda_m\boldsymbol{\omega}_m = \sum_{i=1}^{m} \lambda_i\boldsymbol{\omega}_i, \tag{5.7}$$

where,

$$\sum_{i=1}^{m} \lambda_i = 1, \tag{5.8}$$

and $\lambda_i \geq 0$ for some $m \in \mathbb{N}$, is called a convex combination of $\boldsymbol{\omega}_1, \ldots \boldsymbol{\omega}_m$.

Example 5.3

Determine whether the vector $(0,7)^T$ is a convex combination of the set $\{(3,6)^T, (-6,9)^T, (2,1)^T, (-1,1)^T\}$.

Solution

For these vectors, we have

$$\begin{pmatrix}0\\7\end{pmatrix} = \begin{pmatrix}3\\6\end{pmatrix}\lambda_1 + \begin{pmatrix}-6\\9\end{pmatrix}\lambda_2 + \begin{pmatrix}2\\1\end{pmatrix}\lambda_3 + \begin{pmatrix}-1\\1\end{pmatrix}\lambda_4$$

or

$$3\lambda_1 - 6\lambda_2 + 2\lambda_3 - \lambda_4 = 0 \qquad (*)$$

$$6\lambda_1 + 9\lambda_2 + \lambda_3 + \lambda_4 = 7 \qquad (**)$$

To these equations we add a third condition,

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$$

We must determine whether there exist nonnegative values of $\lambda_1, \lambda_2, \lambda_3$, and λ_4 that simultaneously satisfy (*) and (**). Solving these equations, we obtain

$$\lambda_1 = \frac{2}{3} + \frac{1}{2}\lambda_4, \;\; \lambda_2 = \frac{1}{3} - \frac{5}{16}\lambda_4, \;\; \lambda_3 = -\frac{19}{16}\lambda_4$$

with λ_4 arbitrary. The choice $\lambda_4 = 0$ is forced, giving

$$\lambda_1 = \frac{2}{3}, \;\; \lambda_2 = \frac{1}{3}, \;\; \lambda_3 = 0, \;\; \lambda_4 = 0$$

as an acceptable set of constants. Thus, $(0,7)^T$ is a convex combination of the given set of four vectors.

Next, we proceed with intersections of convex sets.

Theorem 5.4: The intersection of an arbitrary collection of convex sets is convex. Equivalently, let $\{\Omega_\alpha\}_{\alpha\in I}$ be a collection of convex subsets of $\mathbb{R}^n$, where I is an arbitrary index set. Then $\bigcap_{\alpha\in I}\Omega_\alpha$ is also a convex subset of $\mathbb{R}^n$.

Proof:

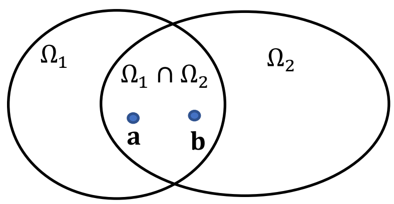

Figure 5.6.

Taking any $\mathbf{a}, \mathbf{b} \in \bigcap_{\alpha\in I}\Omega_\alpha$, we get that $\mathbf{a}, \mathbf{b} \in \Omega_\alpha$ for all $\alpha \in I$, see figure 5.6. The convexity of each Ω_α ensures that

$$\lambda\mathbf{a} + (1-\lambda)\mathbf{b} \in \Omega_\alpha$$

for any $\lambda \in (0,1)$. Thus

$$\lambda\mathbf{a} + (1-\lambda)\mathbf{b} \in \bigcap_{\alpha\in I}\Omega_\alpha$$

and the intersection $\bigcap_{\alpha\in I}\Omega_\alpha$ is convex.

■

Theorem 5.5: Let $\mathbf{b}_i \in \mathbb{R}^n$ and $\beta_i \in \mathbb{R}$ for $i \in I$, where I is an arbitrary index set. Then the set

$$C = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{b}_i\rangle \le \beta_i, i \in I\} \qquad (5.9)$$

is convex.

Note: The theorem 5.5 would still be valid, of course, if some of the inequalities $\le$ were replaced by $\ge, >, <$ or $=$.

Proof:

Let $C_i = \{\mathbf{x} : \langle \mathbf{x}, \mathbf{b}_i\rangle \le \beta_i\}$. Then C_i is a closed half-space and $C = \bigcap_{i\in I} C_i$.

■

Definition: A set which can be expressed as the intersection of finitely many closed half-spaces and hyperplanes of $\mathbb{R}^n$ is called a polyhedral convex set.

$$\mathcal{P} = \left\{\mathbf{x} : \langle \mathbf{a}_j, \mathbf{x}\rangle \le b_j, j = 1, \dots, m, \langle \mathbf{c}_j, \mathbf{x}\rangle = d_j, j = 1, \dots, p\right\}, \qquad (5.10.1)$$

or a polyhedron is defined as the solution set of a finite number of linear equalities and inequalities.

or, a polyhedron $\mathcal{P} \subseteq \mathbb{R}^n$ is defined a the solution set of a system of linear inequalities. Thus, $\mathcal{P}$ has the form

$$\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}, \tag{5.10.2}$$

where $\mathbf{A}$ is a real $m \times n$ matrix, $\mathbf{b} \in \mathbb{R}^m$ and where the vector inequality means it holds for every component.

It is easily shown that polyhedra are convex sets. See figure 5.7. Such sets are considerably better behaved than general convex sets, mostly because of their lack of "curvature."

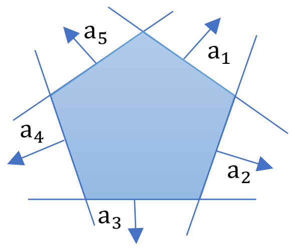

Figure 5.7.

Theorem 5.6: Every polyhedron is a convex set.

Proof:

Consider a polyhedron $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}$ and let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{P}$ and $0 \leq \lambda \leq 1$. Then

$$\mathbf{A}((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) = (1-\lambda)\mathbf{Ax}_1 + \lambda\mathbf{Ax}_2 \leq (1-\lambda)b + \lambda b = b$$

which shows that $(1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in \mathcal{P}$ and the convexity of $\mathcal{P}$ follows.

■

Theorem 5.7: A subset Ω of $\mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.

Proof:

Actually, by definition, a set Ω is convex if and only if $\lambda_1\boldsymbol{\omega}_1 + \lambda_2\boldsymbol{\omega}_2 \in \Omega$ whenever $\boldsymbol{\omega}_1 \in \Omega, \boldsymbol{\omega}_2 \in \Omega, \lambda_1 \geq 0, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. In other words, the convexity of Ω means that Ω is closed undertaking convex combinations with $m = 2$. We must show that this implies Ω is also closed undertaking convex combinations with $m > 2$. Fix now a positive integer $m \geq 2$ and suppose that every convex combination of $k \in \mathbb{N}$ elements from Ω, where $k \leq m$, belongs to Ω. Form the convex combination

$$\mathbf{y} \coloneqq \sum_{i=1}^{m+1} \lambda_i \boldsymbol{\omega}_i\,, \qquad \sum_{i=1}^{m+1} \lambda_i = 1, \qquad \lambda_i \geq 0$$

and observe that if $\lambda_{m+1} = 1$, then $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$, so $\mathbf{y} = \boldsymbol{\omega}_{m+1} \in \Omega$. In the case where $\lambda_{m+1} < 1$ we get the representations

$$\sum_{i=1}^{m} \lambda_i + \lambda_{m+1} = 1$$

and

$$\sum_{i=1}^{m} \lambda_i = 1 - \lambda_{m+1}, \quad \text{i.e.,} \qquad \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} = 1.$$

which imply in turn the inclusion

$$\mathbf{z} \coloneqq \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} \boldsymbol{\omega}_i \in \Omega$$

It yields therefore the relationships

$$\mathbf{y} = (1-\lambda_{m+1})\sum_{i=1}^{m}\left(\frac{\lambda_i}{1-\lambda_{m+1}}\right)\boldsymbol{\omega}_i + \lambda_{m+1}\boldsymbol{\omega}_{m+1} = (1-\lambda_{m+1})\mathbf{z} + \lambda_{m+1}\boldsymbol{\omega}_{m+1} \in \Omega$$

and thus completes the proof of the proposition.

■

Theorem 5.8: Let Ω_1 be a convex subset of $\mathbb{R}^n$ and let Ω_2 be a convex subset of $\mathbb{R}^p$. Then the Cartesian product $\Omega_1 \times \Omega_2$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}^p$.

Proof:

Fix $a = (\mathbf{a}_1, \mathbf{a}_2), b = (\mathbf{b}_1, \mathbf{b}_2) \in \Omega_1 \times \Omega_2$, and $\lambda \in (0,1)$. Then we have $\mathbf{a}_1, \mathbf{b}_1 \in \Omega_1$ and $\mathbf{a}_2, \mathbf{b}_2 \in \Omega_2$. The convexity of Ω_1 and Ω_2 gives us

$$\lambda\mathbf{a}_i + (1-\lambda)\mathbf{b}_i \in \Omega_i \text{ for } i = 1,2.$$

which implies therefore that

$$\lambda a + (1-\lambda)b = \lambda(\mathbf{a}_1, \mathbf{a}_2) + (1-\lambda)(\mathbf{b}_1, \mathbf{b}_2) = (\lambda\mathbf{a}_1 + (1-\lambda)\mathbf{b}_1, \lambda\mathbf{a}_2 + (1-\lambda)\mathbf{b}_2) \in \Omega_1 \times \Omega_2.$$

Thus, the Cartesian product $\Omega_1 \times \Omega_2$ is convex.

■

Theorem 5.9: Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex and let $\lambda \in \mathbb{R}$. Then both sets $\lambda\Omega_1$ and $\Omega_1 + \Omega_2$ are also convex in $\mathbb{R}^n$.

Proof:

Let $\lambda\mathbf{v}_1, \lambda\mathbf{v}_2 \in \lambda\Omega_1$, where $\mathbf{v}_1, \mathbf{v}_2 \in \Omega_1$. Because Ω_1 is convex, we have

$$\alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2 \in \Omega_1,$$

for any $\alpha \in (0,1)$. Hence,

$$\alpha\lambda\mathbf{v}_1 + (1-\alpha)\lambda\mathbf{v}_2 = \lambda(\alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2) \in \lambda\Omega_1$$

and thus $\lambda\Omega_1$ is convex.

Let $\mathbf{v}_1, \mathbf{v}_2 \in \Omega_1 + \Omega_2$. Then $\mathbf{v}_1 = \dot{\mathbf{v}}_1 + \ddot{\mathbf{v}}_1$ and $\mathbf{v}_2 = \dot{\mathbf{v}}_2 + \ddot{\mathbf{v}}_2$, where $\dot{\mathbf{v}}_1, \dot{\mathbf{v}}_2 \in \Omega_1$ and $\ddot{\mathbf{v}}_1, \ddot{\mathbf{v}}_2 \in \Omega_2$. Because Ω_1 and Ω_2 are convex, for all $\alpha \in (0,1)$, we have

$$\mathbf{x}_1 = \alpha\dot{\mathbf{v}}_1 + (1-\alpha)\dot{\mathbf{v}}_2 \in \Omega_1,$$

and

$$\mathbf{x}_2 = \alpha\ddot{\mathbf{v}}_1 + (1-\alpha)\ddot{\mathbf{v}}_2 \in \Omega_2,$$

By definition of $\Omega_1 + \Omega_2$, $\mathbf{x}_1 + \mathbf{x}_2 \in \Omega_1 + \Omega_2$. Now,

$$\alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2 = \alpha(\dot{\mathbf{v}}_1 + \ddot{\mathbf{v}}_1) + (1-\alpha)(\dot{\mathbf{v}}_2 + \ddot{\mathbf{v}}_2) = \mathbf{x}_1 + \mathbf{x}_2 \in \Omega_1 + \Omega_2.$$

Hence, $\Omega_1 + \Omega_2$ is convex.

■

Let us continue with the definition of affine mappings.

Definition: $B: \mathbb{R}^n \to \mathbb{R}^p$ is affine mapping (affine transformation) if and only if

$$B(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \lambda B(\mathbf{x}) + (1-\lambda)B(\mathbf{y}) \text{ for all } \mathbf{x},\mathbf{y} \in \mathbb{R}^n \text{ and } \lambda \in \mathbb{R}. \tag{5.11}$$

Theorem 5.10: A mapping $B: \mathbb{R}^n \to \mathbb{R}^p$ is affine if there exist a linear mapping $A: \mathbb{R}^n \to \mathbb{R}^p$ and an element $\mathbf{b} \in \mathbb{R}^p$ such that

$$B(\mathbf{x}) = A(\mathbf{x}) + \mathbf{b} \tag{5.12}$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Proof:

if $B(\mathbf{x}) = A(\mathbf{x}) + \mathbf{b}$ where A is linear, one has

$$\begin{aligned}
B\big((1-\lambda)\mathbf{x} + \lambda\mathbf{y}\big) &= A\big((1-\lambda)\mathbf{x} + \lambda\mathbf{y}\big) + \mathbf{b} \\
&= (1-\lambda)A(\mathbf{x}) + \lambda A(\mathbf{y}) + \mathbf{b} \\
&= (1-\lambda)A(\mathbf{x}) + \lambda A(\mathbf{y}) + \mathbf{b} + \lambda\mathbf{b} - \lambda\mathbf{b} \\
&= (1-\lambda)A(\mathbf{x}) + \lambda A(\mathbf{y}) + \lambda\mathbf{b} + (1-\lambda)\mathbf{b} \\
&= \big((1-\lambda)A(\mathbf{x}) + (1-\lambda)\mathbf{b}\big) + (\lambda A(\mathbf{y}) + \lambda\mathbf{b}) \\
&= (1-\lambda)(A(\mathbf{x}) + \mathbf{b}) + \lambda(A(\mathbf{y}) + \mathbf{b}) \\
&= (1-\lambda)B(\mathbf{x}) + \lambda B(\mathbf{y})
\end{aligned}$$

Thus, B is affine. Conversely, if B is affine, let $\mathbf{b} = B(\mathbf{0})$ and $A(\mathbf{x}) = B(\mathbf{x}) - \mathbf{b}$. Then A is an affine transformation with $A(\mathbf{0}) = \mathbf{0}$. Then A is actually linear.

■

Now we show that set convexity is preserved under affine operations.

Theorem 5.11: Let $B: \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping. Suppose that Ω is a convex subset of $\mathbb{R}^n$ and Θ is a convex subset of $\mathbb{R}^p$. Then $B(\Omega)$ is a convex subset of $\mathbb{R}^p$ and $B^{-1}(\Theta)$ is a convex subset of $\mathbb{R}^n$.

Proof:

Fix any $\mathbf{a}, \mathbf{b} \in B(\Omega)$ and $\lambda \in (0,1)$. Then $\mathbf{a} = B(\mathbf{x})$ and $\mathbf{b} = B(\mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \Omega$. Since Ω is convex, we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \Omega$. Then

$$\lambda\mathbf{a} + (1-\lambda)\mathbf{b} = \lambda B(\mathbf{x}) + (1-\lambda)B(\mathbf{y}) = B(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \in B(\Omega),$$

which justifies the convexity of the image $B(\Omega)$.

Taking now any $\mathbf{x}, \boldsymbol{y} \in B^{-1}(\Theta)$ and $\lambda \in (0,1)$, we get $B(\mathbf{x})$ and $B(\mathbf{y}) \in \Theta$. This gives us

$$\lambda B(\mathbf{x}) + (1-\lambda)B(\mathbf{y}) = B(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \in \Theta$$

by the convexity of Θ. Thus, we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in B^{-1}(\Theta)$, which verifies the convexity of the inverse image $B^{-1}(\Theta)$.

■

5.2 Convex Hull

Indeed, there are many problems that are comparatively easy to solve for convex sets but very hard in general. We will encounter some particular instances of this phenomenon later in the course. However, not all sets are convex and a discrete set of points is never convex, unless it consists of at most one point only. In such a case it is useful to make a given set P convex, that is, approximate P with or, rather, encompass P within a convex set $H \supseteq P$. Ideally, H differs from P as little as possible, that is, we want H to be a smallest convex set enclosing P.

At this point let us step back for a second and ask ourselves whether this wish makes sense at all: Does such a set H (always) exist? Fortunately, we are on the safe side because the whole space $\mathbb{R}^n$ is certainly convex. It is less obvious, but we will see below that H is actually unique. Therefore, it is legitimate to refer to H as the smallest convex set enclosing P or shortly the convex hull of P.

The convex hull of a shape is the smallest convex set that contains it. The convex hull may be defined either as the intersection of all convex sets containing a given subset of a Euclidean space, or equivalently as the set of all convex combinations of points in the subset. For a bounded subset of the plane, the convex hull may be visualized as the shape enclosed by a rubber band stretched around the subset. (See figure 5.8)

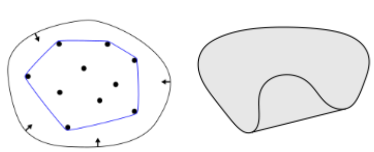

Figure 5.8.

Definition: Let Ω be a subset of $\mathbb{R}^n$. The convex hull of Ω is defined by

$$\text{co}\,\Omega \coloneqq \bigcap\{C: C \text{ is convex and } \Omega \subset C\}$$

From the definition of the convex hull of a given set Ω, convex hull is

- The intersection of all convex sets containing Ω.
- The (unique) minimal convex set containing Ω.
- The set of all convex combinations of points in Ω.

Theorem 5.12: The convex hull $\text{co}\,\Omega$ is the smallest convex set containing Ω.

Proof:

Proof 1: First of all, $\text{co}(\Omega)$ contains Ω: for every $\mathbf{x} \in \Omega$, $1\mathbf{x}$ is a convex combination of size 1, so $\mathbf{x} \in \text{co}(\Omega)$.

Second, $\text{co}(\Omega)$ is a convex set: if we take $\mathbf{x}, \mathbf{y} \in \text{co}(\Omega)$ which are the convex combinations of points in Ω, then $t\mathbf{x} + (1-t)\mathbf{y}$ can be expanded to get another convex combinations of points in Ω.

All convex sets containing Ω must contain $\text{co}(\Omega)$, and $\text{co}(\Omega)$ is itself a convex set containing Ω; therefore it's the smallest such set.

■

Proof 2: The convexity of the set $\text{co}\,\Omega \supset \Omega$ follows from the fact "the intersection of an arbitrary collection of convex sets is convex". On the other hand, for any convex set C that contains Ω we clearly have $\text{co}\,\Omega \subset C$, which verifies the theorem.

■

Proof 3: The intersection of an arbitrary collection of convex sets is convex. Since any set is contained in at least one convex set (the whole vector space in which it sits), it follows that any set, A, is contained in a smallest convex set, namely the intersection of all the convex sets that contain A.

■

Theorem 5.13: For any subset Ω of $\mathbb{R}^n$, its convex hull admits the representation

$$\text{co}\,\Omega \coloneqq \left\{\sum_{i=1}^{m} \lambda_i \mathbf{a}_i : \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0, \mathbf{a}_i \in \Omega, m \in \mathbb{N}\right\}. \tag{5.13}$$

Proof:

Denoting by C the right-hand side of the representation to prove, we obviously have $\Omega \subset C$. Let us check that the set C is convex. Take any $\mathbf{a}, \mathbf{b} \in C$ and get

$$\mathbf{a} := \sum_{i=1}^{p} \alpha_i \mathbf{a}_i, \mathbf{b} := \sum_{i=1}^{q} \beta_i \mathbf{b}_i,$$

where $\mathbf{a}_i, \mathbf{b}_i \in \Omega$, $\beta \geq 0$ with $\sum_{i=1}^{p} \alpha_i = 1$, $\sum_{i=1}^{q} \beta_i = 1$, and $p, q \in \mathbb{N}$. It follows easily that for every number $\lambda \in (0,1)$, we have

$$\begin{aligned}\lambda\mathbf{a} + (1-\lambda)\mathbf{b} &= \sum_{i=1}^{p} \lambda\alpha_i \mathbf{a}_i + \sum_{i=1}^{q} (1-\lambda)\beta_i \mathbf{b}_i \\ &= \lambda\alpha_1\mathbf{a}_1 + \lambda\alpha_2\mathbf{a}_2 + \cdots + \lambda\alpha_p\mathbf{a}_p + (1-\lambda)\beta_1\mathbf{b}_1 + (1-\lambda)\beta_2\mathbf{b}_2 + (1-\lambda)\beta_q\mathbf{b}_q.\end{aligned}$$

Then the resulting equality

$$\sum_{i=1}^{p} \lambda\alpha_i + \sum_{i=1}^{q} (1-\lambda)\beta_i = \lambda \sum_{i=1}^{p} \alpha_i + (1-\lambda) \sum_{i=1}^{q} \beta_i = 1$$

ensures that $\lambda\mathbf{a} + (1-\lambda)\mathbf{b} \in C$, and thus $\operatorname{co}\Omega \subset C$ by the definition of $\operatorname{co}\Omega$.

Fix now any $\mathbf{a} = \sum_{i=1}^{m} \lambda\mathbf{a}_i \in C$ with $\mathbf{a}_i \in \Omega \subset \operatorname{co}\Omega$ for $i = 1, \dots, m$. Since the set $\operatorname{co}\Omega$ is convex, we conclude by the fact "a subset of $\mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements" that $\mathbf{a} \in \operatorname{co}\Omega$ and thus $\operatorname{co}\Omega = C$.

■

5.3 Convex Cone

Definition: A subset K of $\mathbb{R}^n$ is called a cone if it is closed under positive scalar multiplication, i.e., $\lambda\mathbf{x} \in K$ when $\mathbf{x} \in K$ and $\lambda > 0$. Such a set is a union of half-lines emanating from the origin. The origin itself may or may not be included.

Definition: A convex cone is a cone which is a convex set which means that for any $\mathbf{x}_1$, $\mathbf{x}_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in C. \tag{5.14}$$

Points of this form can be described geometrically as forming the two-dimensional pie slice with apex 0 and edges passing through $\mathbf{x}_1$ and $\mathbf{x}_2$. (See figure 5.9)

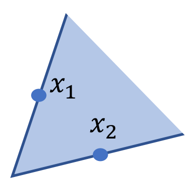

Figure 5.9.

Examples:

1- Subspaces of $\mathbb{R}^n$ are in particular convex cones.
2- So are the open and closed half-spaces corresponding to a hyperplane through the origin.
3- Two of the most important convex cones are the non-negative orthant of $\mathbb{R}^n$,

$$\{\mathbf{x} = (\xi_1, \xi_2, \dots, \xi_n) : \xi_1 \geq 0, \dots, \xi_n \geq 0\}, \tag{5.15}$$

and the positive orthant

$$\{\mathbf{x} = (\xi_1, \xi_2, \dots, \xi_n) : \xi_1 > 0, \dots, \xi_n > 0\}. \tag{5.16}$$

Theorem 5.14: The intersection of an arbitrary collection of convex cones is a convex cone.

Theorem 5.15: Let $\mathbf{b}_i \in \mathbb{R}^n$ for $i \in I$, where I is an arbitrary index set. Then

$$K = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{b}_i \rangle \leq 0, i \in I\} \tag{5.17}$$

is a convex cone.

Theorem 5.16: A subset of $\mathbb{R}^n$ is a convex cone if and only if it is closed under addition and positive scalar multiplication.

Proof:

Let K be a cone. Let $\mathbf{x} \in K$ and $\mathbf{y} \in K$. If K is convex, the vector $\mathbf{z} = \frac{1}{2}(\mathbf{x} + \mathbf{y})$ belongs to K, and hence $\mathbf{x} + \mathbf{y} = \mathbf{2z} \in K$. On the other hand, if K is closed under addition, and if $0 < \lambda < 1$, the vectors $(1 - \lambda)\mathbf{x}$ and $\lambda\mathbf{y}$ belong to K, and hence $(1 - \lambda)\mathbf{x} + \lambda\mathbf{y} \in K$. Thus, K is convex if and only if it is closed under addition.

■

Theorem 5.17: A subset of $\mathbb{R}^n$ is a convex cone if and only if it contains all the positive linear combinations of its elements (i.e., linear combinations $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \lambda_m\mathbf{x}_m$ in which the coefficients are all positive).

Theorem 5.18: Let S be an arbitrary subset of $\mathbb{R}^n$, and let K be the set of all positive linear combinations of S. Then K is the smallest convex cone which includes S.

Proof:

Clearly K is closed under addition and positive scalar multiplication, and $K \supset S$. Every convex cone including S must, on the other hand, include K.

■

A simpler description is possible when S is convex, as follows.

Theorem 5.19: Let C be a convex set, and let

$$K = \{\lambda\mathbf{x} \colon \lambda > 0, \mathbf{x} \in C\}. \tag{5.18}$$

Then K is the smallest convex cone which includes C.

Proof:

This follows from the preceding theorem. Namely, every positive linear combination of elements of C is a positive scalar multiple of a convex combination of elements of C and hence is an element of K.

■

A vector $\mathbf{x}^*$ is said to be normal to a convex set C at a point $\mathbf{a}$, where $\mathbf{a} \in C$, if $\mathbf{x}^*$ does not make an acute angle with any line segment in C with $\mathbf{a}$ as endpoint, i.e. if $\langle \mathbf{v} - \mathbf{a}, \mathbf{x}^* \rangle \leq 0$ for every $\mathbf{x} \in C$. For instance, if C is a half-space $\{\mathbf{x} \colon \langle \mathbf{x}, \mathbf{b} \rangle \leq \beta\}$ and $\mathbf{a}$ satisfies $\langle \mathbf{a}, \mathbf{b} \rangle = \beta$, then $\mathbf{b}$ is normal to C at $\mathbf{a}$. In general, the set of all vectors $\mathbf{x}^*$ normal to C at $\mathbf{a}$ is called the normal cone to C at $\mathbf{a}$. The reader can verify easily that this cone is always convex.

Another easily verified example of a convex cone is the barrier cone of a convex set C. This is defined as the set of all vectors $\mathbf{x}^*$ such that, for some $\beta \in \mathbb{R}$, $\langle \mathbf{x}, \mathbf{x}^* \rangle \leq \beta$ for every $\mathbf{x} \in C$.

5.4 Convex functions

This section collects basic facts about general (extended-real-valued) convex functions including their analytic and geometric characterizations.

Definition: Let $f: \Omega \to \overline{\mathbb{R}}$, ($\overline{\mathbb{R}} = (-\infty, \infty]$), be an extended-real-valued function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then the function f is convex on Ω if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{5.19}$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$ and $\lambda \in (0,1)$.

Geometrically, this inequality means that the line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$, which is the chord from $\mathbf{x}$ to $\mathbf{y}$, lies above the graph of f (see figures 5.10 and 5.11).

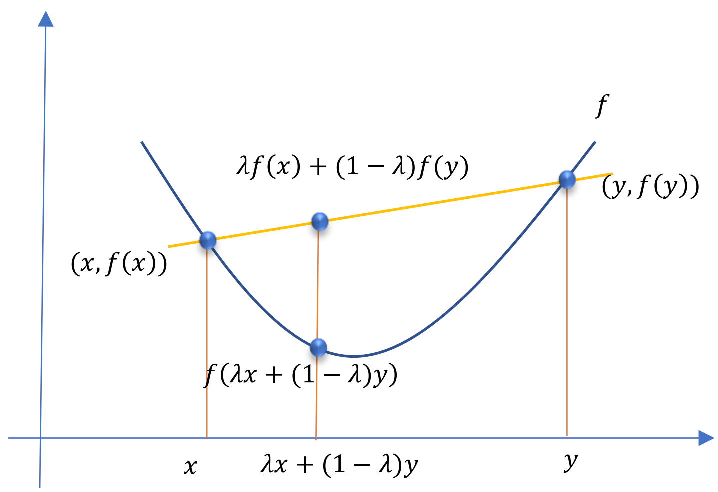

Figure 5.10.

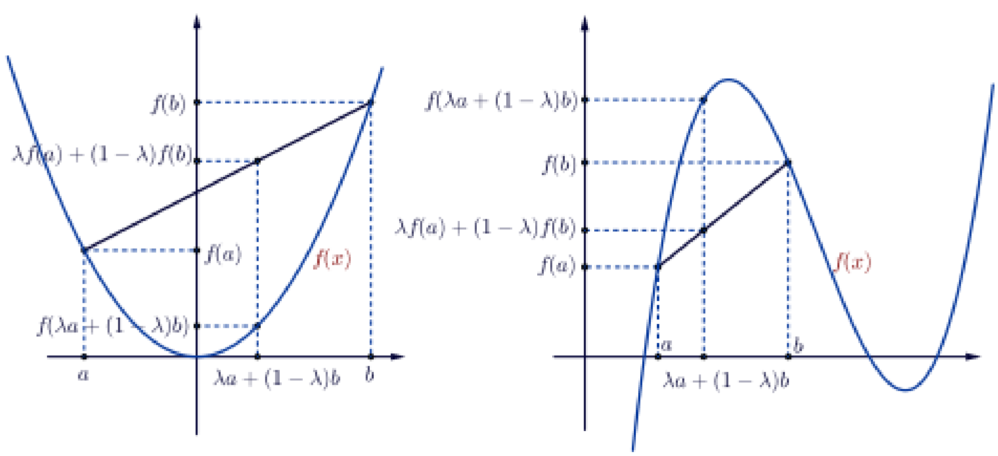

Figure 5.11.

Definition: A function f is strictly convex if strict inequality holds

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{5.20}$$

for all $\mathbf{x} \neq \mathbf{y} \in \Omega$ and $\lambda \in (0,1)$.

A strictly convex function f is a function that the straight line between any pair of points on the curve f is above the curve f except for the intersection points between the straight line and the curve.

We say f is concave if $-f$ is convex, and strictly concave if $-f$ is strictly convex. It is often convenient to extend a convex function to all of $\mathbb{R}^n$ by defining its value to be ∞ outside its domain. If f is convex we define its extended-value extension $\tilde{f}: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \mathbf{x} \in \operatorname{dom} f \\ \infty & \mathbf{x} \notin \operatorname{dom} f \end{cases} \tag{5.21}$$

The extension $\tilde{f}$ is defined on all $\mathbb{R}^n$, and takes values in $\mathbb{R} \cup \{\infty\}$. We can recover the domain of the original function f from the extension $\tilde{f}$ as $\operatorname{dom} f = \{\mathbf{x}: \tilde{f}(\mathbf{x}) < \infty\}$. The extension can simplify notation, since we do not need to explicitly describe the domain, or add the qualifier for all $\mathbf{x} \in \operatorname{dom} \tilde{f}$ every time we refer to $f(\mathbf{x})$. Consider, for example, the basic defining inequality (5.19). In terms of the extension $\tilde{f}$, we can express it as: for $0 < \lambda < 1$,

$$\tilde{f}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda\tilde{f}(\mathbf{x}) + (1-\lambda)\tilde{f}(\mathbf{y}) \tag{5.22}$$

for any $\mathbf{x}$ and $\mathbf{y}$. (For $\lambda = 0$ or $\lambda = 1$ the inequality always holds.)

Let us illustrate the convexity of functions by examples. In the following, we give a few more examples of convex and concave functions. We start with some functions on $\mathbb{R}$, with variable x.

Example 5.4

Is the function $f(x) = |x|$, $x \in \mathbb{R}$ convex? Is it strictly convex?

Solution

To see if f is convex, we need to see if (5.19) is true; to see if it is strictly convex, we need to check (5.20). Furthermore, these inequalities have to be true for every $x_1, x_2 \in \mathbb{R}$, and every $\lambda \in [0,1]$.

$$\begin{aligned} f\big((1-\lambda)x_1 + \lambda x_2\big) &= |(1-\lambda)x_1 + \lambda x_2| \\ &\le |(1-\lambda)x_1| + |\lambda x_2| \\ &= (1-\lambda)|x_1| + \lambda|x_2| \\ &= (1-\lambda)f(x_1) + \lambda f(x_2) \end{aligned}$$

Therefore (5.19) is satisfied, so f is convex. To show that it is strictly convex, we would have to show that the inequality

$$|(1-\lambda)x_1 + \lambda x_2| \le |(1-\lambda)x_1| + |\lambda x_2|$$

can be replaced by a strict inequality $<$. However, we can't do this: for example, if $x_1 = 1$, $x_2 = 2$, $\lambda = 0.5$, the left side of the inequality ($|\frac{1}{2} + \frac{2}{2}| = \frac{3}{2}$) is exactly equal to the right side ($|\frac{1}{2}| + |\frac{2}{2}| = \frac{3}{2}$). So f is not strictly convex.

Note that proving that $f(x)$ is convex requires a general argument, where proving that $f(x)$ was not strictly convex only required a single counterexample. This is because the definition of convexity is a "for all" or "for every" type of argument. To prove convexity, you need an argument that allows for all possible values of x_1, x_2, and λ, whereas to disprove it you only need to give one set of values where the necessary condition does not hold.

Example 5.5

Show that every affine function $f(x) = ax + b$, $x \in \mathbb{R}$ is convex, but not strictly convex.

Solution

$$\begin{aligned} f\big((1-\lambda)x_1 + \lambda x_2\big) &= a\big((1-\lambda)x_1 + \lambda x_2\big) + b \\ &= a\big((1-\lambda)x_1 + \lambda x_2\big) + \big((1-\lambda) + \lambda\big)b \\ &= (1-\lambda)(ax_1 + b) + \lambda(ax_2 + b) \\ &= (1-\lambda)f(x_1) + \lambda f(x_2) \end{aligned}$$

So, we see that inequality (5.19) is in fact satisfied as an equality. That's fine, so every affine function in one varible is convex. However, this means we can't replace the inequality $\le$ with the strict inequality $<$, so affine functions are not strictly convex.

Example 5.6

Show that $f(x) = x^2$, $x \in \mathbb{R}$ is strictly convex.

Solution

Pick x_1, x_2 so that $x_1 \neq x_2$, and pick $\lambda \in (0,1)$.

$$f\big((1-\lambda)x_1+\lambda x_2\big)=\big((1-\lambda)x_1+\lambda x_2\big)^2$$
$$=(1-\lambda)^2x_1^2+\lambda^2x_2^2+2(1-\lambda)\lambda x_1x_2$$

Since $x_1\neq x_2$, $(x_1-x_2)^2>0$. Expanding, this means that $x_1^2+x_2^2>2x_1x_2$. This means that

$$(1-\lambda)^2x_1^2+\lambda^2x_2^2+2(1-\lambda)\lambda x_1x_2<(1-\lambda)^2x_1^2+\lambda^2x_2^2+(1-\lambda)(\lambda)(x_1^2+x_2^2)$$
$$=(1-2\lambda-\lambda^2+\lambda+\lambda^2)x_1^2+(\lambda-\lambda^2+\lambda^2)x_2^2$$
$$=(1-\lambda)x_1^2+\lambda x_2^2$$
$$=(1-\lambda)f(x_1)+\lambda f(x_2)$$

which proves strict convexity.

More examples of convex functions on $\mathbb{R}$ are

- **Exponential**. e^{ax} is convex on $\mathbb{R}$, for any $a\in\mathbb{R}$.
- **Powers**. x^a is convex on $\mathbb{R}_{++}$ when $a\geq 1$ or $a\leq 0$, and concave for $0\leq a\leq 1$.
- **Powers of absolute value**. $|x|^p$, for $p\geq 1$, is convex on $\mathbb{R}$.
- **Logarithm**. $\log x$ is concave on $\mathbb{R}_{++}$ ($\mathbb{R}_{++}$: The set of positive real numbers).

Example 5.7

The following functions are convex:

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

for $\mathbf{x}\in\mathbb{R}^n$, where $\mathbf{a}\in\mathbb{R}^n$ and $b\in\mathbb{R}$.

Solution

Indeed, the function f is convex since

$$f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y})=\langle\mathbf{a},\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\rangle+b$$
$$=\lambda\langle\mathbf{a},\mathbf{x}\rangle+(1-\lambda)\langle\mathbf{a},\mathbf{y}\rangle+b$$
$$=\lambda(\langle\mathbf{a},\mathbf{x}\rangle+b)+(1-\lambda)(\langle\mathbf{a},\mathbf{y}\rangle+b)$$
$$=\lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})$$

for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ and $\lambda\in(0,1)$.

Example 5.8

The following function is convex:

$$g(\mathbf{x}):=\|\mathbf{x}\|\text{ for }\mathbf{x}\in\mathbb{R}^n.$$

Solution

The function g is convex since for $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ and $\lambda\in(0,1)$, we have

$$g(\lambda\mathbf{x}+(1-\lambda)\mathbf{y})=\|\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\|$$
$$\leq\lambda\|\mathbf{x}\|+(1-\lambda)\|\mathbf{y}\|$$
$$=\lambda g(\mathbf{x})+(1-\lambda)g(\mathbf{y})$$

which follows from the triangle inequality and the fact that $\|\alpha\mathbf{u}\|=|\alpha|.\|\mathbf{u}\|$, whenever $\alpha\in\mathbb{R}$ and $\mathbf{u}\in\mathbb{R}^n$.

Example 5.9

Let $\mathbf{A}$ be an $n\times n$ symmetric matrix. It is called positive semidefinite if $\langle\mathbf{Au},\mathbf{u}\rangle\geq 0$ for all $\mathbf{u}\in\mathbb{R}^n$. Let us check that $\mathbf{A}$ is positive semidefinite if and only if the function $f:\mathbb{R}^n\to\mathbb{R}$ defined by

$$f(\mathbf{x})\coloneqq\frac{1}{2}\langle\mathbf{Ax},\mathbf{x}\rangle,\qquad\mathbf{x}\in\mathbb{R}^n$$

is convex.

Solution

Indeed, a direct calculation shows that for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ and $\lambda\in(0,1)$ we have

$$\lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})-f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y})=\frac{1}{2}\lambda(1-\lambda)\langle\mathbf{A}(\mathbf{x}-\mathbf{y}),\mathbf{x}-\mathbf{y}\rangle.$$

If the matrix $\mathbf{A}$ is positive semidefinite, then

$$\langle \mathbf{A}(\mathbf{x}-\mathbf{y}), \mathbf{x}-\mathbf{y} \rangle \geq 0$$

so the function f is convex by (5.19). Conversely, assuming the convexity of f and it is easy to verify that $\mathbf{A}$ is positive semidefinite.
For example, let $h(x) := x^2$ for $x \in \mathbb{R}$. The convexity of the simplest quadratic function h follows from a more general result for the quadratic function on $\mathbb{R}^n$.

The following characterization of convexity is known as the Jensen inequality.

Theorem 5.20: A function $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if for any numbers $\lambda_i \geq 0$ as $i = 1, \ldots, m$ with $\sum_{i=1}^{m} \lambda_i = 1$ and for any elements $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, it holds that

$$f\left(\sum_{i=1}^{m} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{m} \lambda_i f(\mathbf{x}_i). \tag{5.23}$$

Proof:

Since (5.23) immediately implies the convexity of f, we only need to prove that any convex function f satisfies the Jensen inequality (5.23). Arguing by induction and taking into account that for $m = 1$, inequality (5.23) holds trivially and for $m = 2$, inequality (5.23) holds by the definition of convexity, we suppose that it holds for an integer $m := k$ with $k \geq 2$. Fix numbers $\lambda_i \geq 0$, $i = 1, \ldots, k+1$, with $\sum_{i=1}^{k+1} \lambda_i = 1$ and elements $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \ldots, k+1$. We obviously have the relationship

$$\sum_{i=1}^{k} \lambda_i = 1 - \lambda_{k+1}.$$

If $\lambda_{k+1} = 1$, then $\lambda_i = 0$ for all $i = 1, \ldots, k$ and (5.23) obviously holds for $m := k+1$ in this case. Supposing now that $0 \leq \lambda_{k+1} < 1$, we get

$$\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} = 1$$

and by direct calculations based on convexity arrive at

$$\begin{aligned}
f\left(\sum_{i=1}^{k+1} \lambda_i \mathbf{x}_i\right) &= f\left((1-\lambda_{k+1}) \frac{\sum_{i=1}^{k} \lambda_i \mathbf{x}_i}{1-\lambda_{k+1}} + \lambda_{k+1} \mathbf{x}_{k+1}\right) \\
&\leq (1-\lambda_{k+1}) f\left(\frac{\sum_{i=1}^{k} \lambda_i \mathbf{x}_i}{1-\lambda_{k+1}}\right) + \lambda_{k+1} f(\mathbf{x}_{k+1}) \\
&= (1-\lambda_{k+1}) f\left(\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} \mathbf{x}_i\right) + \lambda_{k+1} f(\mathbf{x}_{k+1}) \\
&\leq (1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(\mathbf{x}_i) + \lambda_{k+1} f(\mathbf{x}_{k+1}) \\
&= \sum_{i=1}^{k+1} \lambda_i f(\mathbf{x}_i)
\end{aligned}$$

This justifies inequality (5.23) and completes the proof of the theorem.

■

Definition: The domain and epigraph of $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ are defined, respectively, by

$$\text{dom } f := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < \infty\} \tag{5.24}$$

The graph of a function $f: \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\text{graph } f = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \text{dom } f\}, \tag{5.25}$$

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

$$\begin{aligned} \text{epi } f &:= \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R}: \mathbf{x} \in \mathbb{R}^n, t \geq f(\mathbf{x})\} \\ &= \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R}: \mathbf{x} \in \text{dom } f, t \geq f(\mathbf{x})\} \end{aligned} \tag{5.26}$$

(*i.e.*, all the points in the Cartesian product $\mathbb{R}^n \times \mathbb{R}$ lying on or above its graph.)

('Epi' means 'above' so epigraph means 'above the graph'.) The definition is illustrated in figure 5.12.

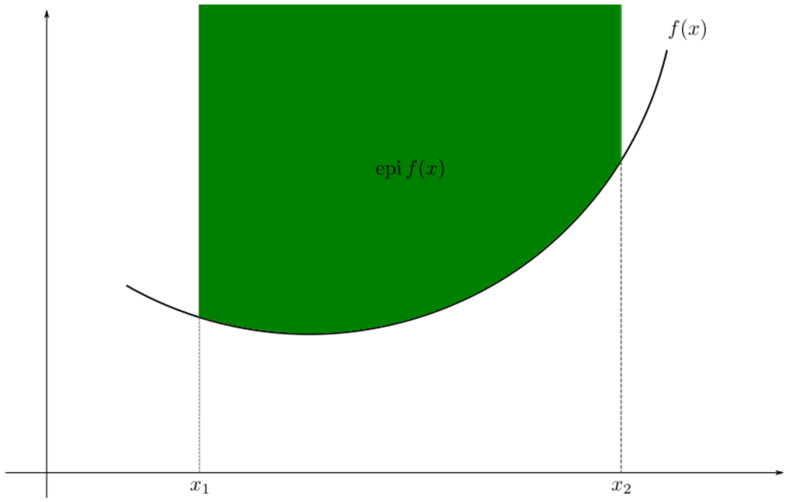

Figure 5.12.

The next theorem gives a geometric characterization of the function convexity via the convexity of the associated epigraphical set.

Theorem 5.21: A function $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its epigraph epi f is a convex subset of the product space $\mathbb{R}^n \times \mathbb{R}$.

Proof:

Assuming that f is convex, fix pairs $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in \text{epi } f$ and a number $\lambda \in (0,1)$. Then we have $f(\mathbf{x}_i) \leq t_i$ for $i = 1,2$. Thus, the convexity of f ensures that

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

This implies therefore that

$$\begin{aligned} \lambda(\mathbf{x}_1, t_1) + (1-\lambda)(\mathbf{x}_2, t_2) &= (\lambda \mathbf{x}_1, \lambda t_1) + \big((1-\lambda)\mathbf{x}_2, (1-\lambda)t_2\big) \\ &= (\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda t_1 + (1-\lambda)t_2) \\ &\in \text{epi } f \end{aligned}$$

and thus, the epigraph epi f is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

Conversely, suppose that the set epi f is convex and fix $\mathbf{x}_1, \mathbf{x}_2 \in \text{dom } f$ and a number $\lambda \in (0,1)$. Then $(\mathbf{x}_1, f(\mathbf{x}_1))$, $(\mathbf{x}_2, f(\mathbf{x}_2)) \in \text{epi } f$. This tells us that

$$\big(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)\big) = \lambda\big(\mathbf{x}_1, f(\mathbf{x}_1)\big) + (1-\lambda)\big(\mathbf{x}_2, f(\mathbf{x}_2)\big) \in \text{epi } f$$

and thus implies the inequality

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

which justifies the convexity of the function f.

■

Now we show that convexity is preserved under some important operations.

Theorem 5.22: Let $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions for all $i = 1, \dots, m$. Then the following functions are convex as well:

(i) The multiplication by scalars λf for any $\lambda > 0$.

(ii) The sum function $\sum_{i=1}^{m} f_i$.
(iii) The maximum function $\max_{1\le i\le m} f_i$.

Proof:

The convexity of λf in (i) follows directly from the definition. It is sufficient to prove (ii) and (iii) for $m = 2$, since the general cases immediately follow by induction.

(ii) Fix any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Then we have

$$\begin{aligned}(f_1 + f_2)(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})\\ &\le \lambda f_1(\mathbf{x}) + (1-\lambda)f_1(\mathbf{y}) + \lambda f_2(\mathbf{x}) + (1-\lambda)f_2(\mathbf{y})\\ &= \lambda(f_1 + f_2)(\mathbf{x}) + (1-\lambda)(f_1 + f_2)(\mathbf{y})\end{aligned}$$

which verifies the convexity of the sum function $f_1 + f_2$.

(iii) Denote $g := max\{f_1, f_2\}$ and get for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in (0,1)$ that

$$\begin{aligned}f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &\le \lambda f_i(\mathbf{x}) + (1-\lambda)f_i(\mathbf{y})\\ &\le \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y})\end{aligned}$$

for $i = 1,2$. This directly implies that

$$\begin{aligned}g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= max\{f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}), f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})\}\\ &\le \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}).\end{aligned}$$

which shows that the maximum function $g(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$ is convex.

■

Convexity conditions, case 1, f is a function of one variable only

Theorem 5.23: Given a convex function $f: \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, assume that its domain is an open interval I. For any $a, b \in I$ and $a < x < b$, we have the inequalities

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \frac{f(b) - f(x)}{b - x} \tag{5.27}$$

Proof:

Fix a, b, x as above and form the numbers $t = \frac{x-a}{b-a} \in (0,1)$. Then

$$\begin{aligned}f(x) &= f\big(a + (x-a)\big)\\ &= f\left(a + \frac{x-a}{b-a}(b-a)\right)\\ &= f\big(a + t(b-a)\big)\\ &= f(tb + (1-t)a).\end{aligned}$$

This gives us the inequalities

$$f(x) \le tf(b) + (1-t)f(a)$$

and

$$\begin{aligned}f(x) - f(a) &\le tf(b) + (1-t)f(a) - f(a)\\ &= t[f(b) - f(a)]\\ &= \frac{x-a}{b-a}(f(b) - f(a))\end{aligned}$$

which can be equivalently written as

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a}$$

Similarly, we have the estimate

$$f(x) - f(b) \le tf(b) + (1-t)f(a) - f(b)$$

$$= (t-1)[f(b) - f(a)]$$
$$= \frac{x-b}{b-a}(f(b) - f(a))$$

which finally implies that

$$\frac{f(b) - f(a)}{b-a} \le \frac{f(b) - f(x)}{b-x}$$

and thus completes the proof of the lemma.

■

Theorem 5.24: Suppose that $f: \mathbb{R} \to \overline{\mathbb{R}}$ is differentiable on its domain, which is an open interval I. Then f is convex if and only if the derivative f' is nondecreasing on I.

Proof:

Suppose that f is convex and fix $a < b$ with $a, b \in I$. By theorem 5.23, we have

$$\frac{f(x) - f(a)}{x-a} \le \frac{f(b) - f(a)}{b-a}$$

for any $x \in (a, b)$. This implies by the derivative definition that

$$f'(a) \le \frac{f(b) - f(a)}{b-a}$$

Similarly, we arrive at the estimate

$$\frac{f(b) - f(a)}{b-a} \le f'(b)$$

and conclude that $f'(a) \le f'(b)$, i.e., f' is a nondecreasing function.

To prove the converse, suppose that f' is nondecreasing on I and fix $x_1 < x_2$ with $x_1, x_2 \in I$ and $t \in (0,1)$. Then

$$x_1 < x_t < x_2 \ \text{ for } \ x_t \coloneqq tx_1 + (1-t)x_2$$

By the mean value theorem, we find c_1, c_2 such that $x_1 < c_1 < x_t < c_2 < x_2$ and

$$f(x_t) - f(x_1) = f'(c_1)(x_t - x_1) = f'(c_1)(1-t)(x_2 - x_1)$$
$$f(x_t) - f(x_2) = f'(c_2)(x_t - x_2) = f'(c_2)t(x_1 - x_2)$$

This can be equivalently rewritten as

$$tf(x_t) - tf(x_1) = f'(c_1)t(1-t)(x_2 - x_1)$$
$$(1-t)f(x_t) - (1-t)f(x_2) = f'(c_2)t(1-t)(x_1 - x_2)$$

Since $f(c_1) \le f(c_2)$, adding these equalities yields

$$f(x_t) \le tf(x_1) + (1-t)f(x_2)$$

which justifies the convexity of the function f.

■

As a result, we have

Theorem 5.25: Let $f: \mathbb{R} \to \overline{\mathbb{R}}$ be a function and let f be differentiable on its domain. Then f is convex if and only if

$$f(x_2) \ge f(x_1) + f'(x_1)(x_2 - x_1) \tag{5.28}$$

for all $x_1, x_2 \in X$.

Proof:

First, we will assume f is convex and try to prove the inequality. Take any $x_1, x_2 \in \mathbb{R}$, and assume $x_1 \neq x_2$ because otherwise the inequality is already satisfied: it just says $f(x) \ge f(x)$. We have

$$f((1-t)x_1 + tx_2) \le (1-t)f(x_1) + tf(x_2)$$

whenever $t \in (0,1)$, which we can rewrite as

$$f\big(x_1 + t(x_2 - x_1)\big) \le (1-t)f(x_1) + tf(x_2)$$
$$\Rightarrow f\big(x_1 + t(x_2 - x_1)\big) - f(x_1) \le tf(x_2) - tf(x_1)$$
$$\Rightarrow \frac{f(x_1 + t(x_2 - x_1)) - f(x_1)}{t(x_2 - x_1)}(x_2 - x_1) \le f(x_2) - f(x_1)$$

If we take the limit as $t \to 0$, then $t(x_2 - x_1) \to 0$ as well, which means the left-hand side of this inequality approaches $f'(x_1)(x_2 - x_1)$. The right-hand side does not depend on t, so it remains the same, and we get

$$f'(x_1)(x_2 - x_1) \le f(x_2) - f(x_1)$$
$$\Rightarrow f(x_2) \ge f(x_1) + f'(x_1)(x_2 - x_1).$$

Next, we will assume that the inequality holds, and try to prove that f is convex. Let $u, v \in \mathbb{R}$ and let $w = tu + (1-t)v$ with $t \in (0,1)$. Then we have

$$f(u) \ge f(w) + f'(w)(u - w)$$
$$f(v) \ge f(w) + f'(w)(v - w)$$

So, if we add t times the first inequality and $(1-t)$ times the second inequality, we get

$$\begin{aligned} tf(u) + (1-t)f(v) &\ge tf(w) + (1-t)f(w) + f'(w)(tu - tw + (1-t)v - (1-t)w) \\ &= f(w) + f'(w)(tu + (1-t)v - w) \\ &= f(w) + f'(w)(w - w) = f(w) \end{aligned}$$

and since $w = tu + (1-t)v$, this is exactly the inequality

$$tf(u) + (1-t)f(v) \ge f(tu + (1-t)v)$$

that proves that f is convex.

■

Theorem 5.26: Let $f: \mathbb{R} \to \overline{\mathbb{R}}$ be twice differentiable on its domain, which is an open interval I. Then f is convex if and only if $f''(x) \ge 0$ for all $x \in I$.

Proof:

Since $f''(x) \ge 0$ for all $x \in I$ if and only if the derivative function f' is nondecreasing on this interval. Then the conclusion follows directly from theorem 5.25.

■

Much simpler! If f is differentiable (or, better yet, twice differentiable) checking these conditions is almost always easier.

Equivalent conditions for strict convexity can be obtained in a natural way, changing $\ge$ to $>$ and requiring that x_1, and x_2 be distinct in theorem 5.25. Essentially, theorem 5.25 says that f lies above its tangent lines, figure 5.13, while theorem 5.26 says that f is always "curving upward." (A convex function lies above its tangents, but below its secants). These conditions are usually easier to verify than that of the definition of convexity.

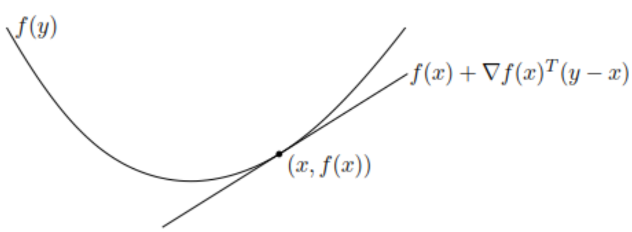

Figure 5.13.

Example 5.10

Show that the following function is convex on $\mathbb{R}$

$$f(x) := \begin{cases} 1/x & \text{if } x > 0 \\ \infty & \text{otherwise} \end{cases}$$

Solution

To verify its convexity, we get that $f''(x) = \frac{2}{x^3} > 0$ for all x belonging to the domain of f, which is $I = (0, \infty)$. Thus, this function is convex on $\mathbb{R}$ by theorem 5.26.

Example 5.11

Show that $f(x) = x^2$ is strictly convex using theorem 5.25.

Solution

Pick any $x_1, x_2 \in \mathbb{R}$ with $x_1 \neq x_2$. We have $f'(x_1) = 2x_1$, so we need to show that

$$x_2^2 > x_1^2 + 2x_1(x_2 - x_1)$$

Expanding the right-hand side and rearranging terms, we see this is equivalent to

$$x_1^2 - 2x_1x_2 + x_2^2 > 0$$

or

$$(x_1 - x_2)^2 > 0$$

which is clearly true since $x_1 \neq x_2$. Thus, f is strictly convex.

Example 5.12

Show that $f(x) = x^2$ is strictly convex using theorem 5.26.

Solution

$f''(x) = 2 > 0$ for all $x \in \mathbb{R}$, so f is strictly convex.

Convexity conditions, case 2, f is a multi-dimensional function

Theorem 5.27: Suppose f is differentiable (i.e., its gradient ∇f exists at each point in dom f, which is open). Then f is convex if and only if dom f is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), (\mathbf{y} - \mathbf{x}) \rangle. \tag{5.29}$$

holds for all $\mathbf{x}, \mathbf{y} \in$ dom f.

Theorem 5.28: Strict convexity can also be characterized by a first-order condition: f is strictly convex if and only if dom f is convex and for $\mathbf{x}, \mathbf{y} \in$ dom f, $\mathbf{x} \neq \mathbf{y}$, we have

$$f(\mathbf{y}) > f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), (\mathbf{y} - \mathbf{x}) \rangle. \tag{5.30}$$

Proof:

we first consider the case $n = 1$: We show that a differentiable function $f: \mathbb{R} \to \mathbb{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x).$$

for all x and y in dom f.

Assume first that f is convex and $x, y \in$ dom f. Since dom f is convex (*i.e.*, an interval), we conclude that for all $0 < t \leq 1$, $x + t(y - x) \in$ dom f, and by convexity of f,

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

If we divide both sides by t, we obtain

$$f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t}.$$

and taking the limit as $t \to 0$ yields (5.29).

To show sufficiency, assume the function satisfies (5.29) for all x and y in dom f (which is an interval). Choose any $x \neq y$, and $0 \leq \theta \leq 1$, and let $z = \theta x + (1-\theta)y$. Applying (5.29) twice yields

$$f(x) \geq f(z) + f'(z)(x-z), \qquad f(y) \geq f(z) + f'(z)(y-z)$$

Multiplying the first inequality by θ, the second by $1-\theta$, and adding them yields

$$\theta f(x) + (1-\theta)f(y) \geq f(z)$$

which proves that f is convex.

Now we can prove the general case, with $f: \mathbb{R}^n \to \mathbb{R}$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and consider f restricted to the line passing through them, *i.e.*, the function defined by

$$g(t) = f(t\mathbf{y} + (1-t)\mathbf{x}), \qquad \text{so} \qquad g'(t) = \langle \nabla f(t\mathbf{y} + (1-t)\mathbf{x}), (\mathbf{y}-\mathbf{x}) \rangle.$$

First assume f is convex, which implies g is convex, so by the argument above we have $g(1) \geq g(0) + g'(0)$, which means

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), (\mathbf{y}-\mathbf{x}) \rangle$$

Now assume that this inequality holds for any x and y, so if $t\mathbf{y} + (1-t)\mathbf{x} \in \text{dom } f$ and $\tilde{t}\mathbf{y} + (1-\tilde{t})\mathbf{x} \in \text{dom } f$, we have

$$f(t\mathbf{y} + (1-t)\mathbf{x}) \geq f(\tilde{t}\mathbf{y} + (1-\tilde{t})\mathbf{x}) + \langle \nabla f(\tilde{t}\mathbf{y} + (1-\tilde{t})\mathbf{x}), (\mathbf{y}-\mathbf{x})(t-\tilde{t}) \rangle$$

i.e., $g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t-\tilde{t})$. We have seen that this implies that g is convex.

■

Notes:

1- The function of $\mathbf{y}$ given by $f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), (\mathbf{y}-\mathbf{x}) \rangle$ is, of course, the first-order Taylor approximation of f near $\mathbf{x}$. The inequality (5.29) states that for a convex function, the first-order Taylor approximation is in fact a global under-estimator of the function, see figure 5.13. Conversely, if the first-order Taylor approximation of a function is always a global under-estimator of the function, then the function is convex.

2- The inequality (5.29) shows that from local information about a convex function (i.e., its value and derivative at a point) we can derive global information (i.e., a global under-estimator of it). This is perhaps the most important property of convex functions and explains some of the remarkable properties of convex functions and convex optimization problems. As one simple example, the inequality (5.29) shows that if $\nabla f(\mathbf{x}) = 0$, then for all $\mathbf{y} \in \text{dom } f$, $f(\mathbf{y}) \geq f(\mathbf{x})$, *i.e.*, $\mathbf{x}$ is a global minimizer of the function f.

Theorem 5.29: We now assume that f is twice differentiable, that is, its Hessian or second derivative $\nabla^2 f$ exists at each point in dom f, which is open. Then f is convex if and only if dom f is convex and its Hessian is positive semidefinite: for all $\mathbf{x} \in \text{dom } f$,

$$\nabla^2 f \succcurlyeq 0 \tag{5.31}$$

For a function on $\mathbb{R}$, this reduces to the simple condition $f''(\mathbf{x}) \geq 0$ (and dom f convex,), which means that the derivative is nondecreasing. The condition $\nabla^2 f \succcurlyeq 0$ can be interpreted geometrically as the requirement that the graph of the function have positive (upward) curvature at $\mathbf{x}$.

Example 5.13

Determine whether the following functions are convex, concave or neither:
1. $f: \mathbb{R} \to \mathbb{R}$, $f(x) = -8x^2$;
2. $f: \mathbb{R}^3 \to \mathbb{R}$, $f(x) = 4x_1^2 + 3x_2^2 + 5x_3^2 + 6x_1x_2 + x_1x_3 - 3x_1 - 2x_2 + 15$;
3. $f: \mathbb{R}^3 \to \mathbb{R}$, $f(x) = 2x_1x_2 - x_1^2 - x_2^2$.

Solution

1. We use theorem 5.29. We first compute the Hessian, which in this case is just the second derivative:

$$\frac{d^2 f}{dx^2}(x) = -16 < 0$$

for all $x \in \mathbb{R}$. Hence, f is concave over $\mathbb{R}$.
2. The Hessian matrix of f is

$$\mathrm{F}(\mathbf{x}) = \begin{pmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{pmatrix}$$

The leading principal minors of $\mathrm{F}(\mathbf{x})$ are

$$\Delta_1 = 8 > 0$$

$$\Delta_2 = \det \begin{pmatrix} 8 & 6 \\ 6 & 6 \end{pmatrix} = 12 > 0$$

$$\Delta_2 = \det \begin{pmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{pmatrix} = 114 > 0$$

Hence, $F(x)$ is positive definite for all $\mathbf{x} \in \mathbb{R}^3$. Therefore, f is a convex function over $\mathbb{R}^3$.
3. The Hessian of f is

$$\mathrm{F}(\mathbf{x}) = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$$

which is negative semidefinite for all $\mathbf{x} \in \mathbb{R}^2$. Hence, f is concave on $\mathbb{R}^2$.

5.5 Sub-gradient and sub-differential of Convex Functions

If f is a differentiable function, then $\mathbf{y} = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle$, is the equation of a hyperplane that is tangent to the surface $\mathbf{y} = f(\mathbf{x})$ at the point $(\mathbf{a}, f(\mathbf{a}))$. And if f is also convex, then $f(\mathbf{x}) \geq f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), (\mathbf{x} - \mathbf{a}) \rangle$ for all $\mathbf{x}$ in the domain of the function, so the tangent plane lies below the graph of the function and is a supporting hyperplane of the epigraph.

The epigraph of an arbitrary convex function is a convex set, by definition. Hence, through each boundary point belonging to the epigraph of a convex function there passes a supporting hyperplane. The supporting hyperplanes of a convex one-variable function f, defined on an open interval, which says that the line $y = f(x_0) + a(x - x_0)$ supports the epigraph at the point $(x_0, f(x_0))$ if (and only if) $f'_-(x_0) \leq a \leq f'_+(x_0)$. The existence of supporting hyperplanes characterizes convexity, and this is a reason for a more detailed study of this concept.

What if f is not differentiable at $\mathbf{x}_t$ (i.e., $\nabla f\ (\mathbf{x}_t)$ does not exist)? Can we still apply this algorithm? In this section, we are going to answer this question. We will introduce sub-gradient, which is a concept closely related to gradient. And for convex functions, we will show that even if the gradient does not exist, the sub-gradient always exists. Then, in convex optimization problems, when the function is nondifferential at certain point, we can use its sub-gradient as an alternative to the gradient.

First, we give the formal definition of sub-gradient.

Definition: A vector $\mathbf{g} \in \mathbb{R}^d$ is a sub-gradient of $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ at $\mathbf{x} \in \mathrm{dom}\, f$, if for any $\mathbf{y} \in \mathrm{dom} f$, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \qquad (5.32)$$

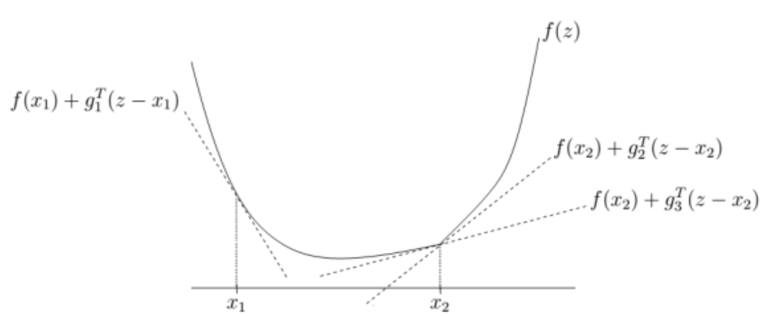

Figure 5.14.

Remark: Clearly, if f is differentiable at $\mathbf{x} \in \mathrm{dom}\, f$, then its sub-gradient at $\mathbf{x}$, $\mathbf{g}$, is equal to $\nabla f(\mathbf{x})$. A function f is called sub-differentiable at $\mathbf{x} \in \mathrm{dom}\, f$ if there exists at least one sub-gradient at $\mathbf{x}$.

Definition: The set of sub-gradients of f at $\mathbf{x} \in \mathrm{dom} f$ is called the sub-differential of f at $\mathbf{x}$, denoted by $\partial f(\mathbf{x})$.

Theorem 5.30: The sub-differential of $f: \mathbb{R}^d \longrightarrow \mathbb{R}$ at $\mathbf{x} \in \mathrm{dom}\, f$ is a closed and convex set.

Proof:

For any $\mathbf{g}_1$, $\mathbf{g}_2 \in \partial f(\mathbf{x})$ and any $\alpha \in [0,1]$, we want to show $\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_2 \in \partial f(\mathbf{x})$. By definition of sub-gradients, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}_1, (\mathbf{y}-\mathbf{x})\rangle,$$
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}_2, (\mathbf{y}-\mathbf{x})\rangle,$$

Multiplying the first inequality by α and the second one by $(1-\alpha)$, and then adding them together leads to

$$\begin{aligned}\alpha f(\mathbf{y}) + (1-\alpha) f(\mathbf{y}) &\geq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{x}) + \langle(\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_2), (\mathbf{y}-\mathbf{x})\rangle \\ &= f(\mathbf{x}) + \langle(\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_2), (\mathbf{y}-\mathbf{x})\rangle.\end{aligned}$$

Thus, by definition, we have $\alpha\mathbf{g}_1 + (1-\alpha)\mathbf{g}_2 \in \partial f(\mathbf{x})$

■

The following theorem is very useful and is simple to proof.

Theorem 5.31: A point $\mathbf{x}^*$ is a minimizer of a convex function $f: \mathbb{R}^d \longrightarrow \mathbb{R}$, if and only if $\mathbf{0} \in \partial f(\mathbf{x}^*)$.

Proof:

If $\mathbf{x}^*$ is the minimizer of f, then for any $\mathbf{y} \in \mathrm{dom}\, f$, we have $f(\mathbf{y}) \geq f(\mathbf{x}^*)$, which can be rewritten as

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \langle \mathbf{0}, (\mathbf{y}-\mathbf{x}^*)\rangle$$

Therefore, by definition, $\mathbf{0}$ is a sub-gradient of f at $\mathbf{x}^*$ and $\mathbf{0} \in \partial f(\mathbf{x}^*)$.

Conversely, if $\mathbf{0} \in \partial f(\mathbf{x}^*)$, then for any $\mathbf{y} \in \mathrm{dom}\, f$, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \langle \mathbf{0}, (\mathbf{y}-\mathbf{x}^*)\rangle = f(\mathbf{x}^*)$$

Clearly, by definition, $\mathbf{x}^*$ is a minimizer of f.

■

Theorem 5.32: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ convex and let $\tilde{\mathbf{x}} \in \mathrm{dom}\, f$ be a local minimizer of f. Then f attains its global minimum at this point.

Proof:

Since $\tilde{\mathbf{x}}$ is a local minimizer of f, there is $\delta > 0$ such that

$$f(\mathbf{u}) \geq f(\tilde{\mathbf{x}})$$

for all $\mathbf{u} \in I\!B(\tilde{\mathbf{x}}, \delta)$. Fix $\mathbf{x} \in \mathbb{R}^n$ and construct a sequence of $\mathbf{x}_k = (1-k^{-1})\tilde{\mathbf{x}} + k^{-1}\tilde{\mathbf{x}}$ as $k \in \mathbb{N}$. Thus, we have $\mathbf{x}_k \in I\!B(\tilde{\mathbf{x}}, \delta)$ when k is sufficiently large. It follows from the convexity of f that

$$f(\tilde{\mathbf{x}}) \leq f(\mathbf{x}_k) \leq (1-k^{-1}) f(\tilde{\mathbf{x}}) + k^{-1} f(\mathbf{x})$$

which readily implies that $k^{-1} f(\tilde{\mathbf{x}}) \leq k^{-1} f(\mathbf{x})$, and hence $f(\tilde{\mathbf{x}}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

■

Theorem 5.33: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be convex and differentiable at $\tilde{\mathbf{x}} \in \mathrm{int}\,(\mathrm{dom}\, f)$. Then we have

$$\partial f(\tilde{\mathbf{x}}) = \{\nabla f(\tilde{\mathbf{x}})\} \tag{5.33}$$

and

$$\langle \nabla f(\tilde{\mathbf{x}}), \mathbf{x} - \tilde{\mathbf{x}}\rangle \geq f(\mathbf{x}) + f(\tilde{\mathbf{x}}) \text{ for all } \mathbf{x} \in \mathbb{R}^n \tag{5.34}$$

Proof:

It follows from the differentiability of f at $\tilde{\mathbf{x}}$ that for any $\epsilon > 0$ there is $\delta > 0$ with

$$-\epsilon\|\mathbf{x}-\tilde{\mathbf{x}}\| \le f(\mathbf{x}) - f(\tilde{\mathbf{x}}) - \langle \nabla f(\tilde{\mathbf{x}}), \mathbf{x}-\tilde{\mathbf{x}}\rangle \le \epsilon\|\mathbf{x}-\tilde{\mathbf{x}}\| \text{ whenever } \|\mathbf{x}-\tilde{\mathbf{x}}\| < \delta \tag{*}$$

Consider further the convex function

$$\varphi(\mathbf{x}) := f(\mathbf{x}) - f(\tilde{\mathbf{x}}) - \langle \nabla f(\tilde{\mathbf{x}}), \mathbf{x}-\tilde{\mathbf{x}}\rangle + \epsilon\|\mathbf{x}-\tilde{\mathbf{x}}\|, \qquad \mathbf{x} \in \mathbb{R}^n$$

and observe that $\varphi(\mathbf{x}) \ge \varphi(\tilde{\mathbf{x}}) = 0$ for all $\mathbf{x} \in IB(\tilde{\mathbf{x}}, \delta)$. The convexity of φ ensures that $\varphi(\mathbf{x}) \ge \varphi(\tilde{\mathbf{x}})$ for all $\mathbf{x} \in \mathbb{R}^n$. Thus

$$\langle \nabla f(\tilde{\mathbf{x}}), \mathbf{x}-\tilde{\mathbf{x}}\rangle \le f(\mathbf{x}) - f(\tilde{\mathbf{x}}) + \epsilon\|\mathbf{x}-\tilde{\mathbf{x}}\|, \quad \text{whenever } \mathbf{x} \in \mathbb{R}^n$$

which yields (5.34) by letting $\epsilon \downarrow 0$.

It follows from (5.34) that $\nabla f(\tilde{\mathbf{x}}) \in \partial f(\tilde{\mathbf{x}})$. Picking now $\mathbf{v} \in \partial f(\tilde{\mathbf{x}})$, we get

$$\langle \mathbf{v}, \mathbf{x}-\tilde{\mathbf{x}}\rangle \le f(\mathbf{x}) - f(\tilde{\mathbf{x}})$$

Then the second part of (*) gives us that

$$\langle \mathbf{v} - \nabla f(\tilde{\mathbf{x}}), \mathbf{x}-\tilde{\mathbf{x}}\rangle \le \epsilon\|\mathbf{x}-\tilde{\mathbf{x}}\|, \quad \text{whenever } \|\mathbf{x}-\tilde{\mathbf{x}}\| < \delta$$

Finally, we observe that $\|\mathbf{v} - \nabla f(\tilde{\mathbf{x}})\| \le \epsilon$, which yields $\mathbf{v} = \nabla f(\tilde{\mathbf{x}})$ since $\epsilon > 0$ was chosen arbitrarily. Thus $\partial f(\tilde{\mathbf{x}}) = \{\nabla f(\tilde{\mathbf{x}})\}$.

■

Theorem 5.34: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a strictly convex function. Then the differentiability of f at $\tilde{\mathbf{x}} \in \operatorname{int}(\operatorname{dom} f)$ implies the strict inequality

$$\langle \nabla f(\tilde{\mathbf{x}}), \mathbf{x}-\tilde{\mathbf{x}}\rangle < f(\mathbf{x}) - f(\tilde{\mathbf{x}}) \quad \text{whenever } \mathbf{x} \ne \tilde{\mathbf{x}} \tag{5.35}$$

Proof:

Since f is convex, we get from theorem 5.33 that

$$\langle \nabla f(\tilde{\mathbf{x}}), \mathbf{u}-\tilde{\mathbf{x}}\rangle < f(\mathbf{u}) - f(\tilde{\mathbf{x}}) \quad \text{for all } \mathbf{u} \in \mathbb{R}^n$$

Fix $\mathbf{x} \ne \tilde{\mathbf{x}}$ and let $\mathbf{u} = \frac{\mathbf{x}+\tilde{\mathbf{x}}}{2}$. It follows from the above that

$$\langle \nabla f(\tilde{\mathbf{x}}), \frac{\mathbf{x}+\tilde{\mathbf{x}}}{2} - \tilde{\mathbf{x}}\rangle \le f\left(\frac{\mathbf{x}+\tilde{\mathbf{x}}}{2}\right) - f(\tilde{\mathbf{x}}) < \frac{1}{2}\mathrm{f}(\mathbf{x}) + \frac{1}{2}\mathrm{f}(\tilde{\mathbf{x}}) - \mathrm{f}(\tilde{\mathbf{x}})$$

Thus, we arrive at the strict inequality (5.35) and complete the proof.

■

Example 5.14

Let $p(\mathbf{x}) := \|\mathbf{x}\|$ be the Euclidean norm function on $\mathbb{R}^n$. Then we have

$$\partial p(\mathbf{x}) = \begin{cases} IB & \text{if } \mathbf{x} = \mathbf{0} \\ \left\{\dfrac{\mathbf{x}}{\|\mathbf{x}\|}\right\} & \text{otherwise} \end{cases}$$

Solution

To verify this, observe first that the Euclidean norm function p is differentiable at any nonzero point with

$$\begin{aligned} \nabla p(\mathbf{x}) &= \nabla (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}} \\ &= \frac{1}{2}(x_1^2 + \cdots + x_n^2)^{-\frac{1}{2}} \times 2 \times \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= \frac{\mathbf{x}}{\|\mathbf{x}\|} \end{aligned}$$

as $\mathbf{x} \ne \mathbf{0}$.

It remains to calculate its subdifferential at $\mathbf{x} = \mathbf{0}$. To proceed by definition (5.32), we have that $\mathbf{v} \in \partial p(\mathbf{0})$ if and only if

$$\langle \mathbf{v}, \mathbf{x}\rangle = \langle \mathbf{v}, \mathbf{x} - \mathbf{0}\rangle \le p(\mathbf{x}) - p(\mathbf{0}) = \|\mathbf{x}\| \text{ for all } \mathbf{v} \in \mathbb{R}^n$$

Letting $\mathbf{x} = \mathbf{v}$ gives us

$$\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2 \le \|\mathbf{v}\|$$
which implies that $\|\mathbf{v}\| \le 1$, i.e., $\mathbf{v} \in IB$. Now take $\mathbf{v} \in IB$ and deduce from the Cauchy-Schwarz inequality that
$$\begin{aligned}\langle \mathbf{v}, \mathbf{x} - \mathbf{0} \rangle &= \langle \mathbf{v}, \mathbf{x} \rangle \\ &\le \|\mathbf{v}\| \cdot \|\mathbf{x}\| \\ &\le \|\mathbf{x}\| \\ &= p(\mathbf{x}) - p(\mathbf{0}) \text{ for all } \mathbf{x} \in \mathbb{R}^n\end{aligned}$$
and thus $\mathbf{v} \in \partial p(\mathbf{0})$, which shows that $\partial p(\mathbf{0}) = IB$.

Theorem 5.35: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a differentiable function on its domain D, which is an open convex set. Then f is convex if and only if
$$\langle \nabla f(\mathbf{u}), \mathbf{x} - \mathbf{u} \rangle \le f(\mathbf{x}) - f(\mathbf{u}) \quad \text{for all } \mathbf{u}, \mathbf{x} \in D \tag{5.36}$$

Proof:

The "only if " part follows from theorem 5.33. To justify the converse, suppose that (5.36) holds and then fix any $\mathbf{x}_1, \mathbf{x}_2 \in D$ and $t \in (0,1)$. Denoting $\mathbf{x}_t \coloneqq t\mathbf{x}_1 + (1-t)\mathbf{x}_2$, we have $\mathbf{x}_t \in D$ by the convexity of D. Then

$$\langle \nabla f(\mathbf{x}_t), \mathbf{x}_1 - \mathbf{x}_t \rangle \le f(\mathbf{x}_1) - f(\mathbf{x}_t)$$
$$\langle \nabla f(\mathbf{x}_t), \mathbf{x}_2 - \mathbf{x}_t \rangle \le f(\mathbf{x}_2) - f(\mathbf{x}_t)$$
It follows furthermore that

$$t\langle \nabla f(\mathbf{x}_t), \mathbf{x}_1 - \mathbf{x}_t \rangle \le tf(\mathbf{x}_1) - tf(\mathbf{x}_t)$$
$$(1-t)\langle \nabla f(\mathbf{x}_t), \mathbf{x}_2 - \mathbf{x}_t \rangle \le (1-t)f(\mathbf{x}_2) - (1-t)f(\mathbf{x}_t)$$
Summing up these inequalities, we arrive at

$$0 \le tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2) - f(\mathbf{x}_t)$$
which ensures that $f(\mathbf{x}_t) \le tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2)$, and so verifies the convexity of f.

■

Theorem 5.36: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be convex and twice continuously differentiable on an open subset V of its domain containing $\tilde{\mathbf{x}}$. Then we have
$$\langle \nabla^2 f(\tilde{\mathbf{x}}), \mathbf{u} \rangle \ge 0 \quad \text{for all } \mathbf{u} \in \mathbb{R}^n \tag{5.37}$$
where $\nabla^2 f(\tilde{\mathbf{x}})$ is the Hessian matrix of f at $\tilde{\mathbf{x}}$.

Proof:

Let $\mathbf{A} := \nabla^2 f(\tilde{\mathbf{x}})$, which is symmetric matrix. Then

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(\tilde{\mathbf{x}} + \mathbf{h}) - f(\tilde{\mathbf{x}}) - \langle \nabla f(\tilde{\mathbf{x}}), \mathbf{h} \rangle - \frac{1}{2}\langle \mathbf{A}\mathbf{h}, \mathbf{h} \rangle}{\|\mathbf{h}\|^2} = 0 \tag{*}$$
It follows from (*) that for any $\epsilon > 0$ there is $\delta > 0$ such that

$$-\epsilon\|\mathbf{h}\|^2 \le f(\tilde{\mathbf{x}} + \mathbf{h}) - f(\tilde{\mathbf{x}}) - \langle \nabla f(\tilde{\mathbf{x}}), \mathbf{h} \rangle - \frac{1}{2}\langle \mathbf{A}\mathbf{h}, \mathbf{h} \rangle \le \epsilon\|\mathbf{h}\|^2 \;\text{ for all } \|\mathbf{h}\| \le \delta$$
By theorem 5.33, we readily have

$$f(\tilde{\mathbf{x}} + \mathbf{h}) - f(\tilde{\mathbf{x}}) - \langle \nabla f(\tilde{\mathbf{x}}), \mathbf{h} \rangle \ge 0$$
Combining the above inequalities ensures that

$$-\epsilon\|\mathbf{h}\|^2 \le \frac{1}{2}\langle \mathbf{A}\mathbf{h}, \mathbf{h} \rangle \;\text{ whenever }\; \|\mathbf{h}\| \le \delta \tag{**}$$
Observe further that for any $\mathbf{0} \ne \mathbf{u} \in \mathbb{R}^n$ the element $\mathbf{h} \coloneqq \frac{\mathbf{u}}{\|\mathbf{u}\|}$ satisfies $\|\mathbf{h}\| \le \delta$ and, being substituted into (**), gives us the estimate

$$-\epsilon\delta^2 \leq \frac{1}{2}\delta^2 \langle \mathbf{A}\frac{\mathbf{u}}{\|\mathbf{u}\|}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\rangle$$

It shows therefore (since the case where $\mathbf{u} = \mathbf{0}$ is trivial) that

$$-2\epsilon\|\mathbf{u}\|^2 \leq \langle \mathbf{Au}, \mathbf{u}\rangle \text{ whenever } \mathbf{u} \in \mathbb{R}^n$$

which implies by letting $\epsilon \downarrow 0$ that $0 \leq \langle \mathbf{Au}, \mathbf{u}\rangle$ for all $\mathbf{u} \in \mathbb{R}^n$.

■

Theorem 5.37: Let $f: \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a function twice continuously differentiable on its domain D, which is an open convex subset of $\mathbb{R}^n$. Then f is convex if and only if $\nabla^2 f(\tilde{\mathbf{x}})$ is positive semidefinite for every $\tilde{\mathbf{x}} \in D$.

Proof:

Taking theorem 5.36 into account, we only need to verify that if $\nabla^2 f(\tilde{\mathbf{x}})$ is positive semidefinite for every $\tilde{\mathbf{x}} \in D$, then f is convex. To proceed, for any $\mathbf{x}_1, \mathbf{x}_2 \in D$ define $\mathbf{x}_t \coloneqq t\mathbf{x}_1 + (1-t)\mathbf{x}_2$ and consider the function

$$\varphi(t) := f(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) - tf(\mathbf{x}_1) - (1-t)f(\mathbf{x}_2), \qquad t \in \mathbb{R}$$

It is clear that φ is well defined on an open interval I containing $(0,1)$. Then

$$\varphi'(t) = \langle \nabla f(\mathbf{x}_t), \mathbf{x}_1 - \mathbf{x}_2\rangle - f(\mathbf{x}_1) + f(\mathbf{x}_2)$$

and

$$\varphi''(t) = \langle \nabla^2 f(\mathbf{x}_t)(\mathbf{x}_1 - \mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2\rangle \geq 0$$

for every $t \in I$ since $\nabla^2 f(\mathbf{x}_t)$ is positive semidefinite. Hence, φ is convex on I. Since $\varphi(0) = \varphi(1) = 0$, for any $t \in (0,1)$ we have

$$\varphi(t) = \varphi(t(1) + (1-t)0) \leq t\varphi(1) + (1-t)\varphi(0) = 0,$$

which implies in turn the inequality

$$f(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \leq tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2)$$

This justifies that the function f is convex on its domain and thus on $\mathbb{R}^n$.

■

We now turn to an important theorem, which shows that for a convex function, its sub-differential is always nonempty set.

Theorem 5.38: (Existence of sub-gradient for convex functions): If $f: \mathbb{R}^d \longrightarrow \mathbb{R}$ is convex, and $\mathbf{x} \in$ int dom f, then $\partial f(\mathbf{x})$ is nonempty.

Now let us see some examples of sub-gradients.

Example 5.15

For $f(x) = |x|$,

$$\partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1.1] & \text{if } x = 0 \end{cases}$$

Solution

Note that $f(x)$ is differentiable when $x > 0$ or $x < 0$, so the subgradient of f is equal to its gradient (1 and -1, respectively). At $x = 0$, for any $y \in \mathbb{R}$, its sub-gradient g should satisfy

$$|y| \geq |0| + g(y - 0). \qquad \text{i.e.,} \qquad g \cdot y \leq |y|$$

Thus, $g \in [-1.1]$.

Example 5.16

Let $f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^{d} |\mathbf{x}_i|$, and $\mathbf{g} = (g_1. \cdots . g_d)^T$ be a sub-gradient of f at $\mathbf{x}$.

Solution

Then we have

$$g_i = \begin{cases} 1 & \text{if } x_i > 0 \\ -1 & \text{if } x_i < 0. i = 1. \cdots . d \\ \in [-1.1] & \text{if } x_i = 0 \end{cases}$$

Finally, consider the subdifferential of point-wise maximum function.

Example 5.17

Let $f(\mathbf{x}) = \max(f_1(\mathbf{x}), f_2(\mathbf{x}))$, where f_1 and f_2 are convex and differentiable.

Solution

Then we have

$$\partial f(\mathbf{x}) = \begin{cases} \{\nabla f_1(\mathbf{x})\} & \text{if } f_1(\mathbf{x}) > f_2(\mathbf{x}) \\ \{\nabla f_2(\mathbf{x})\} & \text{if } f_2(\mathbf{x}) > f_1(\mathbf{x}) \\ \text{conv}(\{\nabla f_1(\mathbf{x}) \cdot \nabla f_2(\mathbf{x})\}) & \text{if } f_1(\mathbf{x}) = f_2(\mathbf{x}) \end{cases}$$

Clearly, when $f_1(\mathbf{x}) \neq f_2(\mathbf{x})$, $f(\mathbf{x})$ is equal to either $f_1(\mathbf{x})$ or $f_2(\mathbf{x})$ and is therefore differentiable.
When $f_1(\mathbf{x}) = f_2(\mathbf{x})$, both $\nabla f_1(\mathbf{x})$ and $\nabla f_2(\mathbf{x})$ are sub-gradients of $f(\mathbf{x})$. Since subdifferential is a convex set, $\partial f(\mathbf{x})$ should be a convex hull of $\nabla f_1(\mathbf{x})$ and $\nabla f_2(\mathbf{x})$.

It is essential to calculate the sub-gradient for non-differential functions. We have the following rules in calculating the sub-gradient.

1. Non-negative Scaling

Suppose $f(\mathbf{x})$ is convex, its domain $\text{dom} f$ and that $\alpha > 0$ then

$$\partial[\alpha f(\mathbf{x})] = \alpha \partial f(\mathbf{x}) \tag{5.38}$$

Note that here " $=$ " indicates set equality, i.e., for two sets A and B, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

2. Summation

Suppose $f(\mathbf{x}) = \sum_{i=1}^{n} f_i(\mathbf{x})$ where all of the $f_i's$ are convex. Then

$$\partial f(\mathbf{x}) = \sum_{i=1}^{n} \partial f_i(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}) + \cdots + \partial f_n(\mathbf{x}) \tag{5.39}$$

Note that here "$+$" indicates set sum, i.e., for two sets A and B, $\mathbf{A} + \mathbf{B} = \{(a + b): a \in \mathbf{A}, b \in \boldsymbol{B}\}$.