Chapter 7
Convex Functions

7.1 ▪ Definition and Examples

In the last chapter we introduced the notion of a convex set. This chapter is devoted to the concept of convex functions, which is fundamental in the theory of optimization.

Definition 7.1 (convex functions). *A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called* **convex** *(or* **convex over** *C) if*

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C, \lambda \in [0,1]. \tag{7.1}$$

The fundamental inequality (7.1) is illustrated in Figure 7.1.

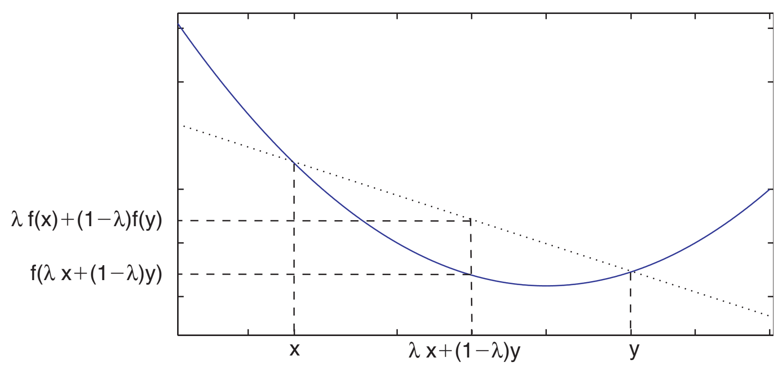

Figure 7.1. *Illustration of the inequality $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$.*

In case when no domain is specified, then we naturally assume that f is defined over the entire space $\mathbb{R}^n$. If we do not allow equality in (7.1) when $\mathbf{x} \neq \mathbf{y}$ and $\lambda \in (0,1)$, the function is called *strictly convex*.

Definition 7.2 (strictly convex functions). *A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called* **strictly convex** *if*

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \text{ for any } \mathbf{x} \neq \mathbf{y} \in C, \lambda \in (0,1).$$

Another important concept is *concavity*. A function is called *concave* if $-f$ is convex. Similarly, f is called *strictly concave* if $-f$ is strictly convex. We can of course write a more direct definition of concavity based on the definition of convexity. A function f is concave over a convex set $C \subseteq \mathbb{R}^n$ if and only if for any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$ we have

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

Equipped only with the definition of convexity, we can give some elementary examples of convex functions. We begin by showing the convexity of affine functions, which are functions of the form $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$, where $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. (If $b = 0$, then f is also called *linear*.)

Example 7.3 (convexity of affine functions). Let $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$, where $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. To show that f is convex, take $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) &= \mathbf{a}^T(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) + b \\ &= \lambda(\mathbf{a}^T\mathbf{x}) + (1-\lambda)(\mathbf{a}^T\mathbf{y}) + \lambda b + (1-\lambda)b \\ &= \lambda(\mathbf{a}^T\mathbf{x} + b) + (1-\lambda)(\mathbf{a}^T\mathbf{y} + b) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}), \end{aligned}$$

and thus in particular $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$, and convexity follows. Of course, if f is an affine function, then so is $-f$, which implies that affine functions (and in fact, as shown in Exercise 7.3, only affine functions) are both convex and concave. ∎

Example 7.4 (convexity of norms). Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. We will show that the norm function $f(\mathbf{x}) = \|\mathbf{x}\|$ is convex. Indeed, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$. Then by the triangle inequality we have

$$\begin{aligned} f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) &= \|\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\| \\ &\leq \|\lambda \mathbf{x}\| + \|(1-\lambda)\mathbf{y}\| \\ &= \lambda\|\mathbf{x}\| + (1-\lambda)\|\mathbf{y}\| \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}), \end{aligned}$$

establishing the convexity of f. ∎

The basic property characterizing a convex function is that the function value of a convex combination of *two* points $\mathbf{x}$ and $\mathbf{y}$ is smaller than or equal to the corresponding convex combination of the function values $f(\mathbf{x})$ and $f(\mathbf{y})$. An interesting result is that convexity implies that this property can be generalized to convex combinations of *any* number of vectors. This is the so-called Jensen's inequality.

Theorem 7.5 (Jensen's inequality). *Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set $C \subseteq \mathbb{R}^n$. Then for any $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in C$ and $\lambda \in \Delta_k$, the following inequality holds:*

$$f\left(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i). \tag{7.2}$$

Proof. We will prove the inequality (7.2) by induction on k. For $k = 1$ the result is obvious (it amounts to $f(\mathbf{x}_1) \leq f(\mathbf{x}_1)$ for any $\mathbf{x}_1 \in C$). The induction hypothesis is that for any k

vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in C$ and any $\lambda \in \Delta_k$, the inequality (7.2) holds. We will now prove the theorem for $k+1$ vectors. Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1} \in C$ and that $\lambda \in \Delta_{k+1}$. We will show that $f(\mathbf{z}) \leq \sum_{i=1}^{k+1} \lambda_i f(\mathbf{x}_i)$, where $\mathbf{z} = \sum_{i=1}^{k+1} \lambda_i \mathbf{x}_i$. If $\lambda_{k+1} = 1$, then $\mathbf{z} = \mathbf{x}_{k+1}$ and (7.2) is obvious. If $\lambda_{k+1} < 1$, then

$$\begin{aligned} f(\mathbf{z}) &= f\left(\sum_{i=1}^{k+1} \lambda_i \mathbf{x}_i\right) \\ &= f\left(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i + \lambda_{k+1}\mathbf{x}_{k+1}\right) && (7.3) \\ &= f\left((1-\lambda_{k+1})\underbrace{\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}}\mathbf{x}_i}_{\mathbf{v}} + \lambda_{k+1}\mathbf{x}_{k+1}\right) && (7.4) \\ &\leq (1-\lambda_{k+1}) f(\mathbf{v}) + \lambda_{k+1} f(\mathbf{x}_{k+1}). && (7.5) \end{aligned}$$

Since $\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = \frac{1-\lambda_{k+1}}{1-\lambda_{k+1}} = 1$, it follows that $\mathbf{v}$ is a convex combination of k points from C, and hence by the induction hypothesis we have that $f(\mathbf{v}) \leq \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(\mathbf{x}_i)$, which combined with (7.5) yields

$$f(\mathbf{z}) \leq \sum_{i=1}^{k+1} \lambda_i f(\mathbf{x}_i). \quad \square$$

7.2 ▪ First Order Characterizations of Convex Functions

Convex functions are not necessarily differentiable, but in case they are, we can replace the Jensen's inequality definition with other characterizations which utilize the gradient of the function. An important characterizing inequality is *the gradient inequality*, which essentially states that the tangent hyperplanes of convex functions are always underestimates of the function.

Theorem 7.6 (the gradient inequality). *Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if*

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C. \tag{7.6}$$

Proof. Suppose first that f is convex. Let $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in (0,1]$. If $\mathbf{x} = \mathbf{y}$, then (7.6) trivially holds. We will therefore assume that $\mathbf{x} \neq \mathbf{y}$. Then

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

and hence

$$\frac{f(\mathbf{x} + \lambda(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\lambda} \leq f(\mathbf{y}) - f(\mathbf{x}).$$

Taking $\lambda \to 0^+$, the left-hand side converges to the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{y}-\mathbf{x}$, so that

$$f'(\mathbf{x}; \mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}).$$

Since f is continuously differentiable, it follows that $f'(\mathbf{x};\mathbf{y}-\mathbf{x}) = \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$, and hence (7.6) follows.

To prove the reverse direction, assume the gradient inequality holds. Let $\mathbf{z},\mathbf{w} \in C$, and let $\lambda \in (0,1)$. We will show that $f(\lambda\mathbf{z}+(1-\lambda)\mathbf{w}) \leq \lambda f(\mathbf{z})+(1-\lambda)f(\mathbf{w})$. Let $\mathbf{u} = \lambda\mathbf{z}+(1-\lambda)\mathbf{w} \in C$. Then

$$\mathbf{z}-\mathbf{u} = \frac{\mathbf{u}-(1-\lambda)\mathbf{w}}{\lambda} - \mathbf{u} = -\frac{1-\lambda}{\lambda}(\mathbf{w}-\mathbf{u}).$$

Invoking the gradient inequality on the pairs $\mathbf{z},\mathbf{u}$ and $\mathbf{w},\mathbf{u}$, we obtain

$$f(\mathbf{u}) + \nabla f(\mathbf{u})^T(\mathbf{z}-\mathbf{u}) \leq f(\mathbf{z}),$$
$$f(\mathbf{u}) - \frac{\lambda}{1-\lambda}\nabla f(\mathbf{u})^T(\mathbf{z}-\mathbf{u}) \leq f(\mathbf{w}).$$

Multiplying the first inequality by $\frac{\lambda}{1-\lambda}$ and adding it to the second one, we obtain

$$\frac{1}{1-\lambda}f(\mathbf{u}) \leq \frac{\lambda}{1-\lambda}f(\mathbf{z}) + f(\mathbf{w}),$$

which after multiplication by $1-\lambda$ amounts to the desired inequality:

$$f(\mathbf{u}) \leq \lambda f(\mathbf{z}) + (1-\lambda)f(\mathbf{w}). \quad \square$$

A modification of the above proof will show that a function is *strictly convex* if and only if the gradient inequality is satisfied with strict inequality for any $\mathbf{x} \neq \mathbf{y}$.

Theorem 7.7 (the gradient inequality for strictly convex function). *Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is strictly convex over C if and only if*

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) < f(\mathbf{y}) \textit{ for any } \mathbf{x},\mathbf{y} \in C \textit{ satisfying } \mathbf{x} \neq \mathbf{y}.$$

Geometrically, the gradient inequality essentially states that for convex functions, the tangent hyperplane is below the surface of the function. A two-dimensional illustration is given in Figure 7.2.

A direct result of the gradient inequality is that the first order optimality condition $\nabla f(\mathbf{x}^*) = \mathbf{0}$ is sufficient for global optimality.

Proposition 7.8 (sufficiency of stationarity under convexity). *Let f be a continuously differentiable function which is convex over a convex set $C \subseteq \mathbb{R}^n$. Suppose that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ for some $\mathbf{x}^* \in C$. Then $\mathbf{x}^*$ is a global minimizer of f over C.*

Proof. Let $\mathbf{z} \in C$. Plugging $\mathbf{x} = \mathbf{x}^*$ and $\mathbf{y} = \mathbf{z}$ in the gradient inequality (7.6), we obtain that

$$f(\mathbf{z}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{z}-\mathbf{x}^*),$$

which by the fact that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ implies that $f(\mathbf{z}) \geq f(\mathbf{x}^*)$, thus establishing that $\mathbf{x}^*$ is a global minimizer of f over C. $\square$

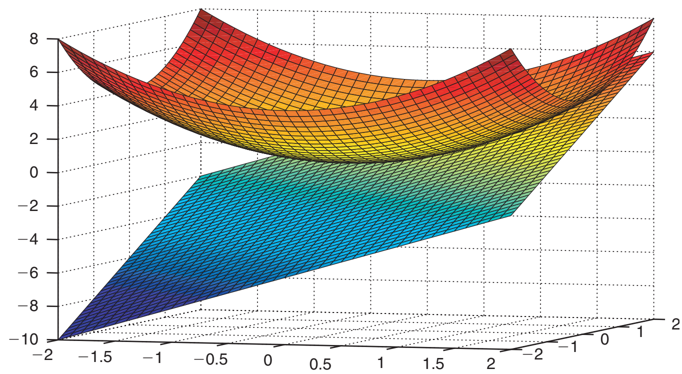

Figure 7.2. *The function $f(x,y)=x^2+y^2$ and its tangent hyperplane at $(1,1)$, which is a lower bound of the function's surface.*

We note that Proposition 7.8 establishes only the sufficiency of the stationarity condition $\nabla f(\mathbf{x}^*)=\mathbf{0}$ for guaranteeing that $\mathbf{x}^*$ is a global optimal solution. When C is not the entire space, this condition is not necessary. However, when $C=\mathbb{R}^n$, then by Theorem 2.6, this is also a necessary condition, and we can thus write the following statement.

Theorem 7.9 (necessity and sufficiency of stationarity). *Let $f:\mathbb{R}^n\rightarrow\mathbb{R}$ be a continuously differentiable convex function. Then $\nabla f(\mathbf{x}^*)=\mathbf{0}$ if and only if $\mathbf{x}^*$ is a global minimum point of f over $\mathbb{R}^n$.*

Using the gradient inequality we can now establish the conditions under which a quadratic function is convex/strictly convex.

Theorem 7.10 (convexity and strict convexity of quadratic functions with positive semidefinite matrices). *Let $f:\mathbb{R}^n\rightarrow\mathbb{R}$ be the quadratic function given by $f(\mathbf{x})=\mathbf{x}^T\mathbf{A}\mathbf{x}+2\mathbf{b}^T\mathbf{x}+c$, where $\mathbf{A}\in\mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b}\in\mathbb{R}^n$, and $c\in\mathbb{R}$. Then f is (strictly) convex if and only if $\mathbf{A}\succeq\mathbf{0}$ ($\mathbf{A}\succ\mathbf{0}$).*

Proof. By Theorem 7.6 the convexity of f is equivalent to the validity of the gradient inequality:

$$f(\mathbf{y})\geq f(\mathbf{x})+\nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \text{ for any } \mathbf{x},\mathbf{y}\in\mathbb{R}^n,$$

which can be written explicitly as

$$\mathbf{y}^T\mathbf{A}\mathbf{y}+2\mathbf{b}^T\mathbf{y}+c\geq\mathbf{x}^T\mathbf{A}\mathbf{x}+2\mathbf{b}^T\mathbf{x}+c+2(\mathbf{A}\mathbf{x}+\mathbf{b})^T(\mathbf{y}-\mathbf{x}) \text{ for any } \mathbf{x},\mathbf{y}\in\mathbb{R}^n.$$

After some rearrangement of terms, we can rewrite the latter inequality as

$$(\mathbf{y}-\mathbf{x})^T\mathbf{A}(\mathbf{y}-\mathbf{x})\geq 0 \text{ for any } \mathbf{x},\mathbf{y}\in\mathbb{R}^n. \tag{7.7}$$

Making the transformation $\mathbf{d}=\mathbf{y}-\mathbf{x}$, we conclude that inequality (7.7) is equivalent to the inequality $\mathbf{d}^T\mathbf{A}\mathbf{d}\geq 0$ for any $\mathbf{d}\in\mathbb{R}^n$, which is the same as saying that $\mathbf{A}\succeq\mathbf{0}$. To prove the strict convexity variant, note that strict convexity of f is the same as

$$f(\mathbf{y})>f(\mathbf{x})+\nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \text{ for any } \mathbf{x},\mathbf{y}\in\mathbb{R}^n \text{ such that } \mathbf{x}\neq\mathbf{y}.$$

The same arguments as above imply that this is equivalent to

$$\mathbf{d}^T\mathbf{A}\mathbf{d} > 0 \text{ for any } 0 \neq \mathbf{d} \in \mathbb{R}^n,$$

which is the same as $\mathbf{A} \succ 0$. □

Examples of convex and nonconvex quadratic functions are illustrated in Figure 7.3.

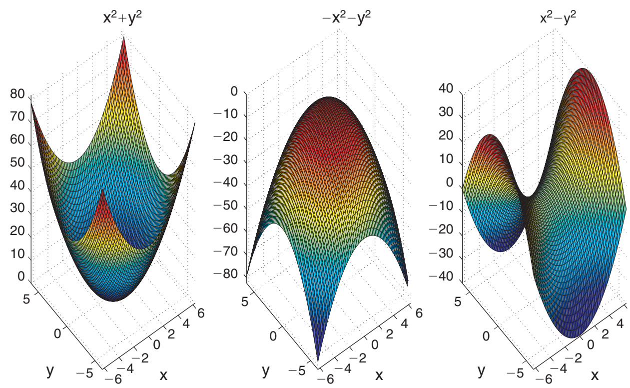

Figure 7.3. *The left quadratic function is convex ($f(x,y) = x^2 + y^2$), while the middle ($-x^2 - y^2$) and right ($x^2 - y^2$) functions are nonconvex.*

Another type of a first order characterization of convexity is the monotonicity property of the gradient. In the one-dimensional case, this means that the derivative is nondecreasing, but another definition of monotonicity is required in the n-dimensional case.

Theorem 7.11 (monotonicity of the gradient). *Suppose that f is a continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if*

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x}-\mathbf{y}) \geq 0 \text{ for any } \mathbf{x},\mathbf{y} \in C. \tag{7.8}$$

Proof. Assume first that f is convex over C. Then by the gradient inequality we have for any $\mathbf{x},\mathbf{y} \in C$

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^T(\mathbf{x}-\mathbf{y}),$$
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}).$$

By summing the two inequalities, the inequality (7.8) follows. To prove the opposite direction, suppose that (7.8) holds and let $\mathbf{x},\mathbf{y} \in C$. Let g be the one-dimensional function defined by

$$g(t) = f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})), \quad t \in [0,1].$$

By the fundamental theorem of calculus we have

$$\begin{aligned} f(\mathbf{y}) = g(1) = g(0) + \int_0^1 g'(t)dt \\ &= f(\mathbf{x}) + \int_0^1 (\mathbf{y}-\mathbf{x})^T \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))dt \\ &= f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \int_0^1 (\mathbf{y}-\mathbf{x})^T(\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}))dt \\ &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}), \end{aligned}$$

where the last inequality follows from the fact that for any $t > 0$ we have by the monotonicity of ∇f that

$$(\mathbf{y}-\mathbf{x})^T(\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x})) = \frac{1}{t}(\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}))^T(\mathbf{x}+t(\mathbf{y}-\mathbf{x})-\mathbf{x}) \geq 0.$$

□

7.3 ▪ Second Order Characterization of Convex Functions

For twice continuously differentiable functions, convexity can be characterized by the positive semidefiniteness of the Hessian matrix.

Theorem 7.12 (second order characterization of convexity). *Let f be a twice continuously differentiable function over an open convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if $\nabla^2 f(\mathbf{x}) \succeq 0$ for any $\mathbf{x} \in C$.*

Proof. Suppose that $\nabla^2 f(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in C$. We will prove the gradient inequality, which by Theorem 7.6 is enough in order to establish convexity. Let $\mathbf{x}, \mathbf{y} \in C$. Then by the linear approximation theorem (Theorem 1.24) we have that there exists $\mathbf{z} \in [\mathbf{x}, \mathbf{y}]$ (and hence $\mathbf{z} \in C$) for which

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{z})(\mathbf{y}-\mathbf{x}). \tag{7.9}$$

Since $\nabla^2 f(\mathbf{z}) \succeq 0$, it follows that $(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{z})(\mathbf{y}-\mathbf{x}) \geq 0$, and hence by (7.9), the inequality $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$ holds.

To prove the opposite direction, assume that f is convex over C. Let $\mathbf{x} \in C$ and let $\mathbf{y} \in \mathbb{R}^n$. Since C is open, it follows that $\mathbf{x} + \lambda \mathbf{y} \in C$ for $0 < \lambda < \varepsilon$, where ε is a small enough positive number. Invoking the gradient inequality we have

$$f(\mathbf{x}+\lambda\mathbf{y}) \geq f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^T \mathbf{y}. \tag{7.10}$$

In addition, by the quadratic approximation theorem (Theorem 1.25) we have that

$$f(\mathbf{x}+\lambda\mathbf{y}) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^T \mathbf{y} + \frac{\lambda^2}{2}\mathbf{y}^T \nabla^2 f(\mathbf{x})\mathbf{y} + o(\lambda^2 \|\mathbf{y}\|^2),$$

which combined with (7.10) yields the inequality

$$\frac{\lambda^2}{2}\mathbf{y}^T \nabla^2 f(\mathbf{x})\mathbf{y} + o(\lambda^2 \|\mathbf{y}\|^2) \geq 0$$

for any $\lambda \in (0, \varepsilon)$. Dividing the latter inequality by λ^2 we have

$$\frac{1}{2}\mathbf{y}^T \nabla^2 f(\mathbf{x})\mathbf{y} + \frac{o(\lambda^2\|\mathbf{y}\|^2)}{\lambda^2} \geq 0.$$

Finally, taking $\lambda \to 0^+$, we conclude that

$$\mathbf{y}^T \nabla^2 f(\mathbf{x})\mathbf{y} \geq 0$$

for any $\mathbf{y} \in \mathbb{R}^n$, implying that $\nabla^2 f(\mathbf{x}) \succeq 0$ for any $\mathbf{x} \in C$. $\square$

We also present the corresponding result for strictly convex functions stating that if the Hessian is positive definite, then the function is strictly convex. The proof of this result is similar to the one given in Theorem 7.12 and is hence left as an exercise (see Exercise 7.6).

Theorem 7.13 (sufficient second order condition for strict convexity). *Let f be a twice continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$, and suppose that $\nabla^2 f(\mathbf{x}) \succ 0$ for any $\mathbf{x} \in C$. Then f is strictly convex over C.*

Note that the positive definiteness of the Hessian is only a sufficient condition for strict convexity and is not necessary. Indeed, the function $f(x) = x^4$ is strictly convex, but its second order derivative $f''(x) = 12x^2$ is equal to zero for $x = 0$. The Hessian test immediately establishes the strict convexity of the one-dimensional functions x^2, e^x, e^{-x} and also of $-\ln(x), x\ln(x)$ over $\mathbb{R}_{++}$. A much more complicated example is that of the so-called *log-sum-exp* function, whose convexity can be shown by the Hessian test.

Example 7.14 (convexity of the log-sum-exp function). Consider the function

$$f(\mathbf{x}) = \ln(e^{x_1} + e^{x_2} + \cdots + e^{x_n}),$$

called the *log-sum-exp* function and defined over the entire space $\mathbb{R}^n$. We will prove its convexity using the Hessian test. The partial derivatives of f are given by

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, \quad i = 1, 2, \ldots, n,$$

and therefore

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\frac{e^{x_i} e^{x_j}}{\left(\sum_{k=1}^n e^{x_k}\right)^2}, & i \neq j, \\ -\frac{e^{x_i} e^{x_i}}{\left(\sum_{k=1}^n e^{x_k}\right)^2} + \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, & i = j. \end{cases}$$

We can thus write the Hessian matrix as

$$\nabla^2 f(\mathbf{x}) = \mathrm{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T,$$

where $w_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$. In particular, $\mathbf{w} \in \Delta_n$. To prove the positive semidefiniteness of $\nabla^2 f(\mathbf{x})$, take $0 \neq \mathbf{v} \in \mathbb{R}^n$ and consider the expression

$$\mathbf{v}^T \nabla^2 f(\mathbf{x})\mathbf{v} = \sum_{i=1}^n w_i v_i^2 - (\mathbf{v}^T\mathbf{w})^2.$$

The latter expression is nonnegative since employing the Cauchy–Schwarz inequality on the vectors $\mathbf{s},\mathbf{t}$ defined by

$$s_i = \sqrt{w_i}v_i, \qquad t_i = \sqrt{w_i}, \quad i = 1,2,\ldots,n,$$

yields

$$(\mathbf{v}^T\mathbf{w})^2 = (\mathbf{s}^T\mathbf{t})^2 \le \|\mathbf{s}\|^2\|\mathbf{t}\|^2 = \left(\sum_{i=1}^n w_i v_i^2\right)\left(\sum_{i=1}^n w_i\right) \stackrel{\mathbf{w}\in\Delta_n}{=} \sum_{i=1}^n w_i v_i^2,$$

establishing the inequality $\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v} \ge 0$. Since the latter inequality is valid for any $\mathbf{v} \in \mathbb{R}^n$, it follows that $\nabla^2 f(\mathbf{x})$ is indeed positive semidefinite. ∎

Example 7.15 (quadratic-over-linear). Let

$$f(x_1,x_2) = \frac{x_1^2}{x_2},$$

defined over $\mathbb{R}\times\mathbb{R}_{++} = \{(x_1,x_2) : x_2 > 0\}$. The Hessian of f is given by

$$\nabla^2 f(x_1,x_2) = 2\begin{pmatrix} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \\ -\frac{x_1}{x_2^2} & \frac{x_1^2}{x_2^3} \end{pmatrix}.$$

By Proposition 2.20, since the Hessian is a 2×2 matrix, to prove that it is positive semidefinite, it is enough to show that the trace and determinant are nonnegative, and indeed,

$$\begin{aligned}
\mathrm{Tr}\left[\nabla^2 f(x_1,x_2)\right] &= 2\left[\frac{1}{x_2} + \frac{x_1^2}{x_2^3}\right] > 0,\\
\det\left[\nabla^2 f(x_1,x_2)\right] &= 4\left[\frac{1}{x_2}\cdot\frac{x_1^2}{x_2^3} - \left(\frac{x_1}{x_2^2}\right)^2\right] = 0,
\end{aligned}$$

establishing the positive semidefiniteness of $\nabla^2 f(x_1,x_2)$ and hence the convexity of f. ∎

7.4 ▪ Operations Preserving Convexity

There are several important operations that preserve the convexity property. First, the sum of convex functions is a convex function and a multiplication of a convex function by a nonnegative number results with a convex function.

Theorem 7.16 (preservation of convexity under summation and multiplication by nonnegative scalars).

(a) *Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and let $\alpha \ge 0$. Then αf is a convex function over C.*

(b) *Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $C \subseteq \mathbb{R}^n$. Then the sum function $f_1 + f_2 + \cdots + f_p$ is convex over C.*

Proof. (a) Denote $g(\mathbf{x}) \equiv \alpha f(\mathbf{x})$. We will prove the convexity of g by definition. Let $\mathbf{x},\mathbf{y} \in C$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} g(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) &= \alpha f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) && \text{(definition of } g) \\ &\le \alpha\lambda f(\mathbf{x})+\alpha(1-\lambda)f(\mathbf{y}) && \text{(convexity of } f) \\ &= \lambda g(\mathbf{x})+(1-\lambda)g(\mathbf{y}) && \text{(definition of } g). \end{aligned}$$

(b) Let $\mathbf{x},\mathbf{y} \in C$ and $\lambda \in [0,1]$. For each $i = 1,2,\ldots,p$, since f_i is convex, we have

$$f_i(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda f_i(\mathbf{x})+(1-\lambda)f_i(\mathbf{y}).$$

Summing the latter inequality over $i = 1,2,\ldots,k$ yields the inequality

$$g(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda g(\mathbf{x})+(1-\lambda)g(\mathbf{y})$$

for all $\mathbf{x},\mathbf{y} \in C$ and $\lambda \in [0,1]$, where $g = f_1+f_2+\cdots+f_p$. We have thus established that the sum function is convex. ☐

Another important operation preserving convexity is affine change of variables.

Theorem 7.17 (preservation of convexity under affine change of variables). *Let $f : C \to \mathbb{R}$ be a convex function defined on a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the function g defined by*

$$g(\mathbf{y}) = f(\mathbf{A}\mathbf{y}+\mathbf{b})$$

is convex over the convex set $D = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y}+\mathbf{b} \in C\}$.

Proof. First of all, note that D is indeed a convex set since it can be represented as an inverse linear mapping of a translation of C (see Theorem 6.8):

$$D = \mathbf{A}^{-1}(C-\mathbf{b}).$$

Let $\mathbf{y}_1,\mathbf{y}_2 \in D$. Define

$$\mathbf{x}_1 = \mathbf{A}\mathbf{y}_1+\mathbf{b}, \tag{7.11}$$
$$\mathbf{x}_2 = \mathbf{A}\mathbf{y}_2+\mathbf{b}, \tag{7.12}$$

which by the definition of D satisfy $\mathbf{x}_1,\mathbf{x}_2 \in C$. Let $\lambda \in [0,1]$. By the convexity of f we have

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

Plugging the expressions (7.11) and (7.12) of $\mathbf{x}_1$ and $\mathbf{x}_2$ into the latter inequality, we obtain that

$$f(\mathbf{A}(\lambda\mathbf{y}_1+(1-\lambda)\mathbf{y}_2)+\mathbf{b}) \le \lambda f(\mathbf{A}\mathbf{y}_1+\mathbf{b})+(1-\lambda)f(\mathbf{A}\mathbf{y}_2+\mathbf{b}),$$

which is the same as

$$g(\lambda\mathbf{y}_1+(1-\lambda)\mathbf{y}_2) \le \lambda g(\mathbf{y}_1)+(1-\lambda)g(\mathbf{y}_2),$$

thus establishing the convexity of g. ☐

Example 7.18 (generalized quadratic-over-linear). Let $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$. We assume that $\mathbf{c} \neq \mathbf{0}$. We will show that the quadratic-over-linear function

$$g(\mathbf{x}) = \frac{\|\mathbf{A}\mathbf{x}+\mathbf{b}\|^2}{\mathbf{c}^T\mathbf{x}+d}$$

is convex over $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x}+d > 0\}$. We begin by proving the convexity of the function

$$h(\mathbf{y},t) = \frac{\|\mathbf{y}\|^2}{t}$$

over the convex set $C \equiv \{\binom{\mathbf{y}}{t} \in \mathbb{R}^{m+1} : \mathbf{y} \in \mathbb{R}^m, t > 0\}$. For that, note that $h = \sum_{i=1}^m h_i$ where

$$h_i(\mathbf{y},t) = \frac{y_i^2}{t}.$$

By the convexity of the quadratic-over-linear function $\varphi(x,z) = \frac{x^2}{z}$ over $\{(x,z) : x \in \mathbb{R}, z > 0\}$ (see Example 7.15), it follows that h_i is convex for any i (specifically, h_i is generated from φ by the linear transformation $x = y_i, z = t$). Hence, h is convex over C. The function f can be represented as

$$f(\mathbf{x}) = h(\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{c}^T\mathbf{x}+d).$$

Consequently, since f is the function h which has gone through an affine change of variables, it is convex over the domain $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x}+d > 0\}$. ∎

Example 7.19. Consider the function

$$f(x_1,x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + e^{x_1}.$$

To prove the convexity of f, note that $f = f_1 + f_2$, where

$$f_1(x_1,x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2,$$
$$f_2(x_1,x_2) = e^{x_1}.$$

The function f_1 is convex since it is a quadratic function with an associated matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$ which is positive semidefinite since $\mathrm{Tr}(\mathbf{A}) = 4 > 0, \det(\mathbf{A}) = 2 > 0$. The function f_2 is convex since it is generated from the one-dimensional convex function $\varphi(t) = e^t$ by the linear transformation $t = x_1$. ∎

Example 7.20. The function $f(x_1,x_2,x_3) = e^{x_1-x_2+x_3} + e^{2x_2} + x_1$ is convex over $\mathbb{R}^3$ as a sum of three convex functions: the function $e^{x_1-x_2+x_3}$, which is convex since it is constructed by making the linear change of variables $t = x_1 - x_2 + x_3$ in the one-dimensional function $\varphi(t) = e^t$. For the same reason, e^{2x_2} is convex. Finally, the function x_1, being linear, is convex. ∎

Example 7.21. The function $f(x_1,x_2) = -\ln(x_1x_2)$ is convex over $\mathbb{R}^2_{++}$ since it can be written as

$$f(x_1,x_2) = -\ln(x_1) - \ln(x_2),$$

and the convexity of $-\ln(x_1)$ and $-\ln(x_2)$ follows from the convexity of $\varphi(t) = -\ln(t)$ over $\mathbb{R}_{++}$. ∎

In general, convexity is not preserved under composition of convex functions. For example, let $g(t) = t^2$ and $h(t) = t^2 - 4$. Then g and h are convex. However, their composition

$$s(t) = g(h(t)) = (t^2 - 4)^2$$

is not convex, as illustrated in Figure 7.4. (This can also be seen by the fact that $s''(t) = 12t^2 - 16$ and hence $s''(t) < 0$ for all $|t| < \sqrt{\frac{4}{3}}$.) The next result shows that convexity is preserved in the case of a composition of a *nondecreasing* convex function with a convex function.

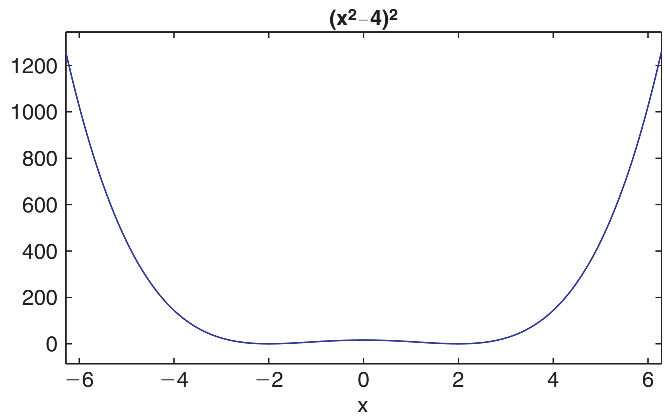

Figure 7.4. *The nonconvex function* $(t^2 - 4)^2$.

Theorem 7.22 (preservation of convexity under composition with a nondecreasing convex function). *Let $f : C \to \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Let $g : I \to \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $I \subseteq \mathbb{R}$. Assume that the image of C under f is contained in I: $f(C) \subseteq I$. Then the composition of g with f defined by*

$$h(\mathbf{x}) \equiv g(f(\mathbf{x})), \quad \mathbf{x} \in C,$$

is a convex function over C.

Proof. Let $\mathbf{x}, \mathbf{y} \in C$ and let $\lambda \in [0, 1]$. Then

$$\begin{aligned} h(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= g(f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})) && \text{(definition of } h) \\ &\leq g(\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})) && \text{(convexity of } f \text{ and monotonicity of } g) \\ &\leq \lambda g(f(\mathbf{x})) + (1-\lambda)g(f(\mathbf{y})) && \text{(convexity of } g) \\ &= \lambda h(\mathbf{x}) + (1-\lambda)h(\mathbf{y}) && \text{(definition of } h), \end{aligned}$$

thus establishing the convexity of h. □

Example 7.23. The function $h(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$ is convex since it can be represented as $h(\mathbf{x}) = g(f(\mathbf{x}))$, where $g(t) = e^t$ is a nondecreasing convex function and $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is a convex function. ∎

Example 7.24. The function $h(\mathbf{x}) = (\|\mathbf{x}\|^2 + 1)^2$ is a convex function over $\mathbb{R}^n$ since it can be represented as $h(\mathbf{x}) = g(f(\mathbf{x}))$, where $g(t) = t^2$ and $f(\mathbf{x}) = \|\mathbf{x}\|^2 + 1$. Both f and g are convex, but note that g is not a nondecreasing function. However, the image of $\mathbb{R}^n$ under f is the interval $[1, \infty)$ on which the function g is nondecreasing. Consequently, the composition $h(\mathbf{x}) = g(f(\mathbf{x}))$ is convex. ∎

Another important operation that preserves convexity is the pointwise maximum of convex functions.

Theorem 7.25 (pointwise maximum of convex functions). *Let $f_1, \ldots, f_p : C \to \mathbb{R}$ be p convex functions over the convex set $C \subseteq \mathbb{R}^n$. Then the maximum function*

$$f(\mathbf{x}) \equiv \max_{i=1,2,\ldots,p} f_i(\mathbf{x})$$

is a convex function over C.

Proof. Let $\mathbf{x}, \mathbf{y} \in C$ and let $\lambda \in [0, 1]$. Then

$$\begin{aligned}
f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \max_{i=1,2,\ldots,p} f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) && \text{(definition of } f)\\
&\le \max_{i=1,2,\ldots,p} \{\lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})\} && \text{(convexity of } f_i)\\
&\le \lambda \max_{i=1,2,\ldots,p} f_i(\mathbf{x}) + (1-\lambda) \max_{i=1,2,\ldots,p} f_i(\mathbf{y}) && (*)\\
&= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) && \text{(definition of } f).
\end{aligned}$$

The inequality (*) follows from the fact that for any two sequences $\{a_i\}_{i=1}^p, \{b_i\}_{i=1}^p$ one has

$$\max_{i=1,2,\ldots,p} (a_i + b_i) \le \max_{i=1,2,\ldots,p} a_i + \max_{i=1,2,\ldots,p} b_i. \quad \square$$

Example 7.26 (convexity of the maximum function). Let

$$f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}.$$

Then since f is the maximum of n linear functions, which are in particular convex, it follows by Theorem 7.25 that it is convex. ∎

Example 7.27 (convexity of the sum of the k largest values). Given a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$. Let $x_{[i]}$ denote the ith largest value in $\mathbf{x}$. In particular, $x_{[1]} = \max\{x_1, x_2, \ldots, x_n\}$ and $x_{[n]} = \min\{x_1, x_2, \ldots, x_n\}$. As stated in the previous example, the function $h(\mathbf{x}) = x_{[1]}$ is convex. However, in general the function $h(\mathbf{x}) = x_{[i]}$ is not convex. On the other hand, the function

$$h_k(\mathbf{x}) = x_{[1]} + x_{[2]} + \cdots + x_{[k]},$$

that is, the function producing the sum of the k largest components, is in fact convex. To see this, note that h_k can be rewritten as

$$h_k(\mathbf{x}) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_k} : i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, n\} \text{ are different}\},$$

so that h_k, as a maximum of linear (and hence convex) functions, is a convex function. ∎

Another operation preserving convexity is partial minimization.

Theorem 7.28. *Let $f : C \times D \to \mathbb{R}$ be a convex function defined over the set $C \times D$ where $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$ are convex sets. Let*

$$g(\mathbf{x}) = \min_{\mathbf{y} \in D} f(\mathbf{x},\mathbf{y}), \quad \mathbf{x} \in C,$$

where we assume that the minimal value in the above definition is real. Then g is convex over C.

Proof. Let $\mathbf{x}_1,\mathbf{x}_2 \in C$ and $\lambda \in [0,1]$. Take $\varepsilon > 0$. Then there exist $\mathbf{y}_1,\mathbf{y}_2 \in D$ such that

$$f(\mathbf{x}_1,\mathbf{y}_1) \le g(\mathbf{x}_1) + \varepsilon, \tag{7.13}$$

$$f(\mathbf{x}_2,\mathbf{y}_2) \le g(\mathbf{x}_2) + \varepsilon. \tag{7.14}$$

By the convexity of f we have

$$\begin{aligned} f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) &\le \lambda f(\mathbf{x}_1,\mathbf{y}_1) + (1-\lambda) f(\mathbf{x}_2,\mathbf{y}_2) \\ &\overset{(7.13),(7.14)}{\le} \lambda(g(\mathbf{x}_1)+\varepsilon) + (1-\lambda)(g(\mathbf{x}_2)+\varepsilon) \\ &= \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2) + \varepsilon. \end{aligned}$$

By the definition of g we can conclude that

$$g(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \le \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2) + \varepsilon.$$

Since the above inequality holds for any $\varepsilon > 0$, it follows that $g(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \le \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2)$, and the convexity of g is established. □

Note that in the above theorem, we only assumed that the minimal value is real, but we did not assume that it is attained.

Example 7.29 (convexity of the distance function). Let $C \subseteq \mathbb{R}^n$ be a convex set. The *distance* function defined by

$$d(\mathbf{x},C) = \min\{\|\mathbf{x}-\mathbf{y}\| : \mathbf{y} \in C\}$$

is convex since the function $f(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$ is convex over $\mathbb{R}^n \times C$, and thus by Theorem 7.28 it follows that $d(\cdot,C)$ is convex. ∎

7.5 ▪ Level Sets of Convex Functions

We begin with the definition of a *level set*.

Definition 7.30 (level sets). *Let $f : S \to \mathbb{R}$ be a function defined over a set $S \subseteq \mathbb{R}^n$. Then the* **level set** *of f with level α is given by*

$$\mathrm{Lev}(f,\alpha) = \{\mathbf{x} \in S : f(\mathbf{x}) \le \alpha\}.$$

A fundamental property of convex functions is that their level sets are necessarily convex.

Theorem 7.31 (convexity of level sets of convex functions). *Let $f : C \to \mathbb{R}$ be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the level set* $\mathrm{Lev}(f,\alpha)$ *is convex.*

Proof. Let $\mathbf{x},\mathbf{y} \in \text{Lev}(f,\alpha)$ and $\lambda \in [0,1]$. Then $f(\mathbf{x}), f(\mathbf{y}) \le \alpha$. By the convexity of C we have that $\lambda\mathbf{x}+(1-\lambda)\mathbf{y} \in C$, which combined with the convexity of f yields

$$f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y}) \le \lambda\alpha+(1-\lambda)\alpha = \alpha,$$

establishing the fact that $\lambda\mathbf{x}+(1-\lambda)\mathbf{y} \in \text{Lev}(f,\alpha)$ and subsequently the convexity of $\text{Lev}(f,\alpha)$. □

Example 7.32. Consider the following subset of $\mathbb{R}^n$:

$$D = \left\{ \mathbf{x} : (\mathbf{x}^T\mathbf{Q}\mathbf{x}+1)^2 + \ln\left(\sum_{i=1}^n e^{x_i}\right) \le 3 \right\},$$

where $\mathbf{Q} \succeq 0$ is an $n \times n$ matrix. The set D is convex as a level set of a convex function. Specifically, $D = \text{Lev}(f,3)$, where

$$f(\mathbf{x}) = (\mathbf{x}^T\mathbf{Q}\mathbf{x}+1)^2 + \ln\left(\sum_{i=1}^n e^{x_i}\right).$$

The function f is indeed convex as the sum of two convex functions: the log-sum-exp function, which was shown to be convex in Example 7.14, and the function $g(\mathbf{x}) = (\mathbf{x}^T\mathbf{Q}\mathbf{x}+1)^2$, which is convex as a composition of the nondecreasing convex function $\varphi(t) = (t+1)^2$ defined on $\mathbb{R}_+$ with the convex quadratic function $\mathbf{x}^T\mathbf{Q}\mathbf{x}$. ∎

All convex functions have convex level sets, but the reverse claim is not true. That is, there do exist nonconvex functions whose level sets are all convex. Functions satisfying the property that all their level sets are convex are called *quasi-convex* functions.

Definition 7.33 (quasi-convex functions). *A function $f : C \to \mathbb{R}$ defined over the convex set $C \subseteq \mathbb{R}^n$ is called* **quasi-convex** *if for any $\alpha \in \mathbb{R}$ the set $\text{Lev}(f,\alpha)$ is convex.*

The following example demonstrates the fact that quasi-convex functions may be nonconvex.

Example 7.34. The one-dimensional function $f(x) = \sqrt{|x|}$ is obviously not convex (see Figure 7.5), but its level sets are convex: for any $\alpha < 0$ we have that $\text{Lev}(f,\alpha) = \emptyset$, and for any $\alpha \ge 0$ the corresponding level set is convex:

$$\text{Lev}(f,\alpha) = \{x : \sqrt{|x|} \le \alpha\} = \{x : |x| \le \alpha^2\} = [-\alpha^2, \alpha^2].$$

We deduce that the nonconvex function f is quasi-convex. ∎

Example 7.35 (linear-over-linear). Consider the function

$$f(\mathbf{x}) = \frac{\mathbf{a}^T\mathbf{x}+b}{\mathbf{c}^T\mathbf{x}+d},$$

where $\mathbf{a},\mathbf{c} \in \mathbb{R}^n$ and $b,d \in \mathbb{R}$. To avoid trivial cases, we assume that $\mathbf{c} \neq 0$ and that the function is defined over the open half-space

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x}+d > 0\}.$$

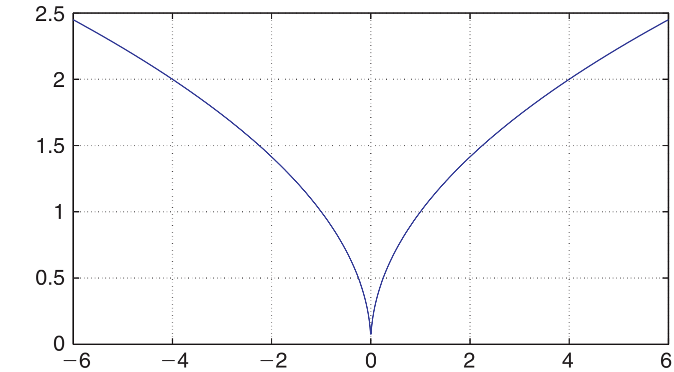

Figure 7.5. *The quasi-convex function* $\sqrt{|x|}$.

In general, f is not a convex function, but it is not difficult to show that it is quasi-convex. Indeed, let $\alpha \in \mathbb{R}$. Then the corresponding level set is given by

$$\text{Lev}(f,\alpha) = \{\mathbf{x} \in C : f(\mathbf{x}) \le \alpha\} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x} + d > 0, (\mathbf{a} - \alpha\mathbf{c})^T\mathbf{x} + (b - \alpha d) \le 0\},$$

which is convex due to the fact that it is an intersection of two half-spaces (which are in particular convex sets) when $\mathbf{a} \neq \alpha\mathbf{c}$, and when $\mathbf{a} = \alpha\mathbf{c}$ it is either a half-space (if $b - \alpha d \le 0$) or the empty set (if $b - \alpha d > 0$). ∎

7.6 ▪ Continuity and Differentiability of Convex Functions

Convex functions are not necessarily continuous when defined on nonopen sets. Let us consider, for example, the function

$$f(x) = \begin{cases} 1, & x = 0, \\ x^2, & 0 < x \le 1, \end{cases}$$

defined over the interval $[0,1]$. It is easy to see that this is a convex function, and obviously it is not a continuous function (as also illustrated in Figure 7.6). The main result is that convex functions are always continuous at interior points of their domain. Thus, for

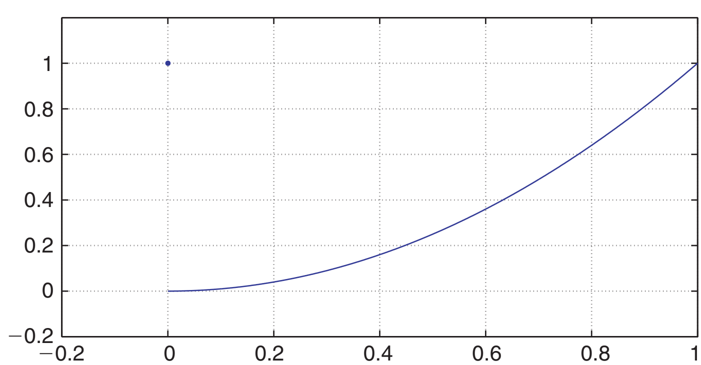

Figure 7.6. *A noncontinuous convex function over the interval* $[0,1]$.

example, functions which are convex over the entire space $\mathbb{R}^n$ are always continuous. We will prove an even stronger result: convex functions are always local Lipschitz continuous at interior points of their domain.

Theorem 7.36 (local Lipschitz continuity of convex functions). *Let $f : C \to \mathbb{R}$ be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \text{int}(C)$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $B[\mathbf{x}_0, \varepsilon] \subseteq C$ and*

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \le L\|\mathbf{x} - \mathbf{x}_0\| \tag{7.15}$$

for all $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$.

Proof. Since $\mathbf{x}_0 \in \text{int}(C)$, it follows that there exists $\varepsilon > 0$ such that

$$B_\infty[\mathbf{x}_0, \varepsilon] \equiv \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_\infty \le \varepsilon\} \subseteq C.$$

Next we show that f is upper bounded over $B_\infty[\mathbf{x}_0, \varepsilon]$. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2^n}$ be the 2^n extreme points of $B_\infty[\mathbf{x}_0, \varepsilon]$; these are the vectors $\mathbf{v}_i = \mathbf{x}_0 + \varepsilon \mathbf{w}_i$, where $\mathbf{w}_1, \ldots, \mathbf{w}_{2^n}$ are the vectors in $\{-1, 1\}^n$. Then obviously, for any $\mathbf{x} \in B_\infty[\mathbf{x}_0, \varepsilon]$ there exists $\lambda \in \Delta_{2^n}$ such that $\mathbf{x} = \sum_{i=1}^{2^n} \lambda_i \mathbf{v}_i$, and hence, by Jensen's inequality,

$$f(\mathbf{x}) = f\left(\sum_{i=1}^{2^n} \lambda_i \mathbf{v}_i\right) \le \sum_{i=1}^{2^n} \lambda_i f(\mathbf{v}_i) \le M,$$

where $M = \max_{i=1,2,\ldots,2^n} f(\mathbf{v}_i)$. Since $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2$ for any $\mathbb{R}^n$ it holds that

$$B_2[\mathbf{x}_0, \varepsilon] = B[\mathbf{x}_0, \varepsilon] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_2 \le \varepsilon\} \subseteq B_\infty[\mathbf{x}_0, \varepsilon].$$

We therefore conclude that $f(\mathbf{x}) \le M$ for any $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$. Let $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$ be such that $\mathbf{x} \ne \mathbf{x}_0$. (The result (7.15) is obvious when $\mathbf{x} = \mathbf{x}_0$.) Define

$$\mathbf{z} = \mathbf{x}_0 + \frac{1}{\alpha}(\mathbf{x} - \mathbf{x}_0),$$

where $\alpha = \frac{1}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\|$. Then obviously $\alpha \le 1$ and $\mathbf{z} \in B[\mathbf{x}_0, \varepsilon]$, and in particular $f(\mathbf{z}) \le M$. In addition,

$$\mathbf{x} = \alpha \mathbf{z} + (1 - \alpha)\mathbf{x}_0.$$

Consequently, by Jensen's inequality we have

$$\begin{aligned} f(\mathbf{x}) &\le \alpha f(\mathbf{z}) + (1 - \alpha) f(\mathbf{x}_0) \\ &\le f(\mathbf{x}_0) + \alpha(M - f(\mathbf{x}_0)) \\ &= f(\mathbf{x}_0) + \frac{M - f(\mathbf{x}_0)}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\|. \end{aligned}$$

We can therefore deduce that $f(\mathbf{x}) - f(\mathbf{x}_0) \le L\|\mathbf{x} - \mathbf{x}_0\|$, where $L = \frac{M - f(\mathbf{x}_0)}{\varepsilon}$. To prove the result, we need to show that $f(\mathbf{x}) - f(\mathbf{x}_0) \ge -L\|\mathbf{x} - \mathbf{x}_0\|$. For that, define $\mathbf{u} = \mathbf{x}_0 + \frac{1}{\alpha}(\mathbf{x}_0 - \mathbf{x})$. Clearly we have $\|\mathbf{u} - \mathbf{x}_0\| = \varepsilon$ and hence $\mathbf{u} \in B[\mathbf{x}_0, \varepsilon]$ and in particular $f(\mathbf{u}) \le M$. In addition, $\mathbf{x} = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})$. Therefore,

$$f(\mathbf{x}) = f(\mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})) \ge f(\mathbf{x}_0) + \alpha(f(\mathbf{x}_0) - f(\mathbf{u})). \tag{7.16}$$

The latter inequality is valid since

$$\mathbf{x}_0 = \frac{1}{1+\alpha}(\mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})) + \frac{\alpha}{1+\alpha}\mathbf{u},$$

and hence, by Jensen's inequality

$$f(\mathbf{x}_0) \le \frac{1}{1+\alpha} f(\mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})) + \frac{\alpha}{1+\alpha} f(\mathbf{u}),$$

which is the same as the inequality in (7.16) (after some rearrangement of terms). Now, continuing (7.16),

$$\begin{aligned} f(\mathbf{x}) &\ge f(\mathbf{x}_0) + \alpha(f(\mathbf{x}_0) - f(\mathbf{u})) \\ &\ge f(\mathbf{x}_0) - \alpha(M - f(\mathbf{x}_0)) \\ &= f(\mathbf{x}_0) - \frac{M - f(\mathbf{x}_0)}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\| \\ &= f(\mathbf{x}_0) - L\|\mathbf{x} - \mathbf{x}_0\|, \end{aligned}$$

and the desired result is established. $\square$

Convex functions are not necessarily differentiable, but on the other hand, as will be shown in the next result, all the directional derivatives at interior points exist.

Theorem 7.37 (existence of directional derivatives for convex functions). *Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \text{int}(C)$. Then for any $\mathbf{d} \neq \mathbf{0}$, the directional derivative $f'(\mathbf{x};\mathbf{d})$ exists.*

Proof. Let $\mathbf{x} \in \text{int}(C)$ and $\mathbf{d} \neq \mathbf{0}$. Then the directional derivative (if exists) is the limit

$$\lim_{t \to 0^+} \frac{g(t) - g(0)}{t}, \tag{7.17}$$

where $g(t) = f(\mathbf{x} + t\mathbf{d})$. Defining $h(t) \equiv \frac{g(t)-g(0)}{t}$, the limit (7.17) can be equivalently written as

$$\lim_{t \to 0^+} h(t).$$

Note that g, as well as h, is defined for small enough values of t by the fact that $\mathbf{x} \in \text{int}(C)$. In fact, we will take an $\varepsilon > 0$ for which $\mathbf{x} + t\mathbf{d}, \mathbf{x} - t\mathbf{d} \in C$ for all $t \in [0, \varepsilon]$. Now, let $0 < t_1 < t_2 \le \varepsilon$. Then

$$\mathbf{x} + t_1\mathbf{d} = \left(1 - \frac{t_1}{t_2}\right)\mathbf{x} + \frac{t_1}{t_2}(\mathbf{x} + t_2\mathbf{d}),$$

and thus, by the convexity of f we have

$$f(\mathbf{x} + t_1\mathbf{d}) \le \left(1 - \frac{t_1}{t_2}\right) f(\mathbf{x}) + \frac{t_1}{t_2} f(\mathbf{x} + t_2\mathbf{d}).$$

The latter inequality can be rewritten (after some rearrangement of terms) as

$$\frac{f(\mathbf{x} + t_1\mathbf{d}) - f(\mathbf{x})}{t_1} \le \frac{f(\mathbf{x} + t_2\mathbf{d}) - f(\mathbf{x})}{t_2},$$

which is the same as $h(t_1) \leq h(t_2)$. We thus conclude that the function h is monotone nondecreasing over $\mathbb{R}_{++}$. All that is left is to prove that h is bounded below over $(0, \varepsilon]$. Indeed, taking $0 < t \leq \varepsilon$, note that

$$\mathbf{x} = \frac{\varepsilon}{\varepsilon + t}(\mathbf{x} + t\mathbf{d}) + \frac{t}{\varepsilon + t}(\mathbf{x} - \varepsilon\mathbf{d}).$$

Hence, by the convexity of f we have

$$f(\mathbf{x}) \leq \frac{\varepsilon}{\varepsilon + t} f(\mathbf{x} + t\mathbf{d}) + \frac{t}{\varepsilon + t} f(\mathbf{x} - \varepsilon\mathbf{d}),$$

which after some rearrangement of terms can be seen to be equivalent to the inequality

$$h(t) = \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \geq \frac{f(\mathbf{x}) - f(\mathbf{x} - \varepsilon\mathbf{d})}{\varepsilon},$$

showing that h is bounded below over $(0, \varepsilon]$. Since h is nondecreasing and bounded below over $(0, \varepsilon]$ it follows that the limit $\lim_{t \to 0^+} h(t)$ exists, meaning that the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists. $\square$

7.7 ▪ Extended Real-Valued Functions

Until now we have discussed functions that are *real-valued*, meaning that they take their values in $\mathbb{R} = (-\infty, \infty)$. It is also quite natural to consider functions that are defined over the entire space $\mathbb{R}^n$ that take values in $\mathbb{R} \cup \{\infty\} = (-\infty, \infty]$. Such a function is called an *extended real-valued function*. One very important example of an extended real-valued function is the *indicator function*, which is defined as follows: given a set $S \subseteq \mathbb{R}^n$, the indicator function $\delta_S : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is given by

$$\delta_S(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in S, \\ \infty & \text{if } \mathbf{x} \notin S. \end{cases}$$

The *effective domain* of an extended real-valued function is the set of vectors for which the function takes a real value:

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

An extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called *proper* if it is not always equal to ∞, meaning that there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $f(\mathbf{x}_0) < \infty$. Similarly to the definition for real-valued functions, an extended real-valued function is convex if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ the following inequality holds:

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

where we use the usual arithmetic with ∞:

$$a + \infty = \infty \text{ for any } a \in \mathbb{R},$$
$$a \cdot \infty = \infty \text{ for any } a \in \mathbb{R}_{++}.$$

In addition, we have the much less obvious rule that $0 \cdot \infty = 0$. The above definition of convexity of extended real-valued functions is equivalent to saying that $\mathrm{dom}(f)$ is a convex set and that the restriction of f to its effective domain $\mathrm{dom}(f)$; that is, the function $g : \mathrm{dom}(f) \to \mathbb{R}$ defined by $g(\mathbf{x}) = f(\mathbf{x})$ for any $\mathbf{x} \in \mathrm{dom}(f)$ is a convex real-valued function over $\mathrm{dom}(f)$. As an example, the indicator function $\delta_C(\cdot)$ of a set $C \subseteq \mathbb{R}^n$ is convex if and only if C is a convex set.

An important set associated with extended real-valued functions is its *epigraph*. Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Then the epigraph set $\mathrm{epi}(f) \subseteq \mathbb{R}^{n+1}$ is defined by

$$\mathrm{epi}(f) = \left\{ \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} : f(\mathbf{x}) \le t \right\}.$$

An example of an epigraph can be seen in Figure 7.7. It is not difficult to show that an extended real-valued (or a real-valued) function f is convex if and only if its epigraph set $\mathrm{epi}(f)$ is convex (see Exercise 7.29). An important property of convex extended real-valued functions that convexity is preserved under the maximum operation. As was already mentioned, we do not use the "sup" notation in this book and we always refer to the maximum of a function or the maximum over a given index set.

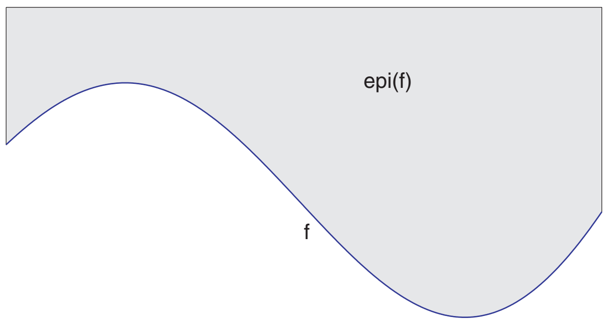

Figure 7.7. *The epigraph of a one-dimensional function.*

Theorem 7.38 (preservation of convexity under maximum). *Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be an extended real-valued convex function for any $i \in I$ (I being an arbitrary index set). Then the function $f(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x})$ is an extended real-valued convex function.*

Proof. The result follows from the fact that $\mathrm{epi}(f) = \bigcap_{i \in I} \mathrm{epi}(f_i)$. The convexity of f_i for any $i \in I$ implies the convexity of $\mathrm{epi}(f_i)$ for any $i \in I$. Consequently, $\mathrm{epi}(f)$, as an intersection of convex sets, is convex, and hence the convexity of f is established. $\square$

The differences between Theorems 7.38 and 7.25 are that the functions in Theorem 7.38 are not necessarily real-valued and that the index set I can be infinite.

Example 7.39 (support functions). Let $S \subseteq \mathbb{R}^n$. The *support function of* S is the function

$$\sigma_S(\mathbf{x}) = \max_{\mathbf{y} \in S} \mathbf{x}^T \mathbf{y}.$$

Since for each $\mathbf{y} \in S$, the function $f_{\mathbf{y}}(\mathbf{x}) \equiv \mathbf{y}^T \mathbf{x}$ is a convex function over $\mathbb{R}^n$ (being linear), it follows by Theorem 7.38 that σ_S is an extended real-valued convex function. $\blacksquare$

As an example of a support function, let us consider the unit (Euclidean) ball $S = B[0,1] = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}\| \leq 1\}$. Let $\mathbf{x} \in \mathbb{R}^n$. We will show that

$$\sigma_S(\mathbf{x}) = \|\mathbf{x}\|. \tag{7.18}$$

Obviously, if $\mathbf{x} = 0$, then $\sigma_S(\mathbf{x}) = 0$ and hence (7.18) holds for $\mathbf{x} = 0$. If $\mathbf{x} \neq 0$, then for any $\mathbf{y} \in S$ and $\mathbf{x} \in \mathbb{R}^n$, we have by the Cauchy–Schwarz inequality that

$$\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\| \leq \|\mathbf{x}\|.$$

On the other hand, taking $\tilde{\mathbf{y}} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} \in S$, we have

$$\mathbf{x}^T\tilde{\mathbf{y}} = \|\mathbf{x}\|,$$

and the desired formula (7.18) follows.

Example 7.40. Consider the function

$$f(\mathbf{t}) = \mathbf{d}^T A(\mathbf{t})^{-1}\mathbf{d},$$

where $\mathbf{d} \in \mathbb{R}^n$ and $A(\mathbf{t}) = \sum_{i=1}^m t_i \mathbf{A}_i$ with $\mathbf{A}_1, \ldots, \mathbf{A}_m$ being $n \times n$ positive definite matrices. We will show that this function is convex over $\mathbb{R}^m_{++}$. Indeed, for any $\mathbf{x} \in \mathbb{R}^n$, the function

$$g_{\mathbf{x}}(\mathbf{t}) = \begin{cases} 2\mathbf{d}^T\mathbf{x} - \mathbf{x}^T A(\mathbf{t})\mathbf{x}, & \mathbf{t} \in \mathbb{R}^m_{++}, \\ \infty & \text{else} \end{cases}$$

is convex over $\mathbb{R}^m$ since it is an affine function over its convex domain. The corresponding max function is

$$\max_{\mathbf{x} \in \mathbb{R}^n} \left\{2\mathbf{d}^T\mathbf{x} - \mathbf{x}^T A(\mathbf{t})\mathbf{x}\right\} = \mathbf{d}^T A(\mathbf{t})^{-1}\mathbf{d}$$

for $\mathbf{t} \in \mathbb{R}^m_{++}$ and ∞ elsewhere. Therefore, the extended real-valued function

$$\tilde{f}(\mathbf{t}) = \begin{cases} \mathbf{d}^T A(\mathbf{t})^{-1}\mathbf{d}, & \mathbf{t} \in \mathbb{R}^m_{++}, \\ \infty & \text{else} \end{cases}$$

is convex over $\mathbb{R}^m$, which is the same as saying that f is convex over $\mathbb{R}^m_{++}$. ∎

7.8 ▪ Maxima of Convex Functions

In the next chapter we will learn that problems consisting of minimizing a convex function over a convex set are in some sense "easy," but in this section we explore some important properties of the much more difficult problem of maximizing a convex function over a convex feasible set. First, we show that the maximum of a nonconstant convex function defined on a convex set C cannot be attained at an interior point of the set C.

Theorem 7.41. *Let $f : C \to \mathbb{R}$ be a convex function which is not constant over the convex set C. Then f does not attain a maximum at a point in* $\text{int}(C)$.

Proof. Assume in contradiction that $\mathbf{x}^* \in \text{int}(C)$ is a global maximizer of f over C. Since the function is not constant, there exists $\mathbf{y} \in C$ such that $f(\mathbf{y}) < f(\mathbf{x}^*)$. Since $\mathbf{x}^* \in \text{int}(C)$, there exists $\varepsilon > 0$ such that $\mathbf{z} = \mathbf{x}^* + \varepsilon(\mathbf{x}^* - \mathbf{y}) \in C$. Since $\mathbf{x}^* = \frac{\varepsilon}{\varepsilon+1}\mathbf{y} + \frac{1}{\varepsilon+1}\mathbf{z}$, it follows by the convexity of f that

$$f(\mathbf{x}^*) \le \frac{\varepsilon}{\varepsilon+1} f(\mathbf{y}) + \frac{1}{\varepsilon+1} f(\mathbf{z}),$$

and hence $f(\mathbf{z}) \ge \varepsilon(f(\mathbf{x}^*) - f(\mathbf{y})) + f(\mathbf{x}^*) > f(\mathbf{x}^*)$, which is a contradiction to the optimality of $\mathbf{x}^*$. □

When the underlying set is also compact, then the next result shows that there exists at least one maximizer that is an extreme point of the set.

Theorem 7.42. *Let $f : C \to \mathbb{R}$ be a convex and continuous function over the convex and compact set $C \subseteq \mathbb{R}^n$. Then there exists at least one maximizer of f over C that is an extreme point of C.*

Proof. Let $\mathbf{x}^*$ be a maximizer of f over C (whose existence is guaranteed by the Weierstrass theorem, Theorem 2.30). If $\mathbf{x}^*$ is an extreme point of C, then the result is established. Otherwise, if $\mathbf{x}^*$ is not an extreme point, then by the Krein–Milman theorem (Theorem 6.35), $C = \operatorname{conv}(\operatorname{ext}(C))$, which means that there exist $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \operatorname{ext}(C)$ and $\lambda \in \Delta_k$ such that

$$\mathbf{x}^* = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i,$$

and $\lambda_i > 0$ for all $i = 1, 2, \ldots, k$. Hence, by the convexity of f we have

$$f(\mathbf{x}^*) \le \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i),$$

or equivalently

$$\sum_{i=1}^{k} \lambda_i (f(\mathbf{x}_i) - f(\mathbf{x}^*)) \ge 0. \tag{7.19}$$

Since $\mathbf{x}^*$ is a maximizer of f over C, we have $f(\mathbf{x}_i) \le f(\mathbf{x}^*)$ for all $i = 1, 2, \ldots, k$. This means that inequality (7.19) states that a sum of nonpositive numbers is nonnegative, implying that each of the terms is zero, that is, $f(\mathbf{x}_i) = f(\mathbf{x}^*)$. Consequently, the extreme points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are all maximizers of f over C. □

Example 7.43. Consider the problem

$$\max\{\mathbf{x}^T \mathbf{Q} \mathbf{x} : \|\mathbf{x}\|_\infty \le 1\},$$

where $\mathbf{Q} \succeq 0$. Since the objective function is convex, and the feasible set is convex and compact, it follows that there exists a maximizer at an extreme point of the feasible set. The set of extreme points of the feasible set is $\{-1, 1\}^n$, and hence we conclude that there exists a maximizer that satisfies that each of its components is equal to 1 or -1. ■

Example 7.44 (computation of $\|\mathbf{A}\|_{1,1}$). Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Recall that (see Example 1.8)

$$\|\mathbf{A}\|_{1,1} = \max\{\|\mathbf{A}\mathbf{x}\|_1 : \|\mathbf{x}\|_1 \le 1\}.$$

Since the optimization problem consists of maximizing a convex function (composition of a norm function with a linear function) over a compact convex set, there exists a maximizer which is an extreme point of the l_1 ball. Note that there are exactly $2n$ extreme points to the l_1 ball: $\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_2, \ldots, \mathbf{e}_n, -\mathbf{e}_n$. In addition,

$$\|\mathbf{A}\mathbf{e}_j\|_1 = \|\mathbf{A}(-\mathbf{e}_j)\|_1 = \sum_{i=1}^m |A_{i,j}|,$$

and thus

$$\|\mathbf{A}\|_{1,1} = \max_{j=1,2,\ldots,n} \|\mathbf{A}\mathbf{e}_j\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^m |A_{i,j}|.$$

This is exactly the maximum absolute column sum norm introduced in Example 1.8. ∎

7.9 ▪ Convexity and Inequalities

Convexity is a powerful tool for proving inequalities. For example, the arithmetic geometric mean (AGM) inequality follows directly from the convexity of the scalar function $-\ln(x)$ over $\mathbb{R}_{++}$.

Proposition 7.45 (AGM inequality). *For any $x_1, x_2, \ldots, x_n \geq 0$ the following inequality holds:*

$$\frac{1}{n}\sum_{i=1}^n x_i \geq \sqrt[n]{\prod_{i=1}^n x_i}. \tag{7.20}$$

More generally, for any $\lambda \in \Delta_n$ one has

$$\sum_{i=1}^n \lambda_i x_i \geq \prod_{i=1}^n x_i^{\lambda_i}. \tag{7.21}$$

Proof. Employing Jensen's inequality on the convex function $f(x) = -\ln(x)$, we have that for any $x_1, x_2, \ldots, x_n > 0$ and $\lambda \in \Delta_n$

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i),$$

and hence

$$-\ln\left(\sum_{i=1}^n \lambda_i x_i\right) \leq -\sum_{i=1}^n \lambda_i \ln(x_i)$$

or

$$\ln\left(\sum_{i=1}^n \lambda_i x_i\right) \geq \sum_{i=1}^n \lambda_i \ln(x_i).$$

Taking the exponential function of both sides we have

$$\sum_{i=1}^n \lambda_i x_i \geq e^{\sum_{i=1}^n \lambda_i \ln(x_i)},$$

which is the same as the generalized AGM inequality (7.21). Plugging in $\lambda_i = \frac{1}{n}$ for all i yields the special case (7.20). We have proven the AGM inequalities only for the case when $x_1, x_2, \ldots, x_n$ are all positive. However, they are trivially satisfied if there exists an i for which $x_i = 0$, and hence the inequalities are valid for any $x_1, x_2, \ldots, x_n \geq 0$. $\square$

A direct result of the generalized AGM inequality is *Young's inequality*.

Lemma 7.46 (Young's inequality). *For any $s, t \geq 0$ and $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$ it holds that*

$$st \leq \frac{s^p}{p} + \frac{t^q}{q}. \tag{7.22}$$

Proof. By the generalized AGM inequality we have for any $x, y \geq 0$

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}.$$

Setting $x = s^p, y = t^q$ in the latter inequality, the result follows. $\square$

We can now prove several important inequalities. The first one is Hölder's inequality, which is a generalization of the Cauchy–Schwarz inequality.

Lemma 7.47 (Hölder's inequality). *For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $p, q \geq 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$ it holds that*

$$|\mathbf{x}^T \mathbf{y}| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q.$$

Proof. First, if $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$, then the inequality is trivial. Suppose then that $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$. For any $i \in \{1, 2, \ldots, n\}$, setting $s = \frac{|x_i|}{||\mathbf{x}||_p}$ and $t = \frac{|y_i|}{||\mathbf{x}||_q}$ in (7.22) yields the inequality

$$\frac{|x_i y_i|}{||\mathbf{x}||_p ||\mathbf{y}||_q} \leq \frac{1}{p} \frac{|x_i|^p}{||\mathbf{x}||_p^p} + \frac{1}{q} \frac{|y_i|^q}{||\mathbf{y}||_q^q}.$$

Summing the above inequality over $i = 1, 2, \ldots, n$ we obtain

$$\frac{\sum_{i=1}^n |x_i y_i|}{||\mathbf{x}||_p ||\mathbf{y}||_p} \leq \frac{1}{p} \frac{\sum_{i=1}^n |x_i|^p}{||\mathbf{x}||_p^p} + \frac{1}{q} \frac{\sum_{i=1}^n |y_i|^q}{||\mathbf{y}||_q^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Hence, by the triangle inequality we have

$$|\mathbf{x}^T \mathbf{y}| \leq \sum_{i=1}^n |x_i y_i| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q. \quad \square$$

Of course, for $p = q = 2$ Hölder's inequality is just the Cauchy–Schwarz inequality. Another inequality that can be deduced as a result of convexity is Minkowski's inequality, stating that the p-norm (for $p \geq 1$) satisfies the triangle inequality.

Lemma 7.48 (Minkowski's inequality). *Let $p \geq 1$. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the inequality*

$$||\mathbf{x} + \mathbf{y}||_p \leq ||\mathbf{x}||_p + ||\mathbf{y}||_p$$

holds.

Proof. For $p = 1$, the inequality follows by summing up the inequalities $|x_i + y_i| \le |x_i| + |y_i|$. Suppose then that $p > 1$. We can assume that $\mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$, and $\mathbf{x} + \mathbf{y} \neq \mathbf{0}$. Otherwise, the inequality is trivial. The function $\varphi(t) = t^p$ is convex over $\mathbb{R}_+$ since $\varphi''(t) = p(p-1)t^{p-2} > 0$ for $t > 0$. Therefore, by the definition of convexity we have that for any $\lambda_1, \lambda_2 \ge 0$ with $\lambda_1 + \lambda_2 = 1$ one has

$$(\lambda_1 t + \lambda_2 s)^p \le \lambda_1 t^p + \lambda_2 s^p.$$

Let $i \in \{1, 2, \ldots, n\}$. Plugging $\lambda_1 = \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}, \lambda_2 = \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}, t = \frac{|x_i|}{\|\mathbf{x}\|_p}$, and $s = \frac{|y_i|}{\|\mathbf{y}\|_p}$ in the above inequality yields

$$\frac{1}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p}(|x_i| + |y_i|)^p \le \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} \frac{|x_i|^p}{\|\mathbf{x}\|_p^p} + \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} \frac{|y_i|^p}{\|\mathbf{y}\|_p^p}.$$

Summing the above inequality over $i = 1, 2, \ldots, n$, we obtain that

$$\frac{1}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p} \sum_{i=1}^n (|x_i| + |y_i|)^p \le \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} + \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} = 1,$$

and hence

$$\sum_{i=1}^n (|x_i| + |y_i|)^p \le (\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p.$$

Finally,

$$\|\mathbf{x} + \mathbf{y}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i + y_i|^p} \le \sqrt[p]{\sum_{i=1}^n (|x_i| + |y_i|)^p} \le \|\mathbf{x}\|_p + \|\mathbf{y}\|_p. \qquad \square$$

Exercises

7.1. For each of the following sets determine whether they are convex or not (explaining your choice).

(i) $C_1 = \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|^2 = 1\right\}$.

(ii) $C_2 = \left\{\mathbf{x} \in \mathbb{R}^n : \max_{i=1,2,\ldots,n} x_i \le 1\right\}$.

(iii) $C_3 = \left\{\mathbf{x} \in \mathbb{R}^n : \min_{i=1,2,\ldots,n} x_i \le 1\right\}$.

(iv) $C_4 = \left\{\mathbf{x} \in \mathbb{R}^n_{++} : \prod_{i=1}^n x_i \ge 1\right\}$.

7.2. Show that the set

$$M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{Q} \mathbf{x} \le (\mathbf{a}^T \mathbf{x})^2, \mathbf{a}^T \mathbf{x} \ge 0\},$$

where $\mathbf{Q}$ is an $n \times n$ positive definite matrix and $\mathbf{a} \in \mathbb{R}^n$ is a convex cone.

7.3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex as well as concave function. Show that f is an affine function; that is, there exist $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ for any $\mathbf{x} \in \mathbb{R}^n$.

7.4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function. Show that for any $\varepsilon > 0$, the function

$$g_\varepsilon(\mathbf{x}) = f(\mathbf{x}) + \varepsilon\|\mathbf{x}\|^2$$

is coercive.

7.5. Let $f : \mathbb{R}^n \to \mathbb{R}$. Prove that f is convex if and only if for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{d} \neq \mathbf{0}$, the one-dimensional function $g_{\mathbf{x},\mathbf{d}}(t) = f(\mathbf{x} + t\mathbf{d})$ is convex.

7.6. Prove Theorem 7.13.

7.7. Let $C \subseteq \mathbb{R}^n$ be a convex set. Let f be a convex function over C, and let g be a strictly convex function over C. Show that the sum function $f + g$ is strictly convex over C.

7.8. (i) Let f be a convex function defined on a convex set C. Suppose that f is *not* strictly convex on C. Prove that there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n (\mathbf{x} \neq \mathbf{y})$ such that f is affine over the segment $[\mathbf{x}, \mathbf{y}]$.

(ii) Prove that the function $f(x) = x^4$ is strictly convex on $\mathbb{R}$ and that $g(x) = x^p$ for $p > 1$ is strictly convex over $\mathbb{R}_+$.

7.9. Show that the log-sum-exp function $f(\mathbf{x}) = \ln\left(\sum_{i=1}^n e^{x_i}\right)$ is *not* strictly convex over $\mathbb{R}^n$.

7.10. Show that the following functions are convex over the specified domain C:

(i) $f(x_1, x_2, x_3) = -\sqrt{x_1 x_2} + 2x_1^2 + 2x_2^2 + 3x_3^2 - 2x_1x_2 - 2x_2x_3$ over $\mathbb{R}^3_{++}$.

(ii) $f(\mathbf{x}) = \|\mathbf{x}\|^4$ over $\mathbb{R}^n$.

(iii) $f(\mathbf{x}) = \sum_{i=1}^n x_i \ln(x_i) - \left(\sum_{i=1}^n x_i\right)\ln\left(\sum_{i=1}^n x_i\right)$ over $\mathbb{R}^n_{++}$.

(iv) $f(\mathbf{x}) = \sqrt{\mathbf{x}^T\mathbf{Q}\mathbf{x} + 1}$ over $\mathbb{R}^n$, where $\mathbf{Q} \succeq \mathbf{0}$ is an $n \times n$ matrix.

(v) $f(x_1, x_2, x_3) = \max\{\sqrt{x_1^2 + x_2^2 + 20x_3^2 - x_1x_2 - 4x_2x_3 + 1}, (x_1^2 + x_2^2 + x_1 + x_2 + 2)^2\}$ over $\mathbb{R}^3$.

(vi) $f(x_1, x_2) = (2x_1^2 + 3x_2^2)\left(\frac{1}{2}x_1^2 + \frac{1}{3}x_2^2\right)$.

7.11. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(\mathbf{x}) = \ln\left(\sum_{i=1}^m e^{\mathbf{A}_i\mathbf{x}}\right),$$

where $\mathbf{A}_i$ is the ith row of $\mathbf{A}$. Prove that f is convex over $\mathbb{R}^n$.

7.12. Prove that the following set is a convex subset of $\mathbb{R}^{n+2}$:

$$C = \left\{ \begin{pmatrix} \mathbf{x} \\ y \\ z \end{pmatrix} : \|\mathbf{x}\|^2 \leq yz, \mathbf{x} \in \mathbb{R}^n, y, z \in \mathbb{R}_+ \right\}.$$

7.13. Show that the function $f(x_1, x_2, x_3) = -e^{(-x_1+x_2-2x_3)^2}$ is not convex over $\mathbb{R}^n$.

7.14. Prove that the geometric mean function $f(\mathbf{x}) = \sqrt[n]{\prod_{i=1}^n x_i}$ is concave over $\mathbb{R}^n_{++}$. Is it strictly concave over $\mathbb{R}^n_{++}$?

7.15. (i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function which is nondecreasing with respect to each of its variables separately; that is, for any $i \in \{1,2,\ldots,n\}$ and fixed $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, the one-dimensional function

$$g_i(y) = f(x_1, x_2, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)$$

is nondecreasing with respect to y. Let $h_1, h_2, \ldots, h_n : \mathbb{R}^p \to \mathbb{R}$ be convex functions. Prove that the composite function

$$r(z_1, z_2, \ldots, z_p) = f(h_1(z_1, z_2, \ldots, z_p), \ldots, h_n(z_1, z_2, \ldots, z_p))$$

is convex.

(ii) Prove that the function $f(x_1, x_2) = \ln(e^{x_1^2+x_2^2} + e^{\sqrt{x_1^2+1}})$ is convex over $\mathbb{R}^2$.

7.16. Let f be a convex function over $\mathbb{R}^n$ and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha > 0$. Define $\mathbf{z} = \mathbf{x} + \frac{1}{\alpha}(\mathbf{x} - \mathbf{y})$. Prove that

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \alpha(f(\mathbf{x}) - f(\mathbf{z})).$$

7.17. Let f be a convex function over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}_1, \mathbf{x}_3 \in C$ and let $\mathbf{x}_2 \in [\mathbf{x}_1, \mathbf{x}_3]$. Prove that if $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are different from each other, then

$$\frac{f(\mathbf{x}_3) - f(\mathbf{x}_2)}{\|\mathbf{x}_3 - \mathbf{x}_2\|} \geq \frac{f(\mathbf{x}_2) - f(\mathbf{x}_1)}{\|\mathbf{x}_2 - \mathbf{x}_1\|}.$$

7.18. Let $\phi : \mathbb{R}_{++} \to \mathbb{R}$ be a convex function. Then the function $f : \mathbb{R}^2_{++} \to \mathbb{R}$ defined by

$$f(x,y) = y\phi\left(\frac{x}{y}\right), \quad x, y > 0,$$

is convex over $\mathbb{R}^2_{++}$.

7.19. Prove that the function $f(x,y) = -x^p y^{1-p} (0 < p < 1)$ is convex over $\mathbb{R}^2_{++}$.

7.20. Let $f : C \to \mathbb{R}$ be a function defined over the convex set $C \subseteq \mathbb{R}^n$. Prove that f is quasi-convex if and only if

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}, \quad \text{for any } \mathbf{x}, \mathbf{y} \in C, \lambda \in [0,1].$$

7.21. Let $f(\mathbf{x}) = \frac{g(\mathbf{x})}{h(\mathbf{x})}$, where g is a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and $h(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ for some $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Assume that $h(\mathbf{x}) > 0$ for all $\mathbf{x} \in C$. Show that f is quasi-convex over C.

7.22. Show an example of two quasi-convex functions whose sum is *not* a quasi-convex function.

7.23. Let $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$, where $\mathbf{A}$ is an $n \times n$ symmetric matrix, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Show that f is *quasi-convex* if and only if $\mathbf{A} \succeq 0$.

7.24. A function $f : C \to \mathbb{R}$ is called *log-concave* over the convex set $C \subseteq \mathbb{R}^n$ if $f(\mathbf{x}) > 0$ for any $\mathbf{x} \in C$ and $\ln(f)$ is a concave function over C.

(i) Show that the function $f(\mathbf{x}) = \frac{1}{\sum_{i=1}^n \frac{1}{x_i}}$ is log-concave over $\mathbb{R}^n_{++}$.

(ii) Let f be a twice continuously differentiable function over $\mathbb{R}$ with $f(x) > 0$ for all $x \in \mathbb{R}$. Show that f is log-concave if and only if $f''(x)f(x) \leq (f'(x))^2$ for all $x \in \mathbb{R}$.

7.25. Prove that if f and g are convex, twice differentiable, nondecreasing, and positive on $\mathbb{R}$, then the product fg is convex over $\mathbb{R}$. Show by an example that the positivity assumption is necessary to establish the convexity.

7.26. Let C be a convex subset of $\mathbb{R}^n$. A function f is called *strongly convex* over C if there exists $\sigma > 0$ such that the function $f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x}\|^2$ is convex over C. The parameter σ is called *the strong convexity parameter*. In the following questions C is a given convex subset of $\mathbb{R}^n$.

(i) Prove that f is strongly convex over C with parameter σ if and only if

$$f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|^2$$

for any $\mathbf{x},\mathbf{y} \in C$ and $\lambda \in [0,1]$.

(ii) Prove that a strongly convex function over C is also strictly convex over C.

(iii) Suppose that f is continuously differentiable over C. Prove that f is strongly convex over C with parameter σ if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{\sigma}{2}\|\mathbf{x}-\mathbf{y}\|^2$$

for any $\mathbf{x},\mathbf{y} \in C$.

(iv) Suppose that f is continuously differentiable over C. Prove that f is strongly convex over C with parameter σ if and only if

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x}-\mathbf{y}) \geq \sigma\|\mathbf{x}-\mathbf{y}\|^2$$

for any $\mathbf{x},\mathbf{y} \in C$.

(v) Suppose that f is twice continuously differentiable over C. Show that f is strongly convex over C with parameter σ if and only if $\nabla^2 f(\mathbf{x}) \succeq \sigma\mathbf{I}$ for any $\mathbf{x} \in C$.

7.27. (i) Show that the function $f(\mathbf{x}) = \sqrt{1+\|\mathbf{x}\|^2}$ is strictly convex over $\mathbb{R}^n$ but is not *strongly convex* over $\mathbb{R}^n$.

(ii) Show that the quadratic function $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$ with $\mathbf{A} = \mathbf{A}^T \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$ is strongly convex if and only if $\mathbf{A} \succ 0$, and in that case the strong convexity parameter is $2\lambda_{\min}(\mathbf{A})$.

7.28. Let $f \in C_L^{1,1}(\mathbb{R}^n)$ be a convex function. For a fixed $\mathbf{x} \in \mathbb{R}^n$ define the function

$$g_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{y}) - \nabla f(\mathbf{x})^T\mathbf{y}.$$

(i) Prove that $\mathbf{x}$ is a minimizer of $g_{\mathbf{x}}$ over $\mathbb{R}^n$.

(ii) Show that for any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$

$$g_{\mathbf{x}}(\mathbf{x}) \leq g_{\mathbf{x}}(\mathbf{y}) - \frac{1}{2L}\|\nabla g_{\mathbf{x}}(\mathbf{y})\|^2.$$

(iii) Show that for any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$

$$f(\mathbf{x}) \leq f(\mathbf{y}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) - \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2.$$

(iv) Prove that for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$

$$\langle\nabla f(\mathbf{x})-\nabla f(\mathbf{y}),\mathbf{x}-\mathbf{y}\rangle\geq\frac{1}{L}\|\nabla f(\mathbf{x})-\nabla f(\mathbf{y})\|^2 \text{ for any } \mathbf{x},\mathbf{y}\in\mathbb{R}^n.$$

7.29. Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ be an extended real-valued function. Show that f is convex if and only if $\text{epi}(f)$ is convex.

7.30. Show that the support function of the set $S=\{\mathbf{x}\in\mathbb{R}^n:\mathbf{x}^T\mathbf{Q}\mathbf{x}\leq 1\}$, where $\mathbf{Q}\succ 0$, is $\sigma_S(\mathbf{y})=\sqrt{\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y}}$.

7.31. Let $S=\{\mathbf{x}\in\mathbb{R}^n:\mathbf{a}^T\mathbf{x}\leq b\}$, where $0\neq\mathbf{a}\in\mathbb{R}^n$ and $b\in\mathbb{R}$. Find the support function σ_S.

7.32. Let $p>1$. Show that the support function of $S=\{\mathbf{x}\in\mathbb{R}^n:\|\mathbf{x}\|_p\leq 1\}$ is $\sigma_S(\mathbf{y})=\|\mathbf{y}\|_q$, where q is defined by the relation $\frac{1}{p}+\frac{1}{q}=1$.

7.33. Let $f_0,f_1,\ldots,f_m$ be convex functions over $\mathbb{R}^n$ and consider the perturbation function

$$F(\mathbf{b})=\min_{\mathbf{x}}\{f_0(\mathbf{x}):f_i(\mathbf{x})\leq b_i, i=1,2,\ldots,m\}.$$

Assume that for any $\mathbf{b}\in\mathbb{R}^m$ the minimization problem in the above definition of $F(\mathbf{b})$ has an optimal solution. Show that F is convex over $\mathbb{R}^m$.

7.34. Let $C\subseteq\mathbb{R}^n$ be a convex set and let $\phi_1,\ldots,\phi_m$ be convex functions over C. Let U be the following subset of $\mathbb{R}^m$:

$$U=\{\mathbf{y}\in\mathbb{R}^m:\phi_1(\mathbf{x})\leq y_1,\ldots,\phi_m(\mathbf{x})\leq y_m \text{ for some } \mathbf{x}\in C\}.$$

Show that U is a convex set.

7.35. (i) Show that the extreme points of the unit simplex Δ_n are the unit-vectors $\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$.

(ii) Find the optimal solution of the problem

$$\begin{array}{ll}\max & 57x_1^2+65x_2^2+17x_3^2+96x_1x_2-32x_1x_3+8x_2x_3+27x_1-84x_2+20x_3\\ \text{s.t.} & x_1+x_2+x_3=1\\ & x_1,x_2,x_3\geq 0.\end{array}$$

7.36. Prove that for any $x_1,x_2,\ldots,x_n\in\mathbb{R}_+$ the following inequality holds:

$$\frac{\sum_{i=1}^n x_i}{n}\leq\sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}.$$

7.37. Prove that for any $x_1,x_2,\ldots,x_n\in\mathbb{R}_{++}$ the following inequality holds:

$$\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i}\leq\sqrt{\frac{\sum_{i=1}^n x_i^3}{\sum_{i=1}^n x_i}}.$$

7.38. Let $x_1,x_2,\ldots,x_n>0$ satisfy $\sum_{i=1}^n x_i=1$. Prove that

$$\sum_{i=1}^n\frac{x_i}{\sqrt{1-x_i}}\geq\sqrt{\frac{n}{n-1}}.$$

7.39. Prove that for any $a, b, c > 0$ the following inequality holds:

$$\frac{9}{a+b+c} \leq 2\left(\frac{1}{a+b} + \frac{1}{b+c} + \frac{1}{c+a}\right).$$

7.40. (i) Prove that the function $f(x) = \frac{1}{1+e^x}$ is strictly convex over $[0, \infty)$.

(ii) Prove that for any $a_1, a_2, \ldots, a_n \geq 1$ the inequality

$$\sum_{i=1}^{n} \frac{1}{1+a_i} \geq \frac{n}{1+\sqrt[n]{a_1 a_2 \cdots a_n}}$$

holds.