

Chapter 2

Optimality Conditions for Unconstrained Optimization

2.1 ■ Global and Local Optima

Although our main interest in this section is to discuss minimum and maximum points of a function over the entire space, we will nonetheless present the more general definition of global minimum and maximum points of a function over a given set.

Definition 2.1 (global minimum and maximum). Let $f : S \rightarrow \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $\mathbf{x}^* \in S$ is called a **global minimum point** of f over S if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in S$,
2. $\mathbf{x}^* \in S$ is called a **strict global minimum point** of f over S if $f(\mathbf{x}) > f(\mathbf{x}^*)$ for any $\mathbf{x}^* \neq \mathbf{x} \in S$,
3. $\mathbf{x}^* \in S$ is called a **global maximum point** of f over S if $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ for any $\mathbf{x} \in S$,
4. $\mathbf{x}^* \in S$ is called a **strict global maximum point** of f over S if $f(\mathbf{x}) < f(\mathbf{x}^*)$ for any $\mathbf{x}^* \neq \mathbf{x} \in S$.

The set S on which the optimization of f is performed is also called *the feasible set*, and any point $\mathbf{x} \in S$ is called *a feasible solution*. We will frequently omit the adjective “global” and just use the terminology “minimum point” and “maximum point.” It is also customary to refer to a global minimum point as a *minimizer* or a *global minimizer* and to a global maximum point as a *maximizer* or a *global maximizer*.

A vector $\mathbf{x}^* \in S$ is called a *global optimum* of f over S if it is either a global minimum or a global maximum. The *maximal value of f over S* is defined as the supremum of f over S :

$$\max\{f(\mathbf{x}) : \mathbf{x} \in S\} = \sup\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

If $\mathbf{x}^* \in S$ is a global maximum of f over S , then the maximum value of f over S is $f(\mathbf{x}^*)$. Similarly the *minimal value of f over S* is the infimum of f over S ,

$$\min\{f(\mathbf{x}) : \mathbf{x} \in S\} = \inf\{f(\mathbf{x}) : \mathbf{x} \in S\},$$

and is equal to $f(\mathbf{x}^*)$ when \mathbf{x}^* is a global minimum of f over S . Note that in this book we will not use the sup/inf notation but rather use only the min/max notation, where the

usage of this notation does not imply that the maximum or minimum is actually attained. As opposed to global maximum and minimum points, minimal and maximal values are always unique. There could be several global minimum points, but there could be only one minimal value. The set of all global minimizers of f over S is denoted by

$$\operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in S\},$$

and the set of all global maximizers of f over S is denoted by

$$\operatorname{argmax}\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

Note that the notation " $f : S \rightarrow \mathbb{R}$ " means in particular that S is the *domain* of f , that is, the subset of \mathbb{R}^n on which f is defined. In Definition 2.1 the minimization and maximization is over the domain of the function. However, later on we will also deal with functions $f : S \rightarrow \mathbb{R}$ and discuss problems of finding global optima points with respect to a subset of the domain.

Example 2.2. Consider the two-dimensional linear function $f(x, y) = x + y$ defined over the unit ball $S = B[0, 1] = \{(x, y)^T : x^2 + y^2 \leq 1\}$. Then by the Cauchy-Schwarz inequality we have for any $(x, y)^T \in S$

$$x + y = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \leq \sqrt{x^2 + y^2} \sqrt{1^2 + 1^2} \leq \sqrt{2}.$$

Therefore, the maximal value of f over S is upper bounded by $\sqrt{2}$. On the other hand, the upper bound $\sqrt{2}$ is attained at $(x, y) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. It is not difficult to see that this is the only point that attains this value, and thus $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is the strict global maximum point of f over S , and the maximal value is $\sqrt{2}$. A similar argument shows that $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ is the strict global minimum point of f over S , and the minimal value is $-\sqrt{2}$. ■

Example 2.3. Consider the two-dimensional function

$$f(x, y) = \frac{x + y}{x^2 + y^2 + 1}$$

defined over the entire space \mathbb{R}^2 . The contour and surface plots of the function are given in Figure 2.1. This function has two optima points: a global maximizer $(x, y) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and a global minimizer $(x, y) = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. The proof of these facts will be given in Example 2.36. The maximal value of the function is

$$\frac{\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}}{(\frac{1}{\sqrt{2}})^2 + (\frac{1}{\sqrt{2}})^2 + 1} = \frac{1}{\sqrt{2}},$$

and the minimal value is $-\frac{1}{\sqrt{2}}$. ■

Our main task will usually be to find and study global minimum or maximum points; however, most of the theoretical results only characterize *local* minima and maxima which are optimal points with respect to a neighborhood of the point of interest. The exact definitions follow.

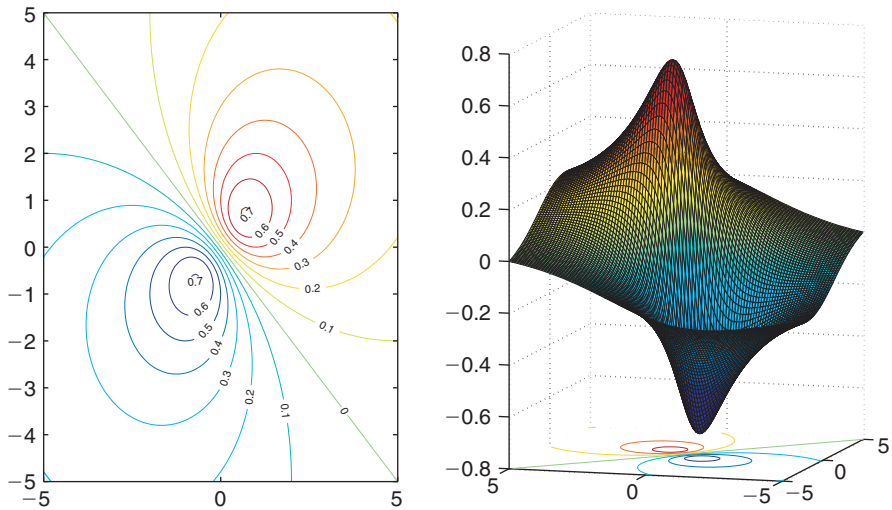


Figure 2.1. Contour and surface plots of $f(x, y) = \frac{x+y}{x^2+y^2+1}$.

Definition 2.4 (local minima and maxima). Let $f : S \rightarrow \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $\mathbf{x}^* \in S$ is called a **local minimum point** of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$,
2. $\mathbf{x}^* \in S$ is called a **strict local minimum point** of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) < f(\mathbf{x})$ for any $\mathbf{x}^* \neq \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$,
3. $\mathbf{x}^* \in S$ is called a **local maximum point** of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$,
4. $\mathbf{x}^* \in S$ is called a **strict local maximum point** of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) > f(\mathbf{x})$ for any $\mathbf{x}^* \neq \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.

Of course, a global minimum (maximum) point is also a local minimum (maximum) point. As with global minimum and maximum points, we will also use the terminology *local minimizer* and *local maximizer* for local minimum and maximum points, respectively.

Example 2.5. Consider the one-dimensional function

$$f(x) = \begin{cases} (x-1)^2 + 2, & -1 \leq x \leq 1, \\ 2, & 1 \leq x \leq 2, \\ -(x-2)^2 + 2, & 2 \leq x \leq 2.5, \\ (x-3)^2 + 1.5, & 2.5 \leq x \leq 4, \\ -(x-5)^2 + 3.5, & 4 \leq x \leq 6, \\ -2x + 14.5, & 6 \leq x \leq 6.5, \\ 2x - 11.5, & 6.5 \leq x \leq 8, \end{cases}$$

described in Figure 2.2 and defined over the interval $[-1, 8]$. The point $x = -1$ is a strict global maximum point. The point $x = 1$ is a nonstrict local minimum point. All the points in the interval $(1, 2)$ are nonstrict local minimum points as well as nonstrict local maximum points. The point $x = 2$ is a local maximum point. The point $x = 3$ is a strict

local minimum, and a non-strict global minimum point. The point $x = 5$ is a strict local maximum and $x = 6.5$ is a strict local minimum, which is a nonstrict global minimum point. Finally, $x = 8$ is a strict local maximum point. Note that, as already mentioned, $x = 3$ and $x = 6.5$ are both global minimum points of the function, and despite the fact that they are strict local minima, they are nonstrict global minimum points. ■

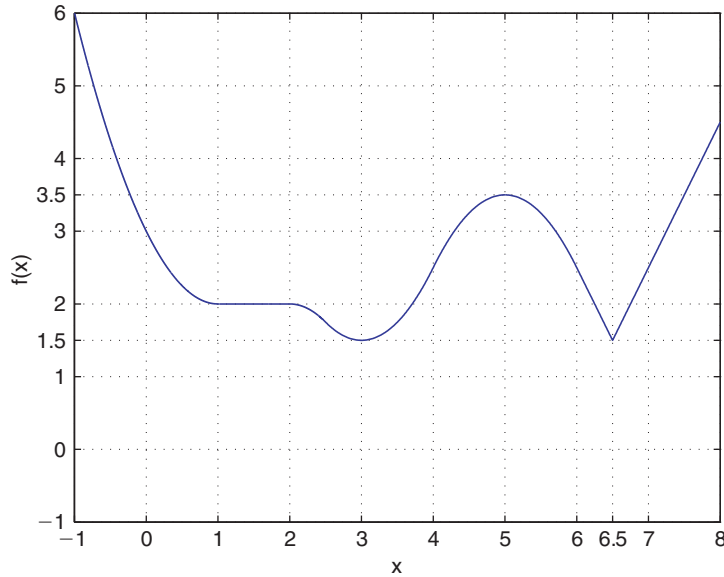


Figure 2.2. Local and global optimum points of a one-dimensional function.

First Order Optimality Condition

A well-known result is that for a one-dimensional function f defined and differentiable over an interval (a, b) , if a point $x^* \in (a, b)$ is a local maximum or minimum, then $f'(x^*) = 0$. This is also known as Fermat's theorem. The multidimensional extension of this result states that the gradient is zero at local optimum points. We refer to such an optimality condition as a *first order optimality condition*, as it is expressed in terms of the first order derivatives. In what follows, we will also discuss *second order optimality conditions* that use in addition information on the second order (partial) derivatives.

Theorem 2.6 (first order optimality condition for local optima points). *Let $f : U \rightarrow \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ is a local optimum point and that all the partial derivatives of f exist at \mathbf{x}^* . Then $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

Proof. Let $i \in \{1, 2, \dots, n\}$ and consider the one-dimensional function $g(t) = f(\mathbf{x}^* + t\mathbf{e}_i)$. Note that g is differentiable at $t = 0$ and that $g'(0) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*)$. Since \mathbf{x}^* is a local optimum point of f , it follows that $t = 0$ is a local optimum of g , which immediately implies that $g'(0) = 0$. The latter equality is exactly the same as $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$. Since this is true for any $i \in \{1, 2, \dots, n\}$, the result $\nabla f(\mathbf{x}^*) = \mathbf{0}$ follows. □

Note that the proof of the first order optimality conditions for multivariate functions strongly relies on the first order optimality conditions for one-dimensional functions.

Theorem 2.6 presents a *necessary* optimality condition: the gradient vanishes at all local optimum points, which are interior points of the domain of the function; however, the reverse claim is not true—there could be points which are not local optimum points, whose gradient is zero. For example, the derivative of the one-dimensional function $f(x) = x^3$ is zero at $x = 0$, but this point is neither a local minimum nor a local maximum. Since points in which the gradient vanishes are the only candidates for local optima among the points in the interior of the domain of the function, they deserve an explicit definition.

Definition 2.7 (stationary points). Let $f : U \rightarrow \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ and that f is differentiable over some neighborhood of \mathbf{x}^* . Then \mathbf{x}^* is called a **stationary point** of f if $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Theorem 2.6 essentially states that local optimum points are necessarily stationary points.

Example 2.8. Consider the one-dimensional quartic function $f(x) = 3x^4 - 20x^3 + 42x^2 - 36x$. To find its local and global optima points over \mathbb{R} , we first find all its stationary points. Since $f'(x) = 12x^3 - 60x^2 + 84x - 36 = 12(x-1)^2(x-3)$, it follows that $f'(x) = 0$ for $x = 1, 3$. The derivative $f'(x)$ does not change its sign when passing through $x = 1$ —it is negative before and after—and thus $x = 1$ is not a local or global optimum point. On the other hand, the derivative does change its sign from negative to positive while passing through $x = 3$, and thus it is a local minimum point. Since the function must have a global minimum by the property that $f(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, it follows that $x = 3$ is the global minimum point. ■

2.2 ■ Classification of Matrices

In order to be able to characterize the second order optimality conditions, which are expressed via the Hessian matrix, the notion of “positive definiteness” must be defined.

Definition 2.9 (positive definiteness).

1. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called **positive semidefinite**, denoted by $\mathbf{A} \succeq \mathbf{0}$, if $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$.
2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called **positive definite**, denoted by $\mathbf{A} \succ \mathbf{0}$, if $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for every $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.

In this book a positive definite or semidefinite matrix is always assumed to be symmetric.

Positive definiteness of a matrix does not mean that its components are positive, as the following examples illustrate.

Example 2.10. Let

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Then for any $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = (x_1, x_2) \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 - 2x_1x_2 + x_2^2 = x_1^2 + (x_1 - x_2)^2 \geq 0.$$

Thus, \mathbf{A} is positive semidefinite. In fact, since $x_1^2 + (x_1 - x_2)^2 = 0$ if and only if $x_1 = x_2 = 0$, it follows that \mathbf{A} is positive definite. This example illustrates the fact that a positive definite matrix might have negative components. ■

Example 2.11. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

This matrix, whose components are all positive, is not positive definite since for $\mathbf{x} = (1, -1)^T$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = (1, -1) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2. \quad \blacksquare$$

Although, as illustrated in the above examples, not all the components of a positive definite matrix need to be positive, the following result shows that the signs of the *diagonal* components of a positive definite matrix are positive.

Lemma 2.12. *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a positive definite matrix. Then the diagonal elements of \mathbf{A} are positive.*

Proof. Since \mathbf{A} is positive definite, it follows that $\mathbf{e}_i^T \mathbf{A} \mathbf{e}_i > 0$ for any $i \in \{1, 2, \dots, n\}$, which by the fact that $\mathbf{e}_i^T \mathbf{A} \mathbf{e}_i = A_{ii}$ implies the result. ■

A similar argument shows that the diagonal elements of a positive semidefinite matrix are nonnegative.

Lemma 2.13. *Let \mathbf{A} be a positive semidefinite matrix. Then the diagonal elements of \mathbf{A} are nonnegative.*

Closely related to the notion of positive (semi)definiteness are the notions of negative (semi)definiteness and indefiniteness.

Definition 2.14 (negative (semi)definiteness, indefiniteness).

1. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called **negative semidefinite**, denoted by $\mathbf{A} \preceq \mathbf{0}$, if $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0$ for every $\mathbf{x} \in \mathbb{R}^n$.
2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called **negative definite**, denoted by $\mathbf{A} \prec \mathbf{0}$, if $\mathbf{x}^T \mathbf{A} \mathbf{x} < 0$ for every $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.
3. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is called **indefinite** if there exist \mathbf{x} and $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ and $\mathbf{y}^T \mathbf{A} \mathbf{y} < 0$.

It follows immediately from the above definition that a matrix \mathbf{A} is negative (semi)definite if and only if $-\mathbf{A}$ is positive (semi)definite. Therefore, we can prove and state all the results for positive (semi)definite matrices and the corresponding results for negative (semi)definite matrices will follow immediately. For example, the following result on negative definite and negative semidefinite matrices is a direct consequence of Lemmata 2.12 and 2.13.

Lemma 2.15.

- (a) Let \mathbf{A} be a negative definite matrix. Then the diagonal elements of \mathbf{A} are negative.
- (b) Let \mathbf{A} be a negative semidefinite matrix. Then the diagonal elements of \mathbf{A} are nonpositive.

When the diagonal of a matrix contains both positive and negative elements, then the matrix is indefinite. The reverse claim is not correct.

Lemma 2.16. Let \mathbf{A} be a symmetric $n \times n$ matrix. If there exist positive and negative elements in the diagonal of \mathbf{A} , then \mathbf{A} is indefinite.

Proof. Let $i, j \in \{1, 2, \dots, n\}$ be indices such that $A_{i,i} > 0$ and $A_{j,j} < 0$. Then

$$\mathbf{e}_i^T \mathbf{A} \mathbf{e}_i = A_{i,i} > 0, \quad \mathbf{e}_j^T \mathbf{A} \mathbf{e}_j = A_{j,j} < 0,$$

and hence \mathbf{A} is indefinite. \square

In addition, a matrix is indefinite if and only if it is neither positive semidefinite nor negative semidefinite. It is not an easy task to check the definiteness of a matrix by using the definition given above. Therefore, our main task will be to find a useful characterization of positive (semi)definite matrices. It turns out that a complete characterization can be given in terms of the eigenvalues of the matrix.

Theorem 2.17 (eigenvalue characterization theorem). Let \mathbf{A} be a symmetric $n \times n$ matrix. Then

- (a) \mathbf{A} is positive definite if and only if all its eigenvalues are positive,
- (b) \mathbf{A} is positive semidefinite if and only if all its eigenvalues are nonnegative,
- (c) \mathbf{A} is negative definite if and only if all its eigenvalues are negative,
- (d) \mathbf{A} is negative semidefinite if and only if all its eigenvalues are nonpositive,
- (e) \mathbf{A} is indefinite if and only if it has at least one positive eigenvalue and at least one negative eigenvalue.

Proof. We will prove part (a). The other parts follow immediately or by similar arguments. Since \mathbf{A} is symmetric, it follows by the spectral decomposition theorem (Theorem 1.10) that there exist an orthogonal matrix \mathbf{U} and a diagonal matrix $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ whose diagonal elements are the eigenvalues of \mathbf{A} , for which $\mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{D}$. Making the linear transformation of variables $\mathbf{x} = \mathbf{U} \mathbf{y}$, we obtain that

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T \mathbf{U}^T \mathbf{A} \mathbf{U} \mathbf{y} = \mathbf{y}^T \mathbf{D} \mathbf{y} = \sum_{i=1}^n d_i y_i^2.$$

We can therefore conclude by the nonsingularity of \mathbf{U} that $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for any $\mathbf{x} \neq \mathbf{0}$ if and only if

$$\sum_{i=1}^n d_i y_i^2 > 0 \text{ for any } \mathbf{y} \neq \mathbf{0}. \quad (2.1)$$

Therefore, we need to prove that (2.1) holds if and only if $d_i > 0$ for all i . Indeed, if (2.1) holds then for any $i \in \{1, 2, \dots, n\}$, plugging $\mathbf{y} = \mathbf{e}_i$ in the inequality implies that $d_i > 0$. On the other hand, if $d_i > 0$ for any i , then surely for any nonzero vector \mathbf{y} one has $\sum_{i=1}^n d_i y_i^2 > 0$, meaning that (2.1) holds. \square

Since the trace and determinant of a symmetric matrix are the sum and product of its eigenvalues respectively, a simple consequence of the eigenvalue characterization theorem is the following.

Corollary 2.18. *Let \mathbf{A} be a positive semidefinite (definite) matrix. Then $\text{Tr}(\mathbf{A})$ and $\det(\mathbf{A})$ are nonnegative (positive).*

Since the eigenvalues of a diagonal matrix are its diagonal elements, it follows that the sign of a diagonal matrix is determined by the signs of the elements in its diagonal.

Lemma 2.19 (sign of diagonal matrices). *Let $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$. Then*

- (a) \mathbf{D} is positive definite if and only if $d_i > 0$ for all i ,
- (b) \mathbf{D} is positive semidefinite if and only if $d_i \geq 0$ for all i ,
- (c) \mathbf{D} is negative definite if and only if $d_i < 0$ for all i ,
- (d) \mathbf{D} is negative semidefinite if and only if $d_i \leq 0$ for all i ,
- (e) \mathbf{D} is indefinite if and only if there exist i, j such that $d_i > 0$ and $d_j < 0$.

The eigenvalues of a matrix give full information on the sign of the matrix. However, we would like to find other simpler methods for detecting positive (semi)definiteness. We begin with an extremely simple rule for 2×2 matrices stating that for 2×2 matrices the condition in Corollary 2.18 is necessary and sufficient.

Proposition 2.20. *Let \mathbf{A} be a symmetric 2×2 matrix. Then \mathbf{A} is positive semidefinite (definite) if and only if $\text{Tr}(\mathbf{A}), \det(\mathbf{A}) \geq 0$ ($\text{Tr}(\mathbf{A}), \det(\mathbf{A}) > 0$).*

Proof. We will prove the result for positive semidefiniteness. The result for positive definiteness follows from similar arguments. By the eigenvalue characterization theorem (Theorem 2.17), it follows that \mathbf{A} is positive semidefinite if and only if $\lambda_1(\mathbf{A}) \geq 0$ and $\lambda_2(\mathbf{A}) \geq 0$. The result now follows from the simple fact that for any two real number $a, b \in \mathbb{R}$ one has $a, b \geq 0$ if and only if $a + b \geq 0$ and $a \cdot b \geq 0$. Therefore, \mathbf{A} is positive semidefinite if and only if $\lambda_1(\mathbf{A}) + \lambda_2(\mathbf{A}) = \text{Tr}(\mathbf{A}) \geq 0$ and $\lambda_1(\mathbf{A})\lambda_2(\mathbf{A}) = \det(\mathbf{A}) \geq 0$. \square

Example 2.21. Consider the matrices

$$\mathbf{A} = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0.1 \end{pmatrix}.$$

The matrix \mathbf{A} is positive definite since

$$\text{Tr}(\mathbf{A}) = 4 + 3 = 7 > 0, \quad \det(\mathbf{A}) = 4 \cdot 3 - 1 \cdot 1 = 11 > 0.$$

As for the matrix \mathbf{B} , we have

$$\text{Tr}(\mathbf{B}) = 1 + 1 + 0.1 = 2.1 > 0, \quad \det(\mathbf{B}) = 0.$$

However, despite the fact that the trace and the determinant of \mathbf{B} are nonnegative, we cannot conclude that the matrix is positive semidefinite since Proposition 2.20 is valid only for 2×2 matrices. In this specific example we can show (even without computing the eigenvalues) that \mathbf{B} is indefinite. Indeed,

$$\begin{aligned} \mathbf{e}_1^T \mathbf{B} \mathbf{e}_1 &> 0, \\ (\mathbf{e}_2 - \mathbf{e}_3)^T \mathbf{B} (\mathbf{e}_2 - \mathbf{e}_3) &= -0.9 < 0. \quad \blacksquare \end{aligned}$$

For any positive semidefinite matrix \mathbf{A} , we can define the square root matrix $\mathbf{A}^{\frac{1}{2}}$ in the following way. Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ be the spectral decomposition of \mathbf{A} ; that is, \mathbf{U} is an orthogonal matrix, and $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ is a diagonal matrix whose diagonal elements are the eigenvalues of \mathbf{A} . Since \mathbf{A} is positive semidefinite, we have that $d_1, d_2, \dots, d_n \geq 0$, and we can define

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\mathbf{E}\mathbf{U}^T,$$

where $\mathbf{E} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n})$. Obviously,

$$\mathbf{A}^{\frac{1}{2}} \mathbf{A}^{\frac{1}{2}} = \mathbf{U}\mathbf{E}\mathbf{U}^T \mathbf{U}\mathbf{E}\mathbf{U}^T = \mathbf{U}\mathbf{E}\mathbf{E}\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{U}^T = \mathbf{A}.$$

The matrix $\mathbf{A}^{\frac{1}{2}}$ is also called *the positive semidefinite square root*. A well-known test for positive definiteness is the *principal minors criterion*. Given an $n \times n$ matrix, the determinant of the upper left $k \times k$ submatrix is called *the k th principal minor* and is denoted by $D_k(\mathbf{A})$. For example, the principal minors of the 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

are

$$D_1(\mathbf{A}) = a_{11}, \quad D_2(\mathbf{A}) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad D_3(\mathbf{A}) = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The principal minors criterion states that a symmetric matrix is positive definite if and only if all its principal minors are positive.

Theorem 2.22 (principal minors criterion). *Let \mathbf{A} be an $n \times n$ symmetric matrix. Then \mathbf{A} is positive definite if and only if $D_1(\mathbf{A}) > 0, D_2(\mathbf{A}) > 0, \dots, D_n(\mathbf{A}) > 0$.*

Note that the principal minors criterion is a tool for detecting positive definiteness of a matrix. It cannot be used in order to detect positive semidefiniteness.

Example 2.23. Let

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & -1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} -4 & 1 & 1 \\ 1 & -4 & 1 \\ 1 & 1 & -4 \end{pmatrix}.$$

The matrix \mathbf{A} is positive definite since

$$D_1(\mathbf{A}) = 4 > 0, \quad D_2(\mathbf{A}) = \det \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix} = 8 > 0, \quad D_3(\mathbf{A}) = \det \begin{pmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix} = 13 > 0.$$

The principal minors of \mathbf{B} are nonnegative: $D_1(\mathbf{B}) = 2, D_2(\mathbf{B}) = D_3(\mathbf{B}) = 0$; however, since they are not positive, the principal minors criterion does not provide any information on the sign of the matrix other than the fact that it is not positive definite. Since the matrix has both positive and negative diagonal elements, it is in fact indefinite (see Lemma 2.16). As for the matrix \mathbf{C} , we will show that it is negative definite. For that, we will use the principal minors criterion to prove that $-\mathbf{C}$ is positive definite:

$$D_1(-\mathbf{C}) = 4 > 0, \quad D_2(-\mathbf{C}) = \det \begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix} = 15 > 0,$$

$$D_3(-\mathbf{C}) = \det \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix} = 50 > 0. \quad \blacksquare$$

An important class of matrices that are known to be positive semidefinite is the class of *diagonally dominant* matrices.

Definition 2.24 (diagonally dominant matrices). Let \mathbf{A} be a symmetric $n \times n$ matrix. Then

1. \mathbf{A} is called **diagonally dominant** if

$$|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$$

for all $i = 1, 2, \dots, n$,

2. \mathbf{A} is called **strictly diagonally dominant** if

$$|A_{ii}| > \sum_{j \neq i} |A_{ij}|$$

for all $i = 1, 2, \dots, n$.

We will now show that diagonally dominant matrices with nonnegative diagonal elements are positive semidefinite and that strictly diagonally dominant matrices with positive diagonal elements are positive definite.

Theorem 2.25 (positive (semi)definiteness of diagonally dominant matrices).

- (a) Let \mathbf{A} be a symmetric $n \times n$ diagonally dominant matrix whose diagonal elements are nonnegative. Then \mathbf{A} is positive semidefinite.
- (b) Let \mathbf{A} be a symmetric $n \times n$ strictly diagonally dominant matrix whose diagonal elements are positive. Then \mathbf{A} is positive definite.

Proof. (a) Suppose in contradiction that there exists a negative eigenvalue λ of \mathbf{A} , and let \mathbf{u} be a corresponding eigenvector. Let $i \in \{1, 2, \dots, n\}$ be an index for which $|u_i|$ is maximal

among $|u_1|, |u_2|, \dots, |u_n|$. Then by the equality $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ we have

$$|A_{ii} - \lambda| \cdot |u_i| = \left| \sum_{j \neq i} A_{ij} u_j \right| \leq \left(\sum_{j \neq i} |A_{ij}| \right) |u_i| \leq |A_{ii}| |u_i|,$$

implying that $|A_{ii} - \lambda| \leq |A_{ii}|$, which is a contradiction to the negativity of λ and the nonnegativity of A_{ii} .

(b) Since by part (a) we know that \mathbf{A} is positive semidefinite, all we need to show is that \mathbf{A} has no zero eigenvalues. Suppose in contradiction that there is a zero eigenvalue, meaning that there is a vector $\mathbf{u} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{u} = \mathbf{0}$. Then, similarly to the proof of part (a), let $i \in \{1, 2, \dots, n\}$ be an index for which $|u_i|$ is maximal among $|u_1|, |u_2|, \dots, |u_n|$, and we obtain

$$|A_{ii}| \cdot |u_i| = \left| \sum_{j \neq i} A_{ij} u_j \right| \leq \left(\sum_{j \neq i} |A_{ij}| \right) |u_i| < |A_{ii}| |u_i|,$$

which is obviously impossible, establishing the fact that \mathbf{A} is positive definite. \square

2.3 ■ Second Order Optimality Conditions

We begin by stating the necessary second order optimality condition.

Theorem 2.26 (necessary second order optimality condition). *Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. Then the following hold:*

- (a) *If \mathbf{x}^* is a local minimum point of f over U , then $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$.*
- (b) *If \mathbf{x}^* is a local maximum point of f over U , then $\nabla^2 f(\mathbf{x}^*) \preceq \mathbf{0}$.*

Proof. (a) Since \mathbf{x}^* is a local minimum point, there exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in B(\mathbf{x}^*, r)$. Let $\mathbf{d} \in \mathbb{R}^n$ be a nonzero vector. For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* \equiv \mathbf{x}^* + \alpha\mathbf{d} \in B(\mathbf{x}^*, r)$, and hence for any such α

$$f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*). \quad (2.2)$$

On the other hand, by the linear approximation theorem (Theorem 1.24), it follows that there exists a vector $\mathbf{z}_\alpha \in [\mathbf{x}^*, \mathbf{x}_\alpha^*]$ such that

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^T (\mathbf{x}_\alpha^* - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x}_\alpha^* - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z}_\alpha) (\mathbf{x}_\alpha^* - \mathbf{x}^*).$$

Since \mathbf{x}^* is a stationary point of f , and by the definition of \mathbf{x}_α^* , the latter equation reduces to

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \frac{\alpha^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d}. \quad (2.3)$$

Combining (2.2) and (2.3), it follows that for any $\alpha \in (0, \frac{r}{\|\mathbf{d}\|})$ the inequality $\mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d} \geq 0$ holds. Finally, using the fact that $\mathbf{z}_\alpha \rightarrow \mathbf{x}^*$ as $\alpha \rightarrow 0^+$, and the continuity of the Hessian, we obtain that $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$. Since the latter inequality holds for any $\mathbf{d} \in \mathbb{R}^n$, the desired result is established.

(b) The proof of (b) follows immediately by employing the result of part (a) on the function $-f$. \square

The latter result is a necessary condition for local optimality. The next theorem states a sufficient condition for strict local optimality.

Theorem 2.27 (sufficient second order optimality condition). *Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. Then the following hold:*

(a) *If $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$, then \mathbf{x}^* is a strict local minimum point of f over U .*

(b) *If $\nabla^2 f(\mathbf{x}^*) \prec \mathbf{0}$, then \mathbf{x}^* is a strict local maximum point of f over U .*

Proof. We will prove part (a). Part (b) follows by considering the function $-f$. Suppose then that \mathbf{x}^* is a stationary point satisfying $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$. Since the Hessian is continuous, it follows that there exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$ for any $\mathbf{x} \in B(\mathbf{x}^*, r)$. By the linear approximation theorem (Theorem 1.24), it follows that for any $\mathbf{x} \in B(\mathbf{x}^*, r)$, there exists a vector $\mathbf{z}_x \in [\mathbf{x}^*, \mathbf{x}]$ (and hence $\mathbf{z}_x \in B(\mathbf{x}^*, r)$) for which

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z}_x)(\mathbf{x} - \mathbf{x}^*). \quad (2.4)$$

Since $\nabla^2 f(\mathbf{z}_x) \succ \mathbf{0}$, it follows by (2.4) that for any $\mathbf{x} \in B(\mathbf{x}^*, r)$ such that $\mathbf{x} \neq \mathbf{x}^*$, the inequality $f(\mathbf{x}) > f(\mathbf{x}^*)$ holds, implying that \mathbf{x}^* is a strict local minimum point of f over U . \square

Note that the sufficient condition implies the stronger property of *strict* local optimality. However, positive definiteness of the Hessian matrix is not a necessary condition for strict local optimality. For example, the one-dimensional function $f(x) = x^4$ over \mathbb{R} has a strict local minimum at $x = 0$, but $f''(0)$ is not positive. Another important concept is that of a *saddle point*.

Definition 2.28 (saddle point). *Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is continuously differentiable over U . A stationary point \mathbf{x}^* is called a **saddle point** of f over U if it is neither a local minimum point nor a local maximum point of f over U .*

A sufficient condition for a stationary point to be a saddle point in terms of the properties of the Hessian is given in the next result.

Theorem 2.29 (sufficient condition for a saddle point). *Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that \mathbf{x}^* is a stationary point. If $\nabla^2 f(\mathbf{x}^*)$ is an indefinite matrix, then \mathbf{x}^* is a saddle point of f over U .*

Proof. Since $\nabla^2 f(\mathbf{x}^*)$ is indefinite, it has at least one positive eigenvalue $\lambda > 0$, corresponding to a normalized eigenvector which we will denote by \mathbf{v} . Since U is an open set, it follows that there exists a positive real $r > 0$ such that $\mathbf{x}^* + \alpha \mathbf{v} \in U$ for any $\alpha \in (0, r)$. By the quadratic approximation theorem (Theorem 1.25) and recalling that $\nabla f(\mathbf{x}^*) = \mathbf{0}$,

we have that there exists a function $g : \mathbb{R}_{++} \rightarrow \mathbb{R}$ satisfying

$$\frac{g(t)}{t} \rightarrow 0 \text{ as } t \rightarrow 0, \quad (2.5)$$

such that for any $\alpha \in (0, r)$

$$\begin{aligned} f(\mathbf{x}^* + \alpha \mathbf{v}) &= f(\mathbf{x}^*) + \frac{\alpha^2}{2} \mathbf{v}^T \nabla^2 f(\mathbf{x}^*) \mathbf{v} + g(\alpha^2 \|\mathbf{v}\|^2) \\ &= f(\mathbf{x}^*) + \frac{\lambda \alpha^2}{2} \|\mathbf{v}\|^2 + g(\|\mathbf{v}\|^2 \alpha^2). \end{aligned}$$

Since $\|\mathbf{v}\| = 1$, the latter can be rewritten as

$$f(\mathbf{x}^* + \alpha \mathbf{v}) = f(\mathbf{x}^*) + \frac{\lambda \alpha^2}{2} + g(\alpha^2).$$

By (2.5) it follows that there exists an $\varepsilon_1 \in (0, r)$ such that $g(\alpha^2) > -\frac{\lambda}{2} \alpha^2$ for all $\alpha \in (0, \varepsilon_1)$, and hence $f(\mathbf{x}^* + \alpha \mathbf{v}) > f(\mathbf{x}^*)$ for all $\alpha \in (0, \varepsilon_1)$. This shows that \mathbf{x}^* cannot be a local maximum point of f over U . A similar argument—exploiting an eigenvector of $\nabla^2 f(\mathbf{x}^*)$ corresponding to a negative eigenvalue—shows that \mathbf{x}^* cannot be a local minimum point of f over U , establishing the desired result that \mathbf{x}^* is a saddle point. \square

Another important issue is the one of deciding on whether a function actually has a global minimizer or maximizer. This is the issue of *attainment* or *existence*. A very well known result is due to Weierstrass, stating that a continuous function attains its minimum and maximum over a compact set.

Theorem 2.30 (Weierstrass theorem). *Let f be a continuous function defined over a nonempty and compact set $C \subseteq \mathbb{R}^n$. Then there exists a global minimum point of f over C and a global maximum point of f over C .*

When the underlying set is not compact, the Weierstrass theorem does not guarantee the attainment of the solution, but certain properties of the function f can imply attainment of the solution even in the noncompact setting. One example of such a property is *coerciveness*.

Definition 2.31 (coerciveness). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function defined over \mathbb{R}^n . The function f is called **coercive** if*

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty.$$

The important property of coercive functions that will be frequently used in this book is that a coercive function always attains a global minimum point on any closed set.

Theorem 2.32 (attainment under coerciveness). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f has a global minimum point over S .*

Proof. Let $\mathbf{x}_0 \in S$ be an arbitrary point in S . Since the function is coercive, it follows that there exists an $M > 0$ such that

$$f(\mathbf{x}) > f(\mathbf{x}_0) \text{ for any } \mathbf{x} \text{ such that } \|\mathbf{x}\| > M. \quad (2.6)$$

Since any global minimizer \mathbf{x}^* of f over S satisfies $f(\mathbf{x}^*) \leq f(\mathbf{x}_0)$, it follows from (2.6) that the set of global minimizers of f over S is the same as the set of global minimizers of f over $S \cap B[\mathbf{0}, M]$. The set $S \cap B[\mathbf{0}, M]$ is compact and nonempty, and thus by the Weierstrass theorem, there exists a global minimizer of f over $S \cap B[\mathbf{0}, M]$ and hence also over S . \square

Example 2.33. Consider the function $f(x_1, x_2) = x_1^2 + x_2^2$ over the set

$$C = \{(x_1, x_2) : x_1 + x_2 \leq -1\}.$$

The set C is not bounded, and thus the Weierstrass theorem does not guarantee the existence of a global minimizer of f over C , but since f is coercive and C is closed, Theorem 2.32 does guarantee the existence of such a global minimizer. It is also not a difficult task to find the global minimum point in this example. There are two options: In one option the global minimum point is in the interior of C , and in that case by Theorem 2.6 $\nabla f(\mathbf{x}) = \mathbf{0}$, meaning that $\mathbf{x} = \mathbf{0}$, which is impossible since the zeros vector is not in C . The other option is that the global minimum point is attained at the boundary of C given by $\text{bd}(C) = \{(x_1, x_2) : x_1 + x_2 = -1\}$. We can then substitute $x_1 = -x_2 - 1$ into the objective function and recast the problem as the one-dimensional optimization problem of minimizing $g(x_2) = (-1 - x_2)^2 + x_2^2$ over \mathbb{R} . Since $g'(x_2) = 2(1 + x_2) + 2x_2$, it follows that g' has a single root, which is $x_2 = -0.5$, and hence $x_1 = -0.5$. Since $(x_1, x_2) = (-0.5, -0.5)$ is the only candidate for a global minimum point, and since there *must* be at least one global minimizer, it follows that $(x_1, x_2) = (-0.5, -0.5)$ is the global minimum point of f over C . \blacksquare

Example 2.34. Consider the function

$$f(x_1, x_2) = 2x_1^3 + 3x_2^2 + 3x_1^2x_2 - 24x_2$$

over \mathbb{R}^2 . Let us find all the stationary points of f over \mathbb{R}^2 and classify them. First,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 6x_1^2 + 6x_1x_2 \\ 6x_2 + 3x_1^2 - 24 \end{pmatrix}.$$

Therefore, the stationary points are those satisfying

$$\begin{aligned} 6x_1^2 + 6x_1x_2 &= 0, \\ 6x_2 + 3x_1^2 - 24 &= 0. \end{aligned}$$

The first equation is the same as $6x_1(x_1 + x_2) = 0$, meaning that either $x_1 = 0$ or $x_1 + x_2 = 0$. If $x_1 = 0$, then by the second equation $x_2 = 4$. If $x_1 + x_2 = 0$, then substituting $x_1 = -x_2$ in the second equation yields the equation $3x_1^2 - 6x_1 - 24 = 0$ whose solutions are $x_1 = 4, -2$. Overall, the stationary points of the function f are $(0, 4), (4, -4), (-2, 2)$. The Hessian of f is given by

$$\nabla^2 f(x_1, x_2) = 6 \begin{pmatrix} 2x_1 + x_2 & x_1 \\ x_1 & 1 \end{pmatrix}.$$

For the stationary point $(0, 4)$ we have

$$\nabla^2 f(0, 4) = 6 \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix},$$

which is a positive definite matrix as a diagonal matrix with positive components in the diagonal (see Lemma 2.19). Therefore, $(0, 4)$ is a strict local minimum point. It is not a global minimum point since such a point does not exist as the function f is not bounded below:

$$f(x_1, 0) = 2x_1^3 \rightarrow -\infty \text{ as } x_1 \rightarrow -\infty.$$

The Hessian of f at $(4, -4)$ is

$$\nabla^2 f(4, -4) = 6 \begin{pmatrix} 4 & 4 \\ 4 & 1 \end{pmatrix}.$$

Since $\det(\nabla^2 f(4, -4)) = -6^2 \cdot 12 < 0$, it follows that the Hessian at $(4, -4)$ is indefinite. Indeed, the determinant is the product of its two eigenvalues – one must be positive and the other negative. Therefore, by Theorem 2.29, $(4, -4)$ is a saddle point. The Hessian at $(-2, 2)$ is

$$\nabla^2 f(-2, 2) = 6 \begin{pmatrix} -2 & -2 \\ -2 & 1 \end{pmatrix},$$

which is indefinite by the fact that it has both positive and negative elements on its diagonal. Therefore, $(-2, 2)$ is a saddle point. To summarize, $(0, 4)$ is a strict local minimum point of f , and $(4, -4), (-2, 2)$ are saddle points. ■

Example 2.35. Let

$$f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2.$$

The gradient of f is given by

$$\nabla f(\mathbf{x}) = 4 \begin{pmatrix} (x_1^2 + x_2^2 - 1)x_1 \\ (x_1^2 + x_2^2 - 1)x_2 + (x_2^2 - 1)x_2 \end{pmatrix}.$$

The stationary points are those satisfying

$$(x_1^2 + x_2^2 - 1)x_1 = 0, \quad (2.7)$$

$$(x_1^2 + x_2^2 - 1)x_2 + (x_2^2 - 1)x_2 = 0. \quad (2.8)$$

By equation (2.7), there are two cases: either $x_1 = 0$, and then by equation (2.8) x_2 is equal to one of the values $0, 1, -1$; the second option is that $x_1^2 + x_2^2 = 1$, and then by equation (2.8), we have that $x_2 = 0, \pm 1$ and hence x_1 is $\pm 1, 0$ respectively. Overall, there are 5 stationary points: $(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1)$. The Hessian of the function is

$$\nabla^2 f(\mathbf{x}) = 4 \begin{pmatrix} 3x_1^2 + x_2^2 - 1 & 2x_1x_2 \\ 2x_1x_2 & x_1^2 + 6x_2^2 - 2 \end{pmatrix}.$$

Since

$$\nabla^2 f(0, 0) = 4 \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} < 0,$$

it follows that $(0, 0)$ is a strict local maximum point. By the fact that $f(x_1, 0) = (x_1^2 - 1)^2 + 1 \rightarrow \infty$ as $x_1 \rightarrow \infty$, the function is not bounded above and thus $(0, 0)$ is not a global maximum point. Also,

$$\nabla^2 f(1, 0) = \nabla^2 f(-1, 0) = 4 \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix},$$

which is an indefinite matrix and hence $(1, 0)$ and $(-1, 0)$ are saddle points. Finally,

$$\nabla^2 f(0, 1) = \nabla^2 f(0, -1) = 4 \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \succeq 0.$$

The fact that the Hessian matrices of f at $(0, 1)$ and $(0, -1)$ are positive semidefinite is not enough in order to conclude that these are local minimum points; they might be saddle points. However, in this case it is not difficult to see that $(0, 1)$ and $(0, -1)$ are in fact *global minimum points* since $f(0, 1) = f(0, -1) = 0$, and the function is lower bounded by zero. Note that since there are two global minimum points, they are *nonstrict* global minima, but they actually are *strict* local minimum points since each has a neighborhood in which it is the unique minimizer. The contour and surface plots of the function are plotted in Figure 2.3. ■

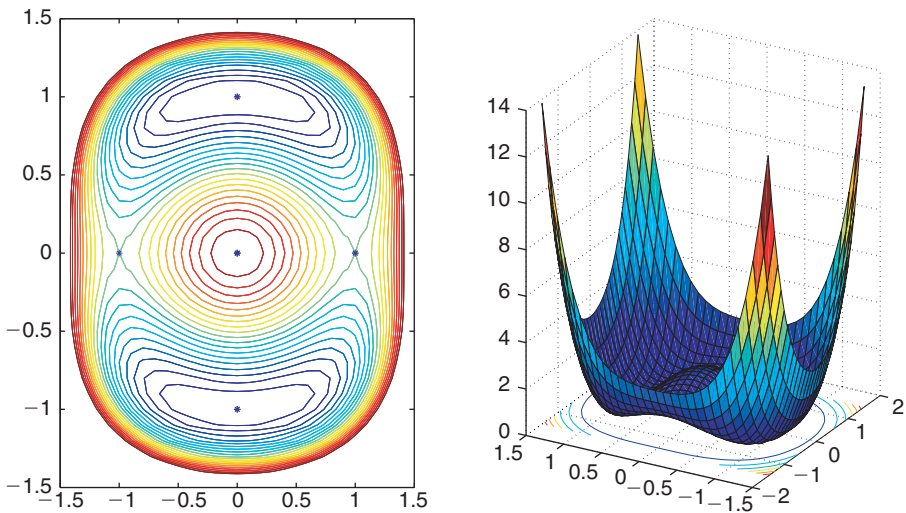


Figure 2.3. Contour and surface plots of $f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$. The five stationary points $(0, 0), (0, 1), (0, -1), (1, 0), (-1, 0)$ are denoted by asterisks. The points $(0, -1), (0, 1)$ are strict local minimum points as well as global minimum points, $(0, 0)$ is a local maximum point, and $(-1, 0), (1, 0)$ are saddle points.

Example 2.36. Returning to Example 2.3, we will now investigate the stationary points of the function

$$f(x, y) = \frac{x + y}{x^2 + y^2 + 1}.$$

The gradient of the function is

$$\nabla f(x, y) = \frac{1}{(x^2 + y^2 + 1)^2} \begin{pmatrix} (x^2 + y^2 + 1) - 2(x + y)x \\ (x^2 + y^2 + 1) - 2(x + y)y \end{pmatrix}.$$

Therefore, the stationary points of the function are those satisfying

$$\begin{aligned} -x^2 - 2xy + y^2 &= -1, \\ x^2 - 2xy - y^2 &= -1. \end{aligned}$$

Adding the two equations yields the equation $xy = \frac{1}{2}$. Subtracting the two equations yields $x^2 = y^2$, which, along with the fact that x and y have the same sign, implies that $x = y$. Therefore, we obtain that

$$x^2 = \frac{1}{2}$$

whose solutions are $x = \pm \frac{1}{\sqrt{2}}$. Thus, the function has two stationary points which are $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. We will now prove that the statement given in Example 2.3 is correct: $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is the global maximum point of f over \mathbb{R}^2 and $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ is the global minimum point of f over \mathbb{R}^2 . Indeed, note that $f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}}$. In addition, for any $(x, y)^T \in \mathbb{R}^2$,

$$f(x, y) = \frac{x + y}{x^2 + y^2 + 1} \leq \sqrt{2} \frac{\sqrt{x^2 + y^2}}{x^2 + y^2 + 1} \leq \sqrt{2} \max_{t \geq 0} \frac{t}{t^2 + 1},$$

where the first inequality follows from the Cauchy-Schwarz inequality. Since for any $t \geq 0$, the inequality $t^2 + 1 \geq 2t$ holds, we have that $f(x, y) \leq \frac{1}{\sqrt{2}}$ for any $(x, y)^T \in \mathbb{R}^2$. Therefore, $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ attains the maximal value of the function and is therefore the global maximum point of f over \mathbb{R}^2 . A similar argument shows that $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ is the global minimum point of f over \mathbb{R}^2 . ■

Example 2.37. Let

$$f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4.$$

The gradient of the function is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} -4x_1 + x_2^2 + 16x_1^3 \\ 2x_1x_2 \end{pmatrix},$$

and the stationary points are those satisfying the system of equations

$$\begin{aligned} -4x_1 + x_2^2 + 16x_1^3 &= 0, \\ 2x_1x_2 &= 0. \end{aligned}$$

By the second equation, we obtain that either $x_1 = 0$ or $x_2 = 0$ (or both). If $x_1 = 0$, then by the first equation, we have $x_2 = 0$. If $x_2 = 0$, then by the first equation $-4x_1 + 16x_1^3 = 0$ and thus $x_1(-1 + 4x_1^2) = 0$, so that x_1 is one of the three values $0, 0.5, -0.5$. The function therefore has three stationary points: $(0, 0), (0.5, 0), (-0.5, 0)$. The Hessian of f is

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -4 + 48x_1^2 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}.$$

Since

$$\nabla^2 f(0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix} \succ 0,$$

it follows that $(0.5, 0)$ is a strict local minimum point of f . It is not a *global* minimum point since the function is not bounded below: $f(-1, x_2) = 2 - x_2^2 \rightarrow -\infty$ as $x_2 \rightarrow \infty$. As for the stationary point $(-0.5, 0)$,

$$\nabla^2 f(-0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & -1 \end{pmatrix},$$

and hence since the Hessian is indefinite, $(-0.5, 0)$ is a saddle point of f . Finally, the Hessian of f at the stationary point $(0, 0)$ is

$$\nabla^2 f(0, 0) = \begin{pmatrix} -4 & 0 \\ 0 & 0 \end{pmatrix}.$$

The fact that the Hessian is negative semidefinite implies that $(0, 0)$ is either a local maximum point or a saddle point of f . Note that

$$f(\alpha^4, \alpha) = -2\alpha^8 + \alpha^6 + 4\alpha^{16} = \alpha^6(-2\alpha^2 + 1 + 4\alpha^{10}).$$

It is easy to see that for a small enough $\alpha > 0$, the above expression is positive. Similarly,

$$f(-\alpha^4, \alpha) = -2\alpha^8 - \alpha^6 + 4\alpha^{16} = \alpha^6(-2\alpha^2 - 1 + 4\alpha^{10}),$$

and the above expression is negative for small enough $\alpha > 0$. This means that $(0, 0)$ is a saddle point since at any one of its neighborhoods, we can find points with smaller values than $f(0, 0) = 0$ and points with larger values than $f(0, 0) = 0$. The surface and contour plots of the function are given in Figure 2.4. ■

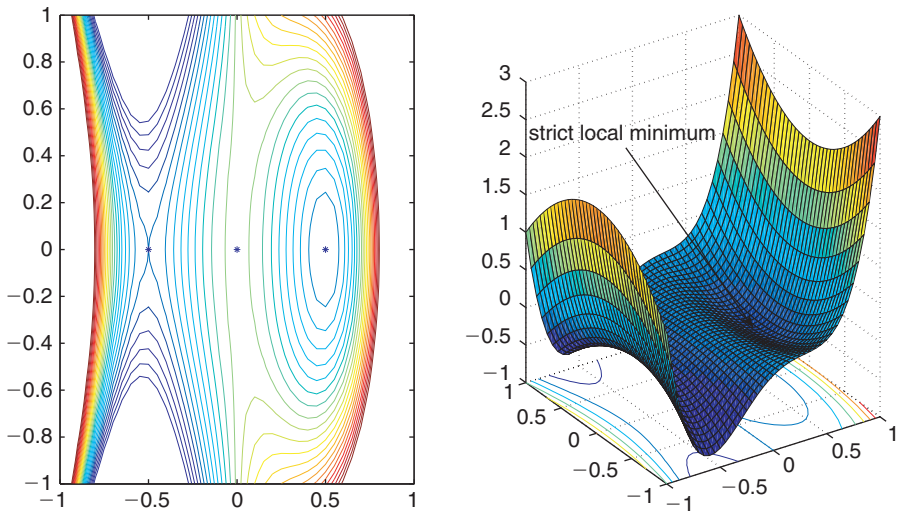


Figure 2.4. Contour and surface plots of $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4$. The three stationary point $(0, 0)$, $(0.5, 0)$, $(-0.5, 0)$ are denoted by asterisks. The point $(0.5, 0)$ is a strict local minimum, while $(0, 0)$ and $(-0.5, 0)$ are saddle points.

2.4 ■ Global Optimality Conditions

The conditions described in the last section can only guarantee—at best—local optimality of stationary points since they exploit only local information: the values of the gradient and the Hessian at a given point. Conditions that ensure global optimality of points must use global information. For example, when the Hessian of the function is always positive semidefinite, all the stationary points are also global minimum points. Later on, we will refer to this property as convexity.

Theorem 2.38. Let f be a twice continuously differentiable function defined over \mathbb{R}^n . Suppose that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f . Then \mathbf{x}^* is a global minimum point of f .

Proof. By the linear approximation theorem (Theorem 1.24), it follows that for any $\mathbf{x} \in \mathbb{R}^n$, there exists a vector $\mathbf{z}_x \in [\mathbf{x}^*, \mathbf{x}]$ for which

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z}_x)(\mathbf{x} - \mathbf{x}^*).$$

Since $\nabla^2 f(\mathbf{z}_x) \succeq \mathbf{0}$, we have that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, establishing the fact that \mathbf{x}^* is a global minimum point of f . \square

Example 2.39. Let

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + (x_1^2 + x_2^2 + x_3^2)^2.$$

Then

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + x_2 + x_3 + 4x_1(x_1^2 + x_2^2 + x_3^2) \\ 2x_2 + x_1 + x_3 + 4x_2(x_1^2 + x_2^2 + x_3^2) \\ 2x_3 + x_1 + x_2 + 4x_3(x_1^2 + x_2^2 + x_3^2) \end{pmatrix}.$$

Obviously, $\mathbf{x} = \mathbf{0}$ is a stationary point. We will show that it is a global minimum point. The Hessian of f is

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_1^2 & 1 + 8x_1x_2 & 1 + 8x_1x_3 \\ 1 + 8x_1x_2 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_2^2 & 1 + 8x_2x_3 \\ 1 + 8x_1x_3 & 1 + 8x_2x_3 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_3^2 \end{pmatrix}.$$

The Hessian is positive semidefinite since it can be written as the sum

$$\nabla^2 f(\mathbf{x}) = \mathbf{A} + \mathbf{B}(\mathbf{x}) + \mathbf{C}(\mathbf{x}),$$

where

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \quad \mathbf{B}(\mathbf{x}) = 4(x_1^2 + x_2^2 + x_3^2)\mathbf{I}_3, \quad \mathbf{C}(\mathbf{x}) = \begin{pmatrix} 8x_1^2 & 8x_1x_2 & 8x_1x_3 \\ 8x_1x_3 & 8x_2^2 & 8x_2x_3 \\ 8x_1x_3 & 8x_2x_3 & 8x_3^2 \end{pmatrix}.$$

The above three matrices are positive semidefinite for any $\mathbf{x} \in \mathbb{R}^3$; indeed, $\mathbf{A} \succeq \mathbf{0}$ since it is diagonally dominant with positive diagonal elements. The matrix $\mathbf{B}(\mathbf{x})$, as a nonnegative multiplier of the identity matrix, is positive semidefinite, and finally,

$$\mathbf{C}(\mathbf{x}) = 8\mathbf{x}\mathbf{x}^T,$$

and hence $\mathbf{C}(\mathbf{x})$ is positive semidefinite as a positive multiply of the matrix $\mathbf{x}\mathbf{x}^T$, which is positive semidefinite (see Exercise 2.6). To summarize, $\nabla^2 f(\mathbf{x})$ is positive semidefinite as the sum of three positive semidefinite matrices (simple extension of Exercise 2.4), and hence, by Theorem 2.38, $\mathbf{x} = \mathbf{0}$ is a global minimum point of f over \mathbb{R}^3 . \blacksquare

2.5 ■ Quadratic Functions

Quadratic functions are an important class of functions that are useful in the modeling of many optimization problems. We will now define and derive some of the basic results related to this important class of functions.

Definition 2.40. A quadratic function over \mathbb{R}^n is a function of the form

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c, \quad (2.9)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

We will frequently refer to the matrix \mathbf{A} in (2.9) as the matrix *associated* with the quadratic function f . The gradient and Hessian of a quadratic function have simple analytic formulas:

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b}, \quad (2.10)$$

$$\nabla^2 f(\mathbf{x}) = 2\mathbf{A}. \quad (2.11)$$

By the above formulas we can deduce several important properties of quadratic functions, which are associated with their stationary points.

Lemma 2.41. Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then

- (a) \mathbf{x} is a stationary point of f if and only if $\mathbf{A}\mathbf{x} = -\mathbf{b}$,
- (b) if $\mathbf{A} \geq \mathbf{0}$, then \mathbf{x} is a global minimum point of f if and only if $\mathbf{A}\mathbf{x} = -\mathbf{b}$,
- (c) if $\mathbf{A} \succ \mathbf{0}$, then $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is a strict global minimum point of f .

Proof. (a) The proof of (a) follows immediately from the formula of the gradient of f (equation (2.10)).

(b) Since $\nabla^2 f(\mathbf{x}) = 2\mathbf{A} \geq \mathbf{0}$, it follows by Theorem 2.38 that the global minimum points are exactly the stationary points, which combined with part (a) implies the result.

(c) When $\mathbf{A} \succ \mathbf{0}$, the vector $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is the unique solution to $\mathbf{A}\mathbf{x} = -\mathbf{b}$, and hence by parts (a) and (b), it is the unique global minimum point of f . \square

We note that when $\mathbf{A} \succ \mathbf{0}$, the global minimizer of f is $\mathbf{x}^* = -\mathbf{A}^{-1}\mathbf{b}$, and consequently the minimal value of the function is

$$\begin{aligned} f(\mathbf{x}^*) &= (\mathbf{x}^*)^T \mathbf{A} \mathbf{x}^* + 2\mathbf{b}^T \mathbf{x}^* + c \\ &= (-\mathbf{A}^{-1}\mathbf{b})^T \mathbf{A} (-\mathbf{A}^{-1}\mathbf{b}) - 2\mathbf{b}^T \mathbf{A}^{-1}\mathbf{b} + c \\ &= c - \mathbf{b}^T \mathbf{A}^{-1}\mathbf{b}. \end{aligned}$$

Another useful property of quadratic functions is that they are coercive if and only if the associated matrix \mathbf{A} is positive definite.

Lemma 2.42 (coerciveness of quadratic functions). Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, 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 coercive if and only if $\mathbf{A} \succ \mathbf{0}$.

Proof. If $\mathbf{A} \succ \mathbf{0}$, then by Lemma 1.11, $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq \alpha \|\mathbf{x}\|^2$ for $\alpha = \lambda_{\min}(\mathbf{A}) > 0$. We can thus write

$$f(\mathbf{x}) \geq \alpha \|\mathbf{x}\|^2 + 2\mathbf{b}^T \mathbf{x} + c \geq \alpha \|\mathbf{x}\|^2 - 2\|\mathbf{b}\| \cdot \|\mathbf{x}\| + c = \alpha \|\mathbf{x}\| \left(\|\mathbf{x}\| - 2\frac{\|\mathbf{b}\|}{\alpha} \right) + c,$$

where we also used the Cauchy–Schwarz inequality. Since obviously $\alpha \|\mathbf{x}\|(\|\mathbf{x}\| - 2\frac{\|\mathbf{b}\|}{\alpha}) + c \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$, it follows that $f(\mathbf{x}) \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$, establishing the coerciveness of f . Now assume that f is coercive. We need to prove that \mathbf{A} is positive definite, or in other words that all its eigenvalues are positive. We begin by showing that there does not exist a negative eigenvalue. Suppose in contradiction that there exists such a negative eigenvalue; that is, there exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ and $\lambda < 0$ such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Then

$$f(\alpha\mathbf{v}) = \lambda \|\mathbf{v}\|^2 \alpha^2 + 2(\mathbf{b}^T \mathbf{v})\alpha + c \rightarrow -\infty$$

as α tends to ∞ , thus contradicting the assumption that f is coercive. We thus conclude that all the eigenvalues of \mathbf{A} are nonnegative. We will show that 0 cannot be an eigenvalue of \mathbf{A} . By contradiction assume that there exists $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \mathbf{0}$. Then for any $\alpha \in \mathbb{R}$

$$f(\alpha\mathbf{v}) = 2(\mathbf{b}^T \mathbf{v})\alpha + c.$$

Then if $\mathbf{b}^T \mathbf{v} = 0$, we have $f(\alpha\mathbf{v}) \rightarrow c$ as $\alpha \rightarrow \infty$. If $\mathbf{b}^T \mathbf{v} > 0$, then $f(\alpha\mathbf{v}) \rightarrow -\infty$ as $\alpha \rightarrow -\infty$, and if $\mathbf{b}^T \mathbf{v} < 0$, then $f(\alpha\mathbf{v}) \rightarrow -\infty$ as $\alpha \rightarrow \infty$, contradicting the coerciveness of the function. We have thus proven that \mathbf{A} is positive definite. \square

The last result describes an important characterization of the property that a quadratic function is nonnegative over the entire space. It is a generalization of the property that $\mathbf{A} \succeq \mathbf{0}$ if and only if $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for any $\mathbf{x} \in \mathbb{R}^n$.

Theorem 2.43 (characterization of the nonnegativity of quadratic functions). *Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then the following two claims are equivalent:*

- (a) $f(\mathbf{x}) \equiv \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.
- (b) $\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix} \succeq \mathbf{0}$.

Proof. Suppose that (b) holds. Then in particular for any $\mathbf{x} \in \mathbb{R}^n$ the inequality

$$\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}^T \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \geq 0$$

holds, which is the same as the inequality $\mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c \geq 0$, proving the validity of (a). Now, assume that (a) holds. We begin by showing that $\mathbf{A} \succeq \mathbf{0}$. Suppose in contradiction that \mathbf{A} is not positive semidefinite. Then there exists an eigenvector \mathbf{v} corresponding to a negative eigenvalue $\lambda < 0$ of \mathbf{A} :

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

Thus, for any $\alpha \in \mathbb{R}$

$$f(\alpha\mathbf{v}) = \lambda \|\mathbf{v}\|^2 \alpha^2 + 2(\mathbf{b}^T \mathbf{v})\alpha + c \rightarrow -\infty$$

as $\alpha \rightarrow -\infty$, contradicting the nonnegativity of f . Our objective is to prove (b); that is, we want to show that for any $\mathbf{y} \in \mathbb{R}^n$ and $t \in \mathbb{R}$

$$\begin{pmatrix} \mathbf{y} \\ t \end{pmatrix}^T \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ t \end{pmatrix} \geq 0,$$

which is equivalent to

$$\mathbf{y}^T \mathbf{A} \mathbf{y} + 2t \mathbf{b}^T \mathbf{y} + ct^2 \geq 0. \quad (2.12)$$

To show the validity of (2.12) for any $\mathbf{y} \in \mathbb{R}^n$ and $t \in \mathbb{R}$, we consider two cases. If $t = 0$, then (2.12) reads as $\mathbf{y}^T \mathbf{A} \mathbf{y} \geq 0$, which is a valid inequality since we have shown that $\mathbf{A} \succeq \mathbf{0}$. The second case is when $t \neq 0$. To show that (2.12) holds in this case, note that (2.12) is the same as the inequality

$$t^2 f\left(\frac{\mathbf{y}}{t}\right) = t^2 \left[\left(\frac{\mathbf{y}}{t}\right)^T \mathbf{A} \left(\frac{\mathbf{y}}{t}\right) + 2\mathbf{b}^T \left(\frac{\mathbf{y}}{t}\right) + c \right] \geq 0,$$

which holds true by the nonnegativity of f . \square

Exercises

- 2.1. Find the global minimum and maximum points of the function $f(x, y) = x^2 + y^2 + 2x - 3y$ over the unit ball $S = B[\mathbf{0}, 1] = \{(x, y) : x^2 + y^2 \leq 1\}$.
- 2.2. Let $\mathbf{a} \in \mathbb{R}^n$ be a nonzero vector. Show that the maximum of $\mathbf{a}^T \mathbf{x}$ over $B[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$ is attained at $\mathbf{x}^* = \frac{\mathbf{a}}{\|\mathbf{a}\|}$ and that the maximal value is $\|\mathbf{a}\|$.
- 2.3. Find the global minimum and maximum points of the function $f(x, y) = 2x - 3y$ over the set $S = \{(x, y) : 2x^2 + 5y^2 \leq 1\}$.
- 2.4. Show that if \mathbf{A}, \mathbf{B} are $n \times n$ positive semidefinite matrices, then their sum $\mathbf{A} + \mathbf{B}$ is also positive semidefinite.
- 2.5. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$ be two symmetric matrices. Prove that the following two claims are equivalent:
 - (i) \mathbf{A} and \mathbf{B} are positive semidefinite.
 - (ii) $\begin{pmatrix} \mathbf{A} & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & \mathbf{B} \end{pmatrix}$ is positive semidefinite.
- 2.6. Let $\mathbf{B} \in \mathbb{R}^{n \times k}$ and let $\mathbf{A} = \mathbf{B}\mathbf{B}^T$.
 - (i) Prove \mathbf{A} is positive semidefinite.
 - (ii) Prove that \mathbf{A} is positive definite if and only if \mathbf{B} has a full row rank.
- 2.7.
 - (i) Let \mathbf{A} be an $n \times n$ symmetric matrix. Show that \mathbf{A} is positive semidefinite if and only if there exists a matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ such that $\mathbf{A} = \mathbf{B}\mathbf{B}^T$.
 - (ii) Let $\mathbf{x} \in \mathbb{R}^n$ and let \mathbf{A} be defined as

$$A_{ij} = x_i x_j, \quad i, j = 1, 2, \dots, n.$$

Show that \mathbf{A} is positive semidefinite and that it is *not* a positive definite matrix when $n > 1$.

2.8. Let $\mathbf{Q} \in \mathbb{R}^{n \times n}$ be a positive definite matrix. Show that the “Q-norm” defined by

$$\|\mathbf{x}\|_{\mathbf{Q}} = \sqrt{\mathbf{x}^T \mathbf{Q} \mathbf{x}}$$

is indeed a norm.

2.9. Let \mathbf{A} be an $n \times n$ positive semidefinite matrix.

(i) Show that for any $i \neq j$

$$A_{ii}A_{jj} \geq A_{ij}^2.$$

(ii) Show that if for some $i \in \{1, 2, \dots, n\}$ $A_{ii} = 0$, then the i th row of \mathbf{A} consists of zeros.

2.10. Let \mathbf{A}^α be the $n \times n$ matrix ($n > 1$) defined by

$$A_{ij}^\alpha = \begin{cases} \alpha, & i = j, \\ 1, & i \neq j. \end{cases}$$

Show that \mathbf{A}^α is positive semidefinite if and only if $\alpha \geq 1$.

2.11. Let $\mathbf{d} \in \Delta_n$ (Δ_n being the unit-simplex). Show that the $n \times n$ matrix \mathbf{A} defined by

$$A_{ij} = \begin{cases} d_i - d_i^2, & i = j, \\ -d_i d_j, & i \neq j, \end{cases}$$

is positive semidefinite.

2.12. Prove that a 2×2 matrix \mathbf{A} is negative semidefinite if and only if $\text{Tr}(\mathbf{A}) \leq 0$ and $\det(\mathbf{A}) \geq 0$.

2.13. For each of the following matrices determine whether they are positive/negative semidefinite/definite or indefinite:

$$(i) \mathbf{A} = \begin{pmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix}.$$

$$(ii) \mathbf{B} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 3 & 3 \\ 2 & 3 & 3 \end{pmatrix}.$$

$$(iii) \mathbf{C} = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{pmatrix}.$$

$$(iv) \mathbf{D} = \begin{pmatrix} -5 & 1 & 1 \\ 1 & -7 & 1 \\ 1 & 1 & -5 \end{pmatrix}.$$

2.14. (Schur complement lemma) Let

$$\mathbf{D} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix},$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$. Suppose that $\mathbf{A} \succ \mathbf{0}$. Prove that $\mathbf{D} \succeq \mathbf{0}$ if and only if $c - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} \geq 0$.

2.15. For each of the following functions, determine whether it is coercive or not:

(i) $f(x_1, x_2) = x_1^4 + x_2^4$.

(ii) $f(x_1, x_2) = e^{x_1^2} + e^{x_2^2} - x_1^{200} - x_2^{200}$.

(iii) $f(x_1, x_2) = 2x_1^2 - 8x_1x_2 + x_2^2$.

(iv) $f(x_1, x_2) = 4x_1^2 + 2x_1x_2 + 2x_2^2$.

(v) $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$.

(vi) $f(x_1, x_2) = x_1^2 - 2x_1x_2^2 + x_2^4$.

(vii) $f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|+1}$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive definite.

2.16. Find a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ which is not coercive and satisfies that for any $\alpha \in \mathbb{R}$

$$\lim_{|x_1| \rightarrow \infty} f(x_1, \alpha x_1) = \lim_{|x_2| \rightarrow \infty} f(\alpha x_2, x_2) = \infty.$$

2.17. For each of the following functions, find all the stationary points and classify them according to whether they are saddle points, strict/nonstrict local/global minimum/maximum points:

(i) $f(x_1, x_2) = (4x_1^2 - x_2)^2$.

(ii) $f(x_1, x_2, x_3) = x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2$.

(iii) $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2x_2$.

(iv) $f(x_1, x_2) = x_1^4 + 2x_1^2x_2 + x_2^2 - 4x_1^2 - 8x_1 - 8x_2$.

(v) $f(x_1, x_2) = (x_1 - 2x_2)^4 + 64x_1x_2$.

(vi) $f(x_1, x_2) = 2x_1^2 + 3x_2^2 - 2x_1x_2 + 2x_1 - 3x_2$.

(vii) $f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 - x_2$.

2.18. Let f be twice continuously differentiable function over \mathbb{R}^n . Suppose that $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Prove that a stationary point of f is necessarily a strict global minimum point.

2.19. Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Suppose that $\mathbf{A} \succeq \mathbf{0}$. Show that f is bounded below¹ over \mathbb{R}^n if and only if $\mathbf{b} \in \text{Range}(\mathbf{A}) = \{\mathbf{A}\mathbf{y} : \mathbf{y} \in \mathbb{R}^n\}$.

¹A function f is bounded below over a set C if there exists a constant α such that $f(\mathbf{x}) \geq \alpha$ for any $\mathbf{x} \in C$.