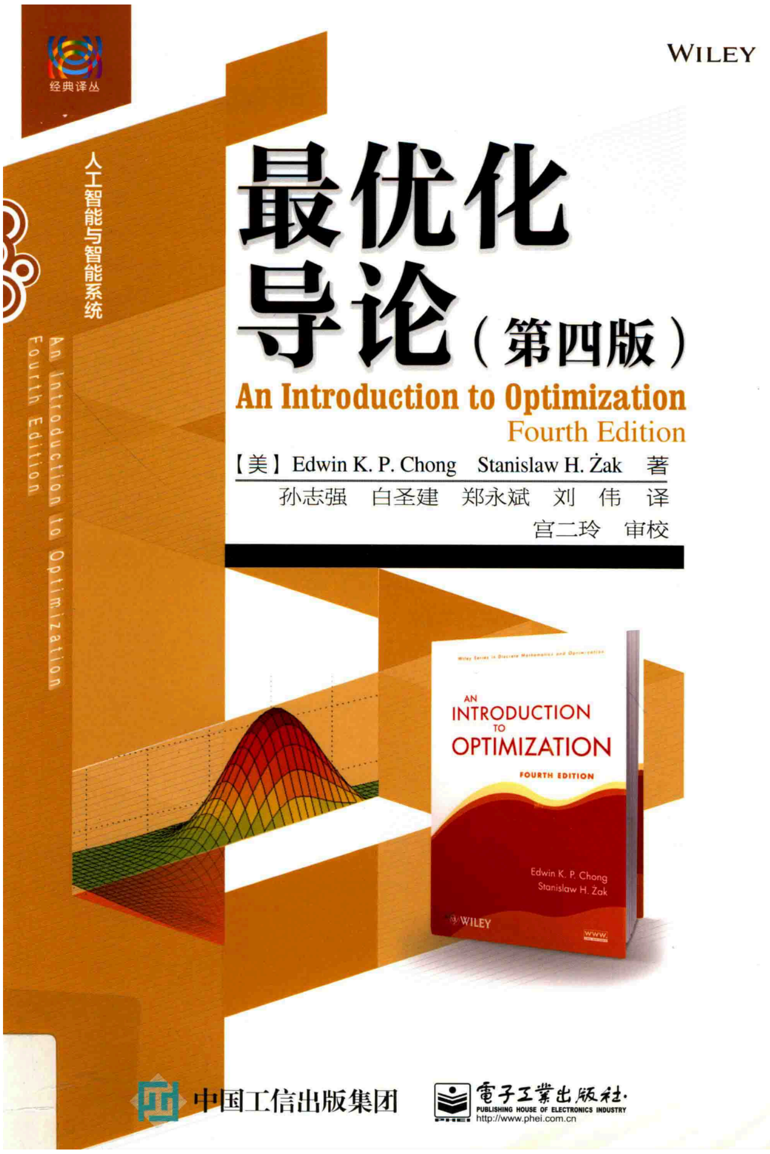
Errata by Minghong Zhao, October 2024

Chinese translation abbreviated as C
English version abbreviated as E
Error Message abbreviated as EM
Corrected version abbreviated as CV
Page abbreviated as P

Chinese Book Details: published by the Publishing House of Electronics Industry ('www.phei.com.cn')



1.

C: P32

4.2 超平面与线性簇

令 $u_1, u_2, \cdots, u_n, v \in \mathbb{R}$，其中至少存在一个 u_i 不为零。由所有满足线性方程

$$u_1x_1 + u_2x_2 + \cdots + u_nx_n = v$$

的点 $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^\top$ 组成的集合称为空间 $\mathbb{R}^n$ 的超平面。超平面可以写为

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} = v\}$$

其中

$$\boldsymbol{u} = [u_1, u_2, \cdots, u_n]^\top$$

注意，超平面不一定是 $\mathbb{R}^n$ 的子空间，因为超平面通常不包含原点。当 $n=2$ 时，超平面方程为 $u_1x_1 + u_2x_2 = v$，刚好是一个直线方程。因此，直线是 $\mathbb{R}^2$ 中的超平面。在 $\mathbb{R}^3$（三维空间）中，超平面是一些普通平面。通过对超平面进行转换，使其包含 $\mathbb{R}^n$ 的原点，就可以将其转换为 $\mathbb{R}^n$ 的子空间（见图 4.2）。由于对应子空间的维数是 $n-1$，因此称超平面是 $n-1$ 维的。

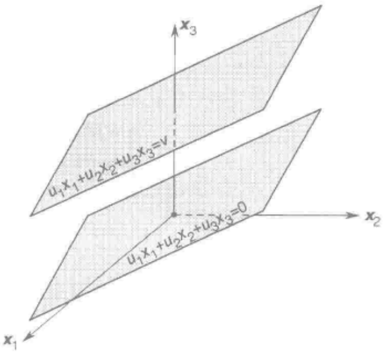


图 4.2 超平面的变换

超平面 $H = \{\boldsymbol{x}: u_1x_1 + \cdots + u_nx_n = v\}$ 将 $\mathbb{R}^n$ 空间分为两半。其中一半包含满足不等式 u_1x_1

E: P46

4.2 Hyperplanes and Linear Varieties

Let $u_1, u_2, \ldots, u_n, v \in \mathbb{R}$, where at least one of the u_i is nonzero. The set of all points $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top$ that satisfy the linear equation

$$u_1x_1 + u_2x_2 + \cdots + u_nx_n = v$$

is called a *hyperplane* of the space $\mathbb{R}^n$. We may describe the hyperplane by

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} = v\},$$

where

$$\boldsymbol{u} = [u_1, u_2, \ldots, u_n]^\top.$$

A hyperplane is not necessarily a subspace of $\mathbb{R}^n$ since, in general, it does not contain the origin. For $n = 2$, the equation of the hyperplane has the form $u_1x_1 + u_2x_2 = v$, which is the equation of a straight line. Thus, straight lines are hyperplanes in $\mathbb{R}^2$. In $\mathbb{R}^3$ (three-dimensional space), hyperplanes are ordinary planes. By translating a hyperplane so that it contains the origin of $\mathbb{R}^n$, it becomes a subspace of $\mathbb{R}^n$ (see Figure 4.2). Because the dimension of this subspace is $n - 1$, we say that the hyperplane has dimension $n - 1$.

The hyperplane $H = \{\boldsymbol{x} : u_1x_1 + \cdots + u_nx_n = v\}$ divides $\mathbb{R}^n$ into two *half-spaces.* One of these half-spaces consists of the points satisfying the inequality $u_1x_1 + u_2x_2 + \cdots + u_nx_n \geq v$, denoted

$$H_+ = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} \geq v\},$$

EM: The translator confused the words “translating” and “transforming”. So he translated it as if it is “by **transforming** a hyperplane…” instead of “by translating a hyperplane…”.

CV: 通过对超平面进行平移...

2.

C: P33

4.3 凸集

已知两点 $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ 之间的线段可表示为集合 $\{\boldsymbol{w} \in \mathbb{R}^n: \boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}, \alpha \in [0, 1]\}$。点 $\boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}(\alpha \in [0, 1])$ 称为点 $\boldsymbol{u}$ 和点 $\boldsymbol{v}$ 的凸组合。

如果对于所有 $\boldsymbol{u}, \boldsymbol{v} \in \Theta$，$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的线段都位于 Θ 内，那么称集合 $\Theta \subset \mathbb{R}^n$ 为凸集。图 4.4 给出了凸集的示例，而图 4.5 则相应地给出了非凸集的示例。注意，当且仅当对于所有 $\boldsymbol{u}, \boldsymbol{v} \in \Theta$，都有 $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{x} \in \Theta$，$\alpha \in (0, 1)$ 时，Θ 是一个凸集。

E: P48

4.3 Convex Sets

Recall that the line segment between two points $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ is the set $\{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}, \alpha \in [0,1]\}$. A point $\boldsymbol{w} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ (where $\alpha \in [0,1]$) is called a *convex combination* of the points $\boldsymbol{u}$ and $\boldsymbol{v}$.

A set $\Theta \subset \mathbb{R}^n$ is *convex* if for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$, the line segment between $\boldsymbol{u}$ and $\boldsymbol{v}$ is in Θ. Figure 4.4 gives examples of convex sets, whereas Figure 4.5 gives examples of sets that are not convex. Note that Θ is convex if and only if $\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v} \in \Theta$ for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$ and $\alpha \in (0,1)$.

EM: **Formula** Error

CV: Only have to correct the **formula**

3.

C: P35

如果不存在两个点 $\boldsymbol{u}$ 和 $\boldsymbol{v}$，使得对于某个 $\alpha \in (0,1)$ 有 $\boldsymbol{x}=\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}$，那么称点 $\boldsymbol{x}$ 是凸集 Θ 的极点。比如，对于图 4.4 中的圆而言，其边界上的任意点都是极点；对于圆右侧的集合而言，其顶点(尖角)是极点；对于最下面的线段而言，其终点也是极点。

4.4 邻域

点 $\boldsymbol{x}\in\mathbb{R}^n$ 的邻域可以表示为

$$\{\boldsymbol{y}\in\mathbb{R}^n : \|\boldsymbol{y}-\boldsymbol{x}\|<\varepsilon\}$$

其中，ε 为某个正数。邻域也可视为半径为 ε、中心为 $\boldsymbol{x}$ 的球体。

E: P50

A point $\boldsymbol{x}$ in a convex set Θ is said to be an *extreme point* of Θ if there are no two distinct points $\boldsymbol{u}$ and $\boldsymbol{v}$ in Θ such that $\boldsymbol{x}=\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}$ for some $\alpha\in(0,1)$. For example, in Figure 4.4, any point on the boundary of the disk is an extreme point, the vertex (corner) of the set on the right is an extreme point, and the endpoint of the half-line is also an extreme point.

4.4 Neighborhoods

A *neighborhood* of a point $\boldsymbol{x}\in\mathbb{R}^n$ is the set

$$\{\boldsymbol{y}\in\mathbb{R}^n : \|\boldsymbol{y}-\boldsymbol{x}\|<\varepsilon\},$$

A point $\boldsymbol{x}$ in a convex set Θ is said to be an *extreme point* of Θ if there are no two distinct points $\boldsymbol{u}$ and $\boldsymbol{v}$ in Θ such that $\boldsymbol{x}=\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}$ for some $\alpha\in(0,1)$. For example, in Figure 4.4, any point on the boundary of the disk is an extreme point, the vertex (corner) of the set on the right is an extreme point, and the endpoint of the half-line is also an extreme point.

4.4 Neighborhoods

A *neighborhood* of a point $\boldsymbol{x}\in\mathbb{R}^n$ is the set

$$\{\boldsymbol{y}\in\mathbb{R}^n : \|\boldsymbol{y}-\boldsymbol{x}\|<\varepsilon\},$$

where ε is some positive number. The neighborhood is also called a *ball* with radius ε and center $\boldsymbol{x}$.

EM: **1.** The translator did not specify u, v, and x are in the convex set Θ
2. did not specify u, v are two distinct points

CV: 如果凸集**Θ**中不存在两个不同的点u和v，使得对于某个凸集**Θ**中的点x有x=αu+(1-α)v，其中α\in(0,1)，那么称x是凸集Θ的极点。

4.

C: P35

4.4 邻域

点 $\boldsymbol{x} \in \mathbb{R}^n$ 的邻域可以表示为

$$\{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\}$$

其中，ε 为某个正数。邻域也可视为半径为 ε、中心为 $\boldsymbol{x}$ 的球体。

在平面 $\mathbb{R}^2$ 中，点 $\boldsymbol{x} = [x_1, x_2]^\top$ 的邻域包含所有以 $\boldsymbol{x}$ 为中心的圆形内部的点。在 $\mathbb{R}^3$ 中，点 $\boldsymbol{x} = [x_1, x_2, x_3]^\top$ 的邻域包含所有以 $\boldsymbol{x}$ 为中心的球体内部的点（见图 4.7）。

如果集合 S 包含 $\boldsymbol{x}$ 的某个邻域，即 $\boldsymbol{x}$ 的某个邻域的所有点都属于 S（见图 4.8），那么点 $\boldsymbol{x} \in S$ 称为集合 S 的内点。S 的所有内点的集合称为 S 的内部。

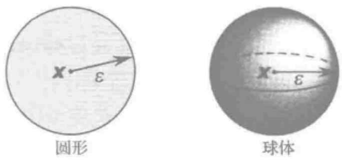


图 4.7 在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 中点 $\boldsymbol{x}$ 的邻域示意图

如果 $\boldsymbol{x}$ 的邻域既包含 S 中的点，也包含 S 外的点（见图 4.8），那么称点 $\boldsymbol{x}$ 为集合 S 的边界点。需要注意的是，S 的边界点可能是 S 中的元素，也可能不是 S 中的元素。S 的所有边界点的集合称为 S 的边界。

如果集合 S 包含它的每个点的邻域，那么称该集合是开集。也就是说，如果 S 的每个点都是内点，或者 S 不包含任何边界点，那么 S 是开集。

如果集合 S 包含边界点，那么称 S 是闭集（见图 4.9）。可以证明，当且仅当一个集合的补是开集，那么该集合是闭集。

如果一个集合可以被一个有限半径的球体所包围，那么该集合称为有界集。如果一个集合既是闭集又是有界集，那么该集合称为紧集。紧集对于优化问题而言是非常重要的，以下定理可以证实这一点。

E: P51

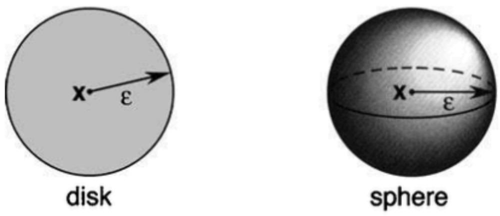


Figure 4.7 Examples of neighborhoods of a point in $\mathbb{R}^2$ and $\mathbb{R}^3$.

A point $\boldsymbol{x} \in S$ is said to be an *interior point* of the set S if the set S contains some neighborhood of $\boldsymbol{x}$; that is, if all points within some neighborhood of $\boldsymbol{x}$ are also in S (see Figure 4.8). The set of all the interior points of S is called the *interior* of S.

A point $\boldsymbol{x}$ is said to be a *boundary point* of the set S if every neighborhood of $\boldsymbol{x}$ contains a point in S and a point not in S (see Figure 4.8). Note that a boundary point of S may or may not be an element of S. The set of all boundary points of S is called the *boundary* of S.

EM: Did not specify **every/all** neighbourhood**(s)** of x. My reasoning is a counterexample that if the set S is finitely large, then there will always be a large enough neighbourhood of x (even though x is an interior point of S or an exterior point of S but not infinitely far away from S) that contains some points in S and some points that are not in S, so we must say all neighbourhoods of x including the very small ones.

CV: 如果 x 的所有邻域既包含 S 中的点，也包含S外的点(见图4.8)，那么称点 x 为集合 S 的边界点。

5.

C: P35

4.4 邻域

点 $\boldsymbol{x} \in \mathbb{R}^n$ 的邻域可以表示为

$$\{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\}$$

其中，ε 为某个正数。邻域也可视为半径为 ε、中心为 $\boldsymbol{x}$ 的球体。

在平面 $\mathbb{R}^2$ 中，点 $\boldsymbol{x} = [x_1, x_2]^\top$ 的邻域包含所有以 $\boldsymbol{x}$ 为中心的圆形内部的点。在 $\mathbb{R}^3$ 中，点 $\boldsymbol{x} = [x_1, x_2, x_3]^\top$ 的邻域包含所有以 $\boldsymbol{x}$ 为中心的球体内部的点(见图 4.7)。

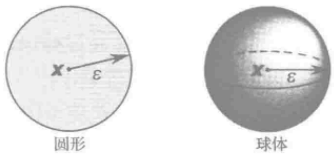


图 4.7　在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 中点 $\boldsymbol{x}$ 的邻域示意图

如果集合 S 包含 $\boldsymbol{x}$ 的某个邻域，即 $\boldsymbol{x}$ 的某个邻域的所有点都属于 S(见图 4.8)，那么点 $\boldsymbol{x} \in S$ 称为集合 S 的内点。S 的所有内点的集合称为 S 的内部。

如果 $\boldsymbol{x}$ 的邻域既包含 S 中的点，也包含 S 外的点(见图 4.8)，那么称点 $\boldsymbol{x}$ 为集合 S 的边界点。需要注意的是，S 的边界点可能是 S 中的元素，也可能不是 S 中的元素。S 的所有边界点的集合称为 S 的边界。

如果集合 S 包含它的每个点的邻域，那么称该集合是开集。也就是说，如果 S 的每个点都是内点，或者 S 不包含任何边界点，那么 S 是开集。

E: P51

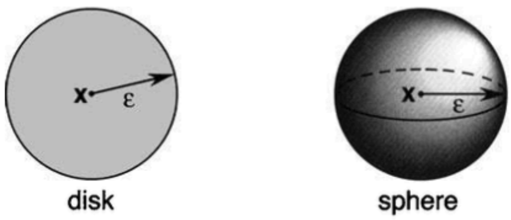


Figure 4.7　Examples of neighborhoods of a point in $\mathbb{R}^2$ and $\mathbb{R}^3$.

A point $\boldsymbol{x} \in S$ is said to be an *interior point* of the set S if the set S contains some neighborhood of $\boldsymbol{x}$; that is, if all points within some neighborhood of $\boldsymbol{x}$ are also in S (see Figure 4.8). The set of all the interior points of S is called the *interior* of S.

A point $\boldsymbol{x}$ is said to be a *boundary point* of the set S if every neighborhood of $\boldsymbol{x}$ contains a point in S and a point not in S (see Figure 4.8). Note that a boundary point of S may or may not be an element of S. The set of all boundary points of S is called the *boundary* of S.

A set S is said to be *open* if it contains a neighborhood of each of its points; that is, if each of its points is an interior point, or equivalently, if S contains no boundary points.

A set S is said to be *closed* if it contains its boundary (see Figure 4.9). We can show that a set is closed if and only if its complement is open.

A set that is contained in a ball of finite radius is said to be *bounded*. A set is *compact* if it is both closed and bounded. Compact sets are important in optimization problems for the following reason.

EM: Did not specify that set S **only has to contain one** neighbourhood of each of its points to be defined as *open*, the Chinese version says “S is *open* if it contains neighbourhoods of each of its points”, there is no **attributive** before neighbourhood(s).

CV: 如果集合S包含它的每个点的一个/某个邻域,那么称该集合是开集。

6.

C: P36

4.5 多面体和多胞形

令 Θ 为一个凸集，$\boldsymbol{y}$ 是 Θ 的一个边界点。某个经过点 $\boldsymbol{y}$ 的超平面将 $\mathbb{R}^n$ 空间分为两个半空间，如果 Θ 完全位于其中一个半空间内，那么称该超平面为集合 Θ 的支撑超平面。

根据定理 4.1 可知，任意多个凸集的交集也是凸集。此处关心的是有限个半空间的交集。由于在 $\mathbb{R}^n$ 中每个半空间 H_+ 或 H_- 都是凸集，因此任意数量的半空间的交集是凸集。

如果一个集合可以表示为有限个半空间的交集，那么称该集合为多面体，如图 4.10 所示。

一个非空有界多面体称为多胞形，如图 4.11 所示。

E: P52

4.5 Polytopes and Polyhedra

Let Θ be a convex set, and suppose that $\boldsymbol{y}$ is a boundary point of Θ. A hyperplane passing through $\boldsymbol{y}$ is called a *hyperplane of support* (or *supporting hyperplane*) of the set Θ if the entire set Θ lies completely in one of the two half-spaces into which this hyperplane divides the space $\mathbb{R}^n$.

EM: The translator mistakenly interchanged the terms “**polytopes**” and “**polyhedra**,” using each in place of the other.

4.5 多面体和多胞形

令 Θ 为一个凸集，$\boldsymbol{y}$ 是 Θ 的一个边界点。某个经过点 $\boldsymbol{y}$ 的超平面将 $\mathbb{R}^n$ 空间分为两个半空间，如果 Θ 完全位于其中一个半空间内，那么称该超平面为集合 Θ 的支撑超平面。

根据定理 4.1 可知，任意多个凸集的交集也是凸集。此处关心的是有限个半空间的交集。由于在 $\mathbb{R}^n$ 中每个半空间 H_+ 或 H_- 都是凸集，因此任意数量的半空间的交集是凸集。

如果一个集合可以表示为有限个半空间的交集，那么称该集合为多面体，如图 4.10 所示。

一个非空有界多面体称为多胞形，如图 4.11 所示。

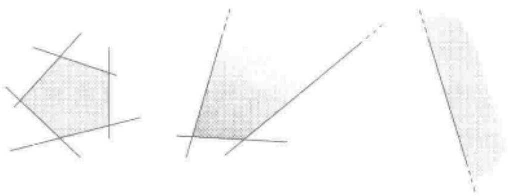


图 4.10　多面体

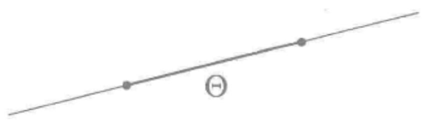


图 4.11　一维多胞形

CV: 多胞形和多面体

7. (**Correct according to Errata**)

C: P36

4.5 多面体和多胞形

令 Θ 为一个凸集，$\boldsymbol{y}$ 是 Θ 的一个边界点。某个经过点 $\boldsymbol{y}$ 的超平面将 $\mathbb{R}^n$ 空间分为两个半空间，如果 Θ 完全位于其中一个半空间内，那么称该超平面为集合 Θ 的支撑超平面。

根据定理 4.1 可知，任意多个凸集的交集也是凸集。此处关心的是有限个半空间的交集。由于在 $\mathbb{R}^n$ 中每个半空间 H_+ 或 H_- 都是凸集，因此任意数量的半空间的交集是凸集。

如果一个集合可以表示为有限个半空间的交集，那么称该集合为多面体，如图 4.10 所示。

一个非空有界多面体称为多胞形，如图 4.11 所示。

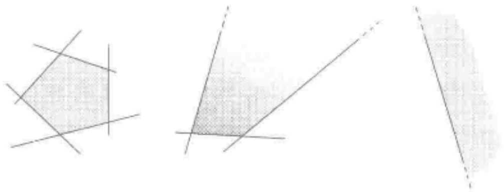


图 4.10　多面体

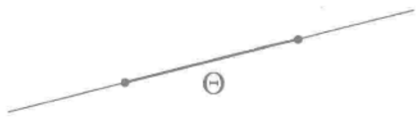


图 4.11　一维多胞形

4.5 多面体和多胞形

令 Θ 为一个凸集，$\boldsymbol{y}$ 是 Θ 的一个边界点。某个经过点 $\boldsymbol{y}$ 的超平面将 $\mathbb{R}^n$ 空间分为两个半空间，如果 Θ 完全位于其中一个半空间内，那么称该超平面为集合 Θ 的支撑超平面。

根据定理 4.1 可知，任意多个凸集的交集也是凸集。此处关心的是有限个半空间的交集。由于在 $\mathbb{R}^n$ 中每个半空间 H_+ 或 H_- 都是凸集，因此任意数量的半空间的交集是凸集。

如果一个集合可以表示为有限个半空间的交集，那么称该集合为多面体，如图 4.10 所示。

一个非空有界多面体称为多胞形，如图 4.11 所示。

E: P52

4.5 Polytopes and Polyhedra

Let Θ be a convex set, and suppose that $\boldsymbol{y}$ is a boundary point of Θ. A hyperplane passing through $\boldsymbol{y}$ is called a *hyperplane of support* (or *supporting hyperplane*) of the set Θ if the entire set Θ lies completely in one of the two half-spaces into which this hyperplane divides the space $\mathbb{R}^n$.

Recall that by Theorem 4.1, the intersection of any number of convex sets is convex. In what follows we are concerned with the intersection of a finite number of half-spaces. Because every half-space H_+ or H_- is convex in $\mathbb{R}^n$, the intersection of any number of half-spaces is a convex set.

A set that can be expressed as the intersection of a finite number of half-spaces is called a *convex polytope* (see Figure 4.10).

4.5 Polytopes and Polyhedra

Let Θ be a convex set, and suppose that $\boldsymbol{y}$ is a boundary point of Θ. A hyperplane passing through $\boldsymbol{y}$ is called a *hyperplane of support* (or *supporting hyperplane*) of the set Θ if the entire set Θ lies completely in one of the two half-spaces into which this hyperplane divides the space $\mathbb{R}^n$.

Recall that by Theorem 4.1, the intersection of any number of convex sets is convex. In what follows we are concerned with the intersection of a finite number of half-spaces. Because every half-space H_+ or H_- is convex in $\mathbb{R}^n$, the intersection of any number of half-spaces is a convex set.

A set that can be expressed as the intersection of a finite number of half-spaces is called a *convex polytope* (see Figure 4.10).

A nonempty bounded polytope is called a *polyhedron* (see Figure 4.11).

EM: **No error**, this part is correct although it is different from the original book, because the **errata** has changed "**convex polytope**" into "**polyhedron**" and "A nonempty bounded **polytope** is called a ***polyhedron***" to "A nonempty bounded **polyhedron** is called a ***polytope***".

Errata:

Errata:
An Introduction to Optimization, Fourth Edition
by
Edwin K. P. Chong and Stanislaw H. Żak

Version: March 6, 2024

Typos and minor changes: Printings 1–3

- p. 12, second line from bottom: Remove one of the two repeated instances of "use."
- p. 15, line 5: Add to the end of the sentence: "Moreover, it follows from properties 1 and 2 that the determinant of a matrix with two identical columns, not necessarily next to each other, is always 0." [Thanks to Zain Khandwala.]
- p. 52, four lines from bottom: Change "convex polytope" to "polyhedron". (Unfortunately, the definitions of *polytope* and *polyhedron* are not universal, so the original text is still consistent with some definitions in the literature.)

Errata:
An Introduction to Optimization, Fourth Edition
by
Edwin K. P. Chong and Stanislaw H. Żak

Version: March 6, 2024

Typos and minor changes: Printings 1–3

- p. 12, second line from bottom: Remove one of the two repeated instances of "use."
- p. 15, line 5: Add to the end of the sentence: "Moreover, it follows from properties 1 and 2 that the determinant of a matrix with two identical columns, not necessarily next to each other, is always 0." [Thanks to Zain Khandwala.]
- p. 52, four lines from bottom: Change "convex polytope" to "polyhedron". (Unfortunately, the definitions of *polytope* and *polyhedron* are not universal, so the original text is still consistent with some definitions in the literature.)
- p. 52, caption of Figure 4.10: Change to "Polyhedra."
- p. 52, three lines from bottom: Change "A nonempty bounded polytope is called a *polyhedron*" to "A nonempty bounded polyhedron is called a *polytope*". (Unfortunately, the definitions of *polytope* and *polyhedron* are not universal, so the original text is still consistent with some definitions in the literature.)

8. (**Correct according to Errata**)

C: P36

4.5 多面体和多胞形

令 Θ 为一个凸集，$\boldsymbol{y}$ 是 Θ 的一个边界点。某个经过点 $\boldsymbol{y}$ 的超平面将 $\mathbb{R}^n$ 空间分为两个半空间，如果 Θ 完全位于其中一个半空间内，那么称该超平面为集合 Θ 的支撑超平面。

根据定理 4.1 可知，任意多个凸集的交集也是凸集。此处关心的是有限个半空间的交集。由于在 $\mathbb{R}^n$ 中每个半空间 H_+ 或 H_- 都是凸集，因此任意数量的半空间的交集是凸集。

如果一个集合可以表示为有限个半空间的交集，那么称该集合为多面体，如图 4.10 所示。

一个非空有界多面体称为多胞形，如图 4.11 所示。

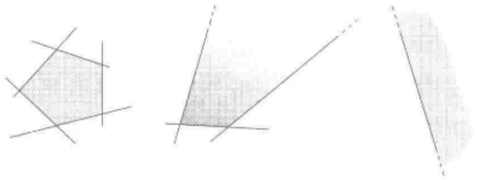


图 4.10 多面体

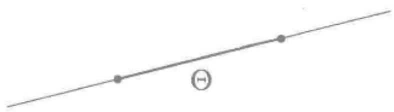


图 4.11 一维多胞形

E: P52,53

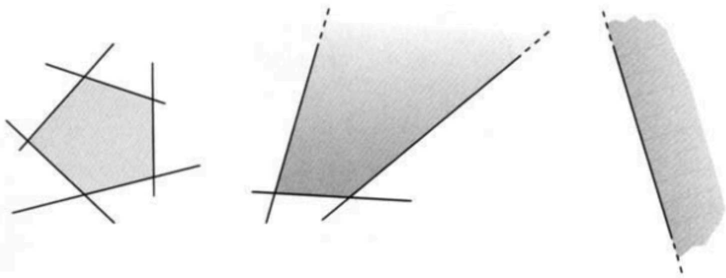


Figure 4.10 Polytopes.

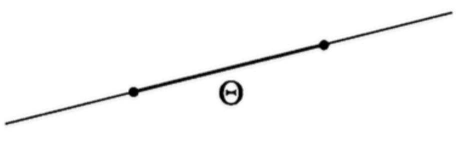


Figure 4.11 One-dimensional polyhedron.

EM: **No error**, this part is correct although it is different from the original book, because the **errata** has changed the names into “polyhedra” and “one-dimensional polytope”.
Errata:

Errata:
An Introduction to Optimization, Fourth Edition
by
Edwin K. P. Chong and Stanislaw H. Żak

Version: March 6, 2024

Typos and minor changes: Printings 1–3

- p. 12, second line from bottom: Remove one of the two repeated instances of "use."
- p. 15, line 5: Add to the end of the sentence: "Moreover, it follows from properties 1 and 2 that the determinant of a matrix with two identical columns, not necessarily next to each other, is always 0." [Thanks to Zain Khandwala.]
- p. 52, four lines from bottom: Change "convex polytope" to "polyhedron". (Unfortunately, the definitions of *polytope* and *polyhedron* are not universal, so the original text is still consistent with some definitions in the literature.)
- p. 52, caption of Figure 4.10: Change to "Polyhedra."
- p. 52, three lines from bottom: Change "A nonempty bounded polytope is called a *polyhedron*" to "A nonempty bounded polyhedron is called a *polytope*". (Unfortunately, the definitions of *polytope* and *polyhedron* are not universal, so the original text is still consistent with some definitions in the literature.)
- p. 53, caption of Figure 4.11: Change "polyhedron" to "polytope".

9. (not a serious problem)

C: P47

5.5 水平集与梯度

函数f: $\mathbb{R}^n \to \mathbb{R}$ 在水平 c 上的水平集定义为

$$S = \{\boldsymbol{x} : f(\boldsymbol{x}) = c\}$$

对于f: $\mathbb{R}^2 \to \mathbb{R}$，水平集 S 是一条曲线，能够直接观察；对于f: $\mathbb{R}^3 \to \mathbb{R}$，水平集 S 通常是一组曲面。

E: P68

5.5 Level Sets and Gradients

The *level set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level c is the set of points

$$S = \{\boldsymbol{x} : f(\boldsymbol{x}) = c\}.$$

For $f : \mathbb{R}^2 \to \mathbb{R}$, we are usually interested in S when it is a curve. For $f : \mathbb{R}^3 \to \mathbb{R}$, the sets S most often considered are surfaces.

EM: The translator paraphrased in his own words that "For f:R2 to R, level set S is a curve, which we can directly observe; for f:R3 to R, level set S is usually a set of curved surfaces."
CV: If directly translated, it should be 对于f:R2 to R 的情况，我们通常关注 S 是曲线的情形。对于 f:R3 to R，通常考虑 S 是曲面。"

10.

C: P158

最后给出一个例子，讨论如何利用最小二乘分析方法导出正交投影算子。

例 12.5 正交投影算子。$\mathcal{V}\subset\mathbb{R}^n$ 为一个子空间，给定一个向量 $\boldsymbol{x}\in\mathbb{R}^n$，可将向量进行正交分解，即

$$\boldsymbol{x}=\boldsymbol{x}_{\mathcal{V}}+\boldsymbol{x}_{\mathcal{V}^{\perp}}$$

其中，$\boldsymbol{x}_{\mathcal{V}}\in\mathcal{V}$ 表示 $\boldsymbol{x}$ 在 $\mathcal{V}$ 上的正交投影，$\boldsymbol{x}_{\mathcal{V}^{\top}}\in\mathcal{V}^{\perp}$ 表示 $\boldsymbol{x}$ 在 $\mathcal{V}^{\perp}$ 上的正交投影(见 3.3 节，且 $\mathcal{V}^{\perp}$ 是 $\mathcal{V}$ 的正交补)。存在矩阵 $\boldsymbol{P}$，使得 $\boldsymbol{x}_{\mathcal{V}}=\boldsymbol{P}\boldsymbol{x}$，$\boldsymbol{P}$ 称为正交投影算子。接下来分别针对 $\mathcal{V}=\mathcal{R}(\boldsymbol{A})$ 和 $\mathcal{V}=\mathcal{N}(\boldsymbol{A})$ 这两种情况，讨论 $\boldsymbol{P}$ 的表达式。

E: P226

Example 12.5 *Orthogonal Projectors.* Let $\mathcal{V}\subset\mathbb{R}^n$ be a subspace. Given a vector $\boldsymbol{x}\in\mathbb{R}^n$, we write the orthogonal decomposition of $\boldsymbol{x}$ as

$$\boldsymbol{x}=\boldsymbol{x}_{\mathcal{V}}+\boldsymbol{x}_{\mathcal{V}^{\perp}},$$

where $\boldsymbol{x}_{\mathcal{V}}\in\mathcal{V}$ is the orthogonal projection of $\boldsymbol{x}$ onto $\mathcal{V}$ and $\boldsymbol{x}_{\mathcal{V}^{\perp}}\in\mathcal{V}^{\perp}$ is the orthogonal projection of $\boldsymbol{x}$ onto $\mathcal{V}^{\perp}$. (See Section 3.3; also recall that $\mathcal{V}^{\perp}$ is

EM: Small formula error

CV: formula correction

11.

C: P339

$\boldsymbol{h}(\boldsymbol{x})=\mathbf{0}$，$\boldsymbol{g}(\boldsymbol{x})\leqslant\mathbf{0}$。其中，$f:\mathbb{R}^n\to\mathbb{R}^m$，$\boldsymbol{h}:\mathbb{R}^n\to\mathbb{R}^m(m\leqslant n)$，$g:\mathbb{R}^n\to\mathbb{R}^p$，且$f,\boldsymbol{h},\boldsymbol{g}\in\mathcal{C}^2$。假设$\boldsymbol{x}^*$是正则点，那么存在$\boldsymbol{\lambda}^*\in\mathbb{R}^m$和$\boldsymbol{\mu}^*\in\mathbb{R}^p$，使得

1. $\boldsymbol{\mu}^*\geqslant\mathbf{0}$，$Df(\boldsymbol{x}^*)+\boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*)+\boldsymbol{\mu}^{*\top}D\boldsymbol{g}(\boldsymbol{x}^*)=\mathbf{0}^\top$，$\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*)=0$；
2. 对于所有$\boldsymbol{y}\in T(\boldsymbol{x}^*)$，都有$\boldsymbol{y}^\top\boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\boldsymbol{y}\geqslant0$成立。 □

证明：第 1 个结论就是 KKT 定理的结果。下面证明第 2 个结论，因为$\boldsymbol{x}^*$是$\{\boldsymbol{x}:\boldsymbol{h}(\boldsymbol{x})=\mathbf{0},\boldsymbol{g}(\boldsymbol{x})\leqslant\mathbf{0}\}$上的局部极小点，所以它也是$\{\boldsymbol{x}:\boldsymbol{h}(\boldsymbol{x})=\mathbf{0},g_j(\boldsymbol{x})=\mathbf{0},j\in J(\boldsymbol{x}^*)\}$上的局部极小点；也就是说，当把起作用约束当作等式约束时，点$\boldsymbol{x}^*$是局部极小点（证明过程留作习题 21.16）。因此，仅含等式约束（见定理 20.4）的极值问题的二阶必要条件在此处是成立的，由此即可完成第 2 个结论的证明。 ■

下面给出不等式约束极值问题的二阶充分条件。该条件中需要使用如下集合：

$$\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)=\{\boldsymbol{y}:D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y}=\mathbf{0},Dg_i(\boldsymbol{x}^*)\boldsymbol{y}=0,i\in\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)\}$$

其中，$\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)=\{i:g_i(\boldsymbol{x}^*)=0,u_i^*>0\}$。注意$\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$是$J(\boldsymbol{x}^*)$的子集，即$\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)\subset J(\boldsymbol{x}^*)$成立。这意味着，$T(\boldsymbol{x}^*)$是$\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$的子集，即$T(\boldsymbol{x}^*)\subset\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$。

E: P498

Theorem 21.2 ***Second-Order Necessary Conditions.*** *Let $\boldsymbol{x}^*$ be a local minimizer of $f:\mathbb{R}^n\to\mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x})=\mathbf{0}$, $\boldsymbol{g}(\boldsymbol{x})\le\mathbf{0}$, $\boldsymbol{h}:\mathbb{R}^n\to\mathbb{R}^m$, $m\le n$, $\boldsymbol{g}:\mathbb{R}^n\to\mathbb{R}^p$, and $f,\boldsymbol{h},\boldsymbol{g}\in\mathcal{C}^2$. Suppose that $\boldsymbol{x}^*$ is regular. Then, there exist $\boldsymbol{\lambda}^*\in\mathbb{R}^m$ and $\boldsymbol{\mu}^*\in\mathbb{R}^p$ such that:*

1. $\boldsymbol{\mu}^*\ge\mathbf{0}$, $Df(\boldsymbol{x}^*)+\boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*)+\boldsymbol{\mu}^{*\top}D\boldsymbol{g}(\boldsymbol{x}^*)=\mathbf{0}^\top$, $\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*)=0$.
2. *For all $\boldsymbol{y}\in T(\boldsymbol{x}^*)$ we have $\boldsymbol{y}^\top\boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\boldsymbol{y}\ge0$.* □

Proof. Part 1 is simply a result of the KKT theorem. To prove part 2, we note that because the point $\boldsymbol{x}^*$ is a local minimizer over $\{\boldsymbol{x}:\boldsymbol{h}(\boldsymbol{x})=\mathbf{0},\ \boldsymbol{g}(\boldsymbol{x})\le\mathbf{0}\}$, it is also a local minimizer over $\{\boldsymbol{x}:\boldsymbol{h}(\boldsymbol{x})=\mathbf{0},\ g_j(\boldsymbol{x})=0,\ j\in J(\boldsymbol{x}^*)\}$; that is, the point $\boldsymbol{x}^*$ is a local minimizer with active constraints taken as equality constraints (see Exercise 21.16). Hence, the second-order necessary conditions for equality constraints (Theorem 20.4) are applicable here, which completes the proof. ■

We now state the second-order sufficient conditions for extremum problems involving inequality constraints. In the formulation of the result, we use the following set:

$$\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)=\{\boldsymbol{y}:D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y}=\mathbf{0},Dg_i(\boldsymbol{x}^*)\boldsymbol{y}=0,i\in\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)\},$$

where $\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)=\{i:g_i(\boldsymbol{x}^*)=0,\mu_i^*>0\}$. Note that $\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$ is a subset of $J(\boldsymbol{x}^*)$: $\tilde{J}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)\subset J(\boldsymbol{x}^*)$. This, in turn, implies that $T(\boldsymbol{x}^*)$ is a subset of $\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$: $T(\boldsymbol{x}^*)\subset\tilde{T}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$.

EM: \mu and u formula error

CV: formula correction

12.

C: P375

例 23.1　某优化问题为

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to} \quad \|\boldsymbol{x}\|^2 = 1 \end{aligned}$$

其中，$\boldsymbol{Q}=\boldsymbol{Q}^\top>0$，要求利用步长固定投影梯度法求解。

a. 推导出此算法的更新方程(即写出 $\boldsymbol{x}^{(k+1)}$ 的计算公式，用 $\boldsymbol{x}^{(k)}$、$\boldsymbol{Q}$ 和固定步长 α 表示)。可假定在计算迭代点 $\boldsymbol{x}^{(k)}$ 的过程中，投影算子的幅角一直都不会为 0。

b. 即使步长 $\alpha>0$ 取任意小的值，该算法是否仍不可能收敛到一个最优解?

c. 证明当 $0<\alpha<1/\lambda_{max}$ (λ_{max} 是 $\boldsymbol{Q}$ 的最大特征值)时，步长固定投影梯度法(步长为 α)收敛到最优解的前提是 $\boldsymbol{x}^{(0)}$ 与 $\boldsymbol{Q}$ 的最小特征值对应的特征向量之间不正交(假定矩阵 $\boldsymbol{A}$ 不存在重复的特征值)。

E: P498

Example 23.1 Consider the problem

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to} \quad \|\boldsymbol{x}\|^2 = 1, \end{aligned}$$

where $\boldsymbol{Q}=\boldsymbol{Q}^\top>0$. Suppose that we apply a *fixed-step-size projected gradient algorithm* to this problem.

a. Derive a formula for the update equation for the algorithm (i.e., write down an explicit formula for $\boldsymbol{x}^{(k+1)}$ as a function of $\boldsymbol{x}^{(k)}$, $\boldsymbol{Q}$, and the fixed step size α). You may assume that the argument in the projection operator to obtain $\boldsymbol{x}^{(k)}$ is never zero.

EM: The translator confused the word “**argument**” in the context of **function/computing**, and he used the Chinese translation of the word “**argument**” in the context of **complex numbers**, which has nearly no connection with the whole book.

CV: …投影算子的参数一直都不会为0。

13.

C: P375

例 23.1 某优化问题为

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1\end{array}$$

其中，$\boldsymbol{Q} = \boldsymbol{Q}^\top > 0$，要求利用步长固定投影梯度法求解。

a. 推导出此算法的更新方程（即写出 $\boldsymbol{x}^{(k+1)}$ 的计算公式，用 $\boldsymbol{x}^{(k)}$、$\boldsymbol{Q}$ 和固定步长 α 表示）。可假定在计算迭代点 $\boldsymbol{x}^{(k)}$ 的过程中，投影算子的幅角一直都不会为 0。

b. 即使步长 $\alpha > 0$ 取任意小的值，该算法是否仍不可能收敛到一个最优解？

c. 证明当 $0 < \alpha < 1/\lambda_{\max}$（$\lambda_{\max}$ 是 $\boldsymbol{Q}$ 的最大特征值）时，步长固定投影梯度法（步长为 α）收敛到最优解的前提是 $\boldsymbol{x}^{(0)}$ 与 $\boldsymbol{Q}$ 的最小特征值对应的特征向量之间不正交（假定矩阵 $\boldsymbol{A}$ 不存在重复的特征值）。

E: P498

Example 23.1 Consider the problem

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1,\end{array}$$

where $\boldsymbol{Q} = \boldsymbol{Q}^\top > 0$. Suppose that we apply a *fixed-step-size projected gradient algorithm* to this problem.

a. Derive a formula for the update equation for the algorithm (i.e., write down an explicit formula for $\boldsymbol{x}^{(k+1)}$ as a function of $\boldsymbol{x}^{(k)}$, $\boldsymbol{Q}$, and the fixed step size α). You may assume that the argument in the projection operator to obtain $\boldsymbol{x}^{(k)}$ is never zero.

b. Is it possible for the algorithm not to converge to an optimal solution even if the step size $\alpha > 0$ is taken to be arbitrarily small?

c. Show that for $0 < \alpha < 1/\lambda_{\max}$ (where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{Q}$), the fixed-step-size projected gradient algorithm (with step size α)

EM: Logical error. The translated sentence means "**Is it still not possible** for the algorithm to converge to an optimal solution even if the step size alpha>0 is taken to be arbitrarily small?"

This sentence suggests that the **before** alpha is specified to be arbitrarily small, the algorithm is **certain to not to converge** to an optimal solution, that is why the translated question asks "is it **still** impossible to converge", but the English version is "is it **possible not to** converge" meaning that before specifying alpha, it **may** converge and **may not**, but small alpha can **improve** the case but the question is whether it is still limited, and **still may not** converge. I believe the main difference is what is accepted before asking the question, "may" or "impossible".

CV: Directly from English, 即使 alpha > 0 取任意小的值，该算法是否仍可能不收敛到一个最优解？

14.

C: P387

例 23.3 考虑问题：

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1 \end{array}$$

其中，$\boldsymbol{Q} = \boldsymbol{Q}^\top > 0$。

a. 利用罚函数 $P(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 - 1)^2$ 和惩罚因子 γ，写出原问题对应的无约束优化问题，使得其最优解 $\boldsymbol{x}_\gamma$ 近似于原问题的解；

b. 证明对于任何 γ，$\boldsymbol{x}_\gamma$ 都是 $\boldsymbol{Q}$ 的一个特征值；

c. 证明当 $\gamma \to \infty$ 时，$\|\boldsymbol{x}_\gamma\|^2 - 1 = O(1/\gamma)$。

E: P567

Example 23.3 Consider the problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1, \end{array}$$

where $\boldsymbol{Q} = \boldsymbol{Q}^\top > 0$.

a. Using the penalty function $P(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 - 1)^2$ and penalty parameter γ, write down an unconstrained optimization problem whose solution $\boldsymbol{x}_\gamma$ approximates the solution to this problem.

b. Show that for any γ, $\boldsymbol{x}_\gamma$ is an eigenvector of $\boldsymbol{Q}$.

c. Show that $\|\boldsymbol{x}_\gamma\|^2 - 1 = O(1/\gamma)$ as $\gamma \to \infty$.

EM: Confused “eigenvalue” and “eigenvector”, should be eigenvector rather than eigenvalue.

CV: 特征向量

15.

C: P23

$$\Delta_1 = q_{11}, \qquad \Delta_2 = \det \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

$$\Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}, \cdots, \quad \Delta_n = \det \boldsymbol{Q}$$

下面证明西尔维斯特准则，该准则可以仅根据 $\boldsymbol{Q}$ 的顺序主子式判定二次型$\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$ 是否正定。

定理 3.6 西尔维斯特准则。给定二次型 $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$，其中 $\boldsymbol{Q} = \boldsymbol{Q}^\top$，该二次型是正定的，当且仅当 $\boldsymbol{Q}$ 的顺序主子式是正定的。 □

E: P32

Theorem 3.6 ***Sylvester's Criterion.*** *A quadratic form $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$, $\boldsymbol{Q} = \boldsymbol{Q}^\top$, is positive definite if and only if the leading principal minors of $\boldsymbol{Q}$ are positive.* □

Proof. The key to the proof of Sylvester's criterion is the fact that a quadratic form whose leading principal minors are nonzero can be expressed in some basis as a sum of squares

$$\frac{\Delta_0}{\Delta_1}\tilde{x}_1^2 + \frac{\Delta_1}{\Delta_2}\tilde{x}_2^2 + \cdots + \frac{\Delta_{n-1}}{\Delta_n}\tilde{x}_n^2,$$

EM: Mistyped positive as positive definite, should delete “definite” as a minor is a number. A number can only be positive or negative instead of positive definite etc.
CV: 正的 instead of 正定的

16.

C: P25

$$x^{\top}Qx=\tilde{x}^{\top}\tilde{Q}\tilde{x}>0$$

对于任意 $x\neq 0$(或对于任意 $\tilde{x}\neq 0$)都成立。

为证明必要性，首先证明对于 $i=1,\cdots,n$，有 $\Delta_i\neq 0$。为此，假定对于某个 k，有 $\Delta_k=0$。注意到 $\Delta_k=\det Q_k$，其中

$$Q_k=\begin{bmatrix} q_{11} & \cdots & q_{1k} \\ \vdots & \ddots & \vdots \\ q_{k1} & \cdots & q_{kk} \end{bmatrix}$$

那么，存在一个向量 $v\in\mathbb{R}^k$，$v\neq 0$，使得 $v^{\top}Q_k=0$。令 $x\in\mathbb{R}^n$，且 $x=[v^{\top},0^{\top}]^{\top}$，可得

$$x^{\top}Qx=v^{\top}Q_kv=0$$

但是已知 $x\neq 0$，这与二次型 f 是正定的这一事实相矛盾。因此，如果 $x^{\top}Qx>0$，那么必定有 $\Delta_i\neq 0$，$i=1,\cdots,n$。根据前面得到的等式：

$$x^{\top}Qx=\tilde{x}^{\top}\tilde{Q}\tilde{x}=\frac{1}{\Delta_1}\tilde{x}_1^2+\frac{\Delta_1}{\Delta_2}\tilde{x}_2^2+\cdots+\frac{\Delta_{n-1}}{\Delta_n}\tilde{x}_n^2$$

其中 $\tilde{x}=[v_1,\cdots,v_n]x$，可知，如果二次型是正定的，那么 Q 的所有顺序主子式都是正定的。 ■

E: P34

In the new basis, the quadratic form can be expressed as a sum of squares

$$x^{\top}Qx=\tilde{x}^{\top}\tilde{Q}\tilde{x}=\frac{1}{\Delta_1}\tilde{x}_1^2+\frac{\Delta_1}{\Delta_2}\tilde{x}_2^2+\cdots+\frac{\Delta_{n-1}}{\Delta_n}\tilde{x}_n^2.$$

We now show that a necessary and sufficient condition for the quadratic form to be positive definite is $\Delta_i>0$, $i=1,\ldots,n$.

Sufficiency is clear, for if $\Delta_i>0$, $i=1,\ldots,n$, then by the previous argument there is a basis such that

$$x^{\top}Qx=\tilde{x}^{\top}\tilde{Q}\tilde{x}>0$$

for any $x\neq 0$ (or, equivalently, any $\tilde{x}\neq 0$).

To prove necessity, we first show that for $i=1,\ldots,n$, we have $\Delta_i\neq 0$. To see this, suppose that $\Delta_k=0$ for some k. Note that $\Delta_k=\det Q_k$,

$$Q_k=\begin{bmatrix} q_{11} & \cdots & q_{1k} \\ \vdots & \ddots & \vdots \\ q_{k1} & \cdots & q_{kk} \end{bmatrix}.$$

Then, there exists a vector $v\in\mathbb{R}^k$, $v\neq 0$, such that $v^{\top}Q_k=0$. Now let $x\in\mathbb{R}^n$ be given by $x=[v^{\top},0^{\top}]^{\top}$. Then,

$$x^{\top}Qx=v^{\top}Q_kv=0.$$

But $x\neq 0$, which contradicts the fact that the quadratic form f is positive definite. Therefore, if $x^{\top}Qx>0$, then $\Delta_i\neq 0$, $i=1,\ldots,n$. Then, using our previous argument, we may write

$$x^{\top}Qx=\tilde{x}^{\top}\tilde{Q}\tilde{x}=\frac{1}{\Delta_1}\tilde{x}_1^2+\frac{\Delta_1}{\Delta_2}\tilde{x}_2^2+\cdots+\frac{\Delta_{n-1}}{\Delta_n}\tilde{x}_n^2,$$

where $\tilde{x}=[v_1,\ldots,v_n]x$. Hence, if the quadratic form is positive definite, then all leading principal minors must be positive. ■

EM: Same as 15

CV: Same as 15

17. (wording)

C: P123

推导过程为

$$\begin{aligned} g^{(1)\top}d^{(0)} &= (Qx^{(1)}-b)^\top d^{(0)} \\ &= x^{(0)\top}Qd^{(0)} - \left(\frac{g^{(0)\top}d^{(0)}}{d^{(0)\top}Qd^{(0)}}\right)d^{(0)\top}Qd^{(0)} - b^\top d^{(0)} \\ &= g^{(0)\top}d^{(0)} - g^{(0)\top}d^{(0)} = 0 \end{aligned}$$

方程 $g^{(1)\top}d^{(0)}=0$ 表示步长 α_0 为 $\alpha_0=\arg\min\phi_0(\alpha)$，其中，$\phi_0(\alpha)=f(x^{(0)}+\alpha d^{(0)})$。下面给出具体的推导过程。

由链式法则，可得

$$\frac{\mathrm{d}\phi_0}{\mathrm{d}\alpha}(\alpha) = \nabla f(x^{(0)}+\alpha d^{(0)})^\top d^{(0)}$$

E: P179

To see this,

$$\begin{aligned} g^{(1)\top}d^{(0)} &= (Qx^{(1)}-b)^\top d^{(0)} \\ &= x^{(0)\top}Qd^{(0)} - \left(\frac{g^{(0)\top}d^{(0)}}{d^{(0)\top}Qd^{(0)}}\right)d^{(0)\top}Qd^{(0)} - b^\top d^{(0)} \\ &= g^{(0)\top}d^{(0)} - g^{(0)\top}d^{(0)} = 0. \end{aligned}$$

The equation $g^{(1)\top}d^{(0)} = 0$ implies that α_0 has the property that $\alpha_0 = \arg\min\phi_0(\alpha)$, where $\phi_0(\alpha) = f(x^{(0)}+\alpha d^{(0)})$. To see this, apply the chain rule to get

$$\frac{d\phi_0}{d\alpha}(\alpha) = \nabla f(x^{(0)}+\alpha d^{(0)})^\top d^{(0)}.$$

EM: The translated sentence in English is “The equation g(1)Td(0) = 0 **represents step size alpha_0 as** alpha_0 = arg min phi_0(alpha), where…” especially in Chinese, we don’t know whether the phrase 表示步长 means “the equation is the representation of the step size” or “the equation represents **that** the step size (as a subject of the clause)… And there is no clue after the verb “represent” which one it is, because the contents after it is hardly a sentence. I did not understand this sentence until I read the English version.

CV: 方程 g(1)Td(0) = 0 意味着步长 alpha_0 具备性质 alpha_0 = arg min phi_0(alpha)，其中...

18. (noun term)

C: P178

Widrow 和 Hoff 将该算法应用到线性神经元的训练过程中(这方面的发展历史，可见参考文献[132])。图 13.5 描述的是带有训练算法的单个神经元，这经常被称为学习机(Adaline)，是自适应线性组件的缩写。

E: P257

The algorithm above was applied to the training of linear neurons by Widrow and Hoff (see [132] for some historical remarks). The single neuron together with the training algorithm above is illustrated in Figure 13.5 and is often called *Adaline*, an acronym for *adaptive linear element*.

EM: I have searched many websites and did not find any source translating Adaline into “Learning Machine” in Chinese.
CV: I suggest giving the name directly as ADALINE or Adaline.