第 5 章 约束优化问题的最优性条件

本章介绍约束优化问题的最优性条件，首先简单介绍约束优化问题最优性条件的基本思想，再介绍仅含不等式约束问题的最优性条件，最后介绍含有一般约束问题的最优性条件。约束优化问题的最优性条件是讨论其算法的基础，本章只讨论一阶条件，不涉及更高阶的最优性条件。

5.1 约束优化问题最优性条件的基本思想

29min

本章考虑的约束优化问题的标准模型为

$$\begin{cases} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \leqslant 0 \quad i=1,2,\cdots,m \\ & h_j(\boldsymbol{x})=0 \quad j=1,2,\cdots,l \end{cases} \tag{5-1}$$

其中，$g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,m$ 称为不等式约束，$h_j(\boldsymbol{x})=0, j=1,2,\cdots,l$ 称为等式约束。有的约束优化问题中还含有箱子集（上下界）约束，这可以归结到不等式约束中。记

$$F=\{\boldsymbol{x}\subseteq \mathbf{R}^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,m; h_j(\boldsymbol{x})=0, j=1,2,\cdots,l\} \tag{5-2}$$

称 F 为问题(5-1)的可行集。

一般来讲，约束优化问题的可行域中不含有对应的无约束优化问题的解，也就是说，目标函数在无约束情况下的稳定点一般不在可行域内，所以无约束优化问题的最优性条件在约束优化问题中已不再适用，需另行研究一套约束优化问题最优性条件的理论。

为了更加直观地描述约束优化问题的最优性条件，先从几何上进行分析。为此，引入下降方向与可行方向的概念。

5.1.1 下降方向

定义 5-1 设 $f(\boldsymbol{x})$是定义在 $\mathbf{R}^n$ 上的实函数，$\bar{\boldsymbol{x}}\in\mathbf{R}^n$，$\boldsymbol{d}$ 是非零向量，若存在 $\delta>0$，使对每个 $\lambda\in(0,\delta)$都有

$$f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<f(\bar{\boldsymbol{x}}) \tag{5-3}$$

则称 $\boldsymbol{d}$ 为函数 $f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处的下降方向。将 $\bar{\boldsymbol{x}}$ 处的所有下降方向组成的集合记作

$$D=\{\boldsymbol{d}\in\mathbf{R}^n \mid \exists\delta>0,\lambda\in(0,\delta),f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<f(\bar{\boldsymbol{x}})\} \tag{5-4}$$

称为下降方向集。

在定义 5-1 中，参数 $\delta>0$ 只需满足存在性，一般不是唯一的，而且将步长限制在 $\lambda\in(0,\delta)$ 的范围是必要的。以函数 $f(\boldsymbol{x})=x_1^2+x_2^2$ 为例，如图 5-1 所示，当 $\lambda\in(0,\delta)$ 时，满足 $f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<f(\bar{\boldsymbol{x}})$，此时 $\boldsymbol{d}$ 是下降方向，但若 λ 过大(此例中若 $\lambda>\delta$)有 $f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})>f(\bar{\boldsymbol{x}})$，此时，$\boldsymbol{d}$ 并未使函数值下降。

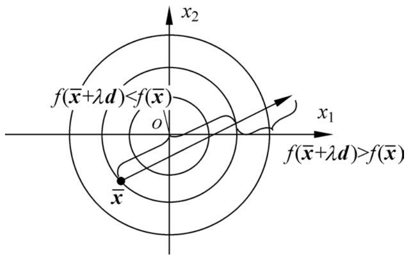

图 5-1　下降方向

如果 $f(\boldsymbol{x})$是可微函数，并且$\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d}<0$，则根据定理 3-1，$\boldsymbol{d}$ 为 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处的下降方向。将在 $\bar{\boldsymbol{x}}$ 点处满足这一条件的下降方向的全体记作

$$D_0=\{\boldsymbol{d}\in\mathbf{R}^n \mid \nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d}<0\} \tag{5-5}$$

显然，D_0 是下降方向集 D 的一个子集，即

$$D_0\subset D \tag{5-6}$$

5.1.2　可行方向

定义 5-2　对可行集 F，若 $\bar{\boldsymbol{x}}\in\mathrm{cl}F$，$\boldsymbol{d}$ 是非零向量，若存在数 $\delta>0$ 使对每个 $\lambda\in(0,\delta)$都有

$$\bar{\boldsymbol{x}}+\lambda\boldsymbol{d}\in F \tag{5-7}$$

则称 $\boldsymbol{d}$ 为可行集 F 在 $\bar{\boldsymbol{x}}$ 的可行方向，其中 cl 表示闭包。集合 F 在 $\bar{\boldsymbol{x}}$ 处的所有可行方向组成的集合记为

$$F_0=\{\boldsymbol{d}\in\mathbf{R}^n \mid \boldsymbol{d}\neq\mathbf{0},\bar{\boldsymbol{x}}\in\mathrm{cl}F,\exists\delta>0,\text{使得 }\forall\lambda\in(0,\delta),\text{有}\bar{\boldsymbol{x}}+\lambda\boldsymbol{d}\in F\} \tag{5-8}$$

集合 F_0 是一个锥，称为在 $\bar{\boldsymbol{x}}$ 处的可行方向锥。

5.1.3　几何描述

我们从几何的角度来描述约束优化问题的最优性条件。由下降方向和可行方向的定义可知，如果 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$在可行集上的局部极小点，则在 $\boldsymbol{x}^*$ 处的可行方向一定不是下降方向。

定理 5-1　考虑约束优化问题(5-1)，设 F 是 $\mathbf{R}^n$ 中的非空集合，$\boldsymbol{x}^*\in F$，$f(\boldsymbol{x})$在 $\boldsymbol{x}^*$ 处可微，如果 $\boldsymbol{x}^*$ 是局部最优解，则 $F_0\cap D=\varnothing$。

证明　用反证法，设存在 $\boldsymbol{d}\neq 0$ 满足

$$\boldsymbol{d}\in D\cap F_0 \tag{5-9}$$

则 $\boldsymbol{d}\in D$，由定义 5-1 存在 $\delta_1>0$，对任意 $\lambda\in(0,\delta_1)$，有

$$f(\boldsymbol{x}^*+\lambda\boldsymbol{d})<f(\boldsymbol{x}^*) \tag{5-10}$$

又有 $\boldsymbol{d}\in F_0$，由定义 5-2 存在 $\delta_2>0$，对任意 $\lambda\in(0,\delta_2)$，有

$$x^* + \lambda d \in F_0 \tag{5-11}$$

令 $\delta = \min\{\delta_1, \delta_2\}$，则对任意 $\lambda \in (0, \delta)$，有式(5-10)和式(5-11)同时成立，这和 x^* 是局部极小点矛盾。定理得证。

由于 $D_0 \subseteq D$，可以得到如下推论。

推论 5-1　考虑约束优化问题(5-1)，设 F 是 $\mathbf{R}^n$ 中的非空集合，$x^* \in F$，$f(x)$ 在 x^* 处可微，如果 x^* 是局部最优解，则 $F_0 \cap D_0 = \varnothing$。

从几何上来考虑，定理 5-1 的结论是显而易见的，以如下问题为例，

$$\begin{cases} \min & f(x) = x_1^2 + x_2^2 \\ \text{s. t.} & x_1 \geqslant 1, x_2 \geqslant 1 \end{cases}$$

显然，该问题的局部最优解是 $x^* = (1,1)^{\mathrm{T}}$，如图 5-2 所示。从 x^* 点出发，任何可行方向(例如 d_1)都使函数值增加；反之，任何能使函数值下降的方向(例如 d_2)都不是 x^* 处的可行方向，也就是说在 x^* 处 $F_0 \cap D = \varnothing$。

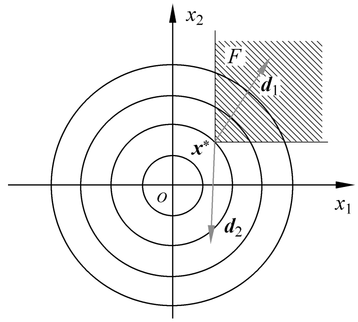

图 5-2　约束最优性条件的几何解释

5.2　不等式约束优化问题的一阶最优性条件

100min

5.2.1　用积极约束表达的几何最优性条件

本节讨论只含不等式约束的优化问题的一阶必要条件，考虑约束优化问题

$$\begin{cases} \min & f(x) \\ \text{s. t.} & g_i(x) \leqslant 0 \quad i = 1, 2, \cdots, m \end{cases} \tag{5-12}$$

其可行域为

$$F = \{x \subseteq \mathbf{R}^n \mid g_i(x) \leqslant 0, i = 1, 2, \cdots, m\} \tag{5-13}$$

为把推论 5-1 中描述的几何条件用代数表示出来，引入以下积极约束的概念。

定义 5-3　对于可行解 $\bar{x} \in F$，称满足条件 $g_i(\bar{x}) = 0$ 的那些约束为积极约束，其指标集记作

$$A(\bar{x}) = \{i \mid g_i(\bar{x}) = 0, i = 1, 2, \cdots, m\} \tag{5-14}$$

反之，称满足条件 $g_i(\bar{x}) < 0$ 的约束为在 $\bar{x}$ 处的非积极约束，其指标集记作

$$\bar{A}(\bar{x}) = \{i \mid g_i(\bar{x}) < 0, i = 1, 2, \cdots, m\} \tag{5-15}$$

从几何上看，积极约束其实是指 $\bar{x}$ 位于某些不等式约束所确定的区域的边界上，而非积极约束是指 $\bar{x}$ 位于某些约束所确定的区域的内部，如图 5-3 所示。对于可行

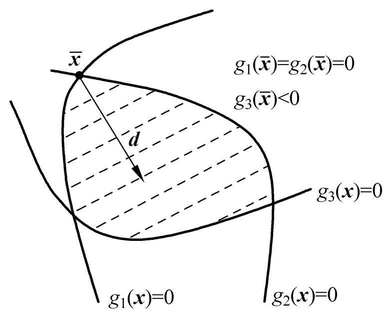

图 5-3　积极约束和非积极约束

解 $\bar{\boldsymbol{x}}$，$g_1(\bar{\boldsymbol{x}})$和 $g_2(\bar{\boldsymbol{x}})$是积极约束，即 $A(\bar{\boldsymbol{x}})=\{1,2\}$，而 $g_3(\bar{\boldsymbol{x}})$是非积极约束，即 $\bar{A}(\bar{\boldsymbol{x}})=\{3\}$。

定义积极约束后，定义 5-2 中的可行方向可以缩小范围，并用积极约束的梯度来表示，即

$$F_1=\{\boldsymbol{d}\in\mathbf{R}^n \mid \nabla g_i(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d}<0,\quad i\in A(\bar{\boldsymbol{x}})\} \tag{5-16}$$

图 5-3 中的 $\boldsymbol{d}$ 就是这样一个方向。

定理 5-2 对可行解 $\bar{\boldsymbol{x}}$，有 $F_1\subseteq F_0$。

证明 设方向 $\boldsymbol{d}\in F_1$，则对 $i\in A(\bar{\boldsymbol{x}})$有$\nabla g_i(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d}<0$，也就是说 $\boldsymbol{d}$ 是 $g_i(\boldsymbol{x})(i\in A(\bar{\boldsymbol{x}}))$在 $\bar{\boldsymbol{x}}$ 处的下降方向，由定义 5-1，存在 $\delta_1>0$，对任意 $\lambda\in(0,\delta_1)$，有

$$g_i(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<g_i(\bar{\boldsymbol{x}})=0,\quad i\in A(\bar{\boldsymbol{x}}) \tag{5-17}$$

又当 $i\in\bar{A}(\bar{\boldsymbol{x}})$时，$g_i(\bar{\boldsymbol{x}})<0$，由于 $g_i(\boldsymbol{x})(i\in\bar{A}(\bar{\boldsymbol{x}}))$在 $\bar{\boldsymbol{x}}$ 处连续，故存在 $\delta_2>0$，当 $\lambda\in(0,\delta_2)$时，

$$g_i(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<0,\quad i\in\bar{A}(\bar{\boldsymbol{x}}) \tag{5-18}$$

令 $\delta=\min\{\delta_1,\delta_2\}$，则对任意 $\lambda\in(0,\delta)$有

$$g_i(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})<0,\quad i=1,2,\cdots,m \tag{5-19}$$

即 $\boldsymbol{d}$ 是一个可行方向，故 $\boldsymbol{d}\in F_0$。定理得证。

根据推论 5-1 和定理 5-2，可以得到约束最优性条件的另一个几何表示。

推论 5-2 设 $\boldsymbol{x}^*\in F$，$f(\boldsymbol{x})$和 $g_i(\boldsymbol{x})(i\in A(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 处可微，$g_i(\boldsymbol{x})(i\in A(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 连续，如果 $\boldsymbol{x}^*$ 是问题(5-12)的局部最优解，则 $F_1\cap D_0=\varnothing$。

5.2.2 Fritz John 条件

先给出如下一个关于不等式的 Gordan 引理。

引理 5-1 (Gordan 引理)设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，那么，$\boldsymbol{Ax}<0$ 有解的充要条件是不存在非零向量 $\boldsymbol{y}\geqslant 0$，使 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=0$。

Gordan 引理的证明要用到凸集分离定理，比较烦琐，已超出本书范围，故在此不加赘述，感兴趣的读者可以参考文献[4]。

Gordan 引理中的条件是充分必要的，故也可以阐述成：$\boldsymbol{Ax}<0$ 无解的充要条件是存在非零向量 $\boldsymbol{y}\geqslant 0$，使 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=0$。下述的 Fritz John 条件的证明便是用的这一阐述。

将推论 5-2 的几何最优性条件用代数式表示，就可以得到下述 Fritz John 条件。

定理 5-3 (Fritz John 条件)设 $\boldsymbol{x}^*\in F$，$A(\boldsymbol{x}^*)=\{i\mid g_i(\boldsymbol{x}^*)=0,i=1,2,\cdots,m\}$和 $\bar{A}(\boldsymbol{x}^*)=\{i\mid g_i(\boldsymbol{x}^*)<0,i=1,2,\cdots,m\}$分别为 $\boldsymbol{x}^*$ 处的积极约束和非积极约束集，f，$g_i(i\in A(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 处可微，$g_i(i\in\bar{A}(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 处连续，如果 $\boldsymbol{x}^*$ 是问题(5-12)的局部最优解，则存在不全为 0 的非负数 ω_0，$\omega_i(i\in A(\boldsymbol{x}^*))$使

$$\omega_0\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)=\mathbf{0} \tag{5-20}$$

证明 根据推论 5-2，在点 $\boldsymbol{x}^*$ 有 $F_1\cap D_0=\varnothing$，即不等式组

$$\begin{cases} \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d} < 0 & i \in A(\boldsymbol{x}^*) \\ \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d} < 0 \end{cases} \tag{5-21}$$

无解，又根据 Gordan 引理，必存在不全为 0 的非负数 $\omega_0, \omega_i (i \in A(\boldsymbol{x}^*))$，使

$$\omega_0 \nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \omega_i \nabla g_i(\boldsymbol{x}^*) = \boldsymbol{0}$$

定理得证。

以下给出两个 Fritz John 条件的案例。

【例 5-1】 考虑约束优化问题

$$\begin{cases} \min & f(\boldsymbol{x}) = (x_1 - 6)^2 + (x_2 - 2)^2 \\ \text{s. t.} & g_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 16 \leqslant 0 \\ & g_2(\boldsymbol{x}) = x_1 + 4x_2 - 4 \leqslant 0 \\ & g_3(\boldsymbol{x}) = x_1 \geqslant 1 \\ & g_4(\boldsymbol{x}) = x_2 \geqslant -1 \end{cases}$$

显然，$\boldsymbol{x}^* = (4,0)^{\mathrm{T}}$ 是该问题的一个局部最优解（如图 5-4 所示），可以验证其满足 Fritz John 条件。

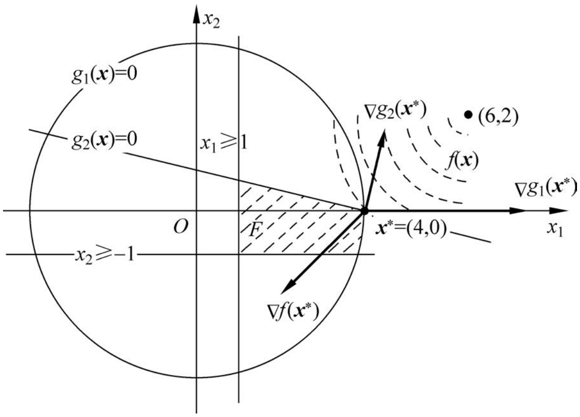

图 5-4 例 5-1 示意图

解 首先，由于 $g_1(\boldsymbol{x}^*) = \boldsymbol{0}$，$g_2(\boldsymbol{x}^*) = \boldsymbol{0}$，故 $A(\boldsymbol{x}^*) = \{1,2\}$，即在 $\boldsymbol{x}^*$ 处，g_1 和 g_2 是积极约束。又由

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 6) \\ 2(x_2 - 2) \end{bmatrix}, \quad \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \quad \nabla g_2(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

得

$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} -4 \\ -4 \end{bmatrix}, \quad \nabla g_1(\boldsymbol{x}^*) = \begin{bmatrix} 8 \\ 0 \end{bmatrix}, \nabla g_2(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

显然，存在 $\omega_0, \omega_i (i \in A(\boldsymbol{x}^*))$ 使式(5-20)成立。例如 $\omega_0 = 8, \omega_1 = 3, \omega_2 = 8$，即

$$8\nabla f(\boldsymbol{x}^*)+3\nabla g_1(\boldsymbol{x}^*)+8g_2(\boldsymbol{x}^*)=\boldsymbol{0}$$

如图 5-4 所示,因此,$\boldsymbol{x}^*=(4,0)^{\mathrm{T}}$ 满足 Fritz John 条件。

【例 5-2】 考虑约束优化问题

$$\begin{cases}\min & f(\boldsymbol{x})=-x_2\\ \text{s.t.} & g_1(\boldsymbol{x})=2x_1-(2-x_2)^3\leqslant 0\\ & g_2(\boldsymbol{x})=-x_1\leqslant 0\end{cases}$$

已知 $\boldsymbol{x}^*=(0,2)^{\mathrm{T}}$ 是该问题的最优点,试验证点 $\boldsymbol{x}^*$ 满足 Fritz John 条件。

解 首先,由于 $g_1(\boldsymbol{x}^*)=0,g_2(\boldsymbol{x}^*)=0$,故 $A(\boldsymbol{x}^*)=\{1,2\}$,即在 $\boldsymbol{x}^*$ 处,g_1 和 g_2 是积极约束。又由

$$\nabla f(\boldsymbol{x})=\begin{bmatrix}0\\-1\end{bmatrix},\quad \nabla g_1(\boldsymbol{x})=\begin{bmatrix}2\\3(2-x_2)^2\end{bmatrix},\quad \nabla g_2(\boldsymbol{x})=\begin{bmatrix}-1\\0\end{bmatrix}$$

得

$$\nabla f(\boldsymbol{x}^*)=\begin{bmatrix}0\\-1\end{bmatrix},\quad \nabla g_1(\boldsymbol{x}^*)=\begin{bmatrix}2\\0\end{bmatrix},\quad \nabla g_2(\boldsymbol{x}^*)=\begin{bmatrix}-1\\0\end{bmatrix}$$

显然,存在 $\omega_0,\omega_i(i\in A(\boldsymbol{x}^*))$使式(5-20)成立,例如 $w_0=0,\omega_1=1,\omega_2=2$,即

$$0\,\nabla f(\boldsymbol{x}^*)+\nabla g_1(\boldsymbol{x}^*)+2g_2(\boldsymbol{x}^*)=\boldsymbol{0}$$

如图 5-5 所示,因此 $\boldsymbol{x}^*=(0,2)^{\mathrm{T}}$ 满足 Firtz John 条件。

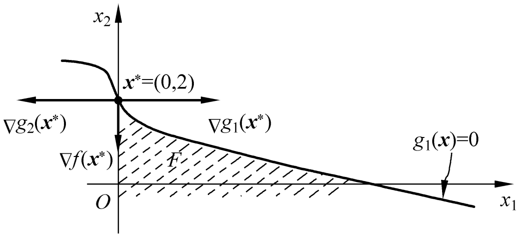

图 5-5 例 5-2 示意图

例 5-2 表明,运用 Fritz John 条件时,可能会出现 $\omega_0=0$ 的情形,这时,Fritz John 条件实际上不包含目标函数的任何数据,只是把积极约束的梯度组合成零向量。这样的条件,对于问题的描述是没有多少价值的。我们感兴趣的是 $\omega_0\neq 0$ 的情形。为了保证 $\omega_0\neq 0$,还需要对约束施加某种限制,这种限制条件通常称为约束规格(Constraint Qualification)。在定理 5-3 中,如果增加积极约束的梯度线性无关的约束规格,则给出不等式约束问题的著名的 KKT 条件。

5.2.3 Karush-Kuhn-Tucker(KKT)条件

定理 5-4 (KKT 条件)设 $\boldsymbol{x}^*\in F,\bar{A}(\boldsymbol{x}^*)=\{i\mid g_i(\boldsymbol{x}^*)=0,i=1,2,\cdots,m\}$和 $\bar{A}(\boldsymbol{x}^*)=\{i\mid g_i(\boldsymbol{x}^*)<0,i=1,2,\cdots,m\}$分别为 $\boldsymbol{x}^*$ 处的积极约束和非积极约束集,$f,g_i(i\in$

$A(\boldsymbol{x}^*)$)在 $\boldsymbol{x}^*$ 处可微，$g_i(i\in\bar{A}(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 处连续，并且$\{\nabla g_i(\boldsymbol{x}^*), i\in A(\boldsymbol{x}^*)\}$线性无关，如果 $\boldsymbol{x}^*$ 是问题(5-12)的局部最优解，则存在非负数 $\omega_i(i\in A(\boldsymbol{x}^*))$使

$$\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0} \tag{5-22}$$

证明　根据定理5-3，存在不全为0的非负数 $\omega_0, \bar{\omega}_i(i\in A(\boldsymbol{x}^*))$，使

$$\omega_0\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\bar{\omega}_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0}$$

显然，$\omega_0\neq 0$，否则由于 $\bar{\omega}_i(i\in A(\boldsymbol{x}^*))$不全为0，必然导致$\{\nabla g_i(\boldsymbol{x}^*), i\in A(\boldsymbol{x}^*)\}$线性相关，于是可令

$$\omega_i=\frac{\bar{\omega}_i}{\omega_0},\quad (i\in A(\boldsymbol{x}^*))$$

从而得到

$$\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0}$$

$$\omega_i\geqslant 0,\quad i\in A(\boldsymbol{x}^*)$$

定理得证。

在定理5-4中，若 $g_i(\boldsymbol{x}),(i\in\bar{A}(\boldsymbol{x}^*))$在 $\boldsymbol{x}^*$ 处可微，当 $i\in\bar{A}(\boldsymbol{x}^*)$，令 $\omega_i=0$，又注意到 $i\in A(\boldsymbol{x}^*)$时，$g_i(\boldsymbol{x}^*)=0$，则 KKT 条件可以等价地写成

$$\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\omega_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0} \tag{5-23}$$

$$\omega_i g_i(\boldsymbol{x}^*)=0,\quad i=1,2,\cdots,m \tag{5-24}$$

$$\omega_i\geqslant 0,\quad i=1,2,\cdots,m \tag{5-25}$$

其中，式(5-24)称为互补松弛条件。

对于约束优化问题(5-12)，式(5-23)和式(5-24)中的 $\boldsymbol{x}^*$ 和 $\omega_i, i=1,2,\cdots,m$ 均为未知的，所以式(5-23)和式(5-24)组成了一个含有 $m+n$ 个未知数，$m+n$ 个方程的方程组，通过求解该方程组得到的解 $\boldsymbol{x}^*$ 称为 KKT 点，KKT 点可能是最优解，也可能不是，但最优解一定是 KKT 点，所以要验证最优解 $\boldsymbol{x}^*$ 满足 KKT 条件，只需验证存在非负数 $\omega_i(i\in A(\boldsymbol{x}^*))$使式(5-22)成立。

以下给出 KKT 条件的两个应用案例。

【例 5-3】　考虑约束优化问题

$$\begin{cases}\min & f(\boldsymbol{x})=(x_1-3)^2+(x_2-1)^2\\ \text{s.t.} & g_1(\boldsymbol{x})=-x_1+x_2\leqslant 0\\ & g_2(\boldsymbol{x})=x_1^2-x_2\leqslant 0\end{cases}$$

试验证 $\boldsymbol{x}_1=(1,1)^{\mathrm{T}}$ 和 $\boldsymbol{x}_2=(0,0)^{\mathrm{T}}$ 是否为 KKT 点。

解　目标函数和约束函数的梯度函数为

$$\nabla f(\boldsymbol{x})=\begin{bmatrix}2(x_1-3)\\2(x_2-1)\end{bmatrix},\quad \nabla g_1(\boldsymbol{x})=\begin{bmatrix}-1\\1\end{bmatrix},\quad \nabla g_2(\boldsymbol{x})=\begin{bmatrix}2x_1\\-1\end{bmatrix}$$

先验证 $\boldsymbol{x}_1=(1,1)^{\mathrm{T}}$,在 $\boldsymbol{x}_1$ 处,由于 $g_1(\boldsymbol{x}_1)=0$,$g_2(\boldsymbol{x}_1)=0$,所以 g_1 和 g_2 均为点 $\boldsymbol{x}_1$ 处的积极约束,即 $A(\boldsymbol{x}_1)=\{1,2\}$,目标函数和积极约束在点 $\boldsymbol{x}_1$ 处的梯度为

$$\nabla f(\boldsymbol{x}_1)=\begin{bmatrix}-4\\0\end{bmatrix},\quad \nabla g_1(\boldsymbol{x}_1)=\begin{bmatrix}-1\\1\end{bmatrix},\quad \nabla g_2(\boldsymbol{x}_1)=\begin{bmatrix}2\\-1\end{bmatrix}$$

设$\nabla f(\boldsymbol{x}_1)+\omega_1\nabla g_1(\boldsymbol{x}_1)+\omega_2\nabla g_2(\boldsymbol{x}_1)=\mathbf{0}$,即

$$\begin{cases}-4-\omega_1+2\omega_2=0\\\omega_1-\omega_2=0\end{cases}$$

解此方程组,得到

$$\omega_1=4,\quad \omega_2=4$$

所以 $\boldsymbol{x}_1=(1,1)^{\mathrm{T}}$ 是 KKT 点。

再验证 $\boldsymbol{x}_2=(0,0)^{\mathrm{T}}$,在 $\boldsymbol{x}_2$ 处,由于 $g_1(\boldsymbol{x}_2)=0$,$g_2(\boldsymbol{x}_2)=0$,所以 g_1 和 g_2 均为点 $\boldsymbol{x}_2$ 处的积极约束,即 $A(\boldsymbol{x}_2)=\{1,2\}$,目标函数和积极约束在点 $\boldsymbol{x}_2$ 处的梯度为

$$\nabla f(\boldsymbol{x}_2)=\begin{bmatrix}-6\\-2\end{bmatrix},\quad \nabla g_1(\boldsymbol{x}_2)=\begin{bmatrix}-1\\1\end{bmatrix},\quad \nabla g_2(\boldsymbol{x}_2)=\begin{bmatrix}0\\-1\end{bmatrix}$$

设$\nabla f(\boldsymbol{x}_2)+\omega_1\nabla g_1(x_1)+\omega_2\nabla g_2(\boldsymbol{x}_2)=0$,即

$$\begin{cases}-6-\omega_1=0\\-2+\omega_1-\omega_2=0\end{cases}$$

解此方程组,得到

$$\omega_1=-6,\quad \omega_2=-8$$

由于 $\omega_1<0$,$\omega_2<0$,故 $\boldsymbol{x}_2=(0,0)^{\mathrm{T}}$ 不是 KKT 点。

例 5-3 的示意图如图 5-6 所示,可以看到 $\boldsymbol{x}_1=(1,1)^{\mathrm{T}}$ 其实是问题的最优解,故其必定为 KKT 点。

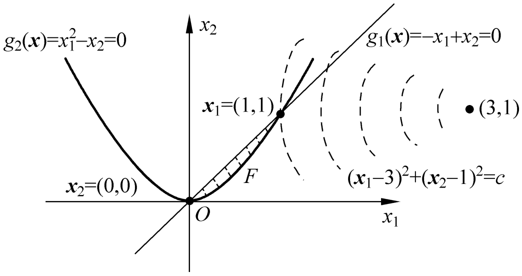

图 5-6　例 5-3 的示意图

【例 5-4】　考虑约束优化问题

$$\begin{cases}\min & f(\boldsymbol{x})=(x_1-2)^2+(x_2-3)^2 \\ \text{s.t.} & g_1(\boldsymbol{x})=-x_1+x_2-1\leqslant 0 \\ & g_2(\boldsymbol{x})=x_1-1\leqslant 0 \\ & g_3(\boldsymbol{x})=-x_2\leqslant 0\end{cases}$$

求满足 KKT 条件的点。

解　为求 KKT 点，需求解由式(5-23)和式(5-24)组成的方程组，目标函数和约束函数的梯度分别为

$$\nabla f(\boldsymbol{x})=\begin{bmatrix}2(x_1-2)\\2(x_2-3)\end{bmatrix},\quad \nabla g_1(\boldsymbol{x})=\begin{bmatrix}-1\\1\end{bmatrix},\quad \nabla g_2(\boldsymbol{x})=\begin{bmatrix}1\\0\end{bmatrix},\quad \nabla g_3(\boldsymbol{x})=\begin{bmatrix}0\\-1\end{bmatrix}$$

则 KKT 条件为

$$\begin{cases}\nabla f(\boldsymbol{x})+\omega_1\nabla g_1(\boldsymbol{x})+\omega_2\nabla g_2(\boldsymbol{x})+\omega_3\nabla g_3(\boldsymbol{x})=\boldsymbol{0}\\ \omega_i g_i(\boldsymbol{x})=0,\quad i=1,2,3\\ \omega_i\geqslant 0,\quad i=1,2,3\end{cases}$$

即

$$\begin{cases}2x_1-4-\omega_1+\omega_2=0\\ 2x_2-6+\omega_1-\omega_3=0\\ \omega_1(-x_1+x_2-1)=0\\ \omega_2(x_1-1)=0\\ -\omega_3x_3=0\end{cases}\tag{5-26}$$

式(5-26)是以 $x_1,x_2,\omega_1,\omega_2,\omega_3$ 为未知数的非线性方程组，问题归结为求这个方程组满足条件 $\omega_1\geqslant0,\omega_2\geqslant0,\omega_3\geqslant0$ 的解。一般来讲，求解非线性方程组是比较困难的，但上述方程组可以结合问题的图像来求解。

从图 5-7 可以看出 $\boldsymbol{x}^*=(1,2)^{\mathrm{T}}$ 是最优解，所以不妨设 $x_1=1,x_2=2$，将其代入式(5-26)中，可以得到

$$\omega_1=2,\quad \omega_2=4,\quad \omega_3=0$$

满足 $\omega_i\geqslant0,i=1,2,3$，说明 $\boldsymbol{x}=(1,2)^{\mathrm{T}}$ 是一个 KKT 点。

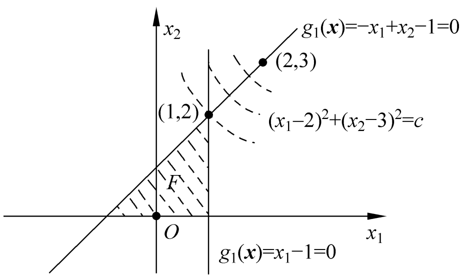

图 5-7　例 5-4 示意图

对 $\boldsymbol{x}=(1,2)^{\mathrm{T}}$，有 $g_1(\boldsymbol{x})=0,g_2(\boldsymbol{x})=0$，$g_3(\boldsymbol{x})=-2$，所以 g_1 和 g_2 是点 $\boldsymbol{x}$ 处的积极约束，而且满足互补松弛条件

$$\omega_i g_i(\boldsymbol{x})=0,\quad i=1,2,3$$

对于一般不等式约束优化问题，KKT 条件只是必要而非充分的，但对凸规划问题，KKT 条件是充分必要的。

定理 5-5 在问题(5-12)中,若 $f, g_i(i=1,2,\cdots,m)$ 是凸函数,设 $\boldsymbol{x}^* \in F, A(\boldsymbol{x}^*)=\{i \mid g_i(\boldsymbol{x}^*)=0, i=1,2,\cdots,m\}$ 和 $\bar{A}(\boldsymbol{x}^*)=\{i \mid g_i(\boldsymbol{x})<0, i=1,2,\cdots,m\}$ 分别是 $\boldsymbol{x}^*$ 处的积极约束和非积极约束,f 和 $g_i(i \in A(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 处可微,$g_i(i \in \bar{A}(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 处连续,并且 $\nabla g_i(\boldsymbol{x}^*)(i \in A(\boldsymbol{x}^*))$ 线性无关,则 $\boldsymbol{x}^*$ 是问题(5-12)的全局最优解的充分必要条件是在 $\boldsymbol{x}^*$ 处 KKT 条件成立。

证明 必要性在定理 5-4 中已证明,以下证明充分条件。

根据定理假设,显然 F 是凸集,f 是凸函数,故问题(5-12)为凸规划问题,由于 f 是凸函数,并且在点 $\boldsymbol{x}^*$ 处可微,根据定理 1-7,对任意的 $\boldsymbol{x} \in F$,有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*)+\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \tag{5-27}$$

又知在点 $\boldsymbol{x}^*$ 处 KKT 条件成立,即存在非负的 KKT 乘子 $\omega_i \geqslant 0(i \in \boldsymbol{x}^*)$,使

$$\nabla f(\boldsymbol{x}^*)=-\sum_{i \in A(\boldsymbol{x}^*)} \omega_i \nabla g_i(\boldsymbol{x}^*) \tag{5-28}$$

将式(5-28)代入式(5-27),得到

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*)-\sum_{i \in A(\boldsymbol{x}^*)} \omega_i \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \tag{5-29}$$

由于 $g_i,(i \in A(\boldsymbol{x}^*))$ 是凸函数,并且在 $\boldsymbol{x}^*$ 处可微,则

$$g_i(\boldsymbol{x}) \geqslant g_i(\boldsymbol{x}^*)+\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*), \quad i \in A(\boldsymbol{x}^*) \tag{5-30}$$

故

$$-\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \geqslant g_i(\boldsymbol{x}^*)-g_i(\boldsymbol{x}), \quad i \in A(\boldsymbol{x}^*) \tag{5-31}$$

由于当 $i \in A(\boldsymbol{x}^*)$ 时,$g_i(\boldsymbol{x}) \leqslant 0, g_i(\boldsymbol{x}^*)=0$,因此有

$$-\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \geqslant 0, \quad i \in A(\boldsymbol{x}^*) \tag{5-32}$$

根据式(5-29)和式(5-32),显然成立

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*) \tag{5-33}$$

即 $\boldsymbol{x}^*$ 是问题(5-12)的全局最优解。

根据定理 5-5,对于凸规划问题,可以用 KKT 条件求得全局最优解,例如例 5-4 中的 $\boldsymbol{x}^*=(1,2)^{\mathrm{T}}$ 就是该问题的全局最优解。

5.3 一般约束优化问题的一阶最优性条件

100min

本节考虑同时带有不等式约束和等式约束的优化问题。

5.3.1 几何最优性条件

考虑一般约束优化问题

$$\begin{cases} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \leqslant 0, \quad i=1,2,\cdots,m \\ & h_j(\boldsymbol{x})=0, \quad j=1,2,\cdots,l \end{cases} \tag{5-34}$$

先重申一些基本概念。

1. 可行集

$$F=\{\boldsymbol{x}\in\mathbf{R}^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,m;\ h_j(\boldsymbol{x})=0, j=1,2,\cdots,l\} \tag{5-35}$$

2. $x\in F$ 点处的积极约束指标集

$$A(\boldsymbol{x})=\{i \mid g_i(\boldsymbol{x})=0, i=1,2,\cdots,m\} \tag{5-36}$$

3. $x\in F$ 点处的非积极约束指标集

$$\overline{A}(\boldsymbol{x})=\{i \mid g_i(\boldsymbol{x})<0, i=1,2,\cdots,m\} \tag{5-37}$$

4. 正则点

定义 5-4　设 $\boldsymbol{x}\in F$，如果向量组 $\{\nabla g_i(\boldsymbol{x}),\nabla h_j(\boldsymbol{x})\,|\,i\in A(\boldsymbol{x}), j=1,2,\cdots,l\}$ 线性无关，就称 $\boldsymbol{x}$ 为约束 $g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,m$ 和 $h_j(\boldsymbol{x})=0, j=1,2,\cdots,l$ 的正则点。

下面将在正则点的前提下，介绍最优解的必要条件。与5.2节思路一样，我们仍然先给出从几何上表达的最优性条件，再将其转换为代数形式。几何上的描述通俗地讲就是满足可行方向集和下降方向集的交集为空的原则。对问题(5-34)，下降方向仍为式(5-5)定义的 D_0，难点在于对可行方向的描述，当 $h_j(\boldsymbol{x}), j=1,2,\cdots,l$ 为非线性函数时，在任何可行点均不存在可行方向，沿任何方向取微小步长都将破坏其可行性，因此，就等式约束 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}$（这里 $\boldsymbol{h}(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x}))^{\mathrm{T}}$，下同）而言，为描述可行移动，需要考虑超曲面 $S=\{\boldsymbol{x}\,|\,\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}\}$ 上的可行曲线。

定义 5-5　点集 $\{\boldsymbol{x}=\boldsymbol{x}(t)\,|\,t_0\leqslant t\leqslant t_1\}$ 称为曲面 $S=\{\boldsymbol{x}\,|\,\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}\}$ 上的一条曲线，如果对所有 $t\in[t_0,t_1]$，则均有 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}$。

显然，曲线上的点是参数 t 的函数，如果导数 $\boldsymbol{x}'(t)=\mathrm{d}\boldsymbol{x}(t)/\mathrm{d}t$ 存在，则称曲线是可微的。曲线 $\boldsymbol{x}(t)$ 的一阶导函数 $\boldsymbol{x}'(t)$ 是曲线在点 $\boldsymbol{x}=\boldsymbol{x}(t)$ 处的切向量，曲面 S 上经过点 $\boldsymbol{x}$ 的所有可微曲线在 $\boldsymbol{x}$ 处的切向量组成的集合，称为曲面 S 在点 $\boldsymbol{x}$ 的切平面，记作 $T(\boldsymbol{x})$。

切平面 $T(\boldsymbol{x})$ 并没有显式的表达式，为了便于表达，定义子空间

$$H(\boldsymbol{x})=\{\boldsymbol{d} \mid \nabla\boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}=0\} \tag{5-38}$$

一般情况下，切平面 $T(\boldsymbol{x})$ 是子空间 $H(\boldsymbol{x})$ 的子集，即 $T(\boldsymbol{x})\subseteq H(\boldsymbol{x})$，但若 $\boldsymbol{x}$ 是约束 $h_j(\boldsymbol{x})=0$，$j=1,2,\cdots,l$ 的正则点，则反之也成立。

定理 5-6　设 $\boldsymbol{x}$ 是曲面 $S=\{\boldsymbol{x}\,|\,\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}\}$ 上的一个正则点，则在点 x 处的切平面 $T(\boldsymbol{x})$ 等于子空间 $H(\boldsymbol{x})$，即 $T(\boldsymbol{x})=H(\boldsymbol{x})$。

定理5-6的证明需要用到隐函数定理，比较烦琐，超出了本书的范围，我们在此先承认它，在此基础上进行后续的讨论。

定理5-6的意义在于给出了切平面的一个几何表示，即

$$T(\boldsymbol{x})=\{\boldsymbol{d} \mid \nabla h_j(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}=0, j=1,2,\cdots,l\} \tag{5-39}$$

根据这一表示，我们给出以下集合最优性条件。

定理 5-7　设在约束优化问题(5-34)中，$\boldsymbol{x}^*\in F$，f 和 $g_i(i\in A(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 可微，$g_i(i\in\overline{A}(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 连续，$h_j(j=1,2,\cdots,l)$ 在点 $\boldsymbol{x}^*$ 连续可微，并且 $\nabla h_1(\boldsymbol{x}^*)$，$\nabla h_2(\boldsymbol{x}^*)$，$\cdots$，

$\nabla h_l(\boldsymbol{x}^*)$线性无关，如果 $\boldsymbol{x}^*$ 是局部最优解，则在 $\boldsymbol{x}^*$ 处有

$$F_1 \cap D_0 \cap H = \varnothing \tag{5-40}$$

其中，F_1 由式(5-16)确定，D_0 由式(5-5)确定，H 由式(5-38)确定。

证明 用反证法，设 $F_1 \cap D_0 \cap H \neq \varnothing$，即存在向量 $\boldsymbol{y} \in F_1 \cap D_0 \cap H$，也就是 $\boldsymbol{y}$ 使

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} < 0 \tag{5-41}$$

$$\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} < 0, \quad i \in A(\boldsymbol{x}^*) \tag{5-42}$$

$$\nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} = 0, \quad j = 1, 2, \cdots, l \tag{5-43}$$

同时成立，根据定理 5-6，$\boldsymbol{y} \in T(\boldsymbol{x}^*)$，即在曲面 $S = \{\boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$上存在经过点 $\boldsymbol{x}^* = \boldsymbol{x}(0)$ 的可微曲线 $\boldsymbol{x}(t)$，其在点 $\boldsymbol{x}^*$ 处的切向量为 $\boldsymbol{y} = \mathrm{d}\boldsymbol{x}(0)/\mathrm{d}t$。下面证明当 $t > 0$ 充分小时，$\boldsymbol{x}(t)$为可行点，并且 $f(\boldsymbol{x}(t)) < f(\boldsymbol{x}^*)$。

当 $i \in A(\boldsymbol{x}^*)$时，

$$\left.\frac{\mathrm{d}g_i(\boldsymbol{x}(t))}{\mathrm{d}t}\right|_{t=0} = \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}} \left.\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\right|_{t=0} = \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} < 0$$

因此，存在 $\delta_1 > 0$，当 $t \in [0, \delta_1)$时

$$g_i(\boldsymbol{x}(t)) \leqslant 0, \quad \forall\, i \in A(\boldsymbol{x}^*) \tag{5-44}$$

当 $i \in \bar{A}(\boldsymbol{x}^*)$时，由于 $g_i(\boldsymbol{x}^*) < 0$，并且 g_i 在 $\boldsymbol{x}^*$ 连续，因此，存在 $\delta_2 > 0$，当 $t \in [0, \delta_2)$ 时，有

$$g_i(\boldsymbol{x}(t)) \leqslant 0, \quad \forall\, i \in \bar{A}(\boldsymbol{x}^*) \tag{5-45}$$

另一方面，由于

$$\left.\frac{\mathrm{d}f(\boldsymbol{x}(t))}{\mathrm{d}t}\right|_{t=0} = \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \left.\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\right|_{t=0} = \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} < 0$$

因此存在 $\delta_3 > 0$，当 $t \in [0, \delta_3)$时，有

$$f(\boldsymbol{x}(t)) < f(\boldsymbol{x}^*) \tag{5-46}$$

令 $\delta = \min\{\delta_1, \delta_2, \delta_3\}$，则 $t \in [0, \delta)$时，式(5-44)～式(5-46)同时成立，并且 $h_j(\boldsymbol{x}(t)) = 0$，$j = 1, 2, \cdots, l$，因此，当 $t \in [0, \delta)$时，$\boldsymbol{x}(t)$为可行点，并且 $f(\boldsymbol{x}(t)) < f(\boldsymbol{x}^*)$，这个结果与 $\boldsymbol{x}^*$ 是局部最优解矛盾，因此有

$$F_1 \cap D_0 \cap H = \varnothing$$

5.3.2 Fritz John 必要条件

下面给出一阶必要条件的代数表达。

定理 5-8 (Fritz John 条件)在问题(5-34)中，设 $\boldsymbol{x}^* \in F$，$A(\boldsymbol{x}^*)$和$\bar{A}(\boldsymbol{x}^*)$分别为积极约束集和非积极约束集，$f$ 和 $g_i(i \in A(\boldsymbol{x}^*))$在点 $\boldsymbol{x}^*$ 可微，$g_i(i \in \bar{A}(\boldsymbol{x}^*))$在点 $\boldsymbol{x}^*$ 连续，$h_j(j = 1, 2, \cdots, l)$在点 $\boldsymbol{x}^*$ 连续可微，如果 $\boldsymbol{x}^*$ 是局部最优解，则存在不全为 0 的数 ω_0，$\omega_i(i \in A(\boldsymbol{x}^*))$和 $\nu_j(j = 1, 2, \cdots, l)$使

$$\omega_0 \nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \omega_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{l} \nu_j \nabla h_j(\boldsymbol{x}^*) = \mathbf{0} \tag{5-47}$$

证明 如果$\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \cdots, \nabla h_l(\boldsymbol{x}^*)$线性相关,则存在不全为 0 的数 $\nu_j(j=1,2,\cdots,l)$使

$$\sum_{j=1}^{l} \nu_j \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}$$

这时,可令 $\omega_0=0, \omega_i=0(i \in A(\boldsymbol{x}^*))$,则得出定理的结论。

如果$\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \cdots, \nabla h_l(\boldsymbol{x}^*)$线性无关,则满足定理 5-7 的条件,必有

$$F_1 \cap D_0 \cap H = \varnothing$$

即不等式

$$\begin{cases} \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} < 0 \\ \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} < 0 & i \in A(\boldsymbol{x}^*) \\ \nabla \boldsymbol{h}_j(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} = 0 & j = 1,2,\cdots,l \end{cases} \tag{5-48}$$

无解。

令 $\boldsymbol{A}$ 是以$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}$ 和$\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(i \in A(\boldsymbol{x}^*))$为行组成的矩阵,$\boldsymbol{B}$ 是以$\nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}}(j=1,2,\cdots,l)$为行组成的矩阵,这样系统(5-48)无解,也就是系统

$$\begin{cases} \boldsymbol{Ad} < \mathbf{0} \\ \boldsymbol{Bd} = \mathbf{0} \end{cases}$$

无解。

现在定义两个集合

$$S_1 = \left\{ \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \middle| \boldsymbol{y}_1 = \boldsymbol{Ad}, \boldsymbol{y}_2 = \boldsymbol{Bd}, \boldsymbol{d} \in \mathbf{R}^n \right\}$$

和

$$S_2 = \left\{ \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \middle| \boldsymbol{y}_1 < \mathbf{0}, \boldsymbol{y}_2 = \mathbf{0} \right\}$$

显然,S_1 和 S_2 均为非空凸集,并且

$$S_1 \cap S_2 = \varnothing$$

根据定理 1-4,存在非零向量

$$\boldsymbol{p} = \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \end{bmatrix}$$

使对每个 $\boldsymbol{d} \in \mathbf{R}^n$ 及每点

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} \in \mathrm{cl} S_2$$

成立

$$\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{Ad}+\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{Bd}\geqslant \boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{y}_1+\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{y}_2 \tag{5-49}$$

令 $\boldsymbol{y}_2=\boldsymbol{0}$,由于 $\boldsymbol{y}_1$ 的每个分量均可为任意负数,因此式(5-49)成立蕴含着

$$\boldsymbol{p}_1\geqslant 0 \tag{5-50}$$

再令

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}\in \mathrm{cl}S_2$$

则由式(5-49)的成立又蕴含着

$$\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{Ad}+\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{Bd}\geqslant 0 \tag{5-51}$$

由于 $\boldsymbol{d}\in\mathbf{R}^n$,可取任何向量,我们令

$$\boldsymbol{d}=-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p}_1+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{p}_2) \tag{5-52}$$

代入式(5-51)得到

$$-\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p}_1+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{p}_2\|^2\geqslant 0$$

由此可知

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p}_1+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{p}_2=\boldsymbol{0} \tag{5-53}$$

把 $\boldsymbol{p}_1$ 的分量记作 ω_0 和 $\omega_i(i\in A(x^*))$,把 $\boldsymbol{p}_2$ 的分量记作 $\nu_j(j=1,2,\cdots,l)$,则式(5-50)和式(5-53)即为

$$\omega_0\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{l}\nu_j\nabla h_j(\boldsymbol{x}^*)=\boldsymbol{0}$$

$$\omega_0,\omega_i\geqslant 0,\quad i\in A(\boldsymbol{x}^*)$$

由于 $\boldsymbol{p}$ 是非零向量,因此数 $\omega_0,\omega_i(i\in A(\boldsymbol{x}^*))$及 $\nu_j(j=1,2,\cdots,l)$不全为 0。

以下给出一个 Fritz John 必要条件的例子。

【例 5-5】 考虑非线性约束优化问题

$$\begin{cases}\min & f(\boldsymbol{x})=x_1^2+(x_2-3)^2\\ \text{s.t.} & g_1(\boldsymbol{x})=(x_1+1)^2+(x_2-1)^2-1\leqslant 0\\ & g_2(\boldsymbol{x})=(x_1-1)^2+(x_2-1)^2-1=0\end{cases}$$

这个问题的唯一可行点是 $\boldsymbol{x}^*=(0,1)^{\mathrm{T}}$,验证在点 $\boldsymbol{x}^*$ 处满足 Fritz John 条件。

解 由于 $g_1(\boldsymbol{x}^*)=0$,故积极约束集为 $A(\boldsymbol{x}^*)=\{1\}$,目标函数和约束函数的梯度分别为

$$\nabla f(\boldsymbol{x})=\begin{bmatrix}2x_1\\ 2(x_2-3)\end{bmatrix},\quad \nabla g_1(\boldsymbol{x})=\begin{bmatrix}2(x_1+1)\\ 2(x_2-1)\end{bmatrix},\quad \nabla g_2(\boldsymbol{x})=\begin{bmatrix}2(x_1-1)\\ 2(x_2-1)\end{bmatrix}$$

故在点 $\boldsymbol{x}^*$ 处

$$\nabla f(\boldsymbol{x}^*)=\begin{bmatrix}0\\ -4\end{bmatrix},\quad \nabla g_1(\boldsymbol{x}^*)=\begin{bmatrix}2\\ 0\end{bmatrix},\quad \nabla g_2(\boldsymbol{x}^*)=\begin{bmatrix}-2\\ 0\end{bmatrix}$$

设

$$\omega_0 \nabla f(\boldsymbol{x}^*) + \omega_1 \nabla g_1(\boldsymbol{x}^*) + \nu_1 \nabla g_2(\boldsymbol{x}^*) = \boldsymbol{0}$$

则

$$\begin{cases} 2\omega_1 - 2\nu_1 = 0 \\ -4\omega_0 = 0 \end{cases}$$

得到 $\omega_0 = 0, \omega_1 = \nu_1 = k$，其中 k 可取任何数，因此在点 $\boldsymbol{x}^*$ 处 Fritz John 条件成立，如图 5-8 所示。

例 5-5 表明，在 Fritz John 条件中，不排除目标函数梯度的系数 ω_0 等于 0 的情形，为了保证 ω_0 不等于 0，需给约束条件施加某种限制（称为约束规格），从而给出一般约束问题的 KKT 必要条件。

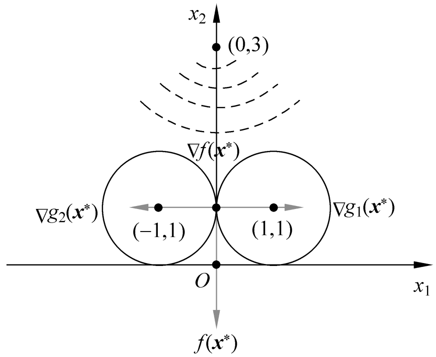

图 5-8　例 5-5 示意图

5.3.3　KKT 必要条件

定理 5-9　（KKT 必要条件）设在问题(5-34)中，$\boldsymbol{x}^* \in F$，$A(\boldsymbol{x}^*)$ 和 $\bar{A}(\boldsymbol{x}^*)$ 分别为 $\boldsymbol{x}^*$ 处的积极约束和非积极约束。f 和 $g_i\,(i \in A(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 处可微，$g_i\,(i \in \bar{A}(\boldsymbol{x}^*))$ 在点 $\boldsymbol{x}^*$ 处连续，$h_j\,(j=1,2,\cdots,l)$ 在点 $\boldsymbol{x}^*$ 处连续可微，向量集合

$$\{\nabla g_i(\boldsymbol{x}^*), \nabla h_j(\boldsymbol{x}^*) \mid i \in A(\boldsymbol{x}^*), j = 1,2,\cdots,l\} \tag{5-54}$$

线性无关，如果 $\boldsymbol{x}^*$ 是局部最优解，则存在数 $\omega_i\,(i \in A(\boldsymbol{x}^*))$ 和 $\nu_j\,(j=1,2,\cdots,l)$ 使

$$\begin{aligned} &\nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \omega_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{l} \nu_j h_j(\boldsymbol{x}^*) = \boldsymbol{0} \\ &\omega_i \geqslant 0, \quad i \in A(\boldsymbol{x}^*) \end{aligned} \tag{5-55}$$

证明　根据定理 5-8，存在不全为 0 的数 ω_0，$\bar{\omega}_i\,(i \in A(\boldsymbol{x}^*))$ 和 $\bar{\nu}_j\,(j=1,2,\cdots,l)$，使

$$\begin{aligned} &\omega_0 \nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \bar{\omega}_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{l} \bar{\nu}_j h_j(\boldsymbol{x}^*) = \boldsymbol{0} \\ &\omega_0, \bar{\omega}_i \geqslant 0 \quad (i \in A(\boldsymbol{x}^*)) \end{aligned} \tag{5-56}$$

若 $\omega_0 = 0$，由于向量组

$$\{\nabla g_i(\boldsymbol{x}^*), \nabla h_j(\boldsymbol{x}^*) \mid i \in A(\boldsymbol{x}^*), j = 1,2,\cdots,l\}$$

线性无关，则由式(5-56)必然有

$$\bar{\omega}_i = 0, \bar{\nu}_j = 0, i \in A(\boldsymbol{x}^*), j = 1,2,\cdots,l$$

这与定理 5-8 矛盾，故一定有 $\omega_0 \neq 0$，令

$$\omega_i = \frac{\bar{\omega}_i}{\omega_0}, \quad i \in A(\boldsymbol{x}^*)$$

$$\nu_j = \frac{\bar{\nu}_j}{\omega_0}, \quad j = 1,2,\cdots,l$$

于是得到

$$\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{l}\nu_j h_j(\boldsymbol{x}^*)=\boldsymbol{0}$$

$$\omega_i\geqslant 0,\quad i\in A(\boldsymbol{x}^*)$$

定理得证。

与只有不等式约束的情形相似,当 $g_i(i\in\overline{A}(\boldsymbol{x}^*))$在点 $\boldsymbol{x}^*$ 也可微时,令其相应的乘子 $\omega_i(i\in\overline{A}(\boldsymbol{x}^*))$等于 0,于是可以将式(5-55)写成

$$\begin{aligned}&\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\omega_i\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{l}\nu_j h_j(\boldsymbol{x}^*)=\boldsymbol{0}\\&\omega_i g_i(\boldsymbol{x}^*)=0,\quad i=1,2,\cdots,m\\&\omega_i\geqslant 0,\quad i=1,2,\cdots,m\end{aligned}\tag{5-57}$$

其中 $\omega_i g_i(\boldsymbol{x}^*)=0(i=1,2,\cdots,m)$仍称为互补松弛条件。

针对问题 5-34,定义广义的拉格朗日函数

$$L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})=f(\boldsymbol{x})+\sum_{i=1}^{m}\omega_i g_i(\boldsymbol{x})+\sum_{j=1}^{l}\nu_j h_j(\boldsymbol{x}^*)\tag{5-58}$$

于是,在定理 5-9 的条件下,若 $\boldsymbol{x}^*$ 为问题(5-34)的局部最优解,则存在乘子向量$\boldsymbol{\omega}^*\geqslant 0$ 和 $\boldsymbol{\nu}^*$,使

$$\nabla_x L(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)=\boldsymbol{0}\tag{5-59}$$

这样,KKT 乘子$\boldsymbol{\omega}^*$ 和$\boldsymbol{\nu}^*$ 也称为拉格朗日乘子。此时,一般情形的一阶必要条件可以表示为

$$\begin{cases}\nabla_{\boldsymbol{x}}L(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)=0\\g_i(\boldsymbol{x}^*)\leqslant 0 & i=1,2,\cdots,m\\h_j(\boldsymbol{x}^*)=0 & j=1,2,\cdots,l\\\omega_i^* g_i(\boldsymbol{x}^*)=0 & i=1,2,\cdots,m\\\omega_i^*\geqslant 0 & i=1,2,\cdots,m\end{cases}\tag{5-60}$$

定理 5-9 是最优解的必要条件,但是在凸规划的假设下,定理 5-9 的条件也是充分的。

定理 5-10 设在问题(5-34)中,$f,g_i(i=1,2,\cdots,m)$是凸函数,$h_j(j=1,2,\cdots,m)$是线性函数,$\boldsymbol{x}^*\in F$,$A(\boldsymbol{x}^*)$和 $A(\boldsymbol{x}^*)$分别为积极约束集和非积极约束集,向量集

$$\{\nabla f_i(\boldsymbol{x}^*),\nabla h_j(\boldsymbol{x}^*)\mid i\in A(\boldsymbol{x}^*),j=1,2,\cdots,l\}$$

线性无关,则 $\boldsymbol{x}^*$ 是问题(5-34)的全局最优解的充分必要条件为在 $\boldsymbol{x}^*$ 处 KKT 条件成立,即存在 $\omega_i\geqslant 0(i\in A(\boldsymbol{x}^*))$及 $\nu_j(j=1,2,\cdots,l)$使

$$\nabla f(\boldsymbol{x}^*)+\sum_{i\in A(\boldsymbol{x}^*)}\omega_i\nabla g_i(\boldsymbol{x}^*)+\sum_{i=1}^{j}\nu_j\nabla h_j(\boldsymbol{x}^*)=\boldsymbol{0}\tag{5-61}$$

证明 必要条件由凸函数的可微性和线性函数的连续性即定理 5-9 可以得到。下面证明充分性。

由定理的假设易知,可行域 F 是凸集,又由目标函数是凸函数,因此问题(5-34)是凸规

划问题。

由于 f 是凸函数，并且在 $\boldsymbol{x}^* \in F$ 可微，根据定理 1-7，对任意的 $\boldsymbol{x} \in F$ 有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \tag{5-62}$$

又由于 $g_i(i \in A(\boldsymbol{x}^*))$ 是凸函数，并且在 $\boldsymbol{x}^* \in F$ 可微，则

$$g_i(\boldsymbol{x}) \geqslant g_i(\boldsymbol{x}^*) + \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*), \quad i \in A(\boldsymbol{x}^*)$$

由于 $\boldsymbol{x} \in F$，故 $g_i(\boldsymbol{x}) \geqslant 0$，$h_j(\boldsymbol{x}) = 0$，因此

$$\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \leqslant 0, \quad i \in A(\boldsymbol{x}^*) \tag{5-63}$$

由于 $h_j(j = 1, 2, \cdots, l)$ 是线性函数，必有

$$h_j(\boldsymbol{x}) = h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*)$$

又因为 $\boldsymbol{x}$ 和 $\boldsymbol{x}^*$ 是可行点，满足

$$h_j(\boldsymbol{x}) = h_j(\boldsymbol{x}^*) = 0, \quad j = 1, 2, \cdots, l$$

于是有

$$\nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0, \quad j = 1, 2, \cdots, l \tag{5-64}$$

将已知条件式(5-61)代入式(5-62)，并注意到式(5-63)和式(5-64)，以及 $\omega_i \geqslant 0(i \in A(\boldsymbol{x}^*))$，得

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*)$$

故 $\boldsymbol{x}^*$ 为全局最优解。

【例 5-6】 求问题

$$\begin{cases} \min & f(\boldsymbol{x}) = (x_1 - 1)^2 + (x_2 - 3)^2 \\ \text{s.t.} & g_1(\boldsymbol{x}) = (x_1 - 1)^2 + x_2^2 - 1 \leqslant 0 \\ & g_2(\boldsymbol{x}) = x_1 + x_2 - 2 \leqslant 0 \end{cases}$$

的最优解。

解　由于 f 和 g_1 是凸函数，g_2 是线性函数，所以该问题是凸规划，定理 5-10 适用于该问题。

目标函数和约束函数的梯度分别为

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 3) \end{bmatrix}, \quad \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 1) \\ 2x_2 \end{bmatrix}, \quad \nabla g_2(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

根据 KKT 条件(5-60)，该问题最优解的充分必要条件为

$$\begin{cases} 2(x_1 - 1) + 2\omega_1(x_1 - 1) + \omega_2 = 0 \\ 2(x_2 - 3) + 2\omega_1 x_2 + \omega_2 = 0 \\ \omega_1((x_1 - 1)^2 + x_2^2 - 1) = 0 \\ \omega_2(x_1 + x_2 - 2) = 0 \\ (x_1 - 1)^2 + x_2^2 - 1 \leqslant 0 \\ x_1 + x_2 - 2 \leqslant 0 \\ \omega_1, \omega_2 \geqslant 0 \end{cases}$$

求解该方程得

$$x_1=1,\quad x_2=1,\quad \omega_1=2,\quad \omega_2=0$$

因此 $\boldsymbol{x}^*=(1,1)^{\mathrm{T}}$ 是该问题的全局最优解,如图 5-9 所示。

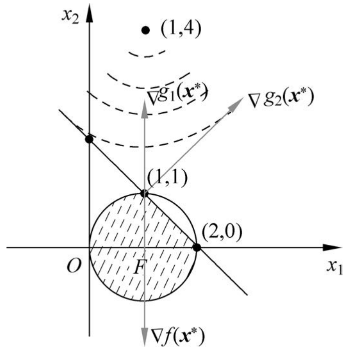

图 5-9 例 5-6 示意图

82min

5.4 对偶问题及鞍点最优性条件

对偶理论是优化理论的重要组成部分,考虑对偶问题可以更深刻地理解最优化原理,某些对偶问题也可以让原问题的求解变得简单。根据对偶函数的不同构造方法,一个问题可以有不同的对偶问题,本节通过著名的拉格朗日(Lagrange)对偶来进一步理解最优性条件。

本节首先介绍拉格朗日对偶问题的定义,然后不加证明地介绍对偶原理,最后介绍基于拉格朗日对偶的鞍点最优性条件及其与前述的 KKT 最优性条件的关系。

5.4.1 拉格朗日对偶问题

考虑一般约束优化问题

$$\begin{cases}\min & f(\boldsymbol{x})\\ \text{s. t.} & g_i(\boldsymbol{x})\leqslant 0, i=1,2,\cdots,m\\ & h_j(\boldsymbol{x})=0, j=1,2,\cdots,l\\ & \boldsymbol{x}\in B\end{cases}\tag{5-65}$$

其中,B 为箱子集,也可以将其整合到不等式约束中。将问题(5-65)视作原问题,定义它的对偶问题为

$$\begin{cases}\max & L(\boldsymbol{\omega},\boldsymbol{\nu})\\ \text{s. t.} & \boldsymbol{\omega}\geqslant 0\end{cases}\tag{5-66}$$

其中,目标为

$$L(\boldsymbol{\omega},\boldsymbol{\nu})=\inf_{\boldsymbol{x}}\left\{f(\boldsymbol{x})+\sum_{i=1}^{m}\omega_i g_i(\boldsymbol{x})+\sum_{j=1}^{l}\nu_j h_j(\boldsymbol{x})\,\middle|\,\boldsymbol{x}\in B\right\}\tag{5-67}$$

当函数 $L(\boldsymbol{\omega},\boldsymbol{\nu})$不存在有限下界时,假设

$$L(\boldsymbol{\omega},\boldsymbol{\nu})=-\infty \tag{5-68}$$

$L(\boldsymbol{\omega},\boldsymbol{\nu})$称为拉格朗日对偶函数。上述拉格朗日对偶问题和拉格朗日对偶函数的定义显得有些突然，但读者可不用先纠结于为什么做这样的定义，随着后续的介绍会慢慢清楚。以下给出一个拉格朗日对偶问题和拉格朗日函数的案例。

【例 5-7】 考虑非线性优化问题

$$\begin{cases}\min & x_1^2+x_2^2\\ \text{s.t.} & -x_1-x_2+4\leqslant 0\\ & x_1\geqslant 0, x_2\geqslant 0\end{cases}$$

很显然，该问题的最优解和最优值为 $\boldsymbol{x}^*=(2,2)^{\mathrm{T}}$ 和 $f^*=8$。

将约束 $x_1\geqslant 0, x_2\geqslant 0$ 视作箱子集约束，即

$$B=\{\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}} \mid x_1\geqslant 0, x_2\geqslant 0\}$$

拉格朗日对偶函数为

$$\begin{aligned}L(\omega)&=\inf_x\{x_1^2+x_2^2+\omega(-x_1-x_2+4) \mid x_1\geqslant 0, x_2\geqslant 0\}\\ &=\inf_{x_1}\{x_1^2-\omega x_1 \mid x_1\geqslant 0\}+\inf_{x_2}\{x_2^2-\omega x_2 \mid x_2\geqslant 0\}+4\omega\end{aligned}$$

当 $\omega\geqslant 0$ 时，有

$$L(\omega)=-\frac{1}{2}\omega^2+4\omega$$

当 $\omega<0$ 时，由于 $x_1\geqslant 0, x_2\geqslant 0$ 故

$$x_1^2-\omega x_1\geqslant 0$$
$$x_2^2-\omega x_2\geqslant 0$$

因此，当 $x_1=x_2=0$ 时，得到极小值

$$L(\omega)=4\omega$$

综上分析，得到拉格朗日函数

$$L(\omega)=\begin{cases}-\frac{1}{2}\omega^2+4\omega & \omega\geqslant 0\\ 4\omega & \omega<0\end{cases}$$

因此，问题(5-65)的拉格朗日对偶问题为

$$\begin{cases}\max & -\frac{1}{2}\omega^2+4\omega\\ \text{s.t.} & \omega\geqslant 0\end{cases}$$

不难求得对偶问题的最优解和最优值分别为 $\omega^*=4$ 和 $L^*=8$

从例 5-7 可以看出，原问题的最优值和对偶问题的最优值正好相等，这种现象在线性规划中是必然的，但对于非线性规划则并不普遍成立。以下介绍针对非线性规划的对偶定理。

5.4.2 对偶定理

对偶定理主要研究原问题和对偶问题之间的关系,具体表现为原问题的最优解和对偶问题的最优解之间的大小关系。

为了阐述方便,使用向量式记号

$$\begin{aligned}&\boldsymbol{g}(\boldsymbol{x})=(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}}\\&\boldsymbol{h}(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x}))^{\mathrm{T}}\\&\boldsymbol{\omega}=(\omega_1,\omega_2,\cdots,\omega_m)^{\mathrm{T}}\\&\boldsymbol{\nu}=(\nu_1,\nu_2,\cdots,\nu_l)^{\mathrm{T}}\end{aligned}\tag{5-69}$$

则原问题(5-65)可改写为

$$\begin{cases}\min & f(\boldsymbol{x})\\ \text{s. t.} & \boldsymbol{g}(\boldsymbol{x})\leqslant\boldsymbol{0}\\ & \boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}\\ & \boldsymbol{x}\in B\end{cases}\tag{5-70}$$

对偶问题(5-66)可改写为

$$\begin{cases}\max & L(\boldsymbol{\omega},\boldsymbol{\nu})\\ \text{s. t.} & \boldsymbol{\omega}\geqslant 0\end{cases}\tag{5-71}$$

其中,对偶函数为

$$L(\boldsymbol{\omega},\boldsymbol{\nu})=\inf_{\boldsymbol{x}}\{f(\boldsymbol{x})+\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})\mid\boldsymbol{x}\in B\}\tag{5-72}$$

定理 5-11 (弱对偶定理)设 $\boldsymbol{x}$ 和$(\boldsymbol{\omega},\boldsymbol{\nu})^{\mathrm{T}}$ 分别为原问题和对偶问题的可行解,则

$$f(\boldsymbol{x})\geqslant L(\boldsymbol{\omega},\boldsymbol{\nu})\tag{5-73}$$

证明 根据拉格朗日函数的定义,有

$$\begin{aligned}L(\boldsymbol{\omega},\boldsymbol{\nu})&=\inf_{\boldsymbol{x}}\{f(\boldsymbol{x})+\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})\mid\boldsymbol{x}\in B\}\\&\leqslant f(\boldsymbol{x})+\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})\end{aligned}\tag{5-74}$$

由于 $\boldsymbol{x}$ 和$(\boldsymbol{\omega},\boldsymbol{\nu})^{\mathrm{T}}$ 分别为原问题和对偶问题的可行解,即满足 $\boldsymbol{g}(\boldsymbol{x})\leqslant\boldsymbol{0}$,$\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}$ 和$\boldsymbol{\omega}\geqslant\boldsymbol{0}$,故有

$$\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})\leqslant 0,\quad \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})=0\tag{5-75}$$

因此得到

$$L(\boldsymbol{\omega},\boldsymbol{\nu})\leqslant f(\boldsymbol{x})\tag{5-76}$$

由定理 5-11 可以得到以下几个推论。

推论 5-3 对原问题和对偶问题,必有

$$\inf_{\boldsymbol{x}}\{f(\boldsymbol{x})\mid\boldsymbol{g}(\boldsymbol{x})\leqslant\boldsymbol{0},\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0},\boldsymbol{x}\in B\}\geqslant\sup_{\boldsymbol{\omega},\boldsymbol{\nu}}\{L(\boldsymbol{\omega},\boldsymbol{\nu})\mid\boldsymbol{\omega}\geqslant\boldsymbol{0}\}\tag{5-77}$$

推论 5-4 如果 $f(\boldsymbol{x}^*)\leqslant L(\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)$,其中

$$\begin{aligned}&\boldsymbol{x}^*\in\{\boldsymbol{x}\in B\mid\boldsymbol{g}(\boldsymbol{x})\leqslant 0,\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}\}\\&\boldsymbol{\omega}^*\geqslant\boldsymbol{0}\end{aligned}\tag{5-78}$$

则 $\boldsymbol{x}^*$ 和 $(\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)$ 分别是原问题和对偶问题的最优解。

推论 5-5　如果

$$\inf\{f(\boldsymbol{x}) \mid \boldsymbol{g}(\boldsymbol{x}) \leqslant \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{x} \in B\} = -\infty \tag{5-79}$$

则对每个 $\boldsymbol{\omega} \geqslant \boldsymbol{0}$，有

$$L(\boldsymbol{\omega},\boldsymbol{\nu}) = -\infty \tag{5-80}$$

推论 5-6　如果

$$\sup\{L(\boldsymbol{\omega},\boldsymbol{\nu}) \mid \boldsymbol{\omega} \geqslant 0\} = \infty \tag{5-81}$$

则原问题没有可行解。

根据推论 5-3，若原问题的最优值为 f^*，对偶问题的最优值为 L^*，则必有

$$f^* \geqslant L^* \tag{5-82}$$

如果严格不等式成立，即 $f^* > L^*$，则称存在“对偶间隙”。线性规划问题一般不会存在对偶间隙，但对非线性规划问题，要想不出现对偶间隙，必须对目标函数和约束函数施加一定的限制，这种限制称为“约束规格”。

定理 5-12　（强对偶定理）设 B 是 $\mathbf{R}^n$ 中的一个非空凸集，f 和 $g_i(i=1,2,\cdots,m)$ 均为凸函数，$h_j(j=1,2,\cdots,l)$ 是 $\mathbf{R}^n$ 上的线性函数，又设存在点 $\bar{\boldsymbol{x}} \in B$，使

$$\boldsymbol{g}(\bar{\boldsymbol{x}}) > \boldsymbol{0}, \quad \boldsymbol{h}(\bar{\boldsymbol{x}}) = \boldsymbol{0}, \quad \boldsymbol{0} \in \operatorname{int} H(B) \tag{5-83}$$

其中，$H(B) = \{\boldsymbol{h}(\boldsymbol{x}) \mid \boldsymbol{x} \in B\}$，则

$$\inf\{f(\boldsymbol{x}) \mid \boldsymbol{g}(\boldsymbol{x}) \leqslant \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{x} \in B\} = \sup\{L(\boldsymbol{\omega},\boldsymbol{\nu}) \mid \boldsymbol{\omega} \geqslant \boldsymbol{0}\} \tag{5-84}$$

定理 5-12 的证明比较烦琐，此处不再赘述，感兴趣的读者可以参考文献[4]。

例 5-7 已经给出了一个强对偶的案例，以下从几何角度直观地解释对偶定理。

为了简单起见，考虑只有一个不等式约束的优化问题

$$\begin{cases} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{g}(\boldsymbol{x}) \leqslant \boldsymbol{0} \\ & \boldsymbol{x} \in B \end{cases} \tag{5-85}$$

其最优值可以表示为

$$f^* = \inf_{\boldsymbol{x}}\{f(\boldsymbol{x}) \mid \boldsymbol{g}(\boldsymbol{x}) \leqslant \boldsymbol{0}\} \tag{5-86}$$

构造集合

$$G = \{(\boldsymbol{g}(\boldsymbol{x}), f(\boldsymbol{x})) \mid \boldsymbol{x} \in B\} \tag{5-87}$$

也就是说 G 是由约束函数和目标函数构成的决策变量空间 B 所对应的函数值空间，如图 5-10 所示。

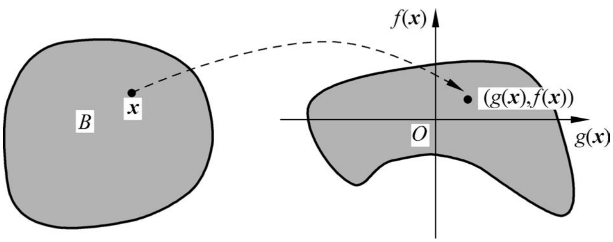

图 5-10　约束函数和目标函数值空间

利用集合 G,可以重写问题(5-85)为

$$f^* = \inf\{t \mid (u,t) \in G, u \leqslant 0\} \tag{5-88}$$

如图 5-11 所示,显然,f^* 为第三象限下方凸出部分顶点的纵坐标。

接下来考虑问题(5-85)的对偶问题,根据式(5-88)的符号,对偶函数为

$$L(\omega) = \inf\{t + \omega u \mid (u,t) \in G\} \tag{5-89}$$

对偶问题为

$$\begin{cases} \max & L(\omega) \\ \text{s.t.} & \omega \geqslant 0 \end{cases} \tag{5-90}$$

从几何上来看,这是一个先从一簇直线中求"最小",然后从众多"最小"中挑出"最大"的过程。

首先固定 ω,求 $L(\omega)$的过程如图 5-12 所示。将 $\omega u+t-b=0$ 绘制在图 5-12 中,显然,当直线和 G 的下端凸出部分相切时,纵截距 b 即为 $b=\omega u+t$ 的最小值,即 $L(\omega)$,而随着斜率 ω 的变化,当最小纵截距 b 不同时,$L(\omega)$不同,而最大的 $L(\omega)$即为和两边下方凸出部分均相切时,此时 f^*-L^* 即为对偶间隙,如图 5-13 所示。

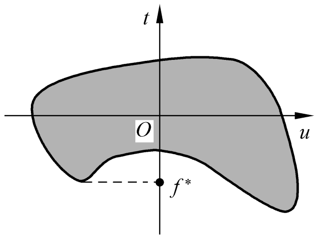

图 5-11 原问题最优值的几何解释

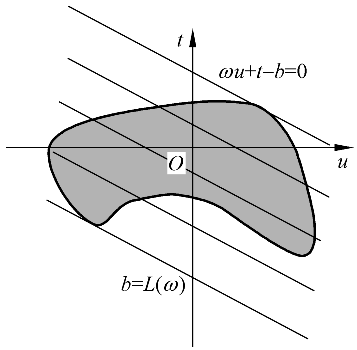

图 5-12 拉格朗日对偶函数的几何解释

从以上分析可以直观地看出,存在对偶间隙的原因是 G 的下方是非凸的,若 G 是一个凸集,显然对偶间隙就不存在了,如图 5-14 所示,例 5-7 也是这种情况。也就是说,G 为凸集是强对偶性的充分条件,但是否是必要条件呢?答案是否定的。图 5-15 给出了一个很简单但能充分说明问题的反例,因此 G 是凸集是强对偶性的充分非必要条件,这就是著名的 Slater 条件。

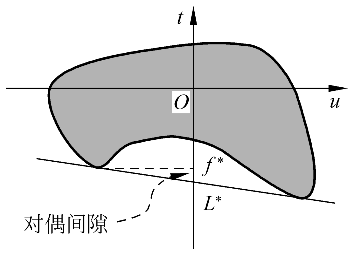

图 5-13 对偶间隙的几何解释

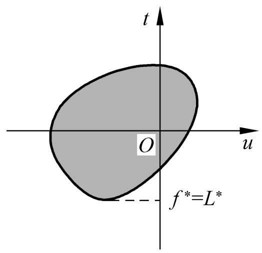

图 5-14 强对偶定理的几何解释

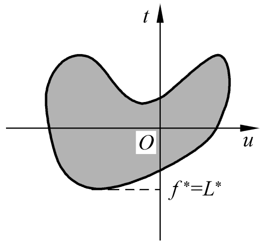

图 5-15　凸规划是强对偶性的充分条件

强对偶性有没有充分必要条件呢？答案是肯定的，5.4.3 节要介绍的 KKT 条件就是强对偶性的一个充分必要条件。

5.4.3　鞍点最优性条件

本节首先定义拉格朗日函数的鞍点，然后讨论拉格朗日函数的鞍点与原问题和对偶问题的最优解之间的关系，给出鞍点最优性条件，最后给出鞍点最优性条件与 KKT 条件之间的关系。

定义 5-6　设 $L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})$为拉格朗日函数，$\boldsymbol{x}^* \in \mathbf{R}^n$，$\boldsymbol{\omega}^* \in \mathbf{R}^m$，$\boldsymbol{\omega}^* \geqslant \mathbf{0}$，$\boldsymbol{\nu}^* \in \mathbf{R}^l$，如果对每个 $\boldsymbol{x} \in \mathbf{R}^n$，$\boldsymbol{\omega} \in \mathbf{R}^m$，$\boldsymbol{\omega} \geqslant \mathbf{0}$ 及$\boldsymbol{\nu} \in \mathbf{R}^l$ 都有

$$L(\boldsymbol{x}^*,\boldsymbol{\omega},\boldsymbol{\nu}) \leqslant L(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*) \leqslant L(\boldsymbol{x},\boldsymbol{\omega}^*,\boldsymbol{\nu}^*) \tag{5-91}$$

则称$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$ 为 $L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})$的鞍点。

由此定义可知，拉格朗日函数的鞍点必是拉格朗日函数关于 $\boldsymbol{x}$ 的极小点及关于$(\boldsymbol{\omega},\boldsymbol{\nu})^{\mathrm{T}}$的极大点，其中$\boldsymbol{\omega}$ 有非负限制，即$\boldsymbol{\omega} \geqslant \mathbf{0}$。

定理 5-13　（鞍点定理）设$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)$是原问题(5-60)的拉格朗日函数 $L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})$的鞍点，则 $\boldsymbol{x}^*$ 和$(\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)$分别是原问题(5-70)和对偶问题(5-71)的最优解。反之，假设 f，g_i $(i=1,2,\cdots,m)$是凸函数，$h_j(j=1,2,\cdots,l)$是线性函数，即 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$，并且 $\boldsymbol{A}$ 满秩，又设存在 $\bar{\boldsymbol{x}}$，使 $\boldsymbol{g}(\bar{\boldsymbol{x}})<\mathbf{0}$，$\boldsymbol{h}(\bar{\boldsymbol{x}})=\mathbf{0}$，如果 $\boldsymbol{x}^*$ 是问题(5-70)的最优解，则存在$(\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$，其中$\boldsymbol{\omega}^* \geqslant \mathbf{0}$，使$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$ 是拉格朗日函数 $L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})$的鞍点。

鞍点定理的前半部分给出了最优解的一种充分条件。定理的后半部分在 Slater 约束规格下，对于凸规划，给出最优解的一种必要条件，但值得注意的是，在一般情形下，当原问题存在最优解时，相应的拉格朗日函数不一定存在鞍点，因此，一般来讲，不能认为鞍点的存在是最优解的必要条件，以下给出一个案例。

【例 5-8】　考虑下列非线性规划问题

$$\begin{cases} \min & f(x)=x^3 \\ \text{s. t.} & x^2 \leqslant 0, x \in \mathbf{R} \end{cases}$$

显然，最优解 $x^*=0$，相应的拉格朗日函数为

$$L(x,\omega)=x^3+\omega x^2$$

现求 $\omega^* \geqslant 0$，使对所有 $x \in \mathbf{R}$，有

$$L(x^*,\omega) \leqslant L(x^*,\omega^*) \leqslant L(x,\omega^*)$$

即满足

$$(x^*)^3+\omega(x^*)^2 \leqslant (x^*)^3+\omega^*(x^*)^2 \leqslant x^3+\omega^* x^2$$

因为 $x^*=0$,所以上式等价地满足

$$x^3+\omega^* x^2 \geqslant 0$$

易知,ω^* 取任何非负数,上式都不能满足。若取 $\omega^*=0$,则当 $x=-1$ 时,上式不成立;若 $\omega^*>0$,则当 $x=-2\omega^*$,上式不成立,因此不存在 $\omega^*\geqslant 0$,使(x^*,ω^*)为函数 $L(x,\omega)$ 的鞍点。

关于鞍点条件与 KKT 条件之间的关系,有下列定理。

定理 5-14 设在问题(5-70)中,可行集为 F,$x^*\in F$ 满足 KKT 条件,即存在乘子

$$\boldsymbol{\omega}^*=(\omega_1^*,\omega_2^*,\cdots,\omega_n^*)^{\mathrm{T}} \geqslant 0 \tag{5-92}$$

和

$$\boldsymbol{\nu}^*=(\nu_1^*,\nu_2^*,\cdots,\nu_l^*)^{\mathrm{T}} \tag{5-93}$$

使

$$\begin{aligned} &\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\omega_i^*\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{l}\nu_j^*\nabla h_j(\boldsymbol{x}^*)=\mathbf{0} \\ &\omega_i^* g_i(\boldsymbol{x}^*)=0, \quad i=1,2,\cdots,m \end{aligned} \tag{5-94}$$

又设 f 和 $g_i(i=1,2,\cdots,m)$是凸函数,$A(\boldsymbol{x}^*)=\{i\,|\,g_i(\boldsymbol{x}^*)=0,i=1,2,\cdots,m\}$为积极约束指标集,当 $\nu_j^*\neq 0$ 时,h_j 是线性函数,则$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$ 是拉格朗日函数 $L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\nu})$的鞍点。反之,若 $f,g_i(i=1,2,\cdots,m)$和 $h_j(j=1,2,\cdots,l)$可微,$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$,$(\boldsymbol{\omega}^*\geqslant\mathbf{0})$是拉格朗日函数的鞍点,则$(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\nu}^*)^{\mathrm{T}}$ 满足 KKT 条件(5-94)。

定理 5-14 表明,如果 $\boldsymbol{x}^*$ 是 KKT 点,则在一定的凸性假设下,KKT 条件中的拉格朗日乘子就是鞍点条件中的乘子;反之,鞍点条件中的乘子也是 KKT 条件中的拉格朗日乘子。