最优化方法及其 Matlab 程序设计

马昌凤

2010 年 3 月 22 日

目 录

第八章　最优性条件

8.1　等式约束问题的最优性条件

本节讨论的最优性条件适合于下面的等式约束问题

$$\begin{aligned}&\min\ f(x),\\&\text{s.t.}\ h_i(x)=0,\ i=1,2,\cdots,l.\end{aligned}\tag{8.1}$$

为了研究问题的方便, 我们作问题 (8.1) 的所谓拉格朗日函数

$$L(x,\lambda)=f(x)-\sum_{i=1}^{l}\lambda_i h_i(x),\tag{8.2}$$

其中 $\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_l)^T$ 称为乘子向量.

下面的拉格朗日定理描述了问题 (8.1) 取极小值的一阶必要条件, 也就是所谓的 KT 条件 (Kuhn-Tucker 条件).

定理 8.1 (**拉格朗日定理**) 假设 x^* 是问题 (8.1) 的局部极小点, $f(x)$ 和 $h_i(x)\,(i=1,2,\cdots,l)$ 在 x^* 的某邻域内连续可微. 若向量组 $\nabla h_i(x^*)\,(i=1,2,\cdots,l)$ 线性无关, 则存在乘子向量 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_l^*)^T$ 使得

$$\nabla_x L(x^*,\lambda^*)=0,$$

即

$$\nabla f(x^*)-\sum_{i=1}^{l}\lambda_i^*\nabla h_i(x^*)=0.$$

证 记

$$H=\big(\nabla h_1(x^*),\nabla h_2(x^*),\cdots,\nabla h_l(x^*)\big).$$

由定理的假设知 H 列满秩. 因此, 若 $l=n$, 则 H 是可逆方阵, 从而矩阵 H

的列构成 $\mathbb{R}^n$ 中的一组基, 故存在 $\lambda^* \in \mathbb{R}^l\,(l=n)$ 使得

$$\nabla f(x^*) = \sum_{i=1}^{l} \lambda_i^* \nabla h_i(x^*),$$

此时结论得证.

下面设 $l<n$. 不失一般性, 可设 H 的前 l 行构成的 l 阶子矩阵 H_1 是非奇异的. 据此, 将 H 分块为

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}.$$

令 $y=(x_1,\cdots,x_l)^T$, $z=(x_{l+1},\cdots,x_n)^T$, 并记 $h(x)=(h_1(x),\cdots,h_l(x))^T$. 则有 $h(y^*,z^*)=0$, 且 $h(y,z)$ 在点 (y^*,z^*) 关于 y 的 Jacobi 矩阵 $H_1^T=\nabla_y h(y^*,z^*)$ 可逆. 故由隐函数定理可知, 在 z^* 附近存在关于 z 的连续可微函数 $y=u(z)$ 使得

$$h(u(z),z)=0.$$

对上式两边关于 z 求导得

$$\nabla_y h(u(z), z)\nabla u(z) + \nabla_z h(u(z), z) = 0,$$

故

$$\nabla u(z^*) = -H_1^{-T} H_2^T. \tag{8.3}$$

在 z^* 附近, 由$h(u(z), z) = 0$ 知 z^* 是无约束优化问题

$$\min_{z \in R^{n-l}} f(u(z), z)$$

的局部极小点, 故有

$$\nabla_z f(u(z^*), z^*) = 0,$$

即

$$\nabla u(z^*)^T \nabla_y f(y^*, z^*) + \nabla_z f(y^*, z^*) = 0.$$

注意到 $x^* = (y^*, z^*)$, 将 (8.3) 代入上式得

$$-H_2 H_1^{-1} \nabla_y f(x^*) + \nabla_z f(x^*) = 0.$$

令 $\lambda^* = H_1^{-1}\nabla_y f(x^*)$, 则有

$$\nabla_y f(x^*) = H_1\lambda^*, \quad \nabla_z f(x^*) = H_2\lambda^*.$$

两式合起来即

$$\nabla f(x^*) = \begin{pmatrix} \nabla_y f(x^*) \\ \nabla_z f(x^*) \end{pmatrix} = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} \lambda^* = \sum_{i=1}^{l} \lambda_i \nabla h_i(x^*).$$

至此已经证明了定理的结论. □

为了讨论等式约束问题的二阶必要条件, 需要用到 (8.2) 定义的拉格朗日函数 $L(x,\lambda)$ 的梯度和关于 x 的 Hesse 阵. 下面, 我们计算出它们的表

达式如下:

$$\nabla L(x,\lambda)=\begin{pmatrix}\nabla_x L(x,\lambda)\\ \nabla_\lambda L(x,\lambda)\end{pmatrix}=\begin{pmatrix}\nabla f(x)-\sum\limits_{i=1}^{l}\lambda_i\nabla h_i(x)\\ -h(x)\end{pmatrix},$$

$$\nabla_{xx}^2 L(x,\lambda)=\nabla^2 f(x)-\sum_{i=1}^{l}\lambda_i\nabla^2 h_i(x).$$

如果目标函数和约束函数都是二阶连续可微的, 则可考虑二阶充分性条件.

定理 8.2 对于等式约束问题 (8.1), 假设 $f(x)$ 和 $h_i(x)\,(i=1,2,\cdots,l)$ 都是二阶连续可微的, 并且存在 $(x^*,\lambda^*)\in\mathbb{R}^n\times\mathbb{R}^l$ 使得 $\nabla L(x^*,\lambda^*)=0$. 若对任意的 $0\neq d\in\mathbb{R}^n$, $\nabla h_i(x^*)^T d=0\,(i=1,2,\cdots,l)$, 均有 $d^T\nabla_{xx}^2 L(x^*,\lambda^*)d>0$, 则 x^* 是问题 (8.1) 的一个严格局部极小点.

证 用反证法. 若 x^* 不是严格局部极小点, 则必存在邻域 $N(x^*,\delta)$ 及

收敛于 x^* 的序列 $\{x_k\}$, 使得 $x_k \in N(x^*, \delta)$, $x_k \neq x^*$, 且有

$$f(x^*) \geq f(x_k), \ \ h_i(x_k) = 0, \ i = 1, 2, \cdots, l, \ k = 1, 2, \cdots$$

令 $x_k = x^* + \alpha_k z_k$, 其中 $\alpha_k > 0$, $\|z_k\| = 1$, 序列 $\{(\alpha_k, z_k)\}$ 有子列收敛于 $(0, z^*)$ 且 $\|z^*\| = 1$.

由泰勒中值公式得

$$0 = h_i(x_k) - h_i(x^*) = \alpha_k z_k^T \nabla h_i(x^* + \theta_{ik}\alpha_k z_k),$$

其中 $\theta_{ik} \in (0, 1)$. 上式两边同除以 α_k, 并令 $k \to \infty$ 得

$$\nabla h_i(x^*)^T z^* = 0, \ \ i = 1, 2, \cdots, l. \tag{8.4}$$

再由泰勒展开式得

$$L(x_k, \lambda^*) = L(x^*, \lambda^*) + \alpha_k \nabla_x L(x^*, \lambda^*)^T z_k + \frac{1}{2}\alpha_k^2 z_k^T \nabla_{xx}^2 L(x^*, \lambda^*) z_k + o(\alpha_k^2).$$

由于 x_k 都满足等式约束, 故有

$$\begin{aligned} 0 \ &\geq \ f(x_k) - f(x^*) = L(x_k, \lambda^*) - L(x^*, \lambda^*) \\ &= \ \frac{1}{2}\alpha_k^2 z_k^T \nabla_{xx}^2 L(x^*, \lambda^*) z_k + o(\alpha_k^2). \end{aligned}$$

上式两边同除以 $\alpha_k/2$, 可得

$$z_k^T \nabla_{xx}^2 L(x^*, \lambda^*) z_k + \frac{o(2\alpha_k^2)}{\alpha_k^2} \leq 0.$$

对上式取极限 $(k \to \infty)$ 即得

$$(z^*)^T \nabla_{xx}^2 L(x^*, \lambda^*) z^* \leq 0.$$

由于 z^* 满足 (8.4), 故得出矛盾. 因此 x^* 一定严格局部极小点. □

8.2　不等式约束问题的最优性条件

本小节我们考虑不等式约束优化问题的最优性条件:

$$\begin{aligned} &\min \ f(x), \\ &\text{s.t. } \ g_i(x) \geq 0, \ i = 1, 2, \cdots, m. \end{aligned} \tag{8.5}$$

记可行域为 $\mathcal{D} = \{x \in \mathbb{R}^n | \, g_i(x) \geq 0, i = 1, 2, \cdots, m\}$, 指标集 $I = \{1, \cdots, m\}$.

不等式约束问题的最优性条件需要用到所谓的有效约束和非有效约束的概念. 对于一个可行点 $\bar{x}$, 即 $\bar{x} \in \mathcal{D}$. 此时可能会出现两种情形. 即有些约束函数满足 $g_i(\bar{x}) = 0$, 而另一些约束函数则满足 $g_i(\bar{x}) > 0$. 对于后一种情形, 在 $\bar{x}$ 的某个邻域内仍然保持 $g_i(\bar{x}) > 0$ 成立, 而前者则不具备这种性质. 因此有必要把这两种情形区分开来.

定义 8.1 若问题 (8.5) 的一个可行点 $\bar{x} \in \mathcal{D}$ 使得 $g_i(\bar{x}) = 0$, 则称不等式约束 $g_i(x) \geq 0$ 为 $\bar{x}$ 的有效约束. 反之, 若有 $g_i(\bar{x}) > 0$, 则称不等式约束 $g_i(x) \geq 0$ 为 $\bar{x}$ 的非有效约束. 称集合

$$I(\bar{x}) = \{i : g_i(\bar{x}) = 0\} \tag{8.6}$$

为 $\bar{x}$ 处的有效约束指标集, 简称 x 处的有效集 (或积极集).

下面的两个引理是研究不等式约束问题最优性条件的基础.

引理 8.1 (**Farkas 引理**) 设 $a, b_i \in \mathbb{R}^n\,(i = 1, \cdots, r)$. 则线性不等式组

$$b_i^T d \geq 0, \quad i = 1, \cdots, r, \ d \in \mathbb{R}^n$$

与不等式

$$a^T d \geq 0$$

相容的充要条件是存在非负实数 $\alpha_1, \cdots, \alpha_r$, 使得 $a = \sum\limits_{i=1}^{r} \alpha_i b_i$.

证 充分性. 即存在非负实数 $\alpha_1, \cdots, \alpha_r$, 使得 $a = \sum\limits_{i=1}^{r} \alpha_i b_i$. 设 $d \in \mathbb{R}^n$ 满足 $b_i^T d \geq 0\,(i = 1, \cdots, r)$. 那么有

$$a^T d = \sum_{i=1}^{r} \alpha_i b_i^T d \geq 0.$$

必要性. 设所有满足 $b_i^T d \geq 0\,(i = 1, \cdots, r)$ 的向量 d 同时也满足 $a^T d \geq 0$. 用反证法. 设结论不成立, 即

$$a \notin C = \left\{x \in \mathbb{R}^n \middle| x = \sum_{i=1}^{r} \alpha_i b_i,\ \alpha_i \geq 0,\ i = 1, \cdots, r\right\}.$$

设 $a_0 \in C$ 是向量 a 在凸锥 C 上的投影, 即

$$\|a_0 - a\|_2 = \min_{x \in C} \|x - a\|_2,$$

则有 $a_0^T(a_0-a)=0$.

(1) 先证明对于任意的 $p\in C$ 必有 $u^T(a_0-a)\geq 0$. 事实上, 若不然, 则存在一个 $u\in C$, 使 $u^T(a_0-a)<0$. 令 $\bar{u}=\dfrac{u}{\|u\|}$, 则 $\bar{u}^T(a_0-a)=-\tau$, $\tau>0$. 注意到 C 是凸锥且 $a_0,\bar{u}\in C$, 故 $a_0+\tau\bar{u}\in C$. 此时有

$$\|a_0+\tau\bar{u}-a\|^2-\|a_0-a\|^2=-\tau^2<0,$$

这与 a_0 是投影的假设矛盾, 故必有 $u^T(a_0-a)\geq 0$, $\forall\, u\in C$.

(2) 现取 $d=a_0-a$. 由于 $b_i\in C$, 那么由 (1) 的结论可得 $b_i^Td\geq 0$. 故由必要性的假设应有 $a^Td\geq 0$. 但另一方面, 有

$$\begin{aligned} a^Td &= a^T(a_0-a)=a^T(a_0-a)-a_0^T(a_0-a)\\ &= -(a_0-a)^T(a_0-a)=-\|a_0-a\|^2<0, \end{aligned}$$

这与假设矛盾, 必要性得证. □

下面的 Gordan 引理可以认为是 Farkas 引理的一个推论.

引理 8.2（**Gordan 引理**）设 $b_i \in \mathbb{R}^n\,(i=1,\cdots,r)$. 线性不等式组

$$b_i^T d < 0,\ \ i=1,2,\cdots,r,\ d \in \mathbb{R}^n \tag{8.7}$$

无解的充要条件是 $b_i\,(i=1,\cdots,r)$ 线性相关, 即存在不全为 0 的非负实数 $\alpha_i\,(i=1,\cdots,r)$, 使得

$$\sum_{i=1}^{r} \alpha_i b_i = 0. \tag{8.8}$$

证 充分性. 用反证法. 设 (8.7) 有解, 即存在某个 d_0, 使得 $b_i^T d_0 < 0,\ \ i=1,\cdots,r$. 于是对于任意不全为 0 的非负实数 $\alpha_i\,(i=1,\cdots,r)$, 有 $\sum\limits_{i=1}^{r}\alpha_i b_i < 0$. 另一方面, 由充分性条件 (8.8) 有 $\sum\limits_{i=1}^{r}\alpha_i b_i = 0$, 矛盾, 故充分性得证.

必要性. 设不等式组 (8.7) 无解. 故对于任意的 $d \in \mathbb{R}^n$, 至少存在一个指标 i 满足 $b_i^T d \ge 0$. 记 $\beta_0 = \max\limits_{1\le i\le r}\{b_i^T d\}$, 则必有 $\beta_0 \ge 0$ 且

$$\beta_0 - b_i^T d \ge 0,\ \ i=1,\cdots,r.$$

下面我们构造 $r+2$ 个 $n+1$ 维向量:

$$\bar{d} = \begin{pmatrix} \beta \\ d \end{pmatrix}, \quad \bar{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \bar{b}_i = \begin{pmatrix} 1 \\ -b_i \end{pmatrix}, \quad i = 1, \cdots, r, \tag{8.9}$$

其中 $\beta \geq \beta_0$. 那么不难验证上述向量满足 Farkas 引理的条件, 即

$$\bar{b}_i^T \bar{d} = \beta - b_i^T d \geq \beta_0 - b_i^T d \geq 0, \ \ i = 1, \cdots, r,$$

且 $\bar{a}^T \bar{d} = \beta \geq 0$. 故由 Farkas 引理, 存在非负实数 $\alpha_1, \cdots, \alpha_r$, 使得

$$\bar{a} = \sum_{i=1}^{r} \alpha_i \bar{b}_i.$$

将 (8.9) 代入上式即得

$$\sum_{i=1}^{r} \alpha_i b_i = 0, \quad \sum_{i=1}^{r} \alpha_i = 1.$$

必要性得证. □

下面的引理可认为是一个几何最优性条件.

引理 8.3 设 x^* 是不等式约束问题 (8.5) 的一个局部极小点, $I(x^*)=\{i|\, g_i(x^*)=0\}$. 假设 $f(x)$ 和 $g_i(x)\,(i\in I(x^*))$ 在 x^* 处可微, 且 $g_i(x)\,(i\in I\backslash I(x^*))$ 在 x^* 处连续. 则问题 (8.5) 的可行方向集 $\mathcal{F}$ 与下降方向集 $\mathcal{S}$ 的交集是空集, 即 $\mathcal{F}\cap\mathcal{S}=\emptyset$, 其中

$$\mathcal{F}=\{d\in\mathbb{R}^n|\,\nabla g_i(x^*)^Td>0,\ i\in I(x^*)\},\quad \mathcal{S}=\{d\in\mathbb{R}^n|\,\nabla f(x^*)^Td<0\}. \tag{8.10}$$

证 用反证法. 设 $\mathcal{F}\cap\mathcal{S}\neq\emptyset$, 则存在 $d\in\mathcal{F}\cap\mathcal{S}$. 显然 $d\neq 0$. 由 $\mathcal{F}$, $\mathcal{S}$ 的定义及函数的连续性知, 存在充分小的正数 $\bar{\varepsilon}$, 使得对任意的 $0<\varepsilon\ll\bar{\varepsilon}$, 有

$$f(x^*+\varepsilon d)<f(x^*),\quad g_i(x^*+\varepsilon d)\geq 0,\ \ i=1,\cdots,m.$$

这与假设矛盾. □

下面我们给出不等式约束问题 (8.5) 的一阶必要条件, 即著名的 KT 条件.

定理 8.3 (**KT 条件**) 设 x^* 是不等式约束问题 (8.5) 的局部极小点, 有效约束集 $I(x^*) = \{i|\, g_i(x^*) = 0\}$. 并设 $f(x)$ 和 $g_i(x)\,(i = 1, \cdots, m)$ 在 x^* 处可微. 若向量组 $\nabla g_i(x^*)\,(i \in I(x^*))$ 线性无关, 则存在向量 $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)^T$ 使得

$$\begin{cases} \nabla f(x^*) - \sum\limits_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \\ g_i(x^*) \geq 0,\ \ \lambda_i^* \geq 0,\ \ \lambda_i^* g_i(x^*) = 0,\ i = 1, \cdots, m. \end{cases}$$

证 因 x^* 是问题 (8.5) 的局部极小点, 故由引理 8.3 知, 不存在 $d \in \mathbb{R}^n$ 使得

$$\nabla f(x^*)^T d < 0, \quad \nabla g_i(x^*)^T d > 0,\ i \in I(x^*),$$

即线性不等式组

$$\nabla f(x^*)^T d < 0, \quad -\nabla g_i(x^*)^T d < 0,\ i \in I(x^*)$$

无解. 于是由 Gordan 引理知, 存在不全为 0 的非负实数 $\mu_0 \geq 0$ 及 $\mu_i \geq$

$0\,(i \in I(x^*))$, 使得

$$\mu_0 \nabla f(x^*) - \sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*) = 0.$$

不难证明 $\mu_0 \neq 0$. 事实上, 若 $\mu_0 = 0$, 则有 $\sum\limits_{i \in I(x^*)} \mu_i \nabla g_i(x^*) = 0$, 由此可知 $\nabla g_i(x^*)\,(i \in I(x^*))$ 线性相关, 这与假设矛盾. 因此必有 $\mu_0 > 0$. 于是可令

$$\lambda_i^* = \frac{\mu_i}{\mu_0},\ \ i \in I(x^*);\ \ \lambda_i^* = 0,\ \ i \in I \backslash I(x^*),$$

则得

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0,$$

及

$$g_i(x^*) \geq 0,\ \ \lambda_i^* \geq 0,\ \ \lambda_i^* g_i(x^*) = 0,\ i = 1, \cdots, m.$$

定理得证. □

8.3 一般约束问题的最优性条件

我们现在考虑一般约束优化问题的最优性条件:

$$
\begin{aligned}
&\min\ f(x),\\
&\text{s.t. } h_i(x)=0,\ i=1,2,\cdots,l,\\
&\qquad g_i(x)\geq 0,\ i=1,2,\cdots,m.
\end{aligned}
\tag{8.11}
$$

记指标集 $E=\{1,\cdots,l\}$, $I=\{1,\cdots,m\}$. 可行域为 $\mathcal{D}=\{x\in\mathbb{R}^n|\,h_i(x)=0,\ i\in E,\ g_i(x)\geq 0,\ i\in I\}$.

把定理 8.1 和定理 8.3 结合起来即得到一般约束问题 (8.11) 的 KT 一阶必要条件.

定理 8.4 (**KT 一阶必要条件**) 设 x^* 是一般约束问题 (8.11) 的局部极小点, 在 x^* 处的有效约束集为

$$S(x^*)=E\cup I(x^*)=E\cup\{i\,|\,g_i(x^*)=0\}. \tag{8.12}$$

并设 $f(x)$, $h_i(x)\,(i \in E)$ 和 $g_i(x)\,(i \in I)$ 在 x^* 处可微. 若向量组 $\nabla h_i(x^*)\,(i \in E)$, $\nabla g_i(x^*)\,(i \in I(x^*))$ 线性无关, 则存在向量 $(\mu^*, \lambda^*) \in \mathbb{R}^l \times \mathbb{R}^m$, 其中 $\mu^* = (\mu_1^*, \cdots, \mu_l^*)^T$, $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)^T$, 使得

$$\begin{cases} \nabla f(x^*) - \sum\limits_{i=1}^{l} \mu_i^* \nabla h_i(x^*) - \sum\limits_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \\ h_i(x^*) = 0,\ i \in E, \\ g_i(x^*) \geq 0,\ \ \lambda_i^* \geq 0,\ \ \lambda_i^* g_i(x^*) = 0,\ \ i \in I. \end{cases} \tag{8.13}$$

注 8.1 (1) 称 (8.13) 为 KT 条件, 满足这一条件的点 x^* 称为 KT 点. 而把 $(x^*, (\mu^*, \lambda^*))$ 称为 KT 对, 其中 (μ^*, λ^*) 称为问题的拉格朗日乘子. 通常 KT 点、 KT 对和 KT 条件可以不加区别的使用.

(2) 称 $\lambda_i^* g_i(x^*) = 0\,(i \in I)$ 为互补性松弛条件. 这意味着 λ_i^* 和 $g_i(x^*)$ 中至少有一个必为 0. 若二者中的一个为 0, 而另一个严格大于 0, 则称之为满足严格互补性松弛条件.

与等式约束问题相仿, 可以定义问题 (8.11) 的拉格朗日函数

$$L(x,\lambda,\mu)=f(x)-\sum_{i=1}^{l}\mu_i h_i(x)-\sum_{i=1}^{m}\lambda_i g_i(x). \tag{8.14}$$

不难求出它关于变量 x 的梯度和 Hesse 阵分别为

$$\nabla_x L(x,\lambda,\mu)=\nabla f(x)-\sum_{i=1}^{l}\mu_i\nabla h_i(x)-\sum_{i=1}^{m}\lambda_i\nabla g_i(x),$$

$$\nabla^2_{xx} L(x,\lambda,\mu)=\nabla^2 f(x)-\sum_{i=1}^{l}\mu_i\nabla^2 h_i(x)-\sum_{i=1}^{m}\lambda_i\nabla^2 g_i(x).$$

与定理 8.2 的证明相类似, 证明问题 (8.13) 的二阶充分条件:

定理 8.5 对于约束优化问题 (8.11), 假设 $f(x)$, $g_i(x)\,(i\in I)$ 和 $h_i(x)\,(i\in E)$ 都是二阶连续可微的, 有效约束集 $I(x^*)$ 由 (8.12) 所定义. 且 $(x^*,(\mu^*,\lambda^*))$ 是问题 (8.11) 的 KT 点. 若对任意的 $0\neq d\in\mathbb{R}^n$, $\nabla g_i(x^*)^T d=0\,(i\in I(x^*)$, $\nabla h_i(x^*)^T d=0\,(i\in E)$, 均有 $d^T\nabla^2_{xx}L(x^*,\lambda^*)d>0$, 则 x^* 是问题 (8.11) 的一个严格局部极小点.

例 8.1 考虑优化问题

$$
\begin{aligned}
\min\ & f(x) = -2x_1^2 - x_2^2, \\
\text{s.t.}\ & x_1^2 + x_2^2 - 2 = 0, \\
& -x_1 + x_2 \geq 0, \\
& x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
$$

试验证 $x^* = (1,1)^T$ 为 KT 点，并求出问题的 KT 对.

解 计算

$$
\nabla f(x^*) = \begin{pmatrix} -4x_1 \\ -2x_2 \end{pmatrix}\Bigg|_{x=x^*} = \begin{pmatrix} -4 \\ -2 \end{pmatrix}, \quad \nabla h(x^*) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \nabla g_1(x^*) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.
$$

令

$$
\nabla f(x^*) - \mu^* \nabla h(x^*) - \lambda_1^* \nabla g_1(x^*) = 0,
$$

解得 $\mu^*=-1.5$, $\lambda_1^*=1$. 再令 $\lambda_2^*=\lambda_3^*=0$, 得

$$\begin{cases}\nabla f(x^*)-\mu^*\nabla h(x^*)-\sum\limits_{i=1}^{3}\lambda_i^*\nabla g_i(x^*)=0,\\ \lambda_i^* g_i(x^*)=0,\ \lambda_i\geq 0,\ i=1,2,3.\end{cases}$$

这表明 x^* 是 KT 点, $(x^*,(\mu^*,\lambda^*))$ 是 KT 对, 其中 $\mu^*=-1.5$, $\lambda^*=(1,0,0)^T$.

一般而言, 问题 (8.11) 的 KT 点不一定是局部极小点. 但如果问题是下面的所谓凸优化问题, 则 KT 点、局部极小点、全局极小点三者是等价的.

下面首先给出约束凸优化问题的定义.

定义 8.2 对于约束最优化问题

$$\begin{aligned}&\min\ f(x),\quad x\in\mathbb{R}^n\\ &\text{s.t.}\ \begin{cases}h_i(x)=0,\ i=1,2,\cdots,l,\\ g_i(x)\geq 0,\ i=1,2,\cdots,m,\end{cases}\end{aligned}\tag{8.15}$$

若 $f(x)$ 是凸函数, $h_i(x)\,(i=1,\cdots,l)$ 是线性函数, $g_i(x)\,(i=1,\cdots,m)$ 是凹函数 (即 $-g_i(x)$ 是凸函数), 那么上述约束优化问题称为凸优化问题.

定理 8.6 设 (x^*,μ^*,λ^*) 是凸优化问题 (8.15) 的 KT 点, 则 x^* 必为该问题的全局极小点.

证 因对于凸优化问题, 其拉格朗日函数

$$L(x,\mu^*,\lambda^*)=f(x)-\sum_{i=1}^{l}\mu_i^*h_i(x)-\sum_{i=1}^{m}\lambda_i^*g_i(x)$$

关于 x 是凸函数. 故对于每一个可行点 x, 我们有

$$\begin{aligned}
f(x) &\geq f(x)-\sum_{i=1}^{l}\mu_i^*h_i(x)-\sum_{i=1}^{m}\lambda_i^*g_i(x)\\
&= L(x,\mu^*,\lambda^*)\\
&\geq L(x^*,\mu^*,\lambda^*)+\nabla_x L(x^*,\mu^*,\lambda^*)^T(x-x^*)\\
&= L(x^*,\mu^*,\lambda^*)\ =\ f(x^*).
\end{aligned}$$

故 x^* 为问题的全局极小点. □

8.4 鞍点和对偶问题

本节介绍约束优化问题的鞍点和对偶等有关概念. 首先给出鞍点的定义.

定义 8.3 对约束优化问题 (8.11), 若存在 x^* 和 (μ^*, λ^*), 其中 $\lambda^* \geq 0$, 满足

$$L(x^*, \mu, \lambda) \leq L(x^*, \mu^*, \lambda^*) \leq L(x, \mu^*, \lambda^*), \ \ \forall (x, \mu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+, \tag{8.16}$$

则称 (x^*, μ^*, λ^*) 为约束优化问题 (8.11) 的拉格朗日函数的鞍点, 通常简称 x^* 为问题 (8.11) 的鞍点.

下面的定理表明, 鞍点 x^* 不仅是 KT 点, 而且是全局极小点.

定理 8.7 设 (x^*, μ^*, λ^*) 是约束优化问题 (8.11) 的鞍点. 则 (x^*, μ^*, λ^*) 不仅是问题 (8.11) KT 点, 而且是它的全局极小点.

证 由鞍点的定义 8.3 可知 x^* 是

$$\min_{x\in\mathbb{R}^n} L(x,\mu^*,\lambda^*)$$

的全局极小点. 由无约束优化问题的最优性条件可得 $\nabla_x L(x^*,\mu^*,\lambda^*)=0$, 这就证明了约束优化问题 (8.11) KT 条件的第一个式子.

另一方面, 再由鞍点的定义可知 (μ^*,λ^*) 是

$$\max_{\lambda_i\geq 0\,(i\in I),\mu\in\mathbb{R}^l} L(x^*,\mu,\lambda)$$

的全局极大点, 等价地, (μ^*,λ^*) 是

$$\min_{\lambda_i\geq 0\,(i\in I),\mu\in\mathbb{R}^l} -L(x^*,\mu,\lambda)$$

的全局极小点. 那么由定理 8.4 可知, 存在乘子向量 $\omega^*=(\omega_1^*,\cdots,\omega_m^*)^T\,(\omega_i^*\geq 0, i=1,\cdots,m)$ 使得

$$\begin{cases} h_i(x^*)=0, & i\in E \\ g_i(x^*)=\omega_i^*\geq 0,\ \lambda_i^*\geq 0,\ \omega_i^*\lambda_i^*=0, & i\in I. \end{cases}$$

从而 x^* 是问题 (8.11) 的可行点, 且 $\lambda_i^* g_i(x^*) = 0,\ i \in I$, 故 (x^*, μ^*, λ^*) 满足 (8.11) 的 KT 条件, 即为 KT 点.

进一步, 由鞍点的定义, 对于问题 (8.11) 的任意可行点 x, 我们有

$$L(x^*, \mu^*, \lambda^*) \le L(x, \mu^*, \lambda^*),$$

即

$$f(x^*) \le f(x) - \sum_{i=1}^{l} \mu_i^* h_i(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \le f(x),$$

从而 x^* 是问题 (8.11) 的全局极小点. □

上述定理说明, 鞍点一定是 KT 点, 但反之不一定成立. 然而, 对于凸优化问题, KT 点、鞍点和全局极小点三者是等价的. 我们有

定理 8.8 设 (x^*, μ^*, λ^*) 是凸优化问题的 KT 点, 则 (x^*, μ^*, λ^*) 为对应的拉格朗日函数的鞍点, 同时 x^* 也是该凸优化问题的全局极小点.

证 注意到对于凸优化问题, 拉格朗日函数

$$L(x,\mu^*,\lambda^*)=f(x)-\sum_{i=1}^{l}\mu_i^*h_i(x)-\sum_{i=1}^{m}\lambda_i^*g_i(x)$$

关于 x 是凸函数, 故由凸函数的性质 (定理 **??**), 有

$$\begin{aligned}L(x,\mu^*,\lambda^*) &\geq L(x^*,\mu^*,\lambda^*)+\nabla_x L(x^*,\mu^*,\lambda^*)^T(x-x^*)\\ &= L(x^*,\mu^*,\lambda^*),\end{aligned}$$

即 $L(x^*,\mu^*,\lambda^*)\leq L(x,\mu^*,\lambda^*)$. 另一方面, 对于任意的 $(\lambda,\mu)\in\mathbb{R}_+^m\times\mathbb{R}^l$, 有

$$\begin{aligned}L(x^*,\mu,\lambda)-L(x^*,\mu^*,\lambda^*) &= -\sum_{i=1}^{l}(\mu_i-\mu_i^*)h_i(x^*)-\sum_{i=1}^{m}(\lambda_i-\lambda_i^*)g_i(x^*)\\ &= -\sum_{i=1}^{m}\lambda_i g_i(x^*)\leq 0,\end{aligned}$$

即 $L(x^*,\mu,\lambda)\leq L(x^*,\mu^*,\lambda^*)$, 故 (x^*,μ^*,λ^*) 为鞍点, 同时 x^* 也是凸优化问题的全局极小点. □

下面讨论约束优化问题的对偶问题. 对于约束优化问题 (8.11), 我们引入如下记号:

$$G(x) = (g_1(x), \cdots, g_m(x))^T, \quad H(x) = (h_1(x), \cdots, h_l(x))^T.$$

令 $y \in \mathbb{R}^m$, $z \in \mathbb{R}^l$, 定义函数

$$L(x, y, z) = f(x) - G(x)^T y - H(x)^T z$$

及

$$\theta(y, z) = \inf_{x \in \mathbb{R}^n} L(x, y, z).$$

易知 $\theta(y, z)$ 关于 (y, z) 是凹函数.

约束问题 (8.11) 的拉格朗日对偶定义为:

$$\begin{cases} \max \ \theta(y, z), \\ \text{s.t.} \ \ y \in \mathbb{R}^m_+, \ z \in \mathbb{R}^l. \end{cases} \tag{8.17}$$

约束问题 (8.11) 的 Wolfe 对偶定义为:

$$
\begin{cases}
\max \ L(x, y, z), \\
\text{s.t.} \ \ \nabla_x L(x, y, z) = 0, \\
\qquad\quad y \in \mathbb{R}_+^m, \ z \in \mathbb{R}^l.
\end{cases} \tag{8.18}
$$

注 8.2 上面的两种对偶在某种意义上是一致的. 事实上, 对于拉格朗日对偶, 由于目标函数 $\theta(y, z)$ 本身就是拉格朗日函数关于 x 的极小值, 所以显然有 $\nabla_x L(x, y, z) = 0$ 成立. 将其合并到拉格朗日对偶的约束当中, 就得到了所谓的 Wolfe 对偶.

对于线性规划问题和凸二次规划问题, 拉格朗日对偶 (8.17) 将会有更为明晰的形式. 我们看下面例子.

例 8.2 设有线性规划问题

$$\begin{cases} \min\ c^T x, \\ \text{s.t.}\quad Ax = b, \\ \qquad\quad x \geq 0. \end{cases} \tag{8.19}$$

试写出它的拉格朗日对偶.

解 其拉格朗日函数为

$$L(x, y, z) = c^T x - y^T x - z^T (Ax - b).$$

对上述函数关于 x 求极小. 令

$$\nabla_x L(x, y, z) = c - y - A^T z = 0.$$

将其代入拉格朗日函数得

$$\begin{aligned}\theta(y,z) &= \inf_{x\in\mathbb{R}^n} L(x,y,z) \\ &= \inf_{x\in\mathbb{R}^n}\{(c-y-A^Tz)^Tx+z^Tb\} \\ &= z^Tb = b^Tz.\end{aligned}$$

注意到 $y=c-A^Tz\ge 0$, 于是有

$$\begin{cases}\max\ b^Tz, \\ \text{s.t.}\quad A^Tz\le c.\end{cases} \tag{8.20}$$

这就是线性规划问题 (8.19) 的对偶规划.

例 8.3 设二次规划问题

$$\begin{cases}\min\ \dfrac{1}{2}x^TBx+c^Tx, \\ \text{s.t.}\quad Ax\le b,\end{cases} \tag{8.21}$$

其中 $B \in \mathbb{R}^{n\times n}$ 对称正定, $A \in \mathbb{R}^{m\times n}$. 试写出二次规划问题 (8.21) 的对偶规划.

解 首先写出该问题的拉格朗日函数为

$$L(x,y) = \frac{1}{2}x^T Bx + c^T x - y^T(b - Ax).$$

对上述函数关于 x 求极小. 由于 B 对称正定, 故函数 $L(x,y)$ 关于 x 为凸函数. 令

$$\nabla_x L(x,y,z) = Bx + c + A^T y = 0,$$

解得 $x=-B^{-1}(c+A^Ty)$. 将其代入拉格朗日函数得

$$\begin{aligned}\theta(y) &= \inf_{x\in\mathbb{R}^n} L(x,y)\\ &= \inf_{x\in\mathbb{R}^n}\left\{(Bx+c+A^Ty)^Tx-y^Tb-\frac{1}{2}x^TBx\right\}\\ &= -b^Ty-\frac{1}{2}[-B^{-1}(c+A^Ty)]^TB[-B^{-1}(c+A^Ty)]\\ &= -b^Ty-\frac{1}{2}(c^T+y^TA)B^{-1}(c+A^Ty)\\ &= (-b-AB^{-1}c)^Ty-\frac{1}{2}y^T(AB^{-1}A^T)y-\frac{1}{2}c^TB^{-1}c.\end{aligned}$$

令

$$d=-b-AB^{-1}c,\quad D=-AB^{-1}A^T,$$

则有

$$\theta(y)=\frac{1}{2}y^TDy+d^Ty-\frac{1}{2}c^TB^{-1}c.$$

注意到乘子向量 $y \geq 0$, 因此二次规划问题 (8.21) 的拉格朗日对偶是

$$\begin{cases} \max \ \dfrac{1}{2} y^T D y + d^T y - \dfrac{1}{2} c^T B^{-1} c, \\ \text{s.t.} \quad y \leq 0. \end{cases}$$

最后, 我们讨论一下原问题与对偶问题的目标函数值之间的关系.

定理 8.9 (**弱对偶定理**) 设 $\bar{x}$ 和 $(\bar{y}, \bar{z})$ 分别是原问题 (8.11) 和对偶问题 (8.17) 的可行解. 则有 $\theta(\bar{y}, \bar{z}) \leq f(\bar{x})$.

证 因 $\bar{x}$ 和 $(\bar{y}, \bar{z})$ 分别是原问题 (8.11) 和对偶问题 (8.17) 的可行解, 故

$$\begin{aligned} \theta(\bar{y}, \bar{z}) &= \inf_{x \in \mathbb{R}^n} \{ f(x) - \bar{y}^T G(x) - \bar{z}^T H(x) \} \\ &\leq f(\bar{x}) - \bar{y}^T G(\bar{x}) - \bar{z}^T H(\bar{x}) = f(\bar{x}). \end{aligned}$$

定理得证. □