第七章 罚函数方法（续）

2018年5月31日 15:44

下面讨论等值约束优化问题在增广 Lagrange 法框架下的充分和必要条件.

引理7.3 设 $A\in R^{n\times n}$, $B\in R^{m\times n}$, 对满足 $Bd=0$、$d\neq 0$ 的 d, $d^TAd>0$ 的充要条件是 $\exists\, \sigma^*\ge 0$, 使得 $\forall\, \sigma\ge\sigma^*$ 和 $d\neq 0$, 有 $d^T(A+\sigma B^TB)d>0$. "A在d方向上正定"
证明见[76]. (或 refbook P517)

定理7.4 设 x^* 是(7.1)的严格局部最优解, λ^* 是对应 Lagrange 乘子, 在 x^*, λ^* 处最优性的二阶充分条件(6.21)满足, 则存在 $\sigma^*\ge 0$, $\forall\, \sigma\ge\sigma^*$, x^* 是增广 Lagrange 函数 $\phi(x,\lambda^*,\sigma)$ 的严格局部极小点; 反之, 若 $c_i(x^*)=0$, $i=1,\cdots,m$, x^* 是 $\phi(x,\lambda^*,\sigma)$ 的局部极小点, 则 x^* 是(7.1)的局部最优解.

证明: 由 $\phi(x,\lambda,\sigma)$ 定义:

$$\nabla_x\phi(x,\lambda,\sigma)=\nabla_x L(x,\lambda)+\sigma A(x)c(x) \qquad (7.31)$$

$$\nabla_x^2\phi(x,\lambda,\sigma)=\nabla_x^2 L(x,\lambda)+\sigma A(x)A(x)^T+\sigma\sum_{i\in E}c_i(x)\nabla^2 c_i(x) \qquad (7.32)$$

由 x^* 是(7.1)的严格局部最优点及 KKT 条件, 及可行性条件 $c_i(x)=0$, 有

$$\nabla_x\phi(x,\lambda,\sigma)=\nabla_x L(x,\lambda)=0$$

$$\nabla_x^2\phi(x,\lambda,\sigma)=\nabla_x^2 L(x,\lambda)+\sigma A(x)A(x)^T$$

由约束优化二阶最优性条件,

$\cancel{\forall d\in F_1^*}$, $d^T\nabla_x^2 L(x^*,\lambda^*)d>0$. $\forall d\in \mathcal{C}^*$, 即 $\nabla c(x^*)^T d=0$

即 $\forall d\neq 0$, $A^*d=0$ $\cancel{(F_1^*)}$, 有 即 $A^*d=0$.

$d^T\nabla_x^2 L(x^*,\lambda^*)d>0$. $\Leftarrow$ 且 $d\neq 0$. (二阶充分)

由引理7.3, $\exists\,\sigma^*\ge 0$, $\forall\,\sigma\ge\sigma^*$, 有

$$\nabla_x^2 L(x^*,\lambda^*)+\sigma A(x^*)A(x^*)^T$$

正定, 从而 x^* 是 $\min\phi(x^*,\lambda^*,\sigma)$ 的严格局部最优解. (存在 $\sigma^*\ge 0$, 使 ϕ 在 $\cancel{F^*}$ $\mathcal{C}$ 方向上是凸的).

反之, 已知 x^* 是可行点, 设 $x\in\Omega$, 且 x 与 x^* 充分接近, 则有

$$\underbrace{\phi(x^*,\lambda^*,\sigma)}_{极小}\le\phi(x,\lambda^*,\sigma)$$

而

$$\phi(x^*,\lambda^*,\sigma)=f(x^*)-\underbrace{\lambda^Tc(x)}_{=0}+\frac{1}{2}\sigma\underbrace{\sum_{i\in E}c_i^2(x)}_{=0}=f(x^*)$$

$$\phi(x,\lambda^*,\sigma)=f(x).$$

即

$$f(x^*)\le f(x).$$

即 x^* 是(7.1)的局部最优解. □

$2x_1+1$

例7.5

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

解)

$$L(x,\lambda)=2x_1^2-x_2^2+x_1-x_2-\lambda(x_1-x_2)$$

$$\nabla_x L(x,\lambda)=\begin{bmatrix}4x_1+1-\lambda\\ -2x_2-1+\lambda\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$\nabla_x \mathcal{L}(x,\lambda)=\begin{bmatrix}4x_1+1-\lambda\\ -2x_2-1+\lambda\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$\Rightarrow \left\{\begin{array}{l} x_1=\frac{\lambda-1}{4}\\ x_2=\frac{\lambda-1}{2}\\ x_1-x_2=0\end{array}\right\} \Rightarrow \frac{\lambda-1}{4}=\frac{\lambda-1}{2} \Rightarrow \lambda=1 = \begin{cases}x_1=0\\ x_2=0\end{cases}$$

在$\lambda^*=1$处，有

$$\phi(x,\lambda^*,\sigma)=2x_1^2-x_2^2+\frac{1}{2}\sigma(x_1-x_2)^2$$

$$\Rightarrow \nabla_x\phi(x,\lambda^*,\sigma)=\begin{bmatrix}(4+\sigma)x_1-\sigma x_2\\ (\sigma-2)x_2-\sigma x_1\end{bmatrix}$$

$$\Rightarrow \nabla_x^2\phi(x,\lambda^*,\sigma)=\begin{bmatrix}4+\sigma & -\sigma\\ -\sigma & \sigma-2\end{bmatrix}$$

特征方程：

$$(4+\sigma-\lambda)(\sigma-2-\lambda)-\sigma^2=0$$
$$\Rightarrow \lambda^2-2\lambda-2\sigma\lambda+2\sigma-8=0$$
$$\Rightarrow \lambda^2-2\lambda(1+\sigma)+(1+\sigma)^2-(1+\sigma)^2+2\sigma-8=0$$
$$\Rightarrow [\lambda-(1+\sigma)]^2=(1+\sigma)^2+8-2\sigma$$
$$\Rightarrow \lambda=1+\sigma\pm\sqrt{(1+\sigma)^2+8-2\sigma}\geq 0$$
$$\Rightarrow \sigma>4.$$

即当$\sigma>4$时，$\nabla_x^2\phi(x,\lambda^*,\sigma)$正定，即$x^*=(0,0)$是$\phi(x,\lambda^*,\sigma)$的严格局部极小值点。

反之，由

$$\phi(x,\lambda,\sigma)=f(x)-\lambda(x_1-x_2)+\frac{1}{2}\sigma(x_1-x_2)^2$$

$$\Rightarrow \nabla_x\phi(x,\lambda,\sigma)=\begin{bmatrix}(4+\sigma)x_1-\sigma x_2-\lambda+1\\ (\sigma-2)x_2-\sigma x_1+\lambda-1\end{bmatrix}=0$$

$$\Rightarrow \left\{\begin{array}{l} x_1=\frac{1-\lambda}{\sigma-4}\\ x_2=2\,\frac{1-\lambda}{\sigma-4}\\ x_1=x_2\end{array}\right\} \Rightarrow \frac{1-\lambda}{\sigma-4}=2\,\frac{1-\lambda}{\sigma-4}\Rightarrow \lambda=1$$

（σ取值并不影响λ）.

$\Rightarrow x_1=x_2=0$. 是问题最优解.

引入松弛变量将一般约束优化问题化为等式约束优化问题：

对(6.1)：

$$\begin{aligned}\min\ & f(x)\\ \text{s.t.}\ & c_i(x)=0,\quad i\in E\\ & c_i(x)\geq 0,\quad i\in I\end{aligned}$$

引入$s_i\geq 0,\ i\in I$，变为：

$$\min\ F(x,s)=f(x) \qquad (7.33a)$$
$$\text{s.t.}\ c_i(x)=0,\ i\in E \qquad (7.33b)$$
$$c_i(x)-s_i=0,\ i\in I \qquad (7.33c)$$
$$s_i\geq 0,\ i\in I \qquad (7.33d)$$

$$\text{s.t.}\quad c_i(x)=0,\ i\in E \qquad (7.33b)$$

$$c_i(x)-s_i=0,\ i\in I \qquad (7.33c)$$

$$s_i\geq 0,\ i\in I \qquad (7.33d)$$

这么做本质上是提高了问题的空间维数，将问题从n维提高到n+m-me维。而对此问题，可用增广Lagrange方法求解：

$$\bar{\Phi}(x,s,\lambda,\sigma)=f(x)-\sum_{i\in E}\lambda_i c_i(x)+\frac{1}{2}\sigma\sum_{i\in E}c_i^2(x)+\sum_{i\in I}\bar{\phi}_i,$$

其中

$$\bar{\phi}_i=-\lambda_i(c_i(x)-s_i)+\frac{1}{2}\sigma(c_i(x)-s_i)^2$$

然后我们求解：

$$\begin{cases}\min\limits_{x,s}\ \bar{\Phi}(x,s,\lambda,\sigma) & (7.34a)\\ \text{s.t.}\quad s_i\geq 0,\ i\in I & (7.34b).\end{cases}$$

注意此问题等价于两重优化问题：

$$\min_x\left\{\begin{array}{l}\min\limits_{s}\ \bar{\Phi}(x,s,\lambda,\sigma)\\ \text{s.t.}\ s_i\geq 0,\ i\in I\end{array}\right\} \qquad (7.35)$$

而对内层问题：

$$\begin{array}{l}\min\ \bar{\Phi}(x,s,\lambda,\sigma)\\ \text{s.t.}\ s_i\geq 0,\ i\in I\end{array} \qquad (7.36)$$

包含s的部分：

$$\bar{\phi}_i=-\lambda_i(c_i(x)-s_i)+\frac{1}{2}\sigma(c_i(x)-s_i)^2$$

当σ充分大时是凸的，即其稳定点：

$$\frac{\partial\bar{\phi}_i}{\partial s_i}=\lambda_i-\sigma c_i(x)+\sigma s_i=0,\quad i\in I$$

$$\Rightarrow\quad s_i=c_i(x)-\eta_i,\quad i\in I$$

其中

$$\eta_i=\frac{\lambda_i}{\sigma}.$$

是其最优值点，即内层问题有解析解：

$$s_i=\max\{c_i(x)-\eta_i,\ 0\},\qquad i\in I. \qquad (7.37)$$

$$\Rightarrow\quad c_i(x)-s_i=\begin{cases}\eta_i, & c_i(x)-\eta_i\geq 0\\ c_i(x), & c_i(x)-\eta_i<0\end{cases} \qquad (7.38)$$

$$\Leftrightarrow\quad c_i(x)-s_i=\min\{c_i(x),\ \eta_i\} \qquad (7.39)$$

代入$\bar{\phi}_i$，得：

$$\bar{\phi}_i=-\lambda_i(c_i(x)-s_i)+\frac{1}{2}\sigma(c_i(x)-s_i)^2$$

$$=\begin{cases}-\lambda_i\eta_i+\frac{1}{2}\sigma\eta_i^2=-\lambda_i\cdot\frac{\lambda_i}{\sigma}+\frac{1}{2}\sigma\cdot\frac{\lambda_i^2}{\sigma^2}=-\frac{1}{2}\sigma\eta_i^2, & c_i(x)-\eta_i\geq 0\\ \frac{1}{2}\sigma\left[(c_i(x)-\eta_i)^2-\eta_i^2\right], & c_i(x)-\eta_i<0\end{cases} \qquad (7.40)$$

$\Leftrightarrow$

$$\bar{\phi}_i=\frac{1}{2}\sigma\left[(\min\{c_i(x)-\eta_i,\ 0\})^2-\eta_i^2\right]:=\phi_i,\quad i\in I.$$

由此得增广Lagrange函数：

$$\phi(x,\lambda,\sigma):=f(x)-\sum_{i\in E}\lambda_i c_i(x)+\frac{1}{2}\sigma\sum_{i\in E}c_i^2(x)+\sum_{i\in I}\phi_i(x,\lambda,\sigma) \qquad (7.41)$$

由此得到了 Lagrange 函数：

$$\phi(x,\lambda,\sigma) := f(x) - \sum_{i\in \mathcal{E}}\lambda_i c_i(x) + \frac{1}{2}\sigma\sum_{i\in \mathcal{E}} c_i^2(x) + \sum_{i\in \mathcal{I}}\phi_i(x,\lambda,\sigma) \qquad (7.41)$$

由算法 7.3，

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - \sigma_k c_i(x_k), \quad i\in \mathcal{E} \qquad (7.42)$$

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - \sigma_k\left[c_i(x_k) - s_i^{(k)}\right], \quad i\in \mathcal{I} \qquad (7.43)$$

而相应的终止准则为

$$\left[\sum_{i\in \mathcal{E}} c_i^2(x_k) + \sum_{i\in \mathcal{I}}\left(\min\{c_i(x), \eta_i^{(k)}\}\right)^2\right]^{\frac{1}{2}} \le \varepsilon$$