优化方法

2016-2017学年工科硕士课程

第4章 约束优化方法

约束优化问题是在自变量满足约束条件的情况下目标函数最小化的问题,

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in S = \{x : g(x) \leq 0, h(x) = 0\}. \end{aligned}$$

约束优化方法大体可分为以下四类:

- 用线性规划或二次规划来逐次逼近非线性规划的方法, 如SLP法, SQP法等;
- 把约束优化问题转化为无约束优化问题来求解的方法,如SUMT外点法,SUMT内点法,乘子法等;
- 对约束优化问题不预先作转换,直接进行处理的分析方法,如可行方向法,梯度投影法,既约梯度法等;
- 对约束优化问题不预先作转换的直接搜索方法,如复形法,随机试验法等.

约束优化问题:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0,\ \ i = 1, 2, \cdots, m \\ & h_j(x) = 0,\ \ j = 1, 2, \cdots, p \\ & x \in X \subset \mathbf{R}^n. \end{array} \tag{1-1}$$

其中$x = (x_1, x_2, \cdots, x_n)^T$,用$S$表示问题(1-1)的可行域,即

$$S = \{x | x \in X, g_i(x) \leq 0,\ \ i = 1, \cdots, m, h_j(x) = 0,\ \ j = 1, \cdots, p\}. \tag{1-2}$$

设x^0为可行点,即$x^0 \in S$,定义

$$I(x^0) = \{i | g_i(x^0) = 0, 1 \le i \le m\}. \tag{1-3}$$

当$i \in I(x^0)$时,对应的约束称为在x^0处的紧约束, 或起作用约束,而$I(x^0)$称为x^0点紧约束的指标集,显然

$$I(x^0) \subset \{1, 2, \cdots, m\}.$$

不引起混淆的情况下将$I(x^0)$简记为I.

例1.1

考察问题

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

可行集S为单位圆在第一象限的部分.最优点为$x^* = (1, 0)$.点x^*处的紧约束指标集为$I(x^*) = \{2, 3\}$,且$g_2(x) \le 0$和$g_3(x) \le 0$为x^*点的紧约束.

最优性的一阶必要条件

定理1.1

设x^*是问题(1-1)的一个可行解,$X \subset \mathbf{R}^n$为非空开集,$f, g_i(i \in I(x^*)), h_j$在点$x^*$可微, $g_i(i \notin I(x^*))$在点x^*处连续,
$\nabla g_i(x^*), i \in I(x^*), \nabla h_j(x^*), j = 1, 2, \cdots, p$线性无关,若$x^*$是(1-1) 的局部最优解,则有实数$u_i, v_j$(其中$i \in I(x^*), j = 1, 2, \cdots, p$)使得

$$\nabla f(x^*) + \sum_{i \in I(x^*)} u_i \nabla g_i(x^*) + \sum_{j=1}^{p} v_j \nabla h_j(x^*) = 0 \tag{1-5a}$$

$$u_i \geq 0, \ \ i \in I(x^*) \tag{1-5b}$$

如果加上$i \notin I(x^*)$时,g_i在点x^*可微,则上述条件可改写为

$$\nabla f(x^*) + \sum_{i=1}^{m} u_i \nabla g_i(x^*) + \sum_{j=1}^{p} v_j \nabla h_j(x^*) = 0 \tag{1-5a'}$$

$$u_i \geq 0, \ \ i = 1, 2, \cdots, m \tag{1-5b'}$$

$$u_i g_i(x^*) = 0, \ \ i = 1, 2, \cdots, m \tag{1-5c}$$

上述条件(1-5)称为Kuhn-Tucker最优性必要条件,满足条件(1-5)的点成为Kuhn-Tucker点,简称为K-T点,有时也称为Karush-Kuhn-Tucker点,简称KKT点.条件(1-5c)称为互补松弛条件.

引入Lagrange函数

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{p} v_j h_j(x)$$

其中$u = (u_1, u_2, \cdots, u_m)^T, v = (v_1, v_2, \cdots, v_p)^T$称为Lagrange乘子向量.

由定理1.1可知:若x^*是问题(1-1)的最优解,则存在$u^* \geq 0, v^*$使(x^*, u^*, v^*)是$L(x, u, v)$的稳定点.

Kuhn-Tucker条件的几何解释

假设(1-1)中没有等式约束,即$p = 0$的情况. 令

$$C = \{y | y = \sum_{i \in I} u_i \nabla g_i(x^*), i \in I, u_i \geq 0, u \neq 0\}.$$

易见C是由紧约束梯度张成的锥.于是Kuhn-Tucker条件

$$-\nabla f(x^*) = \sum_{i \in I} u_i \nabla g(x^*), u_i \geq 0, i \in I, u \neq 0$$

可以解释为:若x^*是问题(1-1)的最优解,则$-\nabla f(x^*)$必位于点x^*的上述锥C之中. 即$-\nabla f(x^*) \in C$.

最优性的充分条件

定理1.2

设$X \subset \mathbf{R}^n$为非空开集,$x^* \in S$,在x^*点K-T条件成立,即存在数$u_i \geq 0(i \in I(x^*))$ 和$v_i(i = 1, 2, \cdots, p)$,使得

$$\nabla f(x^*) + \sum_{i \in I(x^*)} u_i \nabla g_i(x^*) + \sum_{j=1}^{p} v_j \nabla h_j(x^*) = 0$$

令$J = \{i | v_i > 0\}, K = \{i | v_i < 0\}$,设$f$在点$x^*$伪凸,$g_i(i \in I)$在点$x^*$拟凸,$h_i(i \in J)$在点$x^*$拟凸, $h_i(i \in K)$ 在点x^*拟凹,则x^*是问题(1-1)的全局最优解.

若f是可微凸函数,则一定是拟凸和伪凸的;若f是凹函数,则一定是拟凹的.

例1.2

考察问题

$$\begin{cases} \min & f(x) = (x_1-3)^2 + (x_2-2)^2 \\ \text{s.t.} & g_1(x) = x_1^2 + x_2^2 - 5 \le 0 \\ & g_2(x) = -x_1 \le 0 \\ & g_3(x) = -x_2 \le 0 \\ & h_1(x) = x_1 + 2x_2 - 4 = 0. \end{cases}$$

- 易见$x^0 = (2,1)^T \in S$,且只有$g_1(x) \le 0$是紧约束.下面验证x^0是K-T点.

$$\nabla f(x^0) = \begin{pmatrix} -2 \\ -2 \end{pmatrix},\ \nabla g_1(x^0) = \begin{pmatrix} 4 \\ 2 \end{pmatrix},\ \nabla h_1(x^0) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

所以由(1-5a)得

$$\begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + v_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 0$$

可得$u_1 = 1/3, v_1 = 2/3, u_2 = u_3 = 0$,因此$x^0$是K-T点.

- 下面来研究另一点$x^1=(0,2)^T$.这时紧约束只有$g_2(x)\le 0$.

$$\nabla f(x^1)=\begin{pmatrix}-6\\0\end{pmatrix},\ \nabla g_2(x^1)=\begin{pmatrix}-1\\0\end{pmatrix},\ \nabla h_1(x^1)=\begin{pmatrix}1\\2\end{pmatrix}$$

由式(1-5a)得

$$\begin{pmatrix}-6\\0\end{pmatrix}+u_2\begin{pmatrix}-1\\0\end{pmatrix}+v_1\begin{pmatrix}1\\2\end{pmatrix}=0$$

或

$$\begin{cases}v_1-u_2=6\\2v_1=0\end{cases}$$

所以$v_1=0,u_2=-6<0$.因此,x^1不是K-T点.

4.2 惩罚函数法

考虑约束优化问题:

$$
\begin{array}{ll}
\min & f(x) \\
\text{s.t.} & g_i(x) \leq 0,\ \ i = 1, 2, \cdots, m \\
& h_j(x) = 0,\ \ j = 1, 2, \cdots, p \\
& x \in X \subset \mathbf{R}^n.
\end{array}
\tag{2-1}
$$

其中$f, g_i(i = 1, 2, \cdots, m), h_j(j = 1, 2, \cdots, p) \in C^1$.

惩罚函数法的基本思想

把约束优化问题(2-1)转化为一个或一系列无约束问题来求解,所以也称为序列无约束极小化技术,简称SUMT法. 通常把惩罚函数法称为SUMT外点法, 碰壁函数法称为SUMT内点法.

对问题(2-1)定义惩罚函数

$$F(x, M) = f(x) + Mp(x) \tag{2-2}$$

其中$M > 0$为常数,称为惩罚因子,$p(x)$是定义在$\mathbf{R}^n$上的一个函数,称为惩罚项,它满足

1. $p(x)$是连续的;
2. 对任意$x \in \mathbf{R}^n$,有$p(x) \geq 0$;
3. 当且仅当$x \in S$时,$p(x) = 0$.

通常对等式约束,定义

$$g_j^+(x) = (h_j(x))^2,\ j = 1, 2, \cdots, p \tag{2-3}$$

对不等式约束,定义

$$g_{p+i}^+(x) = \begin{cases} 0, & \text{当} g_i(x) \le 0 \\ (g_i(x))^2, & \text{当} g_i(x) > 0 \end{cases} \qquad i = 1, 2, \cdots, m \tag{2-4}$$

令$L = p + m$和$M_k > 0$,于是,惩罚函数为

$$F(x, M_k) = f(x) + M_k \sum_{i=1}^{L} g_i^+(x). \tag{2-5}$$

容易验证,惩罚项$p(x)=\sum_{i=1}^{L} g_i^{+}(x)$满足:

1. 显然$g_i^{+}(x)(i=1,2,\cdots,L)$是连续的,因此$p(x)$连续;
2. 对于所有$x\in \mathbf{R}^n$,显然有$g_i^{+}(x)\geq 0$,所以$p(x)\geq 0$;
3. 当且仅当$x\in S$时,$g_i^{+}(x)=0(i=1,2,\cdots,L)$,所以$p(x)=0$.

容易验证,惩罚项$p(x)=\sum_{i=1}^{L} g_i^+(x)$满足:

1 显然$g_i^+(x)(i=1,2,\cdots,L)$是连续的,因此$p(x)$连续;
2 对于所有$x\in \mathbf{R}^n$,显然有$g_i^+(x)\geq 0$,所以$p(x)\geq 0$;
3 当且仅当$x\in S$时,$g_i^+(x)=0(i=1,2,\cdots,L)$,所以$p(x)=0$.

- 由以上分析可见,当约束条件被损坏,即$x\notin S$时,则至少有一个$i(1\leq i\leq L)$, 使$g_i^+(x)>0$,从而$p(x)>0$.约束条件被破坏得越厉害,则$p(x)>0$取值就越大, 从而$F(x,M)=f(x)+Mp(x)$也就越大,即对于约束条件被破坏是一种惩罚,M越大,则惩罚得越厉害.
- 当约束条件满足时,则$p(x)=0$,不管$M>0$取多么大的值,$F(x,M)\equiv f(x)$,就是说,当约束条件满足时,不受惩罚.
- 关于惩罚因子M_k的取法. 例如, 可以选取$M_{k+1}=M_kc, c\in[4,10]$.

4.2.2 惩罚函数法的计算步骤

(1) 选取$M_1 > 0$,精度$\varepsilon > 0, c \geq 2$,初始点$x^0$,令$k = 1$;

(2) 以x^{k-1}为初始点,求解无约束优化问题

$$\min\ F(x, M_k) = f(x) + M_k \sum_{i=1}^{L} g_i^+(x) \tag{2-6}$$

设其最优解为$x^k = x(M_k)$;

(3) 令$\tau_1 = \max\limits_{1\leq i\leq p} \{|h_i(x^k)|\}, \tau_2 = \max\limits_{1\leq i\leq m} \{g_i(x^k)\}, \tau = \max\{\tau_1, \tau_2\}$.

(4) 若$\tau < \varepsilon$,则迭代结束,取$x^* = x^k$;否则令$M_{k+1} = cM_k, k = k + 1$转回第(2)步.

注

- 在(2-6)式中,惩罚项$p(x)$可以采用(2-3)和(2-4)式的定义,也可用其他的定义, 只要能保证$p(x)$具有前面所述的三条性质;
- 算法的结束准则$\tau < \varepsilon$,也可以改为$M_k p(x^k) < \varepsilon$.

4.2.3 收敛性分析

设$x^k = x(M_k)$为无约束优化问题(2-6)的最优解.分两种情况进行讨论:

- 如果某一个$M_k (k \geq 1)$对应的无约束优化问题(2-6)的最优解$x^k \in S$,则x^k为原来的约束优化问题(2-1)的最优解.
- 如果第一种情况总不发生,这时就得到一个无穷点列$\{x^k\}$, $x^k \notin S$, $k = 1, 2, \cdots$. 这时,可以证明在某些条件下,$\{x^k\}$的任一极限点x^*都是原来约束优化问题(2-1)的最优解.

引理1

对$M_{k+1} > M_k$,下列不等式成立.

(1)

$$F(x^k, M_k) \leq F(x^{k+1}, M_{k+1}) \tag{2-7}$$

(2)

$$p(x^k) \geq p(x^{k+1}) \tag{2-8}$$

(3)

$$f(x^k) \leq f(x^{k+1}) \tag{2-9}$$

引理2

设x^*是约束优化问题(2-1)的一个最优解,则对每一个k,有

$$f(x^*) \geq F(x^k, M_k) \geq f(x^k).$$

定理2.1

设$f(x)\in C, g_i(x)\in C(i=1,2,\cdots,m), h_j(x)\in C(j=1,2,\cdots,p)$,$\{x^k\}$是由惩罚函数法产生的一个序列,则序列的任一个极限点都是约束优化问题(2-1)的一个最优解.

证明.

设$\lim\limits_{i\to\infty} x^{k_i}=\bar{x}$,其中$\{x^{k_i}\}$是$\{x^k\}$一个子序列. 只需证明$\bar{x}\in S$和$f(\bar{x})=\min\limits_{x\in S} f(x)=f(x^*)=f^*$.

(1) 由引理1和引理2知$F(x^k,M_k)$是不减且有上界f^*的序列,因此$\lim\limits_{i\to\infty} F(x^{k_i},M_{k_i})=F^*\le f^*$. 可证$\bar{x}\in S$.

(2) 由引理2知$f(x^{k_i})\le f^*$,令$i\to\infty$,且由$\bar{x}\in S$,可得$f(\bar{x})=f^*$.

□

例2.1

用惩罚函数法求解

$$\begin{array}{ll} \min & f(x) = (x_1-3)^2 + (x_2-2)^2 \\ \text{s.t.} & h(x) = x_1 + x_2 - 4 = 0. \end{array}$$

解

- 令$F(x,M) = (x_1-3)^2 + (x_2-2)^2 + M(x_1+x_2-4)^2$,

  $$\frac{\partial F}{\partial x_1} = 2(x_1-3)+2M(x_1+x_2-4), \frac{\partial F}{\partial x_2} = 2(x_2-2)+2M(x_1+x_2-4)$$

  令$\frac{\partial F}{\partial x_1} = 0, \frac{\partial F}{\partial x_2} = 0$得

  $$x_1 = \frac{5M+3}{2M+1}, \ x_2 = \frac{3M+2}{2M+1}.$$

- 又因为$\nabla^2 F = \begin{pmatrix} 2(M+1) & 2M \\ 2M & 2(M+1) \end{pmatrix}$且$M>0$,所以$\nabla^2 F$是正定的,因此$F(x,M)$在

$$x(M) = (\frac{5M+3}{2M+1}, \frac{3M+2}{2M+1})^T$$

处取得极小值.

- 令$M \to +\infty$,得$x^* = \lim\limits_{M\to+\infty} x(M) = (5/2, 3/2)^T$,易见$h(x^*)=0$,所以$x^*$为极小点.

例2.2

用惩罚函数法求解

$$\begin{array}{ll} \min & f(x) = (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{s.t.} & h(x) = x_1 + x_2 - 4 \le 0. \end{array}$$

解

- 令

$$h^+(x) = \begin{cases} 0, & h(x) \le 0 \\ (h(x))^2, & h(x) > 0 \end{cases}$$

和

$$p(x, M) = (x_1 - 3)^2 + (x_2 - 2)^2 + Mh^+(x)$$

- $$\frac{\partial p}{\partial x_1} = \begin{cases} 2(x_1-3)+2M(x_1+x_2-4), & h(x)>0 \\ 2(x_1-3), & h(x)\le 0 \end{cases}$$

  $$\frac{\partial p}{\partial x_2} = \begin{cases} 2(x_2-2)+2M(x_1+x_2-4), & h(x)>0 \\ 2(x_2-2), & h(x)\le 0 \end{cases}$$

  令$\nabla p = 0$,得$x_1 = 3, x_2 = 2$(与$h(x)\le 0$矛盾)或者

  $$x_1 = \frac{5M+3}{2M+1},\ x_2 = \frac{3M+2}{2M+1}.$$

- 令$M \to +\infty$,得$x^* = \lim_{M\to+\infty} x(M) = (5/2, 3/2)^T$.显然$h(x^*)\le 0$,所以$x^*$为极小点.

4.3 碰壁函数法(Barrier Function Method)

碰壁函数法适用于形如

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in S \end{array} \tag{3-1}$$

的约束优化问题,其中可行集S的内部(用int$S$表示)非空,而且int$S$ 中的点可以任意地接近于S的任一点,从直观上来看,就是S不能包含孤立点和孤立的弧段, 这样的集合S称为健集(Robust set).

显然S中不能包含有等式约束,即S只能是形如

$$S = \{x | g_i(x) \le 0, i = 1, 2, \cdots, m\}$$

的约束集.

碰壁函数法的基本思想

从一个可行点x^0出发,在可行点之间进行迭代的一种方法.为了使迭代点保持为可行点,在约束集S的边界上建造一道“围墙”,它阻挡迭代点列离开可行集S.

碰壁项

碰壁项定义于int$S$中,满足条件:

(1) $B(x)$是连续的;

(2) $B(x) \geq 0$;

(3) 当x趋近于S的边界时,$B(x) \to \infty$.

通常定义碰壁项为

$$B(x)=\sum_{i=1}^{m} g_i^{+}(x), \tag{3-3}$$

其中$g_i^{+}(x)$取

$$g_i^{+}(x)=-ln(-g_i(x)),\ i=1,2,\cdots,m \tag{3-4}$$

或

$$g_i^{+}(x)=-\frac{1}{g_i(x)},\ i=1,2,\cdots,m \tag{3-5}$$

碰壁函数定义:

$$F(x,r_k)=f(x)+r_kB(x), \tag{3-6}$$

其中$r_k>0$且$r_1>r_2>\cdots>r_k>r_{k+1}>\cdots,\lim\limits_{k\to\infty} r_k=0.$

当x靠近可行域S的边界时,$g_1(x), \cdots, g_m(x)$中,至少有一个$i(1 \le i \le m)$,使$g_i(x) \to 0$.因此,$g_i^+(x) \to \infty$,从而$B(x) \to \infty$. 这样求解问题(3-1)就可转化为求解问题

$$\begin{array}{ll} \min & F(x,r) = f(x) + rB(x) \\ \text{s.t.} & x \in \text{int}S \end{array} \tag{3-7}$$

其中$r > 0$.

当x靠近可行域S的边界时,$g_1(x), \cdots, g_m(x)$中,至少有一个$i(1 \le i \le m)$,使$g_i(x) \to 0$.因此,$g_i^+(x) \to \infty$,从而$B(x) \to \infty$. 这样求解问题(3-1)就可转化为求解问题

$$
\begin{aligned}
\min \quad & F(x, r) = f(x) + rB(x) \\
\text{s.t.} \quad & x \in \text{int}S
\end{aligned}
\tag{3-7}
$$

其中$r > 0$.

注意:问题(3-7)虽然是一个约束问题,但由于在S的边界附近,它的目标函数值趋近于无穷大, 所以只要从S的一个内点开始迭代,并注意控制一维搜索的步长,可使x^k不越出可行域, 因而不必直接与约束打交道,从计算的观点来看,它是无约束问题.

碰壁函数法的计算步骤

(1) 选取$r_1 > 0$(例如$r_1 = 1$),精度$\varepsilon > 0, c \geq 2$(常取$c \in [4, 10]$),

(2) 求可行集S的一个内点$x^0 \in \text{int}S$,令$k = 1$;

(3) 以x^{k-1}为初始点,使用求解无约束优化问题的方法求解

$$\begin{array}{ll} \min & F(x, r_k) = f(x) + r_k B(x) \\ \text{s.t.} & x \in \text{int}S \end{array}$$

设其最优解为$x^k = x(r_k)$;

(4) 若x^k满足下面的结束准则,则迭代结束,取$x^* = x^k$;否则令$r_{k+1} = r_k/c, k = k + 1$转回第(3)步.

结束准则

常用的结束准则有以下四种,可以选用其中之一,或要求其中两种同时满足.准则(4)用得更广泛一些.

1. $\|x^k - x^{k-1}\| < \varepsilon$;
2. $\|f(x^k) - f(x^{k-1})\| < \varepsilon$;
3. $\frac{\|f(x^k) - f(x^{k-1})\|}{f(x^k)} < \varepsilon$;
4. $r_k B(x^k) < \varepsilon$.

例3.1

用碰壁函数法求解

$$\begin{array}{ll} \min & f(x) = x_1^2 + x_2^2 \\ \text{s.t.} & x_1 \geq 1. \end{array}$$

解: 令$F(x, r) = x_1^2 + x_2^2 - r\ln(x_1 - 1)$,则

$$\frac{\partial F}{\partial x_1} = 2x_1 - \frac{r}{x_1 - 1}, \ \frac{\partial F}{\partial x_2} = 2x_2.$$

令$\frac{\partial F}{\partial x_1} = 0, \frac{\partial F}{\partial x_2} = 0$得$2x_1^2 - 2x_1 - r = 0, 2x_2 = 0$,从而

$$x_1 = \frac{2 \pm \sqrt{4 + 8r}}{4} = \frac{1 \pm \sqrt{1 + 2r}}{2}, \ x_2 = 0$$

由$x_1 \geq 1$可得

$$x_1 = \frac{1 + \sqrt{1 + 2r}}{2}$$

当$r \to 0$,得$x^* = (1, 0)^T$为最优解.

收敛性分析

引理1

对$r_{k+1} < r_k$,下列不等式成立.

(1) $F(x^k, r_k) \geq F(x^{k+1}, r_{k+1})$

(2) $B(x^k) \leq B(x^{k+1})$

(3) $f(x^k) \geq f(x^{k+1})$

引理2

设x^*是约束优化问题(3-1)的一个最优解,则对每一个k,有

$$f(x^*) \leq f(x^k) \leq F(x^k, r_k).$$

定理3.1

设$f(x) \in C, g_i(x) \in C(i = 1, 2, \cdots, m)$, $\{x^k\}$是由碰壁函数法产生的一个序列,则序列的任一个极限点都是约束优化问题(3-1)的一个最优解.

证明.

设$\lim\limits_{i\to\infty} x^{k_i} = \bar{x}$,其中$\{x^{k_i}\}$是$\{x^k\}$一个子序列.

只需证明$f(\bar{x}) = \min\limits_{x\in S} f(x) = f(x^*) = f^*$.

- 可证$\lim\limits_{i\to\infty} F(x^{k_i}, r_{k_i}) = f^*$.
- 由引理2知$f(x^*) \le f(x^{k_i}) \le F(x^{k_i}, r_{k_i})$,令$i \to \infty$,得$f(\bar{x}) = f^*$.

□

惩罚函数法和碰壁函数法的优点

- 方法简单,程序易于编制.
- 对目标函数和约束函数要求不高,适用范围较广.

惩罚函数法和碰壁函数法的缺点

- 需要求解一系列的无约束优化问题,计算量大,而且参数M_k, r_k的选取对方法的收敛速度影响比较大.
- 在迭代过程中,外点法中的参数M_k不断增大,内点法中的参数r_k不断减小, 使得求解无约束极小化问题变得十分困难.
- 用外点法求得的近似解x^k往往不是可行解,在某些实际问题中,这样的结果是不能用的;而内点法中, 要求初始点x^0位于可行域的内部,这是困难的,而且内点法不能求解包含有等式约束的优化问题.

混合罚函数法

为了克服最后一个缺点,人们往往将内点法和外点法结合起来使用,当初始点x^0给定以后, 对x^0满足的那些不等式约束,按内点法来构造碰壁项$B(x)$,对x^0不满足的那些不等式约束和等式约束,按外点法构造惩罚项$p(x)$. 即取混合罚函数为

$$F(x, r_k) = f(x) + r_k B(x) + \frac{1}{r_k} p(x)$$

其中, $B(x) = \sum_{i\in I_1} g_i^+(x)$, $p(x) = \sum_{i\in I_2} g_i^+(x) + \sum_{j=1}^{p} (h_j(x))^2$
$I_1 = \{i | g_i(x^{k-1}) < 0, i \in I\}, I_2 = \{i | g_i(x^{k-1}) \geq 0, i \in I\}, I = \{1, 2, \ldots, m\}$
$r_0 > r_1 > r_2 > \cdots > r_k > r_{k+1} > \cdots$, $\lim_{k\to\infty} r_k = 0$
$g_i^+(x)$可按照外点法和内点法中所介绍的方法确定.

混合罚函数法的计算步骤

(1) 选取初始点x^0,精度$\varepsilon > 0, 0 < c < 1$,令$k = 1$;

(2) 以x^{k-1}为初始点,求解无约束优化问题$\min F(x, r_k)$, 设其最优解为$x^k = x(r_k)$;

(3) 若x^k满足结束准则,则迭代结束,取$x^* = x^k$;否则令$r_{k+1} = cr_k$, $k = k + 1$,转回第(2)步.

4.4 可行方向法

最早的一个可行方向法是在1960年由G.Zoutendijk提出的. 现在,可行方向法已发展成为求解约束优化问题的一类重要的方法.

基本思想

在给定一个可行点x^k之后,用某种方法确定一个改进的可行方向d_k, 然后沿方向d_k,求解一个有约束的线搜索问题,得极小点λ_k,令$x^{k+1} = x^k + \lambda_k d_k$, 如果$x^{k+1}$仍不是最优解,则重复上述步骤.

不同的可行方向法的主要区别在于,选择可行方向的策略不同,大体上可以分为三类:

- 用求解一个线性规划问题来确定,如Zoutendijk方法,Frank-Wolfe方法和Topkis-Veinott方法等;
- 利用投影矩阵来直接构造一个改进的可行方向,如Rosen的梯度投影法和Rosen-Polak方法等;
- 利用既约梯度,直接构造一个改进的可行方向,如Wolfe的既约梯度法及其各种改进,凸单纯形法.

可行方向与改进方向

考察非线性规划问题

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in S \subset \mathbf{R}^n \end{array} \tag{4-1}$$

其中S为非空的可行集.

定义

- 非零向量d称为在点$x \in S$的一个可行方向,如果存在一个数$\delta > 0$, 使得对$\forall \lambda \in (0, \delta)$,都有$x + \lambda d \in S$.
- 非零向量d称为在点$x \in S$的一个改进的可行方向,如果存在一个数$\delta > 0$, 使得对$\forall \lambda \in (0, \delta)$,都有$f(x + \lambda d) < f(x)$且$x + \lambda d \in S$.

定理4.1

设x是问题

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & Ax \le b, Ex = e & \quad (4\text{-}2) \\ & x \in \mathbf{R}^n & \end{array}$$

的一个可行解,假定$A_1x = b_1, A_2x < b_2$,其中$A^T = (A_1^T, A_2^T)$, $b^T = (b_1^T, b_2^T)$, A 是$m \times n$矩阵,E是$l \times n$矩阵,$b \in \mathbf{R}^n, e \in \mathbf{R}^l$,则一个非零向量$d$是在$x$点的可行方向,当且仅当$A_1d \le 0, Ed = 0$;如果$\nabla f(x)^Td < 0$,则$d$是一改进方向.其中$f(x)$在$x$点可微.

证明.

1. 设$A_1 d \leq 0, Ed = 0$,证明d为$x \in S$的可行方向. 即要证存在一个数$\delta > 0$,使得对$\forall \lambda \in (0, \delta)$,都有$x + \lambda d \in S$.
2. 设$d \neq 0$为点$x \in S$的可行方向,要证$A_1 d \leq 0, Ed = 0$.
3. 设$\nabla f(x)^T d < 0$,由$f(x + \lambda d)$在x的一阶Taylor展开,易知d为改进方向.

$\square$

定理4.2

设x是问题

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, i = 1, 2, \cdots, m \\ & x \in \mathbf{R}^n \end{array} \tag{4-3}$$

的一个可行解,令$S = \{x | x \in \mathbf{R}^n, g_i(x) \leq 0, i = 1, 2, \cdots, m\}$, 集合$I = \{i | g_i(x) = 0\}$是$x$点的紧约束指标集.设$f$和$g_i (i \in I)$在$x$点可微, $g_i (i \notin I)$在x点连续,如果$\nabla f(x)^T d < 0, \nabla g_i(x)^T d < 0 (i \in I)$, 则$d$是一改进的可行方向.

例4.1

考虑如下问题

$$\begin{aligned} \min \quad & f(x) = (x_1-6)^2 + (x_2-2)^2 \\ \text{s.t.} \quad & -x_1 + 2x_2 \le 4 \\ & 3x_1 + 2x_2 \le 12 \\ & -x_1 \le 0 \\ & -x_2 \le 0, \end{aligned}$$

求点$x = (2,3)^T$处的改进的可行方向.

解 在点$x = (2,3)^T$,

$$I = \{1,2\},\ A_1 = \begin{pmatrix} -1 & 2 \\ 3 & 2 \end{pmatrix}.$$

因此可行方向集

合$D = \{d | A_1 d \le 0\} = \{d | -d_1 + 2d_2 \le 0, 3d_1 + 2d_2 \le 0\}$. 而满足条件$\nabla f(x)^T d = -8d_1 + 2d_2 < 0$ 的方向是改进方向.则改进方向的集合$F = \{d | -8d_1 + 2d_2 < 0\}$.因此改进可行方向为$D \cap F$.

Zoutendijk的可行方向法

对可行方向法来说,关键的一步是如何具体地生成改进的可行方向.对于线性约束问题(4-2),由定理4.1知:可由求解如下的规划问题P_1, P_2或P_3来得到d.设x为(4-2)的可行点.

$$\begin{cases} \min & z = \nabla f(x)^T d \\ \text{s.t.} & A_1 d \leq 0 \\ & Ed = 0 \\ & -1 \leq d_j \leq 1, j = 1, 2, \cdots, n \end{cases} \qquad (P_1)$$

$$
\begin{cases}
\min & z = \nabla f(x)^T d \\
\text{s.t.} & A_1 d \leq 0 \\
 & Ed = 0 \\
 & d^T d \leq 1
\end{cases}
\tag{P_2}
$$

$$
\begin{cases}
\min & z = \nabla f(x)^T d \\
\text{s.t.} & A_1 d \leq 0 \\
 & Ed = 0 \\
 & \nabla f(x)^T d \geq -1
\end{cases}
\tag{P_3}
$$

其中$d = (d_1, d_2, \cdots, d_n)^T$.最后一个约束是为了保证解有界的规格化约束.

线性约束的Zoutendijk方法的计算步骤 I

(1) 求问题(4-2)的一初始可行解x^1,令$k=1$转(2).

(2) 对于可行点x^k,设$A_1x^k=b_1, A_2x^k<b_2$,求解问题P_1, P_2或P_3得最优解d_k, 如果$\nabla f(x^k)^T d_k=0$,计算结束,x^k是K-T点;否则转(3).

(3) 求解线搜索问题

$$\begin{array}{ll} \min & f(x+\lambda d_k) \\ \text{s.t.} & 0\le\lambda\le\lambda_{\max} \end{array} \tag{4-4}$$

其中

$$\lambda_{\max}=\begin{cases} +\infty, & 若\hat{d}\le 0 \\ \min\left\{\hat{b}_i/\hat{d}_i \mid \hat{d}_i>0\right\}, & 否则 \end{cases} \tag{4-5}$$

$$\hat{b}=b_2-A_2x^k, \hat{d}=A_2d_k \tag{4-6}$$

设λ_k为(4-4)的最优解, 令$x^{k+1}=x^k+\lambda_k d_k, k=k+1$,返回(2).

线性约束的Zoutendijk方法的计算步骤 II

- 由于计算总会有误差,所以一般来说$\nabla f(x^k)^T d_k = 0$不可能准确地成立, 可取一个允许误差$\varepsilon_1 > 0$,若$|\nabla f(x^k)^T d_k| < \varepsilon_1$,计算即可结束.
- 若可行点x^k为内点,即$Ax^k < b$,这时不必去求解线性规划,可取$d_k = -\nabla f(x^k)$, 若$|z_k| < \varepsilon_2$,计算结束,其中ε_2为允许误差.

改进的可行方向可通过求解如下的线性规划问题来得到

$$\begin{array}{ll} \min & z \\ \text{s.t.} & \nabla f(x)^T d - z \leq 0 \\ & \nabla g_i(x)^T d - z \leq 0, i \in I(x) \\ & -1 \leq d_j \leq 1, j = 1, \cdots, n \end{array} \tag{4-7}$$

其中x为(4-3)的可行点.
令$(\bar{z}, \bar{d})$为(4-7)的最优解,易见$\bar{z} \leq 0$.若$\bar{z} < 0$,由定理4.2知$\bar{d}$是一改进的可行方向;否则可以证明,在一定条件下,x是K-T点的充要条件为$\bar{z} = 0$.

(1) 选取允许误差$\varepsilon_1 > 0, \varepsilon_2 > 0$,求一初始可行点$x^1$,令$k = 1$.

(2) 确定指标集$I(x^k) = \{i | g_i(x^k) = 0\}$.

(4) 若$I(x^k) = \emptyset$,且$\|\nabla f(x^k)\| \le \varepsilon_1$,计算结束,取$x^* = x^k$;若$I(x^k) = \emptyset$,但$\|\nabla f(x^k)\| > \varepsilon_1$,令$d_k = -\nabla f(x^k)$,转(6);若$I(x^k) \neq \emptyset$,转(4);

(4) 在(4-7)中,令$x = x^k$,求出问题(4-7)的最优解$(\bar{z}_k, \bar{d}_k)$.

(5) 若$|\bar{z}_k| < \varepsilon$,计算结束,取$x^* = x^k$;否则令$d_k = \bar{d}_k$,转(6).

(6) 求出线搜索问题

$$\min_{0 \le \lambda \le \lambda_{max}} \quad f(x^k + \lambda d_k)$$

的最优解λ_k,其中$\lambda_{max} = \max\{\lambda | x^k + \lambda d_k \in S\}$,
令$x^{k+1} = x^k + \lambda_k d_k, k = k + 1$,返回(2).

例4.2

用Zoutendijk方法求解如下问题

$$
\begin{aligned}
\min \quad & f(x) = (x_1 - 6)^2 + (x_2 - 2)^2 \\
\text{s.t.} \quad & -x_1 + 2x_2 \leq 4 \\
& 3x_1 + 2x_2 \leq 12 \\
& -x_1 \leq 0 \\
& -x_2 \leq 0
\end{aligned}
$$

Zoutendijk的可行方向法的缺点

可能出现"锯齿现象"使得收敛速度很慢,甚至不收敛于K-T点. 早在1967年Topkis和Veinott提出了一种修正的可行方向法.可以证明:在一定条件下, 由这种算法产生的点列$\{x^k\}$ 是收敛于K-T点的.

4.5 梯度投影法

本节介绍Rosen的梯度投影法, 这是他在1960年[a]针对线性约束的优化问题首先提出来的,1961年Rosen[b]又将他的方法推广到处理非线性约束的情况,后来这类方法又得到了进一步的发展,成为求解非线性规划问题的一类重要的方法.

[a] J.B. Rosen, *The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints*,J. Soc. Indust. and Appl. Math. (8),181-217,1960

[b] J.B. Rosen, *The Gradient Projection Method for Nonlinear Programming. Part II. Nonlinear Constraints*,J. Soc. Indust. and Appl. Math. (9),514-532,1961

考虑线性约束的优化问题

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax \le b \\ & Ex = e \end{array} \tag{5-1}$$

其中A是$m \times n$矩阵,E是$l \times n$矩阵, $b \in \mathbf{R}^m, e \in \mathbf{R}^l, x \in \mathbf{R}^n, f : \mathbf{R}^n \to \mathbf{R}$, $f \in C^1$. 用Ω表示(5-1)的可行域.

梯度投影法的基本思想

当迭代点x^k在可行域Ω的内部时,取$d = -\nabla f(x^k)$为迭代方向; 当点x^k在Ω的边界上时,取$-\nabla f(x^k)$在这些边界面的交集上的投影为迭代的方向.

定义

称$n \times n$矩阵P为**投影矩阵**,如果$P = P^T$, $PP = P$.

引理1

设P为$n \times n$矩阵

(1) 若P为投影矩阵,则P是半正定的;

(2) P是投影矩阵,当且仅当$I - P$是投影矩阵,其中I为n阶单位阵;

(3) 设P是投影矩阵,令$Q = I - P$,则

$$L = \{Px | x \in \mathbf{R}^n\} \text{与} L^{\perp} = \{Qx | x \in \mathbf{R}^n\}$$

为互相正交的线性子空间,并且任一点$x \in \mathbf{R}^n$可唯一地表示为$x = p + q, p \in L, q \in L^{\perp}$.

引理2

设x为问题(5-1)的一个可行解,且使$A_1x=b_1,A_2x<b_2$,
而$A^T=(A_1^T,A_2^T),b^T=(b_1^T,b_2^T)$,设$f(x)$在$x$点可微,$M^T=(A_1^T,E^T)$满秩, 则

(1) $P=I-M^T(MM^T)^{-1}M$为投影矩阵;

(2) 若$P\nabla f(x)\neq 0$,则$d=-P\nabla f(x)$是x点的一个改进的可行方向.

证明.

(1)容易验证.只需证(2).

- $\nabla f(x)^Td=-\nabla f(x)^TP^TP\nabla f(x)<0$,因此$d$是改进方向.
- $Md=-MP\nabla f(x)=0$,即$A_1d=0,Ed=0$,则d是可行方向.

□

定理5.1

设x是问题(5-1)的一个可行解,且使$A_1x=b_1, A_2x<b_2$,
而$A^T=(A_1^T,A_2^T), b^T=(b_1^T,b_2^T), M^T=(A_1^T,E^T)$满秩,令

$$P=I-M^T(MM^T)^{-1}M, w=-(MM^T)^{-1}M\nabla f(x), w^T=(u^T,v^T).$$

设$P\nabla f(x)=0$,

(1) 若$u\geq 0$,则x是一个K-T点;

(2) 若$u\not\geq 0$,令u_j是u的一个负分量,$\hat{M}^T=(\hat{A}_1^T,E^T)$,其中$\hat{A}_1$ 是由A_1中去掉第j行后得到的矩阵,令

$$\hat{P}=I-\hat{M}^T(\hat{M}\hat{M}^T)^{-1}\hat{M}, d=-\hat{P}\nabla f(x),$$

则d是一个改进的可行方向.

证明 I

(1) 因为

$$\begin{aligned} P\nabla f(x) &= [I - M^T(MM^T)^{-1}M]\nabla f(x) \\ &= \nabla f(x) + M^T w \\ &= \nabla f(x) + A_1^T u + E^T v = 0. \end{aligned} \tag{5-2}$$

上面最后一个等式即为$\nabla_x L(x,u,v) = 0$.若$u \geq 0$,则x为$K-T$点.

(2) 首先证明:$\hat{P}\nabla f(x) \neq 0$(反证法). 假设$\hat{P}\nabla f(x) = 0$.
令$\hat{w} = -(\hat{M}\hat{M}^T)^{-1}\hat{M}\nabla f(x)$. 则

$$\hat{P}\nabla f(x) = \nabla f(x) + \hat{M}^T\hat{w} = 0. \tag{5-3}$$

由于$A_1^T u + E^T v = \hat{M}^T\bar{w} + u_j a_j^T$,于是由(5-2)式可得

$$\nabla f(x) + \hat{M}^T\bar{w} + u_j a_j^T = 0. \tag{5-4}$$

其中a_j是A_1的第j行,$\bar{w}$为w去掉第j个分量. 将(5-3)和(5-4)两式相减可得$\hat{M}^T(\hat{w} - \bar{w}) - u_j a_j^T = 0$. 而$u_j < 0$,上式与$M$满秩矛盾.

(3) 易知$\hat{P}$为投影矩阵,则$d=-\hat{P}\nabla f(x)$为改进方向,且

$$\begin{bmatrix} \hat{A}_1 \\ E \end{bmatrix} d = \hat{M}d = -\hat{M}\hat{P}\nabla f(x) = 0 \tag{5-5}$$

要证明d是一个可行方向,只需证明$a_j d \le 0$.
用$a_j\hat{P}$乘以(5-4)式两端,并注意$\hat{P}\hat{M}^T=0$,可得

$$a_j\hat{P}\nabla f(x) + a_j\hat{P}(\hat{M}^T\bar{w} + u_j a_j^T) = -a_j d + u_j a_j \hat{P} a_j^T = 0. \tag{5-6}$$

由引理1可知:$\hat{P}$是半正定的,因此$a_j d \le 0$.

在问题(5-1)中,设

$$A=\begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{pmatrix}, E=\begin{pmatrix} a_{m+1}^T \\ a_{m+2}^T \\ \vdots \\ a_{m+l}^T \end{pmatrix}, B=\begin{pmatrix} b \\ e \end{pmatrix}=\begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_{m+l} \end{pmatrix}$$

$J_k=J(x^k)=\{j|a_j^Tx^k=B_j\}$.$J_k$称为$x^k$点的指标集.$M_k$表示以$a_j^T(j\in J_k)$为行所组成的矩阵.

计算步骤 I

1. 选取x^1为(5-1)的一个可行解,给定计算精度$\varepsilon > 0$, 令k=1.
2. 计算$\nabla f(x^k)$, $J_k = \{j | a_j^T x^k = B_j\}$,若$\|\nabla f(x^k)\| \le \varepsilon$, 则$x^k$为近似的K-T点,计算结束;否则
   (1) 若$J_k = \emptyset$(空集),令P=I,
   (2) 若$J_k \neq \emptyset$,令$P = I - M_k^T (M_k M_k^T)^{-1} M_k$.
3. 若$P\nabla f(x^k) \neq 0$,令$d_k = -P\nabla f(x^k)$,转4; 若$P\nabla f(x^k) = 0$,令

$$w = -(M_k M_k^T)^{-1} M_k \nabla f(x^k) = \begin{pmatrix} u \\ v \end{pmatrix},$$

- 若$u \ge 0$, 则x^k为K-T点,计算结束;
- 若u有某个分量$u_j < 0$,令$\hat{M}_k$是M_k中去掉与u_j 对应的第j行而得到的矩阵.

令$\hat{P} = I - \hat{M}_k^T (\hat{M}_k \hat{M}_k^T)^{-1} \hat{M}_k$, $d_k = -\hat{P}\nabla f(x^k)$,转4.

计算步骤 II

4. 计算

$$\lambda_M = \begin{cases} \min\{\frac{B_i - a_i^T x^k}{a_i^T d_k} | i \notin J_k, a_i^T d_k > 0\} \\ +\infty, \text{ 若} \forall i, a_i^T d_k \leq 0. \end{cases} \tag{5-7}$$

设λ_k是线搜索问题

$$\begin{array}{ll} \min & f(x^k + \lambda d_k) \\ \text{s.t.} & 0 \leq \lambda \leq \lambda_M \end{array}$$

的解,令$x^{k+1} = x^k + \lambda_k d_k$, $k = k + 1$, 返回2.

用上述的梯度投影法求解如下的优化问题

$$\begin{aligned} \min \quad & f(x)=2x_1^2+2x_2^2-2x_1x_2-4x_1-6x_2 \\ \text{s.t.} \quad & x_1+x_2\le 2 \\ & x_1+5x_2\le 5 \\ & -x_1\le 0,\ \ -x_2\le 0 \end{aligned} \tag{5-9}$$

解: $\nabla f(x)=(4x_1-2x_2-4,4x_2-2x_1-6)^T$.因为本例没有等式约束,所以$M=A_1,w=u$.

第一次迭代 取$x^1=(0,0)^T$.

1 确定搜索方向

$$
\begin{aligned}
&\nabla f(x^1) = (-4,-6)^T \\
&M_1 = A_1 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},\ J_1 = \{3,4\} \\
&P = I - A_1^T(A_1A_1^T)^{-1}A_1 = 0 \\
&P\nabla f(x^1) = (0,0)^T
\end{aligned}
$$

则$u = -(A_1A_1^T)^{-1}A_1\nabla f(x^1) = (-4,-6)^T$,取$u_4 = -6$,从$A_1$中去掉与$u_4$对应的行得到

$$
\begin{aligned}
&\hat{A}_1 = (-1,0),\ \hat{P} = I - \hat{A}_1^T(\hat{A}_1\hat{A}_1^T)^{-1}\hat{A}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&d_1 = -\hat{P}\nabla f(x^1) = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -4 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}
\end{aligned}
$$

2 进行线搜索 $f(x^1+\lambda d_1)=72\lambda^2-36\lambda, \lambda_M=\min\{2/6, 5/30\}=1/6.$
求解

$$\min_{0\leq\lambda\leq 1/6} f(x^1+\lambda d_1)=72\lambda^2-36\lambda$$

得$\lambda_1=1/6,$于是$x^2=x^1+\lambda_1 d_1=(0,1)^T.$

第二次迭代

1 确定搜索方向

$$\begin{aligned}
&\nabla f(x^2)=(-6,-2)^T\\
&M_2=A_2=\begin{pmatrix}1 & 5\\ -1 & 0\end{pmatrix},\ J_2=\{2,3\}\\
&P=I-A_2^T(A_2A_2^T)^{-1}A_2=0\\
&P\nabla f(x^2)=(0,0)^T
\end{aligned}$$

则$u=-(A_2A_2^T)^{-1}A_2\nabla f(x^2)=(2/5,-28/5)^T$,取$u_3=-28/5<0$,从$A_2$中去掉与$u_3$对应的行得到

$$\hat{A}_2=(1,5),\ \ \hat{P}=I-\hat{A}_2^T(\hat{A}_2\hat{A}_2^T)^{-1}\hat{A}_2=\begin{pmatrix}25/26 & -5/26\\ -5/26 & 1/26\end{pmatrix}$$

$$d_2=-\hat{P}\nabla f(x^2)=\begin{pmatrix}70/13\\ -14/13\end{pmatrix}$$

由于d_2的模的大小不重要，不妨取$d_2=(5,-1)^T$.

2 进行线搜索 $f(x^2+\lambda d_2)=62\lambda^2-28\lambda-4,\lambda_M=\min\{1/4,1\}=1/4$. 求解

$$\min_{0\le\lambda\le 1/4} 62\lambda^2-28\lambda-4$$

得$\lambda_2=7/31$,于是$x^3=x^2+\lambda_2 d_2=(35/31,24/31)^T$.

第三次迭代

$$
\begin{aligned}
&\nabla f(x^3) = (-32/31, -160/31)^T \\
&M_3 = A_3 = (1,5),\ \ J_3 = \{2\} \\
&P = I - A_3^T(A_3A_3^T)^{-1}A_3 = 1/26 \begin{pmatrix} 25 & -5 \\ -5 & 1 \end{pmatrix} \\
&d_3 = -P\nabla f(x^3) = (0,0)^T \\
&u = -(A_3A_3^T)^{-1}A_3\nabla f(x^3) = 32/31 > 0
\end{aligned}
$$

所以x^3是K-T点,又因为$f(x)$是凸函数,因此x^3是全局最优解.

4.6 既约梯度法(Reduced Gradient Method)

基本思想

利用约束条件将问题中的某些变量用其他的一组独立变量来表示,从而使问题的维数降低. 利用既约梯度,直接构造出一个改进的可行方向然后沿此方向进行线搜索,从而求得一个新点, 这样一步步来逼近原问题的最优解.

4.7 乘子法

乘子法(Multiplier Method),也叫增广拉格朗日方法(Augmented Lagrangian Method).乘子法是由Powell [1]和Hestenes[2]于1969年彼此独立地对等式约束的优化问题首先提出来的. 1973年,Rockafellar[3]将其推广到不等式约束的优化问题,十多年来获得了迅速的发展, 成为求解约束优化问题的一类重要而有效的方法.Bertsekas在乘子法方面做过较多的工作,其中专著“Constrained Optimization and Lagrange Multiplier Methods”对乘子法的理论作了系统的论述.

---

[1]M.J.D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization ed. by R. Fletcher,Academic Press,New York,NY,1969,283 – 298

[2]M.R. Hestenes,*Multiplier and gradient methods*,Journal of Optimization Theory and Applications,4(5),303-320,1969

[3]R.T. Rockafellar,*A dual approach to solving nonlinear programming problems by unconstrained optimization*,Mathematical Programming,5(1),354-373,1973

4.7.1 Hestenes的乘子方法

乘子法思想的来源

惩罚函数法的主要缺点之一是:当罚因子M越来越大时,惩罚目标函数的海赛矩阵越来越病态,使无约束优化方法的计算难以进行下去，而乘子法克服了这个缺点.

罚因子必须无限增大的原因

考察等式约束的优化问题

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h_i(x)=0, i=1,2,\cdots,p \end{aligned} \tag{7-1}$$

引入罚函数

$$F(x,M)=f(x)+M\sum_{i=1}^{p}[h_i(x)]^2 \tag{7-2}$$

若x^*是问题(7-1)的最优解,则$h_j(x^*)=0, j=1,2,\cdots,p$.

于是罚函数在x^*点的梯度为

$$\nabla_x F(x^*, M) = \nabla f(x^*) + 2M \sum_{j=1}^{p} h_j(x^*) \nabla h_j(x^*) = \nabla f(x^*),$$

而由K-T条件应有

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) = 0 \tag{7-3}$$

所以,一般说来$\nabla f(x^*) \neq 0$,所以找不到一个有限的M, 使$\nabla_x F(x^*, M) = 0$成立, 只能期望$\lim_{M\to\infty} \nabla_x F(x^*, M) = 0$.

Hestenes经过分析,引入增广Lagrange函数:

$$\varphi(x,\mu)=f(x)+\sum_{j=1}^{p}\mu_j h_j(x)+\frac{c}{2}\sum_{j=1}^{p}[h_j(x)]^2 \tag{7-4}$$

其中$\mu_1,\mu_2,\cdots,\mu_p$为Lagrange乘子,这里将M改写为$\frac{c}{2}$.(7-4)式也可改写为

$$\varphi(x,\mu)=L(x,\mu)+\frac{c}{2}\sum_{j=1}^{p}[h_j(x)]^2$$

其中$L(x,\mu)=f(x)+\sum_{j=1}^{p}\mu_j h_j(x)$为(7-1)的Lagrange函数.

由K-T条件可知:存在μ^*,使(x^*, μ^*)为$L(x, \mu)$的稳定点,即

$$\nabla_x L(x^*, \mu^*) = 0 \tag{7-5}$$

而附加项$\frac{c}{2}\sum_{j=1}^{p}[h_j(x)]^2$在$x^*$点的梯度为零.因此

$$\nabla_x \varphi(x^*, \mu^*) = 0,$$

这说明x^*是$\varphi(x, \mu^*)$的稳定点.可以证明:存在$c_1 > 0$,当$c \geq c_1$时,x^*将是$\varphi(x, \mu^*)$的极小点,而求解问题(7-1)便可转化为对某个μ^*, 求$\varphi(x, \mu^*)$的无约束极小点了.

μ^*的求法

首先给定一个足够大的正数c,然后在迭代过程中,逐步调整μ^k,如何调整呢? 假设已求得μ^k,求解$\min \varphi(x,\mu^k)$得最优解x^k,则

$$\nabla_x \varphi(x^k,\mu^k) = \nabla f(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla h_j(x^k) + c\sum_{j=1}^{p} h_j(x^k)\nabla h_j(x^k) = 0$$

即

$$\nabla f(x^k) + \sum_{j=1}^{p} (\mu_j^k + c h_j(x^k))\nabla h_j(x^k) = 0 \qquad (7\text{-}6)$$

为了使μ^k, x^k分别逼近μ^*, x^*,将(7-6)式与(7-3)式进行比较,于是我们应采用公式

$$\mu_j^{k+1} = \mu_j^k + c h_j(x^k), j = 1, 2, \cdots, p \qquad (7\text{-}7)$$

来调整μ^k.

μ^k的停止准则

定理7.1

设x^k是无约束优化问题

$$\min \ \varphi(x, \mu^k) = f(x) + \sum_{j=1}^{p} \mu_j^k h_j(x) + \frac{c}{2} \sum_{j=1}^{p} [h_j(x)]^2$$

的最优解,则x^k是问题(7-1)的最优解,μ^k为其相应的Lagrange乘子的充要条件是$h_j(x^k) = 0, j = 1, 2, \cdots, p$.

迭代计算可采用如下的结束准则:选定计算精度$\varepsilon > 0$,

- 若x^k 满足$\|h(x^k)\| < \varepsilon$,则停止迭代,否则继续进行迭代.
- 若在迭代过程中,μ^k收敛太慢或不收敛,可适当增大c.

乘子法的迭代步骤

(1) 给定初始点x^0,初始乘子向量μ^1(若无其他信息,可取$\mu^1 = 0$), 计算精度$\varepsilon > 0$,取$c > 0, 0 < r < 1$(如取$r = 0.25$),$\alpha > 1$(如取$\alpha = 2 \sim 10$), 令$k = 1$.

(2) 以x^{k-1}为初始点,求解$\min \ \varphi(x, \mu^k)$,得解x^k,其中$\varphi(x, \mu)$ 由(7-4)式确定.

(3) 若$\|h(x^k)\| < \varepsilon$,计算结束,取x^k为(7-1)的最优解;否则计算$\beta = \|h(x^k)\|/\|h(x^{k-1})\|$,若$\beta \le r$,转到第(4)步,否则令$c = \alpha c$, 转到第(4).

(4) 计算$\mu_j^{k+1} = \mu_j^k + ch_j(x^k)(j = 1, 2, \cdots, p)$,令$k = k + 1$,返回第(1)步.

4.7.2 Powell的乘子方法

Powell考虑含有两组参数的罚函数

$$M(x,\sigma,\alpha) = f(x) + \frac{1}{2}\sum_{j=1}^{p}\sigma_i[h_j(x)+\alpha_j]^2 \tag{7-9}$$

在上式中,若令$\sigma_j = c, \alpha_j = \frac{\mu_j}{c}$,则得

$$\begin{aligned} M(x,\sigma,\alpha) &= f(x) + \tfrac{c}{2}\sum_{j=1}^{p}([h_j(x)]^2 + 2\tfrac{\mu_j}{c}h_j(x) + (\tfrac{\mu_j}{c})^2) \\ &= f(x) + \tfrac{c}{2}\sum_{j=1}^{p}[h_j(x)]^2 + \sum_{j=1}^{p}\mu_j h_j(x) + \tfrac{1}{2c}\sum_{j=1}^{p}\mu_j^2, \end{aligned}$$

将上式与(7-4)对比,可得

$$M(x,\sigma,\alpha)=\varphi(x,\mu)+\frac{1}{2c}\sum_{j=1}^{p}\mu_j^2.$$

注意:$M(x,\sigma,\alpha)$与$\varphi(x,\mu)$仅差一个常数项. 因此,Powell方法是一种比Hestenes的方法更广泛一些的乘子方法,但两者的出发点不同.

定理7.2

设对某两组参数σ和α, $x^*(\sigma,\alpha)$是$M(x,\sigma,\alpha)$的无约束极小点, 则$x^*(\sigma,\alpha)$也是问题

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h_j(x) = h_j(x^*(\sigma,\alpha)), j = 1,2,\cdots,p \end{array} \tag{7-10}$$

的极小点.

若能找到两组参数σ和α,使

$$h_j(x^*(\sigma,\alpha)) = 0, j = 1,2,\cdots,p \tag{7-11}$$

就可使$x^*(\sigma,\alpha)$就是问题(7-1)的最优解. 为此我们期望产生一列参数值σ^k和α^k,使$\lim\limits_{k\to\infty} h_j(x(\sigma^k,\alpha^k)) = 0, j = 1,2,\cdots,p$. 则当$k\to\infty$时,就可使$x(\sigma^k,\alpha^k)$的极限$x^*(\sigma,\alpha)$满足(7-11)式.

为了得到$\alpha_j^k(j=1,2,\cdots,p)$的调整公式,求出$M(x,\sigma,\alpha)$ 对x的梯度

$$\nabla_x M(x,\sigma,\alpha)=\nabla f(x)+\sum_{j=1}^{p}\sigma_j[h_j(x)+\alpha_j]\nabla h_j(x) \tag{7-12}$$

将(7-12)与(7-3)对比,并注意到$\alpha_j=\frac{\mu_j}{c},\sigma_j=c$,即知: 对固定的充分大的$\sigma_j(j=1,2,\cdots,p)$,可采用公式

$$\alpha_j^{k+1}=\alpha_j^k+h_j(x(\sigma,\alpha^k))),j=1,2,\cdots,p \tag{7-13}$$

来调整α_j^k.若α_j^k收敛太慢或不收敛,可增大$\sigma_j,j=1,2,\cdots,p$.
关于Powell乘子方法的计算步骤与前述的Hestenes乘子方法类似,不再详述.

1973年,Rockafellar将乘子方法推广到不等式约束的优化问题:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0, i = 1, 2, \cdots, p \end{array}$$

首先,引入松弛变量$z_i (i = 1, 2, \cdots, p)$,将不等式约束化为等式约束,即

$$g_i(x) - z_i^2 = 0, i = 1, 2, \cdots, p. \tag{7-15}$$

然后,利用前面讲述的等式约束优化问题的结果推出相应的结果.

考虑在等式约束(7-15)下的增广Lagrange函数:

$$\Phi(x,z,\mu)=f(x)+\sum_{j=1}^{p}\mu_j[g_j(x)-z_j^2]+\frac{c}{2}\sum_{j=1}^{p}[g_j(x)-z_j^2]^2 \tag{7-16}$$

将$\Phi(x,z,\mu)$关于z求极小,令

$$\varphi(x,\mu)=\min_z \Phi(x,z,\mu),\quad \nabla_z\Phi(x,z,\mu)=0,$$

得

$$z_j[cz_j^2-(\mu_j+cg_j(x))]=0, j=1,2,\cdots,p$$

由

$$z_j[cz_j^2 - (\mu_j + cg_j(x))] = 0, j = 1, 2, \cdots, p$$

- 若$\mu_j + cg_j(x) \leq 0$,则$z_j^2 = 0$;
- 若$\mu_j + cg_j(x) > 0$,则$z_j^2 = \frac{1}{c}(\mu_j + cg_j(x))$.

因此,

$$g_j(x) - z_j^2 = \begin{cases} g_j(x), & 若\mu_j + cg_j(x) \leq 0 \\ -\frac{\mu_j}{c} & 若\mu_j + cg_j(x) > 0, \end{cases} \tag{7-17}$$

当$\mu_j + cg_j(x) \le 0$时,有

$$\begin{aligned} \mu_j[g_j(x) - z_j^2] + \tfrac{c}{2}[g_j(x) - z_j^2]^2 &= \mu_j g_j(x) + \tfrac{c}{2} g_j(x)^2 \\ &= \tfrac{1}{2c}[(\mu_j + cg_j(x))^2 - \mu_j^2]. \end{aligned}$$

当$\mu_j + cg_j(x) > 0$时,有

$$\begin{aligned} \mu_j[g_j(x) - z_j^2] + \tfrac{c}{2}[g_j(x) - z_j^2]^2 &= -\tfrac{\mu_j^2}{c} + \tfrac{c}{2}(-\tfrac{\mu_j}{c})^2 \\ &= -\tfrac{1}{2c}\mu_j^2. \end{aligned}$$

所以

$$\varphi(x,\mu) = \min_z \Phi(x,z,\mu) = f(x) + \frac{1}{2c}\sum_{j=1}^{p}\{[\min(0, \mu_j + cg_j(x))]^2 - \mu_j^2\}.$$

利用等式约束优化问题乘子迭代的修正公式(7-7)和(7-17) 可得乘子迭代公式为

$$\mu_j^{k+1} = \min\{0, \mu_j^k + cg_j(x^k)\}. \tag{7-19}$$

结束准则:

$$(\sum_{j=1}^{p}[\min(g_j(x), -\frac{\mu_j}{c})]^2)^{\frac{1}{2}} < \varepsilon$$

其中$\varepsilon > 0$为计算精度.

对于一般的非线性最优化问题

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & h_i(x) = 0, i = 1, 2, \cdots, m \\
& g_j(x) \geq 0, j = 1, 2, \cdots, p
\end{aligned}
\tag{7-21}
$$

与上述类似,可令

$$
\begin{aligned}
\varphi(x, \lambda, \mu) = \quad & f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{c}{2} \sum_{i=1}^{m} [h_i(x)]^2 \\
& + \frac{1}{2c} \sum_{j=1}^{p} \{[\min(0, \mu_j + c g_j(x))]^2 - \mu_j^2\}
\end{aligned}
\tag{7-22}
$$

这时的乘子迭代公式与(7-7)式和(7-19)式类似,而迭代计算的结束准则可采用

$$\sum_{i=1}^{m} h_i^2(x^k) + \sum_{j=1}^{p}[\min(g_j(x^k), -\frac{\mu_j^k}{c})]^2 < \varepsilon^2 \tag{7-23}$$

例7.1

用乘子法求解问题

$$\begin{array}{ll} \min & f(x)=x_1^2+x_2^2 \\ \text{s.t.} & x_1-1\geq 0 \end{array}$$

解 增广Lagrange函数为

$$\begin{aligned} \varphi(x,\mu) &= x_1^2+x_2^2+\frac{1}{2c}\{[\min(0,\mu+c(x_1-1))]^2-\mu^2\} \\ &= \begin{cases} x_1^2+x_2^2+\mu(x_1-1)+\frac{c}{2}(x_1-1)^2, & x_1\leq 1-\frac{\mu}{c} \\ x_1^2+x_2^2-\frac{\mu^2}{2c}, & x_1>1-\frac{\mu}{c} \end{cases} \end{aligned}$$

为求解$\min\varphi(x,\mu)$,一般需用无约束优化方法求解,这里函数比较简单,可用解析法求解.

$$\frac{\partial\varphi}{\partial x_1}=\begin{cases} 2x_1+\mu+c(x_1-1), & x_1\leq 1-\frac{\mu}{c}, \\ 2x_1, & x_1>1-\frac{\mu}{c}, \end{cases} \qquad \frac{\partial\varphi}{\partial x_2}=2x_2$$

取$c=4, \mu^1=0$. 令$\nabla_x \varphi(x,\mu)=0$,得$\varphi(x,\mu)$的极小点为

$$x_1=\frac{c-\mu}{c+2}, \quad x_2=0.$$

则$x_1^1=2/3, x_2^1=0$,于是

$$\mu^2=\min\{0, 0+4(2/3-1)\}=-4/3$$

一般的$x^k=\left(\frac{4-\mu^k}{6}, 0\right)^T$,

$$\mu^{k+1}=\min\{0, \mu^k+4(x_1^k-1)\}=\frac{\mu^k-4}{3}$$

所以,当$k\to+\infty$时,$\mu^k\to-2, x^k\to(1,0)^T=x^*$.

4.8 二次逼近法

线性规划和二次规划都比较容易求解,人们很自然想到:

(1) 将求解的非线性约束优化问题线性化,利用线性规划方法来逐步求其近似解,这种方法称为线性逼近法或序列线性规划法, 简写为SLP法. 注意:线性逼近法逼近的精度差,收敛速度慢.

(2) 将求解的非线性约束优化问题, 用二次规划来逐步逼近. 这种方法称为二次逼近法或序列二次规划法,简写为SQP法.SQP法是当前世界上最流行的约束优化算法之一.

二次规划的概念

所谓二次规划是指在变量$x_1, x_2, \cdots, x_n$的线性等式和线性不等式约束下，求二次函数$Q(x)$的极小值问题:

$$
\begin{array}{ll}
\min & Q(x) = \frac{1}{2}x^T Gx + g^T x \\
\text{s.t.} & a_i^T x = b_i, i = 1, 2, \cdots, m \\
 & a_i^T x \leq b_i, i = m+1, \cdots, p
\end{array}
\qquad (8\text{-}1)
$$

其中G为n阶对称矩阵,$g, a_1, \cdots, a_p$均为n维列向量，假设$a_1, \cdots, a_m$ 线性无关,$x \in \mathbf{R}^n$.$b_1, b_2, \cdots, b_p$为已知常数,$m \leq n, p \geq m$. 用S表示(8-1)的可行集.

- 二次规划问题是最简单的一类非线性规划问题,它不仅有自己的实际背景,而且是SQP法的基础.
- 若$G \geq 0$(半正定),则问题(8-1)就是一个凸规划问题,它的任何局部最优解,也是全局最优解,简称整体解.
- 若$G > 0$(正定),则$Q(x)$为严格凸函数,问题(8-1)若存在整体解,则它是唯一的.
- 问题(8-1)的约束可能不相容,也可能没有有限的最小值,这时称问题QP无解.

例8.1

考虑二次规划问题

$$
\begin{aligned}
\min \quad & z = x_1^2 + x_2^2 - 2x_1 - 4x_2 \\
\text{s.t.} \quad & x_1 + x_2 = 1 \\
& x_1, x_2 \geq 0
\end{aligned}
\tag{8-2}
$$

将$z = Q(x)$写成$z = Q(x) = \frac{1}{2}x^T Gx + g^T x$,其中

$$g = (-2, -4)^T, \; x = (x_1, x_2)^T$$

而$G = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$,可行集

$$S = \{x | x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}.$$

易知:问题(8-2)为严格凸二次规划问题,其唯一整体解为$x^* = (0, 1)^T$.

定义

设$\bar{x}$是问题(8-1)的可行解,若某个$i \in \{1, 2, \cdots, p\}$,使得$a_i^T \bar{x} = b_i$成立, 称它为$\bar{x}$点处的有效约束,称在$\bar{x}$点处所有有效约束的指标组成的集合$J = J(\bar{x}) = \{i | a_i^T \bar{x} = b_i\}$为$\bar{x}$点处的有效约束指标集, 简称为$\bar{x}$点处的有效集.

注意:

- 显然对于任何可行点,所有等式约束都是有效约束,只有不等式约束才可能是非有效约束.
- 注意有效约束和紧约束的区别.

正定二次规划问题解的性质

在问题(8-1)中假设$G > 0$,且令$E = \{1, 2, \cdots, m\}$, $\quad I = \{m+1, \cdots, p\}$.

定理8-1

点x^*是正定二次规划问题(8-1)的严格整体解的充要条件是x^*为K-T点，即有乘子向量$\lambda^* = (\lambda_1^*, \lambda_m^*, \lambda_{m+1}^*, \cdots, \lambda_p^*)^T$, 使得

$$
\begin{array}{l}
Gx^* + g + \sum\limits_{i\in E} \lambda_i^* a_i + \sum\limits_{i \in I} \lambda_i^* a_i = 0 \\
a_i^T x^* - b_i = 0, \quad i \in E \\
a_i^T x^* - b_i \leq 0, \quad i \in I \\
\lambda_i^* \geq 0, i \in I; \lambda_i^* = 0, i \in I \backslash J^*
\end{array} \tag{8-3}
$$

其中J^*为x^*处的有效集.

K-T点x^*与其相应的乘子向量λ^*,称为问题(8-1)的K-T对.

定理8.2

若正定QP问题(8-1)有可行解,则它必有最优解,且最优解是唯一的.

定理8.3

设x^*是正定问题(8-1)的最优解且在x^*点处的有效集为J^*,则x^* 是如下等式约束问题的唯一解.

$$\begin{array}{ll} \min & Q(x) = \frac{1}{2}x^T G x + g^T x \\ \text{s.t.} & a_i^T x = b_i, i \in J^* \end{array} \tag{8-4}$$

仅含等式约束的正定二次规划

下面讨论仅含等式约束的正定QP问题:

$$\begin{array}{ll} \min & Q(x) = \frac{1}{2}x^T Gx + g^T x \\ \text{s.t.} & a_i^T x = b_i, i \in E = \{1, 2, \cdots, m\} \end{array} \tag{8-6}$$

这里假设$a_1, \cdots, a_m$ 线性无关,G为n阶对称正定矩阵,$m \leq n$.
令$A = (a_1, a_2, \cdots, a_m), b = (b_1, b_2, \cdots, b_m)^T$.

定理8.4

在上述假设下,问题(8-6)的K-T对(x^*, λ^*)是存在唯一的,且(x^*, λ^*)为(8-6)的K-T对的充要条件是它们满足如下的方程组:

$$\begin{pmatrix} G & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -g \\ b \end{pmatrix} \tag{8-7}$$

例8.2

求解

$$\begin{aligned} \min \quad & Q(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t.} \quad & x_1 + 2x_2 - x_3 = 4 \\ & x_1 - x_2 + x_3 = -2. \end{aligned}$$

这里$n = 3, m = 2, G = I_3, g = 0$.

解

$$A = (a_1, a_2) = \begin{pmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 1 \end{pmatrix}, \ b = \begin{pmatrix} 4 \\ -2 \end{pmatrix}, \ \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \ x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

与(8-7)相应的方程组为

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 2 & -1 \\ 0 & 0 & 1 & -1 & 1 \\ 1 & 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4 \\ -2 \end{pmatrix}$$

求解上面方程组,得到唯一解

$$x_1 = 2/7, x_2 = 10/7, x_3 = -6/7, \lambda_1 = -4/7, \lambda_2 = 2/7.$$

因此原问题最优解$x^* = (2/7, 10/7, -6/7)^T$,相应的乘子向量为$\lambda^* = (-4/7, 2/7)^T$.

一般正定二次规划的有效集法

基本思想

- 对一般正定QP问题(8-1),由定理(8-3)得知:只要能找到解x^*满足的有效集J^*, 就可通过求解等式约束的QP问题得到x^*.
- 先求出问题(8-1)的一个可行点x^1,并求出x^1点处的有效集J_1. 然后按照使目标函数值减少的原则对有效集不断进行调整,直到得到J^*为止.

方法的导出

设x^1为(8-1)的初始可行解, 其有效集为J_1,它含有t个元素, 不妨设$J_1 = \{1, 2, \cdots, m, m+1, \cdots, t\}$,且$a_i(i \in J_1)$线性无关.设问题

$$\begin{array}{ll} \min & Q(x) = \frac{1}{2}x^T Gx + g^T x \\ \text{s.t.} & a_i^T x = b_i, i \in J_1 \end{array}$$

的最优解为$\bar{x}^1$,相应的乘子向量为λ^1.下面分别对$\bar{x}^1 = x^1$和$\bar{x}^1 \neq x^1$进行讨论.

$\bar{x}^1 \neq x^1$的情况

a. 若$\bar{x}^1$为原问题(8-1)的可行解, 即$\bar{x}^1 \in S$,令$x^2 = \bar{x}^1$,并求出x^2点处的有效集J_2, 用x^2和J_2代替x^1和J_1,重复前述步骤.

b. 若$\bar{x}^1 \notin S$,而$x^1 \in S$,令$d_1 = \bar{x}^1 - x^1$,易见, 从x^1出发沿方向d_1前进到$\bar{x}^1$的过程中,$Q(x)$是减少的,而且在到达$\bar{x}^1$ 之前必然会遇到某个不等式约束$a_i^T x \le b_i, t+1 \le i \le p$的边界, 设最先遇到的那个约束的指标为$l$,相应的交点记为$x^2$,则有$Q(x^2) \le Q(x^1)$, x^2点处的有效集记为J_2,则x^2满足$a_i^T x = b_i, i \in J_2$,以x^2和J_2代替x^1和J_1,重复前述步骤.

$\bar{x}^1 = x^1$的情况

a. 若乘子$\lambda^1_{m+1}, \cdots, \lambda^1_t \geq 0$,则由定理8.1知$\bar{x}^1 = x^1$ 是问题(8-1)的最优解,计算结束.

b. 若乘子$\lambda^1_{m+1}, \cdots, \lambda^1_t$中有负值, 令

$$\lambda^1_q = \min\{\lambda^1_i, i = m+1, \cdots, t\},$$

则$\lambda^1_q < 0$. 这时取

$$x^2 = \bar{x}^1, J_2 = J_1 \backslash \{q\}.$$

以x^2和J_2代替x^1和J_1,重复前述步骤. 这时将出现$\bar{x}^2 \neq x^2$.

从上面的分析中得知:在方法的计算过程中,需要在已知x^k为

$$\begin{array}{ll} \min & Q(x) = \frac{1}{2}x^T Gx + g^T x \\ \text{s.t.} & a_i^T x = b_i, i \in J_k \end{array} \tag{8-10}$$

的可行点的条件下,求问题(8-10)的解$\bar{x}^k$.若令$d^k = \bar{x}^k - x^k$, 则可把问题转化为寻求d^k.事实上由$\bar{x}^k = x^k + d^k$可得

$$\begin{aligned} Q(\bar{x}^k) &= \frac{1}{2}(x^k + d^k)^T G(x^k + d^k) + g^T(x^k + d^k) \\ &= \frac{1}{2}(d^k)^T Gd^k + (Gx^k + g)^T d^k + Q(x^k) \end{aligned}$$

而$\bar{x}^k$满足约束$a_i^T x = b_i, i \in J_k$与d^k满足$a_i^T d^k = 0, i \in J_k$等价.

因此d^k是正定QP问题

$$\begin{array}{ll} \min & q(x) = \frac{1}{2} d^T G d + (G x^k + g)^T d \\ \text{s.t.} & a_i^T d = 0, i \in J_k \end{array} \tag{8-12}$$

的最优解.由d^k即可求得(8-10)的最优解$\bar{x}^k$.

正定二次规划问题(8-1)有效集法的计算步骤

1. 选取(8-1)的初始可行点x^1,确定x^1点处的有效集J_1，使$a_i(i \in J_1)$线性无关,令$k = 1$.
2. 求解仅含等式约束的正定二次规划问题(8-12),设其解为d^k.
3. 若$d^k = 0$,则计算相应的乘子$\lambda_i^k(i \in J_k)$,转4.;否则转5.
4. 若$i \in J_k \cap I$,都有$\lambda_i^k \geq 0$，则x^k为问题(8-1)的最优解,计算结束;否则求出$\lambda_q^k = \min\{\lambda_i^k | i \in J_k \cap I\}$,
   令$x^{k+1} = x^k, J_{k+1} = J_k \backslash \{q\}, k = k + 1$,返回2.
5. 若$\bar{x}^k = x^k + d^k$满足$a_i^T x \leq b_i, i \in I \backslash J_k$，则令$x^{k+1} = \bar{x}^k$,并求出$x^{k+1}$点处的有效集$J_{k+1}$,令$k = k + 1$,返回2.; 否则转6.
6. 计算步长$\alpha_k = \min\{-\frac{a_i^T x^k - b_i}{a_i^T d^k} | i \in I \backslash J_k, a_i^T d^k > 0\}$,令$x^{k+1} = x^k + \alpha_k d^k$,并求$x^{k+1}$点处的有效集$J_{k+1}$,令$k = k + 1$, 返回2.

例8.3 I

求解如下的QP问题

$$
\begin{array}{ll}
\min & Q(x) = x_1^2 + x_2^2 - 2x_1 - 4x_2 \\
\text{s.t.} & -x_1 \le 0 \\
& -x_2 \le 0 \\
& x_1 + x_2 - 1 \le 0
\end{array}
\tag{8-13}
$$

解 第一次迭代

1. 取初始可行点$x^1 = (0,0)^T, J_1 = \{1,2\}$
2. 因为$G = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, g = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$, 所以与(8-12)相应的问题为

$$
\begin{array}{ll}
\min & z = d_1^2 + d_2^2 - 2d_1 - 4d_2 \\
\text{s.t.} & -d_1 = 0, \quad -d_2 = 0
\end{array}
$$

它的解显然为$d_1^1 = 0, d_2^2 = 0.$

3. $d^1=0$,其对应的乘子向量为$\lambda^1=(-2,-4)^T$
4. 计算$\lambda_q^1=\min\{\lambda_1^1,\lambda_2^1\}=-4$,令

$$x^2=x^1=(0,0)^T, J_2=J_1\backslash\{2\}=\{1\}.$$

第二次迭代 求解问题

$$\begin{array}{ll}\min & z=d_1^2+d_2^2-2d_1-4d_2\\ \text{s.t.} & -d_1=0\end{array}$$

显然$d^2=(0,2)^T$.因为$d^2\neq 0$,转到步5,令$\bar{x}^2=x^2+d^2=(0,2)^T$.由于$\bar{x}^2$不是可行点,转步6,计算步长

$$\alpha_2=\min\{-(a_i^Tx^2-b_i)/a_i^Td^2|i=2,3,a_i^Td^2>0\}=\frac{1}{2}$$

令$x^3=x^2+\alpha_2d^2=(0,1)^T$,求出$J_2=\{1,3\}$,转入迭代三.

例8.3 III

第三次迭代 求解问题

$$\begin{array}{ll} \min & z = d_1^2 + d_2^2 - 2d_1 - 4d_2 \\ \text{s.t.} & -d_1 = 0,\ d_1 + d_2 = 0 \end{array}$$

得解$d^3 = (0,0)^T, \lambda^3 = (0,2)^T$.所以$x^3 = (0,1)^T$即为最优解,其相应的乘子向量为$\lambda^* = (0,0,2)^T$.

考虑一般的约束优化问题

$$
\begin{array}{lll}
\min & f(x) & \\
\text{s.t.} & c_i(x)=0, i\in E=\{1,2,\cdots,m\} & \\
 & c_i(x)\le 0, i\in I=\{m+1,\cdots,p\}, &
\end{array}
\tag{8-14}
$$

其中$x=\{x_1,x_2,\cdots,x_n\}^T, f, c_i(i\in E\cup I):\mathbf{R}^n\to\mathbf{R}^1$. 假设$f\in C^2, c_1,\cdots,c_p\in C^1$.用$\Omega$表示可行集,假设$\Omega\neq\emptyset$.

二次逼近法的基本思想

把问题(8-14)转化为求解一系列的二次规划子问题:假定在第k次迭代已知近似解x^k和近似乘子向量λ^k, 根据它们给出第k个二次规划子问题(P_k), 求解(P_k)得到新的近似解x^{k+1}, 并确定相应的乘子向量λ^{k+1},重复上述过程,直到获得问题(8-14)的最优解.

若令$d^k = x^{k+1} - x^k$,则求解问题(P_k)来获得x^{k+1}可转化为求d^k 的二次规划子问题.如何来得到相应的二次规划子问题呢? 可以设想对问题(8-14), 相应的二次规划问题的约束可以用将$c_i(x)$在x^k点线性化的方法得到,即应考虑如下的QP问题

$$\begin{array}{ll} \min & Q(d) = d^T B_k d/2 + \nabla f(x^k)^T d \\ \text{s.t.} & \nabla c_i(x^k)^T d + c_i(x^k) = 0, i \in E \\ & \nabla c_i(x^k)^T d + c_i(x^k) \leq 0, i \in I. \end{array} \tag{8-15}$$

但是这样做有时会遇到困难.例如,对下面的例8.4,将导致相应的可行域为空集, 但原问题的可行域非空.

例8.4

考察约束条件为

$$\begin{cases} c_1(x) = x - 1 \le 0 \\ c_2(x) = -x^2 \le 0 \end{cases}$$

的问题.

设$x^0 = 3$,则与(8-15)相应的约束条件为

$$\begin{cases} \nabla c_1(x^0)^T d + c_1(x^0) = d + 2 \le 0 \\ \nabla c_2(x^0)^T d + c_2(x^0) = -6d - 9 \le 0 \end{cases}$$

显然这两个约束是互相矛盾的,即可行域是空集,而原问题的可行域为$(-\infty, 1]$.

为克服上述缺点,可以这样做:把(8-15)的约束分为两类:

$$I_1 = \{i | c_i(x^k) \geq 0, i \in I\}, \quad I_2 = \{i | c_i(x^k) < 0, i \in I\}$$

然后用如下约束(8-16)来代替问题(8-15)中的约束,其中ξ是一个待定参数.

$$\left\{ \begin{array}{l} \nabla c_i(x^k)^T d + \xi c_i(x^k) = 0, i \in E \\ \nabla c_i(x^k)^T d + \xi c_i(x^k) \leq 0, i \in I_1 \\ \nabla c_i(x^k)^T d + c_i(x^k) \leq 0, i \in I_2 \end{array} \right. \tag{8-16}$$

选择ξ的原则是使(8-16)的可行域非空,且尽量接近问题(8-15)中的约束.

具体办法是:考察以$\xi, d_1, \cdots, d_n$为变量的LP问题(8-17).

$$
\begin{aligned}
\min \quad & z = -\xi \\
\text{s.t.} \quad & \nabla c_i(x^k)^T d + \xi c_i(x^k) = 0, i \in E \\
& \nabla c_i(x^k)^T d + \xi c_i(x^k) \le 0, i \in I_1 \\
& \nabla c_i(x^k)^T d + c_i(x^k) \le 0, i \in I_2 \\
& 0 \le \xi \le 1.
\end{aligned}
\tag{8-17}
$$

其中$d = (d_1, \cdots, d_n)^T$. 易见$\xi = 0, d = 0$是问题(8-17)的一个可行解,所以它总有最优解$(\xi^*, d^*)^T$.把(8-16)式中的ξ 取为ξ^*,就符合上述选择ξ的原则.当原来的二次规划问题(8-15) 有可行解$\bar{d}$ 时,易见$\xi = 1, d = \bar{d}$就是LP问题(8-17)的最优解, 即$\xi^* = 1, d^* = \bar{d}$.这时约束条件(8-16)与问题(8-15)的约束条件相同.

综合上述可见:确定搜索方向d^k的QP问题应取为

$$
\begin{array}{ll}
\min & \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d \\
\text{s.t.} & \nabla c_i(x^k)^T d + \xi^* c_i(x^k) = 0, i \in E \\
& \nabla c_i(x^k)^T d + \xi^* c_i(x^k) \le 0, i \in I_1 \\
& \nabla c_i(x^k)^T d + c_i(x^k) \le 0, i \in I_2
\end{array}
\tag{8-18}
$$

其中E, I_1, I_2由前面所述各式确定,ξ^*由求解问题(8-17)确定.

例8.5

对例8.4中的问题,确定QP问题(8-18)的约束条件和可行域.

解 因为$x^0=3, c_1(3)=2>0, c_2(3)=-9<0$,所以$I_1=\{1\}, I_2=\{2\}$. 注意到:

$$\nabla c_1(x)=1, \quad \nabla c_2(x)=-2x,$$

所以与(8-17)对应的LP问题为

$$\begin{array}{ll} \min & z=-\xi \\ \text{s.t.} & d+2\xi \le 0 \\ & -6d-9 \le 0 \\ & 0 \le \xi \le 1 \end{array}$$

易见其最优解为$(\xi^*, d^*)=(3/4, -3/2)^T$.因此对应于(8-18)的约束条件为

$$\begin{cases} \nabla c_1(x^0)^T d+\xi^* c_1(x^0)=d+3/2 \le 0 \\ \nabla c_2(x^0)^T d+c_2(x^0)=-6d-9 \le 0 \end{cases}$$

其可行解为$d=-3/2$.

计算步骤 I

1. 选取初始点x^1,初始正定对称矩阵B_1(例如取B_1为单位阵, 计算精度$\varepsilon > 0$,令$k = 1$.
2. 求解LP问题(8-17),设其最优解为$(\xi^*, d^{*T})^T$.
3. 求解正定QP问题(8-18)得最优解d^k及相应的乘子向量μ^k.若$d^k = 0$,则取$x^{k+1} = x^k + d^k$为问题(8-14)的近似最优解,计算结束;否则转4.
4. 从x^k出发沿方向d^k,对目标函数$\varphi(\alpha) = \omega(x^k + \alpha d^k, \lambda^k)$ 进行一维搜索,确定步长α_k,其中

$$\omega(x, \lambda) = f(x) + \sum_{i \in E} \lambda_i |c_i(x)| + \sum_{i \in I} \lambda_i \max\{0, c_i(x)\},$$

$$\lambda^k = (\lambda_1^k, \cdots, \lambda_m^k, \lambda_{m+1}^k, \cdots, \lambda_p^k)^T.$$

由μ^k和λ^{k-1}根据下式确定:

$$\begin{cases} \lambda_i^1 = |\mu_i^1|, i \in E \cup I \\ \lambda_i^k = \max\{|\mu_i^k|, \frac{1}{2}(\lambda_i^{k-1} + |\mu_i^k|)\}, i \in E \cup I, k > 1 \end{cases} \tag{8-20}$$

计算步骤 II

5. 令$x^{k+1} = x^k + \alpha_k d^k$.
6. 若$\|x^{k+1} - x^k\| < \varepsilon$,则取$x^{k+1}$为问题(8-14)的近似最优解, 计算结束;否则转7.
7. 按BFGS公式将B_k修正为B_{k+1},即令

$$B_{k+1} = B_k - \frac{B_k s^k (s^k)^T B_k}{(s^k)^T B_k s^k} + \frac{\eta^k (\eta^k)^T}{(\eta^k)^T s^k} \tag{8-21}$$

其中

$$\begin{cases} s^k = x^{k+1} - x^k \\ \eta^k = \theta y^k + (1-\theta) B_k s^k \\ y^k = \nabla_x L(x^{k+1}, \mu^k) - \nabla_x L(x^k, \mu^k) \\ L(x, \mu^k) = f(x) + \sum\limits_{i=1}^{p} \mu_i^k c_i(x) \\ \mu^k = (\mu_1^k, \cdots, \mu_p^k)^T, \end{cases}$$

计算步骤 III

$$\theta = \begin{cases} 1 & (y^k)^T s^k \geq 0.2(s^k)^T B_k s^k \\ \frac{0.8(s^k)^T B_k s^k}{(s^k)^T B_k s^k - (y^k)^T s^k}, & \text{否则} \end{cases}$$

8. 令$k = k + 1$,转2.