最优化方法/实用优化算法
第四章　约束最优化方法

教师　邱松强

中国矿业大学　数学学院

July 6, 2021

目录

约束优化问题

$$
\begin{aligned}
&\min\ f(x)\\
&\text{s.t. } c_i(x)=0,\ \ i\in\mathcal{E}=\{1,2,\cdots,l\},\\
&\qquad c_i(x)\geq 0,\ \ i\in\mathcal{I}=\{l+1,\cdots,m\}.
\end{aligned}
$$

可行域

$$\Omega=\{x\mid c_i(x)=0, i\in\mathcal{E}; c_i(x)\geq 0,\ \ i\in\mathcal{I}\}.$$

Lagrange 函数

$$L(x,\lambda)=f(x)-\sum_{i=1}^{l}\lambda_i c_i(x)-\sum_{i=l+1}^{m}\lambda_i c_i(x),$$

其中, $\lambda_i\geq 0,\ i\in\mathcal{I}$.

Lagrange 函数关于x 的梯度

$$\nabla_x L(x,\lambda) = \nabla f(x) - \sum_{i=1}^{l} \lambda_i \nabla c_i(x) - \sum_{i=l+1}^{m} \lambda_i \nabla c_i(x),$$

其中, $\lambda_i \geq 0,\ i \in \mathcal{I}.$

Lagrange 函数关于x Hesse 矩阵

$$\nabla^2_{xx} L(x,\lambda) = \nabla^2_{xx} f(x) - \sum_{i=1}^{l} \lambda_i \nabla^2_{xx} c_i(x) - \sum_{i=l+1}^{m} \lambda_i \nabla^2_{xx} c_i(x),$$

其中, $\lambda_i \geq 0,\ i \in \mathcal{I}.$

等式约束优化问题的最优性条件

等式约束优化问题

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

Lagrange 函数

$$\begin{aligned} L(x, \lambda) &= f(x) - \lambda^T c(x) \\ &= f(x) - \sum_{i=1}^{l} \lambda_i c_i(x). \end{aligned}$$

其中，$\lambda_i,\ i = 1, 2, \cdots, m$ 称为Lagrange 乘子.

等式约束优化问题的最优性条件

最小化Lagrange 函数

$$\min \;\; L(x, \lambda).$$

相应的一阶必要性条件

$$\nabla_{x,\lambda} L(x, \lambda) = 0.$$

也就是

$$\begin{aligned} \nabla_x L(x, \lambda) &= \nabla f(x) - \nabla c(x)\lambda \\ &= \nabla f(x) - \sum_{i=1}^{l} \lambda_i \nabla c_i(x) = 0 \\ \nabla_\lambda L(x, \lambda) &= -c(x) = 0. \end{aligned}$$

等式约束优化问题的最优性条件

定理1 (一阶必要条件)

若

(i) x^* 是原问题(1) 的局部最优解;

(ii) $f(x)$ 与$c_i(x)$, $i = 1, \cdots, l$ 在x^* 的某邻域内连续可微;

(iii) $\nabla c_i(x^*)$, $i = 1, \cdots, l$ 线性无关（线性无关约束规范, *LICQ*）.

则存在一组实数$\lambda_1^*, \cdots, \lambda_l^*$ 使得

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

等式约束优化问题的最优性条件

例2.1

考虑问题

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

的最优性条件.

【解：】该问题的Lagrange 函数为

$$L(x,\lambda) = f(x) - \lambda c(x).$$

等式约束优化问题的最优性条件

由一阶必要性条件,

$$\nabla_x L(x,\lambda) = \nabla f(x) - \lambda \nabla c(x) = 0$$
$$\Rightarrow \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} - \lambda \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} = 0,$$
$$c(x) = 0 \Rightarrow x_1^2 + 2x_2^2 = 1$$

解之得

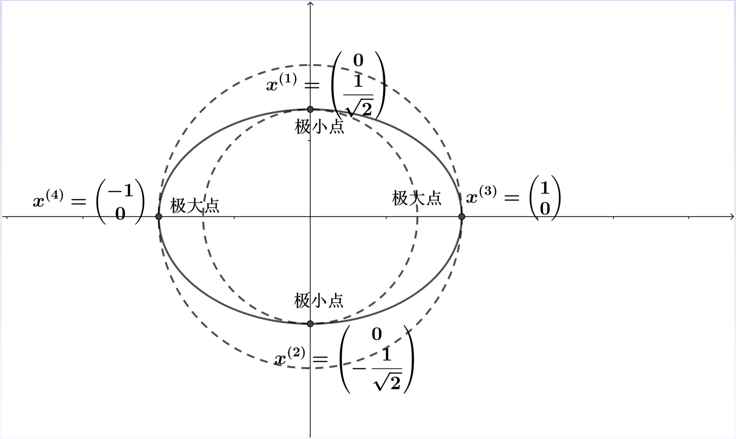
$$x_1 = 0, \lambda = \frac{1}{2}, x_2 = \pm\frac{1}{\sqrt{2}},$$

或

$$x_1 = \pm 1, \lambda = 1, x_2 = 0.$$

等式约束优化问题的最优性条件

如图所示



因此, $x^{(1)}$ 和$x^{(2)}$ 为极小点; $x^{(3)}$ 和$x^{(4)}$ 为极大点.

例题

例2.2 (LICQ 不成立的例子)

证明在等式约束问题

$$\begin{aligned} \min\ & f(x) = x_1 + \frac{1}{2}x_2^2 \\ s.t.\ & c_1(x) = -x_1^2 - x_2^2 + x_3^2 = 0, \\ & c_2(x) = -x_1x_3 = 0. \end{aligned} \tag{2}$$

的解处不存在满足定理1 的*Lagrange* 乘子.

【证明：】首先，容易看出该问题有唯一解$x^* = (0,0,0)^T$.

$\boxed{\text{为什么？}}$

例题

在点$x^* = (0,0,0)^T$ 处,

$$\nabla f(x^*) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \nabla c_1(x^*) = \nabla c_2(x^*) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

所以对任意的$\lambda_1, \lambda_2 \in R$,

$$\nabla f(x^*) - \lambda_1 \nabla c_1(x^*) - \lambda_2 \nabla c_2(x^*) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

所以在x^* 处不存在满足定理1 的Lagrange 乘子.

注解

前面两个例题说明，我们不能去掉定理1 中的条件(3).
实际上，条件(3) 可以换成其它条件.
条件(3) 这样的条件保证满足定理1 的Lagrange 乘子存在. 这种条件称为约束规格.

Fritz John 条件

定理2 (等式约束问题的Fritz John 条件)

若

(i) x^* 是原问题(1) 的局部最优解;

(ii) $f(x)$ 与$c_i(x)$, $i = 1, \cdots, l$ 在x^* 的某邻域内连续可微;

则存在不全为零的一组实数$\lambda_0^*, \lambda_1^*, \cdots, \lambda_l^*$ 使得

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^{l} \lambda_i^* \nabla c_i(x^*) = 0,$$
$$c_i(x^*) = 0, i = 1, \cdots, l.$$

显然定理1 是这个定理的特例.

例2.3 (例2.2 续)

若取$\lambda_0 = 0$, 并取任意不全为*0* 的λ_1, λ_2, 则有

$$
\begin{aligned}
&\lambda_0 \nabla f(x^*) - \lambda_1 \nabla c_1(x^*) - \lambda_2 \nabla c_2(x^*) \\
&= \lambda_0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}
$$

等式约束优化问题的最优性条件

定理3 (二阶充分条件)

在等式约束问题(1) 中，若

(i) $f(x)$ 与$c_i(x)$, $1 \le i \le l$ 是二阶连续可微函数;

(ii) 存在$x^* \in R^n$ 与$\lambda^* \in R^l$ 是*Lagrange* 函数的梯度为零，即

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

(iii) 对于任意非零向量$s \in R^n$ 且

$$s^T \nabla_{xx}^2 L(x^*, \lambda^*) s > 0,$$

即方阵$\nabla_{xx}^2 L(x^*, \lambda^*)$ 正定.

则x^* 是问题(1) 的严格局部极小点.

二阶充分条件在分析约束优化问题的收敛速度时，起着至关重要的作用.

基本概念

不等式约束问题:

$$\begin{aligned} &\min_{x\in R^n}\ f(x) \\ &\text{s.t. } c_i(x)\geq 0,\ i\in\{1,\cdots,m\}. \end{aligned} \tag{3}$$

可行域

$$\Omega=\{x\in R^n\mid c_i(x)\geq 0,\ i=1,\cdots,m\}.$$

基本概念：回顾下降方向与可行方向

定义1 (下降方向)

对于向量$0 \neq d \in R^n$, 和某个点x 处. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, 有

$$f(x + \alpha d) < f(x),$$

则称d 为x 处的一个下降方向.

定义2 (可行方向)

对于向量$0 \neq d \in R^n$, 和可行域Ω 中的某个点$x \in \Omega$. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, $x + \alpha d$ 仍然属于Ω, 即$x + \alpha d \in \Omega$, 则称d 为x 处的一个可行方向.

基本概念

例2.4

考虑优化问题

$$\begin{aligned} \min\ & f(x) = -(x_1 + x_2) \\ s.t.\ & c_1(x) = 1 - x_1^2 - 4x_2^2 \geq 0, \\ & x_1 \geq 0, \\ & x_2 \geq 0. \end{aligned}$$

初始点为$x_0 = \left[\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]$. 试分别判断方向向量$d_1 = [1,0]^T$, $d_2 = [1,-0.5]^T$, $d_3 = [0,-1]^T$ 是否是初始点处的下降方向？是否是可行方向？是否是可行下降方向？

例题

【解:】有$\nabla f(x_0)=[-1,-1]^T$, $\nabla c_1(x_0)=-\frac{1}{\sqrt{5}}[2,8]^T$. 算得

$\nabla f(x_0)^T d_1=-1<0$, $\nabla f(x_0)^T d_2=-0.5<0, \nabla f(x_0)^T d_3=1.$

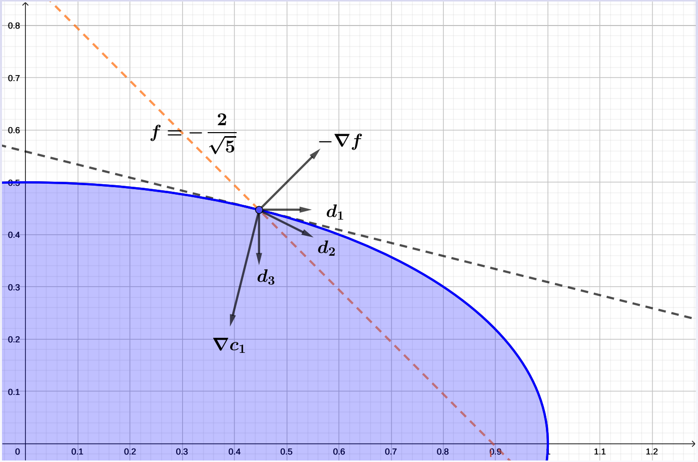
故而d_1, d_2 是下降方向。

$$\nabla c_1(x_0)^T d_1=-\frac{2}{\sqrt{5}}<0,\ \nabla c_1(x_0)^T d_2=\frac{2}{\sqrt{5}}>0,$$

$$\nabla c_1(x_0)^T d_3=\frac{8}{\sqrt{5}}.$$

从而d_2, d_3 是可行方向.(事实上，容易验证，对任意$\alpha>0$, $c_1(x_0+\alpha d_1)<0.$)
只有d_2 是可行下降方向.

如图


$f = -\frac{2}{\sqrt{5}}$
$-\nabla f$
d_1
d_2
d_3
∇c_1
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.1
1.2

基本概念

定义3 (有效约束和有效集)

设$\tilde{x}$ 是问题(3) 是一个可行点. 若第j 个约束条件满足

$$c_j(\tilde{x}) = 0,$$

则称$c_j(x) \geq 0$ 为关于$\tilde{x}$ 的有效约束.
称所有在$\tilde{x}$ 处的有效约束指标组成的集合

$$\tilde{\mathcal{I}} = \mathcal{I}(\tilde{x}) = \{i \mid c_i(\tilde{x}) = 0\}$$

为$\tilde{x}$ 处的有效约束指标集, 简称为$\tilde{x}$ 处的有效集.

基本概念

例如，下图中

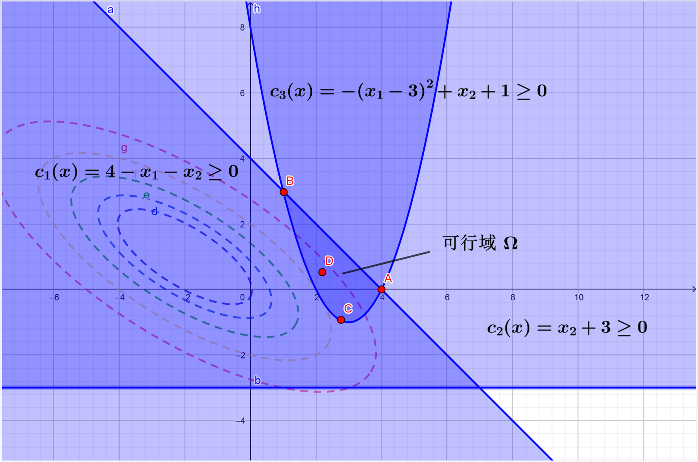


图: 有效约束和有效集

例题

例2.5

问题

$$\begin{aligned} &\min\ x_1^2+x_2^2, \\ &s.t.\ (x_1-1)^2+x_2^2\le 1, \\ &\qquad x_2^2-x_1+1\le 0 \end{aligned} \tag{4}$$

的最优解是$x^*=(1,0)^T$. 在点x^* 处, $\mathcal{I}^*=\{2\}$. 如图

例题

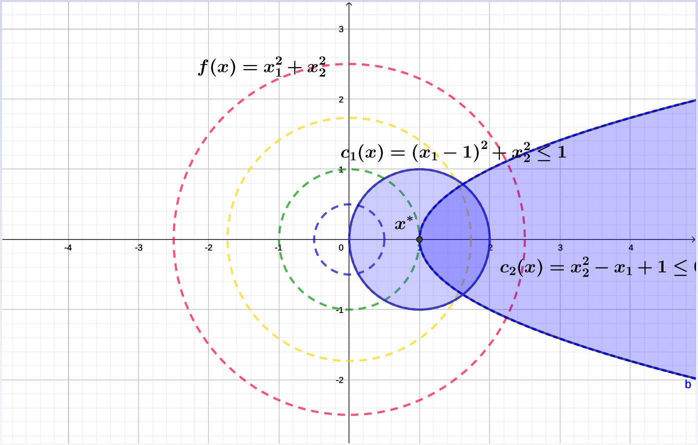
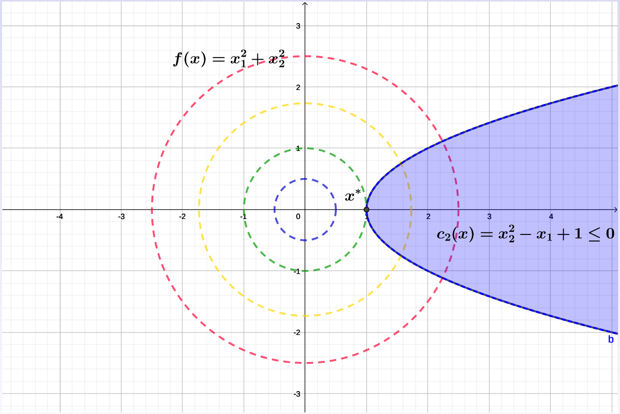

$f(x) = x_1^2 + x_2^2$
$c_1(x) = (x_1 - 1)^2 + x_2^2 \leq 1$
x^*
$c_2(x) = x_2^2 - x_1 + 1 \leq ($
b
3
2
1
0
-1
-2
-4
-3
-2
-1
1
2
3
4

图: 问题及其有效集

例题

去掉约束中的第一个约束（非积极约束）$(x_1-1)^2+x_2^2\leq 1$，问题的解不变，如下图



图：去掉一个非积极约束后的问题

定义4

设集合$C \subset R^n$ 且$C \neq \emptyset$. 若$\boldsymbol{x} \in C$, 对任意$d \in R^n$, 当$\boldsymbol{x} + d \in C$ 时必有$\boldsymbol{x} + td \in C (t \geq 0)$, 则称$C$ 为以$\boldsymbol{x}$ 为顶点的锥. 当C 为凸集时, 称为凸锥.

较为常见的是以原点0 为顶点的锥,即

定义5

设集合$C \subset R^n$ 且$C \neq \emptyset$. 若对任意$\boldsymbol{x} \in C$ 和$\lambda > 0$, 均有$\lambda \boldsymbol{x} \in C$, 则称$C$ 为锥(无须说明顶点). 若对任意的$\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ 和$\lambda_1, \lambda_2 \geq 0$, 均有

$$\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 \in C,$$

则称C 为凸锥.

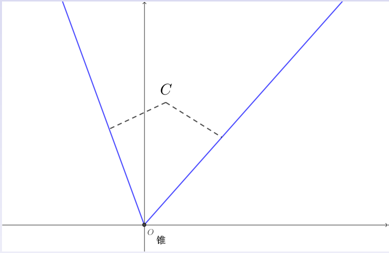


图: 锥

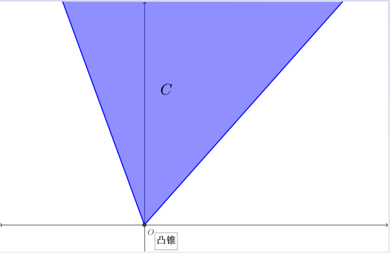


图: 凸锥

引理1 (Farkas 引理)

设a_1, a_2, $\cdots$, a_r 和b 为n 维向量，则所有满足$a_i^T d \geq 0$, $i = 1, 2, \cdots, r$ 的向量$d \in R^n$ 同时也满足不等式$b^T d \geq 0$ 的充要条件是，存在非负实数λ_1, λ_2,$\cdots$, λ_r, 使得$b = \sum_{i=1}^{r} \lambda_i a_i$.

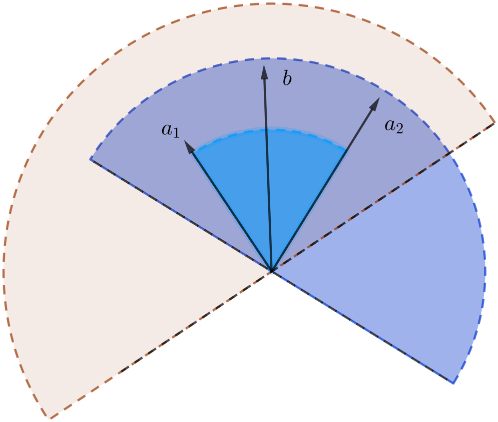


图: Farkas 引理

引理2 (Gordan 引理)

设$a_1, a_2, \cdots, a_r$ 是n 维向量，则不存在向量$d \in R^n$ 使得

$$a_i^T d < 0,\ i = 1, 2, \cdots, r$$

成立的充要条件是，存在不全为零的非负实数组λ_i, $i = 1, 2, \cdots, r$ 是

$$\sum_{i=1}^{r} \lambda_i \alpha_i = 0.$$

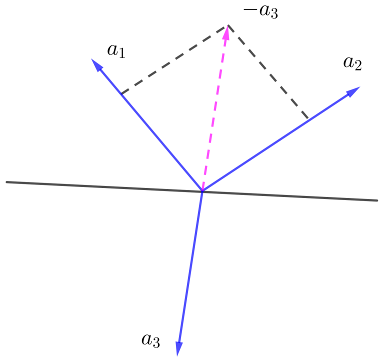


图: Gordan 引理

引理3

在问题(3) 中，假设

(i) $\boldsymbol{x}^*$ 为问题(3)的局部最优解且$\mathcal{I}^* = \{i | c_i(\boldsymbol{x}^*) = 0, 1 \le i \le m\}$;

(ii) $f(\boldsymbol{x})$ 和$c_i(\boldsymbol{x})$ $i \in \mathcal{I}^*$ 在点$\boldsymbol{x}^*$ 可微;

(iii) $c_i(\boldsymbol{x})$ $(i \in \mathcal{I} \backslash \mathcal{I}^*)$ 在点$\boldsymbol{x}^*$ 连续.

则

$$S = \{p \in R^n | \nabla f(\boldsymbol{x}^*)^T p < 0\}$$

与

$$G = \{p \in R^n | \nabla c_i(\boldsymbol{x}^*)^T p > 0, i \in \mathcal{I}^*\}$$

的交是空集，即$G \cap S = \emptyset$.

定理4

设$\boldsymbol{x}^*$ 是可行点, $\mathcal{I}(\boldsymbol{x}^*)$ 是其有效约束指标集. $f(\boldsymbol{x})$ 和$c_i(\boldsymbol{x})$ $(i \in \mathcal{I}(\boldsymbol{x}^*))$ 在点$\boldsymbol{x}^*$ 处可微, $c_i(\boldsymbol{x})$ $(i \notin \mathcal{I}(\boldsymbol{x}^*))$ 处连续. 如果$\boldsymbol{x}^*$ 是约束极值问题(3) 的局部极小点, 则在$\boldsymbol{x}^*$ 处没有可行下降方向.

KKT条件

定理5 (不等式约束优化问题的KKT条件)

在不等式约束问题(3) 中，若

(i) $\boldsymbol{x}^*$ 为局部最优解，有效集$\mathcal{I}^* = \{i \mid c_i(\boldsymbol{x}^*) = 0, i = 1, 2, \cdots, m\}$;

(ii) $f(\boldsymbol{x})$ 及$c_i(\boldsymbol{x}), i = 1, 2, \cdots, m$ 在点$\boldsymbol{x}^*$ 出可微;

(iii) 向量集$\{\nabla c_i(\boldsymbol{x}^*) | i \in \mathcal{I}^*\}$ 线性无关（线性无关约束规范，*LICQ*），则存在向量

$$\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^T$$

使得

$$\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = 0;$$
$$c_i(\boldsymbol{x}^*) \geq 0,\ \lambda_i^* \geq 0,\ \lambda_i^* c_i(\boldsymbol{x}^*) = 0,\ i = 1, 2, \cdots, m.$$

KKT条件

注：

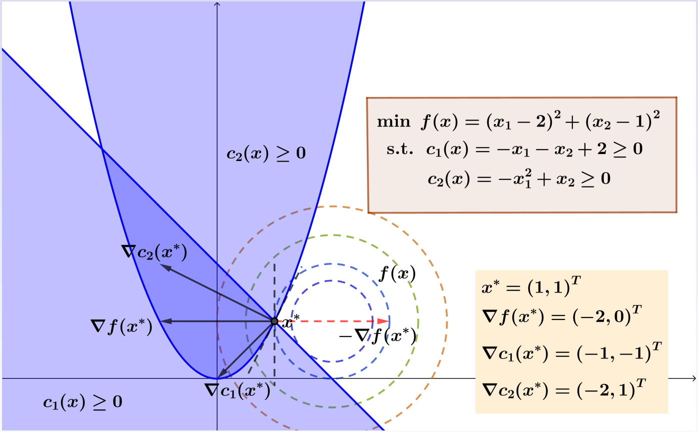
- 定理中LICQ 是一种约束规范. 约束规范有多种, LICQ 可以替换为其它形式的约束规范. (感兴趣的同学可以参考M. Fukushima 的著作《非线性最优化基础》)
- 若没有约束规范条件, 则约束优化问题的局部最优解不一定满足KKT 条件.
- 最后一个条件称为互补条件. 从某种意义上讲, 不等式约束优化问题之所以难以求解, 是因为这个条件.
- 如果对任意$i \in \mathcal{I}$, λ_i^* 和$c_i(\boldsymbol{x}^*)$ 有且仅有一个等于0, 即

$$\lambda_i^* > 0, \forall i \in \mathcal{I}^*.$$

则称问题在$\boldsymbol{x}^*$ 处**严格互补条件** 成立.

基本概念

例如，下图中显然，$\nabla f(x^*)$ 可以表示为$\nabla c_1(x^*)$, $\nabla c_2(x^*)$ 的线性组合，且系数均为正数.



例题

例2.6

求问题

$$
\begin{aligned}
\min\ & f(\boldsymbol{x}) = (x_1-1)^2 + (x_2-1)^2, \\
s.t.\ & c_1(\boldsymbol{x}) = 1 - x_1 - x_2 \geq 0, \\
& c_2(\boldsymbol{x}) = x_1 \geq 0, \\
& c_3(\boldsymbol{x}) = x_2 \geq 0.
\end{aligned}
$$

的KKT 点.

例题

【解：】

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2x_1 - 2 \\ 2x_2 - 2 \end{pmatrix}, \nabla c_1(\boldsymbol{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix},$$

$$\nabla c_2(\boldsymbol{x}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \nabla c_3(\boldsymbol{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

设$\boldsymbol{x}$ 为问题的极小点，则存在$\lambda_1 \geq 0$, $\lambda_2 \geq 0$ 和$\lambda_3 \geq 0$ 使得

例题

$$\begin{pmatrix} 2x_1 - 2 \\ 2x_2 - 2 \end{pmatrix} - \lambda_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} - \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \lambda_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0, \tag{5a}$$

$$1 - x_1 - x_2 \geq 0, \lambda_1 \geq 0, \lambda_1(1 - x_1 - x_2) = 0, \tag{5b}$$

$$x_1 \geq 0, \lambda_2 \geq 0, \lambda_2 x_1 = 0, \tag{5c}$$

$$x_2 \geq 0, \lambda_3 \geq 0, \lambda_3 x_2 = 0, \tag{5d}$$

讨论: 若$\lambda_1 = 0$, 则由(5a) 有

$$\begin{cases} 2x_1 - 2 - \lambda_2 = 0, \\ 2x_2 - 2 - \lambda_3 = 0. \end{cases} \tag{6}$$

例题

两式相加得

$$2(x_1 + x_2 - 2) - (\lambda_2 + \lambda_3) = 0,$$

则

$$\lambda_2 + \lambda_3 = 2(x_1 + x_2 - 2) < 0.$$

这与$\lambda_2, \lambda_3 \geq 0$ 矛盾. 从而$\lambda_1 > 0$.

例题

由(5b) 知

$$1 - x_1 - x_2 = 0. \tag{7}$$

若$x_1 = 0$, 则$x_2 = 1$, 于是相应的$\lambda_3 = 0$. 由(5a) 的第二式知$\lambda_3 = \lambda_1 > 0$, 矛盾. 故$x_1 > 0$. 同理, $x_2 > 0$.
由互补条件(5c) 和(5d) 知, $\lambda_2 = \lambda_3 = 0$. 再由(5a) 知,

$$\begin{cases} 2x_1 - 2 + \lambda_1 = 0, \\ 2x_2 - 2 + \lambda_1 = 0. \end{cases}$$

故$x_1 = x_2$. 从而由(7) 知, $x_1 = x_2 = \frac{1}{2}$.
再回代入上式, 得$\lambda_1 = 1$.

例题

由上述讨论知, 问题的KKT 对为

$$x = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

例题

【注：】将本例中的问题改写为

例2.7

$$\begin{aligned}
\min\ f(\boldsymbol{x}) &= (x_1-1)^2+(x_2-1)^2,\\
s.t.\ c_1(\boldsymbol{x}) &= (1-x_1-x_2)^3 \geq 0,\\
c_2(\boldsymbol{x}) &= x_1 \geq 0,\\
c_3(\boldsymbol{x}) &= x_2 \geq 0.
\end{aligned}$$

显然，该问题与原问题是等价的．所以其极小点也是$x^*=(1/2,1/2)^T$．但是x^*不是KKT 点．事实上，这个问题没有KKT 点．

例题

导致这个现象的原因是，在极小点$(1/2,1/2)^T$ 处，有效约束$c_1(x)$ 的梯度

$$\nabla c_1(\boldsymbol{x}^*) = 3\left(1-\frac{1}{2}-\frac{1}{2}\right)^2\begin{pmatrix}-1\\-1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

它不是线性无关的，即不满足LICQ.

例题

例2.8

用*matlab* 分别求解上述两个问题. (如：可以调用*matlab*的*fmincon* 函数).

数值算法求解这两个问题时的表现是不同的.

不等式约束问题的Fritz John 条件

定理6

在不等式约束问题(3) 中，若

(i) $\boldsymbol{x}^*$ 为局部最优解，有效集$\mathcal{I}^* = \{i \mid c_i(\boldsymbol{x}^*) = 0, i = 1, 2, \cdots, m\}$;

(ii) $f(\boldsymbol{x})$ 及$c_i(\boldsymbol{x})$, $i = 1, 2, \cdots, m$ 在点x^* 处可微; 则存在非零向量

$$\lambda^* = (\lambda_0^*, \lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^T$$

使得

$$\lambda_0^* \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0;$$
$$c_i(\boldsymbol{x}^*) \geq 0,\ \lambda_i^* c_i(\boldsymbol{x}^*) = 0,\ i = 1, 2, \cdots, m.$$
$$\lambda_i^* \geq 0,\ i = 0, 1, \cdots, m.$$

证明:设$\boldsymbol{x}^*$ 处的有效集为$\mathcal{I}^* = \mathcal{I}(\boldsymbol{x}^*)$. 因$\boldsymbol{x}^*$ 为(3) 的局部最优解, 由引理3 知, 不存在$d \in R^n$ 使得

$$\begin{aligned}
&\nabla f(\boldsymbol{x}^*)^T d < 0,\\
&d^T \nabla c_i(\boldsymbol{x}^*) > 0,\ \ i \in \mathcal{I}^*.
\end{aligned}$$

把$\nabla f(\boldsymbol{x}^*)$ 和$-\nabla c_i(x^*), i \in I^*$ 依次视为引理2 中的$a_1, a_2, \cdots$. 则存在不全为零的数

$$\lambda_0^* \geq 0,\ \lambda_i^* \geq 0,\ \ i \in I^*,$$

使得

$$\lambda_0^* \nabla f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{I}^*} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = 0.$$

如果对于任意$i \in \mathcal{I} \backslash \mathcal{I}^*$, 规定$\lambda_i^* = 0$, 则上面二式可写成

$$\lambda_0^* \nabla f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{I}} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = 0.$$

上述规定等价于

$$\lambda_i^* c_i(\boldsymbol{x}^*) = 0, \ i = 1, \cdots, m.$$

显然, 例2.7 中, 取

$$\lambda^* = (0, 1, 0, 0)^T,$$

则$(\boldsymbol{x}^*, \lambda^*)$ 满足Fritz John 条件.

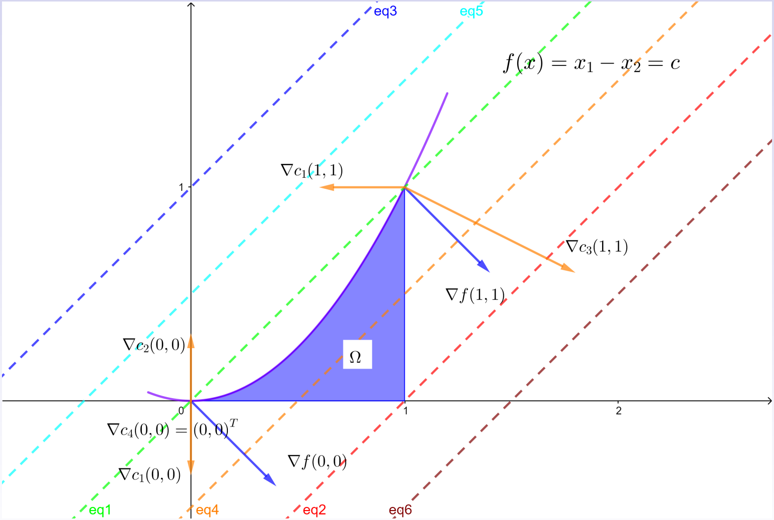
例题

例2.9

考虑问题

$$
\begin{aligned}
\min\ & f(\boldsymbol{x}) = x_1 - x_2 \\
s.t.\ & c_1(\boldsymbol{x}) = x_1^2 - x_2 \geq 0, \\
& c_2(\boldsymbol{x}) = x_2 \geq 0, \\
& c_3(\boldsymbol{x}) = 1 - x_1 \geq 0, \\
& c_4(\boldsymbol{x}) = x_1^3 \geq 0.
\end{aligned}
$$

例题


eq3
eq5
$f(x) = x_1 - x_2 = c$
$\nabla c_1(1,1)$
1
$\nabla c_3(1,1)$
$\nabla f(1,1)$
$\nabla c_2(0,0)$
Ω
0
1
2
$\nabla c_4(0,0) = (0,0)^T$
$\nabla f(0,0)$
$\nabla c_1(0,0)$
eq1
eq4
eq2
eq6

例题

本题有两个最优解：$(1,1)^T$ 和$(0,0)^T$.

- 对$\boldsymbol{x}^* = (1,1)^T$, 有效集为$\mathcal{I} = \{1,3\}$ 且

$$\nabla c_1(\boldsymbol{x}^*) = \begin{pmatrix} 2 \\ -1 \end{pmatrix},\ \nabla c_2(\boldsymbol{x}^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

是线性无关的（即满足LICQ）. 通过计算, 可取

$$\lambda^* = (1,0,1,0),$$

则$(\boldsymbol{x}^*, \lambda^*)$ 满足KKT 条件.

例题

- 对$\boldsymbol{x}^* = (0,0)^T$, 有效集为$\mathcal{I} = \{1,2,4\}$ 且

$$\nabla c_1(\boldsymbol{x}^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \ \nabla c_2(\boldsymbol{x}^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \nabla c_3(\boldsymbol{x}^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

是线性相关的. 并且显然不存在$\lambda^* \geq 0$ 满足KKT 条件.

- 但是若令

$$(\lambda_0^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*)^T = (0,1,1,0,0)^T,$$

则Fritz John 条件成立.

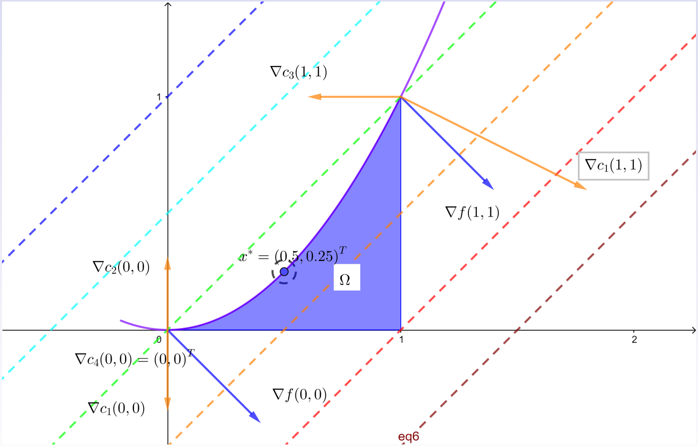
例题

注：在本题的约束条件下，$f(\boldsymbol{x})$ 还有一个稳定点$\left(\frac{1}{2}, \frac{1}{4}\right)^T$ 和一个最大值点$(1,0)$.

- 对$\boldsymbol{x}^* = \left(\frac{1}{2}, \frac{1}{4}\right)^T$, 存在$\lambda^* = (1,0,0,0)^T$ 满足KKT 条件. 但$\boldsymbol{x}^*$ 不是极小点, 也不是极大点.(见下一张幻灯片的图).
- 对$\boldsymbol{x}^* = (1,0)^T$, 不存在$\lambda^* \geq 0$ 满足KKT 条件, 在该点处Fritz John 条件也不成立.

例题

下图中，在$\boldsymbol{x}^* = (0.5, 0.25)$ 的任一邻域内，既有函数值大于$f(\boldsymbol{x}^*)$ 的点，也有小于$f(\boldsymbol{x}^*)$ 的点.



1. 约束优化问题的最优性条件
   - 等式约束优化问题的最优性条件
   - 不等式约束问题的最优性条件
     - 锥和Farkas 引理
   - 一般约束优化问题的最优性条件
2. 罚函数法与乘子法
3. 增广Lagrange 函数法(乘子法)
4. 投影梯度法和简约梯度法
5. 序列二次规划算法（SQP）

一般约束优化问题的最优性条件

$$
\begin{aligned}
&\min\ f(x) \\
&\text{s.t. } c_i(x) = 0,\ i \in \mathcal{E} = \{1, \cdots, l\} \\
&\qquad c_i(x) \geq 0,\ i \in \mathcal{I} = \{l+1, \cdots, m\}.
\end{aligned} \tag{8}
$$

KKT条件

定理7 (KKT 条件)

在问题(8) 中，若

(1) x^* 为局部最优解，其有效集为

$$\mathcal{I}^* = \{i \mid c_i(x^*) = 0,\ i \in \mathcal{I}\};$$

(2) $f(x), c_i(x), i = 1, \cdots, m$ 在点x^* 可微;

(3) $\{\nabla c_i(x^*) \mid i \in \mathcal{E} \cup \mathcal{I}^*\}$ 线性无关，

则存在向量$\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)$ 使得

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

$$c_i(x^*) = 0,\ i \in \mathcal{E},$$

$$\lambda_i^* \geq 0, c_i(x^*) \geq 0, \lambda_i^* c_i(x^*) = 0, i \in \mathcal{I}.$$

例题

例2.10

已知约束问题

$$
\begin{aligned}
\min\ & f(x) = -3x_1^2 - x_2^2 - 2x_3^2, \\
s.t.\ & c_1(x) = x_1^2 + x_2^2 + x_3^2 - 3 = 0, \\
& c_2(x) = -x_1 + x_2 \geq 0, \\
& c_3(x) = x_1 \geq 0, \\
& c_4(x) = x_2 \geq 0, \\
& c_5(x) = x_3 \geq 0.
\end{aligned}
$$

试验证最优解$x^* = (1,1,1)^T$ 为KKT 点.

例题

【解：】$\mathcal{I}^* \cup \mathcal{E} = \{1, 2\}$.

$$\nabla f(x^*) = (-6, -2-4)^T,$$
$$\nabla c_1(x^*) = (2, 2, 2)^T,$$
$$\nabla c_2(x^*) = (-1, 1, 0)^T.$$

设λ_i, $i = 1, 2, \cdots, 5$ 为Lagrange 乘子，其中，$\lambda_i \geq 0$, $i = 2, \cdots, 5$. 由互补条件，有$\lambda_3 = \lambda_4 = \lambda_5 = 0$. 于是有

$$\begin{pmatrix} -6 \\ -2 \\ -4 \end{pmatrix} - \lambda_1 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} - \lambda_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

例题

得$\lambda_1 = -2,\ \lambda_2 = 2$，即

$$\nabla f(x^*) - (-2)\nabla c_1(x^*) - 2\nabla c_2(x^*) = 0,$$

$$\lambda_2 c_2(x^*) = 0, \lambda_2 > 0.$$

故$x^* = (1,1,1)^T$ 为KKT 点.

二阶充分条件

定理8 (二阶充分条件)

设$f(x)$ 和$c_i(x)$, $(i \in \mathcal{E} \cup \mathcal{I})$ 是二阶连续可微函数, (x^*, λ^*) 满足(8) 的KKT 条件, 若对锥

$$\mathcal{F} = \left\{ d \,\middle|\, \begin{array}{c} \nabla c_i(x^*)^T d = 0, i \in \mathcal{I}^* \text{且} \lambda_i^* > 0, \\ \nabla c_i(x^*)^T d \geq 0, i \in \mathcal{I}^* \text{且} \lambda_i^* = 0, \\ \nabla c_i(x^*)^T d = 0, i \in \mathcal{E}^* \end{array} \right\}$$

中的非零向量d, 均有

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0,$$

则x^* 为问题(8) 的严格局部极小点.

凸规划

问题

$$\begin{aligned} &\min\ f(x), \\ &s.t.\ c_i(x) = a_i^T x + b_i = 0, i \in \mathcal{E} = \{1, \cdots, l\}, \\ &\qquad c_i(x) \geq 0, i \in \mathcal{I} = \{l+1, \cdots, m\}, \end{aligned} \tag{9}$$

其中$f(x)$ 为凸函数, $c_i(x), i \in \mathcal{I}$ 为凹函数. 则该问题为凸规划问题.

凸规划问题的KKT条件

定理9 (凸规划问题的KKT条件)

设凸规划问题(9) 中的$f(x)$ 和$c_i(x), i \in \mathcal{I}$ 为可微函数. 则可行点x^*为问题(9) 的整体最优点当且仅当x^* 满足KKT 条件.

将复杂的问题转换为较为简单的问题去解决.

将约束优化问题转换为一系列无约束最优化问题来求解.

等式约束优化问题的外罚函数法

等式约束优化问题

$$\begin{aligned} &\min\ f(x), \\ &\text{s.t. } c_i(x) = 0,\ i \in \mathcal{E}. \end{aligned} \tag{10}$$

定义罚函数

$$P(x, \sigma) = f(x) + \sigma \tilde{P}(x),$$

其中, $\tilde{P}(x) = \sum_{i \in \mathcal{E}} c_i^2(x)$, 等式右端第二项称为惩罚项，$\sigma > 0$ 称为罚因子.

上述罚函数的特点

- 对非可行点而言，当σ 变大时，惩罚项在罚函数中的比重加大，在对罚函数求极小时，相当于迫使其极小点向可行域靠拢.
- 在可行域中，$P(x,\sigma)$ 的极小点与约束最优化问题(10) 的最优解相同.

外罚函数法SUMT

【步 1】 给定$\sigma_1 > 0$, $c > 1$, $\epsilon_1 > 0$, $\epsilon > 0$, x_0, $k := 1$;

【步 2】 以x_{k-1} 为初始点，近似求解

$$\min P(x, \sigma_k)$$

得$x(\sigma_k)$. 当$\|\nabla P(x(\sigma_k), \sigma_k)\| \leq \epsilon_1$ 时停止.

【步 3】 当$\|c(x(\sigma_k))\| \leq \epsilon$ 时，迭代停止;

【步 4】 令$x_k = x(\sigma_k)$, 选$\sigma_{k+1} = c\sigma_k$. 令$k := k + 1$, 转步2.

关于外罚函数法的几点说明

- 算法的第2 步为内层子迭代，我们可以用第三章中介绍的某一无约束优化算法求解

$$\min \ P(x, \sigma_k). \tag{11}$$

- 在第2 步中，我们假定$P(x, \sigma_k)$ 的局部极小点$x(\sigma_k)$ 是存在的．若非如此，增大σ_k 后重新求解．
- 如何选取递增序列$\{\sigma_k\}$ 将直接影响到算法的有效性．如果让该序列增长很快，会影响无约束子问题的求解．而如果我们让该序列增长缓慢，无约束最优化问题固然可以更好地求解，然而迭代的速度必然会收到影响．一般地，我们可选该序列为$\{10^k\}$.

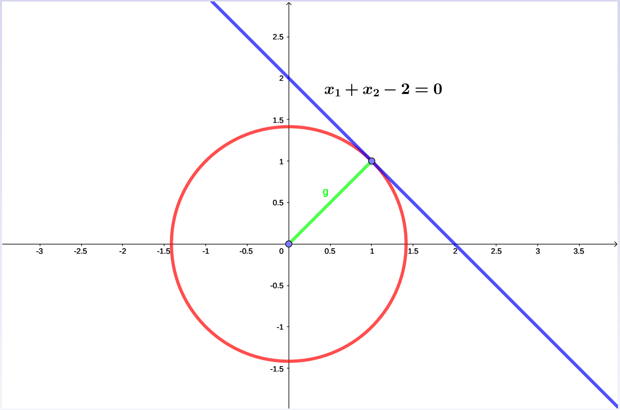
例题

例3.1

求解约束优化问题

$$\min\ f(x) = x_1^2 + x_2^2,$$
$$s.t.\ x_1 + x_2 - 2 = 0.$$

例题



本问题就是求原点到直线$x_1 + x_2 - 2 = 0$上的点的最短距离.显然最优解为$(1,1)^T$.

例题

罚函数为

$$P(x,\sigma) = x_1^2 + x_2^2 + \sigma(x_1 + x_2 - 2)^2.$$

令$\nabla P(x,\sigma) = 0$, 得

$$\begin{cases} 2x_1 + 2\sigma(x_1 + x_2 - 2) = 0, \\ 2x_2 + 2\sigma(x_1 + x_2 - 2) = 0, \end{cases} \tag{12}$$

解之得

$$x_1(\sigma) = x_2(\sigma) = \frac{2\sigma}{1+2\sigma}.$$

令$\sigma \to \infty$, 得$x^* = (1,1)^T$.

注：

因为$x(\sigma)$ 为外罚函数的一个最小值点，那么有

$$\nabla f(x(\sigma)) + 2\sigma \sum_{i\in\mathcal{E}} c_i(x(\sigma))\nabla c_i(x(\sigma)) = 0.$$

则若

$$\lim_{\sigma\to+\infty} -2\sigma c_i(x(\sigma)) = \lambda_i^*, \tag{13}$$

那么λ_i^* 就是相对于$c_i(x)$ 的Lagrange 乘子.
所以, (13) 是外罚函数法中计算乘子的公式.

注：

验证一下. 这一题中$\nabla f(x) = 2(x_1, x_2)^T$, $\nabla c(x) = (1,1)^T$. 而且

$$-2\sigma c(x(\sigma)) = -2\sigma\left(\frac{4\sigma}{1+2\sigma} - 2\right) = -\frac{-4\sigma}{1+2\sigma}$$
$$\to 2(\sigma \to +\infty).$$

显然

$$\nabla f(x^*) - 2\nabla c(x^*) = 0.$$

所以$\lambda^* = 2$ 是对应的Lagrange 乘子.

注：

罚函数$P(x,\sigma)$ 的Hesse 矩阵为

$$H_P = \begin{pmatrix} 2+2\sigma & 2\sigma \\ 2\sigma & 2+2\sigma \end{pmatrix}.$$

当$\sigma = 10,\ 100,\ 1000,\ 10000$ 时, 矩阵的条件数分别是21, 201, 2001, 20001.

SUMT

算法的数值困难

- 当$\sigma_k \to \infty$ 时，步2 中对无约束优化问题的求解会变得越来越困难；
- $\nabla_{xx}^2 P$ 的条件数会趋于无穷.
- 当σ 足够大时，算法即使继续迭代，x_k 的精度也不会有多大改善.

例题

例3.2

用外罚函数法求解

$$\min\ (x_1-2)^4+(x_1-2x_2)^2$$
$$s.t.\ x_1^2-x_2=0.$$

终止条件为$\|c(x_k)\|\le\epsilon$, 分别取$\epsilon=10^{-3},10^{-5},10^{-6}$ 观察算法的表现.

【注：】这是教材上的例题，但是教材上题目的目标函数是

$$f(x)=(x_1-2)^4+(x_2-2x_1)^2$$

这似乎是印刷错误.

外罚函数法

对不等式约束问题

$$\begin{aligned} \min_{x \in R^n} \ & f(x) \\ \text{s.t. } & c_i(x) \geq 0, \ i \in \{1, \cdots, m\}. \end{aligned} \tag{14}$$

它的外罚函数定义为

$$P(x, \sigma) = f(x) + \sigma \tilde{P}(x),$$

其中, $\tilde{P}(x) = \sum_{i \in \mathcal{I}} [\min\{c_i(x), 0\}]^2$.

例题

例3.3

考虑不等式约束最优化问题

$$\min\ x_1^2 + x_2^2,$$
$$s.t.\ x_1 + x_2 - 2 \geq 0.$$

【解：】该问题的罚函数为

$$P(x,\sigma) = x_1^2 + x_2^2 + \sigma\left[\min\{x_1 + x_2 - 2, 0\}\right]^2.$$

也就是

$$P(x,\sigma) = \begin{cases} x_1^2 + x_2^2 + \sigma(x_1 + x_2 - 2)^2, & x_1 + x_2 - 2 < 0 \\ x_1^2 + x_2^2, & x_1 + x_2 - 2 \geq 0. \end{cases}$$

例题

罚函数的偏导数为

$$\frac{\partial P}{\partial x_1} = \begin{cases} (2+2\sigma)x_1 + 2\sigma x_2 - 4\sigma, & x_1 + x_2 - 2 < 0, \\ 2x_1, & x_1 + x_2 - 2 \geq 0. \end{cases}$$

$$\frac{\partial P}{\partial x_2} = \begin{cases} 2\sigma x_1 + (2+2\sigma x_2) - 4\sigma, & x_1 + x_2 - 2 < 0, \\ 2x_2, & x_1 + x_2 - 2 \geq 0. \end{cases}$$

只需求解

$$\frac{\partial P}{\partial x_1} = 0 = \frac{\partial P}{\partial x_2}.$$

当$x_1+x_2-2\geq 0$时，得$x_1=x_2=0$，这与$x_1+x_2-2\geq 0$矛盾．当$x_1+x_2-2<0$时，得

$$\begin{cases}(2+2\sigma)x_1+2\sigma x_2-4\sigma=0,\\ 2\sigma x_1+(2+2\sigma x_2)-4\sigma=0.\end{cases}$$

解得

$$x_1=x_2=\frac{2\sigma}{1+2\sigma}.$$

令$\sigma\to\infty$，得$x^*=(1,1)^T$．

注：

类似于等式约束的情形，若

$$\lim_{\sigma\to+\infty} \max\{-2\sigma c_i(x(\sigma)), 0\} = \lambda_i^*,$$

则λ_i^* 是在x^* 处相对于$c_i(x)$ 的Lagrange 乘子.

【注：】对任意的σ, 我们所得的解

$$\frac{2\sigma}{1+2\sigma}+\frac{2\sigma}{1+2\sigma}-2<0,$$

即罚函数的最优解在原问题的可行域外，这是“外罚函数法”这个名字的由来.

$$
\begin{aligned}
&\min\ f(x) \\
&\text{s.t. } c_i(x) = 0,\ i \in \mathcal{E} = \{1, \cdots, l\} \\
&\qquad c_i(x) \geq 0,\ i \in \mathcal{I} = \{l+1, \cdots, m\}.
\end{aligned} \tag{15}
$$

它的外罚函数定义为

$$P(x, \sigma) = f(x) + \sigma \tilde{P}(x),$$

其中,

$$\tilde{P}(x) = \sum_{i \in \mathcal{E}} c_i^2(x) + \sum_{i \in \mathcal{I}} [\min\{c_i(x), 0\}]^2.$$

例3.4

用$SUMT$ 算法求解如下问题

$$\begin{aligned} \min\ & -x_1x_2 \\ s.t.\ & c_1 = -x_1 - x_2^2 + 1 \geq 0, \\ & c_2 = x_1 + x_2 \geq 0. \end{aligned}$$

构造罚函数

$$P(x,\sigma) = -x_1x_2 + \sigma(\min(0, -x_1 - x_2^2 + 1))^2 + \sigma(\min(0, x_1 + x_2))^2.$$

取初始罚因子$\sigma = 1$, $c = 10$. 利用BFGS 算法求罚函数的最小值.

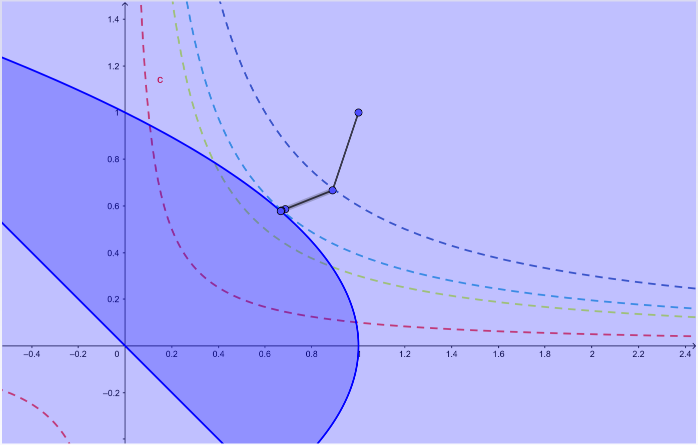
例题

罚函数的近似极小点$x(\sigma)$ 演化路径如下

迭代	初始点x_0	σ	$x(\sigma)$
0	$(1,1)^T$	1	$(0.899,0.667)^T$
1	$(0.899,0.667)^T$	10	$(0.686,0.586)^T$
2	$(0.686,0.586)^T$	100	$(0.669,0.578)^T$
3	$(0.669,0.578)^T$	1000	$(0.667,0.577)^T$

例题

如下图



收敛性

引理4

对于由外点函数法产生的点列$\{\boldsymbol{x}_k\}$, $k \geq 0$, 总有

$$
\begin{aligned}
&P(\boldsymbol{x}_{k+1}, \sigma_{k+1}) \geq P(\boldsymbol{x}_k, \sigma_k),\\
&\tilde{P}(\boldsymbol{x}_k) \geq \tilde{P}(\boldsymbol{x}_{k+1}),\\
&f(\boldsymbol{x}_{k+1}) \geq f(\boldsymbol{x}_k).
\end{aligned}
$$

证明:(1) 因为$\boldsymbol{x}_k$ 是$P(\boldsymbol{x}, \sigma_k)$ 的极小点, 且$\sigma_{k+1} > \sigma_k$, 故

$$
\begin{aligned}
P(\boldsymbol{x}_{k+1}, \sigma_{k+1}) &= f(\boldsymbol{x}_{k+1}) + \sigma_{k+1}\tilde{P}(\boldsymbol{x}_{k+1})\\
&\geq f(\boldsymbol{x}_{k+1}) + \sigma_k\tilde{P}(\boldsymbol{x}_{k+1}) = P(\boldsymbol{x}_{k+1}, \sigma_k) \geq P(\boldsymbol{x}_k, \sigma_k).
\end{aligned}
$$

(2)因为$\boldsymbol{x}_{k+1}$ 是$P(\boldsymbol{x},\sigma_{k+1})$ 的极小点, 将$\boldsymbol{x}=\boldsymbol{x}_k$ 代入$P(\boldsymbol{x},\sigma_{k+1})$ 中得

$$f(\boldsymbol{x}_k)+\sigma_{k+1}\tilde{P}(x_k)\geq f(\boldsymbol{x}_{k+1})+\sigma_{k+1}\tilde{P}(\boldsymbol{x}_{k+1}).$$

又因为$\boldsymbol{x}_k$ 是$P(\boldsymbol{x},\sigma_k)$ 的极小点, 将$\boldsymbol{x}=\boldsymbol{x}_{k+1}$ 代入$P(\boldsymbol{x},\sigma_k)$ 中得

$$f(\boldsymbol{x}_{k+1})+\sigma_k\tilde{P}(\boldsymbol{x}_{k+1})\geq f(\boldsymbol{x}_k)+\sigma_k\tilde{P}(\boldsymbol{x}_k).$$

将上述两式移项可得

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \leq \sigma_{k+1}(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})),$$
$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \geq \sigma_k(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})).$$

联立这两式,得

$$\begin{aligned} & \sigma_{k+1}(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})) \geq f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \\ \geq \ & \sigma_k(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})), \end{aligned}$$

即

$$(\sigma_{k+1} - \sigma_k)(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})) \geq 0.$$

因为$\sigma_{k+1} > \sigma_k$, 故而$\tilde{P}(\boldsymbol{x}_k) \geq \tilde{P}(\boldsymbol{x}_{k+1})$.

(3) 因$\sigma_k \geq 0$ 和(2) 及其的证明, 我们有

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \geq \sigma_k(\tilde{P}(\boldsymbol{x}_k) - \tilde{P}(\boldsymbol{x}_{k+1})) \geq 0,$$

即$f(\boldsymbol{x}_{k+1}) \geq f(\boldsymbol{x}_k)$.

收敛性

定理10 (外罚函数收敛性定理1)

设约束问题(15) 和无约束问题(11) 的整体最优解为$\boldsymbol{x}^*$ 和$\boldsymbol{x}_k$ $(\forall k \geq 1)$, 对正数序列$\{\sigma_k\}$, $\sigma_{k+1} \geq \sigma_k$ 且$\sigma_k \to +\infty$, 则由外罚函数法产生的点列$\{\boldsymbol{x}_k\}$ 的任何聚点$\tilde{\boldsymbol{x}}$ 必是(10) 的整体最优解.

证明:不妨设$\boldsymbol{x}_k \to \tilde{x}$. 因$\boldsymbol{x}^*$ 和$\boldsymbol{x}_k$ 分别是(15) 和(11) 的整体最优解, 故而$\tilde{P}(x^*) = 0$, 且

$$f(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + \sigma_k \tilde{P}(\boldsymbol{x}^*) \geq f(\boldsymbol{x}_k) + \sigma_k \tilde{P}(\boldsymbol{x}_k) = P(\boldsymbol{x}_k, \sigma_k).$$

由引理4及上式和$P(x_k, \sigma_k)$ 的单调有界性知, $\{P(x_k, \sigma_k)\}$ 收敛. 设其极限为p^0.

又由引理4, $f(\boldsymbol{x}_k)$ 单调增加且

$$f(\boldsymbol{x}_k) \le P(\boldsymbol{x}_k, \sigma_k) \le f(\boldsymbol{x}^*). \tag{16}$$

故而$\{f(\boldsymbol{x}_k)\}$ 也收敛. 设$f(\boldsymbol{x}_k) \to f^0$. 于是

$$\lim_{k\to+\infty} \sigma_k \tilde{P}(\boldsymbol{x}_k) = \lim_{k\to+\infty} [P(\boldsymbol{x}_k, \sigma_k) - f(\boldsymbol{x}_k)] = p^0 - f^0.$$

从而$\lim_{k\to+\infty} P(\boldsymbol{x}_k) = 0$.

由于$\boldsymbol{x}_k \to \tilde{\boldsymbol{x}}$, 并且$\tilde{P}(\boldsymbol{x})$ 连续, 所以$\tilde{P}(\tilde{x}) = 0$.
所以$\tilde{\boldsymbol{x}}$ 为可行解, 根据$\boldsymbol{x}^*$ 的定义有

$$f(\boldsymbol{x}^*) \leq f(\tilde{\boldsymbol{x}}).$$

对(16) 取极限,得

$$f(\tilde{\boldsymbol{x}}) = \lim_{k\to+\infty} f(\boldsymbol{x}_k) \leq f(\boldsymbol{x}^*).$$

所以有$f(\boldsymbol{x}^*) = f(\tilde{\boldsymbol{x}})$.

内罚函数法也叫障碍函数法或内点法.

- 迭代点始终保持在可行域内;
- 当迭代点靠近可行域边界时, 对其施以很大的惩罚, 阻止迭代点穿越边界.
- 该惩罚策略只能用于不等式约束问题, 并要求可行域有非空的内部.

问题

不等式约束优化问题

$$\begin{aligned} &\min\ f(x), \\ &s.t.\ c_i(x) \geq 0, i = 1, \cdots, m \end{aligned} \tag{17}$$

内罚函数（障碍函数）

$$B(x, r) = f(x) + r\tilde{B}(x),$$

其中，

$$\tilde{B}(x) = \sum_{i=1}^{m} \frac{1}{c_i(x)} \text{ 或} \tilde{B}(x) = -\sum_{i=1}^{m} \ln(c_i(x)).$$

分别称为倒数障碍函数或对数障碍函数.

取一个正数序列$\{r_k\}$, $\lim_{k\to+\infty} r_k = 0$, 将不等式约束优化问题(17) 转化为一些列无约束优化问题

$$\min \ B(x, r_k) = f(x) + r_k \tilde{B}(x). \tag{18}$$

内罚函数（障碍函数）法

已知约束优化问题(17), 其可行域的内部$D_0 \neq \emptyset$, 取控制误差$\epsilon > 0$ 和罚因子的缩小系数$0 < c < 1$.

【步 1】 选定初始点$x_0 \in D_0$, 给定$r_1 > 0$, 令$k = 1$.

【步 2】 以x_{k-1} 为初始点，求解无约束优化问题(18), 得最优解$x_k = x(r_k)$.

【步 3】 若$r_k \tilde{B}(x_k) < \epsilon$, 则$x_k$ 为(17) 的近似最优解, 停. 否则, 令$r_{k+1} = cr_k$, $k = k + 1$ 转步2.

例题

例3.5

用障碍函数法

$$
\begin{aligned}
&\min\ f(x_1,x_2)=\frac{1}{3}(x_1+1)^3+x_2\\
&s.t.\ 1-x_1\le 0,\\
&\qquad x_2\ge 0.
\end{aligned}
$$

【解：】增广目标函数为

$$B(x_1,x_2,r)=\frac{1}{3}(x_1+1)^3+x_2+r\left(\frac{1}{x_1-1}+\frac{1}{x_2}\right)$$

例题

令

$$\frac{\partial B}{\partial x_1} = (x_1+1)^2 - \frac{r}{(x_1-1)^2} = 0,$$
$$\frac{\partial B}{\partial x_2} = 1 - \frac{r}{x_2^2} = 0.$$

解得

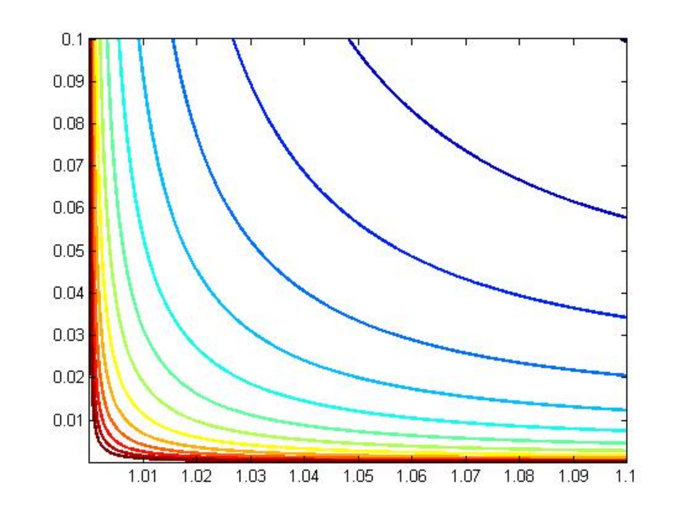
$$x(r) = (\sqrt{1+\sqrt{r}}, \sqrt{r})^T.$$

当$r \to 0$ 时，得

$$x^* = (1,0)^T,\ f^* = \frac{8}{3}.$$

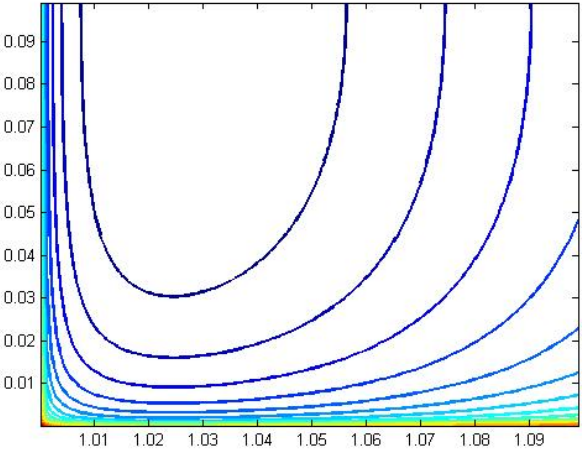
本题的障碍函数等高线图（取对数障碍函数）

图: 障碍函数等高线图$r = 1$



本题的障碍函数等高线图（取对数障碍函数）

图：障碍函数等高线图$r = 0.1$



例题

例3.6

用障碍函数法求解上题. 初始点$x_0=(3,4)^T$.

	倒数障碍		对数障碍		
k	r	$\|x_k-x^*\|$	k	r	$\|x_k-x^*\|$
1	10	3.3290	1	10	10.0565
2	1	1.0824	2	1	1.0209
3	0.1	0.3488	3	0.1	0.1209
⋮	⋮	⋮	⋮	⋮	⋮
12	1.00e-11	1.1180e-05	7	1.00e-05	1.0308e-05
13	1.00e-12	3.5355e-06	8	1.00e-06	1.0308e-06

收敛性

引理5

对于由内罚函数法产生的点列$\{\boldsymbol{x}_k\}$, 则有

(1)
$$B(\boldsymbol{x}_{k+1}, r_{k+1}) \leq B(\boldsymbol{x}_k, r_k),$$
即增广目标函数$B(\boldsymbol{x}_k, r_k)$ 关于k 是单调减小的;

(2) $\tilde{B}(\boldsymbol{x}_k) \leq \tilde{B}(\boldsymbol{x}_{k+1})$;

(3) $f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k)$.

收敛性

定理11

设可行域的内点集

$$D_0 = \{\boldsymbol{x} \in R^n \mid c_i(\boldsymbol{x}) > 0, i \in \mathcal{I}\}$$

非空，$f(\boldsymbol{x})$ 在可行域上存在极小点$\boldsymbol{x}^*$. 对严格单调递减趋向于*0* 的正数列$\{r_k\}$，内点法产生的点列$\{\boldsymbol{x}_k\}$ 的任何聚点$\tilde{\boldsymbol{x}}$ 是约束问题(17) 的最优解.

证明: 因$\boldsymbol{x}_k \in D_0 \subset D$, $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}_k)$, 故

$$B(\boldsymbol{x}_k, r_k) \geq f(\boldsymbol{x}_k) \geq f(\boldsymbol{x}^*).$$

因为$B(\boldsymbol{x}_k, r_k)$ 单调递减有下界，所以它有极限$\tilde{B}$,即

$$\lim_{k\to+\infty} B(\boldsymbol{x}_k, r_k) = \tilde{B} \geq f(\boldsymbol{x}^*).$$

下证

$$\lim_{k\to+\infty} B(\boldsymbol{x}_k, r_k) = f(\boldsymbol{x}^*).$$

由$f(\boldsymbol{x})$ 的连续性可知，对于任意小的正数ϵ, 存在$\delta > 0$ 和$\bar{\boldsymbol{x}} \in D_0$ 满足

$$\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\| < \delta, \ f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) < \epsilon/2.$$

由于$\lim_{k\to+\infty} r_k = 0$, 故存在$K > 0$, 当$k \geq K$ 时有

$$r_k \tilde{B}(\bar{\boldsymbol{x}}) < \epsilon/2.$$

由于$\boldsymbol{x}_k$ 是$B(\boldsymbol{x}, r_k)$ 的最优解, 故

$$B(\boldsymbol{x}_k, r_k) \leq B(\bar{\boldsymbol{x}}, r_k) = f(\bar{\boldsymbol{x}}) + r_k \tilde{B}(\bar{\boldsymbol{x}});$$

$$B(\boldsymbol{x}_k, r_k) - f(\boldsymbol{x}^*) \leq f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) + r_k \tilde{B}(\bar{\boldsymbol{x}}) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

- 内点法适于解仅含不等式约束的问题，且每次迭代的点都是可行点.
- 为寻找一个可行的初始点，往往需要较大的计算量.
- 为了防止迭代过程中，迭代点过于接近可行域边界. 在内循环中可要求$c(x_j) \geq \tau c_{j-1}$, τ 为很小的正数.
- 当x^* 处在边界处时，随着r 的减小，障碍函数的Hesse 矩阵的条件数也会趋向于无穷.
- 对一般约束优化问题，可以将外罚函数法和内罚函数法结合起来用.
- 一般地说，对数障碍函数计算效果更好.

等式约束的情形

先介绍等式约束的情形

$$\begin{aligned} &\min\ f(\boldsymbol{x}), \\ &\text{s.t. } c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E}. \end{aligned} \tag{19}$$

在Lagrange 函数上增加一个惩罚项, 我们称之为增广Lagrange 函数:

$$\begin{aligned} M(\boldsymbol{x}, \lambda, \sigma) &= L(\boldsymbol{x}, \lambda) + \frac{\sigma}{2}\|c(\boldsymbol{x})\|^2 \\ &= f(\boldsymbol{x}) - \lambda^T c(\boldsymbol{x}) + \frac{\sigma}{2}\|c(\boldsymbol{x})\|^2. \end{aligned} \tag{20}$$

基本思想

通过求解一系列无约束优化问题

$$\min \ M(\boldsymbol{x}, \lambda_k, \sigma_k)$$

来求解原问题. 每次迭代（外循环）同时修正$\boldsymbol{x}_k$ 和λ_k. 其一阶必要条件为

$$\begin{aligned}&\nabla_{\boldsymbol{x}} M(\boldsymbol{x}, \lambda_k, \sigma_k)\\ &\quad= \nabla f(\boldsymbol{x}) - \nabla c(\boldsymbol{x})\lambda_k + \sigma_k \nabla c(\boldsymbol{x}) c(\boldsymbol{x})\\ &\quad= \nabla f(\boldsymbol{x}) - \nabla c(\boldsymbol{x})(\lambda_k - \sigma_k c(\boldsymbol{x})).\end{aligned} \tag{21}$$

乘子的修正

设$\boldsymbol{x}_k$ 为无约束优化问题

$$\min \ M(\boldsymbol{x}, \lambda_k, \sigma_k)$$

的解. 则有

$$\nabla_{\boldsymbol{x}} M(\boldsymbol{x}_k, \lambda_k, \sigma_k) = \nabla f(\boldsymbol{x}_k) - \nabla c(\boldsymbol{x}_k)(\lambda_k - \sigma_k c_k) = 0.$$

从而可利用

$$\lambda_{k+1} = \lambda_k - \sigma_k c_k$$

来修正乘子.

例题

例4.1

$$
\begin{aligned}
&\min\ x_1 + x_2 \\
&s.t.\ x_1^2 + x_2^2 - 2 = 0.
\end{aligned}
$$

他的最优解为$\boldsymbol{x}^* = (-1, -1)^T$. 对应的*Lagrange* 乘子为$\lambda^* = -0.5$.

在求解之前先比较一下它的外罚函数和增广Lagrange 函数的等高线图.

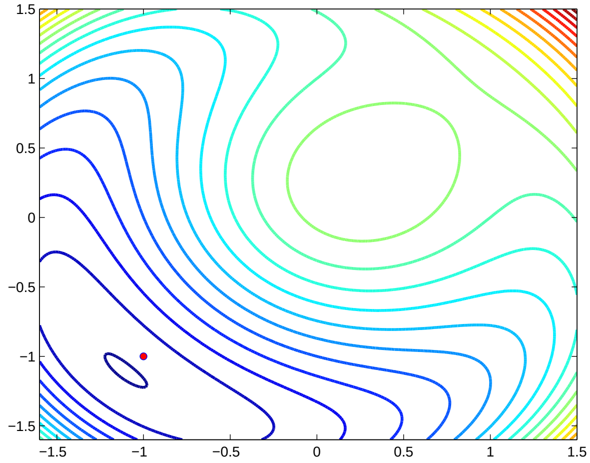
取罚因子$\rho = 0.5$, 定义外罚函数

$$
P(\boldsymbol{x}, \rho) = x_1 + x_2 + \rho(x_1^2 + x_2^2 - 2)^2.
$$

该罚函数的最优解约为$\boldsymbol{x}^* = (-1.1072, -1.1073)^*$.

例题

其等高线图如下



例题

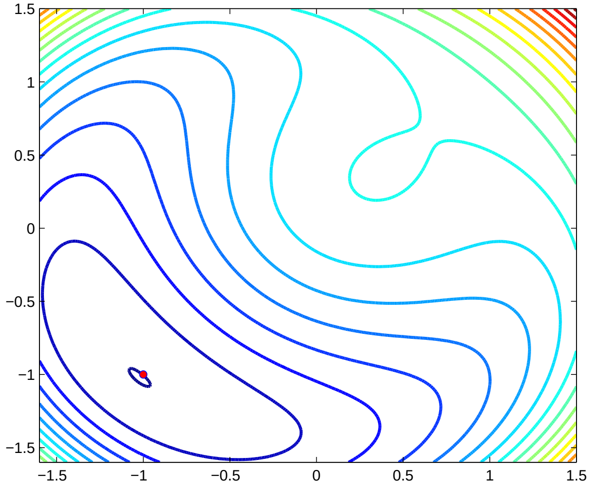
罚因子仍取$\rho = 0.5$, 另取乘子估计$\lambda = -0.4$.
定义增广Language 函数

$$M(\boldsymbol{x}, \lambda, \rho) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2) + \rho(x_1^2 + x_2^2 - 2)^2.$$

该增广Language 函数的最优解约为$\boldsymbol{x}^* = (-1.0221, -1.0221)^T$.

例题

其等高线图如下



例题

例4.2

考虑等式约束最优化问题

$$\min\ x_1^2+x_2^2,$$
$$s.t.\ x_1+x_2-2=0.$$

问题的增广Lagrange 函数为

$$M(\boldsymbol{x},\lambda,\sigma)=x_1^2+x_2^2-(x_1+x_2-2)\lambda+\frac{\sigma}{2}(x_1+x_2-2)^2.$$

令

$$\begin{cases}\dfrac{\partial M}{\partial x_1}=(2+\sigma)x_1+\sigma x_2-(\lambda+2\sigma)=0,\\ \dfrac{\partial M}{\partial x_2}=\sigma x_1+(2+\sigma)x_2-(\lambda+2\sigma)=0.\end{cases}$$

例题

解之得

$$x_1 = x_2 = \frac{\lambda + 2\sigma}{2 + 2\sigma}. \tag{22}$$

当$\sigma > 0 (> -1)$ 时,

$$\nabla^2_{\boldsymbol{xx}} M(\boldsymbol{x}, \lambda, \sigma) = \begin{pmatrix} 2+\sigma & \sigma \\ \sigma & 2+\sigma \end{pmatrix}$$

正定. 从而上述解为增广Lagrange 函数的严格局部极小点. 将(22) 代入乘子迭代公式

$$\lambda_{k+1} = \lambda_k - \sigma c_k,$$

得

$$\lambda_{k+1} = \frac{2}{2+2\sigma}\lambda_k + \frac{4\sigma}{2+2\sigma}.$$

例题

对任意的$\sigma > 0$, 由这个公式确定的数列$\{\lambda_k\}$ 收敛, 设其极限为λ^*.
两边取极限, 得

$$\lambda^* = \frac{2}{2+2\sigma}\lambda^* + \frac{4\sigma}{2+2\sigma}.$$

解之得, $\lambda^* = 2$. 代入(22) 中得解

$$x^* = (1,1)^T.$$

从而,原问题的解为$\boldsymbol{x}^* = (1,1)^T$, 相应的Lagrange 乘子为$\lambda^* = 2$.

例题

确定λ^* 的另一方法：

若要使该解满足约束条件，则有

$$\lambda^* = 2.$$

稍后介绍上述方法的原理.

例题

- 由上述问题的解题过程可知，当$\sigma > 0 (> -1)$ 时，增广Lagragne 函数的Hesse矩阵正定. 从而可以用无约束优化算法求解.
- 同时, 对足够大的σ(本题中, 只要取$\sigma \geq 0$), 增广Lagrange函数的局部极小点就是原问题的局部极小点.
- 无须使$\sigma \to \infty$.

算法

等式约束优化问题的增广Lagrange 函数法

【步 1】 选定初始点$\boldsymbol{x}_0$、初始乘子向量λ_1、初始罚因子σ_1及放大系数$c > 1$、误差$\epsilon > 0$ 与常数$\theta \in (0, 1)$，令$k = 1$.

【步 2】 以$\boldsymbol{x}_{k-1}$ 为初始点求解无约束问题

$$\min \ M(\boldsymbol{x}, \lambda_k, \sigma_k) = f(\boldsymbol{x}) - \lambda_k^T c(\boldsymbol{x}) + \frac{\sigma_k}{2}\|c(\boldsymbol{x})\|^2$$

得（近似）最优解.

【步 3】 当$\|c_k\| < \epsilon$ 时，x_k 为近似最优解，停. 否则，转步4.

【步 4】 当$\|c_k\|/\|c_{k-1}\| \le \theta$ 时，转步5，否则，令$\sigma_{k+1} = c\sigma_k$，转步5.

【步 5】 令$\lambda_{k+1} = \lambda_k - \sigma_k c_k$，$k = k + 1$，转步2.

收敛性

引理6

设$A \in R^{n\times n}$, $B \in R^{n\times m}$, 对满足$B^Td = 0$, $d \neq 0$ 的d, $d^TAd > 0$的充分必要条件是存在$\sigma^* \geq 0$, 使得对任给$\sigma \geq \sigma^*$和$d \neq 0$, 有

$$d^T(A + \sigma BB^T)d > 0. \tag{23}$$

证明:充分性. 因$B^Td = 0$, 故而$d^TBB^Td = 0$, 由式(23) 知, $d^TAd > 0$.

必要性. 先证明存在一个数$\sigma^* > 0$, 对任意的$d \in R^n$ 均有

$$d^T(A + \sigma^* BB^T)d > 0 (d \neq 0). \tag{24}$$

用反证法. 假设(24) 不成立. 必存在向量d_k 且$\|d_k\| = 1$ 使得

$$d_k^T(A + kBB^T)d_k \leq 0, \tag{25}$$

由于$\{d_k\}$ 为有界序列, 必有收敛子列d_{k_i}, 其极限为$\bar{d}$, $\|\bar{d}\| = 1$. 对于$\{d_{k_i}\}$ 由式(25) 得

$$d_{k_i}^T(A + k_i BB^T)d_{k_i} \leq 0.$$

对于前式两端取极限, $k_i \to +\infty$ 得

$$\bar{d}^T A \bar{d} + \lim_{k_i \to +\infty} \|B^T d_{k_i}\|^2 \leq 0.$$

这意味着

$$\lim_{k_i \to +\infty} B^T d_{k_i} = B^T \bar{d} = 0,$$

故有$\bar{d}^T A \bar{d} \leq 0$, $\bar{d} \neq 0$.
这与引理中的条件矛盾. 证毕.

收敛性

定理12

设$\boldsymbol{x}^*$ 是等式约束优化问题(19) 的严格局部最优解，λ^* 是相应的Lagrange 乘子，在$\boldsymbol{x}^*$，λ^* 处二阶充分条件（定理2）成立．则存在$\sigma^* \geq 0$，对任给$\sigma \geq \sigma^*$，$\boldsymbol{x}^*$ 是增广Lagrange 函数$M(\boldsymbol{x}, \lambda^*, \sigma)$ 的严格局部极小点；反之，若$c_i(\boldsymbol{x}^*) = 0$，$i = 1, 2, \cdots, m$，$\boldsymbol{x}^*$ 是$M(\boldsymbol{x}, \lambda^*, \sigma)$ 的局部极小点，则$\boldsymbol{x}^*$ 是等式约束优化问题(19) 的局部最优解.

证明: 由$M(\boldsymbol{x}, \lambda, \sigma)$ 的定义有

$$
\begin{aligned}
\nabla_{\boldsymbol{x}} M(\boldsymbol{x}, \lambda, \sigma) &= \nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) + \sigma \nabla c(\boldsymbol{x}) c(\boldsymbol{x}) \\
&= \nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) + \sigma \sum_{i=1}^{l} \nabla c_i(\boldsymbol{x}) c_i(\boldsymbol{x}). \qquad (26)
\end{aligned}
$$

从而

$$\nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}, \lambda, \sigma) = \nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}, \lambda) + \sigma\left(\sum_{i=1}^{l} c_i(\boldsymbol{x})\nabla_{\boldsymbol{x}}^2 c_i(\boldsymbol{x}) + \nabla c_i(\boldsymbol{x})\nabla c_i(\boldsymbol{x})^T\right).$$

由于$c_i(\boldsymbol{x}^*) = 0$, 故而

$$\begin{aligned}
&\nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}^*, \lambda^*, \sigma) \\
&= \nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}^*, \lambda) + \sigma\sum_{i=1}^{l}\nabla c_i(\boldsymbol{x}^*)\nabla c_i(\boldsymbol{x}^*)^T \\
&= \nabla_{\boldsymbol{x}}^2 L(\boldsymbol{x}^*, \lambda) + \sigma\nabla c(\boldsymbol{x}^*)\nabla c(\boldsymbol{x}^*)^T.
\end{aligned}$$

由二阶充分条件, 对每个满足

$$\nabla c(\boldsymbol{x}^*)^T z = 0$$

的非零向量z, 有

$$z^T \nabla^2_{\boldsymbol{x}} L(\boldsymbol{x}^*, \lambda^*) z > 0.$$

由引理6 知, 存在$\sigma^* > 0$, 使得当$\sigma \geq \sigma^*$ 且$z \neq 0$ 时有

$$z^T M(\boldsymbol{x}^*, \lambda^*, \sigma) z > 0.$$

由式(26) 和$c(\boldsymbol{x}^*) = 0$ 知

$$\nabla_{\boldsymbol{x}} M(\boldsymbol{x}^*, \lambda^*, \sigma) = \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \lambda^*) = 0.$$

由无约束优化问题的二阶充分条件, $\boldsymbol{x}^*$ 为$M(x, \lambda^*, \sigma)$ 的严格局部极小点.

反之, 因$\boldsymbol{x}_0$ 是$M(\boldsymbol{x},\lambda_0,\sigma)$ 的局部极小点且$c(\boldsymbol{x}_0)=0$, 故对任意与$\boldsymbol{x}_0$ 充分靠近的可行解$\bar{\boldsymbol{x}}$, 有

$$M(\boldsymbol{x}_0,\lambda_0,\sigma)\le M(\bar{\boldsymbol{x}},\lambda_0,\sigma).$$

但$c(\boldsymbol{x}_0)=0$, $c(\bar{\boldsymbol{x}})=0$, 故而

$$M(\boldsymbol{x}_0,\lambda_0,\sigma)=f(\boldsymbol{x}_0),\ M(\bar{\boldsymbol{x}},\lambda_0,\sigma)=f(\bar{\boldsymbol{x}}).$$

从而

$$f(\boldsymbol{x}_0)\le f(\bar{\boldsymbol{x}}).$$

即$\boldsymbol{x}_0$ 为约束优化问题(1) 的局部极小点.

定理13

设$\boldsymbol{x}_k$ 是(20) 的最优解，则λ_k 为相应的*Lagrange* 乘子向量的充要条件是$c(\boldsymbol{x}_k)=0$.

证明:必要性是显然的. 下证充分性.

设$\boldsymbol{x}_k$ 是(20) 的最优解且$c(\boldsymbol{x}^*)=0$, 则对任意的

$$\boldsymbol{x}\in D=\{\boldsymbol{x}\in R^n\,|c(\boldsymbol{x})=0\}$$

有

$$f(\boldsymbol{x})=M(\boldsymbol{x},\lambda_k,\sigma)\geq M(\boldsymbol{x}_k,\lambda_k,\sigma)=f(\boldsymbol{x}_k),$$

即$\boldsymbol{x}_k$ 为(1) 的最优解. 又因$c(\boldsymbol{x}_k)=0$ 和(21), 有

$$\nabla f(\boldsymbol{x}_k)-\nabla c(\boldsymbol{x}_k)\lambda_k=0.$$

即λ_k 是与最优解$\boldsymbol{x}_k$ 相应的Lagrange 乘子.

数值算例

例4.3

用增广 *Lagrange* 函数法求解约束问题

$$
\begin{aligned}
&\min \ f(\boldsymbol{x}) = (x_1-2)^4 + (x_1-2x_2)^2 \\
&s.t. \ x_1^2 - x_2 = 0.
\end{aligned}
$$

$$
\begin{aligned}
&\min\ f(x) \\
&\text{s.t. } c_i(x) = 0,\ i \in \mathcal{E} = \{1, \cdots, l\} \\
&\qquad c_i(x) \geq 0,\ i \in \mathcal{I} = \{l+1, \cdots, m\}.
\end{aligned} \tag{27}
$$

先把它转化为如下标准形式.

$$
\begin{aligned}
&\min\ f(x)\\
&\text{s.t. } c_i(x)=0,\ i\in\mathcal{E}=\{1,\cdots,l\}\\
&\qquad c_i(x)-s_i=0,\ i\in\mathcal{I}=\{l+1,\cdots,m\},\\
&\qquad s_i\ge 0,\ i\in\mathcal{I}.
\end{aligned}
\tag{28}
$$

暂时忽略不等式约束，有增广Lagrange 函数

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

其中,

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

考虑

$$\begin{aligned} \min_{x,s}\ & M(x,s,\lambda,\sigma), \\ s.t.\ & s_i \ge 0, i \in \mathcal{I}. \end{aligned}$$

先化简这个问题，转化为

$$\begin{aligned} \min_{x}\min_{s}\ & M(x,s,\lambda,\sigma), \\ s.t.\ & s_i \ge 0, i \in \mathcal{I}. \end{aligned}$$

从内层规划问题

$$\min_{s} M(x,s,\lambda,\sigma),$$
$$s.t.\ s_i \geq 0, i \in \mathcal{I}.$$

中求s_i.
得

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

其中, $\eta_i = \lambda_i/\sigma$. 想想为什么? 其实很容易.

则

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

或

$$c_i(x) - s_i = \min\{c_i(x), \eta_i\}.$$

再将其代入$\bar{\phi}$, 得

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

或

$$\phi_i = \frac{1}{2}\sigma\left[(\min\{c_i(x) - \eta_i, 0\})^2 - \eta_i^2\right].$$

从而得到简化的增广Lagrange 函数

$$M(x,\lambda,\sigma) = f(x) - \sum_{i\in\mathcal{E}} \lambda_i(x)c_i(x) + \frac{\sigma}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \sum_{i\in\mathcal{I}} \phi_i.$$

迭代格式和终止准则

- 乘子迭代格式

$$\begin{cases} \lambda_i^{(k+1)} = \lambda_i^{(k)} - \sigma_k c_i(x_k), & i \in \mathcal{E} \\ \lambda_i^{(k+1)} = -\sigma_k \min\{c_i(x_k) - \eta_i^{(k)}, 0\}, & i \in \mathcal{I} \end{cases}$$

- 终止准则

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

将以上两条代入等式约束优化问题的增广Lagrange 函数算法，即得一般约束优化问题的算法.

例题

例4.4

考虑不等式约束最优化问题

$$\begin{aligned} \min\ & x_1^2 + x_2^2, \\ s.t.\ & x_1 + x_2 - 2 \geq 0. \end{aligned}$$

【解】转化为标准格式

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

记$\eta = \lambda/\sigma$,

$$\phi = \begin{cases} -\frac{1}{2}\sigma\eta^2, & x_1 + x_2 - 2 - \eta \geq 0, \\ \frac{1}{2}\sigma[(x_1 + x_2 - 2 - \eta)^2 - \eta^2], & x_1 + x_2 - 2 - \eta < 0 \end{cases}$$

例题

其Lagrange 函数为

$$M(x,\lambda,\sigma) = x_1^2 + x_2^2 + \phi$$
$$= \begin{cases} x_1^2 + x_2^2 - \frac{1}{2\sigma}\lambda^2, & x_1 + x_2 - 2 - \eta \geq 0, \\ x_1^2 + x_2^2 \\ +\frac{1}{2}\sigma\left[\left(x_1 + x_2 - 2 - \frac{\lambda}{\sigma}\right)^2 - \frac{\lambda^2}{\sigma^2}\right], & x_1 + x_2 - 2 - \eta < 0. \end{cases}$$

当$x_1 + x_2 - 2 - \eta \geq 0$ 时, 令

$$\frac{\partial M}{\partial x_1} = 2x_1 = 0, \quad \frac{\partial M}{\partial x_2} = 2x_2 = 0$$

得$\tilde{x} = (0,0)^T$. 当σ 充分大时, $x_1 + x_2 - 2 < \lambda/\sigma$, 矛盾.

例题

当 $x_1 + x_2 - 2 - \eta < 0$ 时,令

$$\frac{\partial M}{\partial x_1} = 2x_1 - [\lambda - \sigma(x_1 + x_2 - 2)] = 0,$$
$$\frac{\partial M}{\partial x_2} = 2x_2 - [\lambda - \sigma(x_1 + x_2 - 2)] = 0.$$

解得

$$x_1 = x_2 = \frac{2\sigma + \lambda}{2\sigma + 2}.$$

且

$$x_1 + x_2 - 2 - \frac{\lambda}{\sigma} < 0.$$

要使这个解满足约束条件，需

$$x_1 + x_2 = \frac{4\sigma + 2\lambda}{2\sigma + 2} \geq 2 \Rightarrow \lambda \geq 2 > 0.$$

由互补条件，此时$x_1 + x_2 - 2 = 0$. 故有$\lambda = 2$. 相应的

$$x_1 = x_2 = 1.$$

从而$\sigma \geq 0$ 时，增广Lagrange 函数的极小点$\tilde{x} = (1,1)^T$ 就是原问题的极小点.

例题

例4.5

考虑不等式约束优化问题

$$\begin{aligned} &\min\ x_1^2 + x_2^2 \\ &s.t.\ x_1 - x_2 + 1 \le 0 \end{aligned}$$

该问题的最优解为$x^* = (-\frac{1}{2}, \frac{1}{2})$，相应的*Lagrange* 乘子为$\lambda^* = 1$.

用外罚函数法，定义外罚函数

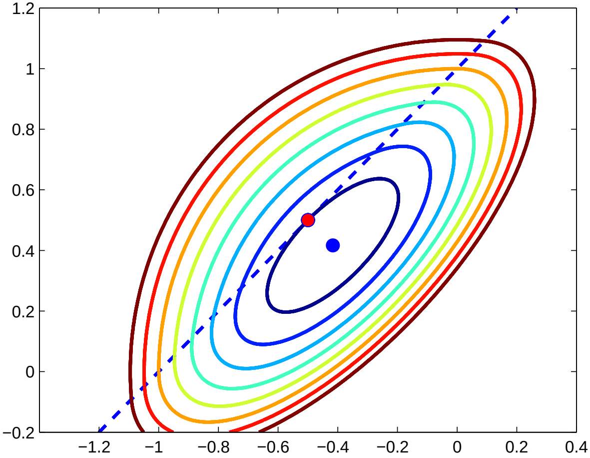
$$P(x,\sigma) = x_1^2 + x_2^2 + \frac{\sigma}{2}(\min(-x_1 + x_2 - 1, 0))^2.$$

外罚函数的最小值为

$$x_P^* = \left(-\frac{\sigma}{2+2\sigma}, \frac{\sigma}{2+2\sigma}\right).$$

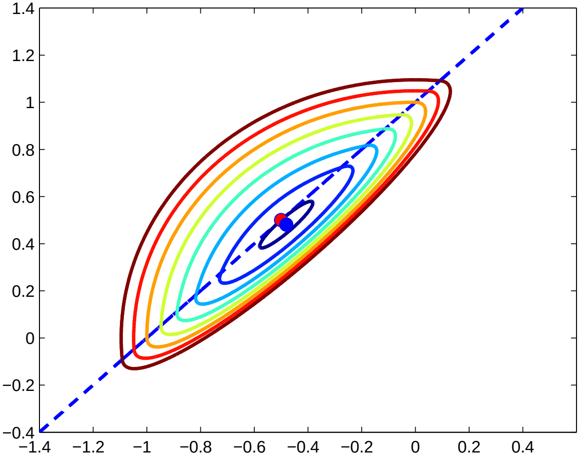
例题

图：外罚函数$\sigma = 5$



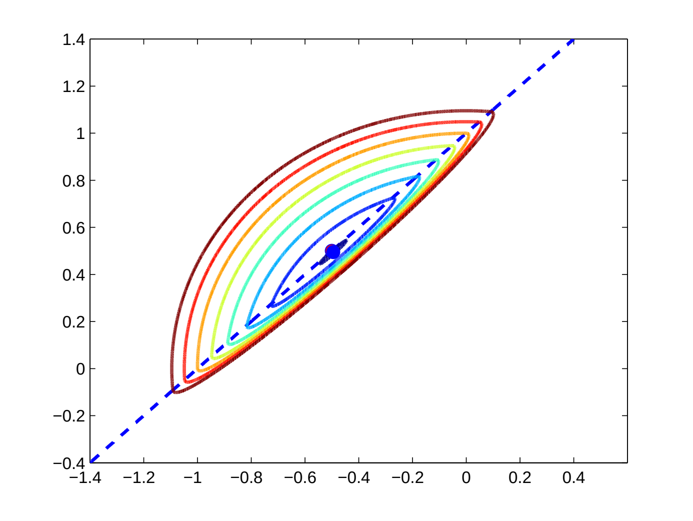
例题

图：外罚函数$\sigma = 25$



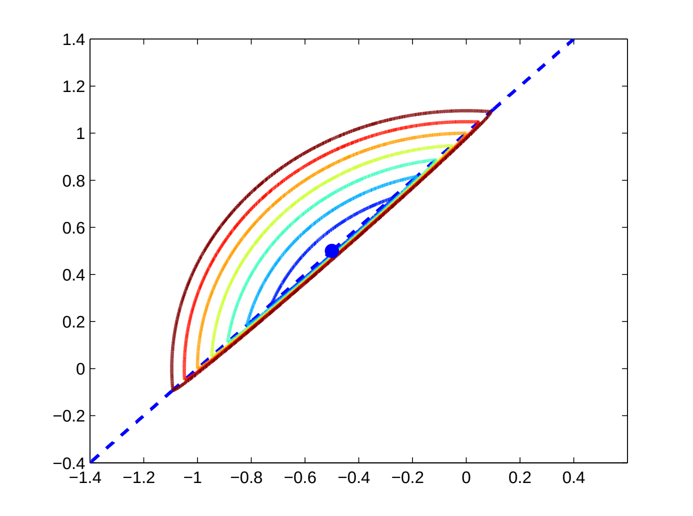
例题

图: 外罚函数$\sigma = 100$



例题

图: 外罚函数$\sigma = 1000$



例题

用内罚函数法（内点法），定义障碍函数

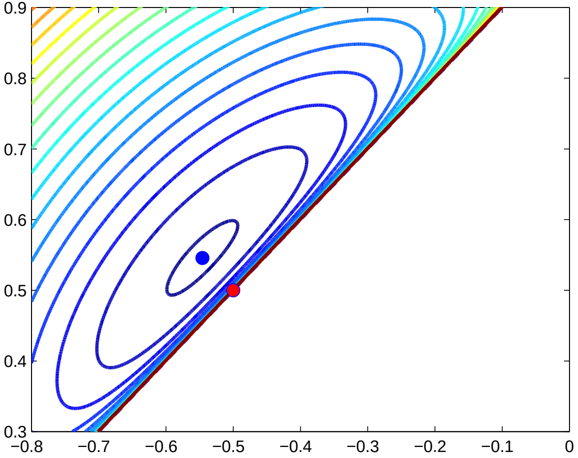
$$b(x,\sigma)=x_1^2+x_2^2-\mu\log(-x_1+x_2-1).$$

分别取障碍因子$\mu=0.1;0.05;0.01;0.001$，障碍函数的最小点x_μ^*分别约为

$$\begin{pmatrix}-0.5458\\0.5458\end{pmatrix},\begin{pmatrix}-0.5239\\0.5239\end{pmatrix},\begin{pmatrix}-0.5050\\0.5050\end{pmatrix},\begin{pmatrix}-0.5005\\0.5005\end{pmatrix}.$$

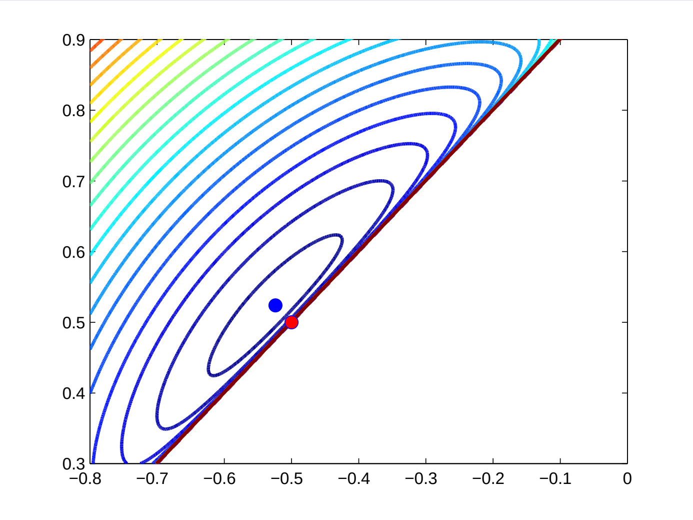
例题

图: 外罚函数$\mu = 0.1$



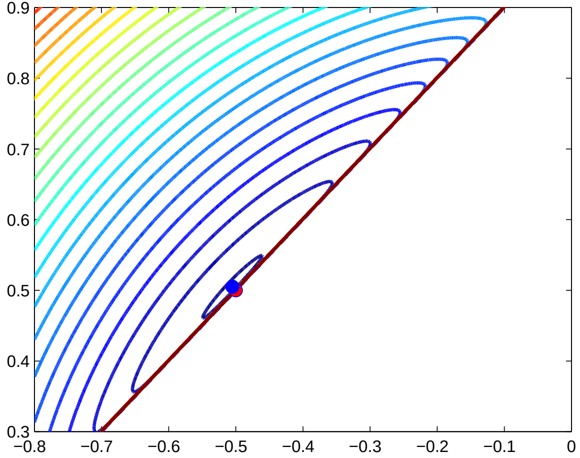
例题

图：内罚函数$\mu = 0.05$



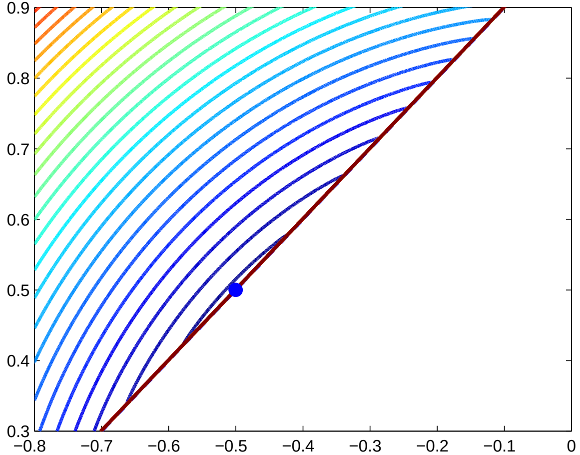
例题

图：内罚函数$\mu = 0.01$



例题

图: 内罚函数$\mu = 0.001$



例题

用增广Lagrange 法, 定义增广Lagrange 函数

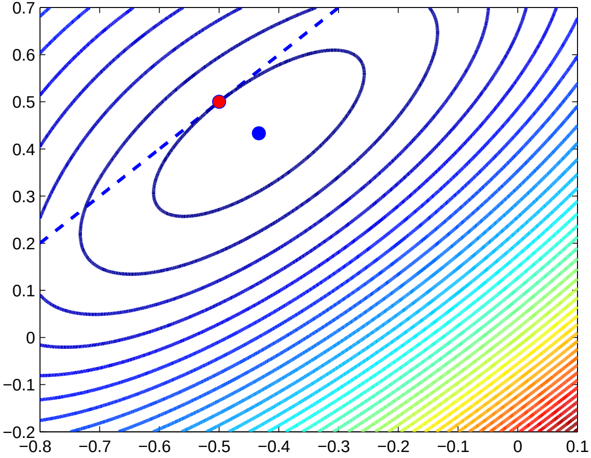
$$M(x,\sigma,\lambda)=x_1^2+x_2^2-(-x_1+x_2-1)\lambda+\frac{\sigma}{2}(\min(-x_1+x_2-1,0))^2.$$

其最优解为

$$x_M^*=\left(-\frac{\sigma+\lambda}{2+2\sigma},\frac{\sigma+\lambda}{2+2\sigma}\right).$$

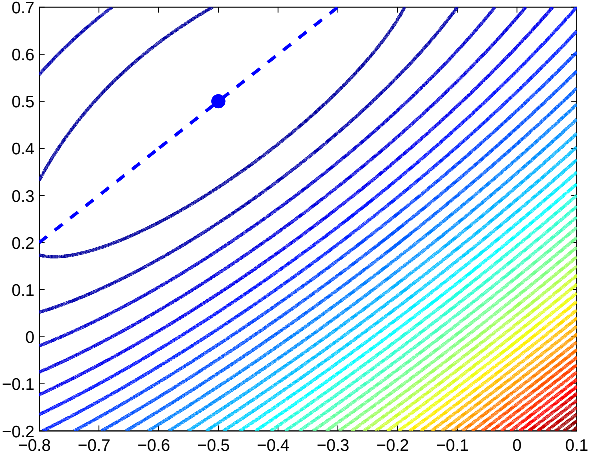
例题

图：增广Lagrange 函数法$\sigma = 5,\ \lambda = 0.2$



例题

图: 增广Lagrange 函数法$\sigma = 25$, $\lambda = 1$



数值试验

问题1

$$\begin{aligned} \min\ & f(x) = -x_1, \\ s.t.\ & x_2^2 - x_1^3 - x_3^2 = 0, \\ & x_1^2 - x_2 - x_4^2 = 0. \end{aligned}$$

解为$x^* = (1,1,0,0)^T$, 初始点为$x^{(0)} = (2,2,2,2)^T$.

问题2

$$\begin{aligned} \min\ & f(x) = \frac{1}{2}x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2, \\ s.t.\ & 25 - 4x_1^2 - x_2^2 \geq 0. \end{aligned}$$

解为$x^* = (2,3)^T$, 初始点$x^{(0)} = (0,0)^T$.

问题	方法	ite	feva	$\sigma_k(r_k)$	$\|x_k - x^*\|$
问题1	外罚	6	235	1.00E+06	2.69E-06
	Aug. Lag.	1	51	1.00E+01	2.4547e-09
问题2	障碍	8	155	3.5527E-07	4.52E-08
	Aug. Lag.	8	229	10000	4.41E-10

考虑仅含不等式约束的问题(3)

$$\min \ f(\boldsymbol{x})$$
$$s.t. \ x \in D,$$

其中, $D = \{x \in R^n \,|c_i(\boldsymbol{x}) \geq 0, \ i = 1, \cdots, m\}$.

基本概念：回顾下降方向与可行方向

定义6 (下降方向)

对于向量$0 \neq d \in R^n$, 和某个点x 处. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, 有

$$f(x+\alpha d) < f(x),$$

则称d 为x 处的一个下降方向.

定义7 (可行方向)

对于向量$0 \neq d \in R^n$, 和可行域Ω 中的某个点$x \in \Omega$. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, $x + \alpha d$ 仍然属于Ω, 即$x + \alpha d \in \Omega$, 则称d 为x 处的一个可行方向.

记

$$\mathcal{A}(\bar{\boldsymbol{x}}) = \{i|c_i(\bar{\boldsymbol{x}}) = 0, i = 1, \cdots, m\}.$$

若$\bar{p}$ 满足

$$\nabla c_i(\bar{\boldsymbol{x}})^T \bar{p} > 0,\ i \in \mathcal{A}(\bar{\boldsymbol{x}}),$$

则$\bar{p}$ 为$\bar{\boldsymbol{x}}$ 处的可行方向.

若$c_i(\boldsymbol{x})$ 为 $\boxed{\text{线性函数}}$, 则当且仅当

$$\nabla c_i(\bar{\boldsymbol{x}})^T \bar{p} \geq 0,\ i \in \mathcal{A}(\bar{\boldsymbol{x}}),$$

时, $\bar{p}$ 为可行方向.

若

$$\nabla f(\bar{\boldsymbol{x}})^T \bar{p} < 0,$$

则$\bar{p}$ 为下降方向.
若$\bar{p}$ 既是可行方向又是下降方向, 则称$\bar{p}$ 为可行下降方向.

例5.1

考察如下约束优化问题在$\bar{\boldsymbol{x}} = (2,3)^T$ 处的可行下降方向

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

定义8 (投影矩阵)

设方阵$P_{n\times n}$ 满足

$$P = P^T \text{ 且} PP = P,$$

则称方阵P 为投影矩阵.

投影矩阵的性质

引理7

设矩阵P 为投影矩阵，则

(i) P 为正半定;

(ii) $Q = I - P$ 亦为投影矩阵;

(iii) 令

$$L = \{Px \mid x \in R^n\}, \ L^{\perp} = \{Qx \mid x \in R^n\},$$

则L 与$L^{\perp}$ 为正交的线性空间，并且对任何$x \in R^n$, x 可唯一地被分解为

$$x = p + q, \ p \in L, \ q \in L^{\perp}.$$

投影梯度法是一种可行方向法.

考虑仅含不等式约束的最优化问题

$$\begin{aligned} &\min\ f(x) \\ &s.t.\ A^T x \geq b, \end{aligned} \tag{29}$$

其中, $f(x)$ 为可微函数, A 为$n \times m$ 矩阵, b 为m 维向量($m \leq n$).

定理14

设$\bar{x}$ 为(29) 的一个可行点, 设在$\bar{x}$ 处有q 个有效约束, 不妨设前q 个为有效约束, 则$A_q^T\bar{x} = b_q$, 其中, A_q 为A 的前q 列组成, b_q 为b 的前q 个分量. 设A_q 列满秩, 则矩阵

$$P_q = I - A_q(A_q^TA_q)^{-1}A_q^T \tag{30}$$

为投影矩阵, 当$P_q\nabla f(\bar{x}) \neq 0$ 时, $\bar{p} = -P_q\nabla f(\bar{x})$ 为$\bar{x}$ 处的一个可行下降方向.

证明:设$Q = A_q(A_q^TA_q)^{-1}A_q^T$. 因$rank(A_q) = q$, 所以$A_q^TA_q$ 为q 阶非奇异方阵, 因而Q 有意义. 又容易验证

$$Q^T = Q, QQ = Q.$$

故而Q 是投影矩阵. 从而$P_q = I - Q$ 也是投影矩阵.

因为

$$\begin{aligned}\bar{p}^T\nabla f(\bar{\boldsymbol{x}}) &= -\nabla f(\bar{x})^T P_q \nabla f(\bar{\boldsymbol{x}}) = -\|P_q\nabla f(\bar{\boldsymbol{x}})\|^2 < 0,\\ A_q^T\bar{p} &= -A_q^T P_q\nabla f(\bar{\boldsymbol{x}})\\ &= -[A_q^T - A_q^T A_q(A_q^T A_q)^{-1}A_q^T]\nabla f(\bar{\boldsymbol{x}}) = 0.\end{aligned}$$

所以$\bar{p}$ 为可行下降方向.

定理15

设$\bar{x}$ 为(29) 的一个可行点，且$P_q\nabla f(\bar{x})=0$，令

$$\lambda=(A_q^TA_q)^{-1}A_q^T\nabla f(\bar{x}), \tag{31}$$

则有

(i) 若$\lambda\geq 0$，则$\bar{x}$ 为KKT 点;

(ii) 若$\lambda\ngeq 0$，设某个分量$\lambda_i<0$，则从A_q 中去掉对应于λ_i 的一列q_i 后得A_{q-1}，令

$$P_{q-1}=I-A_{q-1}(A_{q-1}^TA_{q-1})^{-1}A_{q-1}^T,$$

则P_{q-1} 为投影矩阵，且

$$\bar{p}=-P_{q-1}\nabla f(\bar{x})$$

为$\bar{x}$ 的可行下降方向.

证明:(i) 因

$$
\begin{aligned}
P_q \nabla f(\bar{\boldsymbol{x}}) &= [I - A_q(A_q^T A_q)^{-1} A_q^T] \nabla f(\bar{\boldsymbol{x}}) \\
&= \nabla f(\bar{\boldsymbol{x}}) - A_q \lambda = 0 \qquad (32)
\end{aligned}
$$

且$\lambda \geq 0$, 故由KKT 条件知$\bar{\boldsymbol{x}}$ 为KKT 点.
(ii) 首先证明$P_{q-1}\nabla f(\bar{\boldsymbol{x}}) \neq 0$. 用反证法. 设$P\nabla f(\bar{\boldsymbol{x}}) = 0$, 则

$$
\nabla f(\bar{\boldsymbol{x}}) - A_{q-1}(A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T \nabla f(\bar{\boldsymbol{x}}) = 0.
$$

令

$$
\hat{\lambda} = (A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T \nabla f(\bar{\boldsymbol{x}}),
$$

则$\nabla f(\bar{\boldsymbol{x}}) = A_{q-1}\hat{\lambda}$.

但由(32) 有

$$\nabla f(\bar{\boldsymbol{x}}) = A_q\lambda = (A_{q-1}, a_i)\begin{pmatrix}\bar{\lambda}\\ \lambda_i\end{pmatrix} = A_{q-1}\bar{\lambda} + \lambda_i a_i, \lambda_i < 0, \tag{33}$$

其中$\bar{\lambda}$ 是λ 中除λ_i 以外的分量构成的向量. 故有

$$A_{q-1}\hat{\lambda} = A_{q-1}\bar{\lambda} + \lambda_i a_i,$$

即

$$A_{q-1}(\hat{\lambda} - \bar{\lambda}) - \lambda_i a_i = 0, \lambda_i < 0.$$

上式表明A_q 中各列线性相关, 这与A_q 列满秩相矛盾. 故

$$P_{q-1}\nabla f(\bar{\boldsymbol{x}}) \neq 0.$$

其次,容易证明P_{q-1} 是投影矩阵, 且$\bar{p} = -P_{q-1}\nabla f(\bar{\boldsymbol{x}})$ 为下降方向. 最后证明$\bar{p}$ 为可行方向.
因

$$A_q^T\bar{p} = (A_{q-1}, a_i)^T\bar{p} = \begin{pmatrix} A_{q-1}^T\bar{p} \\ a_i^T\bar{p} \end{pmatrix},$$

由定理14 的证明过程知$A_{q-1}^T\bar{p} = 0$. 又因为P_{q-1} 正半定, $\lambda_i < 0$ 及式(33) 知

$$\begin{aligned} a_i^T\bar{p} &= -a_i^T P_{q-1}\nabla f(\bar{\boldsymbol{x}}) \\ &= -a_i^T P_{q-1}(A_{q-1}\bar{\lambda} + \lambda_i a_i) \\ &= -\lambda_i a_i^T P_{q-1} a_i \geq 0, \end{aligned}$$

所以, $\bar{p}$ 为可行方向.

所以如果可行点x_k 不是KKT 点, 则总可以有投影矩阵确定一个可行下降方向d_k, 再沿d_k 进行线性搜索, 使

$$f(x_k + \alpha_k d_k) = \min_{0 \le \alpha \le \alpha_{\max}} f(x_k + \alpha d_k),$$

其中, $\alpha_{\max}$ 为使得$x_k + \alpha d_k$ 可行的最大步长.

$\alpha_{\max}$ 的计算

记a_i 为矩阵A 的第i 列. 对于$i \notin \mathcal{I}_k$

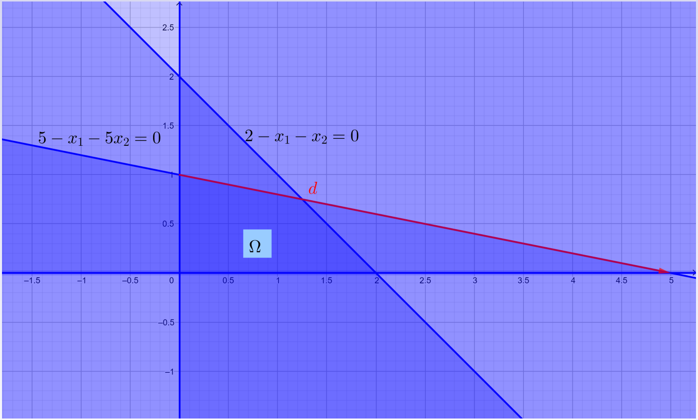
- 当$a_i^T d_k \geq 0$ 时, 对任意$\alpha > 0$, 总有$a_i^T(x_k + \alpha d_k) \geq b_i$;
- 当$a_i^T d_k < 0$ 时, 定义

$$\alpha_i = -\frac{a_i^T x_k - b_i}{a_i^T d_k} > 0,$$

  则$a_i^T(x_k + \alpha d_k) \geq b_i, \forall \alpha \in [0, \alpha_i]$.
- 令

$$\alpha_{\max} = \min_{i \notin \mathcal{I}} \left\{ -\frac{a_i^T x_k - b_i}{a_i^T d_k} \,\middle|\, a_i^T d_k < 0 \right\}, \tag{34}$$

  则当且仅当$\alpha \in [0, \alpha_{\max}]$ 时, $x_k + \alpha d_k$ 可行.


$5 - x_1 - 5x_2 = 0$
$2 - x_1 - x_2 = 0$
d
Ω
–1.5
–1
–0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
2.5
1.5
0.5
–0.5
–1

算法

投影梯度法

【步 1】 给定初始点x_0, 控制误差$\epsilon > 0$, 令$k = 0$;

【步 2】 令$\mathcal{I}_k = \{i \mid a_i^T x_k = b_i,\ i = 1,,2,\cdots,m\}$. 设$A_q$为$a_i(i \in \mathcal{I}_k)$ 为列且列满秩的矩阵.

(i) 若$\mathcal{I}_k = \emptyset$, 则令$P_q = I$;

(ii) 若$\mathcal{I}_k \neq \emptyset$, 则用(30) 计算投影矩阵$P_q$.

【步 3】 令$p_k = -P_q \nabla f_k$, 若$\|p_k\| \le \epsilon$, 则转步5; 否则转步4.

【步 4】 用(34) 计算$\alpha_{\max}$, 并用线性搜索确定步长$\alpha_k \in (0, \alpha_{\max}]$, 令$x_{k+1} = x_k + \alpha_k d_k$, 转步6.

【步 5】 若$\mathcal{I}_k = \emptyset$, 终止; 否则, 用(31) 计算$\lambda$. 若$\lambda \ge 0$, 则$x_k$ 是近似解,停; 否则, 令$\lambda_l = \min_i\{\lambda_i < 0\}$, 从$A_q$ 中去掉a_l 得A_{q-1}. 计算P_{q-1}, 令$p_k = -P_{q-1}\nabla f_k$, 转步4.

【步 6】 令$k = k+1$, 转步2.

例题

例5.2

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

取$x_0 = (0,0)^T$.

【解：】

$$
\nabla f(x) = \begin{pmatrix} 4x_1 - 2x_2 - 4 \\ 4x_2 - 2x_1 - 6 \end{pmatrix},
$$

第一次迭代：$\nabla f_0 = (-4,-6)^T$, 此时, $\mathcal{I}_0 = \{3,4\}$ 为有效集.

例题

$$A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, P_2 = I - A_2(A_2^T A_2)^{-1} A_2^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

则$P_2 \nabla f_0 = (0,0)^T$. 计算

$$\lambda = (A_2^T A_2)^{-1} A_2^T \nabla f_0 = (-4,-6)^T.$$

将$\lambda_2 = -6$ 从A_2 中去掉λ_2 所对应的第二列后得

$$A_1 = A_{2-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$P_1 = P_{2-1} = I - A_1(A_1^T A_1)^{-1} A_1^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$d_0 = -P_1 \nabla f_0 = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -4 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

例题

进行线性搜索：

$$\min_{0\le\alpha\le\alpha_{\max}} \quad f(x_0+\alpha d_0) = 72\alpha^2 - 36\alpha,$$

其中, $\alpha_{\max} = \min\left\{\frac{2}{6}, \frac{5}{30}\right\}$. 解得$\alpha_0 = \frac{1}{6}$.

$$x_1 = x_0 + \alpha_0 d_0 = (0,1)^T.$$

第二次迭代: $\nabla f_1 = (-6,-2)^T$, $\mathcal{I}_1 = \{2,3\}$.

$$A_2 = \begin{pmatrix} -1 & 1 \\ -5 & 0 \end{pmatrix},$$

$$P_2 = I - A_2(A_2^T A_2)^{-1} A_2^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$P_2 \nabla f_1 = (0,0)^T.$$

例题

计算

$$\lambda = (A_2^T A_2)^{-1} A_2^T \nabla f_0 = \left(\frac{2}{5}, -\frac{28}{5} \right)^T.$$

因$\lambda_2 = -\frac{28}{5} < 0$, 从$A_2$ 中去掉相应的列, 得

$$A_1 = A_{2-1} = \begin{pmatrix} -1 \\ -5 \end{pmatrix},$$

$$P_1 = P_{2-1} = I - A_1 (A_1^T A_1)^{-1} A_1^T = \begin{pmatrix} \frac{25}{26} & -\frac{5}{26} \\ -\frac{5}{26} & \frac{1}{26} \end{pmatrix},$$

$$d_1 = -P_1 \nabla f_1 = \begin{pmatrix} \frac{70}{13} \\ -\frac{14}{13} \end{pmatrix}.$$

不妨取$d_1 = (5, -1)^T$.

例题

沿d_1 进行线性搜索

$$\min_{0\le\alpha\le\alpha_{\max}} f(x_1+\alpha d_1) = 62\alpha^2 - 28\alpha - 4,$$

其中, $\alpha_{\max} = \min\left\{\frac{1}{4}, 1\right\} = \frac{1}{4}$. 解得$\alpha_1 = \frac{7}{31}$.

$$x_2 = x_1 + \alpha_1 d_1 = (35/31, 24/31)^T.$$

例题

第三次迭代:$\nabla f(x_2) = (-32/31, -160/31)^T$, $\mathcal{I}_3 = \{2\}$,

$$A_1 = \begin{pmatrix} -1 \\ -5 \end{pmatrix},$$

$$P_1 = I - A_1(A_1^T A_1)^{-1} A_1^T = \begin{pmatrix} \frac{25}{26} & -\frac{5}{26} \\ -\frac{5}{26} & \frac{1}{26} \end{pmatrix},$$

$$d_1 = -P_1 \nabla f_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

计算

$$\lambda = (A_1^T A_1)^{-1} A_1^T \nabla f_2 = \frac{32}{31} > 0.$$

从而$x_2 = (35/31, 24/31)^T$ 为KKT 点.

考虑仅含不等式约束的最优化问题

$$
\begin{aligned}
\min\ & f(x) \\
s.t.\ & Ax = b, \\
& x \geq 0,
\end{aligned} \tag{35}
$$

其中, $f(x)$ 为可微函数, A 为$m \times n$ 矩阵, b 为m 维向量($m \leq n$).

算法原理

- 将x 的分量分成两部分x^B, x^N, $x^B \in R^m$ 称为基向量, $x^N \in R^{n-m}$ 称为非基向量. 不妨设

$$x = \begin{pmatrix} x^B \\ x^N \end{pmatrix}.$$

相应地, $A = (B, N)$, B, N 分别为x^B, x^N 的分量对应的列.

算法原理

- 从而$Ax=b$ 分解为

$$Bx^B+Nx^N=b.$$

- 则

$$\begin{aligned} &x^B=B^{-1}b-B^{-1}Nx^N\geq 0, \\ &x^N\geq 0 \end{aligned} \tag{36}$$

若x 满足上式, 且$x^B>0$, 则称x 为非退化的可行解.

- 给定x^N, 用(36) 计算$x^B = x^B(x^N)$, 记

$$F(x^N) = f(x) = f(x^B, x^N) = f(x^B(x^N), x^N).$$

目标函数变为$F(x^N)$, 变量为x^N.

- 记$r(x^N) = \nabla_{x^N} F(x^N)$, 则有

$$r(x^N) = \nabla_N f(x^B(x^N), x^N) - (B^{-1}N)^T \nabla_B f(x^B(x^N), x^N). \tag{37}$$

- 定义搜索方向

$$(d_k^N)_j = \begin{cases} -(x_k^N)_j r_j(x_k^N), & r_j(x_k^N) > 0 \\ -r_j(x_k^N), & r_j(x_k^N) \le 0. \end{cases}$$

$$d_k^B = -B^{-1} N d_k^N,\ 令 d_k = \begin{pmatrix} d_k^B \\ d_k^N \end{pmatrix}.$$

- d_k 满足

$$d_k^T \nabla f_k < 0,\ (d_k)_j \ge 0, 若 (x_k)_j = 0.$$

即d_k 是可行下降方向.

- 计算最大步长$\alpha_{\max}$:

$$\alpha_{\max} = \min\left\{ -\frac{(x_k)_i}{(d_k)_i} \,\middle|\, (d_k)_i < 0, i \notin \mathcal{I}_k \right\}.$$

则当且仅当$0 \le \alpha \le \alpha_{\max}$ 时, $x_k + \alpha d_k$ 是可行点.

定理16

设问题(35) 中$x = \begin{pmatrix} x^B \\ x^N \end{pmatrix}$ 为非退化可行解. 又设$r(x^N)$ 由式(37) 定义. d^B, d^N 如前面所定义, $d = \begin{pmatrix} d^B \\ d^N \end{pmatrix}$. 则

(i) 当$d \neq 0$ 时, d 为可行下降方向;

(ii) $d = 0$ 当且仅当x 为KKT 点.

算法描述

简约梯度法-RG法

【步 1】 给定初始基可行解$x_0 = \begin{pmatrix} x_0^B \\ x_0^N \end{pmatrix} \geq 0$, 令$k = 0$,

【步 2】 将A 分解为$A = (B, N)$. 计算$r(x_k^N)$, d_k^N 和d_k^B, 令$d_k = \begin{pmatrix} d_k^B \\ d_k^N \end{pmatrix}$.

【步 3】 若$d_k = 0$, 则x_k 为KKT 点, 停; 否则, 计算$\alpha_{\max}$, 用线性搜索求$\alpha_k \in (0, \alpha_{\max}]$.

【步 4】 若$x_{k+1}^B > 0$, 则基向量不变, 令$k = k + 1$, 转步2; 若有某个j 使得$(x_{k+1}^B)_j = 0$, 将$(x_{k+1}^B)_j$ 换出基, 而已x_{k+1}^N 中具有最大分量的变量换入基, 构成新的基向量x_{k+1}^B 与非基向量x_{k+1}^N, 令$k = k + 1$, 转步2.

收敛性

定理17

设问题(35) 中$f(x)$ 连续可微. 若A 的任何m 列向量均线性无关且所有的基可行解都有m 个分量(即非退化的), 则由RG 法产生的点列$\{x_k\}$ 的任意聚点是KKT 点.

例题

例5.3

用简约梯度法求解

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

取$x_0 = (0,0)^T$.

例题

首先，引入松弛变量，化问题为标准式

$$
\begin{aligned}
\min\ & f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\
s.t.\ & x_1 + x_2 + x_3 - 2 = 0, \\
& x_1 + 5x_2 + x_4 - 5 = 0, \\
& x_i \geq 0,\ i = 1, 2, 3, 4
\end{aligned}
$$

此时

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{pmatrix},$$

而

$$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6, 0, 0)^T.$$

取初始可行点$x_0 = (0, 0, 2, 5)^T$.

例题

第一次迭代$k=0$.

$$x^B=(x_3,x_4)^T,\ x^N=(x_1,x_2)^T,$$

$$B=\begin{pmatrix}1&0\\0&1\end{pmatrix},\ N=\begin{pmatrix}1&1\\1&5\end{pmatrix},\ B^{-1}N=\begin{pmatrix}1&1\\1&5\end{pmatrix}.$$

$$\nabla_N f(x_0)=(-4,-6)^T,\ \nabla_B f(x_0)=(0,0)^T,$$

$$r(x_0^N)=\begin{pmatrix}-4\\-6\end{pmatrix}-\begin{pmatrix}1&1\\1&5\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}-4\\-6\end{pmatrix}.$$

例题

于是

$$d^N = \begin{pmatrix} 4 \\ 6 \end{pmatrix},\ d^B = -\begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -10 \\ -34 \end{pmatrix},$$

即

$$d_0 = (4, 6, -10, -34)^T,\ d_0 \neq 0.$$

例题

求$\alpha_{\max} = \min\left\{\frac{2}{10}, \frac{5}{34}\right\} = \frac{5}{34}$. 求解

$$\min\ f(x_0 + \alpha d_0) = 56\alpha^2 - 52\alpha,$$
$$s.t.\ 0 \le \alpha \le \frac{5}{34}.$$

得$\alpha_0 = 5/34$. 故

$$x_1 = x_0 + \alpha_0 d_0 = \left(\frac{10}{17}, \frac{15}{17}, \frac{9}{17}, 0\right)^T.$$

例题

第二次迭代, $k=1$,

$$x^B=(x_2,x_3)^T,\ x^N=(x_1,x_4)^T,$$

$$B=\begin{pmatrix}1&1\\5&0\end{pmatrix},\ N=\begin{pmatrix}1&0\\1&1\end{pmatrix},\ B^{-1}N=\begin{pmatrix}\frac{1}{5}&\frac{1}{5}\\\frac{4}{5}&-\frac{1}{5}\end{pmatrix}.$$

$$\nabla_N f(x_0)=\left(-\frac{58}{17},0\right)^T,\ \nabla_B f(x_0)=\left(-\frac{62}{17},0\right)^T,$$

$$r(x_0^N)=\begin{pmatrix}-\frac{58}{17}\\0\end{pmatrix}-\begin{pmatrix}\frac{1}{5}&\frac{1}{5}\\\frac{4}{5}&-\frac{1}{5}\end{pmatrix}\begin{pmatrix}-\frac{62}{17}\\0\end{pmatrix}=\begin{pmatrix}-\frac{228}{85}\\\frac{62}{85}\end{pmatrix}.$$

例题

于是

$$d^N=\begin{pmatrix}\frac{228}{85}\\ 0\end{pmatrix},\ d^B=-\begin{pmatrix}\frac{1}{5} & \frac{1}{5}\\ \frac{4}{5} & -\frac{1}{5}\end{pmatrix}\begin{pmatrix}\frac{228}{85}\\ 0\end{pmatrix}=\begin{pmatrix}-\frac{228}{425}\\ -\frac{912}{425}\end{pmatrix},$$

即

$$d_1=\left(\frac{228}{85},-\frac{228}{425},-\frac{912}{425},0\right)^T,\ d_1\neq 0.$$

例题

求$\alpha_{\max} = \min\left\{\frac{15}{17}/\frac{228}{425}, \frac{9}{17}/\frac{912}{425}\right\} = \frac{75}{304}$. 求解

$$\min\ f(x_1 + \alpha d_1) = \frac{3223008}{108625}\alpha^2 - \frac{51984}{7225}\alpha,$$
$$s.t.\ 0 \le \alpha \le \frac{75}{304}.$$

得$\alpha_1 = \frac{25}{124}$. 故

$$x_2 = x_1 + \alpha_1 d_1 = \left(\frac{35}{31}, \frac{24}{31}, \frac{3}{31}, 0\right)^T.$$

例题

第三次迭代, $k=2$,

$$x^B=(x_2,x_3)^T,\ x^N=(x_1,x_4)^T,$$

$$B=\begin{pmatrix}1 & 1\\ 5 & 0\end{pmatrix},\ N=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\ B^{-1}N=\begin{pmatrix}\frac{1}{5} & \frac{1}{5}\\ \frac{4}{5} & -\frac{1}{5}\end{pmatrix}.$$

$$\nabla_N f(x_2)=\left(-\frac{32}{31},0\right)^T,\ \nabla_B f(x_0)=\left(-\frac{160}{31},0\right)^T,$$

$$r(x_2^N)=\left(0,\frac{32}{31}\right)^T.$$

$$d^N=(0,0)^T,\ d^B=(0,0)^T.$$

即$d=(0,0,0,0)^T$.

由定理16 知, $x_2=\left(\frac{35}{31},\frac{24}{31},\frac{3}{31},0\right)^T$ 为KKT 点.

原问题的最优解为$x^*=\left(\frac{35}{31},\frac{24}{31}\right)^T$.

目标函数是二次函数,约束函数为线性函数的规划问题称为二次规划,其一般形式为

$$
(QP)\quad
\begin{aligned}
&\min\ f(x)=\frac{1}{2}x^TGx+c^Tx\\
&s.t.\ a_i^Tx-b_i=0, i\in E=\{1,\cdots,l\}\\
&\qquad a_i^Tx-b_i\geq 0, i\in I=\{l+1,\cdots,m\}
\end{aligned}
$$

其中G为n阶对称矩阵.当G正定时,(QP)为严格凸二次规划.显然，我们可以用投影梯度法和简约梯度法求解该问题。

最优性条件

定理18

点x^*是严格凸二次规划(QP)的严格整体最优解的充要条件是x^*满足KKT条件,即存在乘子向量$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)$,使得

$$
\begin{aligned}
&Gx^*+c-\sum_{i\in E}\lambda_i^*a_i-\sum_{i\in I}\lambda_i^*a_i=0\\
&a_i^Tx^*-b_i=0, i\in E\\
&a_i^Tx^*-b_i\geq 0, i\in I\\
&\lambda_i^*\geq 0,\quad i\in I\\
&\lambda_i^*=0,\quad i\in I\backslash I^*
\end{aligned}
$$

其中, I^*是x^*的有效集.

等式约束二次规划

问题

$$
\begin{aligned}
&\min\ f(x)=\frac{1}{2}x^TGx+c^Tx\\
&\text{s.t.}\quad a_i^Tx-b_i=0, i\in E=\{1,\cdots,l\}
\end{aligned}
$$

其中G 正定, $A=(a_1,\cdots,a_l)$ 的秩为l.

显然,x 是它的解的充要条件是:

$$Gx + c - \sum_{i \in E} \lambda_i a_i = 0$$
$$a_i^T x - b_i = 0, i \in E$$

令$b = (b_1, \cdots, b_l)^T$, $\lambda = (\lambda_1, \cdots, \lambda_l)^T$, 上式可以写成

$$\begin{pmatrix} G & -A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}$$

由于G正定, A 列满秩,该线性方程组有唯一解.

例题

例6.1

求解严格凸二次规划

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2 + x_3^2 \\ s.t. \quad & x_1 + 2x_2 - x_3 - 4 = 0 \\ & x_1 - x_2 + x_3 + 2 = 0 \end{aligned}$$

【解】(QP)的最优解及乘子满足

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

例题

该线性方程组有唯一解

$$\left(\frac{2}{7}, \frac{10}{7}, -\frac{6}{7}, \frac{8}{7}, -\frac{4}{7}\right)^T.$$

因此最优解

$$x^* = \left(\frac{2}{7}, \frac{10}{7}, -\frac{6}{7}\right)^T.$$

对应的Lagrange 乘子为

$$\lambda^* = \left(\frac{8}{7}, -\frac{4}{7}\right)^T.$$

一般约束问题

$$
\begin{aligned}
\min\ & f(x) = \frac{1}{2}x^T G x + c^T x \\
\text{s.t.}\ & a_i^T x - b_i = 0, i \in E = \{1, \cdots, l\} \\
& a_i^T x - b_i \geq 0, i \in I = \{l+1, \cdots, m\}
\end{aligned} \tag{38}
$$

最优性条件

定理19

设x^*是(38) 的最优解,且在x^* 处的有效集为I^*, 则x^* 也是以下问题的最优解

$$\begin{aligned} \min\ f(x) &= \frac{1}{2}x^TGx + c^Tx \\ s.t.\quad a_i^Tx - b_i &= 0, i \in E \cup I^*. \end{aligned} \tag{39}$$

反之, 若x^* 是(39) 的最优解, 且相应的$Lagrange$ 乘子λ^* 满足

$$\lambda_i^* \geq 0,\ i \in I^*.$$

则x^* 也是(38) 的解.

一般的,若已知可行解x_k ,有效集I_k,我们求解二次规划

$$\min\ f(x) = \frac{1}{2}x^T Gx + c^T x$$
$$\text{s.t.}\quad a_j^T x - b_i = 0, j \in E \cup I_k$$

令$x = x_k + d$,则它等价于求解

$$\begin{aligned} \min\ q(d) &= \frac{1}{2}d^T Gd + g_k^T d \\ \text{s.t.}\quad a_j^T d &= 0, j \in E \cup I_k \end{aligned} \tag{40}$$

其中, $g_k = Gx_k + c$.

定理20

若严格凸二次规划(40) 的最优解为$d_k = 0$，且$Lagrange$ 乘子$(\lambda_k)_s < 0$，$s \in I^*$，若

$$\begin{aligned} \min\ & q(d) = \frac{1}{2}d^T G d + g_k^T d \\ s.t.\ & a_j^T d = 0, j \in E \cup I_k \setminus \{s\} \end{aligned} \tag{41}$$

的解$p_k \neq 0$，则

$$g_k^T p_k < 0, a_s^T p_k > 0.$$

有效集算法

总结：给定了二次规划可行解x_k,其有效集为I_k,

(1) $d_k = 0$, 且Lagrange 乘子非负,则x_k 为最优解;

(2) $d_k = 0$, 且Lagrange 乘子有负分量,设为$(\lambda_k)_s < 0$ 则令

$$x_{k+1} = x_k, I_{k+1} = I_k \setminus \{s\}.$$

(3) $d_k \neq 0$,取合适的α_k 使得$x_k + a_k d_k$ 为原规划可行解,令$x_{k+1} = x_k + a_k d_k$, 并根据的取值情况判断是否改变有效集:

- 若$a_k = 1$,则有效集不变,$I_{k+1} = I_k$.
- 若$a_k < 1$,则在I_k 中添加指标得到新的有效集I_{k+1}.

有效集算法:计算α_k.

对于非有效约束$i \notin I_k$,

(i) 如果$a_i^T d_k \geq 0$, 则

$$a_i^T(x_k + \alpha d_k) \geq 0$$

肯定成立.

(ii) 只要考虑$a_i^T d_k < 0$, 取

$$\alpha_k = \bar{\alpha}_k = \min_{i \notin I_k} \left\{ \frac{b_i - a_i^T x_k}{a_i^T d_k} \,\middle|\, a_i^T d_k < 0 \right\} \tag{42}$$

(iii) 综合这两种情形,选择

$$\alpha_k = \min\{1, \bar{\alpha}_k\} \tag{43}$$

凸二次规划的有效集算法

Step 1 取初始点x_0, 确定相应的有效集I_0, 令$k=0$.

Step 2 求解(40), 得(d_k, λ_k) 若$d_k \neq 0$, 转Step 4; 否则, 转Step 3.

Step 3 求

$$(\lambda_k)_s = \min_{i \in I_k} (\lambda_k)_i.$$

若$(\lambda_k)_s \geq 0$, 停止. 否则， 令$I_k = I_k \setminus \{s\}$, 转Step 2.

Step 4 由式(43) 确定不成α_k. 令$x_{k+1} = x_k + \alpha_k d_k$.

Step 5 若$\alpha_k < 1$, 则令$I_{k+1} = I_k \cup \{t\}$, 其中t 由(42) 确定; 否则, 令$I_{k+1} = I_k$.

Step 6 令$k = k+1$, 转Step 2.

例题

例6.2

求解

$$\begin{aligned} \min\ & q(\boldsymbol{x}) = (x_1-1)^2 + (x_2-2.5)^2 \\ s.t.\ & x_1 - 2x_2 + 2 \geq 0 \\ & -x_1 - 2x_2 + 6 \geq 0 \\ & -x_1 + 2x_2 + 2 \geq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

【解：】取初始可行点$\boldsymbol{x}_0 = (2,0)^T$, 则有效集为$I_0 = \{3,5\}$. 求解下面的二次规划

$$
\begin{aligned}
\min \quad & \frac{1}{2}\boldsymbol{d}^T G\boldsymbol{d} + g_0^T\boldsymbol{d} \\
\text{s.t.} \quad & -d_1 + 2d_2 = 0 \\
& d_2 = 0
\end{aligned}
$$

其中

$$
g_0 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, G = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.
$$

解得$\boldsymbol{d}_0 = 0$. 接下来判断$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d}_0$ 是否是原问题的解. 因

$$g_1 = (2, -5)^T, a_3 = (-1, 2)^T, a_5 = (0, 1)^T,$$

令

$$g_1 = \lambda_3 a_3 + \lambda_5 a_5,$$

则有$\lambda_3 = -2, \lambda_5 = -1$.
去掉约束3,得有效集$I_1 = \{5\}$, 求解下面的二次规划

$$\begin{array}{ll} \min & \frac{1}{2}\boldsymbol{d}^T G \boldsymbol{d} + g_1^T \boldsymbol{d} \\ \text{s.t.} & d_2 = 0. \end{array}$$

即求解线性方程组

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 0 \end{bmatrix}.$$

解得

$$(d_1, d_2, \lambda)^T = (-1, 0, -5)^T.$$

因此$\boldsymbol{d}_1 = (-1, 0)^T$, $\boldsymbol{x}_1 + \boldsymbol{d}_1 = (1, 0)^T$.

由于$\boldsymbol{x}_1+\boldsymbol{d}_1=(1,0)$ 是可行的. 故而$\alpha_1=1$. $\boldsymbol{x}_2=\boldsymbol{x}_1+\alpha_1\boldsymbol{d}_1=(1,0)^T$ 且有效集不变, 即$I_2=I_1=\{5\}$.
点x_2 对应的新二次规划为

$$\begin{array}{ll}\min & g_2^T\boldsymbol{d}+\frac{1}{2}\boldsymbol{d}^TG\boldsymbol{d}\\ s.t. & d_2=0.\end{array}$$

其中, $g_2=(0,-5)^T$. 它等价于线性方程组

$$\begin{bmatrix}2 & 0 & 0\\ 0 & 2 & -1\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}d_1\\ d_2\\ \lambda\end{bmatrix}=\begin{bmatrix}0\\ 5\\ 0\end{bmatrix}.$$

解得$\boldsymbol{d}_2=(0,0)^T$, $\lambda=-5$.
取$\boldsymbol{x}_3=\boldsymbol{x}_2+\boldsymbol{d}_2=\boldsymbol{x}_2=(1,0)^T$,有效集$I_3=I_2\setminus\{5\}=\emptyset$.

点$\boldsymbol{x}_3$对应二次规划为(无约束)

$$\min \ g_3^T\boldsymbol{d}+\frac{1}{2}\boldsymbol{d}^TG\boldsymbol{d}.$$

解得$\boldsymbol{d}_3=(0,2.5)^T$.此时$\boldsymbol{x}_4=\boldsymbol{x}_3+\boldsymbol{d}_3=(1,2.5)^T$ 不可行. 按最大步长公式, α_3 为满足

$$\begin{aligned}1-5\alpha_3+2\geq 0\\ -1-5\alpha_3+6\geq 0\end{aligned}$$

最大实数, 则$\alpha_3=0.6$. 此时

$$\boldsymbol{x}_4=\boldsymbol{x}_3+\alpha_3\boldsymbol{d}_3=(1,1.5)^T, I_4=\{1\}, g_4=(0,-2)^T.$$

求解对应的二次规划

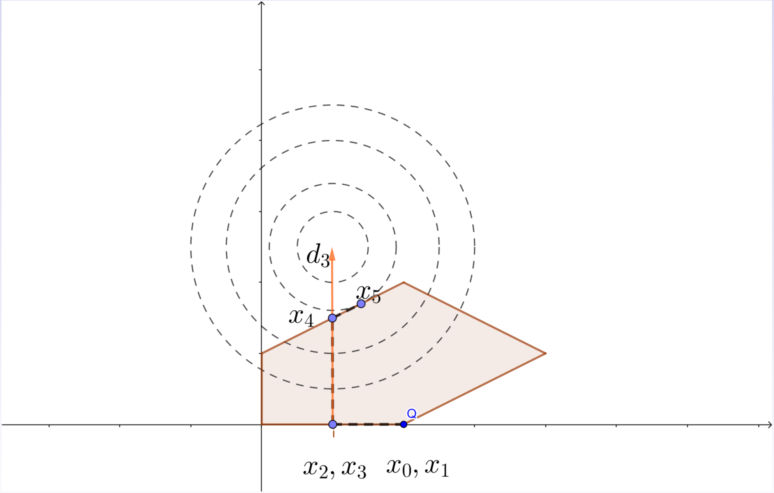
$$
\begin{array}{ll}
\min & \frac{1}{2}\boldsymbol{d}^T G\boldsymbol{d} + g_4^T\boldsymbol{d} \\
\text{s.t.} & d_1 - 2d_2 = 0
\end{array}
$$

解得$\boldsymbol{d}_4 = (0.4, 0.2)^T$. $\boldsymbol{x}_5 = \boldsymbol{x}_4 + \boldsymbol{d}_4 = (1.4, 1.7)^T$.(可行,不改变有效集), $g_5 = (0.8, -1.6)^T$.

求解对应的二次规划

$$\begin{array}{ll} \min & \frac{1}{2}\boldsymbol{d}^T G\boldsymbol{d} + g_5^T\boldsymbol{d} \\ \text{s.t.} & d_1 - 2d_2 = 0 \end{array}$$

其最优解为$\boldsymbol{d}_5 = (0,0)^T$. Lagrange 乘子向量为$\lambda = 0.8 > 0$,算法终止,最优解为$x^* = x_5 = (1.4, 1.7)^T$.


d_3
x_5
x_4
Q
x_2, x_3
x_0, x_1

quadprog 函数

Matlab 求解二次规划函数quadprog. 问题形式

$$\begin{aligned} \min\ & \frac{1}{2}x^T Hx + c^T x \\ \text{s.t.}\ & Ax \le b, \\ & Aeq \cdot x = beq, \\ & lb \le x \le ub. \end{aligned}$$

调用格式

x=quadprog(H,c,A,b, Aeq,beq ,lb,ub,x0,options);

或

[x,fval,exitflag,output,lambda]
=quadprog(H,c,A,b, Aeq,beq ,lb,ub,x0,options);

QP求解举例

例6.3

$$
\begin{aligned}
\min \quad & f(x) = (x_1 - 1)^2 + (x_2 - x_3)^2 + (x_4 - x_5)^2 \\
s.t. \quad & x_1 + x_2 + x_3 + x_4 + x_5 - 5 = 0 \\
& x_3 - 2(x_4 + x_5) + 3 = 0
\end{aligned}
$$

初始点$x_0 = (3, 5, -3, 2, -2)^T$.

【解：】此题中，

$$
H = \begin{pmatrix} 2 & & & & \\ & 2 & -2 & & \\ & -2 & 2 & & \\ & & & 2 & -2 \\ & & & -2 & 2 \end{pmatrix}, c = \begin{pmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
$$

$$Aeq = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 3 & -2 & -2 \end{pmatrix}, \quad beq = \begin{pmatrix} 5 \\ -3 \end{pmatrix}.$$

QP 算例

例6.4

$$
\begin{aligned}
\min \quad & f(x) = 9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 \\
& +2x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 \\
s.t. \quad & 3 - x_1 - x_2 - 2x_3 \geqslant 0 \\
& x_i \geqslant 0, i = 1, 2, 3
\end{aligned} \tag{44}
$$

初始点$x_0 = (0.5, 0.5, 0.5)^T$.

$$G = 2\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix},\ c = \begin{pmatrix} -8 \\ -6 \\ -4 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix},\ b = 3;$$

例6.5

$$
\begin{aligned}
\min \quad & f(x) = x_1 - x_2 - x_3 - x_1x_3 + x_1x_4 + x_2x_3 - x_2x_4 \\
s.t. \quad & 8 - x_1 - 2x_2 \geqslant 0, \\
& 12 - 4x_1 - x_2 \geqslant 0 \\
& 12 - 3x_1 - 4x_2 \geqslant 0, \\
& 8 - 2x_3 - x_4 \geqslant 0 \\
& 8 - x_3 - 2x_4 \geqslant 0, \\
& 5 - x_3 - x_4 \geqslant 0 \\
& x_i \geqslant 0, i = 1, \cdots, 4
\end{aligned}
$$

$$G=\begin{pmatrix}0 & 0 & -1 & 1\\ 0 & 0 & 1 & -1\\ -1 & 1 & 0 & 0\\ 1 & -1 & 0 & 0\end{pmatrix},\ c=\begin{pmatrix}1\\ -1\\ -1\\ 0\end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \ b = \begin{pmatrix} 8 \\ 12 \\ 12 \\ 8 \\ 8 \\ 5 \end{pmatrix};$$

说明：这部分教学内容的编排主要参考J. Nocedal 和S. J. Wright 的名著《Numerical Optimization》.

1. 等式约束问题

- 等式约束问题

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0, \end{aligned}$$

其中, $c(x) = (c_1(x), \cdots, c_l(x))$, 记

$$A(x) = (\nabla c_1(x), \cdots, \nabla c_l(x)).$$

- Lagrange 函数

$$L(x, \lambda) = f(x) - \lambda^T c(x).$$

一阶必要条件

设x^* 为最优解，$A(x^*)$ 列满秩，则

- 存在$\lambda^* = (\lambda_1^*, \cdots, \lambda_l^*)^T$，使得

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

即

$$\nabla f(x^*) - A(x^*)\lambda^* = 0,$$
$$c(x^*) = 0.$$

Newton 方程

用Newton 法求解

$$\begin{pmatrix} H_k & -A_k \\ A_k^T & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda^+ \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) \\ c(x_k) \end{pmatrix}, \tag{45}$$

其中, $H_k = \nabla_x^2 L(x_k, \lambda_k)$. 设该线性方程组的解为$(d_k, \lambda_k^+)$. 新迭代点$x_{k+1} = x_k + d_k$, Lagrange 乘子$\lambda_{k+1} = \lambda_k^+$.

这个方法称为Lagrange-Newton 法.

Lagrange-Newton 法

Lagrange-Newton 法

Step 1 选定初始点x_0 和乘子估计λ_0, 令$k=0$.

Step 2 计算∇f_k, ∇c_k 和$H_k=\nabla_x^2 L(x_k,\lambda_k)$.

Step 3 求解x_k 处的QP子问题(45) 得解(d_k,λ_k^+).

Step 4 令$x_{k+1}=x_k+d_k$, $\lambda_{k+1}=\lambda_k^+$.

Step 5 若终止条件成立, 则终止; 否则, 令$k=k+1$, 转Step 2.

定理21

设$H_k = \nabla^2_{xx} L(x, \lambda)$ 及以下假设条件成立

A1. 矩阵A_k 列满秩;

A2. $d^T H_k d > 0$, $\forall d \neq 0$ 且$A_k^T d = 0$. 即H_k 在A_k^T 的零空间上正定.

则矩阵

$$\begin{pmatrix} H_k & -A_k \\ A_k^T & 0 \end{pmatrix}$$

可逆.

定理22 (Lagrange-Newton 法的收敛性)

设初始点x_0 充分接近解x^*，且假设条件$A1$，$A2$ 成立. 在*Lagrange-Newton* 法收敛，且收敛速度是二次的.

证明略，可参考袁亚湘，孙文瑜《最优化理论与方法》第12章.

QP 子问题

定理23

设H_k, A_k 满足假设条件$A1$, $A2$. 则d 满足(45) 当且仅当d 为二次规划问题

$$\begin{aligned} \min \quad & q_k(d) = \nabla f_k^T d + \frac{1}{2} d^T H_k d \\ s.t. \quad & c_k + A_k^T d = 0 \end{aligned} \tag{46}$$

的极小点.

问题(46) 称为QP 子问题.

一般约束问题的QP子问题

- 问题

$$
\begin{aligned}
\min\quad & f(x) \\
\text{s.t.}\quad & c_i(x) = 0,\ \ i \in E = \{1, 2, \cdots, l\}, \\
& c_i(x) \ge 0,\ \ i \in I = \{l+1, \cdots, m\}.
\end{aligned}
\tag{47}
$$

- 相应的QP子问题

$$
\begin{aligned}
\min\quad & q_k(d) = \nabla f_k^T d + \frac{1}{2} d^T H_k d \\
\text{s.t.}\quad & c_i(x_k) + \nabla c_i(x_k)^T d = 0, i \in E, \\
& c_i(x_k) + \nabla c_i(x_k)^T d \ge 0, i \in I.
\end{aligned}
\tag{48}
$$

算法框架

SQP 算法模型

Step 1　选定初始点x_0, 初始正定矩阵H_0, 令$k=0$.

Step 2　求解x_k 处的QP子问题(48) 得解d_k.

Step 3　令$x_{k+1}=x_k+\alpha_k d_k$, 其中步长α_k 由某种线性搜索方法确定.

Step 4　计算H_{k+1}. 令$k=k+1$, 转Step 2.

修正的BFGS 校正(Powell)

考虑用拟Newton 法求B_k, 使得$B_k \approx H_k$.

(1) $s_k = x_{k+1} - x_k$, $y_k = \nabla_x L(x_{k+1}, \lambda_k) - \nabla_x L(x_k, \lambda_k)$.

(2) $z_k = \theta y_k + (1-\theta) B_k s_k$, $\theta \in [0, 1]$. 其中

$$\theta = \begin{cases} 1, & \text{当 } y_k^T s_k \geq 0.2 s_k^T B_k s_k \text{ 时}, \\ \dfrac{0.8 s_k^T B_k s_k}{s_k^T B_k s_k - y_k^T s_k}, & \text{其它}. \end{cases}$$

(3) 修正公式

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{z_k z_k^T}{z_k^T s_k}.$$

(4) 这样修正B_k 能保证其正定性.

贴士：拟Newton 法的对偶公式.

分别用B_k, H_k 作为Hesse 矩阵及其逆矩阵的近似, 则有

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}, \quad \text{(DFP公式)}$$

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) H_k \left(I - \frac{y_k s_k^T}{s_k^T y_k}\right) + \frac{s_k s_k^T}{s_k^T y_k}, \quad \text{(BFGS公式)}$$

和

$$B_{k+1} = \left(I - \frac{y_k s_k^T}{s_k^T y_k}\right) B_k \left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) + \frac{y_k y_k^T}{s_k^T y_k}, \quad \text{(DFP公式)}$$

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \quad \text{(BFGS 公式)}$$

效益函数

- **效益函数：** 一种既包含目标函数信息又包含约束条件信息在内的函数. 用作线性搜索的辅助函数.
- ℓ_1 **精确罚函数**

$$W(x,\mu) = f(x) + \sigma\Phi(c(x)),$$

其中, $\Phi(c(x)) = \sum_{i\in E} |c_i(x)| + \sum_{i\in I} \max\{0, -c_i(x)\}$.

- 在线性搜索中，减小W 值下降, 相当于兼顾了$f(x)$ 的下降和违反约束的程度的下降. 两者的轻重以罚因子σ 加以调节.
- ℓ_1 精确罚函数与增广Lagrange 函数有一个类似的性质：罚因子σ 不需要趋向与正无穷大. （这也是“精确”二字的含义）

- 线性搜索：利用第一章介绍的非精确线性搜索方法使得

$$W(x_k,\sigma) - W(x_k+\alpha_k d_k,\sigma) \geq -\rho\alpha_k DW(x_k,\sigma;d_k). \tag{49}$$

其中，$\rho \in (0,1)$,

$$DW(x_k,\sigma;d_k) = \nabla f_k^T d_k - \sigma\Phi(c_k).$$

实用的SQP 算法

线性搜索SQP 算法

Step 1 选定初始点x_0, 初始乘子估计λ_0 初始正定矩阵B_0. 给定控制误差$\epsilon > 0$, 令$k = 0$.

Step 2 求解QP 子问题(48)，得d_k 和λ_k^+. 令$\Delta\lambda_k = \lambda_k^+ - \lambda_k$. 若$\|d_k\| \leq \epsilon$, 则算法终止.

Step 3 适当调整罚因子σ, 使得d_k 为$W(x_k, \sigma)$ 的下降方向.

Step 4 令$x_k^+ = x_k + \alpha d_k$, 计算$W(x_k^+, \sigma)$.

Step 5 若(49) 成立, 则令$\alpha_k = \alpha$ 且

$$x_{k+1} = x_k^+, \ \lambda_{k+1} = \lambda_k + \alpha_k \Delta\lambda_k, \ s_k = x_{k+1} - x_k,$$
$$y_k = \nabla_x L(x_{k+1}, \lambda_{k+1}) - \nabla_x L(x_k, \lambda_{k+1}).$$

用修正的BFGS 校正得B_{k+1}. 转Step 2; 否则, 令$\alpha \leftarrow \alpha/2$, 转Step 4.

讨论

SQP 算法中效益函数和罚函数法（内罚函数、外罚函数、增广Lagrange 函数）在算法中的作用是否相同？

d 的下降性

定理24 (d 的下降性)

设d_k 是QP 子问题的KKT 点，λ_k 是相应的$Lagrange$ 乘子，则对ℓ_1 的效益函数$W(x,\sigma)$，有

$$DW(x_k,\mu;d_k) \le -d_k^T B_k d_k - (\sigma - \|\lambda_k\|_\infty)\Phi(c_k).$$

其中,$DW(x_k,\mu;d_k) = \nabla f_k^T d_k - \sigma\Phi(c_k)$. 显然，当$\sigma > \|\lambda_k\|_\infty$ 时，d_k 是$W(x,\mu)$ 的下降方向.

罚因子的选取

- 适当调整罚因子σ, 使得$\sigma > \|\lambda_k\|_\infty$, 从而确保$d_k$ 为$W(x_k,\sigma)$的下降方向.

  实际上, 只需满足$\sigma \geq \frac{\nabla f_k^T d_k}{(1-\rho)\Phi(c_k)}$ 即可，其中

$$\Phi(c_k) = \sum_{i\in E} |c_i(x)| + \sum_{i\in I} \max\{0, -c_i(x)\}.$$

  一个比较简单的方法是令

$$\sigma \leftarrow \begin{cases} \sigma, & \text{若}\nabla f_k^T d_k < \sigma\Phi(c_k), \\ \max\left\{c\sigma, \frac{\nabla f_k^T d_k}{(1-\rho)\Phi(c_k)}\right\}, & \text{其它}, \end{cases}$$

  其中, $c > 1$ 为罚因子放大倍数.

收敛性:假设条件

假设条件1

(A1) $f(x)$ 与$c_i(x)$, $i=1,2,\cdots,m$ 为连续可微函数;

(A2) 存在α, $\beta>0$ 使得对每个k 与任何$x\in R^n$, 有

$$\alpha\|x\|^2\le x^TB_kx\le\beta\|x\|^2;$$

(A3) 对每个k，二次规划(48) 存在一个KKT 点, 且满足

$$\sigma\ge|\lambda_{(k)}|_\infty+\mu,$$

其中, $\mu>0$ 为常数.

收敛性定理

定理25

设假设条件*(A1)-(A3)*成立，则由算法*4.4.2*产生的点列$\{x_k\}$，或者在(47)的一个KKT点处终止，或者任何使

$$s(\bar{x}) = \left\{ d \,\middle|\, \begin{array}{l} c_i(\bar{x}) + \nabla c_i(\bar{x})^T d = 0,\ i \in E, \\ c_i(\bar{x}) + \nabla c_i(\bar{x})^T d > 0,\ i \in I \end{array} \right\}$$

非空的聚点$\bar{x}$为(47)的一个KKT点.

算例

例6.6

用SQP算法求解

$$
\begin{aligned}
&\min\ (x_1-2)^4+(x_1-2x_2)^2\\
&s.t.\ x_1^2-x_2=0.
\end{aligned}
$$

算例

例6.7

$$
\begin{aligned}
\min\ f &= (x_1-1)^2+(x_2-1)^2 \\
s.t.\ & -0.25x_1^2-x_2^2+1\geq 0, \\
& x_1-2x_2+1=0.
\end{aligned}
$$

分别编写程序和利用*fmincon* 函数求解.