最优化理论与方法复习

基础与前沿研究院

谭兵

2018 年 12 月 28 日

1　最优化问题与数学基础

1.1　数学基础

梯度：$L=\left(\frac{\partial f\left(\boldsymbol{X}^{0}\right)}{\partial x_{1}}, \frac{\partial f\left(\boldsymbol{X}^{0}\right)}{\partial x_{2}}, \cdots, \frac{\partial f\left(\boldsymbol{X}^{0}\right)}{\partial x_{n}}\right)^{T}$

梯度方向是函数值上升的方向，负梯度方向是函数值下降的方向，沿梯度方向函数具有最快的变化率。

常见梯度

- $\nabla\left(\boldsymbol{b}^{T} \boldsymbol{X}\right)=\boldsymbol{b}, \boldsymbol{b}=\left(b_{1}, b_{2}, \cdots, b_{n}\right)^{T} \in \boldsymbol{R}^{n}$
- $\nabla\left(\boldsymbol{X}^{T} \boldsymbol{A} \boldsymbol{X}\right)=\boldsymbol{A} \boldsymbol{X}+\boldsymbol{A}^{T} \boldsymbol{X}$

【注】：梯度法则：如果 $f(x): R^{m} \rightarrow R^{p}$，则梯度 $\nabla f(x) \in R^{m \times p}$.

Example 1. $f(x)=\|x-a\|_{2}, a \in R^{2}$，那么 $\nabla f(x)=\frac{x-a}{\|x-a\|_{2}}$.

Definition 1 (**方向导数**). 如果函数 $z=f(x,y)$ 在点 $P(x,y)$ 是可微分的，那么函数在该点沿任一方向的方向导数都存在，且有

$$\frac{\partial f}{\partial l}=\frac{\partial f}{\partial x} \cos \varphi+\frac{\partial f}{\partial y} \sin \varphi$$

其中 φ 为 x 轴到方向 l 的转角。

Example 2. 求函数 $z=x e^{2y}$ 在点 $P(1,0)$ 处沿从点 $P(1,0)$ 到点 $Q(2,-1)$ 方向的方向导数.

Solution. 解：这里方向 l 即为向量 $\overrightarrow{PQ}=\{1,-1\}$，因此 x 轴到方向 l 的转角 $\varphi=-\frac{\pi}{4}$，因为

$$\frac{\partial z}{\partial x}=e^{2y}, \quad \frac{\partial z}{\partial y}=2 x e^{2y}$$

所求方向导数

$$\frac{\partial z}{\partial l}=1 \cdot \cos \left(-\frac{\pi}{4}\right)+2 \cdot \sin \left(-\frac{\pi}{4}\right)=-\frac{\sqrt{2}}{2}$$

【注】：f 在方向 P 处的方向导数：$\frac{\partial f}{\partial P}=\frac{(\nabla f)^{T} P}{\|P\|}$.

Definition 2 (**Hesse 矩阵**).

$$\nabla^2 f(\boldsymbol{X})=\begin{pmatrix}\frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_1} & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_2 \partial x_1}\\ \frac{\partial^2 f(\boldsymbol{X})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_2}\end{pmatrix}$$

【注】: 设 $\phi(t)=f\left(\boldsymbol{X}^0+t\boldsymbol{P}\right)$，则

$$\phi'(t)=\nabla f\left(\boldsymbol{X}^0+t\boldsymbol{P}\right)^T\boldsymbol{P},\quad \phi''(t)=\boldsymbol{P}^T\nabla^2 f\left(\boldsymbol{X}^0+t\boldsymbol{P}\right)\boldsymbol{P}$$

Theorem 1 (**多元函数的 Taylor 展开**).

$$f(\boldsymbol{X})=f\left(\boldsymbol{X}^0\right)+\nabla f\left(\boldsymbol{X}^0\right)^T\left(\boldsymbol{X}-\boldsymbol{X}^0\right)+\frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{X}^0\right)^T\nabla^2 f(\boldsymbol{X})\left(\boldsymbol{X}-\boldsymbol{X}^0\right)+o\left(\left\|\boldsymbol{X}-\boldsymbol{X}^0\right\|^2\right)$$

最优性条件

- **（一阶必要条件）**：若 $\boldsymbol{X}^\star$ 是 $f(\boldsymbol{X})$ 的局部极小点，则 $\nabla f\left(\boldsymbol{X}^\star\right)=0$
- **（二阶必要条件）**：若 $\boldsymbol{X}^\star$ 是 $f(\boldsymbol{X})$ 的局部极小点，则 $\nabla^2 f\left(\mathbf{X}^\star\right)$ 半正定。
- **（二阶充分条件）**：若 $\nabla f\left(\boldsymbol{X}^\star\right)=0$，且 $\nabla^2 f\left(\mathbf{X}^\star\right)$ 是正定矩阵，则 $\boldsymbol{X}^\star$ 是 $f(\boldsymbol{X})$ 的严格局部极小点。

Example 3. 证明，若 $\boldsymbol{X}^\star$ 是 $f(\boldsymbol{X})$ 的局部极小点，则 $\nabla f\left(\boldsymbol{X}^\star\right)=0$

证明. 令 $\boldsymbol{X}=\boldsymbol{X}^\star-\alpha\nabla f\left(\boldsymbol{X}^\star\right)$

利用多元函数的一阶 Taylor 展开

$$\begin{aligned}f(\boldsymbol{X})&=f(\boldsymbol{X}^\star)-\alpha\nabla f\left(\boldsymbol{X}^\star\right)^T\nabla f\left(\boldsymbol{X}^\star\right)\\&=f(\boldsymbol{X}^\star)-\alpha\|\nabla f\left(\boldsymbol{X}^\star\right)\|_2^2\end{aligned}$$

由于 $\boldsymbol{X}^\star$ 是局部极小值，那么

$$-\alpha\|\nabla f\left(\boldsymbol{X}^\star\right)\|_2^2=f(\boldsymbol{X})-f(\boldsymbol{X}^\star)\geq 0$$

因此，$\|\nabla f\left(\boldsymbol{X}^\star\right)\|_2^2=0$，即 $\nabla f\left(\boldsymbol{X}^\star\right)=0$ □

Example 4. 证明，若 x^* 是凸优化中的局部极小点，则 x^* 是全局极小点。

证明.（反证法） 若 x^* 不是全局极小点，则存在 $\bar{x}$ 是全局极小点。使得 $f(\bar{x})\leq f(x^*)$.

令 $z=\alpha\bar{x}+(1-\alpha)x^*$，$\alpha\in(0,1)$

所以，

$$\begin{aligned}f(z)&=f(\alpha\bar{x}+(1-\alpha)x^*)\\&\leq\alpha f(\bar{x})+(1-\alpha)f(x^*)\\&\leq\alpha f(x^*)+(1-\alpha)f(x^*)\\&\leq f(x^*)\end{aligned}$$

当 $\alpha\to 0$，$z=\alpha\bar{x}+(1-\alpha)x^*\in N(x^*)$，其中 $N(x^*)$ 表示 x^* 的某个邻域。

即在 x^* 的某个局部邻域 $N(x^*)$ 中有 z，使得 $f(z)\leq f(x^*)$，这与 x^* 是局部极小点矛盾。

因此，x^* 是全局极小点。 □

1.2 凸集

Definition 3 (凸集). 若对任意的 $X^1, X^2 \in D$ 以及 $\alpha \in [0,1]$，都有 $\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2 \in \boldsymbol{D}$，则称 D 为凸集。

【注】: 几何直观上，若集合 D 中任意两点的连线仍在 D 中，则称 D 为凸集。

Example 5. $D = \{X | AX \le b, X \in R^n, A \in R^{m\times n}\}$ 是凸集。

Example 6. 若 A, B 是 $\boldsymbol{R}^n$ 中的凸集，则 $A \cap B, A+B, A-B$ 也是凸集，但 $A \cup B$ 一般不是凸集。

Example 7. $C = \left\{x \in R_+^2 | x_1 x_2 \ge 1\right\}$，则 C 是凸集。

1.3 凸函数

Definition 4 (凸函数). 设 $f: \boldsymbol{D} \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$\boldsymbol{D}$ 是凸集，若对任意的 $\boldsymbol{X}^1, \boldsymbol{X}^2 \in D$，以及 $\alpha \in (0,1)$，都有

$$f\left(\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2\right) \le \alpha f\left(\boldsymbol{X}^1\right) + (1-\alpha) f\left(\boldsymbol{X}^2\right)$$

则称 $f(\boldsymbol{X})$ 为 $\boldsymbol{D}$ 上的凸函数。

【注】: 从几何上看，曲线上任意两点的连线在相应弧段的上方，即弦在弧之上。

Example 8. 判断 $f(x) = \frac{1}{x^2}$ 是否为凸函数，答案，否，因为其定义域 $D = R - \{0\}$ 不是凸集。

【注】: 判断是否为凸函数，首先应判断定义域是否为凸集。

Example 9. $f(x_1, x_2) = \frac{x_1^2}{x_2}$，$x_2 > 0$，根据 *Hesse* 矩阵的各阶顺序主子式可以判断，它是一个凸函数。

Example 10. $f(x) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$ 是凸函数。

Example 11. $f(x) = e^{x^T A x}$，其中 A 是正定矩阵，它是凸函数。

【注】: $f: R^n \to R^1$ 凸（凹），$g: R^1 \to R^1$ 凸增（凸减），则 $h(x) = g(f(x))$ 是凸函数。

Theorem 2. $f(\boldsymbol{X})$ 在 D 上是凸函数的充要条件是对任意的 $\boldsymbol{X}^1, \boldsymbol{X}^2 \in D$，都有

$$f\left(\boldsymbol{X}^2\right) \ge f\left(\boldsymbol{X}^1\right) + \nabla f\left(\boldsymbol{X}^1\right)^T \left(\boldsymbol{X}^2 - \boldsymbol{X}^1\right)$$

Theorem 3. $f(\boldsymbol{X})$ 在 D 上是凸函数的充要条件是 $\nabla^2 f(\boldsymbol{X})$ 是半正定矩阵。

【注】: 二阶矩阵判断是否是正定矩阵，看顺序主子式均 ≥ 0。

1.4 凸规划

$$
\begin{aligned}
&\min f(\boldsymbol{X}) \\
&s.t. \\
&g_i(\boldsymbol{X}) \geq 0 \quad (i=1,2,\cdots,m) \\
&h_j(\boldsymbol{X}) = 0 \quad (j=1,2,\cdots,p)
\end{aligned}
$$

若 $f(\boldsymbol{X})$ 与 $-g_i(\boldsymbol{X})$ 都是凸函数，$h_j(\boldsymbol{X})$ 是线性函数，则称其为凸规划。

Theorem 4. $\boldsymbol{X}^\star$ 为上面问题的最优解的充要条件是，$\forall \boldsymbol{X} \in \boldsymbol{S}$，有

$$
(\boldsymbol{X}-\boldsymbol{X}^\star)^T \nabla f(\boldsymbol{X}^\star) \geq 0
$$

证明. 充分性：

$$
f(\boldsymbol{X}) \geq f(\boldsymbol{X}^\star) + \nabla f(\boldsymbol{X}^\star)^T (\boldsymbol{X}-\boldsymbol{X}^\star) \geq f(\boldsymbol{X}^\star)
$$

必要性：

若存在 $\boldsymbol{X}^0 \in \boldsymbol{S}$，使得 $\left(\boldsymbol{X}^0-\boldsymbol{X}^\star\right)^T \nabla f(\boldsymbol{X}^\star) < 0$，则

$$
\begin{aligned}
& f\left(\boldsymbol{X}^\star + \alpha\left(\boldsymbol{X}^0-\boldsymbol{X}^\star\right)\right) - f(\boldsymbol{X}^\star) \\
&= \alpha \nabla f(\boldsymbol{X}^\star)^T \left(\boldsymbol{X}^{\mathbf{0}}-\boldsymbol{X}^\star\right) + o\left(\alpha\left\|\boldsymbol{X}^0-\boldsymbol{X}^\star\right\|\right) \\
&= \alpha\left(\left(\boldsymbol{X}^0-\boldsymbol{X}^\star\right)^T \nabla f(\boldsymbol{X}^\star) + \frac{o\left(\alpha\left\|\boldsymbol{X}^0-\boldsymbol{X}^\star\right\|\right)}{\alpha}\right) \\
&\leq 0
\end{aligned}
$$

这与 $\boldsymbol{X}^\star$ 为最优解矛盾。 □

2 线性规划与单纯形法

2.1 基本可行解

基本解：令一些分量为 0，解出其他分量；

基本可行解：既是基本解，又是可行解。

Example 12. 求

$$
\begin{aligned}
&x_1+x_2+x_3=2 \\
&x_1+2x_2+4x_3=6 \\
&x_1,x_2,x_3 \geq 0
\end{aligned}
$$

的基本可行解

Solution.

$$
\boldsymbol{X}^1=(0,1,1)^T, \quad \boldsymbol{X}^2=\left(\frac{2}{3},0,\frac{4}{3}\right)^T, \boldsymbol{X}^3=(-2,4,0)^T(\text{舍去，不是可行解})
$$

【注】：线性规划的顶点集和基本可行解集等价。

2.2 线性规划标准型

线性规划标准型的矩阵形式：

$$
\begin{aligned}
&\min f(\boldsymbol{X})=\boldsymbol{C}^T\boldsymbol{X}\\
&s.t.\\
&\boldsymbol{AX}=b\\
&\boldsymbol{X}\geq 0
\end{aligned}
$$

$$
\begin{aligned}
\sum_{j=1}^{n} a_{pj}x_j + x_{n+p} = b_p \quad (x_{n+p}\geq 0:\text{松弛变量})\\
\sum_{j=1}^{n} a_{qj}x_j - x_{n+q} = b_q \quad (x_{n+q}\geq 0:\text{剩余变量})\\
\exists x_i\in\boldsymbol{R}, x_i = x_i^+ - x_i^- \quad (x_i^+, x_i^-\geq 0:\text{自由变量})
\end{aligned}
$$

Example 13. 将下面的规划化成标准型

$$
\begin{gathered}
\max f(\boldsymbol{X}) = x_1 - 2x_2 + 3x_3\\
s.t.\\
2x_1 - 7x_3 \leq 0\\
3x_1 + 2x_2 \geq 0\\
x_1 + x_2 + x_3 = 5\\
x_1, x_2 \geq 0, x_3 \in R
\end{gathered}
$$

Solution. 标准型为

$$
\begin{gathered}
\min f(\mathbf{X}) = -x_1 + 2x_2 - 3(x_3^+ - x_3^-)\\
s.t.\\
2x_1 - 7(x_3^+ - x_3^-) + x_4 = 0\\
3x_1 + 2x_2 - x_5 = 0\\
x_1 + x_2 + (x_3^+ - x_3^-) = 5\\
x_i \geq 0 (i = 1, 2, 4, 5), x_3^+ > 0, x_3^- > 0
\end{gathered}
$$

2.3 使用表格形式的单纯形法

最小值问题（最大化问题）的表格型单纯形法求解步骤：

1. 最后一行：$z_k - c_k = \max\{z_j - c_j\}$（$z_k - c_k = \min\{z_j - c_j\}$）
2. 选择对应的列作为主列
$$\frac{\bar{b}_r}{y_{ik}} = \min\left\{\frac{\bar{b}_i}{y_{ik}} | y_{ik} > 0\right\}$$
3. 选取交叉元素作为主元。
4. 直到所有判别数均 ≤ 0 时停止（直到所有判别数均 ≥ 0 时停止）

Example 14. 用单纯形法求解下列线性规划问题

$$\begin{aligned}
&\min f(\boldsymbol{X}) = x_2 - 3x_3 + 2x_5 \\
&s.t. \\
&x_1 + 3x_2 - x_3 + 2x_5 = 7 \\
&-2x_2 + 4x_3 + x_4 = 12 \\
&-4x_2 + 3x_3 + 8x_5 + x_6 = 10 \\
&x_j \geq 0, \quad j = 1, \cdots, 6
\end{aligned}$$

Solution. 迭代过程如下

P_1	P_2	P_3	P_4	P_5	P_6	b
1	3	-1	0	2	0	7
0	-2	$\boxed{4}$	1	0	0	12
0	-4	3	0	8	1	10
0	-1	3	0	-2	0	0

P_1	P_2	P_3	P_4	P_5	P_6	b
1	$\boxed{\frac{5}{2}}$	0	$\frac{1}{4}$	2	0	10
0	$-\frac{1}{2}$	1	$\frac{1}{4}$	0	0	3
0	$-\frac{5}{2}$	0	$-\frac{3}{4}$	8	1	1
0	$\frac{1}{2}$	0	$-\frac{3}{4}$	-2	0	-9

P_1	P_2	P_3	P_4	P_5	P_6	b
$\boxed{\frac{2}{5}}$	1	0	$\frac{1}{10}$	$\frac{4}{5}$	0	4
$\frac{1}{5}$	0	1	$\frac{3}{10}$	$\frac{2}{5}$	0	5
1	0	0	$-\frac{1}{2}$	10	1	11
$-\frac{1}{5}$	0	0	$-\frac{4}{5}$	$-\frac{12}{5}$	0	-11

至此，所有判别数小于或等于零，故当前可行解为 $\boldsymbol{X}^* = (0,4,5,0,0,11)^T$，最优值为 $f(\boldsymbol{X}^*) = -11$.

2.4 两阶段法

基本思想：转换为标准形式后，问题中不包含需要的 n 阶单位矩阵，因此需要引进人工变量，用单纯形法求解。

Example 15.

$$\begin{aligned}
&\min \quad x_1 - x_2 \\
&s.t. \\
&-x_1 + 2x_2 + x_3 \leq 2 \\
&-4x_1 + 4x_2 - x_3 = 4 \\
&x_1 - x_3 = 0 \\
&x_i \geq 0, i = 1 \sim 3
\end{aligned}$$

Solution. 引进人工变量 x_5, x_6，构造辅助线性规划

$$\begin{aligned}&\min \quad x_5+x_6\\&s.t.\\&-x_1+2x_2+x_3+x_4=2\\&-4x_1+4x_2-x_3+x_5=4\\&x_1-x_3+x_6=0\\&x_i\geq 0, i=1\sim 6\end{aligned}$$

第一阶段

用单纯形法求解过程如下，极小化目标函数 x_5+x_6

P_1	P_2	P_3	P_4	d_5	d_6	b
-1	$\boxed{2}$	1	1	0	0	2
-4	4	-1	0	1	0	4
1	0	-1	0	0	1	0
-3	4	-2	0	0	0	4
-1/2	1	1/2	1/2	0	0	1
-2	0	-3	-2	1	0	0
$\boxed{1}$	0	-1	0	0	1	0
-1	0	-4	-2	0	0	0

因为人工变量 x_5, x_6 是基变量，所以应该替换出（必须要使人工变量为非基变量）

P_1	P_2	P_3	P_4	d_5	d_6	b
0	1	0	1/2	0	1/2	1
0	0	$\boxed{-5}$	-2	1	2	0
1	0	-1	0	0	1	0
0	1	0	1/2	0	1/2	1
0	0	1	2/5	-1/5	-2/5	0
1	0	0	2/5	-1/5	3/5	0

第一阶段结束后，修改最后的单纯形表，去掉人工变量所在的列，**把最后的判别数行按照原来问题进行修正，其他不变。**

第二阶段

极小化目标函数 x_1-x_2

P_1	P_2	P_3	P_4	b
0	1	0	1/2	1
0	0	1	2/5	0
1	0	0	2/5	0
0	0	0	-1/10	-1

所以 $X^\star=(0,1,0)^T, \quad f(X^\star)=-1.$

【注】: 两阶段法的标准形非常重要，需要满足右端项 $b\geq 0$，且第一阶段的目标函数是加入的人工变量之和。第一阶段中，基变量对应的判别数都为 0.

3 对偶线性规划

3.1 对称形式

原规划

$$(P)\qquad \begin{aligned} &\min \quad C^T\boldsymbol{X} \\ &s.t. \\ &\boldsymbol{AX} \geq b \\ &\boldsymbol{X} \geq 0 \end{aligned}$$

对偶规划

$$(D)\qquad \begin{aligned} &\max \quad b^T\boldsymbol{W} \\ &\boldsymbol{A}^T\boldsymbol{W} \leq \boldsymbol{C} \\ &\boldsymbol{W} \geq 0 \end{aligned}$$

【注:】经验：对偶规划的系数好确定，无非就是需要确定变量和约束符号。口诀：变量符号 $\rightarrow$ 看约束条件符号：小同大反；约束条件符号 $\rightarrow$ 看变量符号：大同小反。

原（对偶）线性规划变换规则表

原规划（对偶规划）	对偶规划（原规划）
min	max
目标中系数	约束条件右端项
约束条件右端项	目标中系数
约束条件 $\geq$	变量 $\geq$
约束条件 $\leq$	变量 $\leq$
约束条件 $=$	变量无约束
变量 $\geq$	约束条件 $\leq$
变量 $\leq$	约束条件 $\geq$
变量无约束	约束条件 $=$

Example 16. 写出对偶线性规划

$$\begin{aligned} &\min \quad f(\boldsymbol{X}) = 8x_1 + 6x_2 + 3x_3 + 6x_4 \\ &s.t. \\ &x_1 + 2x_2 + x_4 = 3 \\ &3x_1 + x_2 + x_3 + x_4 = 6 \\ &x_j \geq 0, j = 1, \cdots, 4 \end{aligned}$$

Solution.

$$\begin{aligned}
&\max \quad g(\boldsymbol{W}) = 3w_1 + 6w_2 + 2w_3 \\
&s.t. \\
&w_1 + 3w_2 \leq 8 \\
&2w_1 + w_2 \leq 6 \\
&w_2 + w_3 \geq 3 \\
&w_1 + w_2 + w_3 = 6 \\
&w_1 \geq 0, w_2 \leq 0, w_3 \in R
\end{aligned}$$

3.2 一般问题的原始对偶理论

原始问题：

$$\begin{aligned}
&\min \quad f(x) \\
&s.t. \\
&h_i(x) = 0,\ i = 1, 2, \cdots, m \\
&g_j(x) \leq 0,\ i = 1, 2, \cdots, p
\end{aligned}$$

Lagrange 函数

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x)$$

需要满足以下两个原则

$$L(x, \lambda, \mu) \leq f(x), \quad \mu \geq 0.$$

关于 x 极小化 $L(x, \lambda, \mu)$

$$q(\lambda, \mu) = \min_x L(x, \lambda, \mu)$$

对偶问题：

$$\begin{cases}
\max \quad q(\lambda, \mu) \\
s.t. \\
\mu \geq 0
\end{cases}$$

3.3 互补松弛性质

3.3.1 对称形式的对偶

对称形式的对偶定义如下：

原问题

$$\begin{aligned}
\min \quad & cx \\
s.t. \quad & Ax \geqslant b \\
& x \geqslant 0
\end{aligned}$$

对偶问题

$$\begin{aligned}
\max \quad & wb \\
s.t. \quad & wA \leqslant 0 \\
& w \geqslant 0
\end{aligned}$$

Theorem 5. 设 $x^{(0)}$ 和 $w^{(0)}$ 分别是上面原问题和对偶问题的可行解，那么 $x^{(0)}$ 和 $w^{(0)}$ 都是最优解的充要条件是，对所有 i 和 j 成立

- 如果 $x_j^{(0)} > 0$，就有 $w^{(0)}p_j = c_j$（第 j 个约束等式成立）
- 如果 $A_i x^{(0)} > b_i$，就有 $w_i^{(0)} = 0$（如果原问题最优解使得第 i 个约束不等式严格成立，就有对偶问题 $w_i^{(0)} = 0$）
- 如果 $w_i^{(0)} > 0$，就有 $A_i x^{(0)} = b_i$
- 如果 $w^{(0)}p_j < c_j$（称作松约束），就有 $x_j^{(0)} = 0$

Example 17. 已知线性规划

$$\begin{aligned}
&\max \quad f(X) = 3x_1 + 4x_2 + x_3 \\
&s.t. \\
&x_1 + 2x_2 + x_3 \leq 10 \\
&2x_1 + 2x_2 + x_3 \leq 16 \\
&x_j \geq 0, j = 1, 2, 3
\end{aligned}$$

的最优解为 $\boldsymbol{X}^\star = (6, 2, 0)^T$，求其对偶规划的最优解？

Solution. 对偶规划为

$$\begin{aligned}
&\min \quad g(Y) = 10y_1 + 16y_2 \\
&s.t. \\
&y_1 + 2y_2 \geq 3 \\
&2y_1 + 2y_2 \geq 4 \\
&y_1 + y_2 \geq 1 \\
&y_j \geq 0, j = 1, 2
\end{aligned}$$

设对偶规划的最优解为 $\boldsymbol{Y}^\star = (y_1^\star, y_2^\star)^T$，由松弛定理得（因为 $x_3 = 0,$ 那么对偶规划中去掉第三个不等式）

$$\begin{cases} y_1^\star + 2y_2^\star = 3 \\ 2y_1^\star + 2y_2^\star = 4 \end{cases} \Rightarrow y_1^\star = y_2^\star = 1$$

所以对偶规划的最优解为 $\boldsymbol{Y}^\star = (1, 1)^T.$

Example 18. 原问题

$$\begin{aligned}
&\min \;\; 2x_1 + 3x_2 + x_3 \\
&\;s.t. \\
&3x_1 - x_2 + x_3 \geqslant 1 \\
&x_1 + 2x_2 - 3x_3 \geqslant 2 \\
&x_1, x_2, x_3 \geqslant 0
\end{aligned}$$

对偶问题

$$\begin{aligned}&\max\ \ w_1+2w_2\\&\ s.t.\\&3w_1+w_2\le 2\\&-w_1+2w_2\leqslant 3\\&w_1-3w_2\leqslant 1\\&w_1,w_2\geqslant 0\end{aligned}$$

设对偶问题的最优解为 $\overline{\boldsymbol{w}}=(\boldsymbol{w}_1,\boldsymbol{w}_2)=\left(\frac{1}{7},\frac{11}{7}\right)$，试求原问题的最优解。

Solution. 由于在最优解 $\overline{\boldsymbol{w}}$ 处，对偶问题的第三个约束成立严格不等式，因此在原问题中第三个变量 $x_3=0$，又由于 $\overline{\boldsymbol{w}}$ 的两个分量均大于令零，因此在原问题中前两个约束在最优解处成立等式，即

$$\begin{cases}3x_1-x_2+x_3=1\\x_1+2x_2-3x_3=2\end{cases}$$

把 $x_3=0$ 带入上述方程组，得到

$$\begin{cases}3x_1-x_2=1\\x_1+2x_2=2\end{cases}$$

解得 $x_1=\frac{4}{7},x_2=\frac{5}{7}$，因此原问题的最优解是

$$\overline{x}=(x_1,x_2,x_3)^{\mathrm{T}}=\left(\frac{4}{7},\frac{5}{7},0\right)^{\mathrm{T}}$$

Example 19. 已知下面线性规划问题的最优解为 $X^*=(2,2,4,0)^T$，求其对偶规划的最优解。

$$\begin{aligned}&\max\ \ 2x_1+4x_2+x_3+x_4\\&s.t.\\&x_1+3x_2+x_4\le 8\\&2x_1+x_2\le 6\\&x_2+x_3+x_4\le 6\\&x_1+x_2+x_3\le 9\\&x_i\ge 0,i=1\sim 4\end{aligned}$$

Solution. 对偶问题为：

$$\begin{aligned}&\min\ \ 8w_1+6w_2+6w_3+9w_4\\&s.t.\\&w_1+2w_2+w_4\ge 2\\&3w_1+w_2+w_3+w_4\ge 4\\&w_3+w_4\ge 1\\&w_1+w_3\ge 1\\&w_i\ge 0(i=1,2,3,4)\end{aligned}$$

由于 x_1,x_2,x_3 均大于 0，那么对偶问题的前三个约束等式严格成立，即

$$\begin{aligned}&w_1+2w_2+w_4=2\\&3w_1+w_2+w_3+w_4=4\\&w_3+w_4=1\end{aligned}$$

将原问题的最优解代入原问题中，发现原问题的第四个约束条件的不等式严格成立，那么有 $w_4 = 0$.

综上，求解得到 $w = (\frac{4}{5}, \frac{3}{5}, 1, 0)^T$，最优值为 16.

3.4 对偶单纯形法

求解基本步骤：

$$k = \min\{i|b_i < 0\}$$

$$\frac{\sigma_l}{a_{kl}} = \min_{1\leq j\leq n}\left\{\frac{\sigma_j}{a_{kj}}|a_{kj} < 0\right\}$$

Example 20. 用对偶单纯形法求解下面的问题

$$\begin{aligned}
&\min \quad f(\boldsymbol{X}) = x_1 + 2x_2 \\
&s.t. \\
&x_1 + 2x_2 \geq 4 \\
&x_1 \leq 5 \\
&3x_1 + x_2 \geq 6 \\
&x_j \geq 0, \quad j = 1 \sim 2
\end{aligned}$$

Solution. 转换为下面的标准形

$$\begin{aligned}
&\min \quad f(\boldsymbol{X}) = x_1 + 2x_2 \\
&s.t. \\
&-x_1 - 2x_2 + x_3 = -4 \\
&x_1 + x_4 = 5 \\
&-3x_1 - x_2 + x_5 = -6 \\
&x_j \geq 0, \quad j = 1 \sim 5
\end{aligned}$$

对偶单纯形法求解过程如下：

P_1	P_2	P_3	P_4	P_5	b
$\boxed{-1}$	-2	1	0	0	-4
1	0	0	1	0	5
-3	-1	0	0	1	-6
-1	-2	0	0	0	0

P_1	P_2	P_3	P_4	P_5	b
1	2	-1	0	0	4
0	-2	1	1	0	1
0	5	-3	0	1	6
0	0	-1	0	0	4

这时 $b > 0$，所以当前最优解为 $\boldsymbol{X}^\star = (4, 0, 0, 1, 6)^T$，因此原规划问题的最优解为 $\boldsymbol{X}^\star = (4, 0)^T$. 对偶问题的最优解为 $\boldsymbol{W}^\star = (1, 0, 0)^T$

【注】：对偶规划的最优解的第 l 个分量就是原规划最终单纯形表中剩余变量的判别数的相反数。

4 无约束最优化计算方法

4.1 基本思想

无约束优化问题：

$$\min_{X\in R^n} f(\boldsymbol{X})$$

新点 $X^{k+1}=X^k+t_k\boldsymbol{P}^k$，使得 $f\left(X^{k+1}\right)<f\left(X^k\right)$，其中 t_k 称为步长因子，$\boldsymbol{P}^k$ 表示下降方向，下降方向需满足 $\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k<0.$

【注】：后面的所有方法，就是在确定下降方向 $\boldsymbol{P}^k$ 和步长因子 t_k

确定步长因子 t_k？即选取使得

$$f\left(\boldsymbol{X}^{\boldsymbol{k}}+t_k\boldsymbol{P}^k\right)=\min_t f\left(\boldsymbol{X}^k+t\boldsymbol{P}^k\right)$$

这时，t_k 称为最优步长。

【注】:求一元函数 $\varphi(t)=f\left(\boldsymbol{X}^k+t\boldsymbol{P}^k\right)$ 极小点的迭代法称为一维搜索,$\varphi'(t)=\nabla f\left(\boldsymbol{X}^k+t\boldsymbol{P}^k\right)^T\boldsymbol{P}^k$

Definition 5 (**收敛速度**). 设序列 $\left\{\boldsymbol{X}^k\right\}$ 收敛于 $\boldsymbol{X}^\star$，若

$$\lim_{k\to\infty}\frac{\left\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^\star\right\|}{\left\|\boldsymbol{X}^k-\boldsymbol{X}^\star\right\|}=\beta$$

则 $0<\beta<1$，称 $\left\{\boldsymbol{X}^k\right\}$ 为 β 线性收敛，$\beta=0$ 时称为超线性收敛，$\beta=1$ 时称为次线性收敛。

又若

$$\lim_{k\to\infty}\frac{\left\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^\star\right\|}{\left\|\boldsymbol{X}^k-\boldsymbol{X}^\star\right\|^p}=\beta<+\infty$$

则称 $\left\{\boldsymbol{X}^k\right\}$ 为 p 阶收敛。这里 $p\geq 1.$

Definition 6 (**二次终止性**). 一个算法用于求解具有正定矩阵的二次函数

$$f(X)=\frac{1}{2}X^TAX+b^TX+c$$

时，可以在有限步内达到它的极小点。

4.2 非精确一维搜索的三大准则

Goldstein 准则

$$\begin{cases}(i)f\left(\boldsymbol{X}^k+t_k\boldsymbol{P}^k\right)\leq f\left(\boldsymbol{X}^k\right)+\rho t_k\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k\\(ii)f\left(\boldsymbol{X}^k+t_k\boldsymbol{P}^k\right)\geq f\left(\boldsymbol{X}^k\right)+(1-\rho)t_k\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k\end{cases}$$

其中 $0<\rho<1.$

Wolfe 准则

$$\begin{cases}(i)f\left(\boldsymbol{X}^k+t_k\boldsymbol{P}^k\right)\leq f\left(\boldsymbol{X}^k\right)+\rho t_k\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k\\(ii)\nabla f\left(\boldsymbol{X}^k+t_k\boldsymbol{P}^k\right)^T\boldsymbol{P}^k\geq\sigma\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k\end{cases}$$

其中 $\sigma\in(\rho,1).$

Armijo 准则

给定 $\beta \in (0,1), \quad \rho \in \left(0, \frac{1}{2}\right), \quad \tau > 0$，设 m_k 是使得下式

$$f\left(\boldsymbol{X}^k + \beta^{m_k}\tau\boldsymbol{P}^k\right) \leq f\left(\boldsymbol{X}^k\right) + \rho\beta^{m_k}\tau\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k$$

成立的最小非负整数。

若令 $t_k = \beta^{m_k}\tau$，其就是 Goldstein 准则中的第一个准则

$$f\left(\boldsymbol{X}^k + t_k\boldsymbol{P}^k\right) \leq f\left(\boldsymbol{X}^k\right) + \rho t_k\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{P}^k$$

为了保证算法收敛性，需要保证每次搜索方向 p_k 与其梯度方向 $-g_k = -\nabla f\left(\boldsymbol{X}^k\right)$ 成锐角。

4.3 0.618 法

区间选择准则：原区间为 $[a_k, b_k]$，0.382 处为 λ_k，0.618 处为 μ_k.

- 若 $f(\lambda_k) > f(\mu_k)$，则新区间为 $[\lambda_k, b_k]$.
- 若 $f(\lambda_k) \leqslant f(\mu_k)$，则新区间为 $[a_k, \mu_k]$.

【注】：口诀：哪边大，舍哪边。

4.4 最速下降法

求解

$$\min_{\boldsymbol{X}\in\boldsymbol{R}^n} f(\boldsymbol{X})$$

下降方向：$\boldsymbol{P}^k = -\nabla f\left(\boldsymbol{X}^k\right)$，确定最优步长 t_k 使得

$$f\left(\boldsymbol{X}^k + t_k\nabla f\left(\boldsymbol{X}^k\right)\right) = \min_t f\left(\boldsymbol{X}^k + t\nabla f\left(\boldsymbol{X}^k\right)\right)$$

若 $f(\boldsymbol{X})$ 具有二阶连续偏导数，由 Taylor 公式得

$$\begin{aligned} f\left(\boldsymbol{X}^k - t\nabla f\left(\boldsymbol{X}^k\right)\right) \approx & f\left(\boldsymbol{X}^k\right) - t\nabla f\left(\boldsymbol{X}^k\right)^T\nabla f\left(\boldsymbol{X}^k\right) \\ & + \frac{1}{2}t^2\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{H}\left(\boldsymbol{X}^k\right)\nabla f\left(\boldsymbol{X}^k\right) \end{aligned}$$

其中 $H\left(\boldsymbol{X}^k\right)$ 是 $f(\boldsymbol{X})$ 在 $\boldsymbol{X}^k$ 的 Hesse 矩阵。

令

$$\begin{aligned} \frac{df\left(\boldsymbol{X}^k - t\nabla f\left(\boldsymbol{X}^k\right)\right)}{dt} = & -\nabla f\left(\boldsymbol{X}^k\right)^T\nabla f\left(\boldsymbol{X}^k\right) \\ & + t\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{H}\left(\boldsymbol{X}^k\right)\nabla f\left(\boldsymbol{X}^k\right) = 0 \end{aligned}$$

可得最优步长

$$t_k = \frac{\nabla f\left(\boldsymbol{X}^k\right)^T\nabla f\left(\boldsymbol{X}^k\right)}{\nabla f\left(\boldsymbol{X}^k\right)^T\boldsymbol{H}\left(\boldsymbol{X}^k\right)\nabla f\left(\boldsymbol{X}^k\right)} = \frac{\left(\boldsymbol{g}^k\right)^T\boldsymbol{g}^k}{\left(\boldsymbol{g}^k\right)^T\boldsymbol{H}^k\boldsymbol{g}^k}$$

由此得到第 $k+1$ 步的迭代点为

$$\boldsymbol{X}^{k+1}=\boldsymbol{X}^{k}-\frac{\left(\boldsymbol{g}^{k}\right)^{T}\boldsymbol{g}^{k}}{\left(\boldsymbol{g}^{k}\right)^{T}\boldsymbol{H}^{k}\boldsymbol{g}^{k}}\boldsymbol{g}^{k}$$

【注】：$\varphi'(t)=\nabla f(\boldsymbol{X}+t\boldsymbol{P})^{T}\boldsymbol{P}=0$，得 $\left(\boldsymbol{g}^{k+1}\right)^{T}\boldsymbol{g}^{k}=0$，因此，**最速下降法存在据锯齿现象**

Example 21. 用最速下降法求 $f(\boldsymbol{X})=(x_1-1)^2+(x_2-1)^2$ 的极小点。

Solution.

$$\boldsymbol{g}^{k}=\nabla f(\boldsymbol{X})=\left(2\left(x_1-1\right),2\left(x_2-1\right)\right)^{T}$$

$$\boldsymbol{H}^{k}=\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}$$

$$\boldsymbol{X}^{0}=(0,0)^{T}$$

$$t_0=\frac{\left(\boldsymbol{g}^{0}\right)^{T}\boldsymbol{g}^{0}}{\left(\boldsymbol{g}^{0}\right)^{T}\boldsymbol{H}^{0}\boldsymbol{g}^{0}}=\frac{1}{2}$$

$$X^{1}=X^{0}-t_0g^{0}=(0,0)^{T}-\frac{1}{2}(-2,-2)^{T}=(1,1)^{T}$$

$$\nabla f\left(\boldsymbol{X}^{1}\right)=(0,0)^{T},\quad \left\|\nabla f\left(\boldsymbol{X}^{1}\right)\right\|<\varepsilon$$

$$\boldsymbol{X}^{\star}=\boldsymbol{X}^{1}$$

4.5 牛顿法（二阶收敛）

基本思想：牛顿法是用一个二次函数去近似一个目标函数，然后精确的求出这个二次函数的极小点，以它作为目标函数极小点的近似值。

在点 $\boldsymbol{X}^k$ 处对目标函数按 Taylor 展开

$$\begin{aligned}f(\boldsymbol{X})\approx q(\boldsymbol{X})\triangleq\ & f\left(\boldsymbol{X}^{k}\right)+\boldsymbol{g}\left(\boldsymbol{X}^{k}\right)^{T}\left(\boldsymbol{X}-\boldsymbol{X}^{k}\right)\\ &+\frac{1}{2}\left(\boldsymbol{X}-\boldsymbol{X}^{k}\right)^{T}\boldsymbol{H}\left(\boldsymbol{X}^{k}\right)\left(\boldsymbol{X}-\boldsymbol{X}^{k}\right)\end{aligned}$$

令 $\nabla q(\boldsymbol{X})=\boldsymbol{H}\left(\boldsymbol{X}^{k}\right)\left(\boldsymbol{X}-\boldsymbol{X}^{k}\right)+\boldsymbol{g}\left(\boldsymbol{X}^{k}\right)=0$，得

$$\boldsymbol{H}\left(\boldsymbol{X}^{k}\right)\left(\boldsymbol{X}-\boldsymbol{X}^{k}\right)=-\boldsymbol{g}\left(\boldsymbol{X}^{k}\right)$$

若 Hesse 矩阵正定，则

$$\boldsymbol{X}-\boldsymbol{X}^{k}=-\boldsymbol{H}^{-1}\left(\boldsymbol{X}^{k}\right)\boldsymbol{g}\left(\boldsymbol{X}^{k}\right)$$

则由上式解出来的 $\boldsymbol{X}=\boldsymbol{X}^{k+1}$，就是二次函数 $q(\boldsymbol{X})$ 的极小点，即

$$\boldsymbol{X}^{k+1}=\boldsymbol{X}^{k}-\boldsymbol{H}^{-1}\left(\boldsymbol{X}^{k}\right)\boldsymbol{g}\left(\boldsymbol{X}^{k}\right)$$

【注】：当目标函数是正定二次函数时，牛顿法能够一次到达极小点（一步到位），具有二次终止性。

4.6 共轭梯度法：共轭性与最速下降法的结合，具有二次终止性

最速下降法存在锯齿现象，收敛速度慢；而牛顿法需要计算 Hesse 矩阵，计算量较大，**而共轭梯度法的收敛速度介于两者之间，无需计算 Hesse 矩阵，具有二次收敛性。**

Definition 7 (共轭)**.** 设 $\boldsymbol{Q}$ 是对称正定阵，如果 $\boldsymbol{X}$ 和 $\boldsymbol{QY}$ 正交，即

$$\boldsymbol{X}^T\boldsymbol{QY} = 0$$

则称 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 关于 $\boldsymbol{Q}$ 共轭。

FR 共轭梯度法

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + t_k\boldsymbol{P}^k$$

$$t_k = \frac{-(\boldsymbol{g}^k)^T\boldsymbol{P}^k}{(\boldsymbol{P}^k)^T\boldsymbol{H}(\boldsymbol{P}^k)}$$

$$\boldsymbol{P}^k = \begin{cases} -\boldsymbol{g}^k, & k=0 \\ -\boldsymbol{g}^k + \frac{\|\boldsymbol{g}^k\|^2}{\|\boldsymbol{g}^{k-1}\|^2}\boldsymbol{P}^{k-1}, & k>0 \end{cases}$$

Example 22. 用 *FR* 方法求解 $f(\boldsymbol{X}) = x_1^2 + 2x_2^2$，取初始点 $\boldsymbol{X}^0 = (5,5)^T$.

Solution. 解：

$$\boldsymbol{g} = \nabla f(\boldsymbol{X}) = (2x_1, 4x_2)^T, \boldsymbol{H} = \nabla^2 f(\boldsymbol{X}) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

第一次迭代

$$\boldsymbol{P}^0 = -\boldsymbol{g}^0 = (-10,-20)^T$$

$$\begin{aligned}
\boldsymbol{X}^1 &= \boldsymbol{X}^0 + t_0\boldsymbol{P}_0 = \boldsymbol{X}^0 - t_0\boldsymbol{g}_0 \\
&= (5,5)^T + \frac{-(\boldsymbol{g}^0)^T\boldsymbol{P}^0}{(\boldsymbol{P}^0)^T\boldsymbol{H}\boldsymbol{P}^0}(-10,-20)^T \\
&= (5,5)^T + \frac{5}{18}(-10,-20)^T \\
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T
\end{aligned}$$

第二次迭代

$$\begin{aligned}
\boldsymbol{P}^1 &= -\boldsymbol{g}^1 + \alpha_0\boldsymbol{P}_0 \\
&= \left(-\frac{40}{9}, \frac{20}{9}\right)^T + \frac{\|\boldsymbol{g}^1\|^2}{\|\boldsymbol{g}^0\|^2}(-10,-20)^T \\
&= \left(-\frac{40}{9}, \frac{20}{9}\right)^T + \frac{4}{81}(-10,-20)^T \\
&= \left(-\frac{400}{81}, \frac{100}{81}\right)^T
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{X}^2 &= \boldsymbol{X}^1 + t_1\boldsymbol{P}^1 \\
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T + \frac{-\left(\boldsymbol{g}^1\right)^T \boldsymbol{P}^1}{\left(\boldsymbol{P}^1\right)^T \boldsymbol{H}\boldsymbol{P}^1}\left(-\frac{400}{81}, \frac{100}{81}\right)^T \\
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T + \frac{9}{20}\left(-\frac{400}{81}, \frac{100}{81}\right)^T \\
&= (0,0)^T
\end{aligned}
$$

此时 $\boldsymbol{g}^2 = 0$，已达到极小点 $\boldsymbol{X}^2 = (0,0)^T$.

此例验证了共轭梯度法的二次终止性。

【注】：初始方向应为负梯度方向，否则所产生的一般不为共轭方向。

4.7 拟牛顿法

基本思想：用一个矩阵来近似 Hesse 逆矩阵。

拟牛顿方程：

$$\boldsymbol{H}^{k+1}\boldsymbol{Y}^k = \boldsymbol{S}^k$$

秩 1 具有如下形式：

$$\boldsymbol{H}^{k+1} = \boldsymbol{H}^k + \Delta\boldsymbol{H}^k = \boldsymbol{H}^k + \lambda\boldsymbol{u}\boldsymbol{u}^T$$

秩 1 校正（SR1）

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + t_k\boldsymbol{P}^k$$

$$\boldsymbol{Y}^k \triangleq \boldsymbol{g}^{k+1} - \boldsymbol{g}^k, \quad \boldsymbol{S}^k \triangleq \boldsymbol{X}^{k+1} - \boldsymbol{X}^k, \quad \boldsymbol{Z}^k = \boldsymbol{S}^k - \boldsymbol{H}^k\boldsymbol{Y}^k$$

$$\boldsymbol{H}^{k+1} = \boldsymbol{H}^k + \frac{\boldsymbol{Z}^k\left(\boldsymbol{Z}^k\right)^T}{\left(\boldsymbol{Y}^k\right)^T\boldsymbol{Z}^k}$$

$$\boldsymbol{P}^{k+1} = -\boldsymbol{H}^{k+1}\boldsymbol{g}^{k+1}$$

【注】：SR1 是一种共轭方向法，当然具有二次终止性。对于正定二次函数，SR1 方法至多 n 步终止。但是，不能保证 $\boldsymbol{P}^k = -\boldsymbol{H}^k\boldsymbol{g}^k$ 是下降方向，即 $\boldsymbol{H}^k$ 不正定。

秩 2 具有如下形式：

$$\Delta\boldsymbol{H}^k = \lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T + \lambda_2\boldsymbol{u}_2\boldsymbol{u}_2^T$$

秩 2 校正公式（DFP 算法）

$$H^{k+1} = H^k + \frac{S^k\left(S^k\right)^T}{\left(S^k\right)^T Y^k} - \frac{H^kY^k\left(H^kY^k\right)^T}{\left(H^kY^k\right)^T Y^k}$$

(正定遗传性)：若 $\boldsymbol{H}^0$ 是对称正定矩阵，则 DFP 校正公式所产生的 $\boldsymbol{H}^k$ 是对称正定矩阵。

【注】：一般情况下，$\boldsymbol{H}^0 = \boldsymbol{I}_n$

【注】：DFP 本质上也是一种共轭方向法，也具有二次收敛性。

Example 23. *用 DFP 算法求* $\min 2x_1^2 + x_2^2 - 4x_1 + 2$，*取* $\boldsymbol{X}^0 = (2,1)^T$

Solution.

$$\boldsymbol{g}=\begin{pmatrix} 4\left(x_1-1\right) \\ 2x_2 \end{pmatrix}, \quad \boldsymbol{Q}=\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \boldsymbol{H}^0=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

第一次迭代：

$$\boldsymbol{g}^0=(4,2)^T$$

$$\boldsymbol{P}^0=-\boldsymbol{H}^0\boldsymbol{g}^0=-\boldsymbol{g}^0$$

$$\boldsymbol{X}^1=\boldsymbol{X}^0+t_0\boldsymbol{P}^0=\boldsymbol{X}^0+\frac{-(\boldsymbol{g}^0)^T\boldsymbol{P}^0}{\left(\boldsymbol{P}^0\right)^T\boldsymbol{Q}\boldsymbol{P}^0}\boldsymbol{P}^0=\left(\frac{8}{9},\frac{4}{9}\right)^T$$

第二次迭代：

$$\boldsymbol{g}^1=\left(-\frac{4}{9},\frac{8}{9}\right)^T$$

$$\boldsymbol{S}^0=\boldsymbol{X}^1-\boldsymbol{X}^0=\left(-\tfrac{10}{9},-\tfrac{5}{9}\right)^T$$
$$\boldsymbol{Y}^0=\boldsymbol{g}^1-\boldsymbol{g}^0=\left(-\tfrac{40}{9},-\tfrac{10}{9}\right)^T$$

$$\begin{aligned}\boldsymbol{H}^1&=\boldsymbol{H}^0+\frac{\boldsymbol{S}^0\left(\boldsymbol{S}^0\right)^T}{\left(\boldsymbol{S}^0\right)^T\boldsymbol{Y}^0}-\frac{\boldsymbol{Y}^0\left(\boldsymbol{Y}^0\right)^T}{\left(\boldsymbol{Y}^0\right)^T\boldsymbol{Y}^0}\\&=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+\frac{1}{18}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}-\frac{1}{17}\begin{pmatrix} 16 & 4 \\ 4 & 1 \end{pmatrix}\\&=\frac{1}{306}\begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix}\end{aligned}$$

搜索方向

$$\boldsymbol{P}^1=-\boldsymbol{H}^1\boldsymbol{g}^1=\frac{4}{17}(1,-4)^T$$

步长

$$t_1=\min_{\lambda} f(\boldsymbol{X}^1+t\boldsymbol{P}^1)$$

从 $\boldsymbol{X}^1$ 出发沿 $\boldsymbol{P}^1$ 进行一维搜索，即

$$\boldsymbol{X}^2=\boldsymbol{X}^1+t_1\boldsymbol{P}^1=(1,0)^T$$

第三次迭代：

$$\left\|\boldsymbol{g}^2\right\|=0$$

迭代终止，所以 $\boldsymbol{X}^2$ 是极小点。

4.8 信赖域方法

前面的方法是先确定搜索方法，再确定步长，信赖域方法是先确定一个步长，再确定搜索方法。

实际下降量

$$\Delta f^k=f\left(\boldsymbol{X}^k\right)-f\left(\boldsymbol{X}^k+\boldsymbol{S}^k\right)$$

预测下降量

$$\Delta q^k=f\left(\boldsymbol{X}^k\right)-q^k\left(\boldsymbol{S}^k\right)$$

比值

$$r^k = \Delta f^k / \Delta q^k$$

它衡量二次模型近似目标函数的程度，r^k 越接近于 1，表示近似程度越好。

【注】：当 r^k**<0 时，无效步；**$(\frac{1}{4}, \frac{3}{4})$**，不变；**$(\frac{3}{4}, 1)$**，扩大** h**；**$(0, \frac{1}{4})$**，缩小** h**。**h **表示信赖域半径。**

【注】：简记口诀：比值大，扩大；比值小，缩小。

【注】：在较强的假设下，信赖域方法具有二阶收敛速度

5 最优性条件

5.1 无约束问题的极值条件

（必要条件）：若 $\bar{x}$ 是局部极小点，则梯度 $\nabla f(\bar{x}) = 0$，且 Hesse 矩阵 $\nabla^2 f(\bar{x}) = 0$ 半正定。

（二阶充分条件）：若梯度 $\nabla f(\bar{x}) = 0$，且 Hesse 矩阵 $\nabla^2 f(\bar{x}) = 0$ 正定，则 $\bar{x}$ 是局部极小点。

（充要条件）：设 $f(x)$ 是定义在 $\mathbb{R}^n$ 上的可微**凸函数**，$\bar{x} \in \mathbb{R}^n$，则 $\bar{x}$ 是全局极小点的充分必要条件是梯度 $\nabla f(\bar{x}) = 0$。

5.2 约束问题的极值条件

5.2.1 不等式约束优化问题的一阶最优性条件

下降方向满足：$\nabla f(\overline{x})^{\mathrm{T}} d < 0$

可行方向满足：$\overline{x} + \lambda d \in S$

考虑非线性规划问题：

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

这个问题的可行域

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

K-T 条件：

$$\begin{array}{l} \nabla f(\overline{x}) - \sum_{i=1}^m w_i \nabla g_i(\overline{x}) = 0 \\ w_i g_i(\overline{x}) = 0, \quad i = 1, \cdots, m \\ w_i \geqslant 0, \quad i = 1, \cdots, m \end{array}$$

Example 24.

$$\begin{array}{ll} \min & (x_1 - 2)^2 + x_2^2 \\ s.\ t. & x_1 - x_2^2 \geqslant 0 \\ & -x_1 + x_2 \geqslant 0 \end{array}$$

验证下列两点

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{和} \quad x^{(2)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

是否为 K-T 点。

Solution. 记

$$
\begin{aligned}
f(x) &= (x_1-2)^2+x_2^2 \\
g_1(x) &= x_1-x_2^2 \\
g_2(x) &= -x_1+x_2
\end{aligned}
$$

目标函数和约束函数的梯度是

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

先验证 $x^{(1)}$，在这一点，$g_1(x)\geqslant 0$ 和 $g_2(x)\geqslant 0$ 都是起作用约束，目标函数和约束函数的梯度分别是

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

设

$$
\begin{bmatrix} -4 \\ 0 \end{bmatrix}-w_1\begin{bmatrix} 1 \\ 0 \end{bmatrix}-w_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

即

$$
\begin{aligned}
-4-w_1+w_2 &= 0 \\
-w_2 &= 0
\end{aligned}
$$

解此方程组得到

$$
w_1=-4,\quad w_2=0
$$

由于 $w_1<0$，因此 $x^{(1)}$ 不是 K-T 点。

在验证 $x^{(2)}$，$g_1(x)\geqslant 0$ 和 $g_2(x)\geqslant 0$ 都是起作用约束，目标函数和约束函数的梯度分别是

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

设

$$
\begin{bmatrix} -2 \\ 2 \end{bmatrix}-w_1\begin{bmatrix} 1 \\ -2 \end{bmatrix}-w_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

即

$$
\begin{cases}
-2-w_1+w_2=0 \\
2+2w_1-w_2=0
\end{cases}
$$

解此方程组得到

$$
w_1=0,\quad w_2=2
$$

所以 $x^{(2)}$ 是 K-T 点。

Example 25.

$$
\begin{aligned}
\min \quad & f(x)\stackrel{\text{def}}{=}(x_1-1)^2+x_2 \\
s.\ t. \quad & g_1(x)=-x_1-x_2+2\geqslant 0 \\
& g_2(x)=x_2\geqslant 0
\end{aligned}
$$

求满足 K-T 条件的点。

Solution. 目标函数和约束函数的梯度分别是

$$\nabla f(x)=\left[\begin{array}{c} 2\left(x_1-1\right) \\ 1 \end{array}\right], \quad \nabla g_1(x)=\left[\begin{array}{c} -1 \\ -1 \end{array}\right], \quad \nabla g_2(x)=\left[\begin{array}{c} 0 \\ 1 \end{array}\right]$$

K-T 条件为

$$\begin{aligned} &\nabla f(x)-\textstyle\sum_{i=1}^{2} w_i \nabla g_i(x)=0 \\ &w_i g_i(x)=0, \quad i=1,2 \\ &w_i \geqslant 0, \quad i=1,2 \end{aligned}$$

即

$$\begin{aligned} &2\left(x_1-1\right)+w_1=0 \\ &1+w_1-w_2=0 \\ &w_1\left(-x_1-x_2+2\right)=0 \\ &w_2 x_2=0 \\ &w_1, w_2 \geqslant 0 \end{aligned}$$

解此方程组得到一组解

$$x_1=1, \quad x_2=0, \quad w_1=0, \quad w_2=1$$

因为 w_1 和 w_2 都是非负数，因此得到 *K-T* 点

$$\bar{x}=\left[\begin{array}{l} 1 \\ 0 \end{array}\right]$$

5.2.2 一般约束优化问题的一阶最优性条件

$$\begin{array}{rl} \min & f(x), \quad x \in \mathbb{R}^n \\ \text{s. t.} & g(x) \geq 0 \\ & h(x)=0 \end{array}$$

Lagrange 函数

$$L(x, w, v)=f(x)-\sum_{i=1}^{m} w_i g_i(x)-\sum_{j=1}^{l} v_j h_j(x)$$

KKT 条件如下

$$\begin{aligned} &\nabla_x L(x, w, v)=0 \quad \text{原始最优化条件} \\ &g_i(x) \geqslant 0, \quad i=1, \cdots, m \quad \text{原始可行性条件} \\ &h_j(x)=0, \quad j=1, \cdots, l \quad \text{原始可行性条件} \\ &w_i g_i(x)=0, \quad i=1, \cdots, m \quad \text{互补松弛条件} \\ &w_i \geqslant 0, \quad i=1, \cdots, m \quad \text{对偶可行性条件} \end{aligned}$$

5.3 惩罚函数法

5.3.1 外点罚函数法

对于

$$\begin{aligned}&\min f(\boldsymbol{X})\\&\text{ s.t.}\\&g_i(\boldsymbol{X})\geq 0, i=1\sim m\end{aligned}$$

构造外点罚函数

$$P\left(\boldsymbol{X},m_k\right)=f(\boldsymbol{X})+m_k\sum_{i=1}^{m}\left(\min\left(g_i(\boldsymbol{X}),0\right)\right)^2$$

Example 26. 用外点罚函数法求解下列问题

$$\begin{array}{cl}\min & f(x)=\left(x_1-1\right)^2+x_2^2\\ s.\ t. & g(x)=x_2-1\geqslant 0\end{array}$$

Solution. 定义罚函数

$$\begin{aligned}F(x,\sigma)&=\left(x_1-1\right)^2+x_2^2+\sigma\left[\max\left\{0,-\left(x_2-1\right)\right\}\right]^2\\&=\begin{cases}\left(x_1-1\right)^2+x_2^2, & x_2\geqslant 1\\ \left(x_1-1\right)^2+x_2^2+\sigma\left(x_2-1\right)^2, & x_2<1\end{cases}\end{aligned}$$

用解析法求解

$$\min \quad F(x,\sigma)$$

求偏导，得

$$\begin{aligned}\frac{\partial F}{\partial x_1}&=2\left(x_1-1\right)\\ \frac{\partial F}{\partial x_2}&=\begin{cases}2x_2, & x_2\geqslant 1\\ 2x_2+2\sigma\left(x_2-1\right), & x_2<1\end{cases}\end{aligned}$$

令

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

得到

$$\overline{x}_\sigma=\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\ \frac{\sigma}{1+\sigma}\end{bmatrix}$$

令 $\sigma\to+\infty$，得

$$\overline{x}_\sigma\to\overline{x}=\begin{bmatrix}1\\1\end{bmatrix}$$

5.3.2 内点罚函数法

内点罚函数只适用于以下问题

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

定义障碍函数

$$G(x,r)=f(x)+rB(x)$$

其中 $B(x)$ 是连续函数，当点 x 趋向于可行域边界时，$B(x)\to\infty$，r 是很小的正数。

两种最重要的形式为：

$$B(x)=\sum_{i=1}^{m}\frac{1}{g_i(x)}$$

及

$$B(x)=-\sum_{i=1}^{m}\log g_i(x)$$

Example 27. 用内点罚函数为求解

$$\begin{array}{ll}\min & \frac{1}{12}\left(x_1+1\right)^3+x_2\\ s.\ t. & x_1-1\geqslant 0\\ & x_2\geqslant 0\end{array}$$

Solution. 定义障碍函数

$$G\left(x,r_k\right)=\frac{1}{12}\left(x_1+1\right)^3+x_2+r_k\left(\frac{1}{x_1-1}+\frac{1}{x_2}\right)$$

下面用解析法求解问题

$$\begin{array}{ll}\min & G\left(x,r_k\right)\\ s.\ t. & x\in S\end{array}$$

令

$$\begin{aligned}\frac{\partial G}{\partial x_1}&=\frac{1}{4}\left(x_1+1\right)^2-\frac{r_k}{\left(x_1-1\right)^2}=0\\ \frac{\partial G}{\partial x_2}&=1-\frac{r_k}{x_2^2}=0\end{aligned}$$

解得

$$\overline{x}_{r_k}=\left(x_1,x_2\right)=\left(\sqrt{1+2\sqrt{r_k}},\sqrt{r_k}\right)$$

当 $r_k\to 0$，$\overline{x}_{r_k}\to\overline{x}=(1,0)$，$\overline{x}$ 是问题的最优解。