第4章 非线性规划

无约束最优化方法

无约束最优化方法

- 求解n元函数的无约束最优化问题（UMP）

$$\min f(x)$$

其中$x = (x_1, \cdots, x_n)^T \in \mathbf{R}^n, f: \mathbf{R}^n \mapsto \mathbf{R}$。

1. 无约束最优化问题的最优性条件
2. 最速下降法
3. 共轭方向法

UMP问题的最优性条件

下降方向

定理 4.4.1 设$f:\mathrm{R}^n \mapsto \mathrm{R}$在点$\bar{x}\in\mathrm{R}^n$处可微。若存在$p\in\mathrm{R}^n$，使得$\nabla f(\bar{x})^{\mathrm{T}}p<0$，则向量$\boldsymbol{p}$是$\boldsymbol{f}$在点$\bar{\boldsymbol{x}}$处的下降方向。

向量内积

- 证：因为$\boldsymbol{f}$在$\bar{\boldsymbol{x}}$处可微，由$\boldsymbol{f}$在$\bar{\boldsymbol{x}}$处的 Taylor 展开式，对任意的$\boldsymbol{t>0}$，有：

$$f(\bar{x}+tp)=f(\bar{x})+t\nabla f(\bar{x})^{\mathrm{T}}p+o(\|tp\|)。$$

- 由于$\nabla f(\bar{x})^{\mathrm{T}}p<0$，$\boldsymbol{t>0}$，可知$t\nabla f(\bar{x})^{\mathrm{T}}p<0$。
- 因此，$\exists\delta>0$，$\forall t\in(0,\delta)$，有$t\nabla f(\bar{x})^{\mathrm{T}}p+o(\|tp\|)<0$。
- 即，$\forall t\in(0,\delta)$，$f(\bar{x}+tp)<f(\bar{x})$。可知$\boldsymbol{p}$是$\boldsymbol{f}$在$\bar{\boldsymbol{x}}$处的下降方向（定义 4.1.3）。

下降方向

- 说明：证 $\exists\delta>0$， $\forall t\in(0,\delta)$，有 $t\nabla f(\bar{x})^{\mathrm{T}}p+o(\|tp\|)<0$ 。**(1)**
- 若 $o(\|tp\|)\le 0$，则(1)已经成立。下面假设 $o(\|tp\|)>0$ 。
- 只需证 $\exists\delta>0$， $\forall t\in(0,\delta)$， $\dfrac{o(\|tp\|)}{t}<-\nabla f(\bar{x})^{\mathrm{T}}p$ 。 **(2)**
- 由定理 4.2.3 的证明，可知 $\lim\limits_{t\to 0}\dfrac{o(\|tp\|)}{t}=0$ 。
- 由函数极限的定义，自然有(2)成立。

附：函数极限的定义

- 设 $y = f(x)$在点 x_0 的某个去心邻域内有定义，即存在$\rho > 0$，使得 $O(x_0, \rho) \setminus \{x_0\} \subset D_f$。
- 如果存在实数 A，对于任意给定的$\varepsilon > 0$，可以找到$\delta > 0$，使得当 $0 < |x - x_0| < \delta$时，成立$|f(x) - A| < \varepsilon$，则称 A 是函数 $f(x)$ 在点 x_0 处的极限，记为 $\lim\limits_{x \to x_0} f(x) = A$ 。
- 如果不存在具有上述性质的实数 A，则称函数 $f(x)$在点 x_0 处的极限不存在。

局部最优解的必要条件

定理 4.4.2 设 $f:\mathrm{R}^n \mapsto \mathrm{R}$ 在点 $\bar{x}\in\mathrm{R}^n$ 处可微。若 $\boldsymbol{x}^*$ 是(UMP)的局部最优解，则 $\nabla \boldsymbol{f}(\boldsymbol{x}^*)=\boldsymbol{0}$。注意是必要条件，非充分条件

- 证：因为 $\boldsymbol{x}^*$是局部最优解，由定义，$\exists\delta>0$，
  $\forall x\in N_\delta\left(x^*\right)$，有 $f\left(x^*\right)\le f(x)$。 **(1)**
- 反证。假设 $\nabla f\left(x^*\right)^{\mathrm{T}}=\left(\dfrac{\partial f\left(x^*\right)}{\partial x_1},\dfrac{\partial f\left(x^*\right)}{\partial x_2},\cdots,\dfrac{\partial f\left(x^*\right)}{\partial x_n}\right)\neq 0$。不妨
  设 $\dfrac{\partial f\left(x^*\right)}{\partial x_1}\neq 0$。
- 令 $p^{\mathrm{T}}=\left(-\dfrac{\partial f\left(x^*\right)}{\partial x_1},0,\cdots,0\right)$。则 $\nabla f\left(x^*\right)^{\mathrm{T}}p<0$。

局部最优解的必要条件

- 由定理 **4.4.1**，可知 $\exists\delta' > 0$， $\forall t \in (0,\delta')$，有 $f(x^* + tp) < f(x^*)$。
- 取 t 充分小，可使 $x^* + tp \in N_\delta(x^*)$，与(1)矛盾。

局部最优解的充分条件

定理 4.4.3 设 $f:\mathrm{R}^n \mapsto \mathrm{R}$ 在点 $\bar{x} \in \mathrm{R}^n$ 处的 **Hesse** 矩阵 $\nabla^2 f(x^*)$ 存在。若 $\nabla f\left(x^*\right)=0$，并且 $\nabla^2 f\left(x^*\right)$ 正定，则 x^* 是(**UMP**)的严格局部最优解。

全局最优解的充分条件

定理 4.4.4 设 $f:\mathrm{R}^n \mapsto \mathrm{R}$， $x^* \in \mathrm{R}^n$，f 是 R^n 上的可微凸函数。若有 $\nabla f(x^*)=0$，则 x^* 是(UMP)的全局最优解。

- 证：因为 f 是 R^n 上的可微凸函数，由凸函数的判别定理 4.2.3，可知 $\forall x \in \mathrm{R}^n$，有 $\nabla f\left(x^*\right)\left(x-x^*\right) \le f(x)-f\left(x^*\right)$。
- 由于 $\nabla f(x^*)=0$，因此 $\forall x \in \mathrm{R}^n$， $0 \le f(x)-f\left(x^*\right)$，即 $f\left(x^*\right) \le f(x)$。
- 因此 x^* 是(UMP)的全局最优解。

例4.4.1

例 4.4.1 **求 UMP** $\min\ f(x)=x_1^2+4x_2^2+x_3^2-2x_1$ 的（全局）最优解。

- 解：先求驻点。$f(x)$的梯度向量 $\nabla f(x)=(2x_1-2,\ 8x_2,\ 2x_3)^{\mathrm{T}}$。

  令 $\nabla f(x)=0$，求得 f 的驻点 $\bar{x}=(1\quad 0\quad 0)^{\mathrm{T}}$。

- f 的 Hesse 矩阵 $\nabla^2 f(x)=\begin{pmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 2 \end{pmatrix}$，正定。由定理 4.2.4，$f$ 是（严格）凸函数。

- f 显然可微。再由定理 4.4.4，$\bar{x}=(1\quad 0\quad 0)^{\mathrm{T}}$ 就是 f 的全局最优解。

最速下降法

基本思想

- 设 **UMP** 问题中的目标函数 $f: \mathrm{R}^n \mapsto \mathrm{R}$ 一阶连续可微。
- 最速下降法（**Cauchy, 1867**）的基本思想是：从当前点 x^k 出发，沿 $f(x)$在 x^k 处下降最快的方向搜索下一个点 x^{k+1}，直到当前点处函数 $f(x)$的“下降程度”可以忽略不计（小于等于某个给定的ε > 0）。
- 什么是下降最快的方向？——负梯度方向。
- $f(x)$的 **Taylor** 展开式：$f(x+\Delta x) = f(x) + \nabla f(x)^{\mathrm{T}} \Delta x + o(\|\Delta x\|)$。

  因此，$f(x) - f(x+\Delta x) = -\nabla f(x)^{\mathrm{T}} \Delta x + o(\|\Delta x\|)$。

  忽略无穷小量 $o(\|\Delta x\|)$，可知Δx 取 $-\nabla f(x)$方向（内积最大），函数值下降最多。

最速下降法

（$\varepsilon > 0$ 为给定终止误差）

1 **选取初始点 $\boldsymbol{x}^0$，$\boldsymbol{k} \leftarrow \mathbf{1}$。**

2 **while** $\left\|\nabla f(x^k)\right\| > \varepsilon$ **do**

3 $p^k \leftarrow -\nabla f(x^k)$。

4 **一维搜索求 $\boldsymbol{t_k}$，使得** $f(x^k + t_k p^k) = \min\limits_{t \geq 0} f(x^k + tp^k)$。

5 $x^{k+1} \leftarrow x^k + t_k p^k$，$k \leftarrow k + 1$。

6 **endwhile**

7 **return $\boldsymbol{x}^k$。**

例4.4.2

例 4.4.2 用最速下降法求解 UMP 问题

$$\min \ f(x_1, x_2) = x_1^2 + 25x_2^2,$$

初始点 $\boldsymbol{x}^0 = (2, 2)^{\mathrm{T}}$，终止误差$\varepsilon = 10^{-6}$。

解： $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 50x_2 \end{pmatrix}$， $\left\|\nabla f(x^0)\right\| = \sqrt{4^2 + 100^2} >> 10^{-6}$。

$$p^0 = -\nabla f(x^0) = \begin{pmatrix} -4 \\ -100 \end{pmatrix},\quad x^0 + tp^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + t\begin{pmatrix} -4 \\ -100 \end{pmatrix} = \begin{pmatrix} 2-4t \\ 2-100t \end{pmatrix}。$$

$$f(x^0 + tp^0) = (2-4t)^2 + 25(2-100t)^2。$$

$$\frac{\mathrm{d}f(x^0 + tp^0)}{\mathrm{d}t} = -8(2-4t) - 5000(2-100t)。$$

例4.4.2

令 $\dfrac{\mathrm{d}f\left(x^0+tp^0\right)}{\mathrm{d}t}=0$，计算得 $t_0=\dfrac{10016}{500032}\approx 0.020037$。

$$x^1=x^0+t_0p^0=\begin{pmatrix}2\\2\end{pmatrix}+0.020037\begin{pmatrix}-4\\-100\end{pmatrix}=\begin{pmatrix}1.919878\\-0.003070\end{pmatrix}。$$

k	x^k	$\nabla f(x^k)$	$\lVert\nabla f(x^k)\rVert$	p^k	t_k
0	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}4\\100\end{pmatrix}$	**100.079968**	$\begin{pmatrix}-4\\-100\end{pmatrix}$	**0.020037**
1	$\begin{pmatrix}1.919852\\-0.0037\end{pmatrix}$	$\begin{pmatrix}3.839704\\-0.185\end{pmatrix}$	**3.844158**	$\begin{pmatrix}-3.839704\\0.185\end{pmatrix}$	**0.486464**
2	$\begin{pmatrix}0.051974\\0.086296\end{pmatrix}$	$\begin{pmatrix}0.103948\\4.3148\end{pmatrix}$	**4.316052**	…	…

经过 10 轮迭代，可得到满足误差 10^{-6} 的解。

相邻的两个搜索方向正交

- 考虑例 4.4.2 第 k 步，一维搜索时求到的 t_k 是 $f\left(x^k+tp^k\right)$ 的极小点，因此 $\dfrac{\mathrm{d}f\left(x^k+tp^k\right)}{\mathrm{d}t}=0$。
- 令 $x^k+tp^k=\left(x_1^k+tp_1^k,x_n^k+tp_n^k,\cdots,x_n^k+tp_n^k\right)=\left(u_1,u_2,\cdots,u_n\right)=u$，由复合函数求导法则，

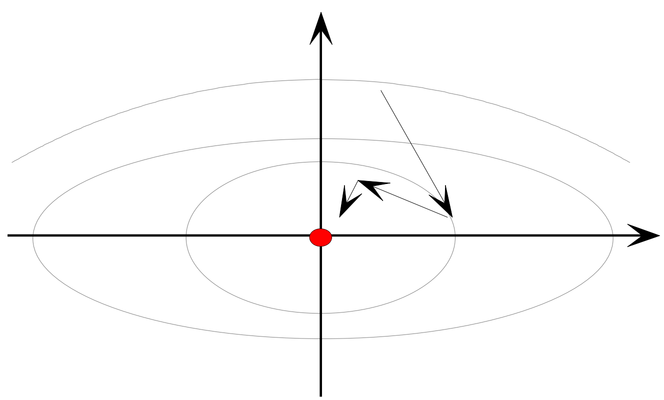
$$\frac{\mathrm{d}f\left(x^k+tp^k\right)}{\mathrm{d}t}=\frac{\partial f(u)}{\partial u_1}\frac{\mathrm{d}u_1}{\mathrm{d}t}+\frac{\partial f(u)}{\partial u_2}\frac{\mathrm{d}u_2}{\mathrm{d}t}+\cdots+\frac{\partial f(u)}{\partial u_n}\frac{\mathrm{d}u_n}{\mathrm{d}t}$$

$$=\frac{\partial f(u)}{\partial u_1}p_1^k+\frac{\partial f(u)}{\partial u_2}p_2^k+\cdots+\frac{\partial f(u)}{\partial u_n}p_n^k$$

$$=\nabla f(u)^{\mathrm{T}}p^k。$$

相邻的两个搜索方向正交

- 因此 $\nabla f\left(x^k+tp^k\right)^{\mathrm{T}}p^k=0$。
- 由于 $p^{k+1}=-\nabla f\left(x^k+tp^k\right)^{\mathrm{T}}$，这表明 $p^{k+1}p^k=0$。



例4.4.2的搜索路径示意图

关于最速下降法的说明

- 因此，“最速下降法”的搜索路径从整体上看是成直角锯齿状（当每个 t_k 都是极小点时）。
- 这表明最速下降法在每个 x^k 处是沿下降最快的方向搜索最优解的，但从整体上看这样做未必最快（“局部最快”不一定导致“全局最快”）。
- 最速下降法具有全局收敛性——无论从任何初始点 x^0 开始，所产生的点列均收敛。

共轭方向法

共轭方向

定义 4.4.1 设 A 为 n 阶实对称矩阵，对于非零向量 $p,q\in \mathrm{R}^n$，若有

$$p^{\mathrm{T}}Aq=0,$$

则称 p 和 q 是相互 A 共轭的。

对于非零向量组 $p^i\in \mathrm{R}^n$，$i=0,1,\cdots,n-1$，若有

$$\left(p^i\right)^{\mathrm{T}}Ap^j=0,\quad i,j=0,1,\cdots,n-1,\quad i\neq j$$

则称 $p^0,p^1,\cdots,p^{n-1}$ 是 A 共轭方向组，也称它们为一组 A 共轭方向。

二次严格凸函数的无约束最优化问题

二次严格凸函数的无约束最优化问题：

$$\min \quad f(x) = \frac{1}{2} x^{\mathrm{T}} A x + b^{\mathrm{T}} x + c \qquad \textbf{(AP)}$$

其中 A 是 n 阶实对称正定矩阵，$b \in \mathrm{R}^n$，$c \in \mathrm{R}$。

定理 4.4.6 对于问题(AP)，若 $p^0, p^1, \cdots, p^{n-1}$ 为任意一组 A 共轭方向，则由任意初始点 $x^0 \in \mathrm{R}^n$ 出发，依次沿 $p^0, p^1, \cdots, p^{n-1}$ 进行精确一维搜索，则最多经 n 次迭代可达到(AP)的整体最优解。

共轭方向法 与 共轭梯度法

- 从任意点$x^0 \in \mathbf{R}^n$出发，依次沿某组共轭方向进行一维搜索求解UMP的方法，叫做共轭方向法。
- 在共轭方向法中，若利用迭代点处的负梯度向量来产生一组共轭方向，则这样的方法叫做共轭梯度法。

F-R共轭梯度法

- 任取初始点 $x^0 \in \mathrm{R}^n$，$\boldsymbol{k} = \mathbf{0}$。一旦 $\nabla f\left(x^k\right)=0$，则停止（表明达到了目标函数的一个驻点）。否则，依次按下式确定第 **k** 次的搜索方向 $\boldsymbol{p}^k$：
- $p^0 = -\nabla f\left(x^0\right)$

  $p^{k+1} = -\nabla f\left(x^{k+1}\right) + \lambda_k p^k$，$\boldsymbol{k} = \mathbf{0, 1, \ldots,} \boldsymbol{n} - \mathbf{2}$，

  其中 $\lambda_k = \dfrac{\left(p^k\right)^{\mathrm{T}} A \nabla f\left(x^{k+1}\right)}{\left(p^k\right)^{\mathrm{T}} A p^k}$。
- 由 λ_k 的取法，易知 $\boldsymbol{p}^{k+1}$ 和 $\boldsymbol{p}^k$ 是 $\boldsymbol{A}$ 共轭的（$\left(p^{k+1}\right)^{\mathrm{T}} A p^k = 0$）。可以证明，在这种取法下，$\boldsymbol{p}^{k+1}$ 和 $\boldsymbol{p}^0, \boldsymbol{p}^1, \ldots, \boldsymbol{p}^k$ 也是 $\boldsymbol{A}$ 共轭的。

F-R共轭梯度法

- 任取初始点 $x^0 \in \mathrm{R}^n$，$\boldsymbol{k=0}$。一旦 $\nabla f\left(x^k\right)=0$，则停止（表明达到了目标函数的一个驻点）。否则，依次按下式确定第 **k** 次的搜索方向 $\boldsymbol{p^k}$：
- $p^0=-\nabla f\left(x^0\right)$

  $p^{k+1}=-\nabla f\left(x^{k+1}\right)+\lambda_k p^k$，$\boldsymbol{k=0, 1, \ldots, n-2}$，

  其中 $\lambda_k=\dfrac{\left(p^k\right)^{\mathrm{T}} A \nabla f\left(x^{k+1}\right)}{\left(p^k\right)^{\mathrm{T}} A p^k}$。
- 由 λ_k 的取法，易知 $\boldsymbol{p^{k+1}}$ 和 $\boldsymbol{p^k}$ 是 $\boldsymbol{A}$ 共轭的（$\left(p^{k+1}\right)^{\mathrm{T}} A p^k=0$）。可以证明，在这种取法下，$\boldsymbol{p^{k+1}}$ 和 $\boldsymbol{p^0, p^1, \ldots, p^k}$ 也是 $\boldsymbol{A}$ 共轭的。

F-R共轭梯度法

- 当求解的问题是 **AP** 问题时，计算搜索方向的参数λ_k可进一步简化为：$$\lambda_k = \frac{\left\|\nabla f\left(x^{k+1}\right)\right\|^2}{\left\|\nabla f\left(x^k\right)\right\|^2}$$，$k = 0, 1, \ldots, n-2$。

F-R共轭梯度法

[Fletcher, Reeves, 1964]

（参数 $\varepsilon > 0$ 为终止误差。）

选取初始点 $\boldsymbol{x}^0$。

while $\left\|\nabla f(x^0)\right\| > \varepsilon$ **do**

    $p^0 \leftarrow -\nabla f(x^0)$，$k \leftarrow 0$。

    while true do

        一维搜索求 $t_k = \arg\min\limits_{t \geq 0} f(x^k + tp^k)$。

        $x^{k+1} \leftarrow x^k + t_k p^k$。

        if $\left\|\nabla f(x^{k+1})\right\| \leq \varepsilon$ **then return** x^{k+1}。

        if $\boldsymbol{k+1=n}$ **then** $x^0 \leftarrow x^n$，退出循环。

F-R共轭梯度法

$$\lambda_k = \frac{\left\|\nabla f(x^{k+1})\right\|^2}{\left\|\nabla f(x^k)\right\|^2}，\quad p^{k+1} \leftarrow -\nabla f(x^{k+1}) + \lambda_k p^k 。$$

$\boldsymbol{k} \leftarrow \boldsymbol{k} + \mathbf{1}$。

endwhile

endwhile

return $\boldsymbol{x}^{\mathbf{0}}$。

例4.4.4

例 4.4.4 用 **F-R** 共轭梯度法求解 $\min\ f(x_1,x_2)=x_1^2+25x_2^2$。初始点 $\boldsymbol{x}^0=(2,2)^{\mathrm{T}}$，终止误差 10^{-6}（仅计算到 1）。

● 解：计算结果列表如下：

k	x^k	$\nabla f(x^k)$	$\|\nabla f(x^k)\|$	p^k	t^k	λ_k
0	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}4\\100\end{pmatrix}$	**100.079968**	$\begin{pmatrix}-4\\-100\end{pmatrix}$	**0.020037**	**0.001475**
1	$\begin{pmatrix}1.919852\\-0.0037\end{pmatrix}$	$\begin{pmatrix}3.839704\\-0.185\end{pmatrix}$	**3.844158**	$\begin{pmatrix}-3.845604\\0.0375\end{pmatrix}$	**0.498283**	
2	$\begin{pmatrix}0.003653\\0.014986\end{pmatrix}$	$\begin{pmatrix}0.007306\\0.7493\end{pmatrix}$	**0.749336**			

最后输出 $x=\begin{pmatrix}0.003653\\0.014986\end{pmatrix}$。

例4.4.4

若干中间计算公式：

- $\nabla f(x)=\begin{pmatrix} 2x_1 \\ 50x_2 \end{pmatrix}$，
- $f(x+tp)=(x_1+tp_1)^2+25(x_2+tp_2)^2$，

$$\frac{\mathrm{d}f(x+tp)}{\mathrm{d}t}=2(x_1+tp_1)p_1+50(x_2+tp_2)p_2$$

$$=(2p_1^2+50p_2^2)t+2p_1x_1+50p_2x_2。$$

例4.4.4

各步计算过程如下：

- $\boldsymbol{k=0}$。
- $x^0=\begin{pmatrix}2\\2\end{pmatrix}$，$\nabla f\left(x^0\right)=\begin{pmatrix}2x_1\\50x_2\end{pmatrix}=\begin{pmatrix}4\\100\end{pmatrix}$，$\left\|\nabla f\left(x^0\right)\right\|=100.079968$。
- $p^0=-\nabla f\left(x^0\right)=\begin{pmatrix}-4\\-100\end{pmatrix}$。
- 解 $\dfrac{\mathrm{d}f}{\mathrm{d}t}=\left(2p_1^2+50p_2^2\right)t+2p_1x_1+50p_2x_2=500032t-10016=0$，

  得 $t_0\approx 0.020037$。
- $x^1=x^0+t_0p^0=\begin{pmatrix}2\\2\end{pmatrix}+\begin{pmatrix}-0.080148\\-2.0037\end{pmatrix}=\begin{pmatrix}1.919852\\-0.0037\end{pmatrix}$

例4.4.4

- $\nabla f\left(x^1\right)=\begin{pmatrix}2x_1\\50x_2\end{pmatrix}=\begin{pmatrix}3.839704\\-0.185\end{pmatrix}$，

  $\left\|\nabla f\left(x_1\right)\right\|=\sqrt{3.839704^2+\left(-0.185\right)^2}=3.844158$

- $\lambda^0=\dfrac{\left\|\nabla f\left(x_1\right)\right\|^2}{\left\|\nabla f\left(x_0\right)\right\|^2}=\dfrac{3.839704^2+\left(-0.185\right)^2}{4^2+100^2}=0.001475$。
- $\boldsymbol{k=1}$。

  $p^1=-\nabla f\left(x^1\right)+\lambda_0 p^0$

- $=\begin{pmatrix}-3.839704\\0.185\end{pmatrix}+\begin{pmatrix}-0.0059\\-0.1475\end{pmatrix}=\begin{pmatrix}-3.845604\\0.0375\end{pmatrix}$。

例4.4.4

- $\frac{\mathrm{d}f}{\mathrm{d}t}=\left(2p_1^2+50p_2^2\right)t+2p_1x_1+50p_2x_2=29.647653t-14.772919$。

  解$\frac{\mathrm{d}f}{\mathrm{d}t}=0$，得$t_1=0.498283$。

- $x^2=x^1+t_1p^1=\begin{pmatrix}1.919852\\-0.0037\end{pmatrix}+\begin{pmatrix}-1.916199\\0.018686\end{pmatrix}=\begin{pmatrix}0.003653\\0.014986\end{pmatrix}$。

- $\nabla f\left(x^2\right)=\begin{pmatrix}2x_1\\50x_2\end{pmatrix}=\begin{pmatrix}0.007306\\0.7493\end{pmatrix}$，

  $\left\|\nabla f\left(x^2\right)\right\|=\sqrt{0.007306^2+0.7493^2}=0.749336$。

关于共轭梯度法的说明

- 满足一定条件时，具有全局收敛性。
- 运算简单，收敛速度快，存储量小。
- 依赖于一维精确搜索。

谢谢大家