

第十章 使用导数的最优化方法

-----研究无约束问题最优化方法

算法 $\left\{ \begin{array}{l} \text{解析法: 在计算过程中要用到目标函数的导数} \\ \text{直接法: 只用到目标函数值。} \end{array} \right.$

精确一维搜索的一个重要性质:

定理: 设 $f(x)$ 具有连续的偏导数, 且 $x^{(k+1)}$ 是从 $x^{(k)}$ 出发沿方向 $d^{(k)}$ 作一维搜索而得到的, 即

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda} f(x^{(k)} + \lambda d^{(k)}),$$

则 $\nabla f(x^{(k+1)})^T d^{(k)} = 0$ 。

最速下降法

最速下降方向

$$\min f(x)$$

$$s.t. \quad x \in E^n$$

$f(x)$ 具有一阶连续偏导数。

取搜索方向: $d^{(k)} = -\nabla f(x^{(k)})$

步骤:

1. 给定初点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$ 。
2. 计算搜索方向 $d^{(k)} = -\nabla f(x^{(k)})$ 。
3. 若 $\|d^{(k)}\| \leq \varepsilon$, 则停止计算; 否则, 从 $x^{(k)}$ 出发, 沿 $d^{(k)}$ 进行一维搜索, 求 λ_k , 使
$$f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda \geq 0} f(x^{(k)} + \lambda d^{(k)}).$$
4. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$, 置 $k := k + 1$, 返回2。

例：求 $\min f(x) = (x_1 - 1)^2 + (x_2 - 1)^2$, 取 $x^{(1)} = (0, 0)^T$, $\varepsilon = 0.1$.

解： $\nabla f(x) = (2(x_1 - 1), 2(x_2 - 1))^T$

第一次迭代

$$d^{(1)} = -(-2, -2)^T = (2, 2)^T, \|d^{(1)}\| = 2\sqrt{2} > \varepsilon$$

\therefore 从 $x^{(1)}$ 出发, 沿方向 $d^{(1)}$ 进行一维搜索, 求步长 λ_1 .

$$\min_{\lambda \geq 0} \varphi(\lambda) = \min_{\lambda \geq 0} f(x^{(1)} + \lambda d^{(1)})$$

$$\because x^{(1)} + \lambda d^{(1)} = (2\lambda, 2\lambda)^T, \therefore \varphi(\lambda) = (2\lambda - 1)^2 + (2\lambda - 1)^2.$$

$$\text{令 } \varphi'(\lambda) = 4(2\lambda - 1) + 4(2\lambda - 1) = 0, \text{ 得 } \lambda_1 = \frac{1}{2}$$

$$\text{令 } x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = (1, 1)^T.$$

第二次迭代

$$d^{(2)} = (0, 0)^T$$

$\because \|d^{(2)}\| = 0 < \varepsilon, \therefore (1, 1)^T$ 为最优解。

例：求 $\min f(x) = x_1^2 + 25x_2^2$, 取 $x^{(1)} = (2, 2)^T$, $\varepsilon = 0.2$.

k	$x^{(k)}$	$d^{(k)}$	λ_k	$\ d^{(k)}\ $
1	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	$\begin{pmatrix} -4 \\ -100 \end{pmatrix}$	0.02	100
2	$\begin{pmatrix} 1.92 \\ -0.003 \end{pmatrix}$	$\begin{pmatrix} -3.84 \\ 0.15 \end{pmatrix}$	0.482	3.84
3	$\begin{pmatrix} 0.07 \\ 0.07 \end{pmatrix}$	$\begin{pmatrix} -0.14 \\ -3.50 \end{pmatrix}$	0.02	3.5
4	$\begin{pmatrix} 0.067 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -0.134 \\ 0 \end{pmatrix}$		$0.134 < \varepsilon$

最速下降法的收敛性

定理：设 $f(x)$ 是连续可微实函数，解集合 $\Omega = \{\bar{x} \mid \nabla f(\bar{x}) = 0\}$,

最速下降算法产生的序列 $\{x^{(k)}\}$ 含于某个紧集，则序列

$\{x^{(k)}\}$ 的每个聚点 $\hat{x} \in \Omega$ 。

证明：最速下降算法 A 可表示为合成映射 $A = MD$

其中 $D(x) = (x, -\nabla f(x))$ 是 $E^n \rightarrow 2^{E^n \times E^n}$ 的映射，算法 M 是 $E^n \times E^n \rightarrow 2^{E^n}$ 的映射(即一维搜索)。

当 $d = -\nabla f(x) \neq 0$ 时，由定理 M 是闭映射。由于 $f(x)$ 是连续可微函数，所以 D 连续， $\Rightarrow A$ 在 $x(\nabla f(x) \neq 0)$ 处是闭的。

其次，当 $x \notin \Omega$ 时， $d = -\nabla f(x) \neq 0$ ，因此对于 $y \in A(x)$ ，必有 $f(y) < f(x)$ ，即 $f(x)$ 是关于 Ω 和 A 的下降函数。

由假设， $\{x^{(k)}\}$ 含于紧集中，所以算法收敛。

二次函数情形

$$f(x) = \frac{1}{2} x^T Q x - b^T x$$

其中 $Q_{n \times n}$ 是正定矩阵, 其特征值为 $0 < a = \lambda_1 \leq \dots \leq \lambda_n = A$

设 $f(x)$ 的唯一极小点为 x^* , 则 $Qx^* - b = 0$. 定义函数

$$E(x) = \frac{1}{2} (x - x^*)^T Q (x - x^*),$$

则 $E(x) = f(x) + \frac{1}{2} x^{*T} Q x^*$.

$$\nabla f(x) = \nabla E(x) = Qx - b \stackrel{\text{def}}{=} g(x)$$

最速下降法表示为

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} = x^{(k)} - \lambda_k g_k$$

$$\text{其中 } g_k = \nabla f(x^{(k)}) = Qx^{(k)} - b.$$

$$\because f(x^{(k)} - \lambda g_k) = \frac{1}{2} (x^{(k)} - \lambda g_k)^T Q (x^{(k)} - \lambda g_k) - b^T (x^{(k)} - \lambda g_k)$$

$$\therefore f(x^{(k)} - \lambda g_k) \text{ 在 } \lambda_k = \frac{g_k^T g_k}{g_k^T Q g_k} \text{ 处达到极小。}$$



$$x^{(k+1)} = x^{(k)} - \frac{g_k^T g_k}{g_k^T Q g_k} g_k$$

引理1 最速下降法满足

$$E\left(x^{(k+1)}\right)=\left[1-\frac{\left(g_k^T g_k\right)^2}{\left(g_k^T Q g_k\right)\left(g_k^T Q^{-1} g_k\right)}\right] E\left(x^{(k)}\right).$$

Kantorovich不等式

设 $Q_{n \times n}$ 是一个正定对称矩阵, 则对任意的 x , 有

$$\frac{\left(x^T x\right)^2}{\left(x^T Q x\right)\left(x^T Q^{-1} x\right)} \geq \frac{4 a A}{(a+A)^2}$$

其中 a 和 A 分别是 Q 的最小和最大特征值.

定理（最速下降法—二次情形）

对任意 $x^{(0)} \in E^n$ ，最速下降法产生的序列收敛于 $f(x)$ 的唯一极小点 x^* ，而且，对任意的 k ，有

$$E(x^{(k+1)}) \leq \left(\frac{A-a}{A+a} \right)^2 E(x^{(k)})$$

其中 $E(x) = \frac{1}{2} (x - x^*)^T Q (x - x^*)$.

非二次情形

定理: 设 $f(x)$ 存在连续二阶偏导数, \bar{x} 是局部极小点, $Hesse$ 矩阵 $\nabla^2 f(\bar{x})$ 的最小特征值 $a > 0$, 最大特征值为 A , 算法产生的序列 $\{x^{(k)}\}$ 收敛于点 \bar{x} , 则目标函数值的序列 $\{f(x^{(k)})\}$ 以不大于

$$\left(\frac{A-a}{A+a} \right)^2$$

的收敛比线性的收敛于 $f(\bar{x})$ 。

令 $r = \frac{A}{a}$, 则

条件数

$$\left(\frac{A-a}{A+a} \right)^2 = \left(\frac{r-1}{r+1} \right)^2 < 1.$$

结论：在相继两次迭代中,梯度方向互相正交.

证明：令

$$\varphi(\lambda) = f(x^{(k)} + \lambda d^{(k)})$$

$$d^{(k)} = -\nabla f(x^{(k)})$$

为求出从 $x^{(k)}$ 出发沿方向 $d^{(k)}$ 的极小点，令

$$\varphi'(\lambda_k) = \nabla f(x^{(k)} + \lambda_k d^{(k)})^T d^{(k)} = 0$$

得
$$-\nabla f(x^{(k+1)})^T \nabla f(x^{(k)}) = 0$$

即方向 $d^{(k+1)} = -\nabla f(x^{(k+1)})$ 与 $d^{(k)} = -\nabla f(x^{(k)})$ 正交。

牛顿法

- **基本思想**: 用一个二次函数去近似目标函数 $f(x)$, 然后精确地求出这个二次函数的极小点.
- ----- 一维搜索函数逼近法中的牛顿法的推广.

设 $x^{(k)}$ 是 $f(x)$ 的极小点 x^* 的第 k 次近似, 将 $f(x)$ 在 $x^{(k)}$ 点作二阶 $Taylor$ 展开, 得

$$f(x) \approx \varphi(x) = f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2} (x - x^{(k)})^T \nabla^2 f(x^{(k)}) (x - x^{(k)})$$

为求 $\varphi(x)$ 的极小点, 令

$$\nabla \varphi(x) = \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) (x - x^{(k)}) = 0$$

设 $\nabla^2 f(x^{(k)})$ 可逆, 则得

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}).$$

$$\text{令 } d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$



牛顿
方向

定理： 设 $f(x)$ 为二次可微函数， $x \in E^n$ ， \bar{x} 满足 $\nabla f(\bar{x}) = 0$ ，且 $\nabla^2 f(\bar{x})^{-1}$ 存在。又设 $x^{(1)}$ 充分接近 \bar{x} ，使得存在 $k_1, k_2 > 0$ ，满足 $k_1 k_2 < 1$ ，且对每一个

$$x \in X = \left\{ x \mid \|x - \bar{x}\| \leq \|x^{(1)} - \bar{x}\| \right\}$$

$$\|\nabla^2 f(x)^{-1}\| \leq k_1 \text{ 和 } \frac{\|\nabla f(\bar{x}) - \nabla f(x) - \nabla^2 f(x)(\bar{x} - x)\|}{\|\bar{x} - x\|} \leq k_2$$

成立，则牛顿法产生的序列收敛于 \bar{x} 。

证明： 牛顿算法映射定义为

$$A(x) = x - \nabla^2 f(x)^{-1} \nabla f(x)$$

定义解集合 $\Omega = \{\bar{x}\}$ ，令函数 $\alpha(x) = \|x - \bar{x}\|$ 。

下证 $\alpha(x)$ 是关于解集合 Ω 和算法 A 的下降函数。

$$\begin{aligned}
& \text{令 } x \in X, \text{ 且 } x \neq \bar{x}, \text{ 又令 } y \in A(x). \text{ 因为 } \nabla f(\bar{x}) = 0 \\
& \therefore y - \bar{x} = x - \nabla^2 f(x)^{-1} \nabla f(x) - \bar{x} \\
& = (x - \bar{x}) - \nabla^2 f(x)^{-1} [\nabla f(x) - \nabla f(\bar{x})] \\
& = \nabla^2 f(x)^{-1} [\nabla f(\bar{x}) - \nabla f(x) - \nabla^2 f(x)(\bar{x} - x)] \\
& \therefore \|y - \bar{x}\| \leq \|\nabla^2 f(x)^{-1}\| \|\nabla f(\bar{x}) - \nabla f(x) - \nabla^2 f(x)(\bar{x} - x)\| \\
& \leq k_1 k_2 \|\bar{x} - x\| < \|\bar{x} - x\|
\end{aligned}$$

因此 $\alpha(x)$ 为下降函数。

由定义 $y \in X$, 故迭代产生的序列 $\{x^{(k)}\} \subset X$ 。易知 X 为紧集, 因此迭代产生的序列含于紧集中。

又算法 A 在紧集 X 上是闭的,
所以迭代产生的序列 $\{x^{(k)}\}$ 必收敛于 \bar{x} 。

步骤:

1. 给定初点 $x^{(0)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 0$ 。

2. 若 $\|\nabla f(x^{(k)})\| < \varepsilon$, 则停止计算; 否则, 转3。

3. 计算

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}),$$

置 $k := k + 1$, 返回2。

$$\text{例 : 求 } \min f(x) = x_1^2 + 25x_2^2$$

解 : 取 $x^{(0)} = (2, 2)^T$, 则

$$\nabla f(x^{(0)}) = \begin{pmatrix} 2x_1 \\ 50x_2 \end{pmatrix} \Big|_{x^{(0)}} = \begin{pmatrix} 4 \\ 100 \end{pmatrix}$$

$$\nabla^2 f(x^{(0)}) = \begin{pmatrix} 2 & 0 \\ 0 & 50 \end{pmatrix} \quad \nabla^2 f(x^{(0)})^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{50} \end{pmatrix}$$

$$x^{(1)} = x^{(0)} - \nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)})$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{50} \end{pmatrix} \begin{pmatrix} 4 \\ 100 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

例:求 $\min f(x) = 4x_1^2 + x_2^2 - x_1^2 x_2$

分别取初始点为 $x_A = (1, 1)^T$, $x_B = (3, 4)^T$,

$x_C = (2, 0)^T$, 精度要求 $\varepsilon = 10^{-3}$.

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

$$\nabla f(x) = (8x_1 - 2x_1x_2, 2x_2 - x_1^2)^T$$

$$\nabla^2 f(x) = \begin{pmatrix} 8 - 2x_2 & -2x_1 \\ -2x_1 & 2 \end{pmatrix}.$$

(1) 取 $x^{(1)} = x_A = (1, 1)^T$, 则

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$\ \nabla f(x^{(k)})\ $	$\nabla^2 f(x^{(k)})$
1	1.0000 1.0000	4.000	6.0000 1.0000	6.0828	6.0000 -2.0000 -2.000 2.0000
2	-0.7500 -1.2500	4.5156	-7.8750 -3.0625	8.4495	10.500 1.5000 1.5000 2.0000
3	-0.1550 -0.1650	0.1273	-1.2911 -0.3540	1.3388	8.3300 0.3100 0.3100 2.0000
4	-0.0057 -0.0111	0.0003	-0.0459 -0.0223	0.0511	8.0222 0.0115 0.0115 2.0000
5	-0.0000 -0.0000	0.0000	-0.0001 -0.0000	0.0001	8.0000 0.0000 0.0000 2.0000

(2) 取 $x^{(1)} = x_B = (3, 4)^T$, 则

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$\ \nabla f(x^{(k)})\ $	$\nabla^2 f(x^{(k)})$
1	3.0000 4.0000	16.000	0.0000 -1.0000	1.0000	0.0000 -6.0000 -6.000 2.0000
2	2.8333 4.0000	16.0000	0.0000 -0.2078	0.0278	0.0000 -5.6667 -5.6667 2.0000
3	2.8284 4.0000	16.0000	0.0000 0.0000	0.0000	0.0000 -5.6569 -5.6569 2.0000

(3) 取 $x^{(1)} = x_C = (2, 0)^T$, 得到

$$\nabla^2 f(x^{(1)}) = \begin{pmatrix} 8 & -4 \\ -4 & 2 \end{pmatrix}$$

由于 $Hesse$ 矩阵不可逆, 无法进行下一步。

用**Newton**法求解无约束问题会出现以下情形:

(1) 收敛到极小点。

(2) 收敛到鞍点。

(3) **Hesse**矩阵不可逆, 无法迭代下去。

优点: (1) **Newton**法产生的点列 $\{x^{(k)}\}$ 若收敛, 则收敛速度快---具有至少二阶收敛速率。

(2) **Newton**法具有二次终止性。

证明: 设 A 为对称, 正定矩阵, 且

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

令 $\nabla f(x) = Ax + b = 0$, 得 $x^* = -A^{-1}b$.

若用 $Newton$ 迭代公式, 从任一点 $x^{(0)}$ 出发, 得

$$\begin{aligned} x^{(1)} &= x^{(0)} - \nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)}) \\ &= x^{(0)} - A^{-1}(Ax^{(0)} + b) = -A^{-1}b = x^*. \end{aligned}$$

缺点:

- (1) 可能会出现某步迭代时, 目标函数值上升.
- (2) 当初始点远离极小点时, 牛顿法产生的点列可能不收敛, 或者收敛到鞍点, 或者 **Hesse** 矩阵不可逆, 无法计算.
- (3) 需要计算 **Hesse** 矩阵的逆矩阵, 计算量大.

阻尼牛顿法

步骤:

1. 给定初点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$ 。

2. 计算 $\nabla f(x^{(k)})$, $\nabla^2 f(x^{(k)})^{-1}$ 。

3. 若 $\|\nabla f(x^{(k)})\| < \varepsilon$, 则停止迭代; 否则, 令

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}).$$

4. 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

$$\min_{\lambda} f(x^{(k)} + \lambda d^{(k)}) = f(x^{(k)} + \lambda_k d^{(k)}),$$

$$\text{令 } x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

5. 置 $k := k + 1$, 转2。

用阻尼牛顿法求解下列问题:

$$\min f(x) = x_1^4 + x_1x_2 + (1+x_2)^2$$

解: 取初始点 $x^{(1)} = (0, 0)^T$.

$$\nabla f(x^{(1)}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \nabla^2 f(x^{(1)}) = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$$

牛顿方向 $d^{(1)} = -\nabla^2 f(x^{(1)})^{-1} \nabla f(x^{(1)})$

$$= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

从 $x^{(1)}$ 出发, 沿 $d^{(1)}$ 作一维搜索, 令

$$\varphi(\lambda) = f(x^{(1)} + \lambda d^{(1)}) = 16\lambda^4 + 1$$

令 $\varphi'(\lambda) = 64\lambda^3 = 0$

则 $\lambda = 0$.

修正牛顿法

步骤:

1. 给定初点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.
2. 计算 $\nabla f(x^{(k)})$, $G_k = \nabla^2 f(x^{(k)})$ 。若 $\|\nabla f(x^{(k)})\| \leq \varepsilon$, 则停止计算, 得点 $x^{(k)}$; 否则转3。
3. 置 $B_k = G_k + \varepsilon_k I$, 其中 ε_k 是一个非负数, 选取 ε_k , 使得 B_k 是对称正定矩阵, 计算修正牛顿方向 $d^{(k)} = -B_k^{-1} \nabla f(x^{(k)})$.
4. 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:
$$\min_{\lambda} f(x^{(k)} + \lambda d^{(k)}) = f(x^{(k)} + \lambda_k d^{(k)}),$$
令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$
5. 置 $k := k + 1$, 转2。

共轭方向法

共轭方向

定义： 设 A 是 $n \times n$ 对称正定矩阵，若 E^n 中的两个方向 $d^{(1)}$ 和 $d^{(2)}$ 满足

$$(d^{(1)})^T A d^{(2)} = 0$$

则称这两个方向关于 A 共轭，或称它们关于 A 正交。

若 $d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 是 E^n 中 k 个方向，它们俩俩关于 A 共轭，即满足

$$(d^{(i)})^T A d^{(j)} = 0, i \neq j, i, j = 1, \dots, k$$

则称这组方向是 A 共轭的，或称它们为 A 的 k 个共轭方向。

例:

$$x^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad y^{(1)} = \begin{pmatrix} -\frac{1}{3} \\ \frac{2}{3} \end{pmatrix} \quad x^{(2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad y^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$x^{(3)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad y^{(3)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\text{则 } x^{(1)T} A y^{(1)} = (1 \ 0) \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -\frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = 0$$

定理1 设 A 是 n 阶对称正定矩阵, $d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 是 k 个 A 共轭的非零向量, 则这 k 个向量线性无关。

证明: 设存在数 $\alpha_1, \alpha_2, \dots, \alpha_k$, 使得

$$\alpha_1 d^{(1)} + \alpha_2 d^{(2)} + \dots + \alpha_k d^{(k)} = 0$$

左乘 $d^{(i)T} A$,

因为 $d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 是 k 个 A 共轭的非零向量,

$$\text{所以有 } \alpha_i d^{(i)T} A d^{(i)} = 0,$$

又因为 A 正定, 所以 $d^{(i)T} A d^{(i)} > 0$

$$\Rightarrow \alpha_i = 0$$

$\therefore d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 线性无关。

定理2: 设有二次函数

$$f(x) = \frac{1}{2} x^T A x + b^T x + c,$$

其中 $A_{n \times n}$ 是对称正定矩阵, $d^{(0)}, d^{(1)}, \dots, d^{(n-1)}$ 是 A 共轭的非零向量, 从任意一点 $x^{(0)} \in E^n$ 出发, 依次沿这组向量进行一维搜索,

$$\min_{\lambda \geq 0} f(x^{(k)} + \lambda d^{(k)}) = f(x^{(k)} + \lambda_k d^{(k)})$$

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} \quad k = 0, 1, \dots, n-1$$

则 $\nabla f(x^{(k+1)})^T d^{(j)} = 0, j = 0, 1, \dots, k$ 。

证明: $\because \nabla f(x) = Ax + b$

$$\therefore \nabla f(x^{(k+1)}) = Ax^{(k+1)} + b = A(x^{(k)} + \lambda_k d^{(k)}) + b$$

$$= Ax^{(k)} + \lambda_k Ad^{(k)} + b$$

$$= A(x^{(k-1)} + \lambda_{k-1} d^{(k-1)}) + \lambda_k Ad^{(k)} + b$$

$$= Ax^{(k-1)} + \lambda_{k-1} Ad^{(k-1)} + \lambda_k Ad^{(k)} + b = \dots$$

$$= Ax^{(j+1)} + \sum_{i=j+1}^k \lambda_i Ad^{(i)} + b = \nabla f(x^{(j+1)}) + \sum_{i=j+1}^k \lambda_i Ad^{(i)}$$

$$\therefore \nabla f(x^{(k+1)})^T = \nabla f(x^{(j+1)})^T + \sum_{i=j+1}^k \lambda_i d^{(i)T} A^T$$

$$\text{右乘 } d^{(j)} : \nabla f(x^{(k+1)})^T d^{(j)} = \nabla f(x^{(j+1)})^T d^{(j)} + \sum_{i=j+1}^k \lambda_i d^{(i)T} A^T d^{(j)}$$

$\because \nabla f(x^{(j+1)})^T d^{(j)} = 0, d^{(0)}, d^{(1)}, \dots, d^{(n-1)}$ 是 A 共轭的

$$\therefore \nabla f(x^{(k+1)})^T d^{(j)} = 0, \quad j = 0, 1, \dots, k.$$

定理（扩张子空间定理,expanding subspace theorem）

设有函数
$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

其中 A 为 n 阶对称正定矩阵, $d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 是 A 共轭的非零向量。以任意的 $x^{(1)} \in E^n$ 为初始点, 依次沿 $d^{(1)}, d^{(2)}, \dots, d^{(k)}$ 进行一维搜索, 得到点 $x^{(2)}, x^{(3)}, \dots, x^{(k+1)}$, 则 $x^{(k+1)}$ 是 $f(x)$ 在线性流形

$$\begin{aligned} M_k \left(x^{(1)}; \{d^{(1)}, d^{(2)}, \dots, d^{(k)}\} \right) &= \left\{ x \left| x = x^{(1)} + \sum_{i=1}^k \mu_i d^{(i)}, \mu_i \in R \right. \right\} \\ &= x^{(1)} + B_k \end{aligned}$$

上的唯一极小点。特别的, 当 $k = n$ 时, $x^{(n+1)}$ 是 $f(x)$ 在 E^n 上的唯一极小点。

证明：由于 $f(x)$ 是严格凸函数，所以对 $\forall y \in M_k$,有

$$f(y) > f(x^{(k+1)}) + \nabla f(x^{(k+1)})^T (y - x^{(k+1)}).$$

$$\because y, x^{(k+1)} \in M_k, \therefore \text{可设 } y = x^{(1)} + \sum_{i=1}^k \mu_i d^{(i)}, x^{(k+1)} = x^{(1)} + \sum_{i=1}^k \alpha_i d^{(i)}$$

$$f(y) > f(x^{(k+1)}) + \nabla f(x^{(k+1)})^T (x^{(1)} - x^{(k+1)}) + \sum_{i=1}^k \mu_i \left[\nabla f(x^{(k+1)})^T d^{(i)} \right]$$

$$= f(x^{(k+1)}) - \sum_{i=1}^k \alpha_i \left[\nabla f(x^{(k+1)})^T d^{(i)} \right] + \sum_{i=1}^k \mu_i \left[\nabla f(x^{(k+1)})^T d^{(i)} \right]$$

若能证明 $\nabla f(x^{(k+1)})^T d^{(i)} = 0$ ，则 $x^{(k+1)}$ 是 M_k 的极小点。

要证明 $x^{(k+1)}$ 是 $f(x)$ 在 $x^{(1)} + B_k$ 上的唯一极小点，只

需证明在 $x^{(k+1)}$ 处， $\nabla f(x^{(k+1)})$ 与子空间 B_k 正交。

$k=1$ 时, 由一维搜索定义知 $\nabla f(x^{(2)}) \perp B_1$.

假设 $k=m < n$ 时, $\nabla f(x^{(m+1)}) \perp B_m$, 下证 $\nabla f(x^{(m+2)}) \perp B_{m+1}$ 。

注意 $\nabla f(x^{(m+2)}) = Ax^{(m+2)} + b$, $x^{(m+2)} = x^{(m+1)} + \lambda_{m+1}d^{(m+1)}$

$$\begin{aligned}\therefore \nabla f(x^{(m+2)}) &= Ax^{(m+2)} + b \\ &= A(x^{(m+1)} + \lambda_{m+1}d^{(m+1)}) + b \\ &= \nabla f(x^{(m+1)}) + \lambda_{m+1}Ad^{(m+1)}\end{aligned}$$

$$\Rightarrow d^{(i)T} \nabla f(x^{(m+2)}) = d^{(i)T} \nabla f(x^{(m+1)}) + \lambda_{m+1}d^{(i)T} Ad^{(m+1)}$$

当 $i=m+1$, 由一维搜索定义有 $d^{(i)T} \nabla f(x^{(m+2)}) = 0$,

当 $1 \leq i \leq m$, 由归纳假设, 有 $d^{(i)T} \nabla f(x^{(m+1)}) = 0$ 。

由于 $d^{(1)}, \dots, d^{(m+1)}$ 关于 A 共轭, 因此有

$$d^{(i)T} Ad^{(m+1)} = 0$$

$$\Rightarrow d^{(i)T} \nabla f(x^{(m+2)}) = 0 \Rightarrow \nabla f(x^{(m+2)}) \perp B_{m+1}.$$

当 $k = n$ 时, $d^{(1)}, d^{(2)}, \dots, d^{(n)}$ 是 E^n 的一组基,

若 $\nabla f(x^{(n+1)}) \neq 0$, 令

$$\nabla f(x^{(n+1)}) = \alpha_1 d^{(1)} + \alpha_2 d^{(2)} + \dots + \alpha_n d^{(n)}$$

两边左乘 $\nabla f(x^{(n+1)})^T$, 得

$$\nabla f(x^{(n+1)})^T \nabla f(x^{(n+1)}) = \alpha_1 \nabla f(x^{(n+1)})^T d^{(1)} + \dots + \alpha_n \nabla f(x^{(n+1)})^T d^{(n)}$$

$$\Rightarrow \nabla f(x^{(n+1)})^T \nabla f(x^{(n+1)}) = 0, \text{ 矛盾,}$$

所以, $\nabla f(x^{(n+1)}) = 0$,

即 $x^{(n+1)}$ 是函数 $f(x)$ 在 E^n 上的唯一极小点。

共轭方向法

从任意点出发，依次沿某组共轭方向进行一维搜索求解非线性规划问题的方法。

步骤（适用于正定二次函数）

1. 给定初始点 $x^{(0)} \in E^n$ ，选定下降方向 $d^{(0)}$ ，令 $k = 0$ 。

2. 从 $x^{(k)}$ 出发，沿 $d^{(k)}$ 方向进行一维搜索，求 λ_k ：

$$\min f(x^{(k)} + \lambda d^{(k)}) = x^{(k)} + \lambda_k d^{(k)}$$

令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ ，计算 $\nabla f(x^{(k+1)})$ 。

3. 若 $\|\nabla f(x^{(k+1)})\| = 0$ ，则停止迭代，否则转 4。

4. 选定共轭方向 $d^{(k+1)}$ ，使满足

$$d^{(j)T} A d^{(k+1)} = 0, j = 0, 1, \dots, k.$$

令 $k = k + 1$ ，返回 2。

结论: $\lambda_k = -\frac{d^{(k)T} (Ax^{(0)} + b)}{d^{(k)T} Ad^{(k)}} = -\frac{\nabla f(x^{(k)})^T d^{(k)}}{d^{(k)T} Ad^{(k)}}.$

证明: 设 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$, 则有

$$\begin{aligned} 0 &= \nabla f(x^{(k+1)})^T d^{(k)} = d^{(k)T} \nabla f(x^{(k+1)}) = d^{(k)T} (Ax^{(k+1)} + b) \\ &= d^{(k)T} (Ax^{(k)} + \lambda_k Ad^{(k)} + b) \Rightarrow \lambda_k = -\frac{\nabla f(x^{(k)})^T d^{(k)}}{d^{(k)T} Ad^{(k)}} \end{aligned}$$

$$\begin{aligned} \text{又由} 0 &= d^{(k)T} (Ax^{(k)} + \lambda_k Ad^{(k)} + b) \\ \Rightarrow d^{(k)T} (Ax^{(k)} + b) &= -\lambda_k d^{(k)T} Ad^{(k)} \end{aligned}$$

将 $x^{(k)} = x^{(0)} + \sum_{i=0}^{k-1} \lambda_i d^{(i)}$ 代入上式, 且 $d^{(i)}$ 共轭, 得

$$d^{(k)T} (Ax^{(0)} + b) = -\lambda_k d^{(k)T} Ad^{(k)} \Rightarrow \lambda_k = -\frac{d^{(k)T} (Ax^{(0)} + b)}{d^{(k)T} Ad^{(k)}}$$

共轭方向法

步骤（适用于正定二次函数）

1. 给定初始点 $x^{(0)} \in E^n$ ，选定下降方向 $d^{(0)}$ ，令 $k = 0$ 。

2. 计算 $\lambda_k = -\frac{d^{(k)T} (Ax^{(0)} + b)}{d^{(k)T} Ad^{(k)}}$,

令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ ，计算 $\nabla f(x^{(k+1)})$ 。

3. 若 $\|\nabla f(x^{(k+1)})\| = 0$ ，则停止迭代，否则转4。

4. 选定共轭方向 $d^{(k+1)}$ ，使满足

$$d^{(j)T} Ad^{(k+1)} = 0, j = 0, 1, \dots, k.$$

令 $k = k + 1$ ，返回2。

共轭方向法

步骤（适用于正定二次函数）

1. 给定初始点 $x^{(0)} \in E^n$ ，选定下降方向 $d^{(0)}$ ，令 $k = 0$ 。

2. 计算
$$\lambda_k = -\frac{\nabla f(x^{(k)})^T d^{(k)}}{d^{(k)T} A d^{(k)}},$$

令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ ，计算 $\nabla f(x^{(k+1)})$ 。

3. 若 $\|\nabla f(x^{(k+1)})\| = 0$ ，则停止迭代，否则转 4。

4. 选定共轭方向 $d^{(k+1)}$ ，使满足

$$d^{(j)T} A d^{(k+1)} = 0, j = 0, 1, \dots, k.$$

令 $k = k + 1$ ，返回 2。

例: $\min x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2$

$$A = \text{diag}(2, 1, 1), x^{(0)} = (1, 1, 1)^T,$$

$$d^{(0)} = (-1, -2, 0)^T, \nabla f = (2x_1, x_2, x_3)^T$$

k	$x^{(k)}$	$d^{(k)}$	$\nabla f(x^{(k)})$	λ_k
0	$(1, 1, 1)^T$	$(-1, -2, 0)^T$	$(2, 1, 1)^T$	$\frac{2}{3}$
1	$\left(\frac{1}{3}, -\frac{1}{3}, 1\right)^T$	$(1, -1, 0)^T$	$\left(\frac{2}{3}, -\frac{1}{3}, 1\right)^T$	$-\frac{1}{3}$
2	$(0, 0, 1)^T$	$(0, 0, 1)^T$	$(0, 0, 1)^T$	-1
3	$(0, 0, 0)^T$		$(0, 0, 0)^T$	

共轭梯度法(Conjugate Gradient Method)(FR法)

记号: $\nabla f(x^{(k)}) = g_k$

在共轭梯度法中, 初始点处的搜索方向取为该点的负梯度方向, 即取

$$d^{(1)} = -\nabla f(x^{(1)}) = -g_1$$

而以下各共轭方向 $d^{(k)}$ 由第 k 次迭代点 $x^{(k)}$ 处的负梯度 $-g_k$ 与已经得到的共轭向量 $d^{(k-1)}$ 的线性组合来确定。

$$\min f(x) = \frac{1}{2} x^T A x + b^T x + c$$

选取初始点 $x^{(1)}$, 取 $d^{(1)} = -g_1$, 从 $x^{(1)}$ 出发, 沿 $d^{(1)}$ 做一维搜索,

$$\lambda_1 = -\frac{\nabla f(x^{(1)})^T d^{(1)}}{d^{(1)T} A d^{(1)}} = \frac{g_1^T g_1}{d^{(1)T} A d^{(1)}}, \quad x^{(2)} = x^{(1)} + \lambda_1 d^{(1)}$$

当 $x^{(1)} \neq x^*$ 时, $g_1 \neq 0 \Rightarrow \lambda_1 \neq 0$ 。

若 $x^{(2)} \neq x^*$ (否则迭代终止), 则 $g_2^T d^{(1)} = 0 \therefore g_2$ 与 $d^{(1)}$ 线性无关。

令 $d^{(2)} = -g_2 + \beta_1 d^{(1)}$

用 $d^{(1)T} A$ 左乘上式, 并令 $d^{(1)T} A d^{(2)} = 0$, 得

$$\beta_1 = \frac{d^{(1)T} A g_2}{d^{(1)T} A d^{(1)}} \Rightarrow d^{(2)} = -g_2 + \frac{d^{(1)T} A g_2}{d^{(1)T} A d^{(1)}} d^{(1)}$$

从 $x^{(2)}$ 出发, 沿 $d^{(2)}$ 做一维搜索, 得

$$\lambda_2 = -\frac{\nabla f(x^{(2)})^T d^{(2)}}{d^{(2)T} A d^{(2)}} = \frac{g_2^T g_2 - \beta_1 \nabla f(x^{(2)})^T d^{(1)}}{d^{(2)T} A d^{(2)}} = \frac{g_2^T g_2}{d^{(2)T} A d^{(2)}}$$

以此类推，得

$$d^{(k)} = -g_k + \beta_{k-1}d^{(k-1)}$$

$$\beta_{k-1} = \frac{d^{(k-1)T} A g_k}{d^{(k-1)T} A d^{(k-1)}}$$

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

$$\lambda_k = \frac{g_k^T g_k}{d^{(k)T} A d^{(k)}}$$

定理：对正定二次函数，FR法在 $m \leq n$ 次一维搜索后终止，且对 $\forall i(1 \leq i \leq m)$ ，下列关系式成立：

$$(1) d^{(i)T} A d^{(j)} = 0, j = 1, 2, \dots, i-1;$$

$$(2) g_i^T g_j = 0, j = 1, 2, \dots, i-1;$$

$$(3) g_i^T d^{(i)} = -g_i^T g_i. (\text{蕴含 } d^{(i)} \neq 0)$$

$$d^{(k)} = -g_k + \beta_{k-1} d^{(k-1)}$$

$$\beta_{k-1} = \frac{d^{(k-1)T} A g_k}{d^{(k-1)T} A d^{(k-1)}}$$

证明：（用归纳法）

$i = 1$ 时， $g_1 = \nabla f(x^{(1)})$, $d^{(1)} = -g_1$, $\therefore (3)$ 成立。

$i = 2$ 时，由构造法， $d^{(2)T} A d^{(1)} = 0$ ，即(1)成立。

由一维搜索性质， $\nabla f(x^{(2)})^T d^{(1)} = 0$ ，而

$$d^{(1)} = -g_1 \Rightarrow g_2^T g_1 = 0 \Rightarrow (2) \text{成立}。$$

$$\therefore d^{(2)} = -g_2 + \beta_1 d^{(1)}$$

$$\therefore g_2^T d^{(2)} = g_2^T (-g_2 + \beta_1 d^{(1)}) = -g_2^T g_2 \therefore (3) \text{成立}。$$

假设对某个 $3 \leq i < m$, (1), (2), (3) 均成立, 下证对 $i+1$ 也成立。

先证(2)

$$(2) \ x^{(i+1)} = x^{(i)} + \lambda_i d^{(i)}$$

$$\nabla f(x^{(i+1)}) = Ax^{(i+1)} + b = Ax^{(i)} + \lambda_i Ad^{(i)} + b = \nabla f(x^{(i)}) + \lambda_i Ad^{(i)}$$

$$\text{即 } g_{i+1} = g_i + \lambda_i Ad^{(i)}$$

$$\text{其中 } \lambda_i = \frac{g_i^T g_i}{d^{(i)T} Ad^{(i)}} \stackrel{\text{归纳假设}}{=} \frac{-g_i^T d^{(i)}}{d^{(i)T} Ad^{(i)}} \neq 0$$

$$(1) \ d^{(i)T} Ad^{(j)} = 0$$

$$(2) \ g_i^T g_j = 0$$

$$(3) \ g_i^T d^{(i)} = -g_i^T g_i.$$

$$\therefore g_{i+1}^T g_j = (g_i^T + \lambda_i d^{(i)T} A^T) g_j$$

$$= g_i^T g_j + \lambda_i d^{(i)T} A (-d^{(j)} + \beta_{j-1} d^{(j-1)})$$

$$= g_i^T g_j - \lambda_i d^{(i)T} Ad^{(j)} + \lambda_i \beta_{j-1} d^{(i)T} Ad^{(j-1)}$$

$$d^{(k)} = -g_k + \beta_{k-1} d^{(k-1)}$$

$$\beta_{k-1} = \frac{d^{(k-1)T} Ag_k}{d^{(k-1)T} Ad^{(k-1)}}$$

$$= \begin{cases} 0 & i \neq j \\ g_i^T g_j - g_i^T g_j + \lambda_i \beta_{j-1} d^{(i)T} Ad^{(j-1)} = 0 & i = j \end{cases}$$

$$(1) d^{(i+1)} = -g_{i+1} + \beta_i d^{(i)}$$

$$d^{(i+1)T} Ad^{(j)} = \left(-g_{i+1} + \beta_i d^{(i)}\right)^T Ad^{(j)} = -g_{i+1}^T Ad^{(j)} + \beta_i d^{(i)T} Ad^{(j)}$$

若 $i = j$, 由 $\beta_i = \frac{d^{(i)T} Ag_{i+1}}{d^{(i)T} Ad^{(i)}} \Rightarrow d^{(i+1)T} Ad^{(i)} = 0$, 以下假设 $i \neq j$ 。

$$\because g_{i+1} - g_i = (Ax^{(i+1)} + b - (Ax^{(i)} + b))$$

$$= A(x^{(i+1)} - x^{(i)}) = A(x^{(i)} + \lambda_i d^{(i)} - x^{(i)}) = \lambda_i Ad^{(i)}$$

$$\therefore Ad^{(i)} = \frac{g_{i+1} - g_i}{\lambda_i}$$

$$\therefore d^{(i+1)T} Ad^{(j)} = -g_{i+1}^T \frac{g_{j+1} - g_j}{\lambda_j} + \beta_i d^{(i)T} Ad^{(j)}$$

$$= -\frac{1}{\lambda_j} g_{i+1}^T g_{j+1} + \frac{1}{\lambda_j} g_{i+1}^T g_j + \beta_i d^{(i)T} Ad^{(j)} = 0$$

$$(1) d^{(i)T} Ad^{(j)} = 0$$

$$(2) g_i^T g_j = 0$$

$$(3) g_i^T d^{(i)} = -g_i^T g_i.$$

$$\begin{aligned}
(3) \quad g_{i+1}^T d^{(i+1)} &= g_{i+1}^T \left(-g_{i+1} + \beta_i d^{(i)} \right) \\
&= -g_{i+1}^T g_{i+1} + \beta_i g_{i+1}^T d^{(i)} \\
&= -g_{i+1}^T g_{i+1}.
\end{aligned}$$

$$(1) \quad d^{(i)T} A d^{(j)} = 0$$

$$(2) \quad g_i^T g_j = 0$$

$$(3) \quad g_i^T d^{(i)} = -g_i^T g_i.$$

定理: $\beta_i = \frac{d^{(i)T} A g_{i+1}}{d^{(i)T} A d^{(i)}} = \frac{\|g_{i+1}\|^2}{\|g_i\|^2} (i \geq 1, g_i \neq 0).$

证明: 由 $A d^{(j)} = \frac{1}{\lambda_j} (g_{j+1} - g_j)$ 得

$$\beta_i = \frac{\frac{1}{\lambda_i} (g_{i+1} - g_i)^T g_{i+1}}{d^{(i)T} \frac{1}{\lambda_i} (g_{i+1} - g_i)} = \frac{g_{i+1}^T g_{i+1} - g_i^T g_{i+1}}{d^{(i)T} g_{i+1} - d^{(i)T} g_i}$$

$$= \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}$$

$$(1) d^{(i)T} A d^{(j)} = 0$$

$$(2) g_i^T g_j = 0$$

$$(3) g_i^T d^{(i)} = -g_i^T g_i.$$

步骤(FR共轭梯度法)

1. 给定初始点 $x^{(1)}$, 置 $k = 1$ 。
2. 计算 $g_k = \nabla f(x^{(k)})$, 若 $\|g_k\| = 0$, 则停止计算, 得点 $\bar{x} = x^{(k)}$; 否则进行下一步。
3. 令 $d^{(k)} = -g_k + \beta_{k-1}d^{(k-1)}$, 其中, 当 $k = 1$ 时,
$$\beta_{k-1} = 0, \text{ 当 } k > 1 \text{ 时, } \beta_{k-1} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}。$$
4. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$, 其中 $\lambda_k = -\frac{g_k^T d^{(k)}}{d^{(k)T} A d^{(k)}}。$
5. 若 $k = n$, 则停止计算, 得点 $\bar{x} = x^{(k+1)}$; 否则, 置 $k = k + 1$, 返回2。

例: $\min x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2$

解: $\nabla f(x) = (2x_1, x_2, x_3)^T$, 取 $x^{(1)} = (1, 1, 1)^T$.

k	$x^{(k)}$	g_k	β_{k-1}	$d^{(k)}$	λ_k
1	$(1, 1, 1)^T$	$(2, 1, 1)^T$	0	$-(2, 1, 1)^T$	$\frac{3}{5}$
2	$\left(-\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right)^T$	$\left(-\frac{2}{5}, \frac{2}{5}, \frac{2}{5}\right)^T$	$\frac{2}{25}$	$\frac{6}{25}(1, -2, -2)^T$	$\frac{5}{6}$
3	$(0, 0, 0)^T$				

例: $\min x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2$

$$A = \text{diag}(2, 1, 1), x^{(1)} = (1, 1, 1)^T,$$

$$d^{(1)} = (-1, -2, 0)^T, \nabla f = (2x_1, x_2, x_3)^T$$

从 $x^{(1)}$ 出发, 沿方向 $d^{(1)}$ 做一维搜索, 得 $\lambda_1 = \frac{2}{3}$,

$$\Rightarrow x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \left(\frac{1}{3}, -\frac{1}{3}, 1\right)^T, g_2 = \left(\frac{2}{3}, -\frac{1}{3}, 1\right)^T.$$

$$\text{令 } d^{(2)} = -g_2 + \beta_1 d^{(1)}, \text{ 其中 } \beta_1 = \frac{d^{(1)T} A g_2}{d^{(1)T} A d^{(1)}} = -\frac{1}{9}$$

$$\Rightarrow d^{(2)} = \left(-\frac{5}{9}, \frac{5}{9}, -1\right)^T$$

从 $x^{(2)}$ 出发, 沿方向 $d^{(2)}$ 做一维搜索, 得 $\lambda_2 = \frac{21}{26}$,

$$\Rightarrow x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \left(-\frac{9}{78}, \frac{9}{78}, \frac{5}{26} \right)^T,$$

$$g_3 = \left(-\frac{18}{78}, \frac{9}{78}, \frac{5}{26} \right)^T.$$

$$\text{令 } d^{(3)} = -g_3 + \beta_2 d^{(2)}, \text{ 其中 } \beta_2 = \frac{d^{(2)T} A g_3}{d^{(2)T} A d^{(2)}} = \frac{45}{676}$$

$$\Rightarrow d^{(3)} = \frac{1}{676} (131, -53, -175)^T$$

结论: $d^{(1)}$ 与 $d^{(2)}$ 关于 A 共轭, $d^{(2)}$ 与 $d^{(3)}$ 关于 A 共轭,
但 $d^{(1)}$ 与 $d^{(3)}$ 关于 A 不共轭。

用于一般函数的共轭梯度法

与原方法的主要区别:

(1) 步长 λ_k 不能再用公式 $\lambda_k = -\frac{\mathbf{g}_k^T \mathbf{d}^{(k)}}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}}$ 计算,

必须用其他一维搜索方法来确定。

(2) 凡用到矩阵 A 之处, 需用现行点处的 $Hessian$ 矩阵 $\nabla^2 f(\mathbf{x}^{(k)})$ 代替。

(3) 有限步迭代达不到极小点。

迭代的延续方法:

(1) 直接延续: 即总用公式 $d^{(k+1)} = -g_{k+1} + \beta_k d^{(k)}$ 构造搜索方向。

(2) 重新开始, 把 n 步作为一轮, 每搜索一轮之后, 取一次最速下降方向, 开始下一轮。

计算参数 β_k 的常用公式:

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T \nabla f(x^{(k+1)})}{\nabla f(x^{(k)})^T \nabla f(x^{(k)})}$$

Fletcher – Reeves (FR)

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T (\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}))}{\nabla f(x^{(k)})^T \nabla f(x^{(k)})}$$

Polak – Ribiere – Polyak (PRP)

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T (\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}))}{d^{(k)T} (\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}))}$$

Crowder – Wolfe

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T \nabla^2 f(x^{(k+1)}) d^{(k)}}{d^{(k)T} \nabla^2 f(x^{(k+1)}) d^{(k)}}$$

Daniel

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T \nabla f(x^{(k+1)})}{d^{(k)T} \nabla f(x^{(k)})}$$

Dixon

$$\beta_k = \frac{\nabla f(x^{(k+1)})^T \nabla f(x^{(k+1)})}{d^{(k)T} (\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}))}$$

Dai – Yuan

步骤(FR共轭梯度法)

1. 给定初始点 $x^{(1)}$, 允许误差 $\varepsilon > 0$, 置 $y^{(1)} = x^{(1)}$,
 $d^{(1)} = -\nabla f(y^{(1)})$, $k = j = 0$ 。

2. 若 $\|\nabla f(y^{(j)})\| < \varepsilon$, 则停止计算; 否则作一维搜索, 求 λ_j

满足 $f(y^{(j)} + \lambda_j d^{(j)}) = \min f(y^{(j)} + \lambda d^{(j)})$,

令 $y^{(j+1)} = y^{(j)} + \lambda_j d^{(j)}$ 。

3. 若 $j < n$, 则转4, 否则转5。

4. 令 $d^{(j+1)} = -\nabla f(y^{(j+1)}) + \beta_j d^{(j)}$, $\beta_j = \frac{\|\nabla f(y^{(j+1)})\|^2}{\|\nabla f(y^{(j)})\|^2}$,

置 $j = j + 1$, 转2。

5. 令 $x^{(k+1)} = y^{(n+1)}$, $y^{(1)} = x^{(k+1)}$, $d^{(1)} = -\nabla f(y^{(1)})$, $j = 1$

$k = k + 1$, 返回2。

变尺度法(Variable Metric Method) 拟牛顿法(Quasi-Newton Method)

这是一种求解无约束极值问题的有效算法，由于它既避免了计算二阶导数、矩阵及其求逆过程，又比最速下降法的收敛速度快，特别是对高维问题具有显著的优越性，所以，它被公认为求解无约束极值问题最有效的算法之一。

牛顿法的缺点：

- **(1)** 可能会出现某步迭代时，目标函数值上升。
- **(2)** 当初始点远离极小点时，牛顿法产生的点列可能不收敛，或者收敛到鞍点，或者**Hesse**矩阵不可逆，无法计算。
- **(3)** 需要计算**Hesse**矩阵的逆矩阵，计算量大。

基本原理:

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

阻尼牛顿法: $d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$

λ_k -----最佳步长。

为了避免计算 $\nabla^2 f(x^{(k)})^{-1}$, 设法构造另一个矩阵 H_k , 用它直接逼近 $\nabla^2 f(x^{(k)})^{-1}$ 。

要求:

(1) 每一步都能以现有的信息来确定下一个搜索方向。

(2) 每一次迭代, 目标函数值均有所下降。

(3) $H_k \rightarrow \nabla^2 f(x^{(k)})^{-1}$.

设在第 k 次迭代后, 得到点 $x^{(k+1)}$.

$$f(x) \approx f(x^{(k+1)}) + \nabla f(x^{(k+1)})^T (x - x^{(k+1)}) + \frac{1}{2} (x - x^{(k+1)})^T \nabla^2 f(x^{(k+1)}) (x - x^{(k+1)})$$

\therefore 在 $x^{(k+1)}$ 附近有

$$\nabla f(x) \approx \nabla f(x^{(k+1)}) + \nabla^2 f(x^{(k+1)}) (x - x^{(k+1)})$$

令 $x = x^{(k)}$, 则有

$$\nabla f(x^{(k)}) \approx \nabla f(x^{(k+1)}) + \nabla^2 f(x^{(k+1)}) (x^{(k)} - x^{(k+1)})$$

$$\text{记 } p^{(k)} = x^{(k+1)} - x^{(k)}, q^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$$

$$\Rightarrow q^{(k)} \approx \nabla^2 f(x^{(k+1)}) p^{(k)}$$

设 $\nabla^2 f(x^{(k+1)})$ 可逆, 则

$$p^{(k)} \approx \nabla^2 f(x^{(k+1)})^{-1} q^{(k)} \quad (*)$$

\therefore 计算出 $p^{(k)}$ 和 $q^{(k)}$ 后, 可以根据(*)估计在 $x^{(k+1)}$ 处的Hesse矩阵的逆矩阵。

$$\text{令 } p^{(k)} = H_{k+1} q^{(k)}$$

拟牛顿
条件

拟牛顿法步骤

1. 给定初始点 $x^{(0)}$, 正定矩阵 H_0 , 允许误差 $\varepsilon > 0$;
计算 $g_0 = \nabla f(x^{(0)})$, 置 $k = 0$ 。

2. 若 $\|g_k\| < \varepsilon$, 则停止计算; 否则, 计算搜索方向

$$d^k = -H_k g_k。$$

3. 作一维搜索, 求 λ_k , 满足

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min f(x^{(k)} + \lambda d^{(k)}),$$

$$\text{令 } x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}。$$

4. 计算 $g_{k+1} = \nabla f(x^{(k+1)})$, 并且对矩阵 H_k 进行校正,
得到 H_{k+1} , 使之满足拟牛顿条件; 令 $k = k + 1$, 返回2。

秩1校正

一般策略： H_1 取为任一个 n 阶对称正定矩阵，
通过修正 H_k 给出 H_{k+1} 。

$$\text{令 } H_{k+1} = H_k + \Delta H_k$$

校正
矩阵

确定 ΔH_k 的方法之一： 令

$$\Delta H_k = \alpha_k Z^{(k)} Z^{(k)T}$$

其中 α_k 为常数， $Z^{(k)}$ 为 n 维列向量，

$\therefore \Delta H_k$ 是秩为1的对称矩阵， $Z^{(k)}$ 的选取应使
拟牛顿条件成立， 即 $p^{(k)} = H_{k+1} q^{(k)}$ 。

$$\Rightarrow p^{(k)} = H_k q^{(k)} + \alpha_k Z^{(k)} Z^{(k)T} q^{(k)} \quad (1)$$

$$\therefore Z^{(k)} = \frac{p^{(k)} - H_k q^{(k)}}{\alpha_k Z^{(k)T} q^{(k)}}$$

$$p^{(k)} = H_{k+1} q^{(k)}$$

另一方面, (1)式两端左乘 $q^{(k)T}$

$$q^{(k)T} (p^{(k)} - H_k q^{(k)}) = \alpha_k (Z^{(k)T} q^{(k)})^2$$

$$\begin{aligned} \therefore \Delta H_k &= \alpha_k Z^{(k)} Z^{(k)T} = \alpha_k \frac{(p^{(k)} - H_k q^{(k)}) (p^{(k)} - H_k q^{(k)})^T}{\alpha_k Z^{(k)T} q^{(k)}} \frac{(p^{(k)} - H_k q^{(k)})^T}{\alpha_k Z^{(k)T} q^{(k)}} \\ &= \frac{(p^{(k)} - H_k q^{(k)}) (p^{(k)} - H_k q^{(k)})^T}{\alpha_k (Z^{(k)T} q^{(k)})^2} = \frac{(p^{(k)} - H_k q^{(k)}) (p^{(k)} - H_k q^{(k)})^T}{q^{(k)T} (p^{(k)} - H_k q^{(k)})} \end{aligned}$$

$$\therefore H_{k+1} = H_k + \frac{(p^{(k)} - H_k q^{(k)}) (p^{(k)} - H_k q^{(k)})^T}{q^{(k)T} (p^{(k)} - H_k q^{(k)})}$$

拟牛顿法步骤

1. 给定初始点 $x^{(0)}$, 正定矩阵 H_0 , 允许误差 $\varepsilon > 0$;

计算 $g_0 = \nabla f(x^{(0)})$, 置 $k = 0$ 。

2. 若 $\|g_k\| < \varepsilon$, 则停止计算; 否则, 计算搜索方向

$$d^k = -H_k g_k。$$

3. 作一维搜索, 求 λ_k , 满足

$$f(x^{(k)} + \lambda_k d^{(k)}) = \min f(x^{(k)} + \lambda d^{(k)}),$$

令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ 。

4. 计算 $g_{k+1} = \nabla f(x^{(k+1)})$, $p^{(k)} = x^{(k+1)} - x^{(k)}$, $q^{(k)} = g_{k+1} - g_k$,

$$H_{k+1} = H_k + \frac{\left(p^{(k)} - H_k q^{(k)}\right) \left(p^{(k)} - H_k q^{(k)}\right)^T}{q^{(k)T} \left(p^{(k)} - H_k q^{(k)}\right)}$$

令 $k = k + 1$, 返回2。

DFP算法(变尺度法)

定义:

$$\Delta H_k = \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}}$$

$$H_{k+1} = H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}}$$

DFP公式

DFP法计算步骤:

1. 给定初始点 $x^{(1)}$, 计算误差 $\varepsilon > 0$ 。
2. 置 $H_1 = I$, 计算 $g_1 = \nabla f(x^{(1)})$, 置 $k = 1$ 。
3. 令 $d^{(k)} = -H_k g_k$ 。
4. 从 $x^{(k)}$ 出发, 沿 $d^{(k)}$ 做一维搜索: $f(x^{(k)} + \lambda_k d^{(k)}) = \min f(x^{(k)} + \lambda d^{(k)})$
令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ 。
5. 若 $\|\nabla f(x^{(k+1)})\| < \varepsilon$, 则停止迭代, 得点 $\bar{x} = x^{(k+1)}$; 否则转6。
6. 若 $k = n$, 则令 $x^{(1)} = x^{(k+1)}$, 返回2; 否则, 进行7。
7. 令 $g_{k+1} = \nabla f(x^{(k+1)})$, $p^k = x^{(k+1)} - x^{(k)}$, $q^k = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$,
$$H_{k+1} = H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}},$$
置 $k = k + 1$, 返回3。

例：用**DFP**方法求解下列问题：

$$\min 2x_1^2 + x_2^2 - 4x_1 + 2$$

解： $\nabla f(x) = (4x_1 - 4, 2x_2)^T$

初始点和初始矩阵分别取为：

$$x^{(1)} = (2, 1)^T, H_1 = I$$

第
一
次
迭
代

在点 $x^{(1)}$ 处的梯度为 $g_1 = (4, 2)^T$,

令搜索方向 $d^{(1)} = -H_1 g_1 = (-4, -2)^T$

从 $x^{(1)}$ 出发，沿方向 $d^{(1)}$ 做一维搜索，得 $\lambda_1 = \frac{5}{18}$,

$$\Rightarrow x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \left(\frac{8}{9}, \frac{4}{9}\right)^T, g_2 = \left(-\frac{4}{9}, \frac{8}{9}\right)^T.$$

第二次迭代

$$p^{(1)} = \lambda_1 d^{(1)} = \left(-\frac{10}{9}, -\frac{5}{9} \right)^T, q^{(1)} = g_2 - g_1 = \left(-\frac{40}{9}, -\frac{40}{9} \right)^T$$

$$H_2 = \frac{1}{360} \begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix}$$

$$\text{令 } d^{(2)} = -H_2 g_2 = \frac{12}{51} \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

$$\begin{aligned} p^k &= x^{(k+1)} - x^{(k)}, \\ q^k &= \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), \end{aligned}$$

从 $x^{(2)}$ 出发，沿方向 $d^{(2)}$ 进行一维搜索，得 $\lambda_2 = \frac{17}{36}$ ，

$$\Rightarrow x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = (1, 0)^T$$

此时， $g_3 = (0, 0)^T, \therefore x^{(3)} = (1, 0)^T$ 为最优解。

例：用共轭梯度法求解下列问题：

$$\min 2x_1^2 + x_2^2 - 4x_1 + 2$$

$$\text{解： } \nabla f(x) = (4x_1 - 4, 2x_2)^T \quad A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\text{初始点取为： } x^{(1)} = (2, 1)^T$$

第
一
次
迭
代

$$\text{在点 } x^{(1)} \text{ 处的梯度为 } g_1 = (4, 2)^T,$$

$$\text{令搜索方向 } d^{(1)} = -g_1 = (-4, -2)^T$$

$$\text{从 } x^{(1)} \text{ 出发，沿方向 } d^{(1)} \text{ 做一维搜索，得 } \lambda_1 = \frac{5}{18},$$

$$\Rightarrow x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \left(\frac{8}{9}, \frac{4}{9} \right)^T, g_2 = \left(-\frac{4}{9}, \frac{8}{9} \right)^T.$$

第二次迭代

$$g_2 = \left(-\frac{4}{9}, \frac{8}{9}\right)^T, \quad \beta_1 = \frac{\|g_2\|^2}{\|g_1\|^2} = \frac{4}{81}$$

$$\text{令 } d^{(2)} = -g_2 + \beta_1 d^{(1)} = \frac{20}{81} \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

从 $x^{(2)}$ 出发，沿方向 $d^{(2)}$ 进行一维搜索，得 $\lambda_2 = \frac{9}{20}$,

$$\Rightarrow x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = (1, 0)^T$$

此时， $g_3 = (0, 0)^T, \therefore x^{(3)} = (1, 0)^T$ 为最优解。

DFP算法(变尺度法)

$$\Delta H_k = \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}}$$

$$H_{k+1} = H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}}$$

定理：设矩阵 H_k 是正定对称矩阵，且 $p^{(k)T}q^{(k)} > 0$,

则由 DFP 公式构造的 H_{k+1} 是正定对称的。

证明：由于 $p^{(k)T}q^{(k)} > 0$ ，所以 $q^{(k)} \neq 0$ ，由 H_k 的正定性，得到

$q^{(k)T}H_kq^{(k)} > 0$ ，所以 DFP 公式有意义。对 $\forall x \neq 0$ ，考察

$$\begin{aligned}x^T H_{k+1} x &= x^T H_k x + \frac{\left(x^T p^{(k)}\right)^2}{p^{(k)T} q^{(k)}} - \frac{\left(x^T H_k q^{(k)}\right)^2}{q^{(k)T} H_k q^{(k)}} \\&= \frac{\left(x^T H_k x\right)\left(q^{(k)T} H_k q^{(k)}\right) - \left(x^T H_k q^{(k)}\right)^2}{q^{(k)T} H_k q^{(k)}} + \frac{\left(x^T p^{(k)}\right)^2}{p^{(k)T} q^{(k)}}\end{aligned}$$

由 $Cuachy - schwarz$ 不等式，知

$$\left(x^T H_k x\right)\left(q^{(k)T} H_k q^{(k)}\right) \geq \left(x^T H_k q^{(k)}\right)^2$$

且等式成立当且仅当存在 $\lambda \neq 0$ ，使得 $x = \lambda q^{(k)}$

$\Rightarrow x^T H_{k+1} x > 0 \Rightarrow H_{k+1}$ 正定。

推论：若 $d^{(k)}$ 是下降方向，且一维搜索是精确的，
并设 H_k 是正定对称矩阵，则 H_{k+1} 也是正定对称阵。

证明：因为一维搜索是精确的，所以有

$$\nabla f\left(x^{(k+1)}\right)^T d^{(k)} = 0$$

而且步长 $\lambda_k > 0$. 由于

$$p^{(k)} = x^{(k+1)} - x^{(k)} = x^{(k)} + \lambda_k d^{(k)} - x^{(k)} = \lambda_k d^{(k)}$$

$$\therefore \nabla f\left(x^{(k+1)}\right)^T p^{(k)} = 0$$

$$\because p^{(k)T} q^{(k)} = p^{(k)T} \left(\nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right) \right)$$

$$= -p^{(k)T} \nabla f\left(x^{(k)}\right) = -\lambda_k d^{(k)T} \nabla f\left(x^{(k)}\right) > 0,$$

由定理知， H_{k+1} 是正定的。

定理：设初始矩阵 H_1 是正定对称矩阵，且一维搜索是精确的，若 $g_k = \nabla f(x^{(k)}) \neq 0$ ，则产生的搜索方向 $d^{(k)}$ 是下降方向。

证明：用归纳法证明 $g_k^T d^{(k)} < 0$ 。

当 $k = 1$ 时， H_1 正定，因此

$$g_1^T d^{(1)} = -g_1^T H_1 g_1 < 0.$$

$$d^{(k)} = -H_k g_k.$$

假设当 $k = m$ 时，命题为真，

$$\text{即 } g_m^T d^{(m)} < 0,$$

由推论， H_{m+1} 正定，因此有

$$g_{m+1}^T d^{(m+1)} = -g_{m+1}^T H_{m+1} g_{m+1} < 0,$$

即当 $k = m+1$ 时，命题为真。

定理：用 DFP 算法求解正定二次函数

$$f(x) = \frac{1}{2} x^T A x + b^T x + c,$$

若一维搜索是精确的，假设已进行了 m 次迭代，则

(1) 搜索方向 $d^{(1)}, \dots, d^{(m)}$ 是 m 个非零的关于 A 共轭的方向；

(2) 对于 $j = 1, 2, \dots, m$, 有

$$H_{m+1} q^{(j)} = p^{(j)}$$

且当 $m = n$ 时，有 $H_{m+1} = A^{-1}$.

其中 $p^{(i)} = x^{(i+1)} - x^{(i)} = \lambda_i d^{(i)}$.

证明：由算法知， $d^{(1)}, \dots, d^{(m)}$ 是非零的。

用数学归纳法证明

$$\begin{cases} p^{(i)T} A p^{(j)} = 0, & 1 \leq j < i \leq m \\ H_{m+1} A p^{(j)} = p^{(j)} & 1 \leq j \leq m. \end{cases}$$

$$\begin{aligned} d^{(k)} &= -H_k g_k \\ p^{(i)} &= \lambda_i d^{(i)} \end{aligned}$$

当 $m=1$ 时，有

$$H_2 A p^{(1)} = \left(H_1 + \frac{p^{(1)} p^{(1)T}}{p^{(1)T} q^{(1)}} - \frac{H_1 q^{(1)} q^{(1)T} H_1}{q^{(1)T} H_1 q^{(1)}} \right) A p^{(1)}$$

$$\text{由于 } A p^{(i)} = A(x^{(i+1)} - x^{(i)}) = g_{i+1} - g_i = q^{(i)}$$

$$\text{所以 } H_2 A p^{(1)} = p^{(1)}.$$

当 $m=2$ 时，有

$$\begin{aligned} p^{(2)T} A p^{(1)} &= \lambda_2 d^{(2)T} A p^{(1)} = -\lambda_2 (H_2 g_2)^T A p^{(1)} \\ &= -\lambda_2 g_2^T H_2 A p^{(1)} = -\lambda_2 g_2^T p^{(1)} = -\lambda_1 \lambda_2 g_2^T d^{(1)} = 0 \end{aligned}$$

由拟牛顿条件，得 $H_3 q^{(2)} = p^{(2)}$,

而 $Ap^{(2)} = q^{(2)}$, 所以, $H_3 Ap^{(2)} = p^{(2)}$.

$$\text{又 } H_3 Ap^{(1)} = \left(H_2 + \frac{p^{(2)} p^{(2)T}}{p^{(2)T} q^{(2)}} - \frac{H_1 q^{(2)} q^{(2)T} H_2}{q^{(2)T} H_2 q^{(2)}} \right) Ap^{(1)}$$

$$= H_2 Ap^{(1)} + \frac{p^{(2)} p^{(2)T} Ap^{(1)}}{p^{(2)T} q^{(2)}} - \frac{H_1 q^{(2)} q^{(2)T} H_2 Ap^{(1)}}{q^{(2)T} H_2 q^{(2)}}$$

$$= p^{(1)} - \frac{H_1 q^{(2)} \left(Ap^{(2)} \right)^T p^{(1)}}{q^{(2)T} H_2 q^{(2)}} = p^{(1)}.$$

$$\begin{cases} p^{(i)T} Ap^{(j)} = 0, \\ H_{m+1} Ap^{(j)} = p^{(j)}. \\ Ap^{(i)} = q^{(i)} \end{cases}$$

假设 $m = k - 1$ 时, 有
$$\begin{cases} p^{(i)T} A p^{(j)} = 0, & 1 \leq j < i \leq k - 1 \\ H_k A p^{(j)} = p^{(j)} & 1 \leq j \leq k - 1. \end{cases}$$

当 $m = k$ 时, 对 $1 \leq j \leq k - 1$

$$\begin{aligned} p^{(k)T} A p^{(j)} &= \lambda_k d^{(k)T} A p^{(j)} = -\lambda_k (H_k g_k)^T A p^{(j)} \\ &= -\lambda_k g_k^T H_k A p^{(j)} = -\lambda_k g_k^T p^{(j)} = -\lambda_k \lambda_j g_k^T d^{(j)} = 0. \end{aligned}$$

对 $1 \leq j \leq k$,

$$H_{k+1} A p^{(j)} = \left(H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}} \right) A p^{(j)}$$

若 $j = k$, 将 $A p^{(k)} = q^{(k)}$ 代入得 $H_{k+1} A p^{(k)} = p^{(k)}$.

$$H_{k+1}Ap^{(j)} = \left(H_k + \frac{p^{(k)} p^{(k)T}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k}{q^{(k)T} H_k q^{(k)}} \right) Ap^{(j)}$$

若 $j < k$, 则

$$\begin{aligned} H_{k+1}Ap^{(j)} &= H_k Ap^{(j)} + \frac{p^{(k)} p^{(k)T} Ap^{(j)}}{p^{(k)T} q^{(k)}} - \frac{H_k q^{(k)} q^{(k)T} H_k Ap^{(j)}}{q^{(k)T} H_k q^{(k)}} \\ &= p^{(j)} - \frac{H_k q^{(k)} \left(Ap^{(k)} \right)^T p^{(j)}}{q^{(k)T} H_k q^{(k)}} = p^{(j)} \end{aligned}$$

\Rightarrow 命题成立

$$\begin{cases} p^{(i)T} Ap^{(j)} = 0, \\ H_{m+1}Ap^{(j)} = p^{(j)}. \\ Ap^{(i)} = q^{(i)} \end{cases}$$

令 $D = (p^{(1)}, p^{(2)}, \dots, p^{(n)})$

由于 $p^{(1)}, p^{(2)}, \dots, p^{(n)}$ 是一组关于 A 共轭的非零向量，
因此它们线性无关，故， D 是可逆矩阵，

由于 $H_{n+1}AD = D$

所以 $H_{n+1} = A^{-1}$.

定理： 若 $g_i \neq 0, i = 1, 2, \dots, n$, 则 DFP 方法构造的矩阵 $H_i (i = 1, 2, \dots, n)$ 为对称正定矩阵。

推论： 若目标函数是正定二次函数，则 DFP 方法经有限步迭代必达极小点。

定理： 如果 $f(x)$ 是 E^n 上的二次连续可微实函数，对任意的 $\hat{x} \in E^n$ ，存在常数 $m > 0$ ，使得当

$$x \in C(\hat{x}) = \{x \mid f(x) \leq f(\hat{x})\}$$

$y \in E^n$ 时，有 $m\|y\|^2 \leq y^T \nabla^2 f(x)y$ ，则 DFP 方法产生的序列 $\{x^{(k)}\}$ 或终止于或收敛于 f 在 E^n 上的唯一极小点。