梯度法和共轭梯度法

1. 无约束最优化问题

2. 梯度法

3. 共轭梯度法

一． 无约束最优化问题

无约束最优化问题

$$\min \quad f(x)$$
$$s.t. \quad x \in R^n$$

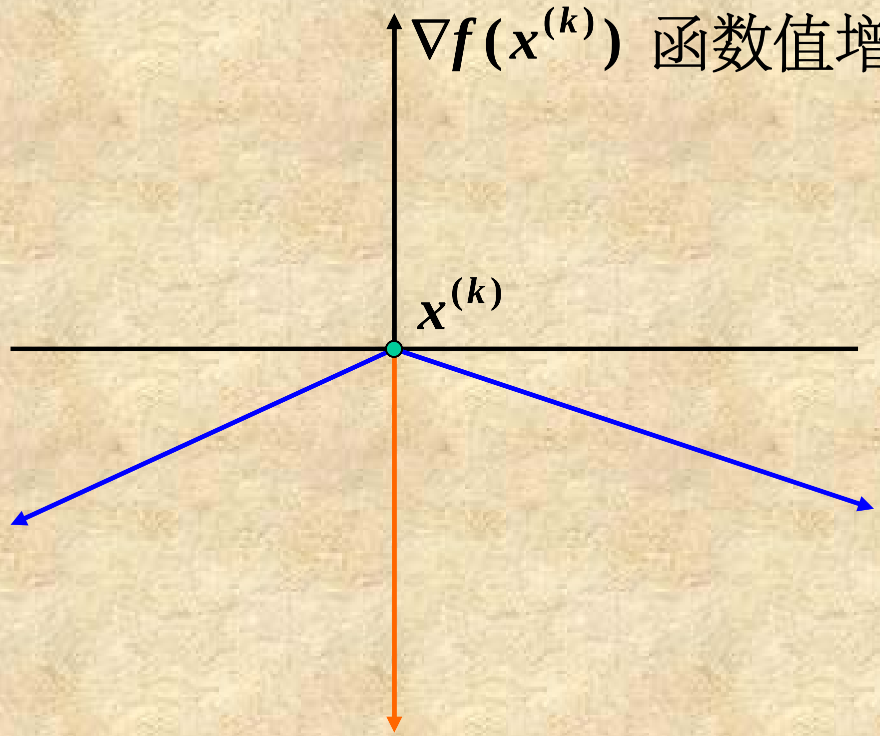
其中$f(x)$有一阶连续偏导数。

解析方法：利用函数的解析性质构造迭代公式使之收敛到最优解。

二. 梯度法（最速下降法）

迭代公式： $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$

如何选择下降最快的方向？



梯度法（最速下降法）：

1. 搜索方向：$d^{(k)} = -\nabla f(x^{(k)})$，也称为最速下降方向；

2. 搜索步长：λ_k 取最优步长，即满足

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

梯度法算法步骤：

1. 任取初始点 $x^{(1)} \in R^n$，允许误差 $\varepsilon > 0$，令 $k = 1$。

2. 计算搜索方向 $d^{(k)} = -\nabla f(x^{(k)})$；

3. 若 $\|d^{(k)}\| \le \varepsilon$，则停止计算，$x^{(k)}$ 为所求极值点；否则，求最优步长 λ_k，使得 $f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda} f(x^{(k)} + \lambda d^{(k)})$。

4. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$，令 $k := k+1$，转2。

例. 用最速下降法求解：$\min \quad f(x)=x_1^2+3x_2^2$，设初始点为 $x^{(1)}=(2,1)^T$，求迭代一次后的迭代点 $x^{(2)}$。

解： $\because \quad \nabla f(x)=(2x_1,6x_2)^T$，

$$\therefore \quad d^{(1)}=-\nabla f(x^{(1)})=(-4,-6)^T .$$

$$\therefore \quad x^{(1)}+\lambda d^{(1)}=(2-4\lambda,1-6\lambda)^T .$$

令 $\varphi(\lambda)=f(x^{(1)}+\lambda d^{(1)})=(2-4\lambda)^2+3(1-6\lambda)^2$，

求解 $\min\limits_{\lambda} \varphi(\lambda)$

令 $\varphi'(\lambda)=-8(2-4\lambda)-36(1-6\lambda)=0 \Rightarrow \lambda_1=\dfrac{13}{62}$

$$\therefore \quad x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=\left(\frac{36}{31},\frac{-8}{31}\right)^T$$

收敛性

性质. 设 $f(x)$ 有一阶连续偏导数，若 步长 λ_k 满足

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

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

证明： 令 $\varphi(\lambda)=f(x^{(k)}+\lambda d^{(k)})$, 所以

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

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

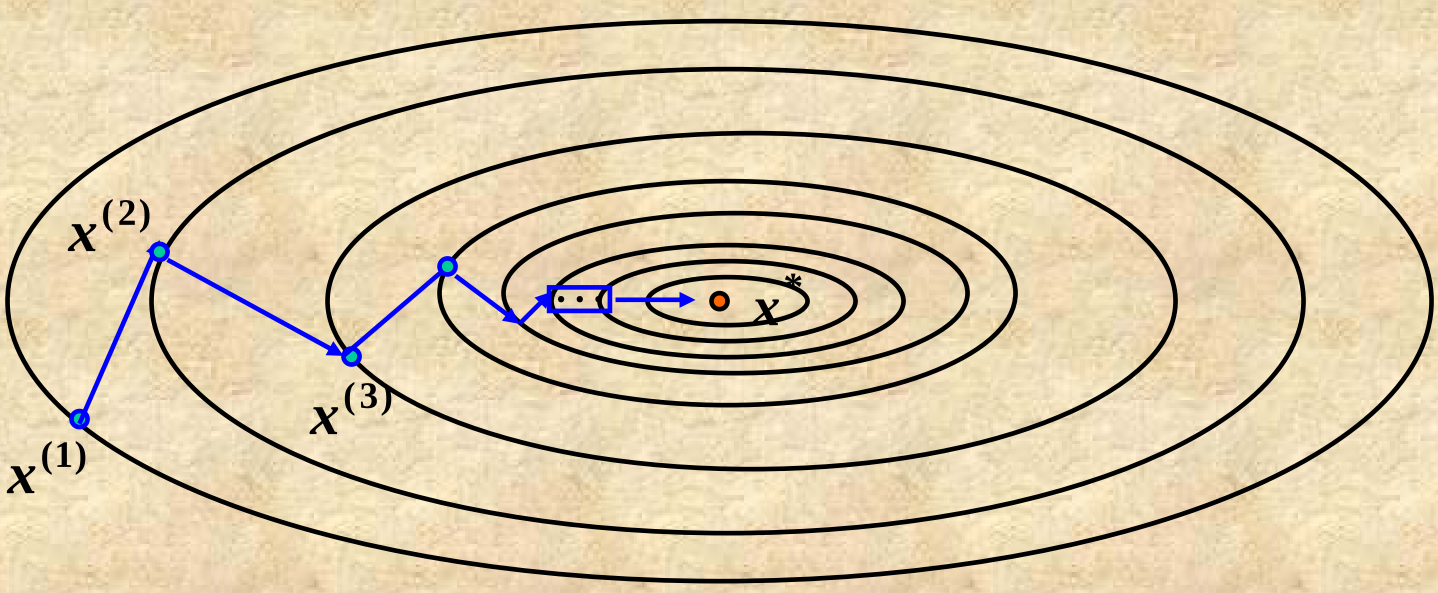
$$\therefore\ \varphi'(\lambda_k)=\nabla f(x^{(k)}+\lambda_k d^{(k)})^T d^{(k)}=0\ .$$

注： 因为梯度法的搜索方向 $d^{(k+1)}=-\nabla f(x^{(k)}+\lambda_k d^{(k)})$, 所以

$$(d^{(k+1)})^T d^{(k)}=0 \Rightarrow d^{(k+1)}\perp d^{(k)}$$ 。

锯齿现象

在极小点附近，目标函数可以用二次函数近似，其等值面近似椭球面。



注 最速下降方向反映了目标函数的一种局部性质。它只是局部目标函数值下降最快的方向。

收敛性：最速下降法是线性收敛的算法。

三. 共轭梯度法

1. 共轭方向和共轭方向法

定义 设 $\boldsymbol{A}$ 是 $\boldsymbol{n}\times\boldsymbol{n}$ 的对称正定矩阵，对于 $\boldsymbol{R}^n$ 中的两个非零向量 $\boldsymbol{d}^{(1)}$ 和 $\boldsymbol{d}^{(2)}$, 若有 $\boldsymbol{d}^{(1)^T}\boldsymbol{A}\boldsymbol{d}^{(2)}=0$，则称 $\boldsymbol{d}^{(1)}$和$\boldsymbol{d}^{(2)}$关于$\boldsymbol{A}$共轭。

设 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)},\cdots,\boldsymbol{d}^{(k)}$ 是 $\boldsymbol{R}^n$ 中一组非零向量，如果 它们两两关于 $\boldsymbol{A}$共轭，即 $\boldsymbol{d}^{(i)^T}\boldsymbol{A}\boldsymbol{d}^{(j)}=0, i\neq j, i,j=1,2,\cdots,k$。

则称这组方向是关于 $\boldsymbol{A}$共轭的，也称它们是一 组$\boldsymbol{A}$共轭方向。

注：如果 $\boldsymbol{A}$是单位矩阵，则

$$\boldsymbol{d}^{(1)^T}\cdot\boldsymbol{I}\cdot\boldsymbol{d}^{(2)}=0\Rightarrow\boldsymbol{d}^{(1)^T}\cdot\boldsymbol{d}^{(2)}=0$$

$$\Rightarrow\boldsymbol{d}^{(1)}\perp\boldsymbol{d}^{(2)}$$

共轭是正交的推广。

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

证明　设存在实数 $\alpha_1,\alpha_2,\cdots,\alpha_k$，使得

$$\sum_{i=1}^{k}\alpha_i d^{(i)}=0,$$

上式两边同时左乘 $d^{(j)^T}A$，则有

$$\sum_{i=1}^{k}\alpha_i d^{(j)^T}\cdot Ad^{(i)}=0,$$

因为 $d^{(1)},d^{(2)},\cdots,d^{(k)}$ 是 k 个 A 共轭的向量，所以上式 可化简为

$$\alpha_j d^{(j)^T}Ad^{(j)}=0.$$

因为$d^{(j)}\neq 0$,而 A是正定矩阵，所以 $d^{(j)^T}Ad^{(j)}>0$,

所以 $\alpha_j=0,\ j=1,2,\cdots,k$。

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

几何意义

设有二次函数

$$f(x)=\frac{1}{2}(x-\bar{x})^T A(x-\bar{x})$$

其中 A 是 $n\times n$ 对称正定矩阵， $\bar{x}$ 是一个定点。

问题：求函数 $f(x)$ 的极小点？

方法一：利用极值点的充要条件计算。

令 $\nabla f(x)=A(x-\bar{x})=0 \Longrightarrow x=\bar{x}$

而 $\nabla^2 f(\bar{x})=A,$

因为A 正定，所以$\nabla^2 f(\bar{x})=A>0,$

因此 $\bar{x}$ 是 $f(x)$ 的极小点。

方法二：利用搜索算法。

$$f(x)=\frac{1}{2}(x-\bar{x})^T A(x-\bar{x})$$

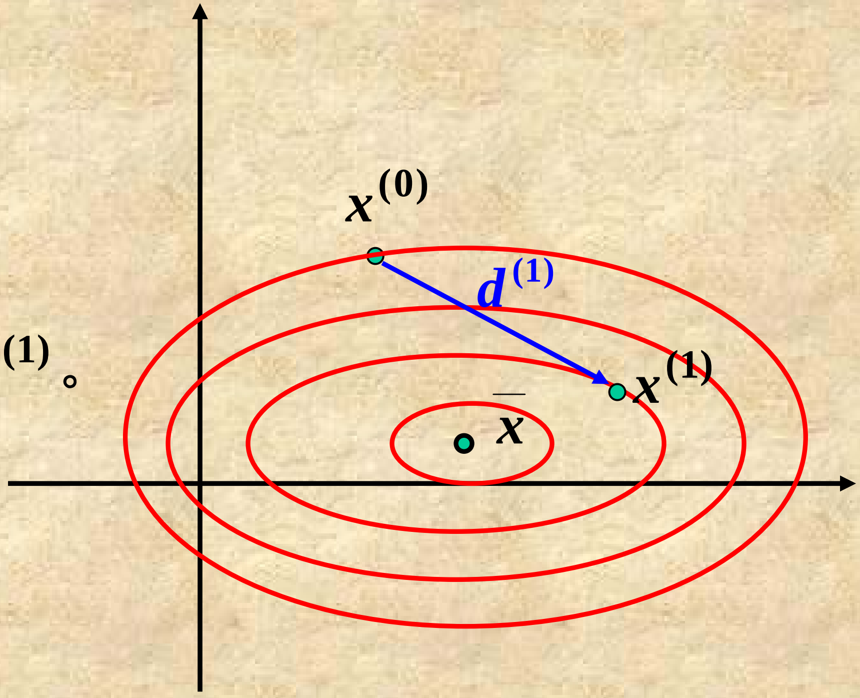
则函数 $f(x)$ 的等值面 $\frac{1}{2}(x-\bar{x})^T A(x-\bar{x})=c$

是以 $\bar{x}$ 为中心的椭球面。

设 $x^{(0)}$ 是在某个等值面上的一点，

$d^{(1)}$ 是 R^n 中的一个方向，

$x^{(0)}$ 沿着 $d^{(1)}$ 以最优步长搜索得到点 $x^{(1)}$。



则 $d^{(1)}$ 是点 $x^{(1)}$ 所在等值面的切向量。

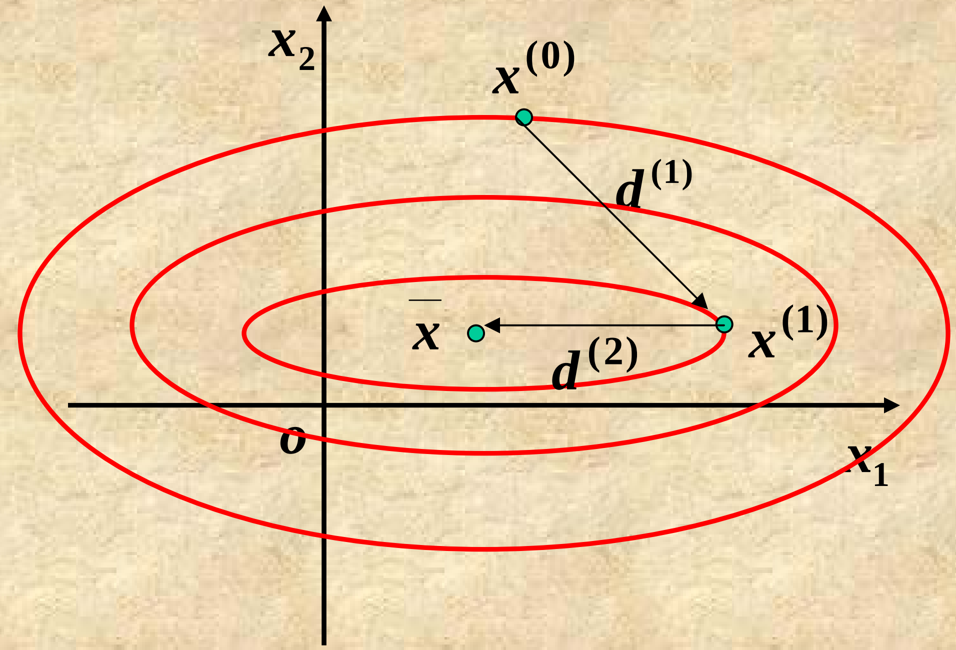
下一个搜索方向如何确定?

该等值面在点 $x^{(1)}$ 处的法向量为

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

则 $d^{(1)}$ 与 $\nabla f(x^{(1)})$ 正交,

即 $d^{(1)T}\nabla f(x^{(1)}) = 0$。



令 $d^{(2)} = \bar{x} - x^{(1)}$,

所以 $d^{(1)T}Ad^{(2)} = 0$。

即等值面上一点处的切 向量与由这一点指向极 小点的向量关于 A 共轭。

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

其中 A 是 n 阶对称正定矩阵。$d^{(1)},d^{(2)},\cdots,d^{(n)}$ 是一组 A 共轭向量。

以任意的 $x^{(1)}\in R^n$ 为初始点，依次沿 $d^{(1)},d^{(2)},\cdots,d^{(n)}$ 进行搜索，

得到点 $x^{(2)},x^{(3)},\cdots,x^{(n+1)}$，则 $x^{(n+1)}$ 是函数 $f(x)$ 在 R^n 上的

$f(x)$ 在 R^n 上的唯一极小点。

推论 在上述定理条件下，必 有

$$\nabla f(x^{(k+1)})^T d^{(i)}=0,\quad i=1,2,\cdots,k。$$

共轭方向法

对于极小化问题

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

其中 A 是正定矩阵，称下述算 法为共轭方向法：

(1) 取定一组 A 共轭方向 $d^{(1)},d^{(2)},\cdots,d^{(n)}$;

(2) 任取初始点 $x^{(1)}$，依次按照下式由 $x^{(k)}$ 确定点 $x^{(k+1)}$，

$$\begin{cases} x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)} \\ f(x^{(k)}+\lambda_k d^{(k)})=\min\limits_{\lambda} f(x^{(k)}+\lambda d^{(k)}) \end{cases}$$

直到某个 $x^{(k)}$ 满足 $\nabla f(x^{(k)})=0$。

注 由定理2可知，利用共轭方向法求解上述极小化问题，

至多经过 n 次迭代必可得到最优解。

- 如何选取一组共轭方向？

2. 共轭梯度法

$Fletcher-Reeves$ 共轭梯度法：

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

其中 $x\in R^n$, A是对称正定矩阵, $b\in R^n$, c 是常数。

基本思想：将共轭性和最速下降方向相结合，利用已知迭 代点处的梯度方向构造一组共轭方向，并沿此方向进行搜索，求出函数的极小点。

以下分析算法的具体步骤。

- 搜索方向的构造：

(1) 任取初始点 $x^{(1)}$，第一个搜索方向取为 $d^{(1)}=-\nabla f(x^{(1)})$；

(2) 设已求得点 $x^{(k+1)}$，若$\nabla f(x^{(k+1)})\neq 0$，令 $g_{k+1}=\nabla f(x^{(k+1)})$，

则下一个搜索方向 $d^{(k+1)}$按如下方式确定：

令 $d^{(k+1)}=-g_{k+1}+\beta_k d^{(k)}$ (1)

如何确定 β_k?

要求$d^{(k+1)}$和$d^{(k)}$关于A共轭。

则在（1）式两边同时左乘 $d^{(k)^T}A$，得

$$0=d^{(k)^T}Ad^{(k+1)}=-d^{(k)^T}Ag_{k+1}+\beta_k d^{(k)^T}Ad^{(k)}$$

解得 $\beta_k=\dfrac{d^{(k)^T}Ag_{k+1}}{d^{(k)^T}Ad^{(k)}}$ (2)

(3) 搜索步长的确定 ：

已知迭代点 $x^{(k)}$ 和搜索方向 $d^{(k)}$，利用一维搜索确定最优步长 λ_k，

即求解 $\min\limits_{\lambda} \quad f(x^{(k)} + \lambda d^{(k)})$。

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

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

即有 $[A(x^{(k)} + \lambda d^{(k)}) + b]^T d^{(k)} = 0$,

令 $g_k = \nabla f(x^{(k)}) = Ax^{(k)} + b$，则有

$$[g_k + \lambda A d^{(k)}]^T d^{(k)} = 0,$$

解得 $$\lambda_k = -\frac{g_k^T d^{(k)}}{d^{(k)T} A d^{(k)}} \qquad (3)$$

定理3 对于正定二次函数 $f(x)=\frac{1}{2}x^TAx+b^Tx+c$，FR算法在 $m\le n$ 次一维搜索后即终止，并 且对所有的 $i(1\le i\le m)$，下列关系成立

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

(2) $g_i{}^Tg_j=0,\ j=1,2,\cdots,i-1;$

(3) $g_i^Td^{(i)}=-g_i^Tg_i$。

注：

（1）由定理3可知搜索方向 $d^{(1)},d^{(2)},\cdots,d^{(m)}$ 是 A 共轭的。

(2) 算法中第一个搜索方向 必须取负梯度方向，否 则构造的搜索方向不能保证共轭性。

(3) 由定理3的（3）可知， $g_i^Td^{(i)}=-g_i^Tg_i=-\|g_i\|^2<0$，所以 $d^{(i)}$ 是迭代点 $x^{(i)}$ 处的下降方向。

(4) 由定理 3 , $\boldsymbol{FR}$算法中 $\boldsymbol{\beta}_i$的计算公式可以简化。

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

$$= \frac{g_{i+1}{}^T A[(x^{(i+1)} - x^{(i)})/\lambda_i]}{d^{(i)^T} A[(x^{(i+1)} - x^{(i)})/\lambda_i]}$$

$$\because \quad g_i = \nabla f(x^{(i)}) = A x^{(i)} + b.$$

$$\therefore \quad \beta_i = \frac{g_{i+1}{}^T (g_{i+1} - g_i)}{d^{(i)^T} (g_{i+1} - g_i)} = \frac{\| g_{i+1} \|^2}{-d^{(i)^T} g_i}$$

$$= \frac{\| g_{i+1} \|^2}{\| g_i \|^2} \qquad \textbf{(4)}$$

FR 算法步骤:

1. 任取初始点 $x^{(1)}$,精度要求ε，令 $k=1$。

2. 令$g_1=\nabla f(x^{(1)})$,若$\|g_1\|<\varepsilon$,停止,$x^{(1)}$为所求极小点;

否则，令$d^{(1)}=-g_1$,利用公式（3）计算λ_1,令$x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}$。

3. 令$g_{k+1}=\nabla f(x^{(k+1)})$,若$\|g_{k+1}\|<\varepsilon$,停止,$x^{(k+1)}$为所求极小点;

否则，令$d^{(k+1)}=-g_{k+1}+\beta_k d^{(k)}$,其中$\beta_k$用公式（4）计算。

令 $k:=k+1$。

4. 利用公式（3）计算λ_k，令 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$,转3。

例　用 ***FR*** 算法求解下述问题：

$$\min \quad f(x)=2x_1^2+x_2^2$$

初始点取为 $x^{(1)}=(2,2)^T$。

解：$\because \quad f(x)=\frac{1}{2}(x_1,x_2)\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}, \quad \therefore \quad A=\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}.$

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

第 1 次迭代：

令 $\quad d^{(1)}=-g_1=(-8,-4)^T,$

而 $\quad \lambda_1=-\dfrac{g_1^T d^{(1)}}{d^{(1)^T}Ad^{(1)}}$

$$=-\frac{(8,4)\begin{bmatrix}-8\\ -4\end{bmatrix}}{(-8,-4)\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}-8\\ -4\end{bmatrix}}=\frac{5}{18}$$

所以 $x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}$

$$=(2,2)^T+\frac{5}{18}(-8,-4)^T=(\frac{-2}{9},\frac{8}{9})^T$$

第 2次迭代:

$\because\quad g_2=(\frac{-8}{9},\frac{16}{9})^T.$

$$\therefore\quad \beta_1=\frac{\|g_2\|^2}{\|g_1\|^2}=\frac{(\frac{-8}{9})^2+(\frac{16}{9})^2}{8^2+4^2}=\frac{4}{81}.$$

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

$$=(\frac{8}{9},\frac{-16}{9})^T+\frac{4}{81}(-8,-4)^T$$

$$=\frac{40}{81}(1,-4)^T$$

$$\because \quad \lambda_2 = -\frac{g_2^T d^{(2)}}{d^{(2)^T} A d^{(2)}}$$

$$= -\frac{\frac{40}{81}\left(\frac{-8}{9}, \frac{16}{9}\right)\begin{bmatrix} 1 \\ -4 \end{bmatrix}}{\left(\frac{40}{81}\right)^2 (1,-4)\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -4 \end{bmatrix}} = \frac{9}{20}$$

$$\therefore \quad x^{(3)} = x^{(2)} + \lambda_2 d^{(2)}$$

$$= \left(\frac{-2}{9}, \frac{8}{9}\right)^T + \frac{9}{20} \times \frac{40}{81}(1,-4)^T = (0,0)^T$$

$$\because \quad g_3 = (0,0)^T$$

$\therefore \quad x^{(3)}$即为所求极小点。

3. 用于一般函数的共轭梯度法

$$\min \quad f(x)$$
$$s.t. \quad x \in R^n$$

对用于正定二次函数的 共轭梯度法进行修改：

(1) 第一个搜索方向仍取最速下降方向，即 $d^{(1)} = -\nabla f(x^{(1)})$。

其它搜索方向按下式计 算：

$$d^{(i+1)} = -\nabla f(x^{(i+1)}) + \beta_i d^{(i)},$$

其中 $\beta_i = \dfrac{\|\nabla f(x^{(i+1)})\|^2}{\|\nabla f(x^{(i)})\|^2}$。

(2) 搜索步长 λ_i 不能利用公式（3）计算，需由一维搜索 确定。

(3) 算法在有限步迭代后不 一定能满足停止条件， 此时可采取如下措施：

以n次迭代为一轮，每次完 成一轮搜索后，如果还 没有求得极小点，则以上一轮 的最后一个迭代点作为 新的初始点，取最速下降方向作为第 一个搜索方向，开始下 一轮搜索。

注 在共轭梯度法中，也可 采用其它形式的公式计 算β_i，如

$$\beta_i = \frac{g_{i+1}^T(g_{i+1} - g_i)}{g_i^T g_i}$$ （PRP共轭梯度法）。