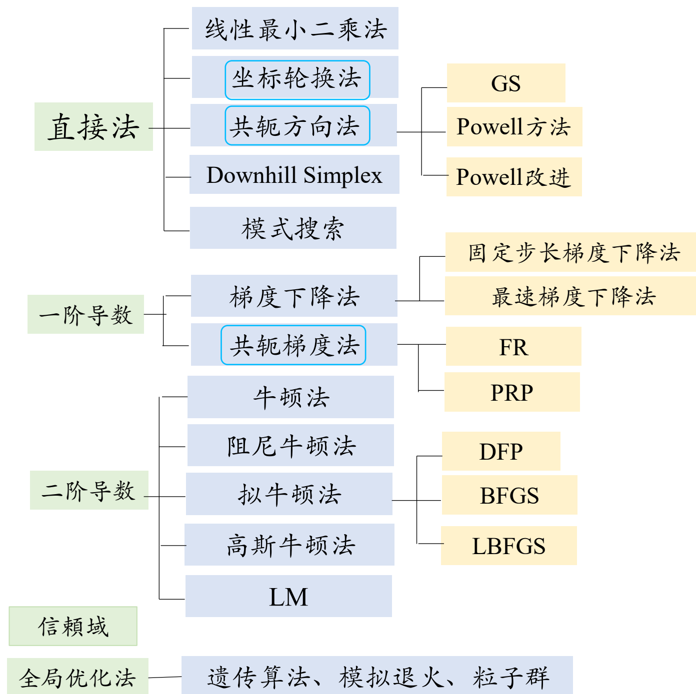
M05M11084 最优化理论、算法与应用

5 无约束优化方法 II

无约束优化方法 II

参考：

1. Numerical optimization, Chapter 5, Jorge Nocedal Stephen J. Wright
2. 最优化导论，第10章， Edwin K.P. Chong, Stanislaw H. Żak 著, 孙志强等译
3. Practical Optimization Algorithms and Engineering Applications, Chapter 6, A. Antoniou, W. LU



1.基本思想

2.共轭向量的定义与性质

3.共轭方向法

4.共轭梯度法

5.Powell方法

坐标轮换法　如果二次型的Hesse阵是对角的正定阵，那么，迭代n次收敛

$$\min f(x) \qquad f(x)=\frac{1}{2}x^T\Lambda x, \Lambda=\mathrm{diag}(\lambda_i)\succ 0,\quad x\in\mathcal{R}^n$$

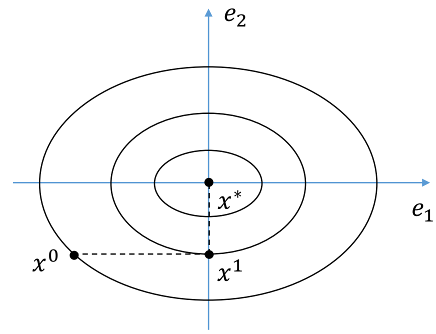
$n=2$　$f(x)=\dfrac{\lambda_1}{2}x_1^2+\dfrac{\lambda_2}{2}x_2^2$

$$x^0 \xrightarrow{e_1} x^1 \xrightarrow{e_2} x^2=x^*$$

$$x^1=x^0+\alpha_1 e_1$$

$$x^2=x^1+\alpha_2 e_2$$

$$x^2=x^*=x^0+\sum_{i=1}^{2}\alpha_i e_i$$



$\forall n>2$　$x^k=x^0+\sum_{i=1}^{k}\alpha_i e_i \quad k=1,\dots,n$　　$x^n=x^*=x^0+\sum_{i=1}^{n}\alpha_i e_i$

二次型的Hesse阵是非对角的正定阵

$f(x)=\dfrac{1}{2}x^TQx \qquad Q=\begin{bmatrix}6 & 2\\ 2 & 8\end{bmatrix}$　非对角阵 $Q\succ 0$　$e_1\to e_2\to$非极小点

正定矩阵Q的分析

$\lambda_1=4.7639$，$\lambda_2=9.2361$

$d_1=\begin{bmatrix}-0.8507\\ 0.5257\end{bmatrix} \quad d_2=\begin{bmatrix}0.5257\\ 0.8507\end{bmatrix}$　特征向量

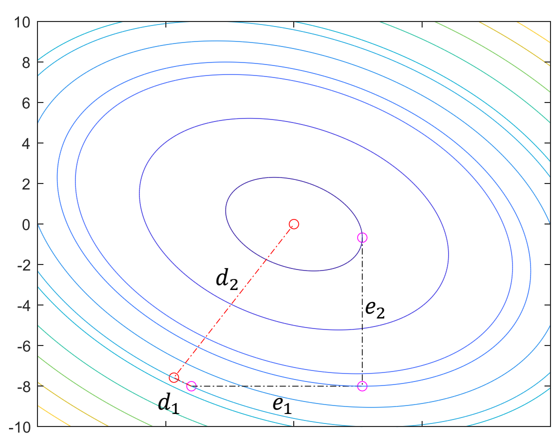
$Q=D\Lambda D^T$，$D^TD=I \qquad \Lambda=\mathrm{diag}(\lambda_i)$

$D=[d_1 \quad d_2]$

$$f(x)=\frac{1}{2}x^T(D\Lambda D^T)x$$

$$=\frac{1}{2}(D^Tx)^T\Lambda(Dx) \qquad \hat{x}=D^Tx, \qquad x=D\hat{x}$$

$$f(\hat{x})=\frac{1}{2}\hat{x}^T\Lambda\hat{x}$$



二次型的Hesse阵是非对角的正定阵

$$f(x)=\frac{1}{2}x^TQx \qquad Q=\begin{bmatrix}6 & 2\\ 2 & 8\end{bmatrix}$$

非对角阵 $Q\succ 0$　$e_1\to e_2\to$非极小点

正定矩阵Q的分析

$\lambda_1=4.7639$，$\lambda_2=9.2361$

$d_1=\begin{bmatrix}-0.8507\\ 0.5257\end{bmatrix}$　$d_2=\begin{bmatrix}0.5257\\ 0.8507\end{bmatrix}$　特征向量

$Q=D\Lambda D^T$，$D^TD=I$　$\Lambda=\mathrm{diag}(\lambda_i)$

$\hat{x}=D^Tx$，　$x=D\hat{x}$　$D=[d_1\ \ d_2]$

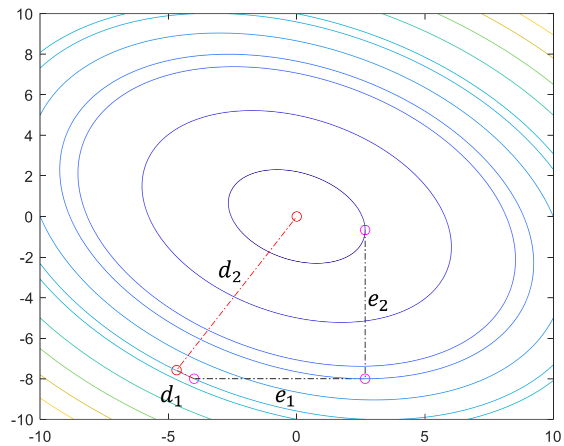
$$f(\hat{x})=\frac{1}{2}\hat{x}^T\Lambda\hat{x}$$

坐标轮换法，得 $\hat{x}^*=(0,0)$

$x^*=D\hat{x}^*$

$x^*=(8.8818\mathrm{e}{-16},2.6645\mathrm{e}{-15})$

$d_1\to d_2\to$极小点



一般地，坐标轮换法如何用？

1. 矩阵正交分解
2. 坐标变换

} 找出特征向量

3. 应用坐标轮换法n次
4. 坐标逆变换→解

} 沿着特征向量迭代n次→解

$$\min f(x) \qquad f(x)=\frac{1}{2}x^TQx, \qquad Q\succ 0, \qquad x\in\mathcal{R}^n$$ 非对角正定阵

$Q=D\Lambda D^T$，$D^TD=I$　$\Lambda=\mathrm{diag}(\lambda_i)$　$D=[d_1\ \cdots\ d_n]$

$\hat{x}=D^Tx$，　$x=D\hat{x}$

$$f(\hat{x})=\frac{1}{2}\hat{x}^T\Lambda\hat{x}$$

$$\hat{x}^k=\hat{x}^0+\sum_{i=1}^{k}\alpha_i\hat{e}_i$$

$$k=1,\dots,n$$

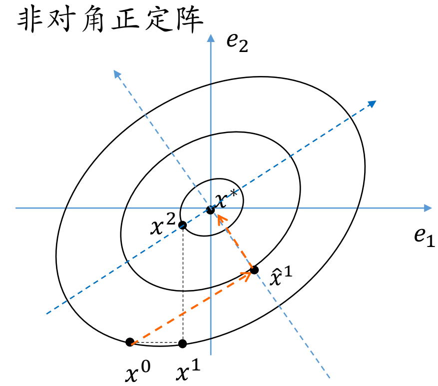
$$\hat{x}^n=\hat{x}^*$$

$$x=D\hat{x}$$

$$D\hat{x}^k=D\hat{x}^0+\sum_{i=1}^{k}\alpha_iD\hat{e}_i$$

$$x^k=x^0+\sum_{i=1}^{k}\alpha_id_i \qquad D\hat{e}_i=d_i$$

$$x^n=x^*$$



分析： $Q = D\Lambda D^T$， $D^T D = I$ $\quad \Lambda = \mathrm{diag}(\lambda_i)$

$D^T Q D = \Lambda$ $\quad$ 记$D = [d_1 \quad \cdots \quad d_n]$

$d_1, \dots, d_n$的性质

$$\begin{bmatrix} d_1^T \\ \vdots \\ d_n^T \end{bmatrix} Q [d_1 \quad \cdots \quad d_n] = \begin{bmatrix} d_1^T Q d_1 & d_1^T Q d_2 & \cdots & d_1^T Q d_n \\ d_2^T Q d_1 & d_2^T Q d_2 & \cdots & d_2^T Q d_n \\ \vdots & \vdots & \ddots & \vdots \\ d_n^T Q d_1 & d_n^T Q d_2 & \cdots & d_n^T Q d_n \end{bmatrix} = \Lambda$$

$$d_i^T Q d_j = \begin{cases} \lambda_i > 0 & i = j \\ 0 & i \neq j \end{cases}$$

称 $d_1, \dots, d_n$为关于Q的一组共轭方向

归纳：

非对角阵 $Q \succ 0$

1. 矩阵正交分解
2. 坐标变换
3. 应用坐标轮换法n次
4. 坐标逆变换→解

$\left.\begin{array}{l}1.\ 2.\end{array}\right\}$ 1.找出矩阵的一组共轭方向

$\left.\begin{array}{l}3.\ 4.\end{array}\right\}$ 2.沿共轭方向搜索迭代n次 → 最优解

1.基本思想

2.共轭向量的定义与性质

n阶正定矩阵的共轭向量组无穷多

共轭向量线性无关

n阶正定矩阵（每组）至多有n个共轭向量

3.共轭方向法

4.共轭梯度法

5.Powell方法

定义

设矩阵$Q \in \mathcal{R}^{n\times n}$为正定矩阵，若存在一组非零向量$d_0, d_1, \dots, d_{n-1} \in \mathcal{R}^n$满足

$$d_i^T Q d_j = 0, \qquad i,j = 0,1,\dots,n-1;\ \ i \neq j$$

则称向量组$d_0, d_1, \dots, d_{n-1}$关于矩阵Q共轭，或称$d_0, d_1, \dots, d_{n-1}$是矩阵Q的一组共轭方向

当$Q = I$时，$d_i^T d_j = 0,\ \ i,j = 0,1,\dots,n-1;\ \ i \neq j$，向量组$d_0, d_1, \dots, d_{n-1}$正交

① n阶正定矩阵Q的共轭向量组有无穷多个 特征向量组只是其中的一组

不妨以$n = 3$为例说明

$n = 3$，一个共轭向量组有3个共轭方向d^0, d^1, d^2 $\quad d^i = [d_1^i \quad d_2^i \quad d_3^i]^T, i = 0,1,2$

共9个未知量

$$d_i^T Q d_j = 0, \qquad i,j = 0,1,2;\ \ i \neq j \quad \Longrightarrow \begin{cases} d_0^T Q d_1 = 0 \\ d_0^T Q d_2 = 0 \\ d_1^T Q d_2 = 0 \end{cases} \quad 共3个方程$$

注意：$d_i^T Q d_j = d_j^T Q d_i$

因而，有无穷多解

一般地，

$$\begin{bmatrix} d_1^T Q d_1 & d_1^T Q d_2 & \cdots & d_1^T Q d_n \\ d_2^T Q d_1 & d_2^T Q d_2 & \cdots & d_2^T Q d_n \\ \vdots & \vdots & \ddots & \vdots \\ d_n^T Q d_1 & d_n^T Q d_2 & \cdots & d_n^T Q d_n \end{bmatrix} = \Lambda$$

定义n个方向有n^2个未知量

$(n-1) + (n-2) + \cdots + 1 = \frac{n(n-1)}{2}$个方程

有无穷多解

例 求矩阵 Q —共轭方向 $Q=\begin{bmatrix}3&0&1\\0&4&2\\1&2&3\end{bmatrix}$, $Q=Q^T$

Q 的特征值: $\lambda_1=1.0968>0, \lambda_2=3.1939>0, \lambda_3=5.7093>0$ $Q\succ 0$

$n=3$，有3个共轭方向

构建矩阵 Q-共轭方向 d^0, d^1, d^2

取 $d^0=[1,0,0]^T$ 设$d^1=(d_1^1, d_2^1, d_3^1)$， $d^2=(d_1^2, d_2^2, d_3^2)$

根据共轭方向的定义， $d^{0T}Qd^1=0$

$$d^{0T}Qd^1=[1\quad 0\quad 0]\begin{bmatrix}3&0&1\\0&4&2\\1&2&3\end{bmatrix}\begin{bmatrix}d_1^1\\d_2^1\\d_3^1\end{bmatrix}=3d_1^1+d_3^1=0$$

$d_1^1=1$， $d_2^1=0$， $d_3^1=-3$ $\Longrightarrow$ $d^1=[1,0,-3]^T$

d^2 与d^0 和 d^1 Q-共轭

$d^{0T}Qd^2=0$ and $d^{1T}Qd^2=0$

$$d^{0T}Qd^2=[1\quad 0\quad 0]\begin{bmatrix}3&0&1\\0&4&2\\1&2&3\end{bmatrix}\begin{bmatrix}d_1^2\\d_2^2\\d_3^2\end{bmatrix}=3d_1^2+d_3^2=0$$

$$d^{1T}Qd^2=[1\quad 0\quad -3]\begin{bmatrix}3&0&1\\0&4&2\\1&2&3\end{bmatrix}\begin{bmatrix}d_1^2\\d_2^2\\d_3^2\end{bmatrix}=-6d_2^2-8d_3^2=0$$

$$\Longrightarrow \begin{cases}3d_1^2+d_3^2=0\\ -6d_2^2-8d_3^2=0\end{cases}$$

令 $d_1^2=1$，解得 $d_2^2=4, d_3^2=-3$ $\Longrightarrow$ $d^2=[1,4,-3]^T$

d^0, d^1和 d^2 关于Q-共轭

$$D=[d_1\quad d_2\quad d_3]\quad D^TQD=\begin{bmatrix}3&0&0\\0&24&0\\0&0&40\end{bmatrix}$$

说明：特征向量只是共轭向量组中的一组

例 同一初始点，不同的共轭方向组

$$f(x)=\frac{1}{2}x^TQx \qquad Q=\begin{bmatrix}1&1\\1&2\end{bmatrix} \qquad x^0=\begin{bmatrix}1.4\\1.4\end{bmatrix}$$

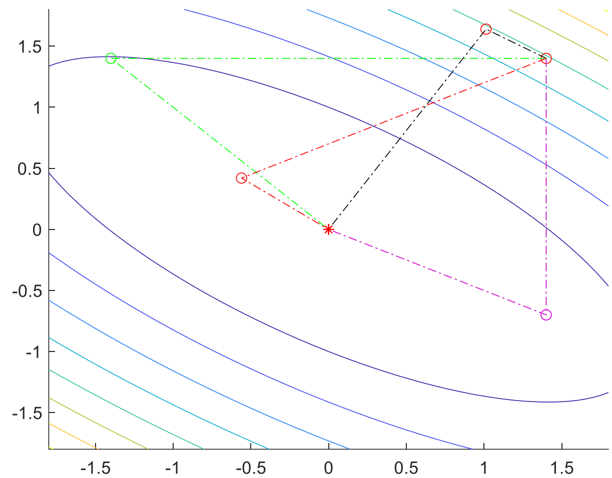
✓ 以特征向量为共轭方向

$$d_1^v=\begin{bmatrix}-0.8507\\0.5257\end{bmatrix}, d_2^v=\begin{bmatrix}0.5257\\0.8507\end{bmatrix}$$

✓ 共轭方向组A $d_1^1=\begin{bmatrix}1\\0\end{bmatrix}, d_2^1=\begin{bmatrix}1\\-1\end{bmatrix}$

✓ 共轭方向组B $d_1^2=\begin{bmatrix}0\\1\end{bmatrix}, d_2^2=\begin{bmatrix}2\\-1\end{bmatrix}$

✓ 共轭方向组C $d_1^3=\begin{bmatrix}1\\0.5\end{bmatrix}, d_2^3=\begin{bmatrix}4\\-3\end{bmatrix}$



例 更换初始点，但，共轭方向组不变

$$x^0=\begin{bmatrix}-0.5\\-1.5\end{bmatrix}$$

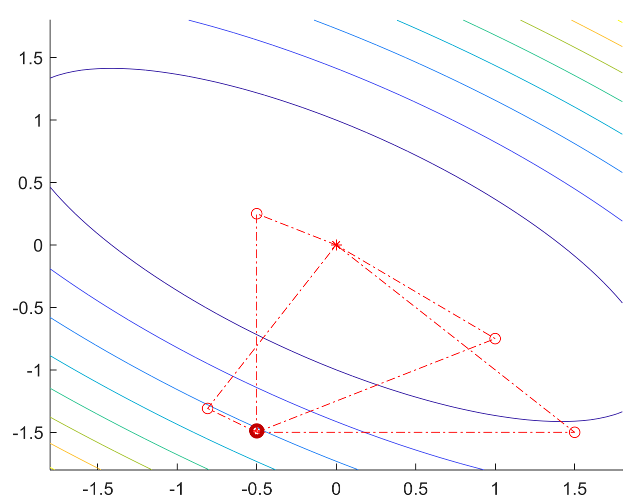
✓ 以特征向量为共轭方向

$$d_1^v=\begin{bmatrix}-0.8507\\0.5257\end{bmatrix}, d_2^v=\begin{bmatrix}0.5257\\0.8507\end{bmatrix}$$

✓ 共轭方向组A $d_1^1=\begin{bmatrix}1\\0\end{bmatrix}, d_2^1=\begin{bmatrix}1\\-1\end{bmatrix}$

✓ 共轭方向组B $d_1^2=\begin{bmatrix}0\\1\end{bmatrix}, d_2^2=\begin{bmatrix}2\\-1\end{bmatrix}$

✓ 共轭方向组C $d_1^3=\begin{bmatrix}1\\0.5\end{bmatrix}, d_2^3=\begin{bmatrix}4\\-3\end{bmatrix}$



不同的初始点，不同的共轭方向组 不影响收敛性

$x^0 = \begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}$ $x^0 = \begin{bmatrix} -0.5 \\ -1.5 \end{bmatrix}$

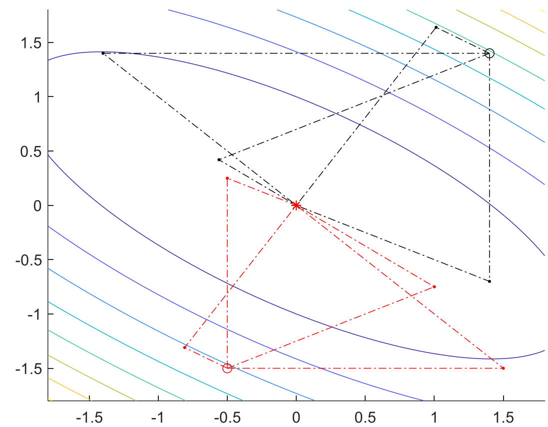
- ✓ 以特征向量为共轭方向

  $d_1^v = \begin{bmatrix} -0.8507 \\ 0.5257 \end{bmatrix}, d_2^v = \begin{bmatrix} 0.5257 \\ 0.8507 \end{bmatrix}$

- ✓ 共轭方向组A $d_1^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, d_2^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

- ✓ 共轭方向组B $d_1^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, d_2^2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

- ✓ 共轭方向组C $d_1^3 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, d_2^3 = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$



② 共轭向量线性无关定理

设$Q \in \mathcal{R}^{n\times n}$为正定矩阵，一向量组$d_0, d_1, \ldots, d_{n-1}$关于$Q$共轭，则该向量组线性无关

证：设 $\alpha_0 d_0 + \alpha_1 d_1 + \cdots + \alpha_{n-1} d_{n-1} = \sum_{i=0}^{n-1} \alpha_i d_i = 0$

两边左乘$d_k^T Q$，$k = 0,1, \ldots, n-1$ $d_k^T Q \sum_{i=0}^{n-1} \alpha_i d_i = \alpha_k d_k^T Q d_k = 0$

$\because Q \succ 0$，$d_k \neq 0 \implies d_k^T Q d_k > 0$

$\implies \alpha_k = 0$，$k = 0,1, \ldots, n-1$

$\implies$ 向量组$d_0, d_1, \ldots, d_{n-1}$线性无关

② 共轭向量线性无关定理

设$Q \in \mathcal{R}^{n\times n}$为正定矩阵，一向量组$d_0, d_1, \dots, d_{n-1}$关于$Q$共轭，则该向量组线性无关

推论 向量组$d_0, d_1, \dots, d_{n-1}$可以构成$\mathcal{R}^n$空间的一组基
$\mathcal{R}^n$空间中任意一点可以由它们的线性组合表示

沿着共轭方向搜索，相当于在以这组方向为基的“坐标系”中，进行“坐标”轮换搜索

③ n阶正定矩阵至多有n个共轭向量 定理

设$Q \in \mathcal{R}^{n\times n}$为正定矩阵，一组关于矩阵$Q$共轭的向量至多有$n$个

证：假设有 $n+1$个向量$d_0, d_1, \dots, d_{n-1}, d_n \in \mathcal{R}^n$关于矩阵$Q$共轭
则 $d_0, d_1, \dots, d_{n-1}$也关于矩阵Q共轭

根据共轭方向线性无关定理，$d_0, d_1, \dots, d_{n-1}$线性无关，构成$\mathcal{R}^n$的一组基
即，$\mathcal{R}^n$中的任一非零向量可由这组基线性表示

∃不全为0的系数α_k，$k = 0,1,\dots,n-1$
使

$$d_n = \sum_{k=0}^{n-1} \alpha_k d_k \qquad 0 \neq d_n^T Q d_n = \sum_{k=0}^{n-1} \alpha_k d_n^T Q d_k = 0 \qquad \text{矛盾}$$

所以，一组关于Q共轭的向量不会超过n个

①基本共轭方向法 采用一组共轭方向作为连续搜索方向

$$\min f(x) \qquad f(x)=\frac{1}{2}x^TQx+x^Tb+c, \qquad Q\succ 0, x\in\mathcal{R}^n$$

设$d_0, d_1, \dots, d_{n-1}$关于Q共轭

给定初始点 $x^0\in\mathcal{R}^n$

$k=0$

While $k<n$

$g^k=Qx^k+b$

$\alpha_k=-\dfrac{d_k^Tg^k}{d_k^TQd_k}$?

$x^{k+1}=x^k+\alpha_kd_k$

$k=k+1$

end

$x^*=x^n$?

由基本共轭方向法，第k次迭代可写成

$$\begin{aligned}x^k&=x^{k-1}+\alpha_{k-1}d_{k-1}\\&=x^{k-2}+\alpha_{k-2}d_{k-2}+\alpha_{k-1}d_{k-1}\\&=\cdots\\&=x^0+\sum_{i=0}^{k-1}\alpha_id_i\end{aligned}$$

$$x^n=x^0+\sum_{i=0}^{n-1}\alpha_id_i$$

二次型的精确步长

$$f(x)=\frac{1}{2}x^TQx+x^Tb+c,\qquad Q\succ 0,x\in\mathcal{R}^n$$

$$x^{k+1}=x^k+\alpha d_k \qquad\qquad g^k=\nabla f(x^k)=Qx^k+b$$

$$\phi(\alpha)=f(x^k+\alpha d_k) \qquad\qquad \alpha_k=-\frac{d_k^Tg^k}{d_k^TQd_k}$$

$$\begin{aligned}0=\phi'(\alpha)&=D_xf(x^k+\alpha d_k)D_\alpha(x^k+\alpha d_k)\\&=[\nabla f(x^k+\alpha d_k)]^Td_k\\&=[Q(x^k+\alpha d_k)+b]^Td_k\\&=(Qx^k+b)^Td_k+\alpha d_k^TQd_k\\&=d_k^T(Qx^k+b)+\alpha d_k^TQd_k\\&=d_k^Tg^k+\alpha d_k^TQd_k\end{aligned}$$

②二次收敛性

$$\min f(x)\qquad\qquad f(x)=\frac{1}{2}x^TQx+x^Tb+c,\qquad Q\succ 0,x\in\mathcal{R}^n$$

对于任意给定 $x^0\in\mathcal{R}^n$，基本共轭方向法都能在n次迭代之内收敛到唯一全局极小点x^*，即， $x^n=x^*$

分析:

对给定 $x^0\in\mathcal{R}^n$， $x^0\neq x^*\implies x^*-x^0\in\mathcal{R}^n$

设向量组$d_0,d_1,\dots,d_{n-1}$关于矩阵Q共轭，构成$\mathcal{R}^n$空间的一组基

$$\implies x^*-x^0=\sum_{i=0}^{n-1}\beta_id_i$$

$$\implies x^*=x^0+\sum_{i=0}^{n-1}\beta_id_i$$

基本算法公式 $x^n=x^0+\sum_{i=0}^{n-1}\alpha_id_i$

只要证明$\beta_i=\alpha_i$ $\quad x^*=x^n$

先证明 $\boxed{d_k^T Q(x^* - x^0) = -d_k^T g^k}$

$$Q(x^* - x^0) = Q\left(x^* - x^k\right) + Q\left(x^k - x^0\right)$$
$$= (Qx^* + b) - (Qx^k + b) + Q\left(x^k - x^0\right)$$
$$= 0 - g^k + Q\left(x^k - x^0\right)$$

$$g^k = Qx^k + b$$
$$g^* = Qx^* + b = 0$$

$$d_k^T Q(x^* - x^0) = -d_k^T g^k + d_k^T Q\left(x^k - x^0\right) = -d_k^T g^k$$

由基本共轭方向法，第k次迭代

$$x^k = x^0 + \sum_{i=0}^{k-1} \alpha_i d_i \longrightarrow x^k - x^0 = \sum_{i=0}^{k-1} \alpha_i d_i$$

左乘$d_k^T Q$ $$d_k^T Q\left(x^k - x^0\right) = \sum_{i=0}^{k-1} \alpha_i d_k^T Q d_i = 0$$

$$x^* - x^0 = \sum_{i=0}^{n-1} \beta_i d_i \qquad \boxed{d_k^T Q(x^* - x^0) = -d_k^T g^k}$$

左乘$d_k^T Q$ $$d_k^T Q(x^* - x^0) = \sum_{i=0}^{n-1} \beta_i d_k^T Q d_i = \beta_k d_k^T Q d_k$$

?

$$Q \succ 0 \Longrightarrow d_k^T Q d_k > 0 \longrightarrow \beta_k = \frac{d_k^T Q(x^* - x^0)}{d_k^T Q d_k} = -\frac{d_k^T g^k}{d_k^T Q d_k} = \alpha_k$$

与基本共轭方向法对比，第k次迭代

$$x^k = x^0 + \sum_{i=0}^{k-1} \alpha_i d_i \qquad \boxed{\alpha_k = -\frac{d_k^T g^k}{d_k^T Q d_k}}$$ 二次型精确步长公式

③搜索方向与梯度的正交关系

在二次型共轭方向法中，对于所有k， $0 \le k \le n-1$， $0 \le i \le k$ ，都有

$$d_i^T g_{k+1} = 0 \qquad g_k = \nabla f(x^k)$$

证明　归纳法

首先，证明$k=0$时，定理成立 即 $d_0^T g_1 = 0$

对于精确步长，有

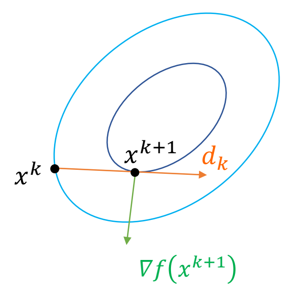
$$\Uparrow \ k = 0$$

$$\left.\begin{aligned} x^{k+1} &= x^k + \alpha_k d_k \\ \alpha_k &= \arg\min_{\alpha \ge 0} f(x^k + \alpha d_k) \end{aligned}\right. \Longrightarrow d_k^T g_{k+1} = 0$$

设 $\phi(\alpha) = f(x^k + \alpha d_k)$

$$\phi'(\alpha) = d_k^T \nabla f(x^k + \alpha d_k)$$

$$0 = \phi'(\alpha_k) = d_k^T \nabla f(x^k + \alpha_k d_k) = d_k^T g_{k+1}$$



在二次型共轭方向法中，对于所有k， $0 \le k \le n-1$， $0 \le i \le k$ ，都有

$$d_i^T g_{k+1} = 0$$

证明　设定理在$k-1$（$k < n-1$）时成立， $d_i^T g_k = 0$，$0 \le i \le k-1$

要证明，在k时也成立；　　　即， $d_i^T g_{k+1} = 0$， $0 \le i \le k$

$$x^{k+1} = x^k + \alpha_k d_k \qquad g_k = Qx^k + b$$

$$Q\alpha_k d_k = Q(x^{k+1} - x^k) = Qx^{k+1} + b - (Qx^k + b) = g_{k+1} - g_k$$

$$\Longrightarrow \ g_{k+1} = g_k + \alpha_k Q d_k$$

固定$k > 0$，考虑$0 \le i < k$　　$d_i^T g_{k+1} = \underbrace{d_i^T g_k}_{\substack{=0 \\ \because 归纳法假设}} + \alpha_k \underbrace{d_i^T Q d_k}_{\substack{=0 \\ \because 共轭向量}} = 0$

$0 \le i \le k-1$

只剩证明$d_k^T g_{k+1} = 0$

在二次型共轭方向法中，对于所有k， $0 \le k \le n-1$， $0 \le i \le k$ ，都有

$$d_i^T g_{k+1} = 0$$

证明 只剩证明$d_k^T g_{k+1} = 0$

$$\begin{aligned} d_k^T g_{k+1} &= d_k^T\left(Qx^{k+1} + b\right) \\ &= d_k^T\left(Qx^k + \alpha_k Q d_k\right) + d_k^T b \\ &= d_k^T\left(Qx^k + b\right) + \alpha_k d_k^T Q d_k \\ &= d_k^T g_k + \left(-d_k^T g_k\right) \\ &= 0 \end{aligned}$$

$$g_k = Qx^k + b$$

$$x^{k+1} = x^k + \alpha_k d_k$$

$$\alpha_k = -\frac{d_k^T g^k}{d_k^T Q d_k}$$ 二次型精确步长公式

证毕

例

$$f(x) = \frac{1}{2} x^T \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} x - x^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \qquad x \in \mathcal{R}^2$$

$x^0 = [0,0]^T$, Q-共轭方向$d_0 = [1,0]^T$ and $d_1 = \left[-\frac{3}{8}, \frac{3}{4}\right]^T$

$$Q = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}, \qquad b = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$\lambda(Q) = 0.7639, 5.2361 > 0 \longrightarrow Q = Q^T \succ 0$

$n = 2 \qquad x^* = x^2$

$k = 0$

$$g^0 = Qx^0 - b = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\alpha_0 = -\frac{d_0^T g^0}{d_0^T Q d_0} = -\frac{\begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = -\frac{1}{4}$$

$$x^1 = x^0 + \alpha_0 d_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix}$$

$k = 1$

$$g^1 = Qx^1 - b = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{3}{2} \end{bmatrix}$$

$$\alpha_1 = -\frac{d_1^T g^1}{d_1^T Q d_1} = -\frac{\begin{bmatrix} 0 & -\frac{3}{2} \end{bmatrix}\begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}}{\begin{bmatrix} -\frac{3}{8} & \frac{3}{4} \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}} = 2$$

$$x^2 = x^1 + \alpha_1 d_1 = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix} + 2\begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}$$

$$x^* = x^2$$

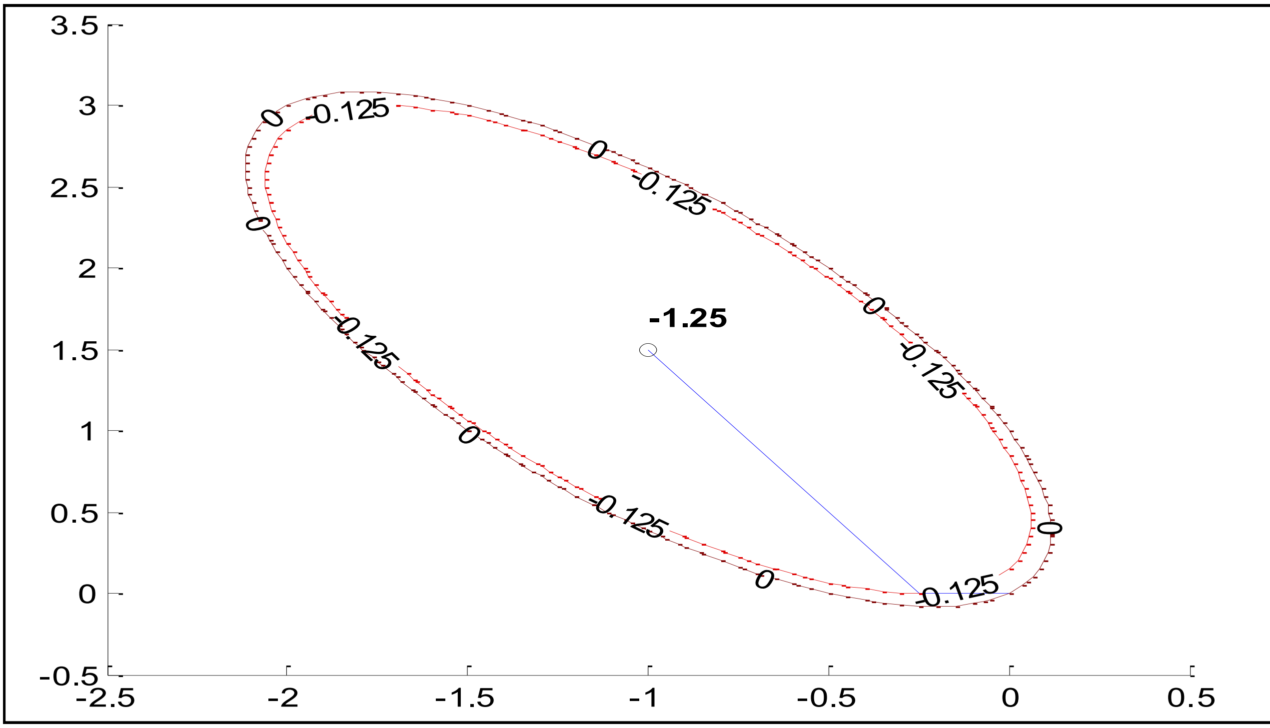
CH10-Example-10.2.xlsx

```
close all
clear
clc
Q = [4,2;2,2];
b = [-1;1];
x0 = [0;0];
d0 = [1;0]; d1 = [-0.375;0.75];
N = 2;

func = @(x)(0.5*x'*Q*x - b'*x);
getx = @(x,d,g)(x - (d'*g)/((d'*Q*d))*d);
getg = @(x)(Q*x - b);

%figure(1);


g0 = getg(x0);

x01 = x0(1);
x02 = x0(2);
v (1)= func(x0);
```

```
hold on;
for i = 1:2
    x1 = getx(x0,d0,g0);
    g1 = getg(x1);

    x01 = [x01 x1(1)];
    x02 = [x02 x1(2)];

    d0 = d1;
    x0 = x1;
    g0 = g1;
    v(i+1) = func(x0);
end
```

```
plot(x01,x02,'b');
f = func(x0);
x = -2.2:0.05:0.2; y = -0.1:0.05:3.2;
[X,Y] = meshgrid(x,y);
Z = 2*X.^2+Y.^2+2*X.*Y+X-Y;
contour (X,Y,Z,v,'ShowText','on');
%contour (X,Y,Z,'ShowText','on')

xx=x0(1)
yy=x0(2)
scatter(xx,yy,‘k’)%标注极小点"o"

text(xx,yy+0.2,{[ num2str(f) ]},‘FontSize’,10,‘FontWeight’,‘bold’);%标出极小值
```

OPT_Example_10_2.m
figure(1);
OPT_Example_10_3.m

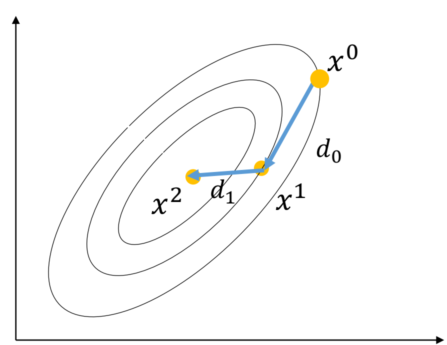


共轭方向法的难点

已知关于矩阵Q的共轭向量组$d_0, d_1, \dots, d_{n-1}$，初始点x^0

$$x^k = x^{k-1} + \alpha_{k-1} d_{k-1} \qquad k = 1,2,\dots,n$$

$$\alpha_k = -\frac{d_k^T g^k}{d_k^T Q d_k} \qquad k = 0,1,2,\dots,n-1$$



迭代前，确定n个共轭方向

但，每次迭代只用1个共轭方向

共轭方向可否随着迭代生成，
即，每次迭代时生成一个?

共轭梯度法

M05M11084 最优化理论、算法与应用

5 无约束优化方法 II

1.基本思想

2.共轭向量的定义与性质

3.共轭方向法

4.共轭梯度法

二次凸目标函数

一般目标函数

5.Powell方法

• 共轭梯度法的基本思想

✓ Hestenes和Stiefel提出解线性方程组

✓ Fletcher和Reeves引入解无约束优化

每一次迭代，产生1个共轭方向

1.第一个方向：$d_0 = -g^0$

2.其余共轭方向：

用当前点x^{k+1}的梯度g^{k+1}和前一点x^k处的搜索方向d^k， 步长α_k?

构造当前搜索方向

$$d_{k+1} = -g^{k+1} + \beta_k d^k, \qquad k = 0,1,\dots,n-2$$

并保证d_{k+1}与之前的搜索方向$d_k, d_{k-1}, \dots, d_0$共轭

3. β_k ?

• 二次凸目标函数的共轭梯度法

$$f(x) = \frac{1}{2}x^TQx + x^Tb, \qquad Q^T = Q \succ 0, \qquad x, b \in \mathcal{R}^n$$

✓ 第一次搜索：在初始点x^0处采用最速下降法

$$d_0 = -g^0, \qquad x^1 = x^0 + \alpha_0 d_0, \qquad \alpha_0 = \operatorname*{argmin}_{\alpha \ge 0} f(x^0 + \alpha d_0)$$

$$\alpha_0 = -\frac{d_0^T g^0}{d_0^T Q d_0}$$

✓ 第k次搜索：x^k出发，沿着d_k方向，确定步长

$$\phi_k(\alpha) = f\left(x^k + \alpha d_k\right)$$

$$\phi_k'(\alpha) = 0$$

$$\boxed{\alpha_k = -\frac{d_k^T g^k}{d_k^T Q d_k}}$$

$$x^{k+1} = x^k + \alpha_k d_k$$

β_k ?

二次凸目标函数的共轭梯度法

$$f(x) = \frac{1}{2}x^TQx + x^Tb, \qquad Q^T = Q \succ 0, \qquad x, b \in \mathcal{R}^n$$

✓ 第一次搜索：在初始点 x^0 处采用最速下降法

$$d_0 = -g^0, \qquad x^1 = x^0 + \alpha_0 d_0, \qquad \alpha_0 = -\frac{d_0^T g^0}{d_0^T Q d_0}$$

✓ 第二次搜索：选择d_0 和g^1的线性组合为d_1，即，$d_1 = -g^1 + \beta_0 d_0$

$$0 = d_1^T Q d_0 = (-g^1 + \beta_0 d_0)^T Q d_0 = -d_0^T Q g^1 + \beta_0 d_0^T Q d_0 \qquad \beta_0 = \frac{d_0^T Q g^1}{d_0^T Q d_0}$$

✓ 第k次搜索：取 $d_{k+1} = -g^{k+1} + \beta_k d_k$

$$0 = d_{k+1}^T Q d_k = \left(-g^{k+1} + \beta_k d_k\right)^T Q d_k = -d_k^T Q g^{k+1} + \beta_k d_k^T Q d_k \qquad \boxed{\beta_k = \frac{d_k^T Q g^{k+1}}{d_k^T Q d_k}}$$

Algorithm FR-CG-Basic

$$f(x) = \frac{1}{2}x^TQx + x^Tb, \qquad Q^T = Q \succ 0, \qquad x, b \in \mathcal{R}^n$$

Given x^0, $k = 0$

Evaluate $f(x^k)$, $g^k = \nabla f(x^k)$

Set $d_k = -g(x^k)$

While $g^k \neq 0$ & $k < n$

  Compute $\alpha_k = -\frac{d_k^T g^k}{d_k^T Q d_k}$ and set $x^{k+1} = x^k + \alpha_k d_k$

  Evaluate $g^{k+1} = \nabla f(x^{k+1})$

    $\beta_k = \frac{d_k^T Q g^{k+1}}{d_k^T Q d_k}$

    $d^{k+1} = -g^{k+1} + \beta_k d_k$

    $k \leftarrow k + 1$

end(while)

例

$$f(x)=\frac{3}{2}{x_1}^2+2{x_2}^2+\frac{3}{2}{x_3}^2+x_1x_3+2x_2x_3-3x_1-x_3$$

给定初始点$x^0=0$，用共轭梯度法求极小值点.

$n=3$ $\quad f(x)=\frac{1}{2}x^TQx-x^Tb,\quad Q=\begin{bmatrix}3&0&1\\0&4&2\\1&2&3\end{bmatrix},\quad b=\begin{bmatrix}3\\0\\1\end{bmatrix}$

$k=0$ $\quad g(x)=\nabla f(x)=Qx-b \quad \longrightarrow g^0=(-3,0,-1) \quad \longrightarrow d_0=-g^0$

$$\alpha_0=-\frac{d_0^Tg^0}{d_0^TQd_0}=\frac{10}{36} \quad \longrightarrow x^1=x^0+\alpha_0d_0=(0.8333,0,0.2778)$$

$k=1$ $\quad g^1=\nabla f(x^1)=[-0.2222 \quad 0.5556 \quad 0.6667]^T$

$$\beta_0=\frac{d_0^TQg^1}{d_0^TQd_0}=0.08025 \quad \longrightarrow d_1=-g^1+\beta_0d_0=[0.4630 \quad -0.5556 \quad -0.5864]^T$$

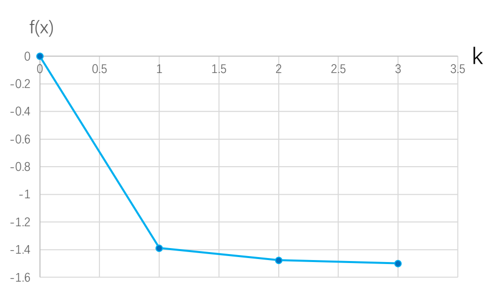
$$\alpha_1=-\frac{d_1^Tg^1}{d_1^TQd_1}=0.2187 \quad \longrightarrow x^2=x^1+\alpha_1d_1=[0.9346 \quad -0.1215 \quad 0.1495]^T$$

$k=2$ $\quad g^2=\nabla f(x^2)=[-0.04673 \quad -0.1869 \quad 0.1402]^T$

$$\beta_1=\frac{d_1^TQg^2}{d_1^TQd_1}=0.07075 \quad \longrightarrow d_2=-g^2+\beta_1d_1=[0.07948 \quad 0.1476 \quad -0.1817]^T$$

$$\alpha_2=-\frac{d_2^Tg^2}{d_2^TQd_2}=0.8231 \quad \longrightarrow x^3=x^2+\alpha_2d_2=[1.000 \quad 0.000 \quad 0.000]^T$$

$$g^3=\nabla f(x^3)=0 \quad \longrightarrow x^*=x^3=[1.000 \quad 0.000 \quad 0.000]^T$$



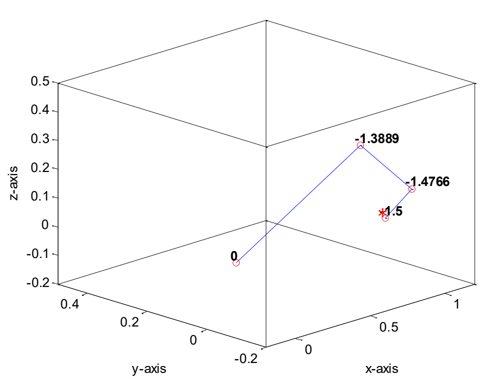
Example 10.3

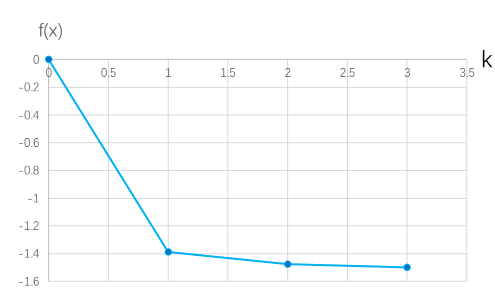
OPT_Example_10_3.m

```
close all
clear
clc
Q = [3,0,1;0,4,2;1,2,3];
b = [3;0;1];
x0 = [0;0;0];

N = 3;

func = @(x)(0.5*x'*Q*x - b'*x);
getd = @(g,d)(-g+(g'*Q*d)/(d'*Q*d)*d);
getx = @(x,d,g)(x - (d'*g)/((d'*Q*d))*d);
getg = @(x)(Q*x - b);

%figure(1);

g0 = getg(x0);
d0 = -g0;
x01 = x0(1);
x02 = x0(2);
x03 = x0(3);
v(1) = func(x0);
```

```
hold on;
for i = 1:N
    x1 = getx(x0,d0,g0);
    g1 = getg(x1);
    d1 = getd(g1,d0);

    x01 = [x01 x1(1)];
    x02 = [x02 x1(2)];
    x03 = [x03 x1(3)];

    d0 = d1;
    x0 = x1;
    g0 = g1;
    v(i+1) = func(x0);
end

plot3(x01,x02,x03,'b');
xlabel('x-axis');ylabel('y-axis');zlabel('z-axis');

axis([-0.2,1.2,-0.2,0.5,-0.2,0.5])
```

```
box on
for i=1:N+1
    scatter3(x01(i),x02(i),x03(i),‘r’)%迭代点"o"
text(x01(i),x02(i)+0.02,x03(i)+0.02,{[ num2str(v(i)) ] },‘FontSize’,10,‘FontWeight’,‘bold’);%迭代点的值
end
```





OPT_Example_10_3.m

• 一般目标函数的共轭梯度法

每一次迭代，产生1个共轭方向

1.第一个方向：$d_0 = -g^0$

2.其余共轭方向：
用当前点x^k的梯度g^k和前一点x^{k-1}处的搜索方向d_{k-1}，非精确一维搜索得步长α_k
构造当前搜索方向

$$d_k = -g^k + \beta_{k-1} d_{k-1}, \qquad k = 1,2,\dots,n-1$$

并保证d_k与之前的搜索方向$d_{k-1}, d_{k-2}, \dots, d_0$共轭

3. β_{k-1}的不同形成若干共轭梯度法

见下表

• 典型的共轭梯度法及其收敛性

名称	β_k的计算公式	提出者	收敛的步长条件
FR	$\dfrac{g^{k\,T} g^k}{g^{k-1\,T}\, g^{k-1}}$	Fletcher-Reeves	强Wolfe条件↓★
PRP	$\dfrac{g^{k\,T}(g^k - g^{k-1})}{g^{k\,T}\, g^{k-1}}$	Polak-Ribiere-Polyak	强Wolfe条件↘
PRP+	$\max\left\{\dfrac{g^{k\,T}(g^k - g^{k-1})}{g^{k\,T}\, g^{k-1}}, 0\right\}$	Powell	强Wolfe条件↓★
DY	$\dfrac{g^{k\,T} g^k}{d_k^T(g^k - g^{k-1})}$	Dai-Yuan	Wolfe条件↓★
HZ	$\left[\gamma^{k-1} - 2d_{k-1}\dfrac{\gamma^{k-1\,T}\gamma^{k-1}}{d_{k-1}^T\gamma^{k-1}}\right]^T \dfrac{g^k}{d_{k-1}^T\gamma^{k-1}}$ $\gamma^k = g^{k+1} - g^k$	Hager-Zhang	Wolfe条件↓★

↓ 共轭方向是下降方向　↘ 共轭方向不一定是下降方向　★ 全局收敛

• 共轭梯度法的特点

1.收敛性

一般形式的目标函数，采用非精确步长法，算法收敛性一般无法保证

典型算法的收敛条件如上表

2.优缺点

优点：

- ✓ 共轭梯度法的收敛速度比最速下降法快
- ✓ 不计算Hesse矩阵及其逆矩阵

缺点：

- ✓ 迭代次数增加
- ✓ 新构造的共轭方向逐渐不精确甚至于不下降，导致收敛速度变慢

改进方法：每迭代n次，令$\beta_k = 0$

Algorithm FR-CG

Given x^0, $k = 0$
Evaluate $f(x^k)$, $g^k = \nabla f(x^k)$;
Set $d_k = -g^k$
While $g^k \neq 0$
  Compute α_k and set $x^{k+1} = x^k + \alpha_k d_k$ 非精线搜，Wolfe准则
  Evaluate $g^{k+1} = \nabla f(x^{k+1})$
    $\beta_{k+1} \leftarrow DY$
      $d_{k+1} = -g^{k+1} + \beta_{k+1} d_k$
    $k \leftarrow k + 1$
    if $k =$ Integer $* n$ $d_{k+1} = -g^{k+1}$
end(while)

$\|x^{k+1} - x^k\| < tol$ 或 $\|g^{k+1}\| < tol$

Conjugate_gradient_DY.m
Wolfe__Search.m

例

例6.1 用Dai-Yuan共轭梯度法求解多维无约束最优化问题

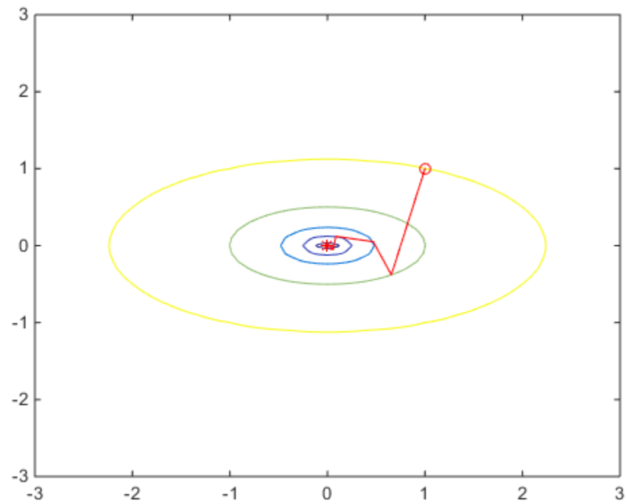
（取初始点$\boldsymbol{x}^0 = (1,1), tol = 1 \times 10^{-6}$）

$$\min f(\boldsymbol{x}) = -0.8e^{-x_1^2-4x_2^2}$$

example_6_1_CH06.m
Conjugate_gradient_DY.m
Wolfe__Search.m

testdata.txt

x_optimal = 1.0e-11 *[-0.7015   0.1358]
f_optimal =-0.8000
k = 10



例6.2 用Dai-Yuan共轭梯度法求解多维无约束最优化问题

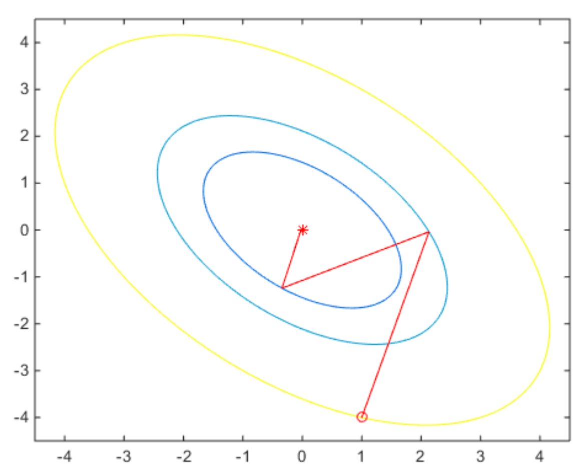
（取初始点$\boldsymbol{x}^0 = (1,-4), tol = 1 \times 10^{-6}$）

$$\min f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_1x_2 + 2$$

example_6_2_CH06.m
Conjugate_gradient_DY.m
Wolfe__Search.m

testdata.txt

x_optimal = 1.0e-17 *[0.2168   -0.520
f_optimal = 2
k = 6



例6.3 用Dai-Yuan共轭梯度法求解多维无约束最优化问题

（取初始点$\boldsymbol{x}^0 = (-4,0,-4,-1,1,1), tol = 1 \times 10^{-6}$）

$$\min f(\boldsymbol{x}) \qquad f(\boldsymbol{x}) = 1 + x_1 + x_2 + x_3 + x_4$$
$$+x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4$$
$$+x_1^2 + x_2^2 + +x_3^2 + x_4^2 - 0.4e^{-x_5^2 - 6x_6^2}$$

```
example_6_3_CH06.m
Conjugate_gradient_DY.m
Wolfe__Search.m
```

x_optimal =[-0.2000 -0.2000
-0.2000 -0.2000
0.0000 0.0000]

f_optimal = 0.2000

k =40

1.基本思想

2.共轭向量的定义与性质

3.共轭方向法

4.共轭梯度法

5.Powell方法

Practical Optimization Algorithms and Engineering Applications,
Chapter 6, A. Antoniou, W. LU

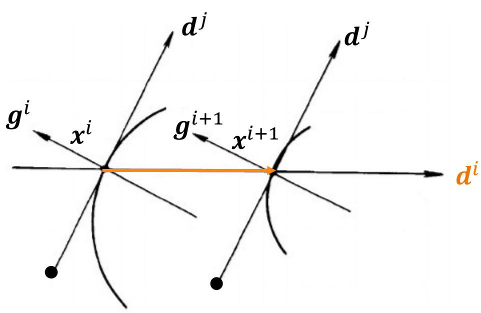
采用一维搜索产生共轭方向

共轭梯度每次迭代计算 $\nabla f(\boldsymbol{x})$

计算量： $n+1$ 函数值 $\Leftrightarrow$ $\nabla f(\boldsymbol{x})$ (n 个分量) + $f(\boldsymbol{x})$ (1个)

Powell每次迭代做一维搜索来产生共轭方向

- 对二次函数，从两个不同的点出发，沿着同一个方向$\boldsymbol{d}^j$，得到两点$\boldsymbol{x}^i$和$\boldsymbol{x}^{i+1}$
- 向量$\boldsymbol{x}^{i+1}-\boldsymbol{x}^i$ 与方向$\boldsymbol{d}^j$是共轭的



无需计算一阶导 $\nabla f(\boldsymbol{x})$

共轭方向生成原理

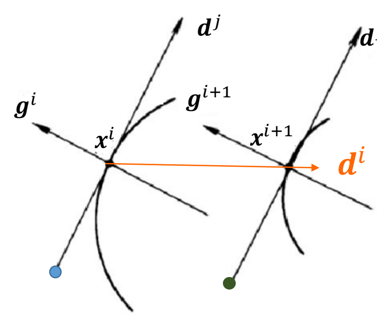
对二次函数，从两个不同的点出发，沿着同一个方向$\boldsymbol{d}^j$，得到两点$\boldsymbol{x}^i$和$\boldsymbol{x}^{i+1}$

$\boldsymbol{d}^i=\boldsymbol{x}^{i+1}-\boldsymbol{x}^i$与$\boldsymbol{d}^j$共轭

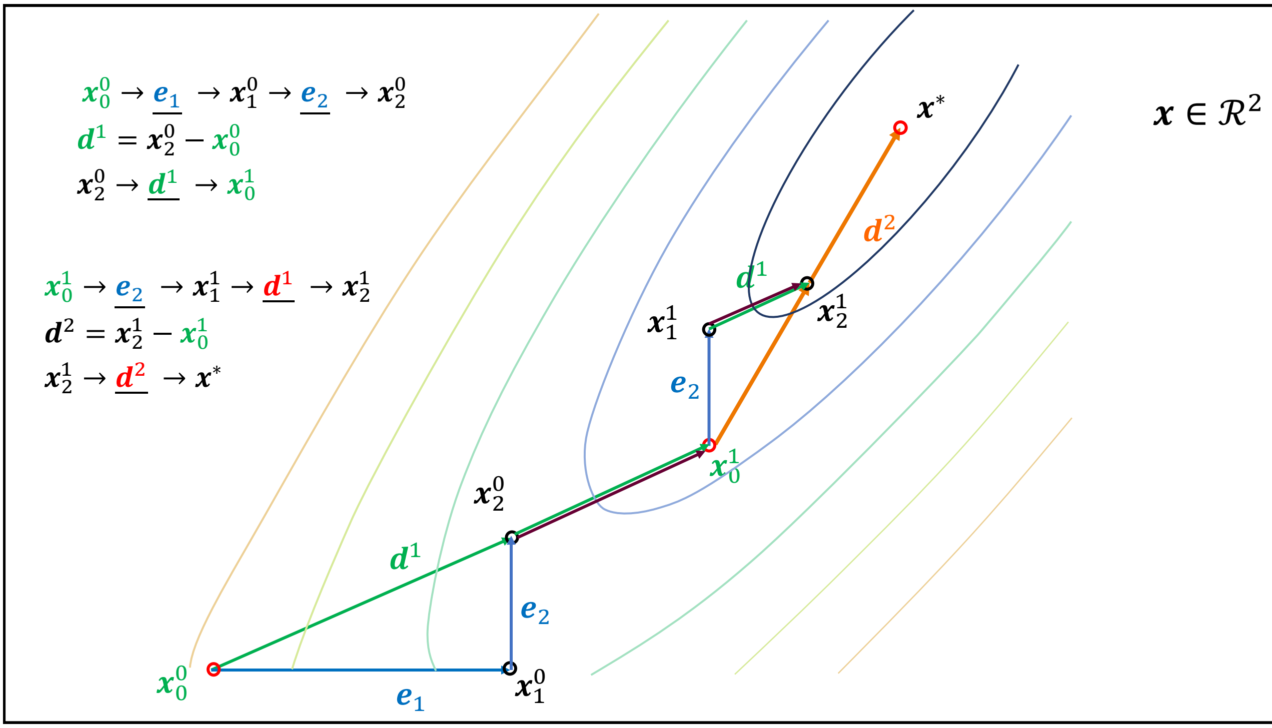
$$\boldsymbol{d}^{j\,T}\boldsymbol{g}^i=0$$
$$\boldsymbol{d}^{j\,T}\boldsymbol{g}^{i+1}=0$$

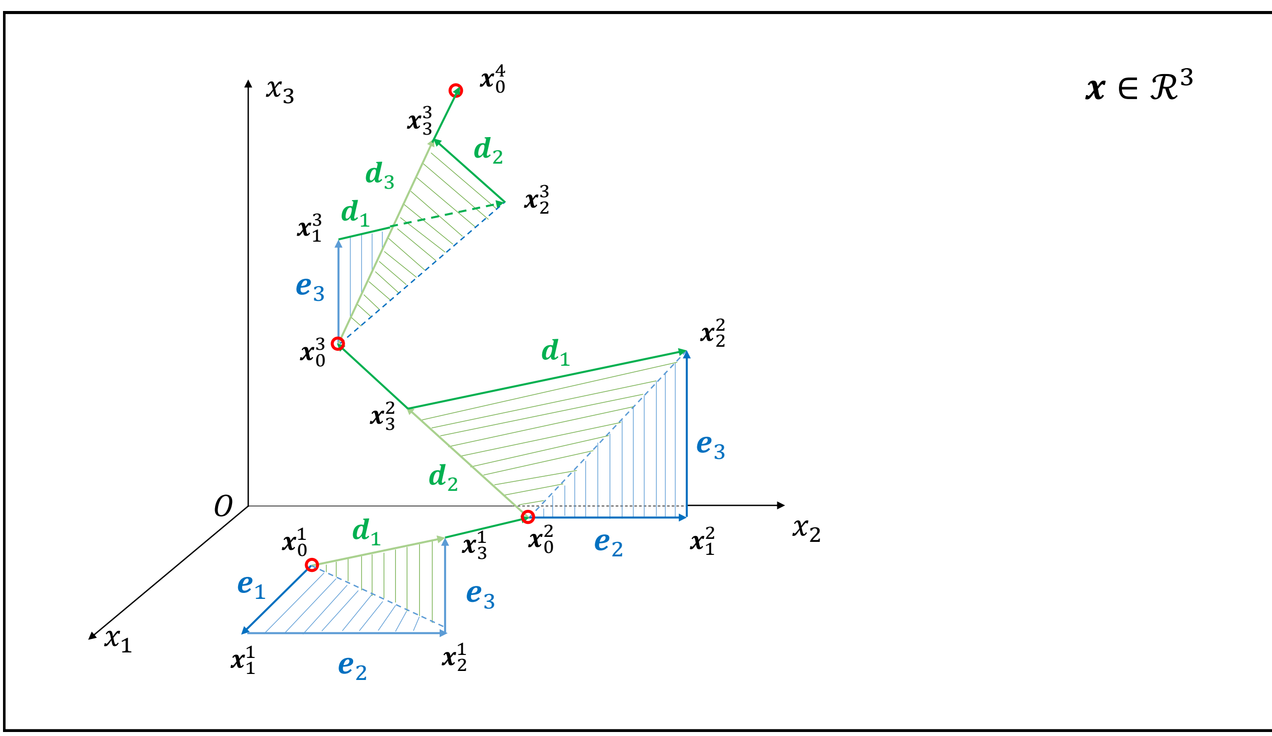
$$\boldsymbol{g}^i=Q\boldsymbol{x}^i-\boldsymbol{b}$$
$$\boldsymbol{g}^{i+1}=Q\boldsymbol{x}^{i+1}-\boldsymbol{b}$$

$$\boldsymbol{g}^{i+1}-\boldsymbol{g}^i=Q\left(\boldsymbol{x}^{i+1}-\boldsymbol{x}^i\right)$$
$$0=\boldsymbol{d}^{j\,T}\left(\boldsymbol{g}^{i+1}-\boldsymbol{g}^i\right)=\boldsymbol{d}^{j\,T}Q\left(\boldsymbol{x}^{i+1}-\boldsymbol{x}^i\right)$$



$$\boldsymbol{d}^i=\boldsymbol{x}^{i+1}-\boldsymbol{x}^i \quad\Rightarrow\quad \boldsymbol{d}^{j\,T}Q\boldsymbol{d}^i=0$$


$x_0^0 \to e_1 \to x_1^0 \to e_2 \to x_2^0$
$d^1 = x_2^0 - x_0^0$
$x_2^0 \to d^1 \to x_0^1$
$x_0^1 \to e_2 \to x_1^1 \to d^1 \to x_2^1$
$d^2 = x_2^1 - x_0^1$
$x_2^1 \to d^2 \to x^*$
$x \in \mathcal{R}^2$
x^*
d^2
d^1
x_1^1
x_2^1
e_2
x_0^1
x_2^0
d^1
e_2
x_0^0
e_1
x_1^0


$x \in \mathcal{R}^3$
x_3
x_0^4
x_3^3
d_2
d_3
x_2^3
x_1^3
d_1
e_3
x_0^3
d_1
x_2^2
x_3^2
e_3
d_2
O
x_2
x_0^1
d_1
x_3^1
x_0^2
e_2
x_1^2
e_1
e_3
x_1
x_1^1
e_2
x_2^1

- Powell算法 $\min_{x} f(\boldsymbol{x}), \boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$

$\{\boldsymbol{d}^i, i = 1, \cdots, n\}$ a given set of linearly independent vectors, $\boldsymbol{x}^0$ an initial approximation to $\boldsymbol{x}^*$

$\boldsymbol{d}^i = \boldsymbol{e}_i$

Step 1
1st Search Basic Search

$$\alpha_i = \arg\min_{\alpha} f(\boldsymbol{x}^{i-1} + \alpha \boldsymbol{d}^i)$$
$$\boldsymbol{x}^i = \boldsymbol{x}^{i-1} + \alpha_i \boldsymbol{d}^i$$

$i = 1, \cdots, n$ $\quad \boldsymbol{d}^1, \boldsymbol{d}^2, \cdots, \boldsymbol{d}^n$

Step 2
Change Direction

$$\boldsymbol{d}^i := \boldsymbol{d}^{i+1}, i = 1, \cdots, n-1$$
$$\boldsymbol{d}^n := \boldsymbol{x}^n - \boldsymbol{x}^0$$

$\boldsymbol{d}^2, \boldsymbol{d}^3, \cdots, \boldsymbol{d}^n, \boldsymbol{x}^n - \boldsymbol{x}^0$

Step 3
2nd Search Accelerated Search

$$\alpha_n = \arg\min_{\alpha} f(\boldsymbol{x}^n + \alpha \boldsymbol{d}^n)$$
$$\boldsymbol{x}^{0\text{ new}} = \boldsymbol{x}^n + \alpha_n \boldsymbol{d}^n = \boldsymbol{x}^n + \alpha_n(\boldsymbol{x}^n - \boldsymbol{x}^0)$$

Go to Step 1 unless Stopping Rules are satisfied

Stopping Rules
$\|\nabla f(\boldsymbol{x}^i)\| < \varepsilon$, $\|\boldsymbol{x}^n - \boldsymbol{x}^0\| < \varepsilon$

Example5.1

c5-Example-Powell-5.1.xlsx

$f(\boldsymbol{x}) = x_1^2 + 2x_2^2 \qquad \varepsilon = 0.03$

$\boldsymbol{x}^0 = [20,20]^T, \boldsymbol{d}^1 = [1,-1]^T, \boldsymbol{d}^2 = [1,1]^T$

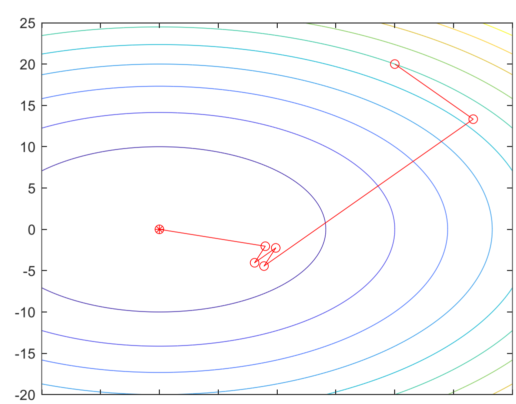
$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$$

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \boldsymbol{b} = \boldsymbol{0}$$

$$\boldsymbol{g}^i = \nabla f(\boldsymbol{x}^i) = Q\boldsymbol{x}^i - \boldsymbol{b}$$

$$\alpha_i = -\frac{\boldsymbol{g}^{i-1\,T} \boldsymbol{d}^i}{\boldsymbol{d}^{i\,T} Q \boldsymbol{d}^i}$$

$$\boldsymbol{x}^i = \boldsymbol{x}^{i-1} + \alpha_i \boldsymbol{d}^i$$



1st Iteration Step 1

i=1 $\boldsymbol{g}^0 = \nabla f(\boldsymbol{x}^0) = Q\boldsymbol{x}^0 - \boldsymbol{b} = \begin{bmatrix} 40 \\ 80 \end{bmatrix}$

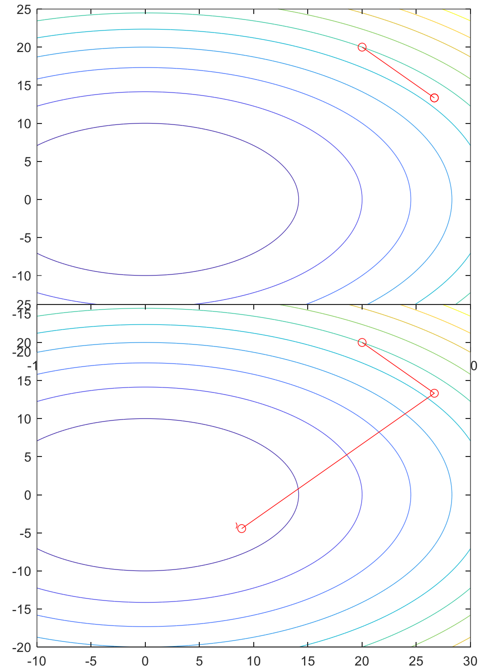
$$\alpha_1 = -\frac{\boldsymbol{g}^{0\,T}\boldsymbol{d}^1}{\boldsymbol{d}^{1\,T}Q\boldsymbol{d}^1} = -\frac{-40}{6} \approx 6.67$$

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 + \alpha_1\boldsymbol{d}^1 = \begin{bmatrix} 26.67 \\ 13.33 \end{bmatrix}$$

i=2 $\boldsymbol{g}^1 = \nabla f(\boldsymbol{x}^1) = Q\boldsymbol{x}^1 - \boldsymbol{b} \approx \begin{bmatrix} 53.34 \\ 53.34 \end{bmatrix}$

$$\alpha_2 = -\frac{\boldsymbol{g}^{1\,T}\boldsymbol{d}^2}{\boldsymbol{d}^{2\,T}Q\boldsymbol{d}^2} \approx -17.78$$

$$\boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha_2\boldsymbol{d}^2 = \begin{bmatrix} 8.89 \\ -4.44 \end{bmatrix}$$

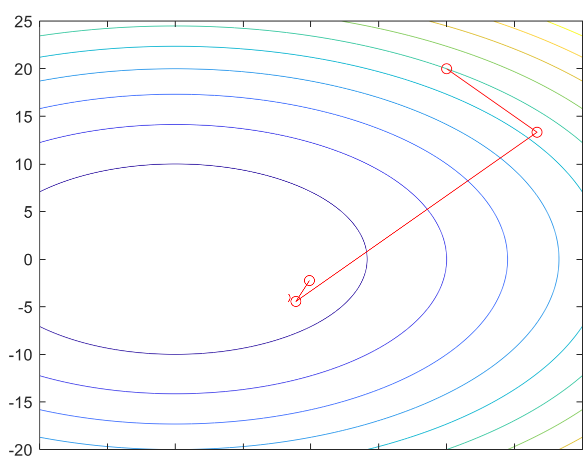


c5-Example-Powell-5.1.xlsx

Step 2

$$\boldsymbol{d}^1 = \boldsymbol{d}^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\boldsymbol{d}^2 = \boldsymbol{x}^2 - \boldsymbol{x}^0 = \begin{bmatrix} -11.11 \\ -24.44 \end{bmatrix}$$



Step 3 $\boldsymbol{g}^2 = \nabla f(\boldsymbol{x}^2) = Q\boldsymbol{x}^2 - \boldsymbol{b} \approx \begin{bmatrix} 17.78 \\ -17.76 \end{bmatrix}$

$$\alpha_{0new} = -\frac{\boldsymbol{g}^{2\,T}\boldsymbol{d}^2}{\boldsymbol{d}^{2\,T}Q\boldsymbol{d}^2} \approx -0.09$$

$$\boldsymbol{x}^{0\,new} = \boldsymbol{x}^2 + \alpha_{0new}\boldsymbol{d}^2 \approx \begin{bmatrix} 9.89 \\ -2.25 \end{bmatrix}$$

$\|\boldsymbol{x}^{0\,new} - \boldsymbol{x}^0\| \approx 2.41 > \varepsilon$

$|f(\boldsymbol{x}^{0\,new}) - f(\boldsymbol{x}^0)| \approx 5.3 > \varepsilon$

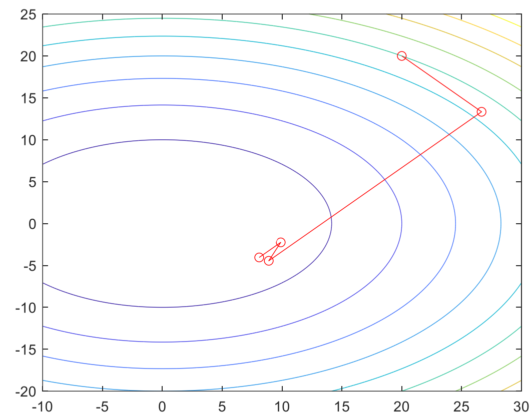
$\|\nabla f(\boldsymbol{x}^{0\,new})\| \approx 22$

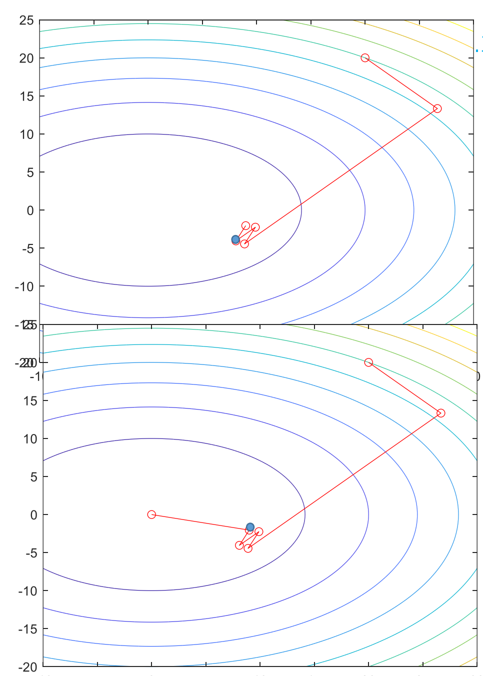
$\boldsymbol{x}^0 = \boldsymbol{x}^{0\,new}$

GoTo Step 1

$\varepsilon = 0.03$

Example_5_1_Powell_Algorithm.m





.1.xlsx

2nd Iteration $\boldsymbol{x}^0 = \begin{bmatrix} 9.89 \\ -2.25 \end{bmatrix}, \boldsymbol{d}^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{d}^2 = \begin{bmatrix} -11.11 \\ -24.44 \end{bmatrix}$

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \quad \boldsymbol{b} = \boldsymbol{0}$$

Step 1

i=1 $\boldsymbol{g}^0 = \nabla f(\boldsymbol{x}^0) = Q\boldsymbol{x}^0 - \boldsymbol{b} = \begin{bmatrix} 19.78 \\ -9 \end{bmatrix}$

$$\alpha_1 = -\frac{\boldsymbol{g}^{0\,T}\boldsymbol{d}^1}{\boldsymbol{d}^{1\,T}Q\boldsymbol{d}^1} \approx -1.80$$

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 + \alpha_1\boldsymbol{d}^1 = \begin{bmatrix} 8.09 \\ -4.04 \end{bmatrix}$$

i=2 $\boldsymbol{g}^1 = \nabla f(\boldsymbol{x}^1) = Q\boldsymbol{x}^1 - \boldsymbol{b} \approx \begin{bmatrix} 16.18 \\ -16.16 \end{bmatrix}$

$$\alpha_2 = -\frac{\boldsymbol{g}^{1\,T}\boldsymbol{d}^2}{\boldsymbol{d}^{2\,T}Q\boldsymbol{d}^2} \approx -0.08$$

$$\boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha_2\boldsymbol{d}^2 = \begin{bmatrix} 9 \\ -2.05 \end{bmatrix}$$

c5-Example-Powell-5.1.xlsx

Step 2

$$\boldsymbol{d}^1 = \boldsymbol{d}^2 = \begin{bmatrix} -11.11 \\ -24.44 \end{bmatrix}$$

$$\boldsymbol{d^2} = \boldsymbol{x}^2 - \boldsymbol{x}^0 = \begin{bmatrix} -0.89 \\ 0.20 \end{bmatrix}$$

Step 3

$$\boldsymbol{g}^2 = \nabla f(\boldsymbol{x}^2) = Q\boldsymbol{x}^2 - \boldsymbol{b} \approx \begin{bmatrix} 18 \\ -8.2 \end{bmatrix}$$

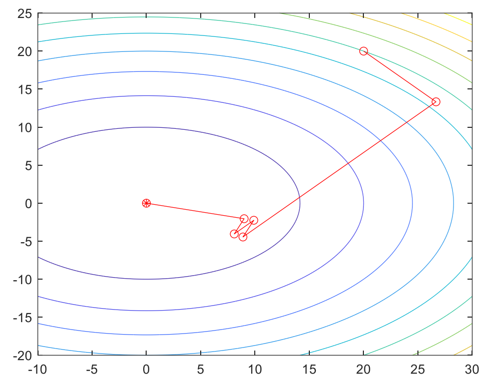
$$\alpha_{0new} = -\frac{\boldsymbol{g}^{2\,T}\boldsymbol{d}^2}{\boldsymbol{d}^{2\,T}Q\boldsymbol{d}^2} \approx 10.1$$

$$\boldsymbol{x}^{0\,new} = \boldsymbol{x}^2 + \alpha_{0new}\boldsymbol{d}^2 \approx \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\|\boldsymbol{x}^{0\,new} - \boldsymbol{x}^0\| \approx 9.22 > \varepsilon$

$|f(\boldsymbol{x}^{0\,new}) - f(\boldsymbol{x}^0)| \approx 44.67 > \varepsilon$

$\varepsilon = 0.03$

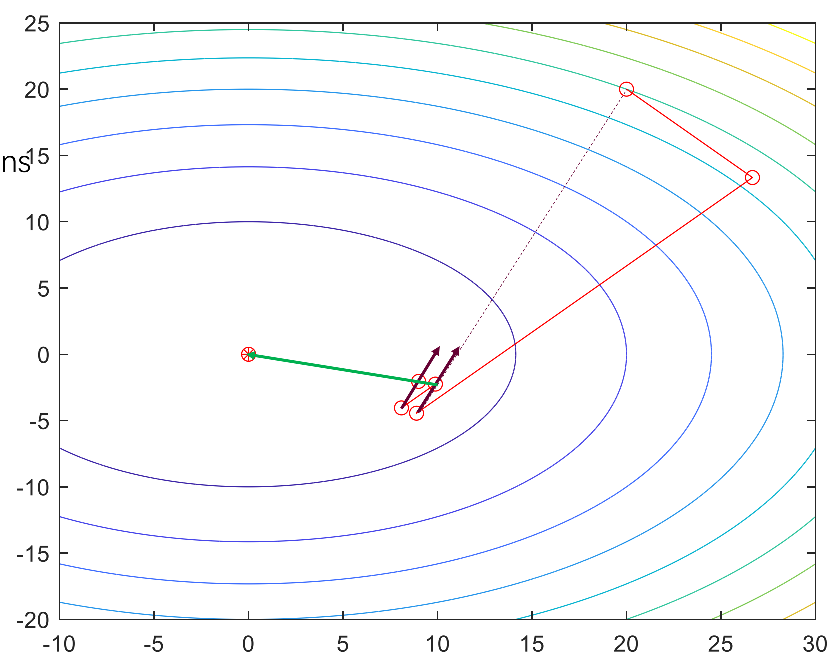


$\|\nabla f(\boldsymbol{x}^{0\,new})\| \approx 0$

Stop

$\boldsymbol{x}^* = \boldsymbol{x}^{0\,new}$

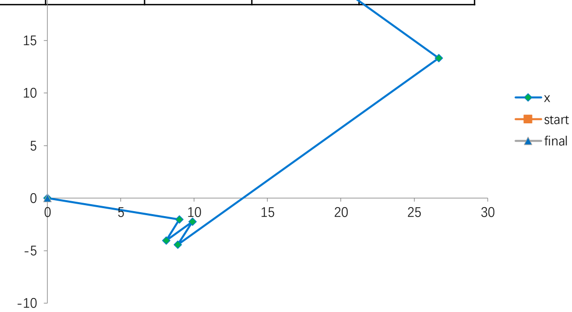
Q-Conjugate Directions



Powell's Algorithm - Example5.1

n	$\boldsymbol{x}$		Q				$\boldsymbol{g}$		$\boldsymbol{d}$		α	$\lVert\boldsymbol{g}\rVert$	$f(\boldsymbol{x})$
0	20	20	2	0	0	4	40	80	1	-1	6.6667	89.4427	1200
1	26.667	13.3333	2	0	0	4	53.3333	53.3333	1	1	-17.7778	75.4247	1066.667
2	8.8889	-4.4444	2	0	0	4	17.7778	-17.7778	-11.1111	-24.4444	-0.0899	25.1416	118.5185
0	9.8876	-2.2472	2	0	0	4	19.7753	-8.9888	1	1	-1.7978	21.7223	107.8652
1	8.0899	-4.0449	2	0	0	4	16.1798	-16.1798	-11.1111	-24.4444	-0.0818	22.8817	98.1694
2	8.9989	-2.0452	2	0	0	4	17.9977	-8.1808	-0.8888	0.2020	10.125	19.7698	89.3452
0	0	4.88E-15	2	0	0	4	0	1.95E-14				1.95E-14	4.77E-29

c5-Example-Powell-5.1.xlsx



Powell's Algorithm - Disadvantage

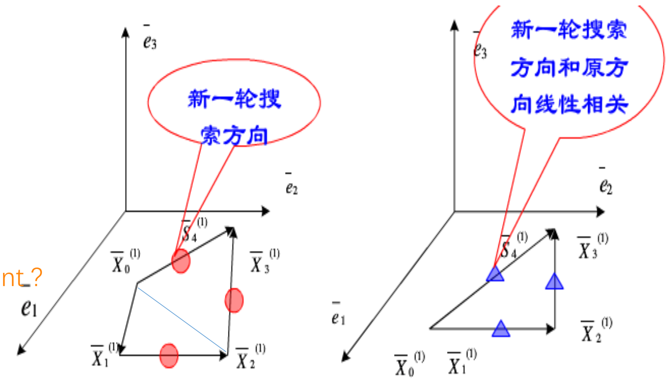
After the new displacement vector becomes a new search vector and is added to the end of the search vector list, the vectors may become linearly dependent and not produce *n*-dimensional space.

$$\boldsymbol{x}^i = \boldsymbol{x}^{i-1} + \alpha_i \boldsymbol{d}^i, i = 1, \cdots, n$$

$$\boldsymbol{x}^n = \boldsymbol{x}^{n-1} + \alpha_n \boldsymbol{d}^n$$

$$= \boldsymbol{x}^{n-2} + \alpha_{n-1} \boldsymbol{d}^{n-1} + \alpha_n \boldsymbol{d}^n$$

$$= \boldsymbol{x}^0 + \underbrace{\alpha_1 \boldsymbol{d}^1 + \cdots + \alpha_n \boldsymbol{d}^n}_{\text{linearly independent or dependent?}}$$



Introduce an **Improved Powell's algorithm**

• Improved Powell's algorithm

1. Determine whether the original search direction set is available for next iteration ?
   If available, just use it; else go to 2.
2. Determine in which direction the function value decrease the largest, and then use the new search vector to replace it.

For kth iteration, denote

$$f_1 = f(\boldsymbol{x}_0^k), f_2 = f(\boldsymbol{x}_n^k), f_3 = f(\boldsymbol{x}_{n+1}^k)$$
$$\Delta_m^k = \max_i\{\Delta_i^k, i = 1,2,\dots,n\}$$
$$\Delta_i^k = f(\boldsymbol{x}_{i-1}^k) - f(\boldsymbol{x}_i^k), \qquad i = 1,2,\dots,n$$
$$\boldsymbol{s}_m^k = \boldsymbol{x}_m^k - \boldsymbol{x}_{m-1}^k$$
$$\boldsymbol{s}^k = \boldsymbol{x}_n^k - \boldsymbol{x}_0^k$$

the reflection point

$$\boldsymbol{x}_{n+1}^k = \boldsymbol{x}^k = 2\boldsymbol{x}_n^k - \boldsymbol{x}_0^k$$

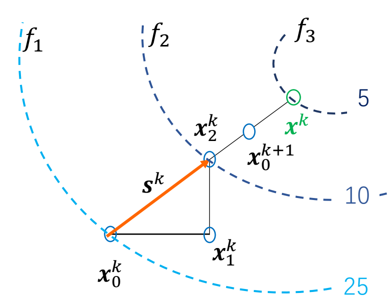
The search vector $\boldsymbol{s}_m^k$ corresponds to Δ_m^k for the function value $f(\boldsymbol{x}_m^k)$ decreases largest of all.

Reflection Point $\quad n = 2, \quad \boldsymbol{x} \in \mathcal{R}^2$

$$\boldsymbol{x}^k = 2\boldsymbol{x}_2^k - \boldsymbol{x}_0^k$$

$$\boldsymbol{x}^k - \boldsymbol{x}_2^k = \boldsymbol{x}_2^k - \boldsymbol{x}_0^k$$

$$f_1 = f(\boldsymbol{x}_0^k), f_2 = f(\boldsymbol{x}_2^k), f_3 = f(\boldsymbol{x}^k)$$
$$\Delta_m^k = \max_{i=1,2}\{\Delta_i^k\}$$
$$\Delta_i^k = f(\boldsymbol{x}_{i-1}^k) - f(\boldsymbol{x}_i^k), \qquad i = 1,2$$
$$\boldsymbol{s}_m^k = \boldsymbol{x}_m^k - \boldsymbol{x}_{m-1}^k$$
$$\boldsymbol{s}^k = \boldsymbol{x}_2^k - \boldsymbol{x}_0^k$$



$$f_1 = 25,\ f_2 = 10,\ f_3 = 5$$
$$\Delta_1^k = 25 - 10 = 15 > \Delta_2^k = 5$$

$$\Delta_1^k > \Delta_2^k, \quad m = 1$$
$$\boldsymbol{s}_1^k = \boldsymbol{x}_2^k - \boldsymbol{x}_0^k \xleftarrow{\text{Replace by}} \boldsymbol{s}^k$$

Powell's condition

$$\text{If} \begin{cases} f_3 < f_1 & \text{Cond}_{\text{PWL 1}} \\ (f_1 - 2f_2 + f_3)\left(f_1 - f_2 - \Delta_m^k\right)^2 < \frac{1}{2}\Delta_m^k (f_1 - f_3)^2 & \text{Cond}_{\text{PWL 2}} \end{cases}$$

then $\boldsymbol{s}_m^k = \boldsymbol{s}^k$ and others $\boldsymbol{s}_i^k, i = 1,2,\dots,m-1,m+1,\dots n$

else $\boldsymbol{s}_i^k, i = 1,2,\dots,m-1,m,m+1,\dots n$ (NO Change)

Improved Powell's algorithm

1. Let, the number of iteration $k = 1$,
   initial point $\boldsymbol{x}_0^1 = \boldsymbol{x}_0$, accuracy $\varepsilon_1, \varepsilon_2$
   initial direction set $\boldsymbol{s}_i^1 = \boldsymbol{e}_i,\ i = 1,2,\dots,n$

$$\lambda^* = -\frac{(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}} \boldsymbol{s}}{\boldsymbol{s}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{s}}$$

2. For i=1,2,…,n, from the point $\boldsymbol{x}_{i-1}^k$ along the search vector $\boldsymbol{s}_i^k$

$$\begin{cases} \lambda_i^* = \underset{\lambda_i}{\operatorname{argmin}} f\left(\boldsymbol{x}_{i-1}^k + \lambda_i \boldsymbol{s}_i^k\right) \\ \boldsymbol{x}_i^k = \boldsymbol{x}_{i-1}^k + \lambda_i^* \boldsymbol{s}_i^k \end{cases}$$

   and let $\boldsymbol{s}^k = \boldsymbol{x}_n^k - \boldsymbol{x}_0^k$

   1st search

   If $\|\boldsymbol{s}^k\| = \|\boldsymbol{x}_n^k - \boldsymbol{x}_0^k\| > \varepsilon$

$$\text{then} \begin{cases} \lambda^* = \underset{\lambda}{\operatorname{argmin}} f\left(\boldsymbol{x}_n^k + \lambda \boldsymbol{s}^k\right) \\ \boldsymbol{x}_{n+1}^k = \boldsymbol{x}_n^k + \lambda^* \boldsymbol{s}^k \end{cases}$$

   else $\boldsymbol{x}^* = \boldsymbol{x}_n^k,\ \ f^* = f(\boldsymbol{x}^*)$, STOP

   from the point $\boldsymbol{x}_n^k$ along the search vector $\boldsymbol{s}^k$

   2nd search

3. Stopping rule $\|\boldsymbol{x}_{n+1}^k - \boldsymbol{x}_0^k\| \le \varepsilon_1$ or $|f(\boldsymbol{x}_{n+1}^k) - f(\boldsymbol{x}_0^k)| \le \varepsilon_2 |f(\boldsymbol{x}_{n+1}^k)|$
   If Stopping rule is satisfied, we get $\boldsymbol{x}^* = \boldsymbol{x}_{n+1}^k$, $f^* = f(\boldsymbol{x}^*)$
   Else go to 4.

4. Calculate $f(\boldsymbol{x}_i^k)$, $i = 1,2,\dots,n$
   $\Delta_m^k = \max_{i=1,2,\dots,n} \{\Delta_i^k = f(\boldsymbol{x}_{i-1}^k) - f(\boldsymbol{x}_i^k)\}$
   Decrease largest along $\boldsymbol{s}_m^k = \boldsymbol{x}_m^k - \boldsymbol{x}_{m-1}^k$

5. Let $\boldsymbol{x}^k = 2\boldsymbol{x}_n^k - \boldsymbol{x}_0^k$, $f_1 = f(\boldsymbol{x}_0^k), f_2 = f(\boldsymbol{x}_n^k), f_3 = f(\boldsymbol{x}^k)$

$$\text{If}\begin{cases} f_3 < f_1 & \text{Cond}_{\text{PWL 1}} \\ (f_1 - 2f_2 + f_3)(f_1 - f_2 - \Delta_m^k)^2 < \frac{1}{2}\Delta_m^k (f_1 - f_3)^2 & \text{Cond}_{\text{PWL 2}} \end{cases}$$

   then GO TO 6
   else GO TO 7

6. $\boldsymbol{x}_0^{k+1} = \boldsymbol{x}_{n+1}^k$
   $s_i^{k+1} = s_i^k, i = 1,\dots,m-1; s_i^{k+1} = s_{i+1}^k, i = m,\dots,n-1; s_n^{k+1} = s^k$
   $k = k+1$
   GO TO 2

$$[s_1^{k+1}, s_2^{k+1}, \dots, s_n^{k+1}] \Leftarrow [s_1^k, \dots, s_{m-1}^k, s_{m+1}^k, s_n^k, \dots, s^k]$$

7. If $f_2 < f_3$, then $x_0^{k+1} = x_n^k$ else $x_0^{k+1} = x^k$
   $s_i^{k+1} = s_i^k, i = 1,\dots,n$
   $k = k+1$
   GO TO 2

OR

6. $\boldsymbol{x}_0^{k+1} = \boldsymbol{x}^k$
   $s_m^{k+1} = s^k$ and others $s_i^{k+1} = s_i^k, i = 1, \ldots, n$ and $i \neq m$
   $k = k + 1$
   GO TO 2

$$\left[s_1^{k+1}, s_2^{k+1}, \ldots, s_n^{k+1}\right] \Longleftarrow \left[s_1^k, \ldots, s_{m-1}^k, s^k, s_{m+1}^k, \ldots, s_n^k\right]$$

7. If $f_2 < f_3$, then $x_0^{k+1} = x_n^k$ else $x_0^{k+1} = x^k$
   $s_i^{k+1} = s_i^k, i = 1, \ldots, n$
   $k = k + 1$
   GO TO 2