无约束优化问题2

翟卫欣 副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院

目录

目录

无约束最优化问题

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

无约束最优化问题

- $f(\boldsymbol{x}) = f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) + o(\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|)$
  记$\boldsymbol{x} - \boldsymbol{x}^{(k)} = \alpha \boldsymbol{d}^{(k)}$，其中$\alpha > 0$，$\boldsymbol{d}^{(k)}$是一个确定的方向向量，上式可写为：
  $f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) = f(\boldsymbol{x}^{(k)}) + \alpha \nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k)} + o(\|\alpha \boldsymbol{d}^{(k)}\|)$
- 如果向量$\boldsymbol{d}^{(k)}$满足$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k)} < 0$，则$\boldsymbol{d}^{(k)}$是函数$f(\boldsymbol{x})$在$\boldsymbol{x}^{(k)}$处的下降方向，并且在所有满足$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k)} < 0$的方向中，$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k)}$越小，则$f(\boldsymbol{x})$的下降幅度越大。

无约束最优化问题

- $-\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(k)} = \| -\nabla f(\boldsymbol{x}^{(k)})\| \|\boldsymbol{d}^{(k)}\| cos\theta_k$
- 其中θ_k是向量$-\nabla f(\boldsymbol{x}^{(k)})$与$\boldsymbol{d}^{(k)}$之间的夹角
- 当α固定时，取$\theta_k = 0$，此时$-\nabla f(\boldsymbol{x}^k)^T\boldsymbol{d}^{(k)}$达到最大值。
- 如果选取搜索方向$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$，相应方法称为最速下降法，迭代形式为$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}$，$\alpha_k$由线性搜索确定。

目录

最速下降法步骤

1. 选定初始点$\boldsymbol{x}^{(1)}$和给定精度要求$\varepsilon > 0$，令$k = 1$；
2. 若$\|\nabla f(\boldsymbol{x}^{(k)})\| < \varepsilon$则停，$\boldsymbol{x}^* = \boldsymbol{x}^{(k)}$；否则，令$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$；
3. 在$\boldsymbol{x}^{(k)}$处沿方向$\boldsymbol{d}^{(k)}$作线性搜索，得到$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$，$k = k + 1$，转到步骤(2)。

- 如果在第(3)步中，采用精确线性搜索，即 $\alpha_k = argminf(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$，就有
  $$\frac{df(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})}{d\alpha}\Big|_{\alpha=\alpha_k} = (\boldsymbol{d}^{(k)})^T \nabla f(\boldsymbol{x}^{(k+1)}) = 0$$
- 此式表明$\boldsymbol{d}^{(k)}$和$\boldsymbol{d}^{(k+1)}$是正交的。

最速下降法步骤

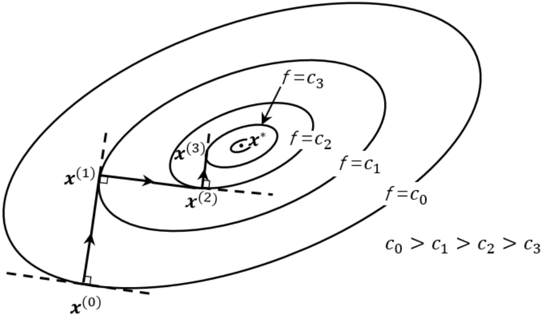

图 1: 最速下降法步骤

最速下降法举例

用最速下降法求解无约束问题

- $min\{f(\boldsymbol{x})=\frac{x_1^2}{2}+x_2^2\}$ 取初始点$\boldsymbol{x}^{(1)}=(2,1)^T$
- 由于$\nabla f(\boldsymbol{x})=(x_1,2x_2)^T$，因此$\nabla f(\boldsymbol{x}^{(1)})=(2,2)^T$，取$\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{x}^{(1)})=(-2,-2)^T$作一维搜索：构造一元函数
- $\phi(\alpha)=f(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{d}^{(1)})=2(1-\alpha)^2+(1-2\alpha)^2$
  求导得，$\phi^{'}(\alpha)=-4(1-\alpha)-4(1-2\alpha)$
- 令$\phi^{'}(\alpha)=0$，得最优步长$\alpha_1=\frac{2}{3}$，
  从而$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha\boldsymbol{d}^{(1)}=(\frac{2}{3},-\frac{1}{3})^T$
  迭代得到$\boldsymbol{x}^{(3)}=(\frac{2}{3^2},\frac{(-1)^2}{3^2})^T,\cdots,\boldsymbol{x}^{(k+1)}=(\frac{2}{3^k},\frac{(-1)^k}{3^k})^T$
- $k\to\infty$时，$\boldsymbol{x}^{(k)}\to(0,0)^T$，为最优解。

目录

Newton法

- 设$\boldsymbol{x}^*$是$\min\limits_{x\in\boldsymbol{R}^n} f(\boldsymbol{x})$的局部解，则$\boldsymbol{x}^*$满足$\nabla f(\boldsymbol{x}^*)=\boldsymbol{0}$。
- 选取初始点$\boldsymbol{x}^{(0)}$，在该处进行Taylor级数展开，取二次近似多项式
  $f(\boldsymbol{x})\approx f(\boldsymbol{x}^{(0)})+\nabla f(\boldsymbol{x}^{(0)})^T(\boldsymbol{x}-\boldsymbol{x}^{(0)})+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{(0)})^T\nabla^2 f(\boldsymbol{x}^{(0)})(\boldsymbol{x}-\boldsymbol{x}^{(0)})$
- 令近似函数的梯度为$\boldsymbol{0}$，得到
  $\nabla f(\boldsymbol{x}^{(0)})+\nabla^2 f(\boldsymbol{x}^{(0)})(\boldsymbol{x}-\boldsymbol{x}^{(0)})=\boldsymbol{0}$
- 当$\nabla^2 f(\boldsymbol{x}^{(0)})$是非奇异矩阵时，求解线性方程组得到
  $\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}-[\nabla^2 f(\boldsymbol{x}^{(0)})]^{-1}\nabla f(\boldsymbol{x}^{(0)})$

Newton法

- 当$f(\boldsymbol{x})$在展开处的Hesse矩阵$\nabla^2 f(\boldsymbol{x}^{(0)})$非奇异时，如果不满足终止条件，可继续迭代：
$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - [\nabla^2 f(\boldsymbol{x}^{(k)})]^{-1} \nabla f(\boldsymbol{x}^{(k)})$$
可简写为$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}$
- 其中，$\boldsymbol{d}^{(k)}$是线性方程组$\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)})$的解向量，通常称为Newton方程。
- 迭代直至满足终止条件。

Newton法举例-画图

$f(x_1,x_2)=x_1^4+2x_2^4$，初始点为$\boldsymbol{x}^{(0)}=[4,4]^T$

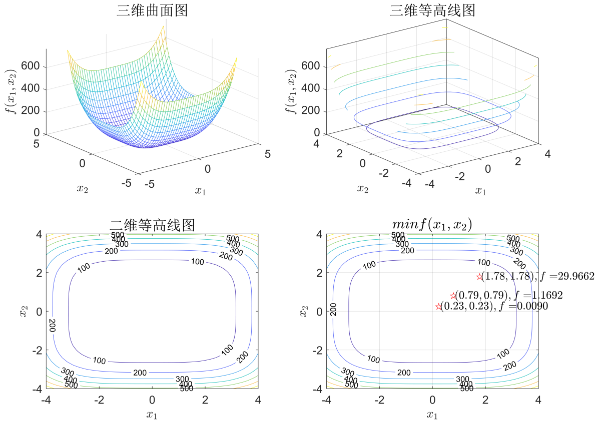

图 2: Newton法举例

Newton法举例

$$f(x_1,x_2,x_3,x_4)=(x_1+10x_2)^2+5(x_3-x_4)^2+(x_2-2x_3)^4+10(x_1-x_4)^4$$

初始点为$\boldsymbol{x}^{(0)}=[3,-1,0,1]^T$

求解：

梯度$\nabla f(\boldsymbol{x})=\begin{bmatrix} 2(x_1+10x_2)+40(x_1-x_4)^3 \\ 20(x_1+10x_2)+4(x_2-2x_3)^3 \\ 10(x_3-x_4)-8(x_2-2x_3)^3 \\ -10(x_3-x_4)-40(x_1-x_4)^3 \end{bmatrix}$

Hesse矩阵$\nabla^2 f(\boldsymbol{x})=$

$$\begin{bmatrix} 2+120(x_1-x_4)^2 & 20 & 0 & -120(x_1-x_4)^2 \\ 20 & 200+12(x_2-2x_3)^2 & -24(x_2-2x_3)^2 & 0 \\ 0 & -24(x_2-2x_3) & 10+48(x_2-2x_3)^2 & -10 \\ -120(x_1-x_4)^2 & 0 & -10 & 10+120(x_1-x_4)^2 \end{bmatrix}$$

Newton法举例

$\boldsymbol{x}^{(0)} = [3, -1, 0, 1]^T$时

梯度$\nabla f(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 306 \\ -144 \\ -2 \\ -310 \end{bmatrix}$

Hesse矩阵

$$\nabla^2 f(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 482 & 20 & 0 & -480 \\ 20 & 212 & -24 & 0 \\ 0 & -24 & 58 & -10 \\ -480 & 0 & -10 & 490 \end{bmatrix}$$

$$[\nabla^2 f(\boldsymbol{x}^{(0)})]^{-1} \nabla f(\boldsymbol{x}^{(0)}) = [1.4127, -0.8413, -0.2540, 0.7460]^T$$

$$x^{(1)} = x^{(0)} - [\nabla^2 f(\boldsymbol{x}^{(0)})]^{-1} \nabla f(\boldsymbol{x}^{(0)}) = [1.5873, -0.1587, 0.2540, 0.2540]^T$$

$$f(\boldsymbol{x}^{(1)}) = 31.8$$

Newton法举例

第2次迭代

$$x^{(2)} = x^{(1)} - [\nabla^2 f(\boldsymbol{x}^{(1)})]^{-1} \nabla f(\boldsymbol{x}^{(1)}) = [1.0582, -0.1058, 0.1694, 0.1694]^T$$

$$f(\boldsymbol{x}^{(2)}) = 6.28$$

第3次迭代

$$x^{(3)} = x^{(2)} - [\nabla^2 f(\boldsymbol{x}^{(2)})]^{-1} \nabla f(\boldsymbol{x}^{(2)}) = [0.7037, -0.0704, 0.1121, 0.1111]^T$$

$$f(\boldsymbol{x}^{(3)}) = 1.24$$

$$\cdots$$

牛顿法计算Hesse矩阵的逆矩阵比较耗时。

目录

共轭及共轭方向

- 对于n维二次型函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$，其中$\boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{Q}$是正定矩阵，即$\boldsymbol{Q} = \boldsymbol{Q}^T > 0$，求$f(\boldsymbol{x})$
- 最直观的方法，令$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}$，即求解方程$\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$即可。
- 但当数据量很大时，这种方法求解速度比较慢。

共轭及共轭方向

- 共轭方向法核心思想：在每次进行迭代时，每个方向上计算一次该方向的步长，对n维空间就进行n次计算。
- 换句话说，在每个方向上都能走到极致，这样就不会走任何的回头路了，效率也最高。

共轭方向

- 设$\boldsymbol{Q}$是$n \times n$正定矩阵，如果$\mathbb{R}^n$中的两个方向$\boldsymbol{d}^{(i)}$与$\boldsymbol{d}^{(j)}$满足$(\boldsymbol{d}^{(i)})^T\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0$，则称$\boldsymbol{d}^{(i)}$与$\boldsymbol{d}^{(j)}$关于$\boldsymbol{Q}$共轭。
- 如果$\boldsymbol{d}^{(0)}$、$\boldsymbol{d}^{(1)}$、$\cdots$、$\boldsymbol{d}^{(k-1)}, (k \leqslant n)$两两关于$\boldsymbol{Q}$共轭，即$(\boldsymbol{d}^{(i)})^T\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0, i \neq j, i, j, = 0, 1, 2, \cdots, k-1$则称$\boldsymbol{d}^{(0)}, \boldsymbol{d}^{(1)}, \cdots, \boldsymbol{d}^{(k-1)} (k \leqslant n)$为$\boldsymbol{Q}$的$k$个共轭方向。
- 若$\boldsymbol{d}^{(i)} \neq \boldsymbol{0}, i, j, = 0, 1, 2, \cdots, k-1$，则称为$\boldsymbol{Q}$的$k$个非零共轭方向。特别，当$\boldsymbol{Q} = \boldsymbol{I}$时，共轭方向即为正交方向。

共轭方向

- 设Q是$n \times n$正定矩阵，如果$\boldsymbol{d}^{(0)}$、$\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$、$\cdots$、$\boldsymbol{d}^{(k-1)}, (k \leqslant n)$非$\mathbf{0}$且两两关于$Q$共轭，则$\boldsymbol{d}^{(0)}$、$\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$、$\cdots$、$\boldsymbol{d}^{(k-1)}(k \leqslant n)$线性无关。

- 证明：
  假设线性相关，则存在一组标量α_0、α_1、α_2、$\cdots$、α_{k-1}，使得$\sum_{i=0}^{k-1} \alpha_i \boldsymbol{d}^{(i)} = \mathbf{0}$
  等式左右同乘以$\boldsymbol{d}^{(j)T}\boldsymbol{Q}, 0 \leqslant j \leqslant k-1$，则$\sum_{i=0,i\neq j}^{k-1} \alpha_i \boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} + \alpha_j \boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = \mathbf{0}$，
  由于共轭性质可知$\sum_{i=0,i\neq j}^{k-1} \alpha_i \boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0$，由于正定性质可知$\boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} > 0$，所以
  $\sum_{i=0,i\neq j}^{k-1} \alpha_i \boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} + \boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = \mathbf{0}$不成立。
  所以，共轭方向线性无关（可构成基底）。

共轭方向举例

求矩阵$\boldsymbol{Q}$的一组共轭向量。

$$\boldsymbol{Q}=\begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

首先验证$\boldsymbol{Q}$是否为正定矩阵：

$$\Delta_1=3>0,\Delta_2=\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}>0,\Delta_3=det\boldsymbol{Q}=20>0$$

所以$\boldsymbol{Q}$的所有顺序主子式都为正数，且$\boldsymbol{Q}=\boldsymbol{Q}^T$，所以$\boldsymbol{Q}$正定。设$\boldsymbol{d}^{(0)}=[1,0,0]^T$，
$\boldsymbol{d}^{(1)}=[x_1,y_1,z_1]^T$，$\boldsymbol{d}^{(2)}=[x_2,y_2,z_2]^T$，则
$\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(1)}=3x_1+z_1=0$，令$x_1=1$则$\boldsymbol{d}^{(1)}=[1,0,-3]^T$；
$\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(2)}=3x_2+z_2=0$，
$\boldsymbol{d}^{(1)T}\boldsymbol{Q}\boldsymbol{d}^{(2)}=-6y_2-8z_2=0$，
令$x_3=1$则$\boldsymbol{d}^{(2)}=[1,4,-3]^T$
因此可得一组共轭向量$\boldsymbol{d}^{(0)}=[1,0,0]^T$、$\boldsymbol{d}^{(1)}=[1,0,-3]^T$、$\boldsymbol{d}^{(2)}=[1,4,-3]^T$。

共轭方向的几何意义1

- 设有二次函数$f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\boldsymbol{Q}(\boldsymbol{x}-\bar{\boldsymbol{x}})$，其中$\boldsymbol{Q}$是$n \times n$对称正定矩阵，$\bar{\boldsymbol{x}}$是一个定点

- 函数$f(\boldsymbol{x})$的等值面$\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\boldsymbol{Q}(\boldsymbol{x}-\bar{\boldsymbol{x}}) = c$是以$\bar{\boldsymbol{x}}$为重心的椭球面

- 由于$\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{Q}(\boldsymbol{x}-\bar{\boldsymbol{x}}) = \boldsymbol{0}$且$\boldsymbol{Q}$正定，所以$f(\boldsymbol{x})$是凸函数且极小值在$\bar{\boldsymbol{x}}$取到

- 设$\boldsymbol{x}^{(1)}$是某个等值面上的一点，该等值面在$\boldsymbol{x}^{(1)}$处的法向量为$\nabla f(\boldsymbol{x}^{(1)}) = \boldsymbol{Q}(\boldsymbol{x}^{(1)}-\bar{\boldsymbol{x}})$，梯度即为法向量的方向，切向量的方向与法向量的方向正交，记切向量为$\boldsymbol{d}^{(1)}$，$\nabla f(\boldsymbol{x}^{(1)})^T\boldsymbol{d}^{(1)} = 0$，令$\boldsymbol{d}^{(2)} = \boldsymbol{x}^{(1)}-\bar{\boldsymbol{x}}$，则有$\boldsymbol{d}^{(1)T}\boldsymbol{Q}\boldsymbol{d}^{(2)} = 0$

共轭方向的几何意义1

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Q共轭

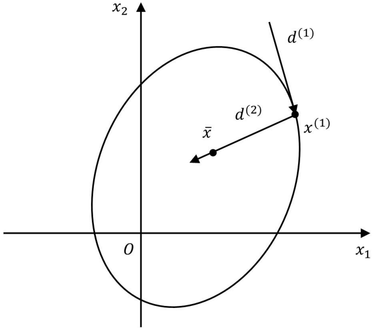

图 3: 共轭方向的几何意义1

共轭方向的几何意义2

- 对于二维向量$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$，其几何变换可以用左乘一个$2\times2$矩阵来表示
- 典型变换包括：缩放$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$，旋转$\begin{bmatrix} cos\alpha & -sin\alpha \\ sin\alpha & cos\alpha \end{bmatrix}$，沿$x$轴剪切$\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$，沿$y$轴剪切$\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}$
- 假设$\boldsymbol{x}$与$\boldsymbol{y}$关于$\boldsymbol{Q}$共轭，其中$\boldsymbol{Q}$正定，则有$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{y} = 0$，可理解为$(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{x})^T(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{y})^T = 0$，$\boldsymbol{Q}^{\frac{1}{2}}$可认为是一种几何变换
- $\boldsymbol{x}$与$\boldsymbol{y}$关于$\boldsymbol{Q}$共轭$\leftrightarrow$ $\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{x}$与$\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{y}$正交

共轭方向的几何意义2

例：$f(x,y)=x^2+xy+y^2$，则可知 $Q=\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$，且$Q$正定

求Q的方法是将Q的特征值λ_1,λ_2和特征向量$\boldsymbol{a}_1,\boldsymbol{a}_2$求出，

$$\begin{bmatrix} \boldsymbol{a}_1,\boldsymbol{a}_2 \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \end{bmatrix}$$

在本例中

$$\boldsymbol{Q}=\begin{bmatrix} \sqrt{2}/2,\sqrt{2}/2 \\ -\sqrt{2}/2,\sqrt{2}/2 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & \sqrt{3}/\sqrt{2} \end{bmatrix}^2\begin{bmatrix} \sqrt{2}/2,-\sqrt{2}/2 \\ \sqrt{2}/2,\sqrt{2}/2 \end{bmatrix}$$

$$=\begin{bmatrix} 1/2,\sqrt{3}/2 \\ -1/2,\sqrt{3}/2 \end{bmatrix}\begin{bmatrix} 1/2,-1/2 \\ \sqrt{3}/2,\sqrt{3}/2 \end{bmatrix}$$

假设一个方向为$[1,0]^T$，则其共轭方向为$[1,-2]^T$

目录

基本共轭方向法

- 给定初始点$\boldsymbol{x}^{(0)}$和一组关于Q共轭的方向$\boldsymbol{d}^{(0)}$、$\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$、$\cdots$、$\boldsymbol{d}^{(n-1)}$，迭代公式为：
  $\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)}$ 如何求解α_i？

- 令$\frac{\partial f(\boldsymbol{x}^{(i+1)})}{\partial \alpha_i} = 0$
  则$f(\boldsymbol{x}^{(i+1)}) = \frac{1}{2}(\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)})^T \boldsymbol{Q}(\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)}) - \boldsymbol{b}^T(\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)})$
  $\frac{\partial f(\boldsymbol{x}^{(i+1)})}{\partial \alpha_i} = \boldsymbol{d}^{(i)T} Q(\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)}) - \boldsymbol{d}^{(i)T}\boldsymbol{b} = \boldsymbol{d}^{(i)T} Q\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)T} Q\boldsymbol{d}^{(i)} - \boldsymbol{d}^{(i)T}\boldsymbol{b} = 0$

- 得到$\alpha_i = -\frac{\boldsymbol{d}^{(i)T}(Q\boldsymbol{x}^{(i)} - \boldsymbol{b})}{\boldsymbol{d}^{(i)T} Q\boldsymbol{d}^{(i)}} = -\frac{\boldsymbol{d}^{(i)T}\nabla f(\boldsymbol{x}^{(i)})}{\boldsymbol{d}^{(i)T} Q\boldsymbol{d}^{(i)}}$
  即$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(0)} + \sum_{i=0}^{k-1} \alpha_i \boldsymbol{d}^{(i)}$

基本共轭方向法

- 求解α_i的另一种方法是$\frac{\partial f(\boldsymbol{x}^{(i+1)})}{\partial \alpha_i} = \nabla f(\boldsymbol{x}^{(i+1)})^T \boldsymbol{d}^{(i)} = (\boldsymbol{Q}\boldsymbol{x}^{(i+1)} - \boldsymbol{b})^T \boldsymbol{d}^{(i)} = 0$
- $[\boldsymbol{Q}(\boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)}) - \boldsymbol{b}]^T \boldsymbol{d}^{(i)} = 0$
- $[\nabla f(\boldsymbol{x}^{(i)}) + \alpha_i \boldsymbol{Q}\boldsymbol{d}^{(i)}]^T \boldsymbol{d}^{(i)} = 0$
- 最终可得到$\alpha_i = -\frac{\boldsymbol{d}^{(i)T} \nabla f(\boldsymbol{x}^{(i)})}{\boldsymbol{d}^{(i)T} \boldsymbol{Q}\boldsymbol{d}^{(i)}}$
- 证明过程中我们注意到$\nabla f(\boldsymbol{x}^{(i+1)})^T \boldsymbol{d}^{(i)} = 0$

基本共轭方向法

- 如何证明基本共轭方向法n次迭代之后的求解结果$\boldsymbol{x}^{(n)}$是局部最小点$\boldsymbol{x}^*$?

- 已知$\boldsymbol{d}^{(i)}, i = 0, 1, \cdots, n-1$是基底，所以$\boldsymbol{x}^* - \boldsymbol{x}^{(0)}$可表示为$\beta_0\boldsymbol{d}^{(0)} + \beta_1\boldsymbol{d}^{(1)} + \cdots + \beta_{n-1}\boldsymbol{d}^{(n-1)}$
- 左乘$\boldsymbol{d}^{(k)T}\boldsymbol{Q}, 0 \leqslant k < n$，由于共轭性质，$\boldsymbol{d}^{(k)T}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}^{(0)}) = \beta_k\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}$
  得$\beta_k = \frac{\boldsymbol{d}^{(k)T}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}^{(0)})}{\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}} = \frac{\boldsymbol{d}^{(k)T}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}^{(k)} + \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(0)})}{\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}}$
- 其中，由共轭性得$\boldsymbol{d}^{(k)T}\boldsymbol{Q}(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(0)}) = \sum_{i=0}^{k-1}\alpha_i\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0$
- 另外，$\boldsymbol{d}^{(k)T}\boldsymbol{Q}(\boldsymbol{x}^* - \boldsymbol{x}^{(k)}) = \boldsymbol{d}^{(k)T}(\boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{b}) - \boldsymbol{d}^{(k)T}(\boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}) = \boldsymbol{d}^{(k)T}(\nabla f(\boldsymbol{x}^*) - \nabla f(\boldsymbol{x}^{(k)})) = -\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^{(k)})$
- 所以$\beta_k = \frac{-\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}} = \alpha_k$，所以$\boldsymbol{x}^{(n)}$是局部最小点$\boldsymbol{x}^*$

基本共轭方向法

- 在共轭方向算法中，对所有的$k, 0 \leqslant k \leqslant n-1, 0 \leqslant i \leqslant k$都有$\nabla f(\boldsymbol{x}^{(k+1)})^T \boldsymbol{d}^{(i)} = 0$

- 证明：
  由于$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$，$\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)}) = ((\boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b}) - (\boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b})) = \boldsymbol{Q}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) = \alpha_k \boldsymbol{Q}\boldsymbol{d}^{(k)}$
  用数学归纳法对本命题继续证明，共分三步，第一步，证明$\nabla f(\boldsymbol{x}^{(1)})^T \boldsymbol{d}^{(0)} = 0$，第二步，证明$\nabla f(\boldsymbol{x}^{(k+1)})^T \boldsymbol{d}^{(i)} = 0, 0 \leqslant i < k$；第三步，证明$\nabla f(\boldsymbol{x}^{(k+1)})^T \boldsymbol{d}^{(k)} = 0$。

- 第一步：已知$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{d}^{(0)} = \boldsymbol{x}^{(0)} - \frac{\boldsymbol{d}^{(0)T} \nabla f(\boldsymbol{x}^{(0)})}{\boldsymbol{d}^{(0)T} \boldsymbol{Q} \boldsymbol{d}^{(0)}} \boldsymbol{d}^{(0)}$所以$\nabla f(\boldsymbol{x}^{(1)})^T \boldsymbol{d}^{(0)} = (\boldsymbol{Q}\boldsymbol{x}^{(1)} - \boldsymbol{b})^T \boldsymbol{d}^{(0)} = (\boldsymbol{x}^{(1)T}\boldsymbol{Q} - \boldsymbol{b}^T)\boldsymbol{d}^{(0)} = (\boldsymbol{x}^{(0)T} + \alpha_0 \boldsymbol{d}^{(0)T})\boldsymbol{Q}\boldsymbol{d}^{(0)} - \boldsymbol{b}^T \boldsymbol{d}^{(0)} = \nabla f(\boldsymbol{x}^{(0)})^T \boldsymbol{d}^{(0)} + \alpha_0 \boldsymbol{d}^{(0)T} \boldsymbol{Q} \boldsymbol{d}^{(0)} = 0$，第一步得证。

基本共轭方向法

- 第二步：假设对任意 $i < k-1$，有$\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(i)} = 0$，考虑到$\boldsymbol{d}^{(k)},\boldsymbol{d}^{(i)}$关于$\boldsymbol{Q}$共轭，那么对任意$i < k$有$\nabla f(\boldsymbol{x}^{(k+1)})^T\boldsymbol{d}^{(i)} = (\nabla f(\boldsymbol{x}^{(k)})^T + \alpha_k\boldsymbol{d}^{(k)T}\boldsymbol{Q})\boldsymbol{d}^{(i)} = \nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(i)} + \alpha_k\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0+0 = 0$，第二步得证。

- 第三步：$\nabla f(\boldsymbol{x}^{(k+1)})^T\boldsymbol{d}^{(k)} = (\boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b})^T\boldsymbol{d}^{(k)}$ $= (\boldsymbol{x}^{(k)} - \frac{\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}}\boldsymbol{d}^{(k)})^T\boldsymbol{Q}\boldsymbol{d}^{(k)} - \boldsymbol{b}^T\boldsymbol{d}^{(k)} = (\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k)}))\boldsymbol{d}^{(k)} = 0$

- 所以，第$k+1$点的梯度与之前所有的$\boldsymbol{d}^{(i)}, 0 \leqslant i \leqslant k$都正交（也可以此方法来证明共轭方向法的解是最优解）。

基本共轭方向法举例

$f(x_1,x_2) = \frac{1}{2}\boldsymbol{x}^T\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\boldsymbol{x} - [-1,1]\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n$ 求$f(\boldsymbol{x})$的极小点，初始点为$\boldsymbol{x}^{(0)} = [0,0]^T$，给定共轭方向$\boldsymbol{d}^{(0)} = [1,0]^T, \boldsymbol{d}^{(1)} = [-\frac{3}{8}, \frac{3}{4}]^T$

- 首先计算$\nabla f(\boldsymbol{x}^{(0)}) = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b} = [1,-1]^T$

$$\alpha_0 = -\frac{\nabla f(\boldsymbol{x}^{(0)})^T\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = -\frac{[1,-1]\begin{bmatrix} 1 \\ 0 \end{bmatrix}}{[1,0]\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = -\frac{1}{4}$$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0\boldsymbol{d}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix}$$

基本共轭方向法举例

- $\nabla f(\boldsymbol{x}^{(1)}) = \boldsymbol{Q}\boldsymbol{x}^{(1)} - \boldsymbol{b} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{3}{2} \end{bmatrix}$

$$\alpha_1 = -\frac{\nabla f(\boldsymbol{x}^{(1)})^T \boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)T}\boldsymbol{Q}\boldsymbol{d}^{(1)}} = -\frac{[0, -\frac{3}{2}]\begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}}{[-\frac{3}{8}, \frac{3}{4}]\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}} = 2$$

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

- 由于函数f是一个二维的二次型函数，故$\boldsymbol{x}^{(2)} = \boldsymbol{x}^*$。

三种方法的对比

对于二次型$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}\boldsymbol{x}$其中$\boldsymbol{x}$为二阶。

最速下降法通过多次迭代可达到最优；Newton法通过1次迭代可达到最优；共轭方向法通过2次迭代可达到最优。

对于共轭方向法，$\boldsymbol{x}^* = \boldsymbol{x}^{(1)} + t_1\boldsymbol{d}^{(1)}$，$\nabla f(\boldsymbol{x}^{(1)})^T\boldsymbol{d}^{(0)} = 0$，

又有$\nabla f(\boldsymbol{x}^{(1)}) = \boldsymbol{Q}\boldsymbol{x}^{(1)} - \boldsymbol{b}$，$\nabla f(\boldsymbol{x}^*) = \boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{b} = \boldsymbol{0} = \boldsymbol{Q}(\boldsymbol{x}^{(1)} + t_1\boldsymbol{d}^{(1)}) - \boldsymbol{b}$

所以$\nabla f(\boldsymbol{x}^{(1)}) = \nabla f(\boldsymbol{x}^*) - t_1\boldsymbol{Q}\boldsymbol{d}^{(1)}$，$\nabla f(\boldsymbol{x}^{(1)})^T\boldsymbol{d}^{(0)} = -t_1\boldsymbol{d}^{(1)T}\boldsymbol{Q}\boldsymbol{d}^{(0)} = 0$

三种方法的对比

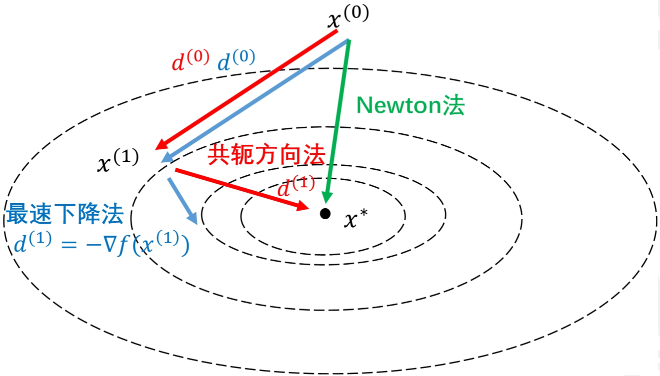

图 4: 三种方法的对比

目录

共轭梯度法

- 如果并未事先指定一组共轭向量，那么如何求解$\boldsymbol{d}^{(k)}$？

- $k=0$时，$\boldsymbol{d}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})$
  每一步的方向都是在起点的负梯度方向的基础上进行的调整，类似施密特正交化$\boldsymbol{b}'=\boldsymbol{b}-\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}$

- $k>0$时，我们希望求得从$\boldsymbol{d}^{(k)}$到$\boldsymbol{d}^{(k+1)}$的迭代公式，
  即构造$\boldsymbol{d}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})+\beta_{k-1}\boldsymbol{d}^{(k-1)}$
  由于$\boldsymbol{d}^{(k)}$与$\boldsymbol{d}^{(k-1)}$共轭，所以$\boldsymbol{d}^{(k-1)T}Q\boldsymbol{d}^{(k)}=0$
  代入得到 $\boldsymbol{d}^{(k-1)T}Q(-\nabla f(\boldsymbol{x}^{(k)})+\beta_{k-1}\boldsymbol{d}^{(k-1)})=-\boldsymbol{d}^{(k-1)T}Q\nabla f(\boldsymbol{x}^{(k)})+\beta_{k-1}\boldsymbol{d}^{(k-1)T}Q\boldsymbol{d}^{(k-1)}=0$，即$\beta_{k-1}=\frac{\boldsymbol{d}^{(k-1)T}Q\nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k-1)T}Q\boldsymbol{d}^{(k-1)}}$

共轭梯度法

- 所以$k > 0$时，每一次的方向为$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) + \frac{\boldsymbol{d}^{(k-1)T} Q \nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k-1)T} Q \boldsymbol{d}^{(k-1)}} \boldsymbol{d}^{(k-1)}$
- 或写为$\boldsymbol{d}^{(k+1)} = -\nabla f(\boldsymbol{x}^{(k+1)}) + \frac{\boldsymbol{d}^{(k)T} Q \nabla f(\boldsymbol{x}^{(k+1)})}{\boldsymbol{d}^{(k)T} Q \boldsymbol{d}^{(k)}} \boldsymbol{d}^{(k)}$。

共轭梯度法

我们先给定共轭梯度法的一般求解步骤，再来证明这种方法的合理性（即$\boldsymbol{d}^{(0)}, \boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n-1)}$共轭）。

一般步骤：

1. 令$k=0$，选择初始值$\boldsymbol{x}^{(0)}$；
2. 计算$\nabla f(\boldsymbol{x}^{(0)})$，如果$\nabla f(\boldsymbol{x}^{(0)}) = 0$则停止迭代，否则，令$\boldsymbol{d}^{(0)} = -\nabla f(\boldsymbol{x}^{(0)})$；
3. 计算$\alpha_k = -\dfrac{\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k)T}Q\boldsymbol{d}^{(k)}}$；
4. 计算$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}$；
5. 计算$\nabla f(\boldsymbol{x}^{(k+1)})$，如果$\nabla f(\boldsymbol{x}^{(k+1)}) = 0$则停止迭代；
6. 计算$\beta_k = \dfrac{\boldsymbol{d}^{(k)T}Q\nabla f(\boldsymbol{x}^{(k+1)})}{\boldsymbol{d}^{(k)T}Q\boldsymbol{d}^{(k)}}$；
7. 计算$\boldsymbol{d}^{(k+1)} = -\nabla f(\boldsymbol{x}^{(k+1)}) + \beta_k\boldsymbol{d}^{(k)}$；
8. 令$k=k+1$，回到第3步。

共轭梯度法

假设共轭梯度法在m次迭代后终止，则对所有的$1 \leqslant i \leqslant m$有如下三个关系成立：

1. $\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0, (j = 1, 2, , \cdots, i-1)$，即$\boldsymbol{d}^{(j)}$与$\boldsymbol{d}^{(i)}$是共轭方向
2. $\nabla f(\boldsymbol{x}^{(i)})^T \nabla f(\boldsymbol{x}^{(j)}) = 0, (j = 1, 2, , \cdots, i-1)$
3. $\nabla f(\boldsymbol{x}^{(i)})^T \boldsymbol{d}^{(i)} = -\nabla f(\boldsymbol{x}^{(i)})^T \nabla f(\boldsymbol{x}^{(i)})$，其中$\boldsymbol{d}^{(i)} \neq \boldsymbol{0}$

共轭梯度法

证明：显然$m \geqslant 1$，现在用归纳法证明上述三个关系。对i归纳。

- 当$i = 1$时，由于$\nabla f(\boldsymbol{x}^{(1)}) = -\boldsymbol{d}^{(1)}$，因此关系（3）成立
- 当$i = 2$时，$\boldsymbol{d}^{(2)} = -\nabla f(\boldsymbol{x}^{(2)}) + \beta_1 \boldsymbol{d}^{(1)}$，所以等式左右各乘以$\boldsymbol{d}^{(1)T}\boldsymbol{Q}$得$\boldsymbol{d}^{(1)T}\boldsymbol{Q}\boldsymbol{d}^{(2)} = 0$，代入$\beta_1$的定义可知（1）成立
- 当$i = 2$时，$\nabla f(\boldsymbol{x}^{(1)})^T \nabla f(\boldsymbol{x}^{(2)}) = \nabla f(\boldsymbol{x}^{(1)})^T [\nabla f(\boldsymbol{x}^{(1)}) + \alpha_1 \boldsymbol{Q}\boldsymbol{d}^{(1)}] = 0$，其中用到了$\nabla f(\boldsymbol{x}^{(i)}) = \boldsymbol{Q}\boldsymbol{x}^{(i)} + \boldsymbol{b}$这个性质以及$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} - \alpha_i \boldsymbol{d}^{(i)}$这个性质，（2）成立
- 当$i = 2$时，$\nabla f(\boldsymbol{x}^{(2)})^T \boldsymbol{d}^{(2)} = \nabla f(\boldsymbol{x}^{(2)})^T (-\nabla f(\boldsymbol{x}^{(2)}) + \beta_1 \boldsymbol{d}^{(1)}) = -\nabla f(\boldsymbol{x}^{(2)})^T \nabla f(\boldsymbol{x}^{(2)})$，（3）成立
- 设对某个$i < m$，这些关系都成立，接下来我们证明对于$i+1$也成立

中国農業大學 China Agricultural University

共轭梯度法

- 先证明（2），因为$\nabla f(\boldsymbol{x}^{(i)}) = \boldsymbol{Q}\boldsymbol{x}^{(i)} - \boldsymbol{b}$这个性质以及$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \alpha_i \boldsymbol{d}^{(i)}$，所以$\nabla f(\boldsymbol{x}^{(i+1)}) = \nabla f(\boldsymbol{x}^{(i)}) + \alpha_i \boldsymbol{Q}\boldsymbol{d}^{(i)}$
- 所以，根据第（3）个关系可知$\alpha_i = -\frac{\boldsymbol{d}^{(i)T}\nabla f(\boldsymbol{x}^{(i)})}{\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i)}} = \frac{\nabla f(\boldsymbol{x}^{(i)})^T\nabla f(\boldsymbol{x}^{(i)})}{\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i)}}$
- 所以$\nabla f(\boldsymbol{x}^{(i+1)})^T\nabla f(\boldsymbol{x}^{(j)}) = [\nabla f(\boldsymbol{x}^{(i)}) + \alpha_i \boldsymbol{Q}\boldsymbol{d}^{(i)}]^T\nabla f(\boldsymbol{x}^{(j)}) = \nabla f(\boldsymbol{x}^{(i)})^T\nabla f(\boldsymbol{x}^{(j)}) + \alpha_i \boldsymbol{d}^{(i)T}\boldsymbol{Q}(-\boldsymbol{d}^{(j)} + \beta_{j-1}\boldsymbol{d}^{(j-1)})$，其中注意$j = 1$时$\nabla f(\boldsymbol{x}^{(i+1)})^T\nabla f(\boldsymbol{x}^{(1)}) = \nabla f(\boldsymbol{x}^{(i)})^T\nabla f(\boldsymbol{x}^{(1)}) - \alpha_i \boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(1)}$
- 当$j = i$时，由归纳法假设$\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i-1)} = 0$以及$\alpha_i = \frac{\nabla f(\boldsymbol{x}^{(i)})^T\nabla f(\boldsymbol{x}^{(i)})}{\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i)}}$可知$\nabla f(\boldsymbol{x}^{(i+1)})^T\nabla f(\boldsymbol{x}^{(i)}) = 0$
- 当$j < i$时，根据归纳法假设可知$\nabla f(\boldsymbol{x}^{(i)})^T\nabla f(\boldsymbol{x}^{(j)}) = 0, \boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0, \boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(j-1)} = 0$
- 因此$\nabla f(\boldsymbol{x}^{(i+1)})^T\nabla f(\boldsymbol{x}^{(j)}) = 0$，（2）得证

共轭梯度法

- 再证明（1）
- $\boldsymbol{d}^{(i+1)}\boldsymbol{Q}\boldsymbol{d}^{(j)} = (-\nabla f(\boldsymbol{x}^{(i+1)}) + \beta_i\boldsymbol{d}^{(i)})\boldsymbol{Q}\boldsymbol{d}^{(j)} = -\nabla f(\boldsymbol{x}^{(i+1)})^T \frac{\nabla f(\boldsymbol{x}^{(j+1)})-\nabla f(\boldsymbol{x}^{(j)})}{\alpha_j} + \beta_i\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(j)}$
- 当$j=i$时，代入$\beta_i = \frac{\boldsymbol{d}^{(i)T}Q\nabla f(\boldsymbol{x}^{(i+1)})}{\boldsymbol{d}^{(i)T}Q\boldsymbol{d}^{(i)}}$可知$\boldsymbol{d}^{(i+1)}\boldsymbol{Q}\boldsymbol{d}^{(i)}=0$
- 当$j<i$时，由已证明的结论和归纳法假设可知等号右边为0
- 因此$\boldsymbol{d}^{(i+1)}\boldsymbol{Q}\boldsymbol{d}^{(j)}=0$，（1）得证

共轭梯度法

- 最后证明（3）
- 由于前边已证明$\nabla f(\boldsymbol{x}^{(i+1)})^T\boldsymbol{d}^{(i)}=0$
- 所以$\nabla f(\boldsymbol{x}^{(i+1)})^T\boldsymbol{d}^{(i+1)} = \nabla f(\boldsymbol{x}^{(i+1)})^T(-\nabla f(\boldsymbol{x}^{(i+1)}) + \beta_i\boldsymbol{d}^{(i)}) = -\nabla f(\boldsymbol{x}^{(i+1)})^T\nabla f(\boldsymbol{x}^{(i+1)})$
- 因此（3）得证

共轭梯度法举例

- 利用共轭梯度法求解$f(x_1,x_2,x_3)$最小值，初始点为$\boldsymbol{x}^{(0)} = [0,0,0]^T$。
  $f(x_1,x_2,x_3) = \frac{3}{2}x_1^2 + 2x_2^2 + \frac{3}{2}x_3^2 + x_1x_3 + 2x_2x_3 - 3x_1 - x_3$
- 函数f可以写为$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$ 其中$\boldsymbol{Q} = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$
- f的梯度为$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b} = [3x_1 + x_3 - 3, 4x_2 + 2x_3, x_1 + 2x_2 + 3x_3 - 1]^T$
  $\nabla f(\boldsymbol{x}^{(0)}) = [-3,0,-1]^T, \boldsymbol{d}^{(0)} = -\nabla f(\boldsymbol{x}^{(0)}), \alpha_0 = -\frac{\nabla f(\boldsymbol{x}^{(0)})^T\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = 0.2778$,
  新的迭代点为$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0\boldsymbol{d}^{(0)} = [0.8333, 0, 0.2778]^T$
- 可得$\nabla f(\boldsymbol{x}^{(1)}) = [-0.2222, 0.5556, 0.6667]^T$ $\beta_0 = \frac{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\nabla f(\boldsymbol{x}^{(1)})}{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = 0.08025$
  $\boldsymbol{d}^{(1)} = -\nabla f(\boldsymbol{x}^{(1)}) + \beta_0\boldsymbol{d}^{(0)} = [0.4630, -0.5556, -0.5864]^T$

共轭梯度法举例

- $\alpha_1 = -\frac{\nabla f(\boldsymbol{x}^{(1)})^T \boldsymbol{d}^{(1)}}{\boldsymbol{d}^{(1)T} \boldsymbol{Q} \boldsymbol{d}^{(1)}} = 0.2187$
- 新的迭代点为$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{d}^{(1)} = [0.9436, -0.1215, 0.1495]^T$
- 可得$\nabla f(\boldsymbol{x}^{(2)}) = [-0.04673, -0.1896, 0.1402]^T$
  $\beta_1 = \frac{\boldsymbol{d}^{(1)T} Q \nabla f(\boldsymbol{x}^{(2)})}{\boldsymbol{d}^{(1)T} \boldsymbol{Q} \boldsymbol{d}^{(1)}} = 0.07075$
  $\boldsymbol{d}^{(2)} = -\nabla f(\boldsymbol{x}^{(2)}) + \beta_1 \boldsymbol{d}^{(1)} = [0.07948, 0.1476, -0.1817]^T$
  $\alpha_2 = -\frac{\nabla f(\boldsymbol{x}^{(2)})^T \boldsymbol{d}^{(2)}}{\boldsymbol{d}^{(2)T} \boldsymbol{Q} \boldsymbol{d}^{(2)}} = 0.8231$
- 最终结果为$\boldsymbol{x}^* = \boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \alpha_2 \boldsymbol{d}^{(2)} = [1.000, 0.000, 0.000]^T$
- 此时$\nabla f(\boldsymbol{x}^{(3)}) = [1, 0, 0]^T$

Thank you for your attention!

翟卫欣　副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院