运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

- 牛顿法 搜索方向$d^k=-\left[\nabla^2 f\left(x^k\right)\right]^{-1}\nabla f\left(x^k\right)$即为牛顿方向
- 修正的牛顿法

摒弃了步长恒为1的做法，而是采用线搜索确定最优步长来构造算法

缺点：

1. 要求海塞矩阵正定；
2. 需要计算海塞矩阵及其逆，计算量大

- 拟牛顿法

拟牛顿方程(基本要求)：

$$\nabla f\left(\boldsymbol{x}^{k+1}\right)-\nabla f\left(\boldsymbol{x}^k\right)=\boldsymbol{B}_{k+1}\left(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\right)$$

搜索方向$\boldsymbol{d}^k=-\boldsymbol{B}_k^{-1}\nabla f\left(\boldsymbol{x}^k\right)$

第二类方法：对$\boldsymbol{B}_k$(或$\mathbf{H}_k$)进行校正，令$\boldsymbol{B}_{k+1}=\boldsymbol{B}_k+\Delta\boldsymbol{B}$

- Rank-2校正，要求$\Delta\boldsymbol{B}$的秩为2：DFP方法，BFGS方法；
- Rank-1校正，要求$\Delta\boldsymbol{B}$的秩为1；

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k-\frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k^T}{\boldsymbol{y}_k^T\boldsymbol{H}_k\boldsymbol{y}_k}+\frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\boldsymbol{s}_k^T\boldsymbol{y}_k}\qquad\text{(DFP)}$$

$$\boldsymbol{B}_{k+1}=\boldsymbol{B}_k+\frac{\boldsymbol{y}_k\boldsymbol{y}_k^T}{\boldsymbol{y}_k^T\boldsymbol{s}_k}-\frac{\boldsymbol{B}_k\boldsymbol{s}_k\boldsymbol{s}_k^T\boldsymbol{B}_k}{\boldsymbol{s}_k^T\boldsymbol{B}_k\boldsymbol{s}_k}\qquad\text{(BFGS)}$$

课程内容

- 共轭方向法
- 共轭梯度法

二次型及二次终结性

- 求解n元二次正定函数极小化问$\min\frac{1}{2}x^TQx+b^Tx, Q\succ 0, b\in R^n$常常用来衡量算法。因为

1. 正定二次函数是容易确定极小值的最简单光滑函数；
2. 一般光滑函数在局部极小点x^*附近可用正定二次函数很好的逼近；
3. 在给定精度下，用二次函数逼近比用线性函数可在较大的区域内有效。

- 二次终结性：如果一个算法用于正定二次函数求极小时，能够通过有限步迭代达到最小点x^*，称该算法具有二次终结性。

- 二次终结性=共轭方向+精确一维搜索

共轭方向法

- 求解n元二次正定函数极小化问题

$$\min \frac{1}{2} x^T Q x + b^T x, Q \succ 0, b \in R^n$$

$$\Leftrightarrow \nabla f(x) = Qx + b = 0$$

迭代法求解上面问题的方法称之为“线性”共轭梯度法

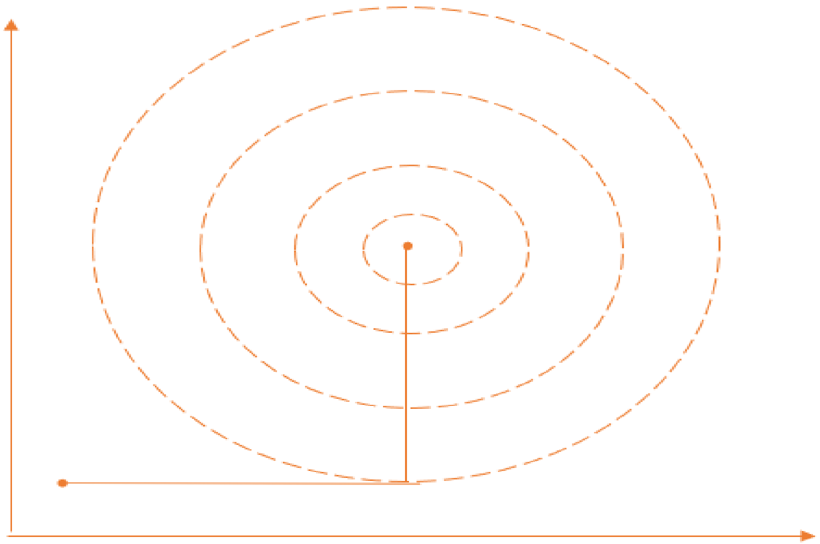
求解最小化一般函数的方法称之为“非线性”共轭梯度法

线性共轭梯度法

➢考虑最优化问题$\min\frac{1}{2}x^TQx+b^Tx,\quad Q\succ 0,b\in R^n$

1. 当矩阵Q是对角阵时，$f(x)=\frac{1}{2}x^T\begin{pmatrix}q_1 & & \\ & \ddots & \\ & & q_n\end{pmatrix}x+\mathrm{b}^Tx$

当$n=2$, $f(x)$等值线如下图



线性共轭梯度法

2. 当Q非对角时，则对于$f(x)=\frac{1}{2}x^TQx+b^Tx$

对Q特征值分解有$Q=P^TDP, D=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$, P为正交矩阵

设$\hat{x}=Px$, 则有$f(\hat{x})=\frac{1}{2}\hat{x}^TD\hat{x}+b^TP^T\hat{x}$

得到最优解$\hat{x}^*$, 则有$P^T\hat{x}^*=x^*$

寻找向量组$(d^0,d^1,\cdots,d^{n-1})=S$可以对角化矩阵$Q$

向量组间的向量共轭

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

共轭方向

- 共轭方向：考虑正定矩阵Q及非零向量d^i, d^j, 如果

$$\left(d^i\right)^T Q d^j = 0$$

则称d^i, d^j关于矩阵Q共轭

- 向量组$d^0, d^1, \cdots, d^k$关于矩阵Q共轭
- 共轭与正交的关系：

如果$d^0, d^1, \cdots, d^k$关于单位矩阵I共轭，则$d^0, d^1, \cdots, d^k$正交

如果$d^0, d^1, \cdots, d^k$关于Q共轭，$Q = P^2, P$正定，则$\left(d^i\right)^T Q d^j = \left(d^i\right)^T P^T P d^j = \left(P d^i\right)^T \left(P d^j\right) = 0$, 即通过$P$变换后可以使它们两两正交

共轭向量组必定线性无关

共轭方向法

- 对于无约束优化问题 $\min f(x)=\frac{1}{2}x^TQx+b^Tx, Q\succ 0$
- 基本思想：

给定初始点x^0及一组关于Q共轭方向$d^0,d^1,\cdots,d^{n-1}$，令

$$x^{k+1}=x^k+\alpha_k d^k, k=0,\cdots,n-1$$

其中，

$$\alpha_k=\arg\min\phi(\alpha)=f(x^k+\alpha d^k)$$

令$\phi'(\alpha_k)=0$

$$\Rightarrow \nabla f\left(x^k+\alpha_k d^k\right)^T d^k=0$$

$$\Rightarrow \left(Q\left(x^k+\alpha_k d^k\right)+b\right)^T d^k=0$$

$$\Rightarrow \left(Qx^k+b\right)^T d^k+\alpha_k\left(d^k\right)^T Qd^k=0$$

也得到$\alpha_k=-\frac{\left(Qx^k+b\right)^T d^k}{\left(d^k\right)^T Qd^k}=-\frac{\nabla f\left(x^k\right)^T d^k}{\left(d^k\right)^T Qd^k}$ （*）

共轭方向法是一类方法，共轭梯度法是其中的一种方法

共轭方向法

➢几何解释：

在生成点列$\{x_0,\cdots,x_n\}$的过程$x^0+\alpha_0 d^0+\alpha_1 d^1+\cdots+\alpha_{n-1}d^{n-1}$中有方向$d^0,d^1,\cdots,d^{n-1}$, $\{d^0,\cdots,d^{n-1}\}=S$，S可逆

1. $S^TQS=\begin{pmatrix}(d^0)^T\\ \vdots\\ (d^{n-1})^T\end{pmatrix}Q(d^0,\cdots,d^{n-1})=\left((d^i)^TQd^j\right)_{n\times n}$ 是对角阵

2. $I=S^{-1}(d^0,\cdots,d^{n-1})=(S^{-1}d^0,\cdots,S^{-1}d^{n-1}),S^{-1}d^i=e_{i+1}\,,i=0,\cdots,n-1$

$$f(x)=\frac{1}{2}x^TQx+b^Tx$$

令$x=S\hat{x}$, 则$\mathrm{f}(\hat{x})=\frac{1}{2}\hat{x}^TS^TQS\hat{x}+(S^Tb)^T\hat{\mathrm{x}}$

$$\begin{aligned}\hat{x}^n&=S^{-1}x\\&=S^{-1}x^0+\alpha_0S^{-1}d^0+\alpha_1S^{-1}d^1+\cdots+\alpha_{n-1}S^{-1}d^{n-1}\\&=S^{-1}x^0+\alpha_0e_1+\alpha_1e_2+\cdots+\alpha_ne_n\end{aligned}$$

共轭方向法的重要特征

➢考虑问题$\min f(x)=\frac{1}{2}x^TQx+b^Tx, Q\succ 0$

给定初始点x^0及一组关于Q共轭方向$d^0,d^1,\cdots,d^{n-1}$, 令

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

其中$\alpha_k=-\frac{\nabla f\left(x^k\right)^T d^k}{\left(d^k\right)^T Q d^k}, k=0,\cdots,n-1$

➢点列$\{x^k\}$具有如下特征：

1. $\nabla f\left(x^k\right)^T d^i=0, i=0,\cdots,k-1;$

2. $\mathrm{x}^{\mathrm{k}}=\arg\min\{\frac{1}{2}x^TQx+b^Tx|x\in X^K\}$，其中

$$X^k=\{x^0+\sum_{\mathrm{i}=0}^{k-1}\alpha_i d^i|\alpha_i\in R, i=1,\cdots,k-1\}$$

当$k=n$时， $\mathrm{x}^{\mathrm{n}}=\arg\min\{\frac{1}{2}x^TQx+b^Tx|x\in X^n\}$, 此时$X^n=R^n$,
表明共轭方向法在n步之内找到最优解

证明参考《运筹学与最优化方法》P95,
《最优化方法》P333-335

共轭方向法的重要特征证明

- 1. $\nabla f(x^k)^T d^i = 0, i = 0, \cdots, k-1;$

证明：将要证明的式子分为两个： $\nabla f(x^k)^T d^{k-1} = 0$和
$\nabla f(x^k)^T d^i = 0, i = 0, \cdots, k-2$

记$\phi(\alpha) = f(x^{k-1} + \alpha d^{k-1})$, 通过解析法对其求解则有$\phi'(\alpha) = 0$,
即$\nabla f(x^{k-1} + \alpha d^{k-1}) d^{k-1} = 0$ 这也就是$\nabla f(x^k)^T d^{k-1} = 0$

对于$\nabla f(x^k)^T d^i = 0, i = 0, \cdots, k-2$

$$
\begin{aligned}
\nabla f(x^k)^T d^i &= (Qx^k + b)^T d^i \\
&= \left(Q\left(x^{i+1} + \alpha_{i+1} d^{i+1} + \cdots + \alpha_{k-1} d^{k-1}\right) + b\right)^T d^i \\
&= (Qx^{i+1} + b)^T d^i \\
&= \nabla f(x^{i+1})^T d^i \\
&= 0
\end{aligned}
$$

共轭方向法的重要特征证明

2. $\mathrm{x}^{\mathrm{k}} = \arg\min\{\frac{1}{2}x^TQx + b^Tx | x \in X^K\}$，其中

$$X^k = \{x^0 + \sum_{\mathrm{i}=0}^{k-1} \alpha_i d^i | \alpha_i \in R, i = 1, \cdots, k-1\}$$

当$k = n$时， $\mathrm{x}^{\mathrm{n}} = \arg\min\{\frac{1}{2}x^TQx + b^Tx | x \in X^n\}$, 此时$X^n = R^n$

证明：记$\phi(\alpha_0, \cdots, \alpha_{k-1}) = \frac{1}{2}\left(x^0 + \sum_{i=0}^{k-1} \alpha_i d^i\right)^T Q\left(x^0 + \sum_{i=0}^{k-1} \alpha_i d^i\right) + b^T\left(x^0 + \sum_{i=0}^{k-1} \alpha_i d^i\right)$

则原问题转化为$(\alpha_0, \cdots, \alpha_{k-1}) = \arg\min \phi(\alpha_0, \cdots, \alpha_{k-1})$

$$\frac{\partial \phi(\alpha_0, \cdots, \alpha_{k-1})}{\partial \alpha_i}\bigg|_{(\alpha_0, \cdots, \alpha_{k-1})} = \left(Q\left(x^0 + \sum_{i=0}^{k-1} \alpha_i d^i\right) + b\right)^T d^i$$

$$= \left(Qx^k + b\right)^T d^i = \nabla f\left(x^k\right)^T d^i = 0, i = 0, \cdots, k-1$$

共轭梯度法

- 在迭代下降过程中，借助当前点x^k的梯度信息构造共轭方向；
- 基本框架：

Step 0：给定初始点x^0, 记$d^0=-\nabla f(x^0), \epsilon>0, k=0$;

Step 1：判断$\|\nabla f(x^k)\|\le\epsilon$是否成立；是，则终止，否则转Step 2;

Step 2:计算步长$\alpha_k=-\frac{\nabla f(x^k)^T d^k}{(d^k)^T Q d^k}$;

Step 3:令$x^{k+1}=x^k+\alpha_k d^k$, 并计算方向

$$d^{k+1}=-\nabla f\left(x^{k+1}\right)+\beta_k d^k, \qquad \beta_k=\frac{\nabla f\left(x^{k+1}\right)Q d^k}{(d^k)^T Q d^k}$$

令$k=k+1$,转Step 1.

$\beta_k d^k$的由来

- 设计$d^{k+1}=-\nabla f\left(x^{k+1}\right)+\beta_0 d^0+\cdots+\beta_k d^k$

为了使$(d^{k+1})^T Q d^i=0, i=0,\cdots,k$

即$\left(-\nabla f\left(x^{k+1}\right)+\beta_0 d^0+\cdots+\beta_k d^k\right)^T Q d^i=0$

$$\Leftrightarrow -\nabla f\left(x^{k+1}\right)^{\mathrm{T}} \mathrm{Qd}^{\mathrm{i}}+\beta_i\left(d^i\right)^T Q d^i=0$$

整理得到$\beta_{\mathrm{i}}=\dfrac{\nabla f\left(x^{k+1}\right)^T Q d^i}{\left(d^i\right)^T Q d^i}, \mathrm{i}=0,\cdots,k$

$$由\alpha_i d^i=\left(x^{i+1}-x^i\right)得到d^i=\frac{x^{i+1}-x^i}{\alpha_i}$$

两边同左乘Q得到$Q d^i=\dfrac{Q x^{i+1}-Q x^i}{\alpha_i}=\dfrac{\nabla f\left(x^{i+1}\right)-\nabla f\left(x^i\right)}{\alpha_i}$

当$i=0,\cdots,k-1, \nabla f\left(x^{k+1}\right)^T Q d^i=\nabla f\left(x^{k+1}\right)^T \dfrac{\nabla f\left(x^{i+1}\right)-\nabla f\left(x^i\right)}{\alpha_i}$

$$\nabla f\left(x^{i+1}\right)=-d^{i+1}+\beta_0 d^0+\cdots+\beta_i d^i$$

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

所以， 构造方向$d^{k+1}=-\nabla f\left(x^{k+1}\right)+\beta_k d^k$

共轭梯度法

➢共轭梯度法的步长公式

$$\alpha_k = -\frac{\nabla f\left(x^k\right)^T d^k}{(d^k)^T Q d^k}$$

可简化为：

$$\alpha_k = \frac{\nabla f\left(x^k\right)^T \nabla f(x^k)}{(d^k)^T Q d^k}$$

为什么？

把$d^k = -\nabla f\left(x^k\right) + \beta_{k-1} d^{k-1}$带入分子$\nabla f\left(x^k\right)^T d^k$得到

$$\begin{aligned}\nabla f\left(x^k\right)^T d^k &= \nabla f\left(x^k\right)^T (-\nabla f\left(x^k\right) + \beta_{k-1} d^{k-1}) \\ &= -\ \nabla f\left(x^k\right)^T \nabla f\left(x^k\right) + \beta_{k-1} \nabla f\left(x^k\right)^T d^{k-1} \\ &= -\nabla f\left(x^k\right)^T \nabla f(x^k)\end{aligned}$$

共轭梯度法

➢共轭梯度法方向公式中的系数

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

可简化为

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

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

$$\nabla f\left(x^{k+1}\right)^T Q d^k = \frac{\nabla f\left(x^{k+1}\right)^T \left(\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right)\right)}{\alpha_k}$$

前面又有$\alpha_k = \frac{\nabla f\left(x^k\right)^T \nabla f\left(x^k\right)}{\left(d^k\right)^T Q d^k}$ 可得到简化式子，此外 $\nabla f\left(x^{k+1}\right)^T \nabla f\left(x^k\right) = 0$

非线性共轭梯度法（FR方法, PRP方法）

Step 0:给定初始点x^0, 记$d^0 = -\nabla f(x^0), \epsilon > 0, k = 0;$

Step 1:判断$\|\nabla f(x^k)\| \le \epsilon$是否成立；是就终止，否则转Step 2;

Step 2:利用线性搜索计算步长α_k;

Step 3:令$x^{k+1} = x^k + \alpha_k d^k$, 并计算方向

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

其中

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

或者

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

令$k = k + 1$, 转Step 1.

非线性共轭梯度法

- 在实践中，为保证每次产生的方向为下降方向，可能会对β_k进行调整
- 对正定二次函数具有二次终止性
- 实现过程中常采用n步重启策略，可达到n步二阶收敛

适合大规模运算

收敛速度介于最速下降法和牛顿法之间

参考文献

- 运筹学与最优化方法，第五章 Page 117-120
- 最优化方法，第4章
- 最优化基础理论与方法，第三章
- 最优化理论与方法　　网课第七讲　　上海财经大学