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

- 非精确的线搜索

- 变量轮换法

- 最速下降法

课程内容

- 牛顿法
- 修正的牛顿法
- 拟牛顿法

牛顿法

- 考虑无约束优化问题

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

$f(\boldsymbol{x})$在$\boldsymbol{x}^k$处具有二阶连续偏导数，且在$\boldsymbol{x}^k$处的海塞矩阵$\nabla^2 f(\boldsymbol{x}^k) \succ \mathbf{0}$， $\boldsymbol{x}^k$是$f(\boldsymbol{x})$的极小点的第k轮估计值

将$f(\boldsymbol{x})$在$\boldsymbol{x}^k$处作二阶泰勒展开：

$$\begin{aligned} f(\boldsymbol{x}) = f(\boldsymbol{x}^k) + \nabla f(\boldsymbol{x}^k)^T (\boldsymbol{x} - \boldsymbol{x}^k) \\ + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}^k)^T \nabla^2 f(\boldsymbol{x}^k)(\boldsymbol{x} - \boldsymbol{x}^k) \\ + O\left(\|\boldsymbol{x} - \boldsymbol{x}^k\|^2\right) \end{aligned}$$

记$Q(\boldsymbol{x}) = f(\boldsymbol{x}^k) + \nabla f(\boldsymbol{x}^k)^T (\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}^k)^T \nabla^2 f(\boldsymbol{x}^k)(\boldsymbol{x} - \boldsymbol{x}^k)$, 有$Q(\boldsymbol{x}) \approx f(\boldsymbol{x})$

令$\nabla Q(\boldsymbol{x}) = 0$, 即$\nabla f(\boldsymbol{x}^k) + \nabla^2 f(\boldsymbol{x}^k)(\boldsymbol{x} - \boldsymbol{x}^k) = 0$, 记$\boldsymbol{x}^{k+1}$为$Q(\boldsymbol{x})$的平稳点，则

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

搜索方向$d^k = -[\nabla^2 f(\boldsymbol{x}^k)]^{-1} \nabla f(\boldsymbol{x}^k)$即为牛顿方向

牛顿法

- $\boldsymbol{d}^k$的几何意义：

因为$Q(\boldsymbol{x})$是一个二次函数， 且有$\nabla Q(\boldsymbol{x}) = \nabla f(\boldsymbol{x}^k) + \nabla^2 f(\boldsymbol{x}^k)(\boldsymbol{x} - \boldsymbol{x}^k)$， 所以

$$\nabla^2 Q(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}^k) \succ \boldsymbol{0}$$

$Q(\boldsymbol{x})$是凸函数，其平稳点$\boldsymbol{x}^{k+1}$就是全局极小点

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

由牛顿方向$-[\nabla^2 f(\boldsymbol{x}^k)]^{-1} \nabla f(\boldsymbol{x}^k) = \boldsymbol{x}^{k+1} - \boldsymbol{x}^k$可见，牛顿方向实质上就是由$\boldsymbol{x}^k$指向$\boldsymbol{x}^{k+1}$的方向，也就是$f(\boldsymbol{x})$的第$k$轮极小点估计值指向近似二次函数$Q(\boldsymbol{x})$的极小点的方向

牛顿法(作业)

➢基本思想：对于当前点$\boldsymbol{x}^k$处选择牛顿方向$\boldsymbol{d}^k = -\left[\nabla^2 f(\boldsymbol{x}^k)\right]^{-1}\nabla f(\boldsymbol{x}^k)$

也可以理解为：对当前点$\boldsymbol{x}^k$处的二次逼近函数进行最小化

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

注意：纯牛顿法规定步长因子$\alpha_k = 1$

基本步骤：

Step 1:给定初始点$\boldsymbol{x}^0$，终止误差$\epsilon > 0$，k=0;

Step 2:计算梯度$\nabla f(\boldsymbol{x}^k)$，判断是否满足终止条件$\|\nabla f(\boldsymbol{x}^k)\| < \epsilon$,满足则终止；否则　转step 3;

Step 3:构造牛顿方向$\boldsymbol{d}^k = -\left[\nabla^2 f(\boldsymbol{x}^k)\right]^{-1}\nabla f(\boldsymbol{x}^k)$

Step 4:计算$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{d}^k$，令$k = k + 1$转step 2

牛顿法

- 问题：牛顿方向$\boldsymbol{d}^k = -\left[\nabla^2 f\left(\boldsymbol{x}^k\right)\right]^{-1}\nabla f\left(\boldsymbol{x}^k\right)$一定是下降方向么？

$$\nabla f\left(\boldsymbol{x}^k\right)^T \boldsymbol{d}^k = -\nabla f\left(\boldsymbol{x}^k\right)^T\left[\nabla^2 f\left(\boldsymbol{x}^k\right)\right]^{-1}\nabla f\left(\boldsymbol{x}^k\right) < 0?$$

如果$\left[\nabla^2 f\left(\boldsymbol{x}^k\right)\right]^{-1} \succ \boldsymbol{0}$，则牛顿方向是下降方向，否则不是

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

优点：

1. 当初始点接近于最优解且海塞矩阵正定时，牛顿法二阶收敛；
2. 对于凸二次函数，只需要一步牛顿法即可找到最优解

修正牛顿法

➢对于步长的修正：

保留牛顿方向为搜索方向但摒弃了步长恒为1的做法，而是采用线搜索确定最优步长来构造算法

➢基本步骤：

Step 1:给定初始点$\boldsymbol{x}^0$，终止误差$\epsilon > 0$，k=0；

Step 2:计算梯度$\nabla f\left(\boldsymbol{x}^k\right)$，判断是否满足终止条件$\left\|\nabla f(\boldsymbol{x}^k)\right\| < \epsilon$，满足则终止；否则 转step 3；

Step 3:构造牛顿方向$\boldsymbol{d}^k = -\left[\nabla^2 f\left(\boldsymbol{x}^k\right)\right]^{-1}\nabla f\left(\boldsymbol{x}^k\right)$；

Step 4:进行线搜索求α_k使

$$f\left(\boldsymbol{x}^k + \alpha_k \boldsymbol{d}^k\right) = \min_{\alpha \geq 0} f\left(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k\right)$$

计算$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{d}^k$，令$k = k + 1$转step 2

修正牛顿法

- 对于方向的修正：选取$\boldsymbol{d}^k = -B_k^{-1}\nabla f\left(\boldsymbol{x}^k\right)$

若$\nabla^2 f\left(\boldsymbol{x}^k\right) \succ \boldsymbol{0}$，则取$\boldsymbol{B}_k = \nabla^2 f(\boldsymbol{x}^k)$；

否则, 采取修正方法：

1. $\boldsymbol{B}_k = \nabla^2 f(\boldsymbol{x}^k) + \lambda \boldsymbol{I}$, λ为适当的正数保证$\boldsymbol{B}_k$正定；

$\nabla^2 f(\boldsymbol{x}^k)$的特征值$\lambda_1, \lambda_2, \cdots, \lambda_n$，要求$\lambda_i + \lambda > 0, i = 1,2,\cdots,n$

只需要取$\lambda > \max\limits_{i=1,2,\cdots,n} \{-\lambda_i\}$

2. 考虑特征值分解$\nabla^2 f\left(\boldsymbol{x}^k\right) = \boldsymbol{Q}^T \Lambda \mathbf{Q}$，其中$\Lambda = diag\,(\lambda_1, \cdots, \lambda_n)$;

令$\boldsymbol{B}_k = \boldsymbol{Q}^T diag\,(\tau_i) \boldsymbol{Q}$,

$$\tau_i = \begin{cases} \lambda_i, if\ \lambda_i \geq \delta \\ \delta, otherwise \end{cases}, \quad \delta\text{为一适当正数}$$

拟牛顿法

- 考虑$f(x)$在当前点x^k处的二次近似函数

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

其中$B_k \succ 0$

利用$\min Q_k(x)$ 得到搜索方向$d^k = -B_k^{-1}\nabla f(x^k)$

拟牛顿法思想：使B_k体现一些二阶信息并且容易获取

拟牛顿法框架：

Step 1:给定初始点x^0，终止误差$\epsilon > 0$，k=0以及$B_0 \succ 0$;

Step 2:计算梯度$\nabla f(x^k)$，判断是否满足终止条件$\|\nabla f(x^k)\| < \epsilon$,满足则终止；否则 转step 3;

Step 3:构造拟牛顿方向$d^k = -B_k^{-1}\nabla f(x^k)$;

Step 4:进行线搜索求α_k使

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

计算$x^{k+1} = x^k + \alpha_k d^k$, 确定$B_{k+1}$ 令$k = k + 1$转step 2

拟牛顿法

- 如何简便获取矩阵$\boldsymbol{B}_{k+1}$?
- 拟牛顿方程(基本要求):

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

如何理解?

已知$\boldsymbol{x}^k$由迭代法算出$\boldsymbol{x}^{k+1}$，然后计算出$\nabla f\left(\boldsymbol{x}^{k+1}\right)$和$\nabla f\left(\boldsymbol{x}^{k}\right)$,

由微分中值定理得到:

$$\nabla f\left(\boldsymbol{x}^{k+1}\right)-\nabla f\left(\boldsymbol{x}^{k}\right)=\nabla^2 f(\xi)\left(\boldsymbol{x}^{k+1}-\boldsymbol{x}^{k}\right)$$

其中$\xi=\lambda\boldsymbol{x}^k+(1-\lambda)\boldsymbol{x}^{k+1}, \lambda\in(0,1)$

为方便书写，记$\boldsymbol{y}_k=\nabla f\left(\boldsymbol{x}^{k+1}\right)-\nabla f\left(\boldsymbol{x}^{k}\right), \boldsymbol{s}_k=\boldsymbol{x}^{k+1}-\boldsymbol{x}^{k}$,

则牛顿方程变为

$$\boldsymbol{y}_k=\boldsymbol{B}_{k+1}\boldsymbol{s}_k$$

如果再记$H_k=B_k^{-1}$, 拟牛顿方程就表示为：$\boldsymbol{s}_k=\boldsymbol{H}_{k+1}\boldsymbol{y}_k$

拟牛顿

- 求$\boldsymbol{B}_{k+1}$ 或$\boldsymbol{H}_{k+1}$:

基于已有信息$(\boldsymbol{y}_k, \boldsymbol{s}_k, \boldsymbol{B}_k)$获取$\boldsymbol{B}_{k+1}$

基于已有信息$(\boldsymbol{y}_k, \boldsymbol{s}_k, \boldsymbol{H}_k)$获取$\boldsymbol{H}_{k+1}$

- 第一类方法：选择满足拟牛顿方程且与$\boldsymbol{B}_k$(或$\mathbf{H}_k$)近似的矩阵：

$$\begin{aligned} &\min \|\boldsymbol{B}-\boldsymbol{B}_k\| \\ &s.t.\ \boldsymbol{B}\boldsymbol{s}_k=\boldsymbol{y}_k \\ &\quad\ \boldsymbol{B}=\boldsymbol{B}^T \end{aligned}$$

或

$$\begin{aligned} &\min \|\boldsymbol{H}-\boldsymbol{H}_k\| \\ &s.t.\ \boldsymbol{H}\boldsymbol{y}_k=\boldsymbol{s}_k \\ &\quad\ \boldsymbol{H}=\boldsymbol{H}^T \end{aligned}$$

- 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；

DFP(Davidon, Fletcher, Powell)方法

➢DFP方法是对$\boldsymbol{H}_k$进行rank-2校正

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

如何得到?

DFP假设$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\boldsymbol{E}_k$，令$\boldsymbol{E}_k=a\boldsymbol{u}_k\boldsymbol{u}_k^T+b\boldsymbol{v}_k\boldsymbol{v}_k^T$，其中$\boldsymbol{u}_k,\boldsymbol{v}_k$都是$n\times 1$的向量，将$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\boldsymbol{E}_k$带入拟牛顿方程$\boldsymbol{s}_k=\boldsymbol{H}_{k+1}\boldsymbol{y}_k$得到

$$\boldsymbol{s}_k=(\boldsymbol{H}_k+\boldsymbol{E}_k)\boldsymbol{y}_k$$

再将$E_k=a\boldsymbol{u}_k\boldsymbol{u}_k^T+b\boldsymbol{v}_k\boldsymbol{v}_k^T$带入上式得到

$$\boldsymbol{s}_k=(\boldsymbol{H}_k+a\boldsymbol{u}_k\boldsymbol{u}_k^T+b\boldsymbol{v}_k\boldsymbol{v}_k^T)\boldsymbol{y}_k$$

$$\Rightarrow a\left(\boldsymbol{u}_k^T\boldsymbol{y}_k\right)\boldsymbol{u}_k+b\left(\boldsymbol{v}_k^T\boldsymbol{y}_k\right)\boldsymbol{v}_k=\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k \qquad (*)$$

已知$\boldsymbol{u}_k^T\boldsymbol{y}_k,\boldsymbol{v}_k^T\boldsymbol{y}_k$是实数， $\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k$为$n\times 1$的向量。

假设$\boldsymbol{u}_k=r\boldsymbol{H}_k\boldsymbol{y}_k,\boldsymbol{v}_k=\theta\boldsymbol{s}_k$，则$\boldsymbol{E}_k=ar^2\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k^T+b\theta^2\boldsymbol{s}_k\boldsymbol{s}_k^T$

将$\boldsymbol{u}_k=r\boldsymbol{H}_k\boldsymbol{y}_k,\boldsymbol{v}_k=\theta\boldsymbol{s}_k$带入上式(*)得到

DFP方法

$$\Rightarrow a[(r\boldsymbol{H}_k\boldsymbol{y}_k)^T\boldsymbol{y}_k](r\boldsymbol{H}_k\boldsymbol{y}_k)+b[(\theta\boldsymbol{s}_k)^T\boldsymbol{y}_k](\theta\boldsymbol{s}_k)=\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k$$

$$\Rightarrow\left[ar^2\boldsymbol{y}_k^T\boldsymbol{H}_k^T\boldsymbol{y}_k+1\right](\boldsymbol{H}_k\boldsymbol{y}_k)+\left[b\theta^2\boldsymbol{s}_k^T\boldsymbol{y}_k-1\right](\boldsymbol{s}_k)=0$$

令$ar^2\boldsymbol{y}_k^T\boldsymbol{H}_k^T\boldsymbol{y}_k+1=0, b\theta^2\boldsymbol{s}_k^T\boldsymbol{y}_k-1=0$, 则

$$\mathrm{ar}^2=-\frac{1}{\boldsymbol{y}_k^T\boldsymbol{H}_k^T\boldsymbol{y}_k}$$

$$b\theta^2=\frac{1}{\boldsymbol{s}_k^T\boldsymbol{y}_k}$$

带入$\boldsymbol{E}_k=ar^2\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k^T+b\theta^2\boldsymbol{s}_k\boldsymbol{s}_k^T$

于是最终的DFP公式为

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k-\frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k^T}{\boldsymbol{y}_k^T\boldsymbol{H}_k\boldsymbol{y}_k}+\frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\boldsymbol{s}_k^T\boldsymbol{y}_k}$$

DFP方法

➢DFP算法步骤：

Step 0:给定步长$\alpha \in (0,1), \sigma \in (0,0.5)$，初始点$\boldsymbol{x}_0 \in R^n$, 终止误差$\epsilon$, 初始对称正定矩阵$H_0$(通常取为$G(x_0)^{-1}$或单位矩阵$I_n$), 令$k = 0$;

Step 1:计算$g_k = \nabla f(x_k)$，若$\|g_k\| \le \epsilon$, 停止计算，输出x_k作为近似极小点;

Step 2: 计算搜索方向:

$$d_k = -H_k g_k;$$

Step 3: 设m_k是满足下列不等式的最小非负整数m:

$$f(x_k + \delta^m d_k) \le f(x_k) + \sigma \delta^m g_k^T d_k$$

令$\alpha_k = \delta^{m_k}, x_{k+1} = x_k + \alpha_k d_k$

Step 4: 由公式$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k - \frac{\boldsymbol{H}_k \boldsymbol{y}_k \boldsymbol{y}_k^T \boldsymbol{H}_k^T}{\boldsymbol{y}_k^T \boldsymbol{H}_k \boldsymbol{y}_k} + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^T}{\boldsymbol{s}_k^T \boldsymbol{y}_k}$确定$\boldsymbol{H}_{k+1}$;

Step 5:令$k = k + 1$, 转入step 1

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

BFGS (Broyden, Fletcher, Goldarb, Shanno) 算法

➢BFGS可看做对$\boldsymbol{B}_k$进行rank-2校正

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

如何得到？

过程和DFS相同，DFS是依据拟牛顿方程$\boldsymbol{s}_k = \boldsymbol{H}_{k+1}\boldsymbol{y}_k$，BFGS则依据拟牛顿方程$\boldsymbol{y}_k = \boldsymbol{B}_{k+1}\boldsymbol{s}_k$。所以只需要互换$s_k$和$y_k$位置

拟牛顿方向需要计算$\boldsymbol{B}_{k+1}^{-1}$，可利用Sherman-Morrison公式显示写出

Sherman-Morrison公式：设$\boldsymbol{A} \in R^n$为非奇异方阵，$\boldsymbol{u}, \boldsymbol{v} \in R^n$，若$1 + \boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{u} \neq 0$，则有

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u}}$$

由Sherman-Morrison公式得到

$$\boldsymbol{B}_{k+1}^{-1} = \left(\boldsymbol{I} - \frac{\boldsymbol{s}_k \boldsymbol{y}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k}\right) \boldsymbol{B}_k^{-1} \left(\boldsymbol{I} - \frac{\boldsymbol{y}_k \boldsymbol{s}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k}\right) + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k}$$

被认为是最有效的拟牛顿法，超线性收敛

拟牛顿法SR1方法

➢SR1方法是对B_k进行rank-1校正

$$B_{k+1}=B_k+\frac{(y_k-B_ks_k)(y_k-B_ks_k)^T}{(y_k-B_ks_k)^Ts_k} \qquad \text{(SR1)}$$

➢如何得到上式？

设$B_{k+1}=B_k+auu^T$需要满足拟牛顿公式$B_{k+1}s_k=y_k$

得到$B_ks_k+auu^Ts_k=y_k$

移项得到 $au(u^Ts_k)=y_k-B_ks_k$

令$u=y_k-B_ks_k$, 得到

$$\mathrm{a}\left(\mathrm{u}^{\mathrm{T}}\mathrm{s}_{\mathrm{k}}\right)=1$$

即$a=\frac{1}{\mathrm{u}^{\mathrm{T}}\mathrm{s}_{\mathrm{k}}}=\frac{1}{(y_k-B_ks_k)^Ts_k}$

将$u=y_k-B_ks_k$和$a=\frac{1}{(y_k-B_ks_k)^Ts_k}$代入$B_{k+1}=B_k+auu^T$，就得到公式SR1

➢SR1算法迭代公式简单，但不能保证正定性；适当条件下能达到n步超线性收敛

$$\lim_{k\to\infty}\frac{\|x^{k+1+n}-x^*\|}{\|x^k-x^*\|}=0$$

参考文献

- 运筹学与最优化方法，第5章 Page114-126
- 最优化方法，第4章
- 最优化基础理论与方法，第三章