最优化理论与方法

第12讲

拟牛顿法

第4章 无约束问题算法(II)——拟Newton法(变尺度法)

第一节 拟Newton法及其性质

1、拟Newton方程与Dennis-Moré 条件
2、对称秩1(SR1)修正公式
3、BFGS修正公式与BFGS算法
4、Broyden族算法及其性质

1、拟Newton法的思想

考虑无约束问题：

$$\min f(x), \quad x \in R^n \tag{4.1}$$

拟Newton法是对Newton法的一种改善

保持其优点：快速收敛性；

克服其缺陷：需计算Hessian矩阵且要求其正定

方　　　　法：构造对称正定矩阵或H_k满足

$$B_k \approx \nabla^2 f(x_k) \text{ 或 } H_k \approx \nabla^2 f(x_k)^{-1}$$

计算搜索方向：$d_k = -B_k^{-1}\nabla f(x_k)$　(拟Newton方向)

或：$d_k = -H_k \nabla f(x_k)$　(拟Newton方向)

构造B_k的要求如下：

(1) $\underline{B_k \approx \nabla^2 f(x_k)}$ 使得产生的拟Newton方向近似Newton方向；

(2) $\underline{\forall k \geq 0, B_k\text{对称正定}}$, 使得产生的拟Newton方向为下降方向；

(3) $\underline{B_k\text{易于计算}}$.

2、拟Newton方程与Dennis-Moré 条件

设函数$f(x)$二次连续可微, 利用Taylor展开式, 我们得到 $f(x)$ 在点x_{k+1}处的近似二次函数为

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

求梯度并令 $x = x_k$得:

$$\nabla f(x_k) \approx \nabla f(x_{k+1}) + \nabla^2 f(x_{k+1})(x_k - x_{k+1})$$

用B_{k+1}取代$\nabla^2 f(x_{k+1})$并取等号得:

$$\nabla f(x_k) - \nabla f(x_{k+1}) = B_{k+1}(x_k - x_{k+1})$$

记 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k), s_k = x_{k+1} - x_k$, 上式简写成:

$$B_{k+1} s_k = y_k \tag{4.2}$$

方程(4.2)称为拟Newton方程或割线方程(或拟Newton条件).

若令$H_{k+1}=B_{k+1}^{-1}$，则(4.2)等价地写成：

$$H_{k+1}y_k=s_k \tag{4.3}$$

为方便起见，我们称(4.3)为第二拟Newton方程，而称(4.2)为第一拟Newton方程

由于产生的矩阵B_k(或它的逆H_k)是Newton法中Hessian矩阵$\nabla^2 f(x_k)$(或它的逆$\nabla^2 f(x_k)^{-1}$)的近似，我们称之为拟Newton矩阵.

现在我们通过求解线性方程组：

$$B_k d+\nabla f(x_k)=0 \tag{4.4}$$

来计算搜索方向d_k，即

$$d_k=-B_k^{-1}\nabla f(x_k)=-H_k\nabla f(x_k)$$

显然该方向是Newton方向的近似，我们称之为拟Newton方向

拟Newton矩阵的这种近似能使拟Newton法保持较快的收敛速度，即超线性收敛速度.

定理4.1.1 设函数$f: R^n \to R$二次连续可微，考察如下迭代过程：

$$x_{k+1} = x_k + d_k, \quad k = 0, 1, 2, \cdots$$

其中d_k是线性方程组

$$B_k d + \nabla f(x_k) = 0 \tag{4.4}$$

的解，设$\{x_k\}$收敛于x^*且$\nabla f(x^*) = 0, \nabla^2 f(x^*)$正定. 则$\{x_k\}$超线性收敛当且仅当

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x^*))d_k\|}{\|d_k\|} = 0 \tag{4.5}$$

该定理是拟Newton法具有快速收敛性及良好数值效果的理论基础

定理的证明：

充分性：由于$\{x_k\}$收敛于x^*, 及$B_k d_k + \nabla f(x_k) = 0$，则

$$\| \nabla f(x_k) + \nabla^2 f(x_k) d_k \| = \| (B_k - \nabla^2 f(x_k)) d_k \|$$

$$\leq \| (B_k - \nabla^2 f(x^*)) d_k \| + \| (\nabla^2 f(x_k) d_k - \nabla^2 f(x^*)) d_k \|$$

$$= \| (B_k - \nabla^2 f(x^*)) d_k \| + o(\| d_k \|)$$

所以, 若(4.5)成立, 则得到

$$\lim_{k\to\infty} \frac{\| \nabla f(x_k) + \nabla^2 f(x_k) d_k \|}{\| d_k \|} \leq \lim_{k\to\infty} \frac{\| (B_k - \nabla^2 f(x^*)) d_k \|}{\| d_k \|} = 0$$

由定理2.5.2知序列$\{x_k\}$超线性收敛于 x^*.

必要性：若$\{x_k\}$超线性收敛于x^*，即

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=0 \qquad (4.6)$$

由三角不等式可推出

$$0\le\left|\frac{\|x_k-x^*\|}{\|x_k-x^*\|}-\frac{\|d_k\|}{\|x_k-x^*\|}\right|\le\frac{\|x_k+d_k-x^*\|}{\|x_k-x^*\|}=\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}$$

即可从(4.6)得：

$$\lim_{k\to\infty}\frac{\|x_k-x^*\|}{\|d_k\|}=1 \qquad (4.7)$$

从而由(4.4)可得

$$\|(B_k-\nabla^2 f(x^*))d_k\|=\|\nabla f(x_k)+\nabla^2 f(x^*)d_k\|$$

$$\le\|\nabla f(x_k)-\nabla f(x^*)-\nabla^2 f(x^*)(x_k-x^*)\|+\|\nabla^2 f(x^*)(x_{k+1}-x^*)\|$$

$$=\mathrm{o}(\|x_k-x^*\|)=\mathrm{o}(\|d_k\|)$$

即得（4.5）成立.

注意到$s_k = x_{k+1} - x_k$，(4.5)可以等价地写成：

$$\lim_{k\to\infty}\frac{\|(B_k - \nabla^2 f(x^*))s_k\|}{\|s_k\|} = 0$$

上式或(4.5)称为Dennis - Moré条件

在下面我们来介绍拟Newton矩阵B_k的计算：

给定初始对称正定矩阵：$B_0 \in R^{n\times n}$，然后由修正公式

$$B_{k+1} = B_k + \Delta_k,\quad k \geq 0, \qquad (4.8)$$

产生拟Newton矩阵序列$\{B_k\}$. 其中Δ_k为低秩对称矩阵，通常其秩为1或2.

注意：Δ_k的构造方式不是惟一的，不同的构造方式对应不同的拟Newton法

三类重要的修正公式：

对称秩1: $SR1$ 修正公式

对称秩2: BFGS 修正公式

DFP 修正公式

2、对称秩1（SR1）修正公式

若在(4.8)中选Δ_k的秩为1，则

$$\Delta_k = \beta_k u_k u_k^T$$

其中$\beta_k \in R, u_k \in R^n$，因而

$$B_{k+1} = B_k + \beta_k u_k u_k^T$$

将其代入第一拟Newton方程(4.2)并整理，得

$$\beta_k u_k u_k^T s_k = y_k - B_k s_k \qquad (4.9)$$

上式说明：$u_k \parallel y_k - B_k s_k$

不妨取$u_k = y_k - B_k s_k$，则 $\beta_k = \dfrac{1}{u_k^T s_k} = \dfrac{1}{(y_k - B_k s_k)^T s_k}$

由此即得对称秩1修正公式：

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k} \qquad (4.10)$$

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k} \qquad (4.10)$$

类似地，利用第二拟Newton方程(4.3)进行对称秩1修正得到：

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k} \qquad (4.11)$$

比较上述两式：

$$B_k \leftrightarrow H_k$$

$$s_k \leftrightarrow y_k$$

特点： ※ 无需搜索($\alpha_k = 1$)，具有二次终止性

※ 遗传性质：$B_k s_i = y_i, \forall i < k$；自对偶性质

缺点：正定性无保证，因$(y_k - B_k s_k)^T s_k > 0$不一定成立

2、 BFGS 修正公式与 BFGS 算法

若在(4.8)中选Δ_k的秩为2,则

$$\Delta_k = a_k u_k u_k^T + b_k v_k v_k^T$$

其中$a_k, b_k \in R, u_k, v_k \in R^n$,因而

$$B_{k+1} = B_k + a_k u_k u_k^T + b_k v_k v_k^T$$

将其代入第一拟Newton方程(4.2)并整理,得

$$a_k u_k u_k^T s_k + b_k v_k v_k^T s_k = y_k - B_k s_k \qquad (4.12)$$

显然,上式中,$a_k, b_k \in R, u_k, v_k \in R^n$的取法不唯一,

简单的取法是两边一对一

$$a_k u_k u_k^T s_k + b_k v_k v_k^T s_k = y_k - B_k s_k \tag{4.12}$$

令 $u_k = y_k$， 得 $a_k = \dfrac{1}{u_k^T s_k} = \dfrac{1}{y_k^T s_k}$;

令 $v_k = B_k s_k$,得 $b_k = -\dfrac{1}{v_k^T s_k} = -\dfrac{1}{s_k^T B_k s_k}$;

将它们代入Δ_k然后再代入(4.8), 得

BFGS修正公式：

$$\boxed{B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \qquad (4.13)}$$

该公式由：Broyden-Fletcher-Goldfarb-Shanno 提出

BFGS公式是迄今为止最好的拟Newton修正公式

特点： ※ 精确搜索下，具有二次终止性

※ 遗传性质： $B_k s_i = y_i, \forall i < k$(二次且精确)

※ 正定性保证： $y_k^T s_k > 0$, 较易成立

命题3.1.1 设B_k对称正定，B_{k+1}由BFGS修正公式(4.13)确定. 则当且仅当$y_k^T s_k > 0$时，B_{k+1}对称正定.

证明： 若B_{k+1}对称正定，则由第一拟Newton方程得到

$$y_k^T s_k = s_k^T B_{k+1} s_k$$

显然，我们有 $y_k^T s_k > 0$.

在下面我们来证明：

$$y_k^T s_k > 0 \text{且} B_k \text{正定} \Rightarrow d^T B_{k+1} d > 0, \forall 0 \neq d \in R^n$$

由BFGS修正公式得

$$d^T B_{k+1} d = d^T B_k d - \frac{d^T B_k s_k s_k^T B_k d}{s_k^T B_k s} + \frac{d^T y_k y_k^T d}{y_k^T s_k}$$

由于B_k正定有正定分解：$B_k = B_k^{1/2} B_k^{1/2}$，然后由Cauchy - Schwarz

不等式得

$$(d^T B_k s_k)^2 = (d^T B_k^{1/2} B_k^{1/2} s_k)^2 \leq \| B_k^{1/2} d \|^2 \| B_k^{1/2} s_k \|^2 = d^T B_k d \;\; s_k^T B_k s_k$$

当且仅当$B_k^{1/2} d \parallel B_k^{1/2} s_k$，即$B_k^{1/2} d = \lambda_k B_k^{1/2} s_k \; (\lambda_k \neq 0)$或$d = \lambda_k s_k$时上

式取等号.

所以，若$d = \lambda_k s_k$，则

$$d^T B_{k+1} d = d^T B_k d - d^T B_k d + \frac{d^T y_k y_k^T d}{y_k^T s_k} = \frac{(d^T y_k)^2}{y_k^T s_k} = \lambda_k^2 y_k^T s_k > 0$$

否则，$\quad d^T B_{k+1} d > d^T B_k d - d^T B_k d + \dfrac{d^T y_k y_k^T d}{y_k^T s_k} \geq 0$

综上分析，$d^T B_{k+1} d > 0, \forall 0 \neq d \in R^n$，即$B_{k+1}$正定.

算法4.1 (BFGS算法)

步1 给定初始点$x_0 \in R^n$,初始对称正定矩阵B_0,精度$\varepsilon > 0$.

令$k = 0$;

步2 若$\|\nabla f(x_k)\| \le \varepsilon$,则得解$x_k$,算法终止. 否则转下一步;

步3 解线性方程组

$$B_k d + \nabla f(x_k) = 0 \tag{4.16}$$

得解d_k;

步4 由线性搜索计算步长α_k;

步5 令 $x_{k+1} = x_k + \alpha_k d_k$, 若$\|\nabla f(x_{k+1})\| \le \varepsilon$, 则得解$x_{k+1}$,

算法 终止. 否则由公式(4.13) 计算B_{k+1};

步6 令$k := k + 1$, 转步3.

与Newton法比较, BFGS算法避免了计算二阶导数, 计算量大幅减少.

由方程(4.16)知，为确保BFGS算法每一步产生的搜索方向是下降的，要求矩阵序列$\{B_k\}$是正定的，进一步由命题3.1.1知，要求对所有 $k \geq 0$, $y_k^T s_k > 0$.

问题：如何保证 $y_k^T s_k > 0$ ？

分析：从函数 f 本身以及迭代过程来分析

由于

$$y_k^T s_k = [\nabla f(x_{k+1}) - \nabla f(x_k)]^{\mathrm{T}} s_k = \alpha_k [\nabla f(x_{k+1}) - \nabla f(x_k)]^{\mathrm{T}} d_k$$

(1) 由Lagrange中值定理可知

$$y_k^T s_k = [\nabla f(x_{k+1}) - \nabla f(x_k)]^{\mathrm{T}} s_k = s_k^T \nabla^2 f(x_k + \theta_k \alpha_k d_k) s_k$$

因此，当$\forall x, \nabla^2 f(x)$正定，即f具有某种凸性（如一致凸）时，$y_k^T s_k > 0$成立，

(2) 由Wolfe - Powell搜索的第二个条件

$$\nabla f(x_{k+1})^{\mathrm{T}} d_k \geq \sigma_2 \nabla f(x_k)^{\mathrm{T}} d_k$$

知

$$y_k^T s_k = \alpha_k [\nabla f(x_{k+1}) - \nabla f(x_k)]^{\mathrm{T}} d_k \geq \alpha_k (\sigma_2 - 1) \nabla f(x_k)^{\mathrm{T}} d_k$$

这里$\nabla f(x_k)^{\mathrm{T}} d_k < 0$. 因此，当采用Wolfe - Powell搜索时，$y_k^T s_k > 0$ 成立.

总结为如下命题：

命题4.1.2 设d_k满足$\nabla f(x_k)^{\mathrm{T}} d_k < 0$.若下面条件之一成立，则

$$y_k^T s_k > 0, \quad \forall k \geq 0$$

(1) 算法中采用Wolfe - Powell搜索或精确搜索；

(2) 函数 f 二次连续可微且$\forall x \in R^n, \nabla^2 f(x)$正定.

BFGS修正公式的改进：

由前面的分析知，当在BFGS算法中使用弱的线性搜索(如Armijo搜索)或目标函数f非凸时，拟Newton矩阵$\{B_k\}$的正定性是没有保证的，为此，人们提出一些改进形式：

第一类方法：在BFGS修正公式中使用某种开关条件.

如 Powell 提出：

$$B_{k+1}=\begin{cases} B_k-\dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}+\dfrac{y_k y_k^T}{y_k^T s_k}, & \text{如果 } \underline{y_k^T s_k>0} \\ B_k, & \text{否则} \end{cases}$$

如 LI - Fukushima 提出：

$$B_{k+1}=\begin{cases} B_k-\dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}+\dfrac{y_k y_k^T}{y_k^T s_k}, & \text{如果 } \underline{\dfrac{y_k^T s_k}{s_k^T s_k}\ge\varepsilon\|\nabla f(x_k)\|^{\beta}} \\ B_k, & \text{否则} \end{cases}$$

其它类改进方法：形式多样.

在BFGS公式中修改y_k确保$y_k^T s_k > 0$

(1) 如Li - Fukushima 提出：

$$y_k \to \hat{y}_k = \nabla f(x_{k+1}) - \nabla f(x_k) + t_k s_k, \qquad (t_k \geq 0)$$

这里参数t_k依据某种开关条件取值. 然后得到

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\hat{y}_k \hat{y}_k^T}{\hat{y}_k^T s_k}$$

(2) 如Cheng - Li提出：

$$y_k \to \hat{y}_k = \tau_k y_k, \qquad \tau_k = \frac{y_k^T s_k}{y_k^T y_k}$$

上面的公式实际修改了拟Newton方程为：

$$B_{k+1} s_k = \hat{y}_k$$

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{4.13}$$

关于BFGS修正公正公式的逆修正：

令$H_k = B_k^{-1}$及$H_{k+1} = B_{k+1}^{-1}$，利用Sherman - Morison公式(1.16)两次，可以得到BFGS公式的逆修正公式为：

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right) H_k \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right)^T + \frac{s_k s_k^T}{y_k^T s_k} \tag{4.18}$$

或简写成

$$H_{k+1} = V_k H_k V_k^T + \rho_k s_k s_k^T$$

这里 $\rho_k = \dfrac{1}{y_k^T s_k}, V_k = I - \rho_k s_k y_k^T$

大家可以自己去推导一下公式(4.18)，以一小时为限

例4.1.1 用精确搜索的BFGS算法求解下面的无约束问题：

$$\min\ f(x)=\frac{1}{2}x_1^2+x_2^2-x_1x_2-2x_1$$

其中初始点$x^{(0)}=(1,1)^T$，初始矩阵$B_0=I$.

解：$\nabla f(x)=\begin{pmatrix}x_1-x_2-2\\-x_1+2x_2\end{pmatrix}$，令$d=\begin{pmatrix}d_1\\d_2\end{pmatrix}$.由$\phi(\alpha)=f(x+\alpha d)$,

并令$\phi'(\alpha)=0$，即可得精确搜索的步长

$$\alpha=\frac{x_1d_1-x_2d_1-2d_1-x_1d_2+2x_2d_2}{(d_1-d_2)^2+d_2^2}$$

第一次迭代：

$$x^{(0)}=\begin{pmatrix}1\\1\end{pmatrix},\nabla f(x^{(0)})=\begin{pmatrix}-2\\1\end{pmatrix},d^{(0)}=-B_0^{-1}\nabla f(x^{(0)})=\begin{pmatrix}2\\-1\end{pmatrix},\alpha_0=\frac{1}{2}$$

所以 $$x^{(1)}=x^{(0)}+\alpha_0d^{(0)}=\begin{pmatrix}2\\1/2\end{pmatrix}$$

第二次迭代：先计算B_1

$$s_0 = x^{(1)} - x^{(0)} = \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}, y_0 = \nabla f(x^{(1)}) - \nabla f(x^{(0)}) = \begin{pmatrix} 3/2 \\ -2 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{4}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{pmatrix} + \begin{pmatrix} \frac{9}{10} & -\frac{6}{5} \\ -\frac{6}{5} & \frac{8}{5} \end{pmatrix} = \begin{pmatrix} \frac{11}{10} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{12}{5} \end{pmatrix}$$

所以 $d^{(1)} = -B_1^{-1} \nabla f(x^{(1)}) = \begin{pmatrix} \frac{9}{10} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{12}{5} \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}, \quad \alpha_1 = 2$

从而 $x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad \nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; x^* = x^{(2)}$

总结

拟Newton法具有超线性收敛性，其理论非常成熟. 拟Newton法在使用精确搜索或非精确的Armijo搜索及Wolfe-Powell搜索时不具有全局收敛性. 因而人们提出了有许多改正的方法. 其全局收敛性理论近年来也取得了重要的进展. 然而，拟Newton法具有非常好的数值效果. 被广泛用来求解无约束问题.

作　业

- **作业：** 习题4： 5(1), 6, 15

- **课后练习**：编写BFGS法的程序，并与Newton法和最速下降法进行比较
- **思考题:** 可否不要求目标函数凸性,获得BFGS的收敛性?Newton算法的收敛性结果？