4-4 拟牛顿法(2)

变度量意义下的最速下降法

牛顿法与拟牛顿法的最速下降对应度量：

- 牛顿法：$\boldsymbol{d}_k = -G_k^{-1}\boldsymbol{g}_k$ – G_k度量意义下的最速下降法

$$|\boldsymbol{g}_k^T\boldsymbol{d}_k| = |\boldsymbol{g}_k^T G_k^{-1} G_k \boldsymbol{d}_k| \leq \|G_k^{-1}\boldsymbol{g}_k\|_{G_k}\|\boldsymbol{d}_k\|_{G_k}$$

  其中$\boldsymbol{d}_k$与$-G_k^{-1}\boldsymbol{g}_k$共线时等式成立。

- 拟牛顿法：$\boldsymbol{d}_k = -H_k\boldsymbol{g}_k$ – H_k^{-1}度量意义下的最速下降法

$$|\boldsymbol{g}_k^T\boldsymbol{d}_k| = |\boldsymbol{g}_k^T H_k H_k^{-1} \boldsymbol{d}_k| \leq \|H_k\boldsymbol{g}_k\|_{H_k^{-1}}\|\boldsymbol{d}_k\|_{H_k^{-1}}$$

  其中$\boldsymbol{d}_k$与$-H_k\boldsymbol{g}_k$共线时等式成立。

以上两种方法每步迭代的度量矩阵随着迭代点变化而变化，因此称这两种方法为**变度量方法**。常规最速下降法是$\|\cdot\|_2$度量意义下的最速下降方法，非变度量方法。

定理4.4.1 对称秩1方法的二次终止性.

若对任意的初始点$\boldsymbol{x}_0 \in \mathbb{R}^n$和任意对称正定矩阵$B_0 \in \mathbb{R}^{n\times n}$，对称秩1公式有定义，$\boldsymbol{s}_0, \cdots, \boldsymbol{s}_{n-1}$线性无关，其中$\boldsymbol{s}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i$，则至多经过$n+1$次迭代，可求得正定二次函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}$的极小点，其中$G = B_n$。

证明： 由于$f(\boldsymbol{x})$为正定二次函数，总有如下关系式成立：

$$\boldsymbol{y}_k = G\boldsymbol{s}_k, k = 0, 1, \ldots. \qquad \text{(a)}$$

用数学归纳法证明：对$k \geq 1$，有：

$$\boldsymbol{y}_j = B_k\boldsymbol{s}_j, j = 0, \cdots, k-1 \qquad \text{(b)}$$

当$k = 1$时，由拟牛顿方程$\boldsymbol{y}_0 = B_1\boldsymbol{s}_0$，知(b)式成立。假定对$k(> 1)$，(b)式成立，下证对$k+1$，(b)式依然成立。

对称秩1方法的性质

证明续： 对于$j < k$，有：

$$B_{k+1}\boldsymbol{s}_j = B_k\boldsymbol{s}_j + \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T\boldsymbol{s}_j}{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T\boldsymbol{s}_k}$$

由(a)与归纳假设有：

$$(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T\boldsymbol{s}_j = \boldsymbol{y}_k^T\boldsymbol{s}_j - \boldsymbol{s}_k^T B_k\boldsymbol{s}_j = \boldsymbol{s}_k^T G\boldsymbol{s}_j - \boldsymbol{s}_k^T\boldsymbol{y}_j = 0$$

则：

$$B_{k+1}\boldsymbol{s}_j = B_k\boldsymbol{s}_j = \boldsymbol{y}_j, j = 0, \ldots, k-1.$$

结合拟牛顿公式$B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k$，所以(b)式对$k+1$亦成立。由数学归纳法，(b)式对一切$k$都成立。

对称秩1方法的性质

证明续： 在(b)式中，令$k=n$, 得：

$$B_n \boldsymbol{s}_j = \boldsymbol{y}_j = G\boldsymbol{s}_j \Rightarrow (B_n - G)\boldsymbol{s}_j = 0, j = 0, \ldots, n-1.$$

由于$\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n-1}$线性无关，则必有$B_n = G$。所以，第$n+1$步迭代是Newton迭代，此时若前面迭代尚未终止，则该次迭代必然达到极小点。 ■

若定理中$\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n-1}$线性相关，易推导知$\boldsymbol{g}_{k+1} = 0$, 迭代停止。

定理4.4.2 DFP校正正定性.

设H_k对称正定，且$\boldsymbol{s}_k^T\boldsymbol{y}_k>0$, 则DFP校正(4.4.17)可构造出$H_{k+1}^{\mathrm{DFP}}$对称正定。

证明： 因为H_k对称正定，对于$\boldsymbol{x}\in\mathbb{R}^n\backslash\{\boldsymbol{0}\}$, 有$\boldsymbol{x}^TH_k\boldsymbol{x}>0$, 由DFP校正公式(4.4.17)与$\boldsymbol{s}_k^T\boldsymbol{y}_k>0$, 则$H_{k+1}^{\mathrm{DFP}}$存在有意义。

下证H_{k+1}^{DFP}正定性。给定$\boldsymbol{x}\in\mathbb{R}^n\backslash\{\boldsymbol{0}\}$, 有：

$$\begin{aligned}\boldsymbol{x}^TH_{k+1}^{\mathrm{DFP}}\boldsymbol{x}&=\boldsymbol{x}^TH_k\boldsymbol{x}+\frac{(\boldsymbol{x}^T\boldsymbol{s}_k)^2}{\boldsymbol{s}_k^T\boldsymbol{y}_k}-\frac{(\boldsymbol{x}^TH_k\boldsymbol{y}_k)^2}{\boldsymbol{y}_k^TH_k\boldsymbol{y}_k}\\&=\frac{(\boldsymbol{x}^TH_k\boldsymbol{x})(\boldsymbol{y}_k^TH_k\boldsymbol{y}_k)-(\boldsymbol{x}^TH_k\boldsymbol{y}_k)^2}{\boldsymbol{y}_k^TH_k\boldsymbol{y}_k}+\frac{(\boldsymbol{x}^T\boldsymbol{s}_k)^2}{\boldsymbol{s}_k^T\boldsymbol{y}_k}\qquad\text{(a)}\end{aligned}$$

DFP校正的正定性

证明续： 在H_k度量意义下，由Cauchy-Schwarz不等式：

$$(\boldsymbol{x}^T H_k \boldsymbol{y}_k)^2 \leq (\boldsymbol{x}^T H_k \boldsymbol{x})(\boldsymbol{y}_k^T H_k \boldsymbol{y}_k)$$

上式等式成立当且仅当$\boldsymbol{x} = \gamma \boldsymbol{y}_k (\gamma \in \mathbb{R}\backslash\{0\})$。

情形1： 当$\boldsymbol{x} \neq \gamma \boldsymbol{y}_k$时，由(a)式有：

$$\boldsymbol{x}^T H_{k+1}^{\mathrm{DFP}} \boldsymbol{x} > \frac{(\boldsymbol{x}^T \boldsymbol{s}_k)^2}{\boldsymbol{s}_k^T \boldsymbol{y}_k} \geq 0.$$

情形2： 当$\boldsymbol{x} = \gamma \boldsymbol{y}_k$时，由(a)式有：

$$\boldsymbol{x}^T H_{k+1}^{\mathrm{DFP}} \boldsymbol{x} = \frac{(\boldsymbol{x}^T \boldsymbol{s}_k)^2}{\boldsymbol{s}_k^T \boldsymbol{y}_k} = \gamma^2 \boldsymbol{s}_k^T \boldsymbol{y} > 0$$

H_{k+1}^{DFP}正定, 其对称性显然成立。 ■

校正矩阵正定性条件保证

由上述定理可知，$\boldsymbol{s}_k^T\boldsymbol{y}_k > 0$对于保持DFP和BFGS校正的正定性是关键的; 同时，这个条件也是实际的，并且是可以满足的。

情形1： 正定二次函数:

$$\boldsymbol{s}_k^T\boldsymbol{y}_k = \boldsymbol{s}_k^T G\boldsymbol{s}_k > 0.$$

情形2：对于一般函数: (采用精确线性搜索)

$$\boldsymbol{s}_k^T\boldsymbol{y}_k = \boldsymbol{g}_{k+1}^T\boldsymbol{s}_k - \boldsymbol{g}_k^T\boldsymbol{s}_k > 0.$$

其中: $\boldsymbol{s}_k$是下降步($\boldsymbol{g}_k^T\boldsymbol{s}_k < 0$）和($\boldsymbol{g}_{k+1}^T\boldsymbol{s}_k = 0$)

校正矩阵正定性条件保证

定理4.4.3 线性搜索正定性条件保证.

采用精确线性搜索或者Wolfe非精确线性搜索准则的DFP方法或BFGS方法，有$\boldsymbol{s}_k^T\boldsymbol{y}_k > 0$成立。

证明：(1) 先考虑精确线性搜索，令α_k是对$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)$的最优步长因子，即有$\boldsymbol{g}_{k+1}^T\boldsymbol{d}_k=0$。另外，

$$\boldsymbol{s}_k=\alpha_k\boldsymbol{d}_k=-\alpha_k H_k\boldsymbol{g}_k. \tag{a}$$

结合H_k的正定性，于是有：

$$\begin{aligned}\boldsymbol{s}_k^T\boldsymbol{y}_k &= \boldsymbol{s}_k^T(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)\\ &= \alpha_k\boldsymbol{d}_k^T\boldsymbol{g}_{k+1}+\alpha_k\boldsymbol{g}_k^T H_k\boldsymbol{g}_k > 0.\end{aligned}$$

校正矩阵正定性条件保证

证明续：(2) 考虑Wolfe非精确线性搜索，有：$\boldsymbol{g}_{k+1}^T\boldsymbol{s}_k \geq \sigma \boldsymbol{g}_k^T\boldsymbol{s}_k, (0 < \sigma < 1)$. 另外，注意到$\boldsymbol{d}_k$是下降方向($\boldsymbol{g}_k^T\boldsymbol{d}_k < 0$). 结合(a)式，有：

$$\begin{aligned}
\boldsymbol{s}_k^T\boldsymbol{y}_k &= \alpha_k\boldsymbol{d}_k^T(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k) \\
&= \alpha_k(\boldsymbol{d}_k^T\boldsymbol{g}_{k+1} - \boldsymbol{d}_k^T\boldsymbol{g}_k) \\
&\geq \alpha_k(\sigma\boldsymbol{d}_k^T\boldsymbol{g}_k - \boldsymbol{d}_k^T\boldsymbol{g}_k) \\
&= -\alpha_k(1-\sigma)\boldsymbol{d}_k^T\boldsymbol{g}_k > 0.
\end{aligned}$$

■

以上定理结论对于强Wolfe准则亦成立，但Goldstein准则无法保证拟牛顿法中正定性条件$\boldsymbol{s}_k\boldsymbol{y}_k^T > 0$。

定理4.4.4 Broyden族公式正定性.

设H_{k+1}^{DFP}是对称正定矩阵，对于$\varphi \geq 0$，由(4.4.21)式得到的Broyden族公式H_{k+1}^{φ}亦对称正定。

对Broyden族公式而言，

- 所有矩阵H_{k+1}^{φ}与矩阵H_{k+1}^{DFP}仅差一个秩1矩阵$\varphi \boldsymbol{v}_k^T \boldsymbol{v}_k (\varphi \geq 0)$;
- 由H_{k+1}^{DFP}的正定性和$\varphi \boldsymbol{v}_k^T \boldsymbol{v}_k (\varphi \geq 0)$的半正定性可以得到$H_{k+1}^{\varphi}$的正定性。

DFP公式意义

引理4.4.5.

设对称矩阵$B_k \in \mathbb{R}^{n\times n}$, $\boldsymbol{s}_k, \boldsymbol{y}_k \in \mathbb{R}^n$, 且$\boldsymbol{s}_k \neq 0$, 定义矩阵序列$\{B_{k+1}^{(j)}\}$:

$$\begin{aligned}
B_{k+1}^{(0)} &= B_k \\
B_{k+1}^{(2j+1)} &= B_{k+1}^{(2j)} + \frac{\left(\boldsymbol{y}_k - B_{k+1}^{(2j)}\boldsymbol{s}_k\right)\boldsymbol{s}_k^T}{\boldsymbol{s}_k^T\boldsymbol{s}_k} \\
B_{k+1}^{(2j+2)} &= \frac{1}{2}\left(B_{k+1}^{(2j+1)} + +B_{k+1}^{(2j+1)^T}\right)
\end{aligned}$$

则$\{B_{k+1}^{(2j+2)}\}$收敛于:

$$B_k + \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)\boldsymbol{s}_k^T + \boldsymbol{s}_k(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T}{\boldsymbol{s}_k^T\boldsymbol{s}_k} - \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T\boldsymbol{s}_k}{(\boldsymbol{s}_k^T\boldsymbol{s}_k)^2}\boldsymbol{s}_k\boldsymbol{s}_k^T.$$

DFP公式意义

PSB(Powell-symmetric-Broyden)公式：

$$B_{k+1}^{\mathrm{PSB}} = B_k + \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)\boldsymbol{s}_k^T + \boldsymbol{s}_k(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T}{\boldsymbol{s}_k^T\boldsymbol{s}_k} - \frac{(\boldsymbol{y}_k - B_k\boldsymbol{s}_k)^T\boldsymbol{s}_k}{(\boldsymbol{s}_k^T\boldsymbol{s}_k)^2}\boldsymbol{s}_k\boldsymbol{s}_k^T.$$

定理4.4.6.

设$B_k \in \mathbb{R}^{n\times n}$对称，$\boldsymbol{s}_k, \boldsymbol{y}_k \in \mathbb{R}^n$, 且$\boldsymbol{s}_k \neq 0$, 则：

$$\begin{array}{ll} \min & \|B - B_k\|_F \\ \text{s.t.} & B\boldsymbol{s}_k = \boldsymbol{y}_k, \quad B - B_k\text{对称} \end{array}$$

的唯一解由以上B_{k+1}^{PSB}公式给出。

若令$B = W^TW$，其中$W \in \mathbb{R}^{n\times n}$为非奇异矩阵, 考虑变换$\hat{\boldsymbol{x}} = W\boldsymbol{x}$, 则可以由关于$\hat{\boldsymbol{x}}$的PSB公式导出DFP公式。(详见教材P74)

定理4.4.7.

设$B_k \in \mathbb{R}^{n\times n}$对称，$\boldsymbol{s}_k, \boldsymbol{y}_k \in \mathbb{R}^n$，且$\boldsymbol{s}_k^T\boldsymbol{y}_k > 0$，则对任给$W \in \mathbb{R}^{n\times n}$，$W^TW = B$满足拟牛顿条件$B_k\boldsymbol{s}_k = \boldsymbol{y}_k$，问题：

$$\min_{B} \quad \|W^{-T}(B - B_k)W^{-1}\|_F \tag{4.4.22a}$$

$$\text{s.t.} \quad B = B^T,\ B\boldsymbol{s}_k = \boldsymbol{y}_k \tag{4.4.22b}$$

的唯一解B_{k+1}^{DFP}由DFP校正公式(4.4.17)给出。

在某种范数的意义上，B_{k+1}^{DFP}是满足拟牛顿条件的所有矩阵中最靠近当前矩阵B_k。类似地，可解释H_{k+1}^{BFGS}为最靠近H_k矩阵。