拟牛顿法

拟牛顿法就是找到一个迭代式子来代替H或H-1。根据泰勒二阶展开的求导，可以有以下等式，称为拟牛顿条件。

$$f(x) = f\left(x^{(k)}\right) + \nabla f\left(x^{(k)}\right)^T \left(x - x^{(k)}\right) + \frac{1}{2}\left(x - x^{(k)}\right)^T H\left(x^{(k)}\right)\left(x - x^{(k)}\right)$$

$$\nabla f(x) = \nabla f\left(x^{(k)}\right) + H_k\left(x - x^{(k)}\right) = 0$$

$$\Rightarrow x^{(k+1)} = x^{(k)} - H_k^{-1} \nabla f\left(x^{(k)}\right) \ (H_k\text{需为正定矩阵})$$

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

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

$$H_k \delta_k = y_k \ (\text{拟牛顿条件})$$

$$\Rightarrow \delta_k = H_k^{-1} y_k$$

也就是说这里没有直接令x点处的梯度为0，而是使用了迭代形式，为了进一步推出Hk的迭代近似。

记 $y_k = g_{k+1} - g_k$，$\delta_k = x^{(k+1)} - x^{(k)}$，则

$$y_k = H_k \delta_k \tag{B.12}$$

或

$$\boxed{H_k^{-1} y_k = \delta_k} \tag{B.13}$$

G_k

式 (B.12) 或式 (B.13) 称为拟牛顿条件。

如果 H_k 是正定的（H_k^{-1} 也是正定的），那么可以保证牛顿法搜索方向 p_k 是下降方向。这是因为搜索方向是 $p_k = -H_k^{-1} g_k$，由式 (B.8) 有

$$x = x^{(k)} + \lambda p_k = x^{(k)} - \lambda H_k^{-1} g_k \tag{B.14}$$

所以 $f(x)$ 在 $x^{(k)}$ 的泰勒展开式 (B.2) 可以近似写成：

$$f(x) = f(x^{(k)}) - \lambda g_k^{\mathrm{T}} H_k^{-1} g_k \tag{B.15}$$

>0 $g_k^{T} H_k^{-1} g_k$ $x^T x$ $-x^2$

因 H_k^{-1} 正定，故有 $g_k^{\mathrm{T}} H_k^{-1} g_k > 0$。当 λ 为一个充分小的正数时，总有 $f(x) < f(x^{(k)})$，也就是说 p_k 是下降方向。

拟牛顿法将 G_k 作为 H_k^{-1} 的近似，要求矩阵 G_k 满足同样的条件。首先，每次迭代矩阵 G_k 是正定的。同时，G_k 满足下面的拟牛顿条件：

$$G_{k+1} y_k = \delta_k \tag{B.16}$$

按照拟牛顿条件选择 G_k 作为 H_k^{-1} 的近似或选择 B_k 作为 H_k 的近似的算法称为拟牛顿法。

按照拟牛顿条件，在每次迭代中可以选择更新矩阵 G_{k+1}：

$$G_{k+1} = G_k + \Delta G_k \tag{B.17}$$

DFP算法

DFP是近似H-1计算的。使用类似待定系数法的方法将G拆成三部分，然后根据拟牛顿条件列出等式解方程。

3. DFP (Davidon-Fletcher-Powell) 算法 (DFP algorithm)

DFP 算法选择 G_{k+1} 的方法是，假设每一步迭代中矩阵 G_{k+1} 是由 G_k 加上两个附加项构成的，即

$$G_{k+1} = G_k + P_k + Q_k \tag{B.18}$$

其中 P_k，Q_k 是待定矩阵。这时，

$$G_{k+1}y_k = G_k y_k + P_k y_k + Q_k y_k \tag{B.19}$$

为使 G_{k+1} 满足拟牛顿条件，可使 P_k 和 Q_k 满足：

$$P_k y_k = \delta_k \tag{B.20}$$

$$Q_k y_k = -G_k y_k \tag{B.21}$$

事实上，不难找出这样的 P_k 和 Q_k，例如取

$$P_k = \frac{\delta_k \delta_k^{\mathrm{T}}}{\delta_k^{\mathrm{T}} y_k} \tag{B.22}$$

$$Q_k = -\frac{G_k y_k y_k^{\mathrm{T}} G_k}{y_k^{\mathrm{T}} G_k y_k} \tag{B.23}$$

这样就可得到矩阵 G_{k+1} 的迭代公式：

$$G_{k+1} = G_k + \frac{\delta_k \delta_k^{\mathrm{T}}}{\delta_k^{\mathrm{T}} y_k} - \frac{G_k y_k y_k^{\mathrm{T}} G_k}{y_k^{\mathrm{T}} G_k y_k} \tag{B.24}$$

称为 DFP 算法。

BFGS算法

跟DFP思路类似，只不过是近似Hk。此时拟牛顿条件就应该转化成

![](C:\Users\28609\AppData\Roaming\Typora\typora-user-images\image-20241115190507604.png

这时，相应的拟牛顿条件是

$$B_{k+1}\delta_k = y_k \tag{B.25}$$

可以用同样的方法得到另一迭代公式。首先令

$$B_{k+1} = B_k + P_k + Q_k \tag{B.26}$$

$$B_{k+1}\delta_k = B_k\delta_k + P_k\delta_k + Q_k\delta_k \tag{B.27}$$

y_k $\quad -B_k\delta_k$

考虑使 P_k 和 Q_k 满足：

$$P_k\delta_k = y_k \tag{B.28}$$

$P_k = \frac{y_k}{\delta_k}$

$$Q_k\delta_k = -B_k\delta_k \tag{B.29}$$

$Q_k = \frac{-B_k\delta_k}{\delta_k}$

找出适合条件的 P_k 和 Q_k，得到 BFGS 算法矩阵 B_{k+1} 的迭代公式：

$$B_{k+1} = B_k + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}}\delta_k} - \frac{B_k\delta_k\delta_k^{\mathrm{T}}B_k}{\delta_k^{\mathrm{T}}B_k\delta_k} \tag{B.30}$$

可以证明，如果初始矩阵 B_0 是正定的，则迭代过程中的每个矩阵 B_k 都是正定的。