第三节　拟牛顿法

一、拟牛顿法简介

二、拟牛顿条件及拟牛顿法的主要步骤

三、对称秩 1 校正

四、**DFP** 和 **BFGS** 校正（对称秩 2 校正）

一、拟牛顿法简介

牛顿法：$\mathbf{d}^{(k)} = -\nabla^2 f(\mathbf{x}^{(k)})^{-1}\nabla f(\mathbf{x}^{(k)})$，最大优点是在迭代点附近收敛快，因为对目标函数进行二次近似时，利用了目标函数的曲率信息（**Hesse** 矩阵）。

使用 **Hesse** 矩阵的缺点：求逆的计算量大、**Hesse** 矩阵可能非正定，导致牛顿方向可能不是下降方向。

愿望：寻找一种算法，既具有牛顿法收敛速度快的优点，又能克服计算工作量大、产生非下降方向等缺点。

拟牛顿法（**quasi-Newton method**）的基本思想：在迭代过程中只利用目标函数$f(\mathbf{x})$和梯度$\mathbf{g}=\nabla f(\mathbf{x})$信息，构造 **Hesse** 矩阵的近似逆矩阵，由此获得一个搜索方向，生成新的迭代点。

近似矩阵的不同构造方式，对应着拟牛顿法的不同变形。

拟牛顿法是一类收敛速度比较快的算法，是公认的比较有效的无约束优化方法。

二、拟牛顿条件及拟牛顿法的主要步骤

设迭代到当前点$\mathbf{x}^{(k+1)}$，将目标函数$f(\mathbf{x})$在点$\mathbf{x}^{(k+1)}$处作二阶 **Taylor** 近似：

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

上式进一步对 $\mathbf{x}$ 求导得到

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

则在点$\mathbf{x}^{(k)}$处的梯度$\nabla f(\mathbf{x}^{(k)})$近似为

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

即 $\nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)}) \approx \nabla^2 f(\mathbf{x}^{(k+1)})(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$

记

$$\mathbf{q}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)});\quad \mathbf{p}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$$

有

$$\mathbf{q}^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)})\mathbf{p}^{(k)}$$

若 Hesse 矩阵$\nabla^2 f(\mathbf{x}^{(k+1)})$可逆，则

$$\mathbf{p}^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)})^{-1}\mathbf{q}^{(k)}$$

当第 k 次迭代结束时，即可计算出$\mathbf{p}^{(k)}$和$\mathbf{q}^{(k)}$。

如果能用上式信息估计出$\nabla^2 f(\mathbf{x}^{(k+1)})^{-1}$，则下一步迭代的牛顿方向：$\mathbf{d}^{(k+1)} = -\nabla^2 f(\mathbf{x}^{(k+1)})^{-1}\nabla f(\mathbf{x}^{(k+1)})$就可以知道。

直接估计$\nabla^2 f(\mathbf{x}^{(k+1)})^{-1}$的好处是可以避免求逆矩阵的计算量、以及保证对$\nabla^2 f(\mathbf{x}^{(k+1)})$的正定性要求。

令 $$\mathbf{H}_{k+1} \approx \nabla^2 f(\mathbf{x}^{(k+1)})^{-1}$$

表示通过估计得到的点$\mathbf{x}^{(k+1)}$处的 Hesse 矩阵的近似逆矩阵，那么应该有

$$\mathbf{p}^{(k)} \approx \mathbf{H}_{k+1}\mathbf{q}^{(k)} \tag{1}$$

（1）式称为拟牛顿条件[①]。

使用近似矩阵$\mathbf{H}_{k+1}$，则$\mathbf{x}^{(k+1)}$处的迭代方向就应该是

$$\mathbf{d}^{(k+1)} = -\mathbf{H}_{k+1}\mathbf{g}_{k+1} = -\mathbf{H}_{k+1}\nabla f(\mathbf{x}^{(k+1)})$$

[①] 意思就是说，$\mathbf{H}_{k+1}$“像”牛顿方向所用的$\nabla^2 f(\mathbf{x}^{(k+1)})^{-1}$。

拟牛顿法的核心公式：

$$\mathbf{p}^{(k)} \approx \mathbf{H}_{k+1}\mathbf{q}^{(k)}$$

$$\mathbf{d}^{(k+1)} = -\mathbf{H}_{k+1}\mathbf{g}_{k+1} = -\mathbf{H}_{k+1}\nabla f(\mathbf{x}^{(k+1)})$$

称为尺度矩阵。若找到了估计尺度矩阵$\mathbf{H}_{k+1}$的方法，就可以构造拟牛顿算法。

因为在迭代的每一步，$\mathbf{H}_{k+1}$都不同，因此拟牛顿法也称为*变尺度法*。

拟牛顿法的计算步骤：

（1）**初始化**：给定初始点$\mathbf{x}^{(1)} \in \mathbb{R}^n$，初始正定阵$\mathbf{H}_1 \in \mathbb{R}^{n\times n}$，允许误差$\varepsilon \in (0,1)$；计算$\mathbf{g}_1 = \nabla f(\mathbf{x}^{(1)})$，置$k = 1$。

（2）**平稳点检验**：若$\|\mathbf{g}_k\| \leq \varepsilon$，则停止计算；否则，计算搜索方向$\mathbf{d}^{(k)} = -\mathbf{H}_k\mathbf{g}_k$。

（3）**一维搜索**：求步长λ_k，令$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k\mathbf{d}^{(k)}$。

（4）**修正拟牛顿方程**：计算$\mathbf{g}_{k+1} = \nabla f(\mathbf{x}^{(k+1)})$，并对矩阵$\mathbf{H}_k$进行校正，得到$\mathbf{H}_{k+1}$，使之满足拟牛顿条件（1）；置$k = k + 1$，转步骤（2）。

关键在于，如何构造近似矩阵$\mathbf{H}_{k+1}$？有很多方法。

三、对称秩 1 校正

$\mathbf{H}_k$与 Hesse 矩阵的逆一样，须是 n 阶对称正定阵。

构造$\mathbf{H}_{k+1}$的一般策略：初始矩阵$\mathbf{H}_1$取为某 n 阶对称正定矩阵，如单位阵$\mathbf{I}_n$，然后修正$\mathbf{H}_k$给出$\mathbf{H}_{k+1}$。

令

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \Delta\mathbf{H}_k$$

$\Delta\mathbf{H}_k$称为校正矩阵。$\Delta\mathbf{H}_k$的选取要求保持$\mathbf{H}_{k+1}$的可逆、对称、正定。

那么$\Delta\mathbf{H}_k$应该是什么形式?

Sherman-Morrison 定理：已知矩阵$\mathbf{A} \in \mathbb{R}^{n\times n}$可逆，对$\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n$，若$1 + \mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a} \neq 0$，那么

（1）$\bar{\mathbf{A}} = \mathbf{A} + \mathbf{a}\mathbf{b}^{\mathrm{T}}$也是可逆的；

（2）$\bar{\mathbf{A}}$的逆矩阵$\mathbf{A}^{-1}$可直接计算为

$$\bar{\mathbf{A}}^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\mathbf{a}\mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}}{1 + \mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a}}$$

注（对称性），如果选择的上述向量$\mathbf{a}, \mathbf{b}$是同一个向量，比如：$\mathbf{z} = \mathbf{a} = \mathbf{b}$，那么若$\mathbf{A}$是对称可逆的，则新矩阵$\bar{\mathbf{A}}$也是对称可逆的。

运用 S-M 定理的结论（1）构造$\mathbf{H}_k$

确定$\Delta\mathbf{H}_k$的方法之一，取

$$\Delta\mathbf{H}_k = \alpha_k \mathbf{z}^{(k)} \mathbf{z}^{(k)\mathrm{T}}$$

$\mathbf{z}^{(k)} \in \mathbb{R}^n$是$n$维列向量，容易证明$\Delta\mathbf{H}_k$对称且秩为 1。

令

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \Delta\mathbf{H}_k$$

根据 **Sherman-Morrison** 定理的结论（1），如果上次迭代使用的矩阵$\mathbf{H}_k$是 **Hesse** 矩阵的逆矩阵，则矩阵$\mathbf{H}_{k+1}$也仍然对称、可逆。

$\Delta\mathbf{H}_k$的选择应满足拟牛顿条件（1）式，即

$$\mathbf{p}^{(k)} = \mathbf{H}_{k+1}\mathbf{q}^{(k)} = \mathbf{H}_k\mathbf{q}^{(k)} + \alpha_k\mathbf{z}^{(k)}\mathbf{z}^{(k)\mathrm{T}}\mathbf{q}^{(k)} \quad （2）$$

得到

$$\mathbf{z}^{(k)} = \frac{\mathbf{p}^{(k)} - \mathbf{H}_k\mathbf{q}^{(k)}}{\alpha_k\mathbf{z}^{(k)\mathrm{T}}\mathbf{q}^{(k)}}$$

于是，$\Delta\mathbf{H}_k = \alpha_k\mathbf{z}^{(k)}\mathbf{z}^{(k)\mathrm{T}}$变换为

$$\Delta\mathbf{H}_k = \alpha_k\mathbf{z}^{(k)}\mathbf{z}^{(k)\mathrm{T}} = \frac{(\mathbf{p}^{(k)} - \mathbf{H}_k\mathbf{q}^{(k)})(\mathbf{p}^{(k)} - \mathbf{H}_k\mathbf{q}^{(k)})^{\mathrm{T}}}{\alpha_k(\mathbf{z}^{(k)\mathrm{T}}\mathbf{q}^{(k)})^2}$$

为求上式分母，（2）式两端左乘$\mathbf{q}^{(k)\mathrm{T}}$，有

$$\mathbf{q}^{(k)\mathrm{T}}\mathbf{p}^{(k)} = \mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)} + \alpha_k\mathbf{q}^{(k)\mathrm{T}}\mathbf{z}^{(k)}\mathbf{z}^{(k)\mathrm{T}}\mathbf{q}^{(k)}$$

于是有

$$\mathbf{q}^{(k)\mathrm{T}}(\mathbf{p}^{(k)} - \mathbf{H}_k\mathbf{q}^{(k)}) = \alpha_k(\mathbf{z}^{(k)\mathrm{T}}\mathbf{q}^{(k)})^2 \qquad （3）$$

最终，得到秩 1 校正公式[②]：

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{(\mathbf{p}^{(k)}-\mathbf{H}_k\mathbf{q}^{(k)})(\mathbf{p}^{(k)}-\mathbf{H}_k\mathbf{q}^{(k)})^{\mathrm{T}}}{\mathbf{q}^{(k)\mathrm{T}}(\mathbf{p}^{(k)}-\mathbf{H}_k\mathbf{q}^{(k)})} \qquad （4）$$

令搜索方向 $\mathbf{d}^{(k+1)} = -\mathbf{H}_{k+1}\nabla f(\mathbf{x}^{(k+1)})$，做下一次的一维搜索即可。

[②] 显然，根据 Sherman-Morrison 定理，若$\mathbf{q}^{(k)\mathrm{T}}(\mathbf{p}^{(k)} - \mathbf{H}_k\mathbf{q}^{(k)}) > 0$，则（4）式决定的$\mathbf{H}_{k+1}$必然可继承$\mathbf{H}_k$的对称正定性。

对称秩 1 校正的收敛性

可以证明：秩 1 校正在一定条件下收敛，且具有二次终止性。

秩 1 校正的局限性

（1）仅当$\mathbf{q}^{(k)\mathrm{T}}(\mathbf{p}^{(k)}-\mathbf{H}_k\mathbf{q}^{(k)})>0$时，由（4）式得到的$\mathbf{H}_{k+1}$才能确保正定性，而这一点是没有保证的。

（2）即使$\mathbf{q}^{(k)\mathrm{T}}(\mathbf{p}^{(k)}-\mathbf{H}_k\mathbf{q}^{(k)})>0$，因计算时舍入误差的影响，也可能导致$\Delta\mathbf{H}_k$无界，从而产生数值计算上的困难。

四、对称秩 2 校正

（一）DFP 算法

DFP 方法由 Davidon 首先提出，后来又被 Fletcher 和 Powell 改进，得到 DFP 校正公式

$$\Delta \mathbf{H}_k = \frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} - \frac{\mathbf{H}_k\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}} \qquad （5）$$

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} - \frac{\mathbf{H}_k\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}} \qquad （6）$$

可验证，该校正得到的矩阵$\mathbf{H}_{k+1}$满足拟牛顿条件（1）。

秩 2 校正

（5）式的第一项提供了一个关于向量$\mathbf{p}^{(k)}$的对称秩 1 校正，第二项也提供了一个关于向量$\mathbf{H}_k\mathbf{q}^{(k)}$的对称秩 1 校正，根据 Sherman-Morrison 定理，若$\mathbf{H}_k$对称可逆，则得到的$\mathbf{H}_{k+1}$也对称可逆。

但总的来看，上述校正方法提供的校正矩阵$\Delta\mathbf{H}_k$的秩为 2，因此称为秩 2 校正。

尺度矩阵的正定性

Sherman-Morrison 定理只保证了每步迭代得到的尺度矩阵是对称可逆的，但不能保证是正定的。

下述定理 **1** 进一步保证了上述系列矩阵$\mathbf{H}_i$总是正定矩阵，所以迭代方向能使函数值下降。

定理 **1**：若$\mathbf{x}^{(k)}$处的梯度$\mathbf{g}_k \neq \mathbf{0}$, $(k = 1, \dots, n)$，则 **DFP** 方法构造的矩阵$\mathbf{H}_k$, $(k = 1, \dots, n)$为对称正定矩阵。

证明：用归纳法可证（略）。

DFP 方法计算步骤：

（1）给定初始点$\mathbf{x}^{(1)}$，允许误差$\varepsilon > 0$。

（2）置$\mathbf{H}_1 = \mathbf{I}_n$，计算$\mathbf{g}_1 = \nabla f(\mathbf{x}^{(1)})$，置$k = 1$。

（3）$\mathbf{d}^{(k)} = -\mathbf{H}_k \mathbf{g}_k = -\mathbf{H}_k \nabla f(\mathbf{x}^{(k)})$，沿$\mathbf{d}^{(k)}$做一维搜索：

$$\lambda_k = \arg\min_{\lambda \geq 0} f(\mathbf{x}^{(k)} + \lambda \mathbf{d}^{(k)}) \to \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}$$

（5）检验是否满足收敛准则，若$\|\nabla f(\mathbf{x}^{(k+1)})\| \leq \varepsilon$，则停止迭代，得到极小点$\bar{\mathbf{x}} = \mathbf{x}^{(k+1)}$；否则，转步骤（6）。

（6）若$k = n$，则令$\mathbf{x}^{(1)} = \mathbf{x}^{(k+1)}$，返回步骤（2）——初点重置，重新启动算法；若$k < n$，转步骤（7）。

（7）求校正矩阵，令

$$\mathbf{p}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$$

$$\mathbf{g}_{k+1} = \nabla f(\mathbf{x}^{(k+1)})$$

$$\mathbf{q}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$$

$$\Delta\mathbf{H}_k = \frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} - \frac{\mathbf{H}_k\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}}$$

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} - \frac{\mathbf{H}_k\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}}$$

置$k = k + 1$，返回步骤（3）。

例 1：DFP 法求解：$\min f(\mathbf{x}) = 2x_1^2 + x_2^2 - 4x_1 + 2$

取初始点$\mathbf{x}^{(1)} = (2,1)^T$，及初始矩阵$\mathbf{H}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$。目标函数的梯度$\mathbf{g} = \nabla f(\mathbf{x}) = \begin{bmatrix} 4(x_1 - 1) \\ 2x_2 \end{bmatrix}$。

第 1 次迭代: $\mathbf{g}_1 = \nabla f(\mathbf{x}^{(1)}) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, $\mathbf{d}^{(1)} = -\mathbf{H}_1\mathbf{g}_1 = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$;

一维搜索:

$$\min_{\lambda \geq 0} f(\mathbf{x}^{(1)} + \lambda \mathbf{d}^{(1)}) = 3 - 20\lambda + 36\lambda^2$$

解得$\lambda_1 = 5/18$。计算新点:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \lambda_1 \mathbf{d}^{(1)} = \begin{bmatrix} 8/9 \\ 4/9 \end{bmatrix}, \quad \mathbf{g}_2 = \begin{bmatrix} -4/9 \\ 8/9 \end{bmatrix}$$

第 2 次迭代： $\mathbf{p}^{(1)} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)} = \lambda_1 \mathbf{d}^{(1)} = \begin{bmatrix} -10/9 \\ -5/9 \end{bmatrix},$

$$\mathbf{q}^{(1)} = \mathbf{g}_2 - \mathbf{g}_1 = \begin{bmatrix} -40/9 \\ -10/9 \end{bmatrix}$$

校正矩阵：

$$\mathbf{H}_2 = \mathbf{H}_1 + \frac{\mathbf{p}^{(1)}\mathbf{p}^{(1)\mathrm{T}}}{\mathbf{p}^{(1)\mathrm{T}}\mathbf{q}^{(1)}} - \frac{\mathbf{H}_1\mathbf{q}^{(1)}\mathbf{q}^{(1)\mathrm{T}}\mathbf{H}_1}{\mathbf{q}^{(1)\mathrm{T}}\mathbf{H}_1\mathbf{q}^{(1)}} = \frac{1}{306}\begin{bmatrix} 86 & -38 \\ -38 & 305 \end{bmatrix}$$

令 $\mathbf{d}^{(2)} = -\mathbf{H}_2\mathbf{g}_2 = \frac{4}{17}\begin{bmatrix} 1 \\ -4 \end{bmatrix}$

从$\mathbf{x}^{(2)}$出发，沿方向$\mathbf{d}^{(2)}$搜索：$\min_{\lambda\geq 0} f\left(\mathbf{x}^{(2)}+\lambda\mathbf{d}^{(2)}\right)$

解得$\lambda_2=17/36$，计算新点

$$\mathbf{x}^{(3)}=\mathbf{x}^{(2)}+\lambda_2\mathbf{d}^{(2)}=\begin{bmatrix}1\\0\end{bmatrix},\ \mathbf{g}_3=\begin{bmatrix}0\\0\end{bmatrix}$$

得到最优解$\bar{\mathbf{x}}=(1,0)^T$。

此例经两次搜索达到极小点，这不是偶然的：可以证明，**DFP** 方法具有二次终止性。

定理 2：设用 DFP 方法求解下列问题：

$$\min f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + c$$

$\mathbf{A}$为n阶对称正定矩阵。取初点$\mathbf{x}^{(1)} \in \mathbb{R}^n$，令$\mathbf{H}_1$是$n$阶对称正定矩阵，则成立：

$$\mathbf{p}^{(i)\mathrm{T}}\mathbf{A}\mathbf{p}^{(j)} = 0,\ \ 1 \le i < j \le k$$

$$\mathbf{H}_{k+1}\mathbf{A}\mathbf{p}^{(i)} = \mathbf{p}^{(i)},\ \ 1 \le i \le k$$

其中$\mathbf{p}^{(i)} = \mathbf{x}^{(i+1)} - \mathbf{x}^{(i)} = \lambda_i \mathbf{d}^{(i)}, (\lambda_i \neq 0, k \le n)$。

证明：用归纳法可证（略）。

推论：在定理 2 的条件下，必有$\mathbf{H}_{n+1} = \mathbf{A}^{-1}$。

注意到，$\mathbf{p}^{(i)}=\mathbf{x}^{(i+1)}-\mathbf{x}^{(i)}=\lambda_i\mathbf{d}^{(i)}$，所以定理 2 给出的$\mathbf{p}^{(1)},\ldots,\mathbf{p}^{(k)}$关于 **A**共轭，就等价于$\mathbf{d}^{(1)},\ldots,\mathbf{d}^{(k)}$关于**A**共轭。可见，**DFP** 方法构造的搜索方向是一组关于矩阵**A**的共轭方向，因此 **DFP** 方法具有二次终止性，即对于正定二次函数，**DFP** 方法经有限步迭代必达全局极小点。

但是，对于一般（非二次）函数，目前还没有成功建立 **DFP** 方法的全局收敛性结果。

而对于一般的二次函数（不一定凸），**DFP** 方法的收敛性如下：

定理 3：$f(\mathbf{x})$是$\mathbb{R}^n$上的二次连续可微实函数，对任意$\hat{\mathbf{x}} \in \mathbb{R}^n$，存在常数$m > 0$，使得当水平截集$C(\hat{\mathbf{x}})$的元素

$$\mathbf{x} \in C(\hat{\mathbf{x}}) = \{\mathbf{x} | f(\mathbf{x}) \le f(\hat{\mathbf{x}})\},$$

以及$\mathbf{y} \in \mathbb{R}^n$时，有

$$m\|\mathbf{y}\|^2 \le \mathbf{y}^{\mathrm{T}} \nabla^2 f(\mathbf{x}) \mathbf{y}$$ ③

则 DFP 方法以一维搜索产生的序列$\{\mathbf{x}^{(k)}\}$或终止于或收敛于$f(\mathbf{x})$在$\mathbb{R}^n$上的惟一极小点。

---

③ $\mathbf{y}^{\mathrm{T}} \nabla^2 f(\mathbf{x}) \mathbf{y} > 0$则$\nabla^2 f(\mathbf{x})$正定；更直观的说，$\mathbf{y}^{\mathrm{T}} \nabla^2 f(\mathbf{x}) \mathbf{y} \ge m\|\mathbf{y}\|^2$是要其“足够正定”。

（二）BFGS 公式

Broyden，Fletcher，Goldfarb 和 Shanno 于 1970 年提出 BFGS 方法，比 DFP 公式要好，应用广泛。

BFGS 不直接考虑拟牛顿条件：

$$\mathbf{p}^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)})^{-1}\mathbf{q}^{(k)} = \mathbf{H}_{k+1}\mathbf{q}^{(k)}$$

而是考虑关系式：

$$\mathbf{q}^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)})\mathbf{p}^{(k)} = \mathbf{B}_{k+1}\mathbf{p}^{(k)} \qquad （7）$$

即先构造$\mathbf{B}_{k+1}$，然后通过求逆确定尺度矩阵$\mathbf{H}_{k+1}$。

$$\mathbf{H}_{k+1} = \mathbf{B}_{k+1}^{-1} \qquad （8）$$

考察 DFP 公式：

$$\mathbf{p}^{(k)} \approx \mathbf{H}_{k+1}\mathbf{q}^{(k)} \text{——DFP}$$

和 BFGS 公式：

$$\mathbf{q}^{(k)} \approx \mathbf{B}_{k+1}\mathbf{p}^{(k)} \text{——BFGS}$$

显然，矩阵$\mathbf{H}_{k+1}$和$\mathbf{B}_{k+1}$、向量$\mathbf{p}^{(k)}$和$\mathbf{q}^{(k)}$彼此分别处于对称的位置，因此只要在$\mathbf{H}_k$的递推公式中互换$\mathbf{p}^{(k)}$和$\mathbf{q}^{(k)}$，并用$\mathbf{B}_{k+1}$和$\mathbf{B}_k$分别取代$\mathbf{H}_{k+1}$和$\mathbf{H}_k$，就能得到$\mathbf{B}_k$的递推公式。

矩阵 B 的 BFGS 修正（DFP 公式的对偶）公式：

$$\mathbf{B}_{k+1}=\mathbf{B}_k+\frac{\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{p}^{(k)}}-\frac{\mathbf{B}_k\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}\mathbf{B}_k}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{B}_k\mathbf{p}^{(k)}} \qquad （9）$$

对比 DFP： $\mathbf{H}_{k+1}=\mathbf{H}_k+\frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}}-\frac{\mathbf{H}_k\mathbf{q}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k}{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}}$。

和 DFP 校正的（5）式类似，BFGS 校正矩阵的第一项提供了一个关于向量$\mathbf{q}^{(k)}$的对称秩 1 校正，第二项提供了一个关于向量$\mathbf{B}_k\mathbf{p}^{(k)}$的对称秩 1 校正，而总体上校正矩阵$\Delta\mathbf{B}_k$的秩为 2，因此 BFGS 方法也是秩 2 校正。

运用 S-M 定理的结论（2）求逆

根据 Sherman-Morrison 定理，若$\mathbf{B}_k$对称可逆，则$\mathbf{B}_{k+1}$也对称可逆，且其逆矩阵$\mathbf{B}_{k+1}^{-1}$可直接由定理的结果（2）给出。记，

$$\mathbf{H}_{k+1}^{\mathrm{BFGS}} = \mathbf{B}_{k+1}^{-1}$$

可得到关于矩阵H的 BFGS 公式为：

$$\mathbf{H}_{k+1}^{\mathrm{BFGS}} = \mathbf{H}_{k}^{\mathrm{BFGS}} + \left(1 + \frac{\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k\mathbf{q}^{(k)}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}}\right)\frac{\mathbf{p}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} - \frac{\mathbf{p}^{(k)}\mathbf{q}^{(k)\mathrm{T}}\mathbf{H}_k + \mathbf{H}_k\mathbf{q}^{(k)}\mathbf{p}^{(k)\mathrm{T}}}{\mathbf{p}^{(k)\mathrm{T}}\mathbf{q}^{(k)}} \qquad （10）$$

例 2：用 **BFGS** 法求解问题

$$\min f(\mathbf{x}) = 2x_1^2 + x_2^2 - 4x_1 + 2$$

取初始点$\mathbf{x}^{(1)} = (2,1)^T$及初始矩阵$\mathbf{H}_1^{\mathrm{BFGS}} = \mathbf{I}_2$，$\varepsilon = 0$。

目标函数的梯度$\mathbf{g} = \nabla f(\mathbf{x}) = \begin{bmatrix} 4(x_1 - 1) \\ 2x_2 \end{bmatrix}$，所以

$$\mathbf{g}_1 = \nabla f\left(\mathbf{x}^{(1)}\right) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

第 1 次迭代：搜索方向 $\mathbf{d}^{(1)} = -\mathbf{H}_1^{\mathrm{BFGS}}\mathbf{g}_1 = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$，从$\mathbf{x}^{(1)}$出发作一维搜索：$\min_{\lambda \geq 0} f\left(\mathbf{x}^{(1)} + \lambda \mathbf{d}^{(1)}\right) = 3 - 20\lambda + 36\lambda^2$

解得$\lambda_1 = 5/18$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \lambda_1 \mathbf{d}^{(1)} = \begin{bmatrix} 8/9 \\ 4/9 \end{bmatrix}, \quad \mathbf{g}_2 = \begin{bmatrix} -4/9 \\ 8/9 \end{bmatrix}$$

$$\mathbf{p}^{(1)} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)} = \begin{bmatrix} -10/9 \\ -5/9 \end{bmatrix}, \quad \mathbf{q}^{(1)} = \mathbf{g}_2 - \mathbf{g}_1 = \begin{bmatrix} -40/9 \\ -10/9 \end{bmatrix}$$

由（10）式计算出

$$\mathbf{H}_2^{\mathrm{BFGS}} = \mathbf{H}_1^{\mathrm{BFGS}} + \left(1 + \frac{\mathbf{q}^{(1)\mathrm{T}}\mathbf{H}_1\mathbf{q}^{(1)}}{\mathbf{p}^{(1)\mathrm{T}}\mathbf{q}^{(1)}}\right)\frac{\mathbf{p}^{(1)}\mathbf{p}^{(1)\mathrm{T}}}{\mathbf{p}^{(1)\mathrm{T}}\mathbf{q}^{(1)}} - \frac{\mathbf{p}^{(1)}\mathbf{q}^{(1)\mathrm{T}}\mathbf{H}_1 + \mathbf{H}_1\mathbf{q}^{(1)}\mathbf{p}^{(1)\mathrm{T}}}{\mathbf{p}^{(1)\mathrm{T}}\mathbf{q}^{(1)}} = \frac{1}{162}\begin{bmatrix} 46 & -22 \\ -22 & 169 \end{bmatrix}$$

第 2 次迭代：搜索方向$\mathbf{d}^{(2)} = -\mathbf{H}_2^{\mathrm{BFGS}}\mathbf{g}_2 = \begin{bmatrix} 20/81 \\ -80/81 \end{bmatrix}$

求解一维搜索：

$$\min_{\lambda\geq 0} f\left(\mathbf{x}^{(2)} + \lambda\mathbf{d}^{(2)}\right) = \frac{2}{729}(9-20\lambda)^2$$

解得$\lambda_2 = 9/20$。

新的迭代点$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} + \lambda_2\mathbf{d}^{(2)} = (1,0)^{\mathrm{T}}$，对应的函数值 $f(\mathbf{x}^{(3)}) = 0$。

$\mathbf{g}_3 = \nabla f(\mathbf{x}^{(3)}) = 0$，所以$\mathbf{x}^{(3)} = (1,0)^{\mathrm{T}}$就是最优解.

拟牛顿法的优点：

（1）迭代中仅需一阶导数，不必计算 **Hesse** 矩阵；

（2）$\mathbf{H}_k$正定，所以算法产生的方向均为下降方向；

（3）具有二次终止性，对于一般情形，具有超线性收敛速率；而且还具有 n 步二级收敛速率。

拟牛顿法的缺点：

BFGS 方法所需存储量较大，但在当今世界存储不是最大的难题；除非面对百万级甚至上亿个变量，这才是需要注意的问题。

（三）关于拟牛顿法的推广

DFP 和 BFGS 校正，都是由$\mathbf{p}^{(k)}$和$\mathbf{H}_k\mathbf{q}^{(k)}$构成的秩 2 校正，特点是它们的线性组合也可作为一种校正。

可将 DFP 和 BFGS 做线性组合，得到新的校正

$$\mathbf{H}_{k+1} = (1-\theta_k)\mathbf{H}_{k+1}^{\mathrm{DFP}} + \theta_k\mathbf{H}_{k+1}^{\mathrm{BFGS}}$$

或者从$\mathbf{B}_k$矩阵入手 $\mathbf{B}_{k+1} = (1-\phi_k)\mathbf{B}_{k+1}^{\mathrm{BFGS}} + \phi_k\mathbf{B}_{k+1}^{\mathrm{DFP}}$。

简单说，上述不同参数组合得到的拟牛顿矩阵不同，形成了一大类校正公式，称为 Broyden 族。前面的秩 1 校正也是这个族的成员之一。

基于 **Sherman-Morrison 定理**的矩阵校正只是其中一种矩阵校正方法，除了 **Broyden** 族校正公式，使用其他矩阵校正方法也能得到不同公式，如 **Huang 族**校正。

Broyden 族需要一个参数θ_k或ϕ_k，而 **Huang** 族需要三个参数。

Broyden 族是 **Huang** 族的一个子族。

作业题

1、分别用 DFP 法和 BFGS 法求解下述问题：

$$\min f(\mathbf{x}) = x_1^2 - 2x_1x_2 + 4x_2^2 + x_1 - 3x_2$$

取初始点$\mathbf{x}^{(1)} = (1,1)^{\mathrm{T}}$。

2、考虑习题 1 中 DFP 和 BFGS 方法各自所构造的搜索方向序列，完成下述计算：

（1）考察各自的搜索方向是否关于 **Hesse** 矩阵共轭？

（2）取初始点 $\mathbf{x}^{(1)} = (1,1)^{\mathrm{T}}$，以相反顺序使用搜索方向重做一维搜索，指出与原搜索过程的异、同点。