

The background of the slide is a faded photograph of the main entrance gate of Jiaotong University. The gate is a large, white, curved structure with the university's name in Chinese characters '交通大学' and English 'JIAOTONG UNIVERSITY' inscribed on it. To the left of the gate is a stone wall with a large, intricate relief sculpture. Several people are visible walking through the gate. The overall scene is bright and clear.

4-4 拟牛顿法(1)

拟牛顿法

牛顿法的缺点: Hesse矩阵涉及大量二阶偏导数计算, 迭代点需要充分靠近极小点, Hesse矩阵正定性无法保证。是否能够构造一类算法与牛顿法一样具有二阶收敛速度, 但不需要计算二阶偏导数?

假定当前迭代点 \boldsymbol{x}_{k+1} , 牛顿法中的牛顿方程为:

$$G_{k+1}\boldsymbol{d} = -\boldsymbol{g}_{k+1}$$

若能够构造出一个正定矩阵 B_{k+1} 作为 G_{k+1} 的近似, 搜索方向 \boldsymbol{d}_{k+1} 由方程组:

$$B_{k+1}\boldsymbol{d} = -\boldsymbol{g}_{k+1}$$

给出。这里仍然需要解线性方程组。

拟牛顿法

或者用相同的信息构造出正定矩阵 H_{k+1} 作为 G_{k+1}^{-1} 的近似，搜索方向 \mathbf{d}_{k+1} 由方程组：

$$\mathbf{d} = -H_{k+1}\mathbf{g}_{k+1}$$

决定。

近似矩阵的构造应该是简单有效，它应具有如下的条件：

- 只需要 f 的一阶导数信息；
- $B_{k+1}(H_{k+1})$ 正定，保证搜索方向的下降性；
- 方法具有较快的收敛速度。

拟牛顿条件

考虑 $f(\mathbf{x})$ 在 \mathbf{x}_{k+1} 附近的二阶泰勒展开为:

$$f(\mathbf{x}) = f(\mathbf{x}_{k+1}) + \mathbf{g}_{k+1}^T(\mathbf{s}) + \frac{1}{2}\mathbf{s}^T G_{k+1} \mathbf{s} + O(\|\mathbf{s}\|^2), \quad (4.4.1)$$

其中 $\mathbf{s} = \mathbf{x} - \mathbf{x}_{k+1}$ 。对 $f(\mathbf{x})$ 求导可得:

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}_{k+1} + G_{k+1} \mathbf{s} + O(\|\mathbf{s}\|^2). \quad (4.4.2)$$

令 $\mathbf{x} = \mathbf{x}_k$,得

$$\mathbf{g}(\mathbf{x}_k) = \mathbf{g}_{k+1} + G_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) + O(\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2) \quad (4.4.3)$$

拟牛顿条件

记:

$$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \quad (4.4.4a)$$

$$\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k \quad (4.4.4b)$$

于是(4.4.3)可写成:

$$G_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k + O(\|\boldsymbol{s}_k\|^2).$$

B_{k+1} 作为 G_{k+1} 的近似矩阵, 应满足:

$$B_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k. \quad (4.4.5)$$

该方程称为**拟牛顿方程**或者**拟牛顿条件**。

拟牛顿条件

若记: $H_{k+1} = B_{k+1}^{-1}$, 则拟牛顿条件为:

$$H_{k+1} \mathbf{y}_k = \mathbf{s}_k. \quad (4.4.6)$$

若 B_{k+1} 满足拟牛顿条件(4.4.5), 则在 \mathbf{x}_{k+1} 点处的二次模型:

$$m_{k+1}(\mathbf{x}) = f(\mathbf{x}_{k+1}) + \mathbf{g}_{k+1}^T (\mathbf{x} - \mathbf{x}_{k+1}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_{k+1})^T B_{k+1} (\mathbf{x} - \mathbf{x}_{k+1}) \quad (4.4.7)$$

满足插值性质:

$$m_{k+1}(\mathbf{x}_{k+1}) = f(\mathbf{x}_{k+1}), \nabla m_{k+1}(\mathbf{x}_{k+1}) = \mathbf{g}_{k+1}, \nabla m_{k+1}(\mathbf{x}_k) = \mathbf{g}_k. \quad (4.4.8)$$

上式中: (1) 第一、第二个等式是显然的. (2) 第三个等式是利用拟牛顿条件可得.

拟牛顿方法

拟牛顿法是指通过 $B_{k+1}\mathbf{d} = -\mathbf{g}_{k+1}$ 确定迭代方向的最优化方法，这里 B_{k+1} 或者 H_{k+1} 应该满足以上拟牛顿条件。

算法4.4.1 – 拟牛顿法

步1: 给出 $\mathbf{x}_0 \in \mathbb{R}^n$, B_0 (或 H_0) $\in \mathbb{R}^{n \times n}$, $\varepsilon > 0$, $k := 0$.

步2: 若停机条件满足, 停止迭代并输出结果。

步3: 计算 $B_k\mathbf{d} = -\mathbf{g}_k$ (或 $\mathbf{d}_k = -H_k\mathbf{g}_k$) 得到搜索方向 \mathbf{d}_k 。

步4: 沿着 \mathbf{d}_k 方向进行线性搜索 $\alpha_k > 0$, 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$ 。

步5: 修正 B_k 得 B_{k+1} (或校正 H_k 得 H_{k+1}), 使得拟牛顿条件仍然成立。令 $k := k + 1$, 转步2。

注意: B_0 通常取为单位矩阵, 即 $B_0 = I$ 。拟牛顿法的第一次迭代等价于一步最速下降迭代。

拟牛顿修正

在拟牛顿法的每一步迭代中，如何修正 H_k 得到 H_{k+1} ，即：

$$H_{k+1} = H_k + \Delta H_k$$

确定修正量 ΔH_k 的方法是多种多样的，但它应当具简单、计算量小、有效的特点。

关于 H_k 经典的修正方式有：

- 对称秩1修正

$$H_{k+1} = H_k + \beta \mathbf{u} \mathbf{u}^T, \quad \mathbf{u} \in \mathbb{R}^n, \beta \in \mathbb{R} \quad (4.4.9)$$

- 对称秩2修正

$$H_{k+1} = H_k + \beta \mathbf{u} \mathbf{u}^T + \gamma \mathbf{v} \mathbf{v}^T, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \beta, \gamma \in \mathbb{R} \quad (4.4.10)$$

对称秩1修正公式

将(4.4.9)代入拟牛顿方程，则有：

$$H_k \mathbf{y}_k + \beta \mathbf{u} \mathbf{u}^T \mathbf{y}_k = \mathbf{s}_k$$

即有：

$$\mathbf{s}_k - H_k \mathbf{y}_k = \beta \mathbf{u} (\mathbf{u}^T \mathbf{y}_k) \quad (4.4.11)$$

上式中表明 \mathbf{u} 与 $\mathbf{s}_k - H_k \mathbf{y}_k$ 共线，从而存在常数 $\gamma \in \mathbb{R}$ ，使得：

$$\mathbf{u} = \gamma (\mathbf{s}_k - H_k \mathbf{y}_k) \quad (4.4.12)$$

将(4.4.12)代入(4.4.11)得：

$$\mathbf{s}_k - H_k \mathbf{y}_k = \beta \gamma^2 [\mathbf{s}_k - H_k \mathbf{y}_k] [(\mathbf{s}_k - H_k \mathbf{y}_k)^T \mathbf{y}_k] \quad (4.4.13)$$

对称秩1修正公式

比较(4.4.13)两边可得:

$$\beta\gamma^2[(\mathbf{s}_k - H_k\mathbf{y}_k)^T\mathbf{y}_k] = 1 \rightarrow \beta\gamma^2 = \frac{1}{(\mathbf{s}_k - H_k\mathbf{y}_k)^T\mathbf{y}_k}.$$

将(4.4.12)与上式代入(4.4.9)可得修正公式:

$$H_{k+1}^{\text{SR1}} = H_k + \frac{(\mathbf{s}_k - H_k\mathbf{y}_k)(\mathbf{s}_k - H_k\mathbf{y}_k)^T}{(\mathbf{s}_k - H_k\mathbf{y}_k)^T\mathbf{y}_k} \quad (4.4.14)$$

利用Shermann-Morrison-Woodbury公式(教材P61), 可得到 H_{k+1}^{SR1} 得逆矩阵校正公式:

$$B_{k+1}^{\text{SR1}} = B_k + \frac{(\mathbf{y}_k - B_k\mathbf{s}_k)(\mathbf{y}_k - B_k\mathbf{s}_k)^T}{(\mathbf{y}_k - B_k\mathbf{s}_k)^T\mathbf{s}_k} \quad (4.4.15)$$

对称秩2修正- DFP公式

将(4.4.10)代入拟牛顿方程得:

$$\mathbf{s}_k = H_k \mathbf{y}_k + \beta \mathbf{u} \mathbf{u}^T \mathbf{y}_k + \gamma \mathbf{v} \mathbf{v}^T \mathbf{y}_k \quad (4.4.16)$$

注意这里 \mathbf{u} 和 \mathbf{v} 并不唯一确定, 但 \mathbf{u} 和 \mathbf{v} 的明显的选择:

$$\mathbf{u} = \mathbf{s}_k, \quad \mathbf{v} = H_k \mathbf{y}_k, \quad \beta \mathbf{u}^T \mathbf{y}_k = 1, \quad \gamma \mathbf{v}^T \mathbf{y}_k = -1$$

从而有:

$$\beta = \frac{1}{\mathbf{u}^T \mathbf{y}_k} = \frac{1}{\mathbf{s}_k^T \mathbf{y}_k}, \quad \gamma = -\frac{1}{\mathbf{v}^T \mathbf{y}_k} = -\frac{1}{\mathbf{y}_k^T H_k \mathbf{y}_k}.$$

对称秩2修正- DFP公式

代入(4.4.10), 得到以下DFP校正公式(关于 H_k).

$$H_{k+1}^{\text{DFP}} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}. \quad (4.4.17)$$

称采用DFP修正公式的拟牛顿法为DFP方法。

和秩1修正公式类似, 利用利用Shermann-Morrison-Woodbury公式可得到 B_{k+1} 的修正公式如下:

$$B_{k+1}^{\text{DFP}} = B_k + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T}{s_k^T y_k} - \left(\frac{y_k s_k^T B_k + B_k s_k y_k^T}{s_k^T y_k}\right). \quad (4.4.18)$$

对称秩2修正- BFGS公式

考虑 B_k 的秩2修正公式:

$$B_{k+1} = B_k + \beta \mathbf{u} \mathbf{u}^T + \gamma \mathbf{v} \mathbf{v}^T$$

采用与DFP方法相似的推导方法可以得到关于 B_k 的BFGS公式:

$$B_{k+1}^{\text{BFGS}} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k}. \quad (4.4.19)$$

利用利用Shermann-Morrison-Woodbury公式可得到 H_{k+1} 的修正公式:

$$H_{k+1}^{\text{BFGS}} = H_k + \left(1 + \frac{\mathbf{y}_k^T H_k \mathbf{y}_k}{\mathbf{y}_k^T \mathbf{s}_k}\right) \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \left(\frac{\mathbf{s}_k \mathbf{y}_k^T H_k + H_k \mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}\right). \quad (4.4.20)$$

对偶关系与Broyden族公式

比较(4.4.19)与(4.4.17), (4.4.20)与(4.4.18), 容易发现如果在(4.4.17)与(4.4.18)中交换 B_k 与 H_k , 以及 s_k 与 y_k , 便可得到公式(4.4.19)与(4.4.20)。由于这种关系, 称DFP与BFGS是互为对偶的方法。注意秩1修正是自对偶方法。

由 H_{k+1}^{DFP} 与 H_{k+1}^{BFGS} 可以构造出一族拟牛顿方法的修正公式, 称其为Broyden族公式:

$$H_{k+1}^{\varphi} = (1 - \varphi)H_{k+1}^{\text{DFP}} + \varphi H_{k+1}^{\text{BFGS}} \quad (4.4.21)$$

这里 $\varphi \geq 0$ 。DFP公式与BFGS公式都是Broyden族公式的特殊情形。

DFP方法算例

例4.4.1 用DFP方法求解: $\min f(\mathbf{x}) = \{x_1^2 + 4x_2^2\}$, 取初始点 $\mathbf{x}_0 = (1, 1)^T$, H_0 取为单位阵。

解: 当 H_0 取单位阵, 第一步迭代为最速下降法, 计算:

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}, \mathbf{g}(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix}$$

$$\mathbf{d}_0 = -\mathbf{g}_0 = \begin{pmatrix} -2 \\ -8 \end{pmatrix}, \alpha_0 = -\frac{\mathbf{d}_0^T \mathbf{g}_0}{\mathbf{d}_0^T Q \mathbf{d}_0} = 0.13077$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 = \begin{pmatrix} 0.73846 \\ -0.04616 \end{pmatrix}, \mathbf{g}_1 = \begin{pmatrix} 1.47692 \\ -0.36923 \end{pmatrix}$$

解续:

$$\mathbf{s}_0 = \mathbf{x}_1 - \mathbf{x}_0 = \begin{pmatrix} -\frac{34}{130} \\ -\frac{136}{130} \end{pmatrix}, \mathbf{y}_0 = \mathbf{g}_1 - \mathbf{g}_0 = \begin{pmatrix} -\frac{68}{130} \\ -\frac{1088}{130} \end{pmatrix}$$

$$\begin{aligned} H_1 &= H_0 + \frac{\mathbf{s}_0 \mathbf{s}_0^T}{\mathbf{s}_0^T \mathbf{y}_0} - \frac{H_0 \mathbf{y}_0 \mathbf{y}_0^T H_0}{\mathbf{y}_0^T H_0 \mathbf{y}_0} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0.00769 & 0.03077 \\ 0.03077 & 0.12308 \end{pmatrix} - \begin{pmatrix} 0.00389 & 0.06226 \\ 0.06226 & 0.99611 \end{pmatrix} \\ &= \begin{pmatrix} 1.00380 & -0.03149 \\ -0.03149 & 0.12697 \end{pmatrix} \end{aligned}$$

解续:

$$\mathbf{d}_1 = -H_1 \mathbf{g}_1 = \begin{pmatrix} -1.49416 \\ 0.09340 \end{pmatrix}, \alpha_1 = -\frac{\mathbf{d}_1^T \mathbf{g}_1}{\mathbf{d}_1^T Q \mathbf{d}_1} = 0.49423$$

$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{g}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

所以极小点为: $\mathbf{x}^* = \mathbf{x}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

□

BFGS方法算例

例4.4.2 用BFGS方法求解: $f(\mathbf{x}) = \min\{\frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1\}$, 取初始点 $\mathbf{x}_0 = (0, 0)^T$, H_0 取为单位阵。迭代终止精度为 $\varepsilon = 0.0001$ 。

解: 当 H_0 取单位阵, 第一步迭代为最速下降法, 计算:

$$Q = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}, \mathbf{g}(\mathbf{x}) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix}, \mathbf{d}_0 = -H_1\mathbf{g}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\alpha_0 = 0.3333, \quad \mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{d}_0 = \begin{pmatrix} 0.6666 \\ 0 \end{pmatrix}, \mathbf{g}_1 = \begin{pmatrix} -0.0002 \\ -0.6666 \end{pmatrix},$$

$$\mathbf{s}_0 = \mathbf{x}_1 - \mathbf{x}_0 = \begin{pmatrix} 0.6666 \\ 0.0000 \end{pmatrix}, \mathbf{y}_0 = \mathbf{g}_1 - \mathbf{g}_0 = \begin{pmatrix} 1.9998 \\ -0.6666 \end{pmatrix}$$

BFGS方法算例

解续：由BFGS校正公式得：

$$H_1 = \begin{pmatrix} 0.3666 & 0.10000 \\ 0.10000 & 0.30000 \end{pmatrix}, \mathbf{d}_1 = -H_1 \mathbf{g}_1 = \begin{pmatrix} 0.0667 \\ 0.2000 \end{pmatrix}$$

$$\alpha_1 = 5, \mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = \begin{pmatrix} 1.0001 \\ 1.0000 \end{pmatrix}, \mathbf{g}_2 = \begin{pmatrix} 0.0003 \\ -0.0001 \end{pmatrix}$$

注意 $\|\mathbf{g}_2\| = 0.00003 < \varepsilon$, 停止迭代, 近似解为:

$$\mathbf{x}^* = \mathbf{x}_2 = \begin{pmatrix} 1.000 \\ 1.000 \end{pmatrix}$$



拟牛顿法对比

例4.4.3 教材P76数值试验1。

$\sigma = 0.1$ 时, SR1, BFGS 与 DFP 三种方法所需的
迭代次数和函数调用次数

n	SR1 方法		BFGS 方法		DFP 方法	
	ite	feva	ite	feva	ite	feva
2	44	274	23	136	27	167
10	65	412	26	137	33	173
20	72	398	27	140	34	177
40	91	560	25	135	35	181
60	114	658	27	140	37	183

$\sigma = 0.9$ 时, SR1, BFGS 与 DFP 三种方法所需的
迭代次数和函数调用次数

n	SR1 方法		BFGS 方法		DFP 方法	
	ite	feva	ite	feva	ite	feva
2	52	194	34	82	40	91
10	64	249	40	101	51	114
20	67	275	42	106	49	114
40	87	431	43	109	46	109
60	73	310	41	106	56	130