

# 牛顿和拟牛顿方法

2025 年 12 月 17 日

文档主要分为两个部分，第一部分是对牛顿、拟牛顿方法的介绍，第二部分介绍具体的拟牛顿数值方法。

## 1 非线性方程组解法——牛顿法

Newton 法又称为 Newton-Raphson 法，是采用函数  $\mathbf{F}(\mathbf{x})$  的 Taylor 级数的前面几项来寻找方程  $\mathbf{F}(\mathbf{x}) = 0$  的根。Newton 法是把非线性方程  $\mathbf{F}(\mathbf{x}) = 0$  线性化的一种方法。对于单变量又称为切线法，对于多变量则是采用目标函数的 Jocobi 矩阵对非线性方程组的解进行迭代，是用超切平面趋近，拟牛顿法采用割线的思想，用近似矩阵趋近 Jocobi 矩阵，然后对非线性方程组的解进行迭代，是用超割平面去趋近。牛顿法为平方收敛，自矫正误差不会传递。但是求导过程较为复杂，并且赋值后的 Jocobi 矩阵可能会稀疏性，奇异性或者病态等问题，导致数值不稳定性。拟牛顿法的收敛速度介于直线收敛之间，速度会慢于牛顿法，但是稳定性很高。割线法有很多种，2 点割线法，n 点割线法，n+1 点割线法（n 是方程的个数），其中 n+1 点割线法的效率是最高的。拟牛顿法具有 n+1 点割线法的效率，同时具备很高的稳定性。

对于给定的一个  $n \times n$  的非线性方程组

$$\mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})]^T = 0, \quad (1)$$

对应的 Jocobi 矩阵为

$$\mathbf{F}'(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad (2)$$

对于牛顿法， $x_0$  给定，然后进行迭代

$$\mathbf{x}^{k+1} = \mathbf{x}^k - [\mathbf{F}'(\mathbf{x}^k)]^{-1} \mathbf{F}(\mathbf{x}^k). \quad (3)$$

令  $A_k = \mathbf{F}'(\mathbf{x}^k)$ ，则原式为

$$\mathbf{x}^{k+1} = \mathbf{x}^k - A_k^{-1} \mathbf{F}(\mathbf{x}^k), \quad (4)$$

## 2 拟牛顿法数值方法的介绍

拟牛顿法的基本思路是不断用一个低秩矩阵对  $A_k$  进行修正，低秩矩阵不同得到的方法也不同。拟牛顿法主要用来解决两类问题，一是  $n$  个方程  $n$  个未知数的系统，二是非约束函数的最小值问题。拟牛顿法解决这两类问题时主要的不同在于生成的近似导数矩阵不同，对于非线性问题，采用的是 Jacobian 矩阵，即目标函数的一阶导矩阵，对于非约束最小值问题，采用的是 Hessian 矩阵，即目标函数的二阶导矩阵。对于非约束最小值问题求解的拟牛顿法，其种类是很多的，并且随着机器学习等学科的兴起而不断的增加，但是这些方法之间的关系并不是十分清楚的 [1]。而对于非线性方程求解的拟牛顿方程，主要采用的是 Broyden 方法 (1965)[2]，而对于非约束最小值问题，则采用 Powell 对称形式的 Broyden 方法 [3]，Davidon-Fletcher-Powell(DFP)[4, 5] 方法以及 Broyden-Fletcher-Goldfarb-Shanno(BFGS) 方法 [6, 7, 8, 9]。

## 2.1 各个方法的迭代表达式

1. Broyden 方法是秩为 1 的算法 [2]

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - B_k^{-1} \mathbf{F}(\mathbf{x}^k), \\ B_{k+1} = B_k + \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)(\mathbf{s}_k)^T}{(\mathbf{s}_k)^T \mathbf{s}_k}, \end{cases} \quad (5)$$

这里  $B$  表示的是目标函数的一阶导矩阵，即 Jacobi 矩阵，其中  $\mathbf{s}_k = \mathbf{x}^{k+1} - \mathbf{x}^k$ ,  $\mathbf{y}_k = \mathbf{F}(\mathbf{x}^{k+1}) - \mathbf{F}(\mathbf{x}^k)$ 。方程 (5) 中要求线性系统  $B_k \mathbf{s}_k = -\mathbf{F}(\mathbf{x}^k)$  的解，这个要求是比较难解的。为了克服这个困难，可以通过 Sherman 和 Morrison 的理论来解决 [10]。该理论得出这样一个定理：

**定理 1.** 令  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$  并假设  $A \in L(\mathbb{R}^n)$  是非奇异的，那么  $A + \mathbf{u}\mathbf{v}^T$  是非奇异的当且仅当  $\sigma = 1 + \mathbf{v}^T A^{-1} \mathbf{u} \neq 0$ . 如果  $\sigma \neq 0$ , 那么

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - (1/\sigma)A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}. \quad (6)$$

根据这一条定理，方程 (5) 可以写成不含求逆项的方程，或者在实际迭代过程中只需要一次求逆的迭代表达式。令  $H_{k+1} = B_{k+1}^{-1}$ ,  $H_k = B_k^{-1}$ , 根据方程 (5) 的第二式以及方程 (6)，我们可以得到

$$\begin{aligned} B_{k+1}^{-1} &= (B_k + \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)(\mathbf{s}_k)^T}{(\mathbf{s}_k)^T \mathbf{s}_k})^{-1} \\ &= B_k^{-1} - (1/\sigma)B_k^{-1} \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)(\mathbf{s}_k)^T}{(\mathbf{s}_k)^T \mathbf{s}_k} B_k^{-1}, \end{aligned}$$

其中

$$\sigma = 1 + \mathbf{s}_k^T B_k^{-1} \frac{\mathbf{y}_k - B_k \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{s}_k}.$$

则有

$$\begin{aligned} B_{k+1}^{-1} &= B_k^{-1} - \frac{B_k^{-1}(\mathbf{y}_k - B_k \mathbf{s}_k)\mathbf{s}_k^T B_k^{-1}}{(1 + \mathbf{s}_k^T B_k^{-1} \frac{\mathbf{y}_k - B_k \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{s}_k}) \mathbf{s}_k^T \mathbf{s}_k} \\ &= B_k^{-1} + \frac{(\mathbf{s}_k - B_k \mathbf{y}_k)\mathbf{s}_k^T B_k^{-1}}{\mathbf{s}_k^T B_k^{-1} \mathbf{y}_k}. \end{aligned}$$

因此可以得到改进的 Broyden 表达式

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - H_k \mathbf{F}(\mathbf{x}^k), \\ H_{k+1} = H_k + \frac{(\mathbf{s}_k - H_k \mathbf{y}_k)(\mathbf{s}_k)^T H_k}{(\mathbf{s}_k)^T H_k \mathbf{y}_k}. \end{cases} \quad (7)$$

对于非约束最小值问题，采用的是 Hessian 矩阵，因此，以下  $B$  代表的是目标函数的二阶导矩阵，即 Hessian 矩阵。

2. Powell 对称形式的 Broyden 方法是一个秩为 2 的算法 [3], 是一个具有对称性的算法

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - B_k^{-1} \mathbf{F}(\mathbf{x}^k), \\ B_{k+1} = B_k + \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)\mathbf{s}_k^T + \mathbf{s}_k(\mathbf{y}_k - B_k \mathbf{s}_k)^T}{\mathbf{s}_k^T \mathbf{s}_k} - \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)^T \mathbf{s}_k \mathbf{s}_k \mathbf{s}_k^T}{[(\mathbf{s}_k)^T \mathbf{s}_k]^2}. \end{cases} \quad (8)$$

3. Davidon-Fletcher-Powell(DFP) 方法 [4, 5]

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - B_k^{-1} \mathbf{F}(\mathbf{x}^k), \\ B_{k+1} = B_k + \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)\mathbf{y}_k^T + \mathbf{y}_k(\mathbf{y}_k - B_k \mathbf{s}_k)^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{(\mathbf{y}_k - B_k \mathbf{s}_k)^T \mathbf{s}_k \mathbf{y}_k \mathbf{y}_k^T}{[(\mathbf{y}_k)^T \mathbf{s}_k]^2}. \end{cases} \quad (9)$$

在对称性的基础上，还具有继承正定性，即当  $B_k$  为正定的时候， $B_{k+1}$  也为正定的。

即如果令  $H_k = B_k^{-1}$ , 可以得到改进的 DFP 方法

$$\begin{cases} \mathbf{x}^{k+1} &= \mathbf{x}^k - H_k \mathbf{F}(\mathbf{x}^k), \\ H_{k+1} &= H_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{H_k \mathbf{y}_k \mathbf{y}_k^T H_k}{(\mathbf{y}_k)^T H_k \mathbf{y}_k}. \end{cases} \quad (10)$$

4. BFGS(Broyden-Fletcher-Goldfarb-Shanno) 方法 [6, 7, 8, 9]

$$\begin{cases} \mathbf{x}^{k+1} &= \mathbf{x}^k - H_k \mathbf{F}(\mathbf{x}^k), \\ H_{k+1} &= (I - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}) H_k (I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}. \end{cases} \quad (11)$$

BFGS 方法又称为互补 DFP 方法。BFGS 方法的更新矩阵  $\bar{H}_{BFGS}$  和 DFP 方法的更新矩阵  $\bar{H}_{DFP}$  满足关系式

$$\bar{H}_{BFGS} = \bar{H}_{DFP} + \mathbf{v} \mathbf{v}^T,$$

其中

$$\mathbf{v} = (\mathbf{y}^T H \mathbf{y})^{1/2} \left[ \frac{\mathbf{s}}{\mathbf{s}^T \mathbf{y}} - \frac{H \mathbf{y}}{\mathbf{y}^T H \mathbf{y}} \right],$$

## 参考文献

- [1] J. E. Dennis, Jr. and Jorge J. Moré. Quasi-Newton Methods, Motivation and Theory. *SIAM Review*, 19(1):46–89, 1977.
- [2] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–577, 1965.
- [3] M.J.D. Powell. A New Algorithm for Unconstrained Optimization. In *Nonlinear Programming*, pages 31–65. Elsevier, 1970.
- [4] William C. Davidon. VARIABLE METRIC METHOD FOR MINIMIZATION. *SIAM Journal on Optimization*, 1(1):1–17, 1959.
- [5] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *The Computer Journal*, 6(2):163–168, 1963.
- [6] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.
- [7] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 1970.
- [8] Donald Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109):23, 1970.
- [9] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111):10, 1970.
- [10] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.