拟牛顿法的本质思想是改善牛顿法每次需要求解复杂的Hessian矩阵的逆矩阵的缺陷，它使用正定矩阵来近似Hessian矩阵，从而简化了运算的复杂度。拟牛顿法和最速下降法一样只要求每一步迭代时直到目标函数的梯度。通过测量梯度的变化，构造一个目标函数的模型使之足以产生超线性收敛。这类方法大大优于最速下降法，尤其对于困难的问题。

另外，因为拟牛顿法不需要二阶导数的信息，所以有时候比牛顿法更为有效。

具体步骤：

拟牛顿法

(1) 第k次迭代 $x^{(k)}$.

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

其中，$g_k = g(x^{(k)}) = \nabla f(x^{(k)})$.

要求

$$\nabla f(x) = 0.$$

即：$\nabla f(x) = g_k + H_k(x-x^{(k)}) = 0$.

$x^k \to x^{k+1}$ 的更新：

$$x^{(k+1)} = x^{(k)} - H_k^{-1} g_k = x^{(k)} + p_k$$

(2) 在k+1次迭代：$x = x^{(k+1)}$

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

即：

$$g_{k+1} - g_k = H_k(x^{(k+1)} - x^{(k)}).$$ 拟牛顿条件.

(3) p_k 是下降方向.

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

f(x) 的泰勒展开式近似为：

$$f(x) = f(x^{(k)}) - \lambda g_k^T H_k^{-1} g_k$$

(4) 将 G_k 作为 H_k^{-1} 的近似.

G_k正定，且：

$$G_{k+1}(g_{k+1} - g_k) = x^{(k+1)} - x^{(k)}$$

更新：$G_{k+1} = G_k + \Delta G_k$.