牛顿法

要求函数二阶连续可导，则H先对x求导和先对y求导结果相同，因为二阶连续可导可推出求导顺序没有区别。

高斯牛顿法是专门用来计算非线性最小二乘问题的，是通过使用使用一阶泰勒近似，用JTJ近似Hessian矩阵。

牛顿法框架

$step0$：给定初始点$x_0, k=0$，终止限ε(默认0)

$stpe1$：计算$\nabla f(x_k), H_k$矩阵，即$Hessen$矩阵，判断$\nabla f(x_k)$

$step2$：置下降方向$d_k = -H_k^{-1}\nabla f(x_k)$

$step3$：步长$\alpha_k = 1$

$step4$：令$x_{k+1} = x_k + \alpha_k d_k, k = k+1$转向$step1$

- 牛顿法推导:
  将$f(x)$在x^k用**泰勒公式二阶**展开，得

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

  将上边公式对x**求导**可得（注意g_k^{T}与$H(x^{(k)})$都是常数项，自变量已经固定为$x^{(k)}$），得到$x=x_k$邻域内的$\nabla f(x)$的近似函数

$$\nabla f(x)=g_k+H_k\left(x-x^{(k)}\right) \qquad (2)$$

  由于$f(x)$ 在$\nabla f(x)=0$的地方得到极值点（假如为凸函数则为最优点），那么直接令$\nabla f(x)$的近似函数(2)等于0即可得到下一次迭代的$x^{(k+1)}$。
  令公式（2）等于0，可得牛顿法的**迭代公式**

$$x^{(k+1)}=x^{(k)}-H_k^{-1} g_k \qquad (3)$$

- 牛顿法的一步到位：**当$f(x)$的二阶导是常数时，那么对于$\nabla f(x)$的泰勒展开式就完全没有误差，可以一步到位了！**
  如当$f(x)=x^2$时，若$x^0=2, g_0=4, H_0^{-1}=0.5$，则$x_1=2-4\times0.5=0$，直接就迭代到了最优解$x=0$。

若记$H(x_0,y_0)=\begin{pmatrix} f''_{xx}(x_0,y_0) & f''_{xy}(x_0,y_0) \\ f''_{xy}(x_0,y_0) & f''_{yy}(x_0,y_0)\end{pmatrix}=\begin{pmatrix} A & B \\ B & C\end{pmatrix}$，我们可以用二次型的正定性将这个结论叙述为：

（1）如果$H(x_0,y_0)$为正定矩阵（$B^2-AC<0$且$A>0$或$C>0$），则$f(x_0,y_0)$为极小值；如果$H(x_0,y_0)$为负定矩阵（$B^2-AC<0$且$A<0$或$C<0$），则$f(x_0,y_0)$为极大值。

（2）如果$H(x_0,y_0)$为不定矩阵，则$f(x_0,y_0)$不是极值。

（3）如果不能判定$H(x_0,y_0)$为不定矩阵或半正定矩阵或半负定矩阵（$B^2-AC=0$），则$f(x_0,y_0)$是否为极值，需进一步讨论才能确定。

当曲线很平缓的时候，H会比较小从而H-1会较大导致加大牛顿法的迭代量。当H比较大而H-1比较小的时候曲线比较陡峭，这个时候要走得小心一点慢一点，牛顿法的迭代量也变小了。