1 牛顿迭代法

1.1 方程求根

首先，选择一个接近函数 $f(x)$ 零点的 x_0，计算相应的 $f(x_0)$ 和切线斜率 $f'(x_0)$（这里 f' 表示函数 f 的导数）。然后我们计算穿过点 $(x_0, f(x_0))$ 并且斜率为 $f'(x_0)$ 的直线和 x 轴的交点的 x 坐标，也就是求如下方程的解：

$f(x_0) = (x_0 - x) \cdot f'(x_0)$ 我们将新求得的点的 x 坐标命名为 x_1，通常 x_1 会比 x_0 更接近方程 $f(x) = 0$ 的解。因此我们现在可以利用 x_1 开始下一轮迭代。迭代公式可化简为如下所示：

$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ 已经证明，如果 f' 是连续的，并且待求的零点 x 是孤立的，那么在零点 x 周围存在一个区域，只要初始值 x_0 位于这个邻近区域内，那么牛顿法必定收敛。并且，如果 $f'(x)$ 不为 0, 那么牛顿法将具有平方收敛的性能. 粗略的说，这意味着每迭代一次，牛顿法结果的有效数字将增加一倍。

1.2 海森矩阵

在数学中，海森矩阵（Hessian matrix 或 Hessian）是一个自变量为向量的实值函数的二阶偏导数组成的方块矩阵，此函数如下：

$f(x_1, x_2, \ldots, x_n)$, 如果 f 所有的二阶导数都存在，那么 f 的海森矩阵即：

$H(f)_{ij}(x) = D_i D_j f(x)$ 其中 $x = (x_1, x_2, \ldots, x_n)$，即

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \\ \vdots & \vdots & \ddots & \vdots \\ \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

当函数 $f : \mathbb{R}^n \to \mathbb{R}$ 二阶连续可导时，Hessian 矩阵 H 在临界点 x_0 上是一个 $n \times n$ 阶的对称矩阵。

当 H 是正定矩阵时，临界点 x_0 是一个局部的极小值。

当 H 是负定矩阵时，临界点 x_0 是一个局部的极大值。

$H=0$, 需要更高阶的导数来帮助判断。

在其余情况下，临界点 x_0 不是局部极值。

1.3 函数最优化

给定二阶导数连续的函数 $f:\mathbb{R}^2\rightarrow\mathbb{R}$，在 $(x+\Delta x)$ 点按 Taylar 展开，得：

$f(x+\Delta x)\approx Q(x)=f(x)+\nabla f(x)^T\Delta x+\frac{1}{2}\Delta x^T H(x)\Delta x$

其中 $\Delta x=x_{n+1}-x_n$，$\nabla f(x)$ 是 $f(x)$ 的梯度，$H(x)$ 是 Hessian 矩阵。

令 $\nabla Q(x)=\nabla f(x)+H(x)\Delta x=0$

得到 $H(x)(x_{n+1}-x_n)=-\nabla f(x)$

所以，n 元函数最优化的牛顿迭代公式为：$x_{n+1}=x_n-H(x)^{-1}\nabla f(x)$