牛顿法

前言

暂时没啥好说的。

正文

牛顿法

由泰勒公式可知，$f(x)$在$x=x_0$出可以近似成如下形式

$$f(x) \approx f(x_0) + f^{'}(x_0)(x-x_0) + \frac{f^{''}(x_0)}{2!}(x-x_0)^2。 \tag{1}$$

假如$f(x)$为最小值点，则必有$f^{'}(x)=0$,因此有

$$0 = f^{'}(x) \approx f^{'}(x_0) + f^{''}(x_0)(x-x_0), \tag{2}$$

$$x \approx x_0 - \frac{f^{'}(x_0)}{f^{''}(x_0)}。 \tag{3}$$

所以可得迭代公式

$$x_{k+1} = x_k - \frac{f^{'}(x_k)}{f^{''}(x_k)}。 \tag{4}$$

从公式（4）到公式（3）该怎么理解呢？先假设x就是最终的最小值点，x_0是其附近的一点，因此可求出公式（3）。可是x并不是真正的最小点，对于真正的最小值点x的作用就相当于x_0相对于x的作用，即将公式（3）中x替换x_0继续计算，最总可抽象成公式（4）。
对于多元函数，其迭代公式为

$$x_{k+1} = x_k - H_k^{-1}J_k, \tag{5}$$

其中，H_k为海森矩阵，J_k为雅克比矩阵。 给公式（5）加入一个迭代步长因子λ就成了阻尼牛顿法

$$x_{k+1} = x_k - \lambda H_k^{-1}J_k。 \tag{6}$$

阻尼牛顿法可以避免牛顿法在待求解值处因收敛速度过快导致函数值增大、结果发散的问题。

拟牛顿法

牛顿法的迭代公式中含有海森矩阵，计算该参数计算量较大，因此就引出了拟牛顿法。拟牛顿法主要思路就是通过近似的迭代计算来减少每次计算。

$$D_k = \varphi(D_{k-1}) \tag{7}$$

$$x_{k+1} = x_k - \lambda D_k J_k \tag{8}$$

拟牛顿法主要的公式就入（7）（8）所示，对于步长的求解和算法的迭代步骤基本跟牛顿法一致。对于公式（7）具体的实现可见拟牛顿法 分析与推导。

PS

1.最优化问题中，牛顿法为什么比梯度下降法求解需要的迭代次数更少?