维基百科
自由的百科全书

拟牛顿法

维基百科，自由的百科全书

拟牛顿法是一种以牛顿法为基础设计的，求解非线性方程组或连续的最优化问题函数的零点或极大、极小值的算法。当牛顿法中所要求计算的雅可比矩阵或Hessian矩阵难以甚至无法计算时，拟牛顿法便可派上用场。

搜索极值

与牛顿法相同，拟牛顿法是用一个二次函数以近似目标函数$f(x)$. $f(x)$的二阶泰勒展开是

$$f(x_k+\Delta x)\approx f(x_k)+\nabla f(x_k)^T\Delta x+\frac{1}{2}\Delta x^T B\Delta x.$$

其中，∇f表示$f(x)$的梯度，B表示Hessian矩阵$\mathbf{H}[f(x)]$的近似. 梯度∇f可进一步近似为下列形式

$$\nabla f(x_k+\Delta x)\approx \nabla f(x_k)+B\Delta x.$$

令上式等于$\mathbf{0}$，计算出**Newton步长**Δx，

$$\Delta x=-B^{-1}\nabla f(x_k).$$

然后构造$\mathbf{H}[f(x)]$的近似B满足

$$\nabla f(x_k+\Delta x)=\nabla f(x_k)+B\Delta x.$$

上式称作**割线方程组**. 但当$f(x)$是定义在多维空间上的函数时，从该式计算B将成为一个*不定问题*（未知数个数比方程式个数多）. 此时，构造B，根据**Newton步长**更新当前解的处理需要回归到求解割线方程. 几乎不同的拟牛顿法就有不同的选择割线方程的方法. 而大多数的方法都假定B具有对称性（即满足$B=B^{\mathrm{T}}$）. 另外，下表所示的方法可用于求解B_{k+1}；在此，B_{k+1}于某些范数与B_k尽量接近. 即对于某些正定矩阵V，以以下方式更新B：

$$B_{k+1}=\arg\min_B \|B-B_k\|_V.$$

近似Hessian矩阵一般以单位矩阵等作为初期值[1]. 最优化问题的解x_k由根据近似所得的B_k计算出的**Newton步长**更新得出.

以下为该算法的总结：

- $\Delta x_k=-\alpha B_k^{-1}\nabla f(x_k)$
- $x_{k+1}=x_k+\Delta x_k$
- 计算新一个迭代点下的梯度$\nabla f(x_{k+1})$
- 令$y_k=\nabla f(x_{k+1})-\nabla f(x_k)$
- 利用y_k，直接近似Hessian矩阵的逆矩阵B_{k+1}^{-1}. 近似的方法如下表：

Method	$B_{k+1} =$	$H_{k+1} = B_{k+1}^{-1} =$
DFP法	$\left(I - \frac{y_k\,\Delta x_k^T}{y_k^T\,\Delta x_k}\right) B_k \left(I - \frac{\Delta x_k y_k^T}{y_k^T\,\Delta x_k}\right) + \frac{y_k y_k^T}{y_k^T\,\Delta x_k}$	$H_k + \frac{\Delta x_k \Delta x_k^T}{y_k^T\,\Delta x_k} - \frac{H_k y_k y_k^T H_k^T}{y_k^T H_k y_k}$
BFGS法	$B_k + \frac{y_k y_k^T}{y_k^T \Delta x_k} - \frac{B_k \Delta x_k (B_k \Delta x_k)^T}{\Delta x_k^T B_k\,\Delta x_k}$	$\left(I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k}\right)^T H_k \left(I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k}\right) + \frac{\Delta x_k \Delta x_k^T}{y_k^T\,\Delta x_k}$
Broyden法	$B_k + \frac{y_k - B_k \Delta x_k}{\Delta x_k^T\,\Delta x_k}\,\Delta x_k^T$	$H_k + \frac{(\Delta x_k - H_k y_k)\Delta x_k^T H_k}{\Delta x_k^T H_k\,y_k}$
Broyden族	$(1-\varphi_k)B_{k+1}^{BFGS} + \varphi_k B_{k+1}^{DFP}, \qquad \varphi \in [0,1]$	
SR1法	$B_k + \frac{(y_k - B_k\,\Delta x_k)(y_k - B_k\,\Delta x_k)^T}{(y_k - B_k\,\Delta x_k)^T\,\Delta x_k}$	$H_k + \frac{(\Delta x_k - H_k y_k)(\Delta x_k - H_k y_k)^T}{(\Delta x_k - H_k y_k)^T y_k}$

与逆矩阵的关联

若f是一个凸二次函数，且Hessian矩阵B正定，我们总是希望由拟牛顿法生成的矩阵H_k收敛于Hessian矩阵的逆$H = B^{-1}$。这是基于迭代值更新最小 (least-change update) 的拟牛顿法系列的一个实例。[2]

实现

拟牛顿法是现在普遍使用的一种最优化算法,存在多种编程语言的实现方法。

参见

- 牛顿法
- 应用于最优化的牛顿法

参考文献

1. William H. Press. Numerical Recepes. Cambridge Press. 2007: 521–526. ISBN 978-0-521-88068-8.
2. Robert Mansel Gower; Peter Richtarik. Randomized Quasi-Newton Updates are Linearly Convergent Matrix Inversion Algorithms. 2015. arXiv:1602.01768 [math.NA].

-