BFGS算法

quandy2020@126.com

拟牛顿算法

牛顿法

牛顿法（经典牛顿法）的迭代表达式：

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

但是，牛顿法过程中 Hessian 矩阵 $\nabla^2 f(x^k)^{-1}$ 的计算和存储的代价很高，对于条件数较多的问题很难求解。因此，引入 **拟牛顿法**。

拟牛顿法

拟牛顿法 的核心思路在于，在牛顿法的迭代过程中，用 **近似解** 计算第 k次迭代下的 Hessian 矩阵 $\nabla^2 f(x^k)$，近似值记为B^k，即有 $B^k \approx \nabla^2 f(x^k)$，称为 **拟牛顿矩阵**。

用 **近似值**B^k 代替牛顿法中的 $\nabla^2 f(x^k)$，得：

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

在近似 Hessian 矩阵时，也需要通过 **某种映射关系** 并 **不断迭代** 得到。但是依然需要求近似矩阵的逆，为了避免计算逆矩阵的开销，我们可以 **直接近似** Hessian **矩阵的逆**，记$H^k = (B^k)^{-1}$。故我们有：

$$x^{k+1} = x^k - H^k \nabla f(x^k)$$
$$H^{k+1} = g(H^k)$$

其中 g为 **近似** Hessian **矩阵的逆** 的映射函数。一般有 $H^{k+1} = H^k + C^k$，其中 C^k被称为 **修正矩阵**。

拟牛顿法基本过程

拟牛顿法：

- 令 $H^0 = I$，任选初始点 $x^0 \in \mathbb{R}^n$，令$k = 0$
- 计算 **梯度** $\nabla f(x^k)$，如果满足终止条件 $||\nabla f(x^k)|| < \epsilon$，取 $x^* = x^k$，并结束整个算法
- 计算 **搜索方向** $d^k = -H^k \nabla f(x^k)$，H^k为当前 x^k处的Hessian 矩阵的近似
- 迭代更新x：$x^{k+1} = x^k + d^k$
- 更新 H：$H^{k+1} = g(H^k)$根据 x^k点的信息进行简单修正

拟牛顿法H^k的确定

设 $f(x)$是二阶连续可微函数，对 $\nabla f(x)$ 在点 x^{k+1} 处进行一阶泰勒近似，得：

$$\nabla f(x) = \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1}) + O(||x - x^{k+1}||^2)$$

令 $x = x^k$，设 $s^k = x^{k+1} - x^k$为 **点差**，$y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$为 **梯度差**，得：

$$\nabla^2 f(x^{k+1}) s^k + O(||s^k||^2) = y^k$$

忽略高阶项$O(||s^k||^2)$，由此可以得到：

$$\nabla^2 f(x^{k+1})s^k = y^k$$

所以，我们希望 **近似** Hessian **矩阵**B^{k+1} 满足方程：

$$B^{k+1}s^k = y^k$$

因此 **近似** Hessian **矩阵的逆** H^{k+1}满足：

$$H^{k+1}y^k = s^k$$

上述的两个方程被称为 **割线方程**。

SR1方法

SR1 定义

SR1 方法 （秩一更新 Symmetric Rank-One）的核心思路很简单，即 根据 x^k处的信息得到修正量 ΔH^k来更新 H^k，即：

$$H^{k+1} = H^k + \Delta H^k$$

我们希望 $H^k \approx \nabla^2 f(x^k)^{-1}$，$H^{k+1} \approx \nabla^2 f(x^{k+1})^{-1}$故有：

$$\Delta H^k \approx \nabla^2 f(x^{k+1})^{-1} - \nabla^2 f(x^k)^{-1}$$

需要保证H^k和 H^{k+1}都是对称的，故显然 ΔH^k也是对称的。所以令 $\beta \in \mathbb{R}^n$, $u \in \mathbb{R}^n$，使得 $\Delta H^k = \beta\mu\mu^T$，故 H 的迭代更新表达式为：

$$H^{k+1} = H^k + \beta\mu\mu^T$$

显然 $\beta\mu\mu^T$ 是一个 $n \times n$ 的 **对称矩阵**。β是待定的标量，μ是待定的向量。

SR1 更新公式

根据 **割线方程**$H^{k+1}y^k = s^k$，代入 SR1 更新的结果，得到：

$$(H^k + \beta\mu\mu^T)y^k = s^k$$

整理可得：

$$\beta\mu\mu^T y^k = (\beta\mu^T y^k)\mu = s^k - H^k y^k$$

其中可以得出$\beta\mu^T y^k$是一个 **标量**，因此上式表明 **向量** μ和$s^k - H^k y^k$**同向**。故有：

$$\mu = \frac{1}{\beta\mu^T y^k}(s^k - H^k y^k)$$

记 $\frac{1}{\beta\mu^T y^k} = \gamma$，得：

$$\mu = \gamma(s^k - H^k y^k)$$

将 μ回代到 $\beta\mu\mu^T y^k = s^k - H^k y^k$，得：

$$s^k - H^k y^k = \beta\gamma^2(s^k - H^k y^k)(s^k - H^k y^k)^T y^k$$

由于 $\beta\gamma^2$ 和 $(s^k - H^k y^k)^T y^k$ 都是 **标量**，上式可以写成：

$$s^k - H^k y^k = [\beta\gamma^2(s^k - H^k y^k)^T y^k](s^k - H^k y^k)$$

显然只有在 $\beta\gamma^2(s^k - H^k y^k)^T y^k = 1$时，等式成立。

因此，我们可以得到：

$$\beta\gamma^2 = \frac{1}{(s^k - H^k y^k)^T y^k}$$

将上式 $\beta\gamma^2$回代到 **迭代更新表达式**$H^{k+1} = H^k + \beta\mu\mu^T$：

$$\begin{aligned} H^{k+1} &= H^k + \beta\mu\mu^T \\ &= H^k + \beta\gamma^2(s^k - H^k y^k)(s^k - H^k y^k)^T \\ &= H^k + \frac{\beta\gamma^2(s^k - H^k y^k)(s^k - H^k y^k)^T}{(s^k - H^k y^k)^T y^k} \end{aligned}$$

记 $v = s^k - H^k y^k$，那么上述更新表达式可以化简为：

$$H^{k+1} = H^k + \frac{vv^T}{v^T y^k}$$

由此得到了最终 SR1 **方法** 的 **更新公式**。

SR1 的缺点

- 在迭代过程中 无法保证B^k正定，也就是说 **搜索方向不一定下降**。而且即使B^k**正定**，也 **不一定保证** B^{k+1}
- **无法保证** $v^T y^k$**恒大于 0**，因此也可能会导致后续的 B^{k+1}**非正定**

BFGS 方法

BFGS 定义

BFGS方法考虑的是 对 B^k进行秩二更新。对于拟牛顿矩阵 $B^k \in \mathbb{R}^{n\times n}$，设 $\mu \neq 0, \nu \neq 0, \mu, \nu \in \mathbb{R}^n$ 以及 $a, b \in \mathbb{R}$，其中设定的向量和标量都是待定的，则有 **秩二更新表达式**：

$$B^{k+1} = B^k + a\mu\mu^T + b\nu\nu^T$$

显然 $a\mu\mu^T$ **和** $b\nu\nu^T$**都是对称的**。

BFGS 更新公式

根据 **割线方程** $B^{k+1}s^k = y^k$，代入 **待定参量**，得：

$$B^{k+1} = (B^k + a\mu\mu^T + b\nu\nu^T)s^k = y^k$$

整理可得：

$$a\mu\mu^T s^k + b\nu\nu^T s^k = (a\mu^T s^k)\mu + (b\nu^T s^k)\nu = y^k - B^k s^k$$

可以得出 $a\mu^T s^k$ 和 $b\nu^T s^k$为 **标量**，不妨取$(a\mu^T s^k)\mu = y^k, (b\nu^T s^k)\nu = -B^k s^k$，所以可以得到如下取值

$$a\mu^T s^k = a, \mu = y^k, b\nu^T s^k = -1, \nu = B^k s^k$$

化简可得所有 **待定参量的取值**：

$$a = \frac{1}{\mu^T s^k} = \frac{1}{(y^k)^T s^k}$$

$$b = -\frac{1}{\nu^T s^k} = -\frac{1}{(B^k s^k)^T s^k} = \frac{1}{(s^k)^T B^k s^k}$$

将上述取值回代到 **更新表达式** $B^{k+1} = B^k + a\mu\mu^T + b\nu\nu^T$，得:

$$B^{k+1} = B^k + \frac{y^k (y^k)^T}{(y^k)^T s^k} - \frac{B^k s^k (s^k)^T B^k}{(s^k)^T B^k s^k}$$

参考

- https://www.cnblogs.com/MAKISE004/p/17904431.html
- https://zhuanlan.zhihu.com/p/144736223
- https://www.cnblogs.com/MAKISE004/p/17904431.html