一、Newton 法

1. 问题

$$\min_{x\in R^n} f(x)$$

$f(x)$是R^n上二次连续可微函数，

即$f(x)\in C^2(R^n)$。

2. 算法思想

$$x^{(0)} \to x^{(1)} \to \cdots \to x^{(k)} \to x^{(k+1)} \to \cdots$$

为了由$x^{(k)}$产生$x^{(k+1)}$，用二次函数$Q(x)$近似$f(x)$，并用$Q(x)$的极值点近似 $f(x)$的极值点。

$$f(x) \approx Q(x) = f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)})$$

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

$$= f(x^{(k)}) + g_k^T \cdot (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T G_k (x - x^{(k)})$$

其中 $g_k = \nabla f(x^{(k)})^T, G_k = \nabla^2 f(x^{(k)})$。

令 $\nabla Q(x)=g_k+G_k(x-x^{(k)})=0$

若 *Hesse* 矩阵 G_k 正定，即 $G_k>0$，则 $G_k^{-1}>0$，此时有

$$x^{(k+1)}=x^{(k)}-G_k^{-1}g_k$$

这就是 *Newton* 迭代公式 。

比较迭代公式 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$，有

$$d^{(k)}=-G_k^{-1}g_k\ ,$$

$$\lambda_k=1\ 。$$

3. 算法步骤

$step1.$ 给定初始点 $x^{(0)}$，精度 $\varepsilon > 0, k := 0$

$step2.$ 计算$g_k = \nabla f(x^{(k)})$和$G_k = \nabla^2 f(x^{(k)})$

当 G_k 可逆时，$x^{(k+1)} = x^{(k)} - G_k^{-1} g_k$ 。

$step3.$ 由方程组$\nabla Q(x) = g_k + G_k(x - x^{(k)}) = 0$ 解出$x^{(k+1)}$。

$step4.$ 若 $\|\nabla f(x^{(k+1)})\| < \varepsilon$，停止，$x^* = x^{(k+1)}$；

否则，令$k := k+1$，转$step\,2$。

例 用牛顿法求解下列问题

$$\min \quad f(x)=2x_1^2+4x_2^2-2x_1x_2$$

初始点为 $x^{(0)}=(2,1)^T$ 。

解： $\nabla f(x)=[4x_1-2x_2, 8x_2-2x_1]^T$

$$\therefore \quad g_0=\nabla f(x^{(0)})=[6,4]^T$$

$$\because G_0=\nabla^2 f(x)=\begin{bmatrix} 4 & -2 \\ -2 & 8 \end{bmatrix} \qquad \therefore \quad G_0^{-1}=\begin{bmatrix} \frac{2}{7} & \frac{1}{14} \\ \frac{1}{14} & \frac{1}{7} \end{bmatrix}$$

$$\therefore \quad x^{(1)}=x^{(0)}-G_0^{-1}g_0$$

$$=(2,1)^T-\begin{bmatrix} \frac{2}{7} & \frac{1}{14} \\ \frac{1}{14} & \frac{1}{7} \end{bmatrix}\begin{bmatrix} 6 \\ 4 \end{bmatrix}=(0,0)^T$$

4. 算法特点

- 收敛速度快，为二阶收敛。
- 初始点要选在极值点附近。

5. 存在缺点及修正

(1) $f(x^{(k+1)}) < f(x^{(k)})$?

（2） 初始点的选取困难，甚至无法实施。

(3) G_k^{-1}的存在性和计算量问题 。

问题一：　如何使得 $f(x^{(k+1)}) < f(x^{(k)})$ ？

在 Newton 法中，有

$$x^{(k+1)} = x^{(k)} - G_k^{-1} g_k = x^{(k)} + d^{(k)}$$

当 $G_k > 0$ 时，有 $\nabla f(x^{(k)})^T d^k = -\nabla f(x^{(k)})^T G_k^{-1} g_k$

$$= -g_k^T G_k^{-1} g_k < 0,$$

$\therefore$ 当 $G_k > 0$ 时，$d^{(k)}$ 是下降方向。

阻尼牛顿法：

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$

$$\lambda_k : f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda} f(x^{(k)} + \lambda d^{(k)})$$

则有：$f(x^{(k+1)}) < f(x^{(k)})$。

注：阻尼牛顿法有全局收敛性，即收敛性和初始点的选取无关。

问题二：如何克服缺点 (3)？

二、拟 *Newton* 算法　(变尺度法)

1. 先考虑 *Newton* 迭代公式:

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

在 *Newton* 迭代公式中，如果我们 用正定矩阵 H_k 替代 G_k^{-1}，则有:

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

2. 考虑更一般的形式：

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

$H_k \equiv I$时 $\Rightarrow$ 梯度法

$H_k = G_k^{-1}$时 $\Rightarrow$ *Newton* 法

H_K近似G_k^{-1} $\Rightarrow$ 拟*Newton*算法

3. 如何对 H_k 附加某些条件使得：

（1）迭代公式具有下降性 质 $\leftrightarrow$ $H_k > 0$

（2）H_k 的计算量要小 $\leftrightarrow$ $H_{k+1} = H_k + \Delta H_k$

$(\Delta H_k = H_{k+1} - H_k)$

（3）收敛速度要快 $\leftrightarrow$ $H_k \approx G_k^{-1}$

如何保证 $H_k > 0$ 和 $H_k \approx G_k^{-1}$?

如何确定 ΔH_k?

拟$Newton$条件

拟$Newton$条件 $\leftrightarrow H_k \approx G_k^{-1}$

分析: G_k^{-1} 需满足的条件，并利用 此条件确定 H_k。

记$g(x) = \nabla f(x)$, $g_k = \nabla f(x^{(k)})$, $G_k = \nabla f^2(x^{(k)})$, 则

$$f(x) \approx f(x^{(k+1)}) + \nabla f(x^{(k+1)})^T (x - x^{(k+1)})$$

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

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

代入 $x = x^{(k)}$ 可得

$$g_k \approx g_{k+1} + G_{k+1}(x^{(k)} - x^{(k+1)})$$

则有 $G_{k+1}^{-1}(g_{k+1}-g_k)=x^{(k+1)}-x^{(k)}$。

由上式构造条件:

$$H_{k+1}(g_{k+1}-g_k)=x^{(k+1)}-x^{(k)}$$

记$y_k=g_{k+1}-g_k, s_k=x^{(k+1)}-x^{(k)}$,则有

$H_{k+1}y_k=s_k \Rightarrow$ 拟 *Newton* 条件或 拟 *Newton* 方程。

4、拟 *Newton* 算法 (变尺度法)的一般步骤；

Step 1. 给定初始点 $x^{(0)}$, 正定矩阵 H_0, 精度 $\varepsilon > 0$,

$k := 0$。

step 2. 计算搜索方向 $d^{(k)} = -H_k \nabla f(x^{(k)})$;

step 3. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$,

其中 $f(x^{(k)} + \lambda_k d^{(k)}) = \min f(x^{(k)} + \lambda d^{(k)})$。

*Step*4. 判断 $x^{(k+1)}$ 是否满足终止准则:

yes: 计算 stop, $x^* := x^{(k+1)}$。

No : 转step 5 。

step 5. 令 $g_{k+1}=\nabla f(x^{(k+1)})$, $g_k=\nabla f(x^{(k)})$,

$y_k=\nabla f(x^{(k+1)})-\nabla f(x^{(k)})=g_{k+1}-g_k$,

$s_k=x^{(k+1)}-x^{(k)}$。

按照校正公式 $H_{k+1}=H_k+\Delta H_k$,

计算H_{k+1}使得H_{k+1}满足 拟*Newton* 条件

或拟*Newton* 方程： $H_{k+1}y_k=s_k$。

令 $k=:k+1$, 转*step* 2.

H_k 的确定？

三、*DFP*算法

1. *DFP*算法的提出：

（1） 1959年*Davidon* 首次提出；

（2） 1963年*Fletcher*和*Powell*做了改进；

（3） 多变量无约束优化问题 的一个重要算法。

2. 如何确定 H_k?

$$H_{k+1} = H_k + \Delta H_k$$
$$= H_k + \alpha_k u_k u_k^T + \beta_k v_k v_k^T$$

待定： α_k， $\beta_k \in R, u_k$， $v_k \in R^n$

根据 拟*Newton*条件： $H_{k+1}y_k = s_k$，有

$$(H_k + \alpha_k u_k u_k^T + \beta_k v_k v_k^T)y_k = s_k$$

即： $\alpha_k u_k u_k^T y_k + \beta_k v_k v_k^T y_k = s_k - H_k y_k$

满足上述方程的解很多，可以如下确定一组解：

令 $$\alpha_k u_k u_k^T y_k = s_k\,,$$

$$\beta_k v_k v_k^T y_k = -H_k y_k\,,$$

则可以取：

$$u_k = s_k\,,\qquad \alpha_k u_k^T y_k = 1,$$

$$v_k = H_k y_k\,,\quad \beta_k v_k^T y_k = -1。$$

即：　$u_k = s_k$，　$\alpha_k = \dfrac{1}{s_k^T y_k}$，

$v_k = H_k y_k$，　$\beta_k = -\dfrac{1}{y_k^T H_k y_k}$。

根据上述推导，可得 H_k的DFP的校正公式：

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

$-$ DFP校正公式

性质：　若$H_0 > 0$，则$\forall k$，有 $H_k > 0$。

3、*DFP*算法的步骤；

将变尺度法的第5步改为：

step 5.　按照*DFP*的校正公式：

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

计算 H_k, $k := k+1$, 转 *step* 2.

例. 请用***DFP***算法求解 $\min f(x)=x_1^2+4x_2^2$, 初始点 $x^{(0)}=\begin{pmatrix}1\\1\end{pmatrix}$.

解: 取 $H_0=I$, $\nabla f(x)=\begin{pmatrix}2x_1\\8x_2\end{pmatrix}$。

第一步***DFP***算法与梯度法相同,

$$\because \quad x^{(0)}-\lambda\nabla f(x^{(0)})=\begin{pmatrix}1-2\lambda\\1-8\lambda\end{pmatrix},$$

$$f(x^{(0)}-\lambda_0\nabla f(x^{(0)}))=\min f(x^{(0)}-\lambda\nabla f(x^{(0)}))$$

$$=\min \quad (1-2\lambda)^2+4(1-8\lambda)^2$$

解得 $\lambda_0=0.13$, 所以 $x^{(1)}=\begin{pmatrix}0.73846\\-0.04616\end{pmatrix}$。

$$\because s_0 = x^{(1)} - x^{(0)} = \begin{pmatrix} -0.26154 \\ -1.04616 \end{pmatrix},$$

$$y_0 = \nabla f(x^{(1)}) - \nabla f(x^{(0)}) = g_1 - g_0 = \begin{pmatrix} -0.52308 \\ -8.36923 \end{pmatrix}。$$

按照DFP的校正公式：

$$H_1 = H_0 + \frac{s_0^T s_0}{s_0^T y_0} - \frac{H_0 y_0 y_0^T H_0}{y_0^T H_0 y_0} = \begin{pmatrix} 1.00380 & -0.03149 \\ -0.03149 & 0.12697 \end{pmatrix}。$$

$$\therefore \text{搜索方向} \ d^{(1)} = -H_1 \nabla f(x^{(1)}) = \begin{pmatrix} -1.49416 \\ 0.09340 \end{pmatrix}。$$

求解 $f(x^{(1)}+\lambda_1 d^{(1)})=\min\limits_{\lambda} f(x^{(1)}+\lambda d^{(1)})$,

解得 $\lambda_1=0.49423$。

$$\therefore x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=x^{(1)}+0.49423 d^{(1)}=\begin{pmatrix}0.0000\\0.0000\end{pmatrix}。$$

因为$\nabla f(x^{(2)})=0$，所以$x^{(2)}$是极小点。