$$x^{k+1} = x^k + \alpha^k d^k \qquad d^k = -\nabla f(x^k)$$

$$d^k = x^{k+1} - x^k$$

$$f(x^k + d^k) = f(x^k) + \langle \nabla f(x^k), d^k \rangle + o(\cdot\cdot)$$

$$\Rightarrow d^k = -\nabla f(x^k)$$

牛顿法

二阶展开：$f(x^k + d^k) = f(x^k) + \langle \nabla f(x^k), d^k \rangle + \frac{1}{2} \langle d^k, \nabla^2 f(x^k) d^k \rangle + o(\cdot\cdot)$

牛顿方程：$\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$

牛顿方向：$d^k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$

经典牛顿方法：$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$

步长：经典方法：$\alpha_k = 1$

也可以使用线搜索步长

牛顿法的收敛性

梯度下降法：Q-线性收敛

牛顿法　：Q-二次收敛

$|\nabla^2 f(y) - \nabla^2 f(x)| \le L_3 |y-x|$

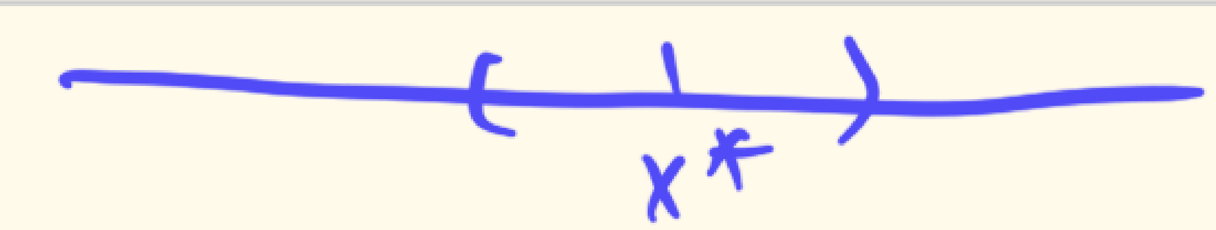
$|\nabla f(y) - \nabla f(y)| \le L_2 |y-x|$

$|f(y) - f(x)| \le L_1 |y-x|$

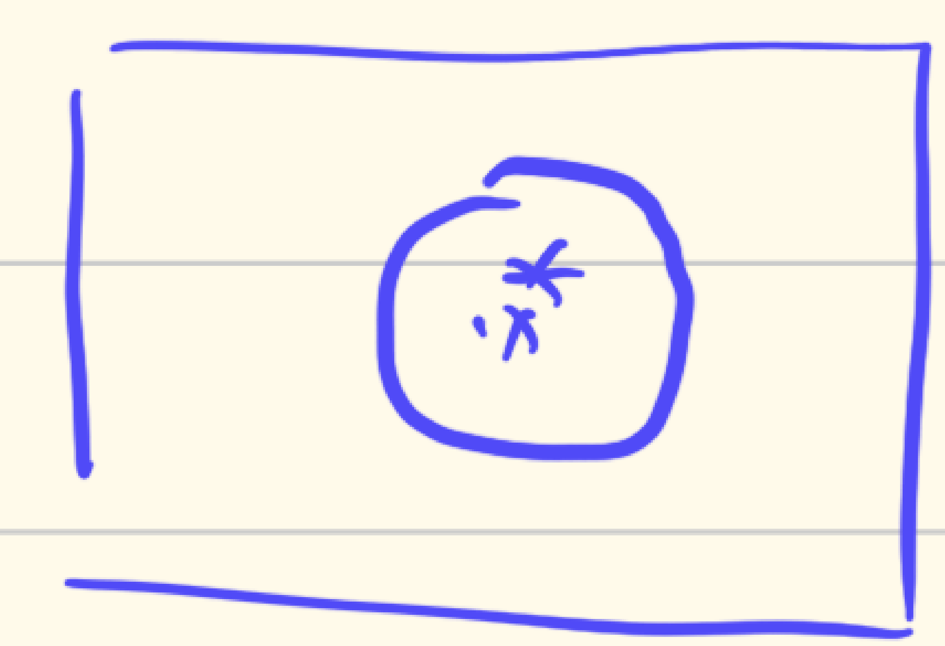
Thm：设 f 是二阶连续可微的函数，且 Hessian 矩阵在最优值点 x^* 的一个邻域 $N_\delta(x^*)$ 是 Lipschitz 连续的.

$N_\delta(x^*) = \{x \mid |x-x^*| < \delta\}$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L\|x-y\|,\ \forall x,y \in N_\delta(x^*)$$

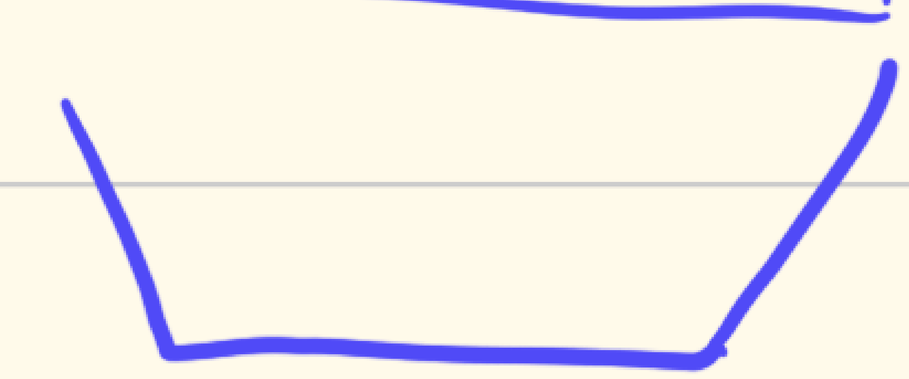


如果 $f(x)$ 在 x^* 满足 $\nabla f(x^*)=0,\ \nabla^2 f(x^*) \succ 0$



(1) 如果初始点 x^0 距离 x^* 足够近，则牛顿法产生的迭代点列 $\{x^k\}$ 收敛到 x^*.

(2) $\{x^k\}$ 收敛到 x^* 是 Q-二次的



(3) $\{\|\nabla f(x^k)\|\}$ Q-二次收敛到 0.

Proof: 由于是最优值点 x^*，有 $\nabla f(x^*)=0$

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

$$=[\nabla^2 f(x^k)]^{-1}\left(\nabla^2 f(x^k)(x^k-x^*)-(\nabla f(x^k)-\nabla f(x^*))\right)$$

$$\nabla f(x^k)-\nabla f(x^*) = -\nabla f(x^k+t(x^*-x^k))\Big|_0^1$$

$$= \int_0^1 \nabla^2 f(x^k+t(x^*-x^k))(x^k-x^*)\,dt$$

范数的性质：$\left\|\int_a^b x(t)dt\right\| \le \int_a^b \|x(t)\|dt$

$$\|\nabla^2 f(x^k)(x^k-x^*)-(\nabla f(x^k)-\nabla f(x^*))\|$$

$$\left\|\int_0^1[\nabla^2 f(x^k+t(x^*-x^k))-\nabla^2 f(x^k)](x^k-x^*)dt\right\|$$

$$\le \int_0^1 \|(\nabla^2 f(x^k+t(x^*-x^k))-\nabla^2 f(x^k))(x^k-x^*)\|\,dt$$

$$\le \int_0^1 \|\nabla^2 f(x^k+t(x^*-x^k))-\nabla^2 f(x^k)\|\,\|x^k-x^*\|\,dt$$

$$\le \int_0^1 L\|x^k+t(x^*-x^k)-x^k\|dt\cdot\|x^k-x^*\|$$

$\leq \int_0^1 L\|x^k - x^*\| t\,dt\, \|x^k - x^*\|$

$= \frac{L}{2}\|x^k - x^*\|^2$ $\qquad \|x^{k+1} - x^*\| \leq C\|x^k - x^*\|^2$

$\nabla^2 f(x)$ 在 $N_\delta(x^*)$ 是 Lipschitz 连续

Lemma：若 f 是二阶连续可微，$\nabla^2 f(x^*)$ 是非奇异，则存在 r，使得对任意

$\|x - x^*\| \leq r$，有 $\|\nabla^2 f(x)^{-1}\| \leq 2\|\nabla^2 f(x^*)^{-1}\|$

Proof：令 $A = \nabla^2 f(x^*)$，$E = \nabla^2 f(x) - \nabla^2 f(x^*)$，当 $\|x - x^*\| \leq r$ 时，

$\|E\| = \|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq L\|x - x^*\| \leq Lr$

令 r 满足 $Lr \leq \frac{1}{2\|A^{-1}\|}$，根据矩阵逆的扰动公式

$\|(A+E)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|} \leq \frac{\|A^{-1}\|}{1 - \frac{1}{2}} = 2\|A^{-1}\|$

$A + E = \nabla^2 f(x)$ $\quad A = \nabla^2 f(x^*)$ $\quad \Rightarrow \|\nabla^2 f(x)^{-1}\| \leq 2\|\nabla f(x^*)^{-1}\|$

$\|x^{k+1} - x^*\| \leq \|\nabla^2 f(x^k)^{-1}\| \, \|\nabla^2 f(x^k)(x^k - x^*) - (\nabla f(x^k) - \nabla f(x^*))\|$

$\leq L\|\nabla^2 f(x^*)^{-1}\| \, \|x^k - x^*\|^2$

因此，当初始点满足

$$\|x^0 - x^*\| \le \min\left\{\delta,\ r,\ \frac{1}{2L\|\nabla^2 f(x^*)^{-1}\|}\right\}$$

$\{x^k\}$ 是 Q-二次收敛到 x^*.

$\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$

$$\begin{aligned}
\|\nabla f(x^{k+1})\| &= \|\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^k) d^k\| \\
&= \left\|\int_0^1 \nabla f(x^k + t d^k) d^k dt - \int_0^1 \nabla^2 f(x^k) d^k dt\right\| \\
&\le \int_0^1 \|\nabla f(x^k + t d^k) - \nabla^2 f(x^k)\| \|d^k\| dt \\
&\le L\int_0^1 \|t d^k\| dt \, \|d^k\| \\
&= \frac{L}{2}\|d^k\|^2 \le \frac{L}{2}\|\nabla^2 f(x^k)^{-1}\|^2 \|\nabla f(x^k)\|^2 \\
&\le 2L\|\nabla^2 f(x^*)^{-1}\| \|\nabla f(x^k)\|^2
\end{aligned}$$

$d^k = [\nabla f(x^k)]^{-1} \nabla f(x^k)$

$\|\nabla f(x^{k+1})\| \le c' \|\nabla f(x^k)\|^2$

① 牛顿法的收敛速度不太会受到目标函数条件数的影响.

② 但条件数会影响牛顿法的收敛域

修正牛顿法.

困难：Hessian 矩阵的逆不容易计算，也不易储存.

$$\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$$

修正：令 $B^k = \nabla^2 f(x^k) + E^k$

使得 B^k 正定，且条件数较小，

再求解修正牛顿方程：$\nabla f(x^k) + B^k d^k = 0$

① 令 $E^k = \tau_k I$ 是一个常用方法，但需要选择合适的参数 τ_k

② 利用 cholesky 分解得到隐式的 E^k

Cholesky 分解：$A = LDL^T$

$D = \mathrm{diag}(d_1, d_2 \cdots, d_n)$

L：对角线元素为 1 的下三角矩阵

保证 d_i 不要太小即可

非精确牛顿法:

原始: $\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$

非精确: $\nabla f(x^k) + \nabla^2 f(x^k) d^k = r^k$ 残差

假设残差满足 $\|r^k\| \le \eta_k \|\nabla f(x^k)\|$ $\eta_k = O(\|\nabla f(x^k)\|)$

则非精确牛顿法是 Q-2次收敛的

共轭梯度法满足此条件

牛顿共轭梯度法:

1. 给定初始点 x_0
2. while $\|\nabla f(x^k)\| > \varepsilon$, do
3. 用共轭梯度法非精确求解 d^k: $\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$
4. 使用线搜索确定步长

5. 更新 $x^{k+1} = x^k + \alpha_k d^k$

6. End while

7. 输出 x^{k+1}