Newton's method：二阶近似 (局部收敛)

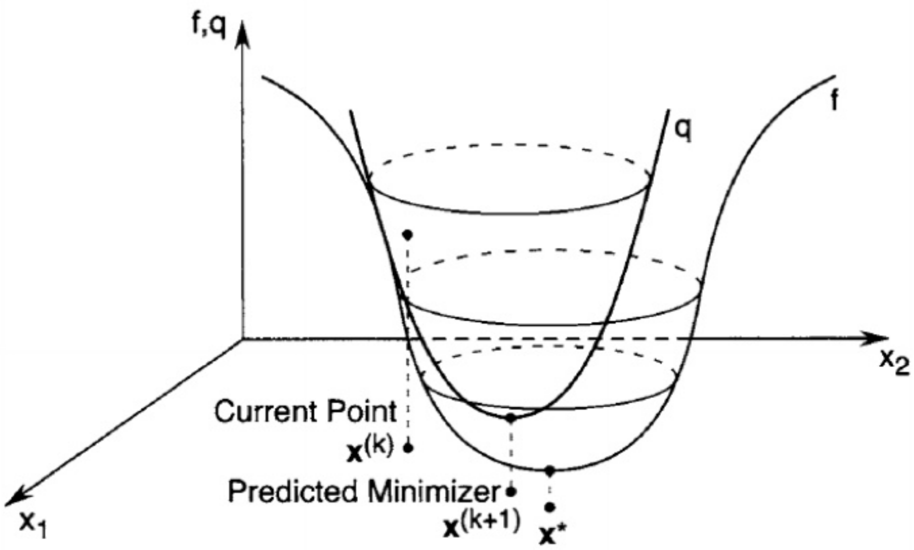
已知 x^k，objective Function：$f \in C^2$ twice differentiable.

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

$$令\ q'(x^k) = 0 \Leftrightarrow \nabla f(x^k) + \nabla^2 f(x^k)^T (x^{k+1}-x^k) = 0$$

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

前提是 $\nabla^2 f(x^k)$ 可逆.



对于二次问题，newton method 一步到达 x^*

eq. min：$\frac{1}{2}x^TQx - bx$ $\qquad x^* = Q^{-1}b$

for a random start point. x_0

$$x_1 = x_0 - \nabla^2 f(x^0)^{-1} \nabla f(x_0)$$
$$= x_0 - Q^{-1}(Qx_0 - b)$$
$$= Q^{-1}b$$
$$= x^*$$

一个重要的等式:

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla_2 f(x+t(y-x))(y-x)\,dt.$$

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

$$x^{k+1} - x^* = x^k - x^* - \nabla_2 f(x^k)^{-1}\nabla f(x^k)$$

$$= \nabla_2 f(x^k)^{-1}\left(\nabla_2 f(x^k)(x^k - x^*) - (\nabla f(x^k) - \nabla f(x^*))\right)$$

$$= \nabla_2 f(x^k)^{-1}\left(\nabla_2 f(x^k)(x^k - x^*) - \int_0^1 \nabla_2 f(x^* + t(x^k - x^*))(x^k - x^*)\,dt\right)$$

$$= \nabla_2 f(x^k)^{-1}\int_0^1 (\nabla_2 f(x^k) - \nabla_2 f(x^* + t(x^k - x^*)))\,dt\,(x^k - x^*)$$

$$\|x^{k+1} - x^*\| \le \|\nabla_2 f(x^k)^{-1}\| \cdot \int_0^1 \|\nabla_2 f(x^k) - \nabla_2 f(x^* + t(x^k - x^*))\|\,dt\,\|x^k - x^*\|$$

$$\therefore \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \le \|\nabla_2 f(x^k)^{-1}\| \cdot \int_0^1 \|\nabla_2 f(x^k) - \nabla_2 f(x^* + t(x^k - x^*))\| \cdot dt.$$

若 $\nabla_2 f(x)$ 满足 Lipschitz 连续, 即 $\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L\|x-y\|$ $\forall x, y$.

$$\therefore \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \le \|\nabla_2 f(x)^{-1}\| \cdot \int_0^1 L \cdot \|x^k - x^* - t(x^k - x^*)\| \cdot dt.$$

$$\le \|\nabla_2 f(x)^{-1}\| \cdot L\|x^k - x^*\| \cdot \int_0^1 (1-t)\,dt.$$

$$= \frac{L}{2}\|\nabla_2 f(x)^{-1}\| \cdot \|x^k - x^*\|.$$

注意到一阶梯度法本质上是个一阶近似加上个正则项.

即 $x^{k+1} := \arg\min f(x^k) + \nabla f(x^k)(x-x^k) + \underbrace{\frac{t}{2}\|x-x^k\|^2}$.

$$= x^k - \frac{1}{t}\nabla f(x^k)$$

保证了 x^{k+1} 与 x^k 不会太远, 从而有全局收敛.

同理的, 为改进 Newton method 的局部收敛的 drawback.

可类似改进

$$x^{k+1} := \arg\min f(x^k) + \nabla f(x^k)(x-x^k) + \frac{1}{2}(x-x^k)^T \nabla^2 f(x^k)(x-x^k) + \frac{t}{3}\|x-x^k\|^3$$

设 $f(x)$ 的在 x^k 处附近的 d 维近似为 $T_d(x; x^k)$

则可构造. $x_{k+1} := \arg\min T_d(x, x^k) + \frac{t}{d}\|x-x^k\|^d$.