Newton's Method

- search direction: $p_k^N = -(\nabla^2 f_k)^{-1} \cdot \nabla f_k$ (1)
- $x_{k+1} = x_k + \underbrace{\alpha_k}_{=1} (\nabla^2 f_k)^{-1} \nabla f_k$
- Since the Hessian $\nabla^2 f_k$ may not always be p.d., p_k^N may not always be a descent direction
  (we may have: $\left(-(\nabla^2 f_k)^{-1} \cdot f_k\right)^T f_k \not< 0$)
  $\Rightarrow$ many of the ideas discussed so far no longer apply.
- We discuss <u>local</u> rate-of-convergence properties of Newton's method
- If $\nabla^2 f(x^*)$ is p.d., then $\nabla^2 f(x)$ will also be p.d. for all x in the vicinity of the solution x^*. Newton's method will be well-defined in this region, and will converge quadratically if $\alpha_k = 1 \ \forall k$.

<u>Theorem</u>: Suppose that f is twice differentiable and that the Hessian $\nabla^2 f(x)$ is Lipschitz-continuous ⊛ in a neighborhood of the solution x^* at which the following conditions are satisfied:

- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*)$ is p.d.

Consider the iteration $x_{k+1} = x_k + p_k^N$ with p_k^N given by (1). Then:

(i) If the starting point x_0 is sufficiently close to x^*, the sequence of iterates converge to x^*.

(ii) The rate of convergence of $\{x_k\}$ is quadratic.

(iii) the sequence of gradient norms $\{\|\nabla f_k\|\}$

converges quadratically to zero.

(*) $\nabla^2 f$ is Lipschitz-continuous on N if there exists a constant $L > 0$ s.t.

$$\underbrace{\| \nabla^2 f(x_1) - \nabla^2 f(x_2) \|}_{\text{Matrix norm: } \|A\| := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}} \leq L \| x_1 - x_2 \| \quad \forall x_1, x_2 \in$$

Proof: We have that

$$x_k + p_k^N - x^* \overset{(1)}{=} x_k - x^* - (\nabla^2 f_k)^{-1} f_k$$

$$= \underbrace{(\nabla^2 f_k)^{-1} [\nabla^2 f_k (x_k - x^*)}_{= x_k - x^*} \underbrace{- (\nabla f_k - \nabla f_*)]}_{= -(\nabla^2 f_k)^{-1} \nabla f_k + (\nabla^2 f_k)^{-1} \cdot \underbrace{\nabla f_*}_{=0}}$$

By Taylor's theorem, we have:

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

Therefore,

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

$$= \| \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k))] \cdot (x_k - x^*) \|$$

$$\leq \int_0^1 \| \nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k)) \| \cdot \| x_k - x^* \| \, dt$$

$$= \| x_k - x^* \| \cdot \int_0^1 \underbrace{\| \nabla^2 f(x_k) - \nabla f(x_k + t(x^* - x_k)) \|}_{\leq L \cdot \| x_k - (x_k + t(x^* - x_k)) \| = t \cdot (x^* - x_k)} \, dt$$

Lipschitz

$$\leq \| x_k - x^* \|^2 \cdot \int_0^1 L \cdot t \, dt$$

$$= \| x_k - x^* \|^2 \cdot L \cdot \left[\frac{1}{2} t^2 \right]_0^1 = \frac{1}{2} L \| x_k - x^* \|^2$$

since $\nabla^2 f(x^*)$ is nonsingular, there is a radius $r > 0$ s.t. $\|(\nabla^2 f_k)^{-1}\| \leq 2\|\nabla^2 f(x^*)^{-1}\|$ for all x_k with $\|x_k - x^*\| \leq r$.

Now we have:

$$\|x_k + p_k^N - x^*\| \leq \frac{1}{2} L \|x_k - x^*\| \cdot \|\nabla^2 f_k^{-1}\|$$

$$\leq L \cdot \|(\nabla^2 f_*)^{-1}\| \cdot \|x_k - x^*\|^2$$

$$= \tilde{L} \cdot \|x_k - x^*\|^2$$

$$\tilde{L} := L \cdot \|(\nabla^2 f(x^*))^{-1}\|$$

Choosing x_0 s.t. $\|x_0 - x^*\| \leq \min(r, \frac{1}{\tilde{L}})$, we can use this inequality inductively to deduce that the sequence converges to x^*, and the rate of convergence is quadratic.

Moreover, we have

$$\|\nabla f(x_{k+1})\| = \|\nabla f(x_{k+1}) \underbrace{- \nabla f_k - \nabla^2 f(x_k) \cdot p_k^N}_{=0, \text{ because } \nabla f_k + \nabla^2 f_k p_k^N = \nabla f_k + \nabla^2 f_k \cdot (-\nabla^2 f_k)^{-1} \cdot \nabla f_k}\|$$

$$= \| \int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k)) \cdot (\underbrace{x_{k+1} - x_k}_{p_k^N}) \, dt - \nabla^2 f(x_k) p_k^N \|$$

$$= \| \int_0^1 \nabla^2 f(x_k + t p_k^N)(\underbrace{x_{k+1} - x_k}_{= p_k^N}) \, dt - \nabla^2 f(x_k) p_k^N$$

$$\leq \int_0^1 \| \nabla^2 f(x_k + t p_k^N) - \nabla^2 f(x_k) \| \cdot \|p_k^N\| \, dt$$

$$\leq \frac{1}{2} L \|p_k^N\|^2$$

$$\leq \frac{1}{2} L \|\nabla^2 f(x_k)^{-1}\|^2 \cdot \|\nabla f_k\|^2$$

$$\uparrow$$

$$p_k^N = -(\nabla^2 f_k)^{-1} \cdot \nabla f_k$$

$$\leq 2L \| \nabla f(x) \| \cdot \| V t_k \|$$

Thus, the gradient norms converge to zero quadratically.

■