Newton's Method in n-dimn.

We are used to writing

$$f(b)-f(a)=\int_a^b f'(x)\,dx$$

for functions of a single variable. For functions of many variables, we can still write

$$f(\vec{y})-f(\vec{x})=\int_0^1 \underbrace{\underbrace{Df(t\vec{y}+(1-t)\vec{x})}_{\text{matrix}}\cdot\underbrace{(\vec{y}-\vec{x})}_{\text{vector}}}_{\text{vector}}\,dt$$

This means that

$$f(\vec{y})-f(\vec{x})-Df(\vec{x})(y-x)$$

$$=\int_0^1 \big(Df(t\vec{y}+(1-t)\vec{x})-Df(\vec{x})\big)(\vec{y}-\vec{x})\,dt$$

But $\vec{x}$ and $t\vec{y}+(1-t)\vec{x}$ are within $\|\vec{y}-\vec{x}\|$ of each other, so

~~the~~

$$\|Df(t\vec{y}+(1-t)\vec{x})-Df(\vec{x})\| \leq K\|x-y\|$$

This means that

$$\|f(\vec{y})-f(\vec{x})-Df(\vec{x})(\vec{y}-\vec{x})\|$$

$$=\left\|\int_0^1 (Df(t\vec{y}+(1-t)\vec{x})-Df(\vec{x}))(\vec{y}-\vec{x})\,dt\right\|$$

$$\leq \int_0^1 \|Df(t\vec{y}+(1-t)\vec{x})-Df(\vec{x})\|\,\|\vec{y}-\vec{x}\|$$

$$\leq \int_0^1 K\|\vec{y}-\vec{x}\|^2\,dt$$

$$\leq K\|\vec{y}-\vec{x}\|^2.$$

Subbing in $\vec{y}=N(\vec{x})$ and $\vec{x}=\vec{x}$, we have proved the lemma.

Lemma 4. The Newton sequence $\vec{X}_K$ is well-defined and majorized by

$$t_{K+1} = t_K - \frac{(\beta K/2)t_K^2 - t_K + \eta}{\beta K t_K - 1}, \quad t_0 = 0$$

Further, $t_K \to t_* = \frac{1}{\beta K}(1 - \sqrt{1-2h})$.

Proof. First, it's clear that the t_K are Newton iterates for $p(t) = \frac{(\beta K)}{2} t^2 - t + \eta$, which has roots

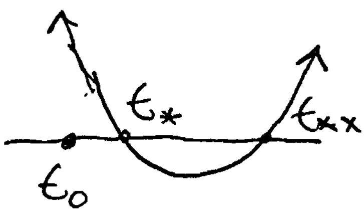
$$t_* = \frac{1 - \sqrt{1 - 2\beta K\eta}}{\beta K}$$

and

$$t_{**} = \frac{1 + \sqrt{1 - 2\beta K\eta}}{\beta K}$$

Since we assumed $\beta K\eta = h < 1/2$, the discriminant is positive and these roots are real and distinct.

Further, both are positive and $t_* < t_{**}$.



It's a homework exercise to show that the Newton iterates $t_k \to t_*$ and that the sequence is increasing.

Now we consider the first step in the Newton iteration from $\vec{x}_0$:

$$\vec{x}_1 = \vec{x}_0 - Df^{-1}(\vec{x}_0)\, f(\vec{x}_0).$$

We have assumed that $Df^{-1}(x_0)$ exists &

$$\| Df^{-1}(\vec{x}_0) f(\vec{x}_0) \| < \eta$$

so $\vec{x}_1$ exists and is within η of $\vec{x}_0$.

Now

$$t_1 = t_0 - \frac{(\beta K/2)t_0^2 - t_0 + \eta}{\beta K t_0 - 1}$$

$$= -\frac{\eta}{-1} = \eta, \text{ since } t_0 = 0.$$

Thus $\vec{x}_1$ exists and

$$\| \vec{x}_1 - \vec{x}_0 \| \leq t_1 - t_0.$$

We now proceed by induction.

Suppose that

$\vec{x}_1, \ldots, \vec{x}_k$ exist

and

$$\|\vec{x}_i - \vec{x}_{i-1}\| \leq t_i - t_{i-1}$$

for $i = 1, \ldots, k$. By the triangle inequality

$$\|\vec{x}_k - \vec{x}_0\| \leq t_k - t_0 \leq t_* - t_0 = t_*$$

This means that

$$\vec{x}_k \in B_{t_*}(\vec{x}_0) \underset{\substack{\uparrow \\ \text{assumption}}}{\subset} D_0$$

Now $t_* < \frac{1}{BK}$ so our first Lemma says that $Df(\vec{x}_k)$ is invertible.

Thus $\vec{x}_{k+1}$ exists. Further

$$\|\vec{x}_{k+1}-\vec{x}_k\| = \|N(N(\vec{x}_{k-1})) - N(\vec{x}_{k-1})\|$$

$$\le \frac{1}{2}\,\frac{\beta K\,\|\vec{x}_{k-1}-\vec{x}_k\|^2}{1-\beta K\,\|\vec{x}_0-\vec{x}_k\|}$$

Since $\|\vec{x}_k-\vec{x}_{k-1}\| \le t_k - t_{k-1}$ and $\|\vec{x}_k-\vec{x}_0\| \le t_{*k}$ by the inductive hypothesis,

$$\le \frac{1}{2}\,\frac{\beta K\,(t_k - t_{k-1})^2}{1-\beta K\,t_{*k}}$$

Now

$$t_{k+1} = t_k - \frac{(\beta K/2)\,t_k^2 - t_k + \eta}{\beta K\,t_k - 1}$$

by definition. So

$$t_{k+1} - t_k = \frac{(\beta K/2)\,t_k^2 - t_k + \eta}{1-\beta K\,t_k}$$

and

$$t_K = t_{K-1} - \frac{(\beta K/2)\, t_{K-1}^2 - t_{K-1} - \eta}{1 - \beta K t_K}$$

Claim.

$$t_{K+1} - t_K = \frac{(\beta K/2)(t_K - t_{K-1})^2}{1 - \beta K t_K}.$$

Proof. We can (laboriously!) reduce both sides to an expression in t_{K-1} using defns.

(Extra credit: Find a better proof!)

We can now assemble all the lemmas to prove the NK theorem.

By the last lemma, the sequence t_K dominates the sequence x_K and so $x_k \to x_*$.

We now show x_* is a solution.

$$\|f(\vec{x}_K)\| = $$

$$= \|Df(\vec{x}_k)(x_k - x_{k+1})\|$$

(defn of newton iteration

$$\vec{x}_{k+1} = \vec{x}_k - Df^{-1}(\vec{x}_k) f(\vec{x}_k) \ )$$

$$\leq \|Df(\vec{x}_k)\|_{op} \ \|x_k - x_{k+1}\|$$

$$\leq \left(\|Df(\vec{x}_0)\|_{op} + \|Df(\vec{x}_k) - Df(\vec{x}_k)\|_{op}\right) * \|x_k - x_{k+1}\|$$

(triangle inequality for op norm)

$$\leq \left[\|Df(\vec{x}_0)\|_{op} + K\|\vec{x}_k - \vec{x}_0\| \right] \|x_k - x_{k+1}\|$$

(assumption that Df is Lipschitz)

$$\leq \left(\|Df(\vec{x}_0)\|_{op} + K(t_k - t_0) \right) \|x_k - x_{k+1}\|$$

(t_k dominates $\vec{x}_k$)

$$\leq \left(\|Df(\vec{x}_0)\|_{op} + Kt_* \right) \|x_k - x_{k+1}\|$$

($t_* - t_0 \geq t_k - t_0$)

$$\leq \left(\|Df(\vec{x}_0)\|_{op} + Kt_* \right) |t_k - t_{k+1}|$$

and the rhs $\to 0$ as $k \to \infty$. Since $f(\vec{x})$ is a continuous function, this shows $f(\vec{x}_*) = 0$.

We now show that if $h < 1/2$, the convergence is at least quadratic.

~~If so, the~~

We note that $\vec{x}_k$ converges at least as fast as t_k and that if $\cancel{h} < 1/2$, the roots t_* and t_{**} of

$$(\beta K/2)t^2 - t + \eta$$

are distinct and hence since (homework) the convergence of $t_k \to t_*$ is quadratic, so is the convergence of $\vec{x}_k \to \vec{x}_*$.

The last claim is that $\vec{x}_*$ is the only root in $B_{t_{**}}(\vec{x}_0)$.