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

\[ f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \left( Df(t \mathbf{y} + (1-t) \mathbf{x}) \right) (\mathbf{y} - \mathbf{x}) \, dt \]

This means that

\[ f(\mathbf{y}) - f(\mathbf{x}) = Df(\mathbf{x}) (\mathbf{y} - \mathbf{x}) \]

\[ = \int_0^1 \left( Df(t \mathbf{y} + (1-t) \mathbf{x}) - Df(\mathbf{x}) \right) (\mathbf{y} - \mathbf{x}) \, dt \]

But \( \mathbf{x} \) and \( t \mathbf{y} + (1-t) \mathbf{x} \) are within \( \| \mathbf{y} - \mathbf{x} \| \) of each other, so
\[ \| Df(t\hat{y} + (1-t)\hat{x}) - Df(\hat{x}) \| \leq K \| x - y \| \]

This means that

\[ \| f(\hat{y}) - f(\hat{x}) - Df(\hat{x})(\hat{y} - \hat{x}) \| \]

\[ \leq \int_0^1 \| Df(t\hat{y} + (1-t)\hat{x}) - Df(\hat{x}) \| \| \hat{y} - \hat{x} \| \, dt \]

\[ \leq \int_0^1 K \| \hat{y} - \hat{x} \|^2 \, dt \]

\[ \leq K \| \hat{y} - \hat{x} \|^2. \]

Subbing in \( \hat{y} = N(\hat{x}) \) and \( \hat{x} = \hat{x} \), we have proved the lemma.
Lemma 4. The Newton sequence $\tilde{x}_k$ is well-defined and majorized by
\[
\tilde{x}_{k+1} = \tilde{x}_k - \frac{(BK/2)\tilde{x}_k^2 - \tilde{x} + \eta}{BK\tilde{x}_k - 1}, \tilde{x}_0 = 0
\]
Further, $\tilde{x}_k \to \tilde{x} = \frac{1}{BK} \left( 1 - \sqrt{1 - 2\eta} \right)$.

Proof. First, it's clear that the $\tilde{x}_k$ are Newton iterates for
\[
p(t) = (BK)\tilde{x}_k^2 - \tilde{x} + \eta,
\]
which has roots
\[
\tilde{x}_* = \frac{1 - \sqrt{1 - 2BK\eta}}{BK}
\]
and
\[
\tilde{x}_{**} = \frac{1 + \sqrt{1 - 2BK\eta}}{BK}
\]
Since we assumed $BK\eta = h < \frac{1}{2}$, the discriminant is positive.
and these roots are real and distinct. Further, both are positive and \( \varepsilon_x < \varepsilon_{xx} \).

It's a homework exercise to show that the Newton iterates \( \varepsilon_k \rightarrow \varepsilon_x \) and that the sequence is increasing.

Now we consider the first step in the Newton iteration formula \( \dot{x}_0 \):

\[
\dot{x}_{1} = \dot{x}_0 - Df^{-1}(\dot{x}_0) f(\dot{x}_0).
\]
We have assumed that $Df^{-1}(x_0)$ exists and so $\|Df^{-1}(x_0)f(x_0)\| < \eta$ so $x_1$ exists and is within $\eta$ of $x_0$.

Now

$$t_1 = t_0 - \frac{(BK/2)t_0^2 - t_0 + \eta}{BKt_0 - 1}$$

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

Thus $x_1$ exists and

$$\|x_1 - x_0\| \leq t_1 - t_0.$$ 

We now proceed by induction.
Suppose that 
\[
\tilde{x}_1, \ldots, \tilde{x}_k
\]
exist and 
\[
\| \tilde{x}_i - \tilde{x}_{i-1} \| \leq \varepsilon_i - \varepsilon_{i-1}
\]
for \( i = 1, \ldots, k \). By the triangle inequality
\[
\| \tilde{x}_k - \tilde{x}_0 \| \leq \varepsilon_k - t_0 \leq \varepsilon_k - t_0 = t_x
\]
This means that 
\[
\tilde{x}_k \in B_{t_x}(\tilde{x}_0) \subset D_0
\]
up
assumption
Now \( t_x < \frac{1}{BK} \) so our first lemma says that \( D_f(\tilde{x}_k) \) is invertible.
Thus \( \hat{x}_{k+1} \) exists. Further,

\[
\| \hat{x}_{k+1} - \hat{x}_k \| = \| N(\hat{x}_{k-1}) - N(\hat{x}_{k-1}) \|
\]

\[
\leq \frac{1}{2} \frac{BK \| \hat{x}_{k-1} - \hat{x}_k \|^2}{1 - BK \| \hat{x}_0 - \hat{x}_k \|}
\]

Since \( \| \hat{x}_k - \hat{x}_{k-1} \| \leq t_k - t_{k-1} \) and \( \| \hat{x}_k - \hat{x}_0 \| \leq t_k \) by the inductive hypothesis,

\[
\leq \frac{1}{2} \frac{BK (\| t_k - t_{k-1} \|^2)}{1 - BK t_k}
\]

Now,

\[
t_{k+1} = t_k - \frac{(BK/2) t_k^2 - t_{k+1} + N}{BK t_k - 1}
\]

by definition. So

\[
t_{k+1} - t_k = \frac{(BK/2) t_k^2 - t_{k+1} + N}{1 - BK t_k}
\]
and
\[ t_K = t_{K-1} - \frac{(BK/2)t_{K-1}^2 - t_{K-1} - n}{1 - BKt_K} \]

Claim.
\[ t_{K+1} - t_K = \frac{(BK/2)(t_K - t_{K-1})^2}{1 - BKt_K} . \]

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 $X_k \to x_k$.

We now show $x_k$ is a solution.

$$
\|f(x_k)\| = \|Df(x_k) (x_k - x_{k+1})\|
$$

(defn of newton iteration

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

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

$$
\leq (\|Df(x_0)\|_{op} + \|Df(x_k) - Df(x_{k+1})\|)_{op}
$$

(triangle inequality for op norm)
\[
\leq \left[ \| Df(x_0) \|_{op} + K \| \dot{x}_k - \dot{x}_0 \| \right] \| x_k - x_{k+1} \|
\]
(assumption that \( Df \) is Lipschitz)
\[
\leq (\| Df(x_0) \|_{op} + K(t_k-t_0)) \| x_k - x_{k+1} \|
\]
(\( t_k \) dominates \( \dot{x}_k \))
\[
\leq (\| Df(x_0) \|_{op} + K t_*) \| x_k - x_{k+1} \|
\]
(\( t_* - t_0 \geq t_k - t_0 \))
\[
\leq (\| Df(x_0) \|_{op} + K t_*) | t_k - t_{k+1} |
\]

and the rhs \( \to 0 \) as \( K \to \infty \).

Since \( f(\dot{x}) \) is a continuous function, this shows \( f(\dot{x}_*) = 0 \).

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

Indeed, we note that \( \hat{x}_k \) converges at least as fast as \( t_k \) and that if \( b < \frac{1}{2} \), the roots \( t_* \) and \( t_{**} \) of

\[(Bk/2)\xi^2 - \xi + \eta\]

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

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