Math 346 Lecture #9
7.3 Newton's Method

Newton's method, for finding a zero of a function, is quite simple: use linear approximations to generate a sequence of successive approximations.

For a Banach space X, the linear approximation of a differentiable $f : X \to X$ at a point $\mathrm{x}_n \in X$ is

$$L(\mathrm{x}) = f(\mathrm{x}_n) + Df(\mathrm{x}_n)(\mathrm{x} - \mathrm{x}_n).$$

If $Df(\mathrm{x}_n) \in \mathscr{B}(X)$ is invertible, then L has a unique zero at

$$\mathrm{x}_{n+1} = \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n).$$

Starting with a guess $\mathrm{x}_0 \in X$, we form a sequence of successive approximations $(\mathrm{x}_n)_{n=0}^\infty$ which will converge to a zero $\bar{\mathrm{x}}$ of f if x_0 is close enough to $\bar{\mathrm{x}}$ and $Df(\mathrm{x})$ has bounded inverse for all x in an open ball centered at $\bar{\mathrm{x}}$.

7.3.1 Convergence

For a sequence $(\mathrm{x}_n)_{n=0}^\infty$ in a normed linear space $(X, \|\cdot\|)$ converging to $\bar{\mathrm{x}} \in X$, we quantify two different rates of convergence.

Definition 7.3.1. For $(\mathrm{x}_n)_{n=0}^\infty$ converging to $\bar{\mathrm{x}}$ in a normed linear space $(X, \|\cdot\|)$, denote the error between x_n and $\bar{x}$ by

$$\epsilon_n = \|\mathrm{x}_n - \bar{\mathrm{x}}\|.$$

The sequence $(\mathrm{x}_n)_{n=0}^\infty$ converges linearly with rate $\mu \in [0, 1)$ if for all $n = 0, 1, 2, 3, \ldots$ there holds

$$\epsilon_{n+1} \leq \mu \epsilon_n.$$

The sequence $(\mathrm{x}_n)_{n=0}^\infty$ converges quadratically with rate $k \geq 0$ (not necessarily smaller than 1) if for all $n = 0, 1, 2, 3, \ldots$ there holds

$$\epsilon_{n+1} \leq k \epsilon_n^2.$$

For convergent sequences of real numbers, linear convergence with rate μ adds about $\log_{10} \mu$ digits of accuracy each iteration, while quadratic convergence with rate k doubles the number of digits of accuracy with each iteration.

7.3.2 Newton's Method: Scalar Version

Convergence of Newton's method is a consequence of the Contraction Mapping Principle.

Lemma 7.3.2. Let $f : [a, b] \to \mathbb{R}$ be C^2. If there is $\bar{x} \in (a, b)$ such that $f(\bar{x}) = 0$ and $f'(\bar{x}) \neq 0$, then there exists $\delta > 0$ such that $[\bar{x} - \delta, \bar{x} + \delta] \subset [a, b]$ and the function

$$\phi(x) = x - \frac{f(x)}{f'(x)}$$

maps $[\bar{x} - \delta, \bar{x} + \delta]$ into $[\bar{x} - \delta, \bar{x} + \delta]$ and is a contraction on $[\bar{x} - \delta, \bar{x} + \delta]$.

Proof. Continuity of f' at $\bar{x}$ and $f'(\bar{x}) \neq 0$ imply the existence of $\delta_1 > 0$ such that $(\bar{x} - \delta_1, \bar{x} + \delta_1) \subset (a, b)$ and $|f'(x)| > |f(\bar{x})|/2 > 0$ for all $x \in (\bar{x} - \delta_1, \bar{x} + \delta_1)$.

Since f is C^2, the function ϕ is C^1 on $(\bar{x}-\delta_1, \bar{x}+\delta_1)$ with derivative

$$\phi'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}.$$

Continuity of f'' on $[a,b]$ implies the existence of $M > 0$ such that $|f''(x)| \leq M$ on $[a,b]$. This together with $|f'(x)| > |f(\bar{x})|/2$ on $(\bar{x}-\delta_1, \bar{x}+\delta_1)$ gives for all $x \in (\bar{x}-\delta_1, \bar{x}+\delta_1)$ that

$$|\phi'(x)| \leq \frac{2M}{|f(\bar{x})|^2}|f(x)|.$$

Continuity of f at $\bar{x}$ and $f(\bar{x}) = 0$ gives the existence of $\delta \in (0, \delta_1)$ such that for all $x \in [\bar{x}-\delta, \bar{x}+\delta]$ there holds

$$|f(x)| \leq \frac{|f'(\bar{x})|^2}{2M}\frac{9}{10}.$$

Thus on $[\bar{x}-\delta, \bar{x}+\delta]$ we have

$$|\phi'(x)| \leq \frac{9}{10}.$$

By the Mean Value Theorem, for any $[x,y] \subset [\bar{x}-\delta, \bar{x}+\delta]$ there exists $c \in (x,y)$ such that

$$|\phi(x) - \phi(y)| = |\phi'(c)|\,|x-y| \leq \frac{9}{10}|x-y|.$$

Since $\bar{x}$ is a fixed point of ϕ, then for any $x \in [\bar{x}-\delta, \bar{x}+\delta]$ there holds

$$|\phi(x) - \bar{x}| = |\phi(x) - \phi(\bar{x})| \leq \frac{9}{10}|x-\bar{x}| \leq \frac{9\delta}{10} < \delta.$$

Therefore ϕ maps $[\bar{x}-\delta, \bar{x}+\delta]$ into $[\bar{x}-\delta, \bar{x}+\delta]$ (which is what Remark 7.3.3 says) and is a contraction on $[\bar{x}-\delta, \bar{x}+\delta]$. □

Theorem 7.3.4 (Newton's Method–Scalar Version). If $f : [a,b] \to \mathbb{R}$ is C^2, and there is $\bar{x} \in (a,b)$ such that $f(\bar{x}) = 0$ and $f'(\bar{x}) \neq 0$, then the sequence $(x_n)_{n=0}^{\infty}$ defined iteratively by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

converges to $\bar{x}$ quadratically whenever x_0 is sufficiently close to $\bar{x}$.

Proof. Since f is C^2, the derivative f' is locally Lipschitz at $\bar{x}$ by Proposition 6.3.7: there exists $\delta_1 > 0$ and $L > 0$ such that for all $|h| < \delta_1$ there holds

$$|f'(\bar{x}+h) - f'(\bar{x})| \leq L|h|.$$

By Lemma 7.3.2, there exists $\delta_2 > 0$ such that the function $\phi(x) = x - f(x)/f'(x)$ is a contraction on $[\bar{x}-\delta_2, \bar{x}+\delta_2]$.

Choose $\delta < \min\{\delta_1, \delta_2\}$.

For an initial condition $x_0 \in [\bar{x}-\delta, \bar{x}+\delta]$, the sequence $(x_n)_{n=0}^{\infty}$ defined iteratively by $x_{n+1} = \phi(x_n)$ converges to $\bar{x}$ (as $\bar{x}$ is a fixed point of f and the Contraction Mapping Principle guarantees a unique fixed point, so the limit of $(x_n)_{n=0}^{\infty}$ must be $\bar{x}$).

Set $\epsilon_n = x_n - \bar{x}$.

The function $h \to f(\bar{x} + h\epsilon_{n-1})$ is continuous on $h \in [0,1]$ and differentiable on $(0,1)$.

By the Mean Value Theorem (and $f(\bar{x}) = 0$) there exists $\eta \in (0,1)$ such that

$$f(\bar{x} + \epsilon_{n-1}) = f(\bar{x} + \epsilon_{n-1}) - f(\bar{x}) = f'(\bar{x} + \eta\epsilon_{n-1})\epsilon_{n-1}.$$

This applies even when $\epsilon_{n-1} < 0$.

From the iterative definition of $(x_n)_{n=0}^{\infty}$ we have

$$\begin{aligned}
|\epsilon_n| &= |x_n - \bar{x}| \\
&= \left| x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} - \bar{x} \right| \\
&= \left| \epsilon_{n-1} - \frac{f(\bar{x} + \epsilon_{n-1})}{f'(\bar{x} + \epsilon_{n-1})} \right| \\
&= \left| \frac{f'(\bar{x} + \epsilon_{n-1})\epsilon_{n-1} - f(\bar{x} + \epsilon_{n-1})}{f'(\bar{x} + \epsilon_{n-1})} \right| \\
&= \left| \frac{f'(\bar{x} + \epsilon_{n-1})\epsilon_{n-1} - f'(\bar{x} + \eta\epsilon_{n-1})\epsilon_{n-1}}{f'(\bar{x} + \epsilon_{n-1})} \right| \\
&= \left| \frac{f'(\bar{x} + \epsilon_{n-1}) - f'(\bar{x} + \eta\epsilon_{n-1})}{f'(\bar{x} + \epsilon_{n-1})} \right| |\epsilon_{n-1}| \\
&= \left| \frac{f'(\bar{x} + \epsilon_{n-1}) - f'(\bar{x}) + f'(\bar{x}) - f'(\bar{x} + \eta\epsilon_{n-1})}{f'(\bar{x} + \epsilon_{n-1})} \right| |\epsilon_{n-1}| \\
&\le \left\{ \left| \frac{f'(\bar{x} + \epsilon_{n-1}) - f'(\bar{x})}{f'(\bar{x} + \epsilon_{n-1})} \right| + \left| \frac{f'(\bar{x}) - f'(\bar{x} + \eta\epsilon_{n-1})}{f'(\bar{x} + \epsilon_{n-1})} \right| \right\} |\epsilon_{n-1}| \\
&\le \left\{ \frac{L|\epsilon_{n-1}|}{|f'(\bar{x} + \epsilon_{n-1})|} + \frac{L|\eta\epsilon_{n-1}|}{|f'(\bar{x} + \epsilon_{n-1})|} \right\} |\epsilon_{n-1}| \\
&\le \left\{ \frac{L|\epsilon_{n-1}|}{|f'(\bar{x} + \epsilon_{n-1})|} + \frac{L|\epsilon_{n-1}|}{|f'(\bar{x} + \epsilon_{n-1})|} \right\} |\epsilon_{n-1}| \\
&= \frac{2L}{|f'(\bar{x} + \epsilon_{n-1})|} |\epsilon_{n-1}|^2.
\end{aligned}$$

Since $|f'| \ge |f'(\bar{x})|/2$ on $[\bar{x} - \delta, \bar{x} + \delta]$ (by the choice of δ in Lemma 7.3.2), the quantity

$$M = \inf\{|f'(t)| : \bar{x} - \delta \le t \le \bar{x} + \delta\}$$

is finite and positive.

With $|f'(t)| \ge M$ for $t \in [\bar{x} - \delta, \bar{x} + \delta]$, we thus have

$$|\epsilon_n| \le \frac{2L}{M} |\epsilon_{n-1}|^2,$$

giving quadratic convergence. $\square$

Example 7.3.5. The function

$$g(x) = \frac{1}{2}\left(x + \frac{b}{x}\right)$$

that gives the square root of $b \geq 1$ as the limit of the sequence $(x_n)_{n=0}^\infty$ where $x_0 \geq \sqrt{b/2}$ and $x_{n+1} = g(x_n)$, is Newton's method applied to $f(x) = x^2 - b$ because $f'(x) = 2x$ so that

$$x_{n+1} = x_n - \frac{x_n^2 - b}{2x_n} = x_n - \frac{x_n}{2} + \frac{b}{2x_n} = \frac{1}{2}\left(x_n + \frac{b}{x_n}\right).$$

The function $f(x) = x^2 - b$ is C^2 on $[\sqrt{b/2}, b]$ with $f(\bar{x}) = 0$ for $\bar{x} = \sqrt{b} \in [\sqrt{b/2}, b]$, and $f'(\bar{x}) \neq 0$.

By Theorem 7.3.4, the convergence of $(x_n)_{n=0}^\infty$ to $\bar{x} = \sqrt{b}$ is quadratic.

Remark 7.3.6. If $f'(\bar{x}) = 0$, then Newton's method is not necessarily quadratic in convergence and it may not even converge!

When $f'(\bar{x}) = 0$, we say that f has a multiple zero at $\bar{x}$.

When $f'(\bar{x}) = 0$, we say that f has a simple, or isolated, zero at $\bar{x}$, i.e., there is no other zero of f in a open ball centered at $\bar{x}$.

Remark 7.3.8. The sequence arising in Newton's method may not converge if the initial guess x_0 is not close enough to $\bar{x}$. Unexample 7.3.9 gives an example of an initial guess x_0 for which $|x_n| \to \infty$.

7.3.3 A Quasi-Newton Method: Vector Version

The Quasi-Newton method is similar to the Newton method, but it depends on knowing a priori (knowing before hand) the fixed point. The sequence arising from the Quasi-Newton method converges but not necessarily quadratically. The Quasi-Newton method plays a key role in the proof of the Implicit Function Theorem of Section 7.4.

Definition. For a Banach space $(X, \|\cdot\|)$, an operator $A \in \mathscr{B}(X)$ is said to have bounded inverse if A is invertible, and its inverse $A^{-1} \in \mathscr{L}(X)$ has finite operator norm, i.e.,

$$\|A^{-1}\| = \sup\left\{\frac{\|A^{-1}\mathrm{x}\|}{\|\mathrm{x}\|} : \mathrm{x} \in X, \mathrm{x} \neq 0\right\},$$

so that $A^{-1} \in \mathscr{B}(X)$. In this case we have $AA^{-1} = I = A^{-1}A$.

For X finite dimensional, every invertible operator has bounded inverse, but this is not true when X is infinite dimensional (think integration versus differentiation).

Theorem 7.3.10. Let $(X, \|\cdot\|)$ be a Banach space, $f : X \to X$ a C^1 function, and U an open neighbourhood of $\bar{x} \in X$. If $f(\bar{x}) = 0$ and $Df(\bar{x}) \in \mathscr{B}(X)$ has bounded inverse, then there exists $\delta > 0$ such that $\overline{B(\bar{\mathrm{x}}, \delta)} \subset U$ and

$$\phi(\mathrm{x}) = \mathrm{x} - Df(\bar{\mathrm{x}})^{-1} f(\mathrm{x})$$

is a contraction on $\overline{B(\bar{\mathrm{x}}, \delta)}$.

Proof. By the hypothesized continuity of Df, there exists $\delta > 0$ such that $\overline{B(\bar{x}, \delta)} \subset U$ and for all $x \in \overline{B(\bar{x}, \delta)}$ there holds

$$\|Df(\bar{x}) - Df(x)\| < \frac{1}{2\|Df(x)^{-1}\|}.$$

Hence for every $x \in \overline{B(\bar{x}, \delta)}$ we have

$$\begin{aligned}
\|D\phi(x)\| &= \|I - Df(\bar{x})^{-1}Df(x)\| \\
&= \|Df(\bar{x})^{-1}Df(\bar{x}) - Df(\bar{x})^{-1}Df(x)\| \\
&\leq \|Df(\bar{x})^{-1}\| \, \|Df(x) - Df(x)\| \\
&< \frac{\|Df(\bar{x})^{-1}\|}{2\|Df(\bar{x})^{-1}\|} \\
&= \frac{1}{2}.
\end{aligned}$$

For $x, y \in \overline{B(\bar{x}, \delta)}$, applying the Integral Mean Value Theorem along the line segment $\ell(x, y)$ gives

$$\begin{aligned}
\|\phi(x) - \phi(y)\| &= \left\| \int_0^1 D\phi((1-t)x + ty)(x - y)\ dt \right\| \\
&\leq \int_0^1 \|D\phi((1-t)x + ty)\| \, \|x - y\| \, dt \\
&\leq \int_0^1 \frac{\|x - y\|}{2}\ dt \\
&= \frac{1}{2}\|x - y\|.
\end{aligned}$$

This shows that ϕ is a contraction mapping on $\overline{B(\bar{x}, \delta)}$. □

If we happen to know the fixed point $\bar{x}$ of ϕ, then we can compute $Df(\bar{x})^{-1}$ and use ϕ to determine a sequence $(x_n)_{n=0}^{\infty}$ defined iteratively by

$$x_{n+1} = \phi(x_n) = x_n - Df(\bar{x})^{-1}f(x_n).$$

If we do not know $\bar{x}$, then we might be able to find a good approximation of $Df(\bar{x})^{-1}$ that we can use in ϕ to determine $(x_n)_{n=0}^{\infty}$.

Lemma 7.3.11. For a Banach space $(X, \|\cdot\|)$ suppose that $g : X \to \mathscr{B}(X)$ is continuous. For $\bar{x} \in X$, if $g(\bar{x})$ has bounded inverse, then there exists a $\delta > 0$ such that for all $x \in B(\bar{x}, \delta)$ there holds

$$\|g(x)^{-1}\| < 2\|g(\bar{x})^{-1}\|.$$

Proof. Proposition 5.7.7 (from Subsection 5.7.4 that we skipped) states that the function $A \to A^{-1}$ on

$$\mathrm{GL}(X) = \{A \in \mathscr{B}(X) : A^{-1} \text{ exists and belongs to } \mathscr{B}(X)\}$$

is a continuous function.

Since $g(\bar{\mathrm{x}})$ has bounded inverse, i.e., $g(\bar{\mathrm{x}}) \in \mathrm{GL}(X)$, the continuity of $g : X \to \mathscr{B}(X)$ at $\bar{\mathrm{x}}$ implies for $\epsilon = \|g(\bar{\mathrm{x}})^{-1}\|$ the existence of $\delta > 0$ such that for all $\mathrm{x} \in B(\bar{\mathrm{x}}, \delta)$ there holds

$$\|g(\mathrm{x})^{-1} - g(\bar{\mathrm{x}})^{-1}\| < \epsilon.$$

Thus

$$\begin{aligned} \|g(\mathrm{x})^{-1}\| &= \|g(\mathrm{x})^{-1} - g(\bar{\mathrm{x}})^{-1} + g(\bar{\mathrm{x}})^{-1}\| \\ &\leq \|g(\mathrm{x})^{-1} - g(\bar{\mathrm{x}})^{-1}\| + \|g(\bar{\mathrm{x}})^{-1}\| \\ &< \epsilon + \|g(\bar{\mathrm{x}})^{-1}\| \\ &= 2\|g(\bar{\mathrm{x}})^{-1}\|. \end{aligned}$$

whenever $\mathrm{x} \in B(\bar{\mathrm{x}}, \delta)$. □.

7.3.4 Newton's Method: Vector Version

We extend the scalar version of Newton's method to the general Banach space setting.

Theorem 7.3.12 (Newton's Method–Vector Version. Let $(X, \|\cdot\|)$ be a Banach space and $f : X \to X$. Suppose there is an open neighbourhood U of $\bar{\mathrm{x}} \in X$ for which $f \in C^1(U, X)$ and $f(\bar{\mathrm{x}}) = 0$. If $Df(\bar{\mathrm{x}})$ has bounded inverse and Df is Lipschitz on U, then for $\mathrm{x}_0 \in U$ chosen sufficiently close to $\bar{\mathrm{x}}$, the sequence $(\mathrm{x}_n)_{n=0}^{\infty}$ defined iteratively by

$$\mathrm{x}_{n+1} = \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n)$$

converges quadratically to $\bar{\mathrm{x}}$.

Proof. The assumed Lipschitz of Df on U implies the continuity if Df on U.

By Lemma 7.3.11 there exists $\delta > 0$ such that $B(\bar{\mathrm{x}}, \delta) \subset U$ and for all $\mathrm{x} \in B(\bar{\mathrm{x}}, \delta)$ there holds

$$\|Df(\mathrm{x})^{-1}\| < 2\|Df(\bar{\mathrm{x}})^{-1}\|.$$

For $\mathrm{x}_0 \in B(\bar{\mathrm{x}}, \delta)$, form the sequence $(\mathrm{x}_n)_{n=0}^{\infty}$ by

$$\mathrm{x}_{n+1} = \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n).$$

By the Integral Mean Value Theorem (which is the first-order Taylor expansion) we have

$$\begin{aligned} f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) &= \int_0^1 Df(\bar{\mathrm{x}} + t(\mathrm{x}_n - \bar{\mathrm{x}}))(\mathrm{x}_n - \bar{\mathrm{x}})\ dt \\ &= Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}}) + \int_0^1 \big(Df(\bar{\mathrm{x}} + t(\mathrm{x}_n - \bar{\mathrm{x}})) - Df(\bar{\mathrm{x}})\big)(\mathrm{x}_n - \bar{\mathrm{x}})\ dt. \end{aligned}$$

This gives

$$f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}}) = \int_0^1 \big(Df(\bar{\mathrm{x}} + t(\mathrm{x}_n - \bar{\mathrm{x}})) - Df(\bar{\mathrm{x}})\big)(\mathrm{x}_n - \bar{\mathrm{x}})\ dt.$$

Again by the assumed Lipschitz of Df on U, there is $k > 0$ such that for all $\mathrm{x}, \mathrm{y} \in U$ there holds

$$\|Df(\mathrm{x}) - Df(\mathrm{y})\| \leq k\|\mathrm{x} - \mathrm{y}\|.$$

Applying this gives

$$\begin{aligned}
&\|f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})\| \\
&\quad\leq \int_0^1 \|\big(Df(\bar{\mathrm{x}} + t(\mathrm{x}_n - \bar{\mathrm{x}})) - Df(\bar{\mathrm{x}})\big)(\mathrm{x}_n - \bar{\mathrm{x}})\| \; dt \\
&\quad\leq \int_0^1 \|Df(\bar{\mathrm{x}} + t(\mathrm{x}_n - \bar{\mathrm{x}})) - Df(\bar{\mathrm{x}})\| \, \|\mathrm{x}_n - \bar{\mathrm{x}}\| \; dt \\
&\quad\leq \int_0^1 k\|t(\mathrm{x}_n - \bar{\mathrm{x}})\| \, \|\mathrm{x}_n - \bar{\mathrm{x}}\| \, dt \\
&\quad= \int_0^1 kt\|\mathrm{x}_n - \bar{\mathrm{x}}\| \, \|\mathrm{x}_n - \bar{\mathrm{x}}\| \; dt \\
&\quad= k\|\mathrm{x}_n - \bar{\mathrm{x}}\|^2 \int_0^1 t \; dt \\
&\quad= \frac{k}{2}\|\mathrm{x}_n - \bar{\mathrm{x}}\|^2.
\end{aligned}$$

From the inductive definition of $(\mathrm{x}_n)_{n=0}^\infty$ and $f(\bar{\mathrm{x}}) = 0$ we have

$$\begin{aligned}
\mathrm{x}_{n+1} - \bar{\mathrm{x}} &= \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n) - \bar{\mathrm{x}} \\
&= \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n) - \bar{\mathrm{x}} + Df(\mathrm{x}_n)^{-1} f(\bar{\mathrm{x}}) \\
&= \mathrm{x}_n - \bar{\mathrm{x}} - Df(\mathrm{x}_n)^{-1}(f(\mathrm{x}_n) - f(\bar{\mathrm{x}})) \\
&= \mathrm{x}_n - \bar{\mathrm{x}} - Df(\mathrm{x}_n)^{-1}\big(Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}}) + f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})\big) \\
&= Df(\mathrm{x}_n)^{-1} Df(\mathrm{x}_n)(\mathrm{x}_n - \bar{\mathrm{x}}) \\
&\qquad - Df(\mathrm{x}_n)^{-1}\big(Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}}) + f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})\big) \\
&= Df(\mathrm{x}_n)^{-1}\big(Df(\mathrm{x}_n) - Df(\bar{\mathrm{x}})\big)(\mathrm{x}_n - \bar{\mathrm{x}}) \\
&\qquad - Df(\mathrm{x}_n)^{-1}\big(f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})\big).
\end{aligned}$$

Using the triangle inequality and the estimates above we have

$$\begin{aligned}
\epsilon_{n+1} &= \|x_{n+1} - \bar{\mathrm{x}}\| \\
&\leq \|Df(\mathrm{x}_n)^{-1}\big(Df(\mathrm{x}_n) - Df(\bar{\mathrm{x}})\big)(\mathrm{x}_n - \bar{\mathrm{x}})\| \\
&\qquad + \|Df(\mathrm{x}_n)^{-1}\big(f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})\big)\| \\
&\leq \|Df(\mathrm{x}_n)^{-1}\| \, \|Df(\mathrm{x}_n) - Df(\bar{\mathrm{x}})\| \, \|\mathrm{x}_n - \bar{\mathrm{x}}\| \\
&\qquad + \|Df(\mathrm{x}_n)^{-1}\| \, \|f(\mathrm{x}_n) - f(\bar{\mathrm{x}}) - Df(\bar{\mathrm{x}})(\mathrm{x}_n - \bar{\mathrm{x}})]\| \\
&< 2k\|Df(\bar{\mathrm{x}})^{-1}\| \, \|\bar{x}_n - \bar{\mathrm{x}}\|^2 + k\|Df(\bar{\mathrm{x}})^{-1}\| \, \|\bar{x}_n - \bar{\mathrm{x}}\|^2 \\
&= 3k\|Df(\bar{\mathrm{x}})^{-1}\|\epsilon_n^2.
\end{aligned}$$

With $M = 3k\|Df(\bar{\mathrm{x}})^{-1}\|$ we have obtained $\epsilon_{n+1} \leq M\epsilon_n^2$. □

Remark 7.3.13. There is a way to know whether the initial guess x_0 is close enough to $\bar{\mathrm{x}}$ so that the sequence $(x_n)_{n=0}^{\infty}$ converges to $\bar{\mathrm{x}}$.

The Newton-Kantorovich Theorem, which generalizes Lemma 7.3.2, states that, under the hypothesis of Theorem 7.3.12, if

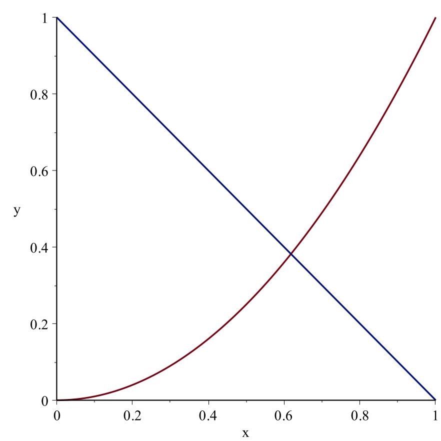
$$k\|f(\mathrm{x}_0)\|\,\|Df(\mathrm{x}_0)^{-1}\|^2 \leq \frac{1}{2},$$

where k is the Lipschitz constant of $Df : U \to \mathscr{B}(X)$, then x_0 is close enough.

Example (in lieu of 7.3.14). We illustrate the use of the Newton-Kantorovich Theorem for the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x,y) = \begin{bmatrix} x^2 - y \\ x + y - 1 \end{bmatrix}.$$

This function has a zero $\bar{\mathrm{x}}$ in the first quadrant where the curves $x^2 - y = 0$ and $x + y - 1 = 0$ intersect. See the following graph.



A starting guess for Newton's method is $\mathrm{x}_0 = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}^{\mathrm{T}}$.

Using the ∞-norm on $\mathbb{R}^2$ we have

$$\|f(\mathrm{x}_0)\|_\infty = \left\| \begin{bmatrix} -1/4 \\ 0 \end{bmatrix} \right\|_\infty = \frac{1}{4}.$$

Since

$$Df(x,y) = \begin{bmatrix} 2x & -1 \\ 1 & 1 \end{bmatrix}$$

we have

$$Df(x,y)^{-1} = \frac{1}{2x+1} \begin{bmatrix} 1 & 1 \\ -1 & 2x \end{bmatrix}.$$

This gives

$$Df(\mathrm{x}_0)^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

for which (by Theorem 3.5.20) we have

$$\|Df(\mathrm{x}_0)^{-1}\|_\infty = 1.$$

It remains to find the value of k in the Lipschitz condition for Df on some neighbourhood U of $\bar{\mathrm{x}}$.

We use second derivative of f to get k.

By the Integral Mean Value Theorem we have

$$\|Df(\mathrm{x}) - Df(\mathrm{y})\|_\infty \leq \sup_{\mathrm{c}\in\ell(\mathrm{y},\mathrm{x})} \|D^2 f(\mathrm{c})\| \, \|\mathrm{x} - \mathrm{y}\|_\infty.$$

With $f = (f_1, f_2)$ we have

$$Df_1(x,y) = \begin{bmatrix} 2x & 1 \end{bmatrix} \text{ and } Df_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

Thus

$$D^2 f_1(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } D^2 f_2(x,y) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

With $D^2 f(x,y) \in \mathscr{B}(\mathbb{R}^2, \mathbb{R}^2; \mathbb{R}^2)$ and $\mathrm{h}_1, \mathrm{h}_2 \in \mathbb{R}^2$, we have

$$D^2 f(x,y)(\mathrm{h}_1, \mathrm{h}_2) = \begin{bmatrix} \mathrm{h}_1^{\mathrm{T}} D^2 f_1(x,y)\mathrm{h}_2 \\ \mathrm{h}_1 D^2 f_2(x,y)\mathrm{h}_2 \end{bmatrix}.$$

The second entry here is always $0 \in \mathbb{R}$, while for the first entry, with

$$\mathrm{h}_1 = \begin{bmatrix} a \\ b \end{bmatrix} \text{ and } \mathrm{h}_2 = \begin{bmatrix} c \\ d \end{bmatrix},$$

we have

$$\mathrm{h}_1^{\mathrm{T}} D^2 f_1(x,y)\mathrm{h}_2 = 2ac \in \mathbb{R}.$$

From this we get for $\mathrm{h}_1, \mathrm{h}_2 \neq 0$ that

$$\frac{\|\mathrm{h}_1^{\mathrm{T}} D^2 f(x,y)\mathrm{h}_2\|_\infty}{\|\mathrm{h}_1\|_\infty \|\mathrm{h}_2\|_\infty} = \frac{2|ac|}{\sup\{|a|,|b|\}\sup\{|c|,|d|\}} \leq 2.$$

This implies for all (x,y) that

$$\|D^2 f(x,y)\| \leq 2.$$

Thus by the Integral Mean Value Theorem we have for all $\mathrm{x}, \mathrm{y} \in \mathbb{R}^2$ that

$$\|Df(\mathrm{x}) - Df(\mathrm{y})\| \leq 2\|\mathrm{x} - \mathrm{y}\|_\infty$$

and so the Lipschitz constant for Df on $U = \mathbb{R}^2$ is $k = 2$.

The initial guess of $\mathrm{x}_0 = (1/2, 1/2)$ satisfies the Newton-Kantorovich condition because

$$k\|f(\mathrm{x}_0)\| \, \|Df(\mathrm{x}_0)^{-1}\|^2 = 2(1/4)(1)^2 = \frac{1}{2}.$$

The sequence of successive approximations $(\mathrm{x}_n)_{n=0}^\infty$ defined by

$$\mathrm{x}_{n+1} = \mathrm{x}_n - Df(\mathrm{x}_n)^{-1} f(\mathrm{x}_n)$$

therefore converges quadratically to the root of f in the first quadrant.

Here

$$\begin{aligned}
\mathrm{x}_1 &= \mathrm{x}_0 - Df(\mathrm{x}_0)^{-1} f(\mathrm{x}_0) \\
&= \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} - \frac{1}{2(1/2)+1} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1/4 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1/4 \\ 1/4 \end{bmatrix} \\
&= \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 1/8 \\ -1/8 \end{bmatrix} \\
&= \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix} = \begin{bmatrix} 0.625 \\ 0.375 \end{bmatrix}.
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{x}_2 &= \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix} - \frac{1}{2(5/8)+1} \begin{bmatrix} 1 & 1 \\ -1 & 2(5/8) \end{bmatrix} \begin{bmatrix} (5/8)^2 - 3/8 \\ 5/8 + 3/8 - 1 \end{bmatrix} \\
&= \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix} - \frac{4}{9} \begin{bmatrix} 1 & 1 \\ -1 & 5/4 \end{bmatrix} \begin{bmatrix} 1/64 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix} - \frac{4}{9} \begin{bmatrix} 1/64 \\ -1/64 \end{bmatrix} \\
&= \begin{bmatrix} 5/8 - 1/144 \\ 3/8 + 1/144 \end{bmatrix} = \begin{bmatrix} 0.61805556 \\ 0.3819444 \end{bmatrix}.
\end{aligned}$$

We can explicitly compute the limit $\bar{\mathrm{x}} = (\bar{x}, \bar{y})$ of $(\mathrm{x}_n)_{n=0}^\infty$ because we can algebraically solve the system of equations

$$\begin{aligned}
x^2 - y &= 0, \\
x + y - 1 &= 0.
\end{aligned}$$

The second gives $-y = x - 1$ and substitution of this into the first gives the quadratic

$$x^2 + x - 1 = 0.$$

By the quadratic formula we have

$$\bar{x} = \frac{-1 + \sqrt{5}}{2} \cong 0.618033988749895$$

so that by $y = 1 - x$ we have

$$\bar{y} = \frac{3 - \sqrt{5}}{2} \cong 0.381966011250105.$$

The iterate x_2 has four correct digits in both entries!