Nonlinear equations

Norms for $\mathbb{R}^n$

Assume that X is a vector space. A **norm** $\|\cdot\|$ is a mapping $X \to \mathbb{R}$ with $\|x\| \geq 0$ such that for all $x, y \in X$, $\alpha \in \mathbb{R}$

- $\|x\| = 0 \implies x = 0$
- $\|\alpha x\| = |\alpha| \, \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$

We define the following norms on the vector space $\mathbb{R}^n$:

- $\|x\|_1 = |x_1| + \cdots + |x_n|$
- $\|x\|_2 = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{1/2}$
- $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$

A matrix $A \in \mathbb{R}^{n\times n}$ corresponds to a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. If $\|x\|$ denotes a vector norm for $x \in \mathbb{R}^n$ we can define the **matrix norm** $\|A\|$ as follows:

$$\|A\| = \sup_{x\in\mathbb{R}^n} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|=1}} \|Ax\|$$

Note that $S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is a compact set, hence there exists a point $x \in S$ with $\|Ax\| = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|=1}} \|Ax\|$, and we can write "max" instead of "sup".

Lemma 1. *For $E \in \mathbb{R}^{n\times n}$ with $\|E\| < 1$ the matrix $I + E$ is invertible and* $\left\|(I + E)^{-1}\right\| \leq \dfrac{1}{1 - \|E\|}$.
Proof: $\|x\| = \|(1 + E)x - Ex\| \leq \|(1 + E)x\| + \|E\| \, \|x\|$, *hence* $\|(I + E)x\| \geq (1 - \|E\|) \, \|x\|$.

Convergence orders for iterative methods

We want to approximate $x_* \in \mathbb{R}^n$. An iterative method gives a sequence $x_0, x_1, x_2, x_3, \ldots$.

We say the method **converges** if $\lim\limits_{k\to\infty} x_k = x_*$. We can specify more precisely how quickly it converges by looking at quotients

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^\alpha}$$

for $\alpha \geq 1$. This leads to so-called **"q-orders" of convergence** (where "q" stands for quotient):

- the method converges with **q-order 1** or **q-linearly** if there exists $C < 1$ such that for all k

$$\|x_{k+1} - x_*\| \leq C \, \|x_k - x_*\|$$

- the method converges with **q-order 2** or **q-quadratically** if $\lim_{k\to\infty} x_k = x_*$ and there exists C such that for all k

$$\|x_{k+1} - x_*\| \leq C \, \|x_k - x_*\|^2$$

the method converges **q-superlinearly** if

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0$$

Example: We consider a nonlinear equation $f(x) = 0$ where $x \in D \subset \mathbb{R}$. Assume

- $f(x_*) = 0$
- the derivatives f' and f'' exist around x_* and are continuous
- $f'(x_*) \neq 0$

Then there exists $\epsilon > 0$ such that

- for $|x_0 - x_*| < \epsilon$ the **Newton method** converges **q-quadratically**
- for $|x_0 - x_*| < \epsilon$ and $|x_1 - x_*| < \epsilon$ the **secant method** converges with **q-order** $\alpha = \frac{\sqrt{5}+1}{2} \approx 1.618$. Hence it converges **q-superlinearly**.

In many cases we cannot prove that the error improves for every single step from k to $k+1$, but we still have upper bounds

$$\|x - x_*\| \leq E_k$$

where E_k converges with a certain q-order to zero. This leads to so-called **"r-orders" of convergence**:

- a method converges with **r-order 1** or **r-linearly** if $\|x - x_*\| \leq E_k$ where E_k converges with q-order 1 to 0
- a method converges with **r-order 2** or **r-quadratically** if $\|x - x_*\| \leq E_k$ where E_k converges with q-order 2 to 0
- a method converges **r-superlinearly** if $\|x - x_*\| \leq E_k$ where E_k converges with superlinearly to 0

Example: We consider a nonlinear equation $f(x) = 0$ where

- f is continuous on the interval $[a_0, b_0]$
- $f(a_0) \cdot f(b_0) < 0$

Then the **bisection method** gives a sequence of intervals $[a_k, b_k]$. Let $c_k = (a_k + b_k)/2$. Then the sequence c_k converges **r-linearly** to a point x_* with $f(x_*) = 0$:

$$|c_k - x_*| \leq E_k = \frac{b_0 - a_0}{2^{k+1}}$$

Note that in general we do *not* have q-linear convergence of c_k for the bisection method.

Derivatives

Let X and Y denote normed vector spaces. Consider a mapping f from $D \subset X$ to Y. We say that f is Frechet differentiable at a point $x_0 \in D$ if there exists a linear mapping $A\colon X \to Y$ such that

$$\frac{\|f(x) - f(x_0) - A(x - x_0)\|}{\|x - x_0\|} \to 0 \qquad x \to x_0$$

Lipschitz conditions

We say a function f satisfies a **Lipschitz condition with constant** L on D if

$$\|f(x) - f(y)\| \leq \|y - x\| \qquad \text{for all } x, y \in D$$

Lemma 2. *Assume*

- *D is convex*
- *$\nabla f(x)$ exists and is continuous on D*
- *$\|\nabla f(x)\| \leq L$ for $x \in D$*

  Then f satisfies a Lipschitz condition on D with constant L on D.

Taylor remainder terms

For a function $f\colon D \to \mathbb{R}$ with $D \subset \mathbb{R}$ we have estimates

$$|f(x+h) - f(x)| \le \left(\max_{u \in \mathrm{conv}(x,y)} |f'(u)| \right) |h| \qquad \text{if } f \in C^1(D)$$

$$|f(x+h) - f(x) - f'(x)h| \le \left(\max_{u \in \mathrm{conv}(x,y)} |f''(u)| \right) \frac{1}{2} |h|^2 \qquad \text{if } f \in C^2(D).$$

We would like to have analogous estimates for a function $f\colon D \to \mathbb{R}^n$ with $D \subset \mathbb{R}^n$. Assume that$f \in C^1(D)$. Then the derivative exists in the Frechet sense:

$$f(x+h) = f(x) + Df(x)\langle h\rangle + R, \qquad \|R\| = o(\|h\|) \quad \text{for } h \to 0$$

where $Df(x)\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear mapping $h \mapsto Df(x)\langle h\rangle$. Since linear mappings correspond to matrices in $\mathbb{R}^{n\times n}$ we use the notation

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}, \qquad Df(x) = f'(x) = \begin{bmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_n(x) & \cdots & \partial_n f_n(x) \end{bmatrix}$$

$$y = Df(x)\langle h\rangle \iff y_i = \sum_{j=1}^n \partial_j f_i(x) h_j$$

$$|y_i| \le \left(\sum_{j=1}^n |\partial_j f_i(x)| \right) \|h\|_\infty, \qquad \|y\|_\infty \le \max_{i=1,\ldots,n} \left(\sum_{j=1}^n |\partial_j f_i(x)| \right) \|h\|_\infty$$

$$\|Df(x)\| = \max_{i=1,\ldots,n} \left(\sum_{j=1}^n |\partial_j f_i(x)| \right)$$

with $\partial_i = \frac{\partial}{\partial x_i}$. Assume that D is convex and $x, y \in D$. Let $h := y - x$ and $\mathrm{conv}(x,y) \subset D$ denote the line segment betwen x and y. We have by the fundamental theorem of calculus and the chain rule for

$$F(t) := f(x+th), \qquad F'(t) = Df(x+th)\langle h\rangle$$

$$f(y) - f(x) = F(1) - F(0) = \int_0^1 F'(t)dt = \int_0^1 f'(x+th)h\,dt$$

$$\|f(y) - f(x)\| \le \int_0^1 \|f'(x+th\| \, \|h\| \, dt \le \left(\max_{u \in \mathrm{conv}(x,y)} \|f'(u)\| \right) \|y - x\|.$$

This shows that the Lipschitz constant for f in D is bounded by $\max_{u\in D} \|f'(u)\|$.

Note that we first choose a vector norm, e.g., $\|v\|_\infty = \max_{j=1,\ldots,n} |v_j|$. For $A \in \mathbb{R}^{n\times n}$ this induces a matrix norm:

$$\|A\|_\infty = \max_{\|v\|_\infty = 1} \|Av\|_\infty = \max_i \sum_j |A_{ij}|.$$

$$\|f'(x)\|_\infty = \max_{\|v\|_\infty = 1} \|f'(x)v\| = \max_i \sum_j |\partial_j f_i|.$$

We can read this as an estimate for the Taylor remainder term of order 1: $f(y) = f(x) + R_1$, $\|R_1\| \le \cdots$.

Now we want to consider the next Taylor term and remainder: $f(y) = f(x) + f'(x)(y-x) + R_2$:

$$R_2 = f(y) - f(x) - f'(x)(y-x) = \int_0^1 \left[f'(x + t(y-x)) - f'(x) \right](y-x)\,dt.$$

Assume that f' satisfies a Lipschitz condition for $x, y \in D$:

$$\|f'(y) - f'(x)\| \le \gamma \|y - x\| \quad \text{for } x, y \in D,$$

then we obtain

$$\|R_2\| \le \int_0^1 \gamma t \|y-x\|^2 \, dt = \frac{\gamma}{2} \|y-x\|^2 .$$

Now we assume that $f \in C^2(D)$: For a fixed $u \in \mathbb{R}^n$ let $G(x) := Df(x) \langle u \rangle$. Then $DG(x) \langle v \rangle = D^2 f(x) \langle u, v \rangle$ where

$$y = D^2 f(x) \langle u, v \rangle \iff y_i = \sum_{j,k=1}^{n} (\partial_{jk} f_i) \, u_j v_k$$

$$|y_i| \le \left(\sum_{j,k=1}^{n} |\partial_{jk} f_i| \right) \|u\|_\infty \|v\|_\infty , \qquad \|y\|_\infty \le \max_{i=1,\dots,n} \left(\sum_{j,k=1}^{n} |\partial_{jk} f_i| \right) \|u\|_\infty \|v\|_\infty$$

$$\left\| D^2 f(x) \right\|_\infty = \max_{i=1,\dots,n} \left(\sum_{j,k=1}^{n} |\partial_{jk} f_i| \right) .$$

Now we have for any $u \in \mathbb{R}^n$ and $h = y - x$

$$G(t) := Df(x+th) \langle u \rangle , \qquad G'(t) = D^2 f(x+th) \langle u, h \rangle$$

$$[Df(y) - Df(x)] \langle u \rangle = G(1) - G(0) = \int_0^1 G'(t) dt = \int_0^1 D^2 f(x+th) \langle u, h \rangle \, dt$$

$$\|Df(y) - Df(x) \langle u \rangle\| \le \int_0^1 \left\| D^2 f(x+th) \right\| \|u\| \|h\| \, dt, \qquad \|Df(y) - Df(x)\| \le \left(\max_{z \in D} \left\| D^2 f(z) \right\| \right) \|y-x\| .$$

Local convergence results for fixed point iteration

Assume that D is open so that $x_* \in D$ implies that a ball $B_\delta \subset D$.

Theorem 3. *Assume $g\colon D \to \mathbb{R}^n$ with $D \subset \mathbb{R}^n$, and $g(x_*) = x_*$ for $x_* \in D$. If $g \in C^1(D)$ and $\|g'(x_*)\| < 1$ then the fixed point iteration $x^{k+1} := g(x^k)$ is* locally convergent*: There exists $\delta > 0$ so that for $\left\|x^0 - x_*\right\| < \delta$ there holds $\lim_{k\to\infty} x^k = x_*$.*

Proof: Since g' is continuous there exists $\delta > 0$, $q < 1$ so that $\|g'(x)\| \le q$ for $x \in B_\delta := \{x \mid \|x - x_*\| < \delta\}$. Assume that $x^k \in B_\delta$. Then

$$\left\|x^{k+1} - x_*\right\| = \left\|g(x^k) - g(x_*)\right\| \le \left(\max_{u \in B_\delta} \|g'(u)\| \right) \left\|x^k - x_*\right\| \le q \left\|x^k - x_*\right\| .$$

Hence $x^{k+1} \in B_\delta$. Then $x^0 \in B_\delta$ implies $\left\|x^k - x_*\right\| \le q^k \left\|x^0 - x_*\right\|$ and therefore $\lim_{k\to\infty} x^k = x_*$.

Remark: For a chosen vector norm (e.g., $\|\cdot\|_\infty$) and the induced matrix norm one may have $\|g'(x_*)\| > 1$, but for a different vector norm and the induced matrix norm one could still have $\|g'(x_*)\| > 1$. The *spectral radius* $\rho(A) := \max\{|\lambda_j(A)|\}$ denotes the maximum absolute values of eigenvalues of A and satisfies [see e.g. Stoer-Bulirsch Theorem (6.9.2)]

- $\rho(A) \le \|A\|$ for any choice of vector norm and induced matrix norm
- for any $\epsilon > 0$ there exists a vector norm so that its induced matrix norm satisfies $\|A\| \le \rho(A) + \epsilon$.

Therefore we can replace the condition $\|g'(x_*)\| < 1$ by $|\rho(g'(x_*))| < 1$ and the conclusion of the theorem still holds.

Theorem: Assume $g\colon D \to \mathbb{R}^n$ with $D \subset \mathbb{R}^n$, and $g(x_*) = x_*$ for $x_* \in D$. If $g'(x_*) = 0$ and g' is Lipschitz in D then the fixed point iteration $x^{k+1} := g(x^k)$ is *locally convergent* of order 2: There exists $\delta > 0$ so that for $\left\|x^0 - x_*\right\| < \delta$ there holds $\lim_{k\to\infty} x^k = x_*$ and

$$\left\|x^{k+1} - x_*\right\| \le C \left\|x^k - x_*\right\|^2 .$$

Proof: Assume that we have $\|f'(y) - f'(x)\| \le \gamma \|y - x\|$ for $x, y \in D$.

$$\left\|x^{k+1} - x_*\right\| = \left\|g(x^k) - g(x_*) - g'(x_*)(x^k - x_*)\right\| \le \frac{\gamma}{2} \left\|x^k - x_*\right\|^2 .$$

Newton method

We have local convergence of order 2 as long as $\nabla f(x_*)$ is a nonsingular matrix:

Theorem 4. *Assume that $f(x_*) = 0$ and*

- *∇f exists and satisfies a Lipschitz condition in a neighborhood of x_**
- *$\nabla f(x_*)$ is nonsingular*

Then there exists $\delta > 0$ such that for an initial guess x_0 with $\|x_0 - x_*\| \leq \delta$

- *the Newton iterates x_k exist for $k = 1, 2, 3, \ldots$ since $\nabla f(x_{k-1})$ is nonsingular*
- $\lim\limits_{k\to\infty} x_k = x_*$
- *the convergence is of q-order 2*

Note that we have with $d_k = x_{k+1} - x_k = -\nabla f(x_k)^{-1} f(x_k)$

$$f(x_{k+1}) - f(x_k) = \int_0^1 \nabla f(x_k + t \cdot d_k) d_k dt$$

and using $\nabla f(x_k) d_k = -f(x_k)$ that

$$f(x_{k+1}) = \int_0^1 \left[\nabla f(x_k + t \cdot d_k) - \nabla f(x_k)\right] d_k dt$$

implying

$$\|f(x_{k+1})\| \leq \int_0^1 \underbrace{\|\nabla f(x_k + t \cdot d_k) - \nabla f(x_k)\|}_{\leq Lt\|d_k\|} \|d_k\| \, dt \leq \frac{L}{2} \|d_k\|^2$$

$$f(x_{k+1}) = -\int_0^1 \left[\nabla f(x_k + t \cdot d_k) - \nabla f(x_k)\right] \nabla f(x_k)^{-1} f(x_k) dt = f(x_{k+1}) = -\int_0^1 \left[\nabla f(x_k + t \cdot d_k) \nabla f(x_k)^{-1} - I\right] f(x_k) dt$$

Newton-Kantorovich theorem

The Newton-Kantorovich theorem does **not** assume that a solution x_* with $f(x_*)$ exists.

Theorem 5. *Assume*

- $\|f(x_0)\| \leq C_0$, $\left\|\nabla f(x_0)^{-1}\right\| \leq C_1$
- *$\nabla f(x)$ is Lipschitz with constant L for $\|x - x_0\| \leq \rho$*

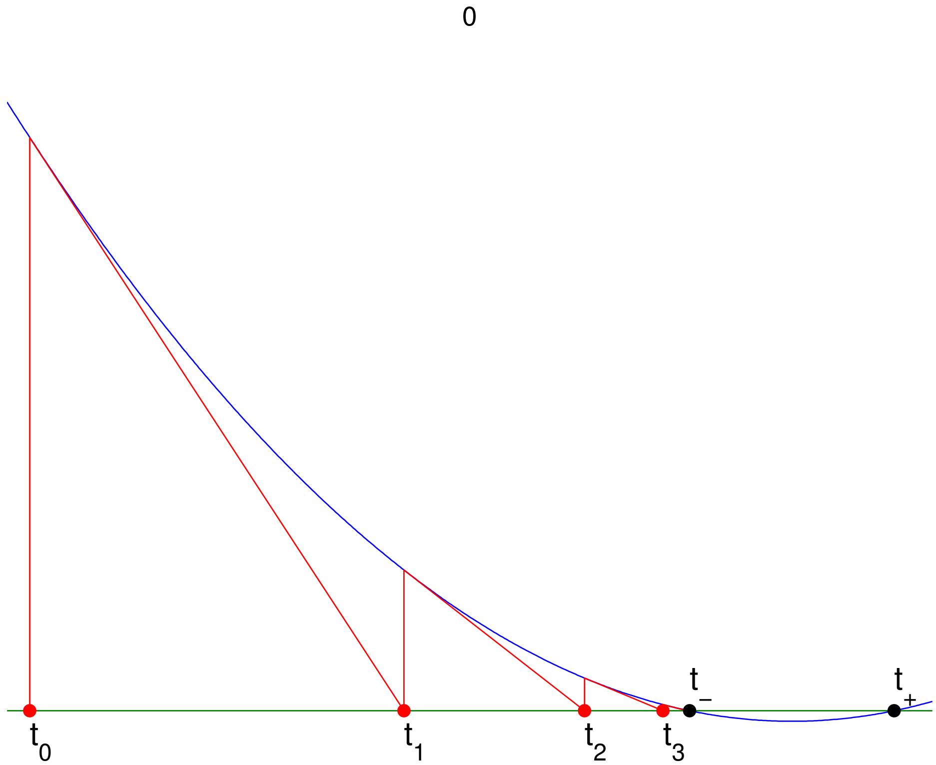
where

$$h_0 := C_0 C_1^2 L \leq \frac{1}{2}, \qquad \rho := \frac{1 - \sqrt{1 - 2h_0}}{C_1 L}$$

Then

- *there exists a unique solution $x_* \in \overline{B}_\rho(x_0)$ with $f(x_*) = 0$*
- *the Newton iterates x_k are well-defined and converge to x_**
- *the convergence is r-quadratic for $h_0 < \frac{1}{2}$, and r-linear for $h_0 = \frac{1}{2}$.*

In preparation for the proof we consider the Newton method for the quadratic equation $f(t) = t^2 - 2t + 2h_0 = 0$:



Lemma 6. *Assume $\omega_0 \ge 0$ and $h_0 \le \frac{1}{2}$ and define ω_k, h_k for $k = 1, 2, \dots$ by*

$$\omega_{k+1} := \frac{\omega_k}{1 - h_k}, \qquad h_{k+1} := \frac{h_k^2}{2(1-h_k)^2} \tag{1}$$

Then

$$\sum_{j=0}^{\infty} \frac{h_k}{\omega_k} = \frac{1 - \sqrt{1 - 2h_0}}{\omega_0}$$

and the sum converges q-quadratically for $h_0 < \frac{1}{2}$, and q-linearly for $h_0 = \frac{1}{2}$.

Proof: It is sufficient to consider the case $\omega_0 = 1$. Consider the quadratic equation $f(t) = t^2 - 2t + 2h_0 = 0$ which has the two roots $t_\pm = 1 \pm \sqrt{1 - 2h_0}$ with $0 < t_- < 1 < t_+$ for $h_0 < \frac{1}{2}$ and $0 < t_- = t_+ = 1$ for $h_0 = \frac{1}{2}$. We define the sequence t_k using the Newton method with $t_0 = 0$. We have $t_k < t_{k+1} < t_-$ for all k, this implies that $\lim_{k\to\infty} = t_-$. In the case of $h_0 < \frac{1}{2}$ we have quadratic convergence, in the case of $h_0 = \frac{1}{2}$ we have linear convergence. Note that for a quadratic function $f(t) = t^2 + a_1 t + a_0$ and its tangent line $p(t)$ at $t = c$ we have $f(t) - p(t) = (t - c)^2$ (since it must be a quadratic function with leading coefficient 1 and minimum 0 at $t = c$). Hence for $c = t_k$ we have by the definition of the Newton method $p(t_{k+1}) = 0$ and $(t_{k+1} - t_k)^2 = f(t_{k+1}) - p(t_{k+1}) = f(t_{k+1})$ for all k so that

$$t_{k+1} - t_k = -\frac{f(t_k)}{f'(t_k)} = -\frac{(t_k - t_{k-1})^2}{2t_k - 2} = \frac{1}{2}\frac{(t_k - t_{k-1})^2}{1 - t_k}. \tag{2}$$

We now want to show that for $k = 0, 1, 2, \dots$ we have

$$t_{k+1} - t_k = \omega_k^{-1} h_k, \qquad 1 - t_k = \omega_k^{-1} \tag{3}$$

We use induction: For $k = 0$ we have $t_0 = 0$,

$$t_1 - t_0 = -f(0)/f'(0) = -2h_0/(-2) = h_0 = \omega_0 h_0$$

and $1 - t_0 = 1 = \omega_0^{-1}$ as $\omega_0 = 1$.

Induction step: Assume that (3) holds.

We have

$$\omega_{k+1}^{-1} \overset{(D)}{=} \omega_k^{-1}(1-h_k) \overset{(I)}{=} \omega_k^{-1} - (t_{k+1}-t_k) \overset{(I)}{=} (1-t_k) - (t_{k+1}-t_k) = 1 - t_{k+1}$$

which is the second equation in (3) with $k+1$ instead of k.

We get from (2) that

$$t_{k+2} - t_{k+1} = \frac{1}{2}\frac{(t_{k+1}-t_k)^2}{1-t_{k+1}} = \frac{1}{2}\frac{(h_k/\omega_k)^2}{\omega_{k+1}^{-1}} = \omega_{k+1}^{-1}\frac{1}{2}\frac{h_k^2}{(\omega_k/\omega_{k+1})^2} = \omega_{k+1}^{-1}\frac{1}{2}\frac{h_k^2}{(1-h_k)^2} = \omega_{k+1}^{-1}h_{k+1}$$

Note that

$$\sum_{k=0}^{N}\frac{h_k}{\omega_k} = \sum_{k=0}^{N}(t_{k+1}-t_k) = t_{N+1} \to t_- = 1-\sqrt{1-2h_0} \qquad \text{as } N\to\infty$$

We use the following notation for open and closed balls:

$$B_r(x_0) := \{x \mid \|x-x_0\| < r\}, \qquad \overline{B}_r(x_0) := \{x \mid \|x-x_0\| \le r\}$$

The **Newton method** uses the iteration

$$x_{k+1} = x_k + d_k \qquad \text{with } d_k := -f'(x_k)^{-1}f(x_k)$$

Assumption: There are constants $\alpha, \omega_0 > 0$ such that $h_0 := \alpha\omega_0 \le \frac{1}{2}$ and

$$\left\|f'(x_0)^{-1}f(x_0)\right\| \le \alpha \tag{4}$$

$$\left\|f'(x_0)^{-1}\left(f'(y)-f'(x)\right)\right\| \le \omega_0\|y-x\| \quad \text{for } x,y\in D \tag{5}$$

$$\overline{B}_\rho(x_0) \subset D \qquad \text{with } \rho := \frac{1-\sqrt{1-2h_0}}{\omega_0}$$

Then:

- there exists a unique solution $x_* \in \overline{B}_\rho(x_0)$ with $f(x_*) = 0$
- the Newton iterates x_k are well-defined and converge to x_*
- the convergence is r-quadratic for $h_0 < \frac{1}{2}$, and r-linear for $h_0 = \frac{1}{2}$.

Induction: We claim that for $k = 0, 1, 2, \ldots$:

$$(A_k): \qquad \omega_k\|d_k\| \le h_k \tag{6}$$

$$(B_k): \qquad \left\|f'(x_k)^{-1}\left(f'(y)-f'(x)\right)\right\| \le \omega_k\|y-x\| \qquad \text{for } x,y\in D \tag{7}$$

$$x_k \in \overline{B}_\rho(x_0) \tag{8}$$

where

$$\omega_{k+1} := \frac{\omega_k}{1-h_k}, \qquad h_{k+1} := \frac{h_k^2}{2(1-h_k)^2} \tag{9}$$

Proof: For $k=0$ the statements (6), (7) are just the assumptions (4), (5).

Induction step: Assume that (6), (7), (8) and we have to prove the corresponding statements for $k+1$.

We first prove (8): From (A_k) and the Lemma we obtain that

$$\|x_{k+1}-x_0\| \le \sum_{j=0}^{k}\|d_j\| \le \sum_{j=0}^{\infty}\frac{h_j}{\omega_j} = \rho. \tag{10}$$

We then prove (6) for $k+1$: Let $M_k := f'(x_k)^{-1} f'(x_{k+1})$, then $f'(x_{k+1})^{-1} = M_k^{-1} f'(x_k)^{-1}$ and

$$\left\| f'(x_{k+1})^{-1} \left(f'(y) - f'(x) \right) \right\| \le \left\| M_k^{-1} \right\| \left\| f'(x_k)^{-1} \left(f'(y) - f'(x) \right) \right\| \overset{(B_k)}{\le} \left\| M_k^{-1} \right\| \omega_k \left\| y - x \right\|$$
$$\le \frac{1}{1-h_k} \omega_k \left\| y - x \right\| = \omega_{k+1} \left\| y - x \right\|$$

since

$$M_k = I + E, \qquad E := f'(x_k)^{-1} \left(f'(x_{k+1}) - f'(x_k) \right), \qquad \|E\| \overset{(B_k)}{\le} \omega_k \|d_k\| \overset{(A_k)}{\le} h_k$$
$$\left\| M_k^{-1} \right\| = \left\| (I+E)^{-1} \right\| \le \frac{1}{1 - \|E\|} \le \frac{1}{1-h_k}$$

We now prove (6) for $k+1$: With $x_{k+1} = x_k + d_k$ and $f(x_k) = -f'(x_k) d_k$ we have

$$d_{k+1} = -f'(x_{k+1})^{-1} f(x_{k+1}) = -f'(x_{k+1})^{-1} \left[f(x_{k+1}) - f(x_k) + f(x_k) \right] = -\int_0^1 f'(x_{k+1})^{-1} \left[f'(x_k + t d_k) - f'(x_k) \right] d_k dt$$

$$\|d_{k+1}\| \overset{(B_{k+1})}{\le} \int_0^1 \omega_{k+1} t \|d_k\| \|d_k\| \, dt = \frac{1}{2} \omega_{k+1} \frac{h_k^2}{\omega_k^2}$$

$$\omega_{k+1} \|d_{k+1}\| \le \frac{1}{2} \frac{h_k^2}{(\omega_k / \omega_{k+1})^2} = \frac{1}{2} \frac{h_k^2}{(1-h_k)^2} = h_{k+1}.$$

This concludes the induction proof. We now obtain from (10) that x_k is a Cauchy sequence and therefore has a limit $x_* \in B_\rho(x_0) \subset D$. By taking the limit in $f'(x_k)(x_{k+1} - x_k) = -f(x_k)$ we obtain by the continuity of f' and f that $f(x_*) = 0$.

We skip the proof of the uniqueness.