2.4 Das Newton–Verfahren

Ziel: Wir suchen die Nullstellen einer Funktion $\mathbf{f} : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}$$

- Wir kennen bereits die Fixpunktiteration

$$\mathbf{x}^{k+1} := \Phi(\mathbf{x}^k)$$

  mit Startwert $\mathbf{x}^0$ und Iterationsvorschrift $\Phi : \mathbb{R}^n \to \mathbb{R}^n$.

- Konvergenzaussagen liefert der Banachsche Fixpunktsatz.

Vorteil: Dieses Verfahren ist **ableitungsfrei**.

Nachteile:

- das numerische Verfahren konvergiert zu langsam (nur linear),
- es gibt keine eindeutige Iterationsvorschrift.

Banachsche Fixpunktsatz (BFS)

$f(x) = 0$ $\quad \boxed{x = x + f(x) = \Phi_1(x)}$

wähle x^0

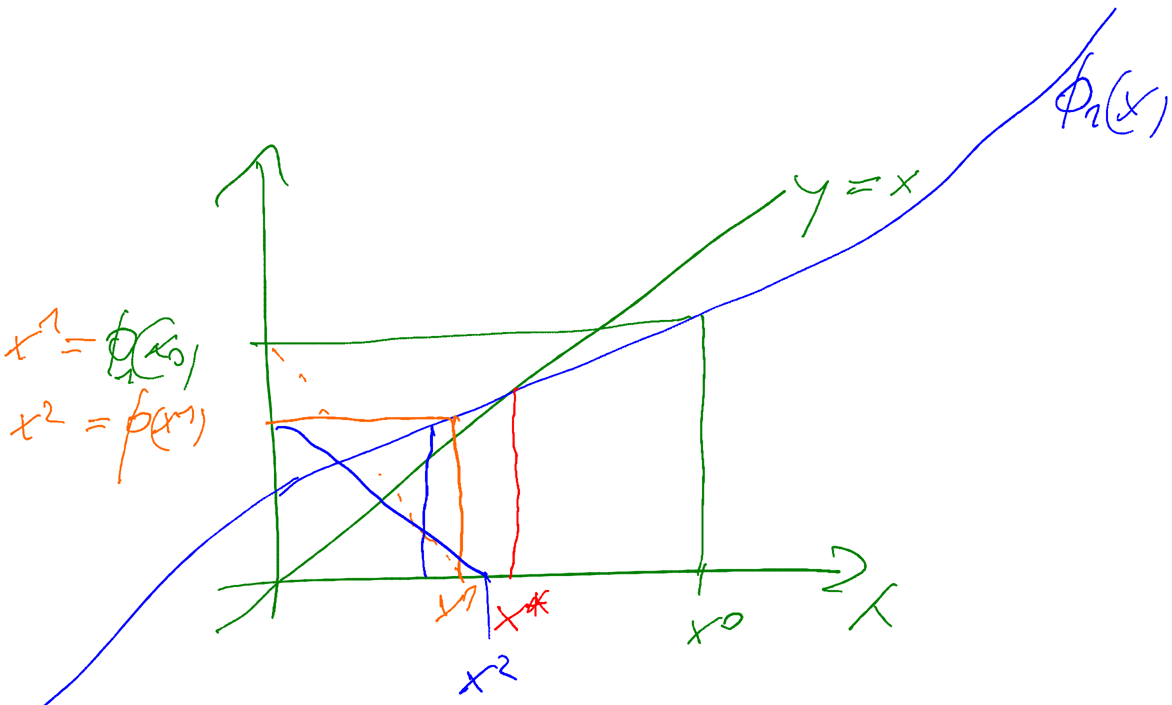
$$x^1 = \Phi_1(x_0)$$
$$x^2 = \Phi_1(x^1)$$
$$\vdots$$

Hoffnung: $x^n \xrightarrow{n \to \infty} x^*$ $\quad x^* = \Phi_1(x^*)$

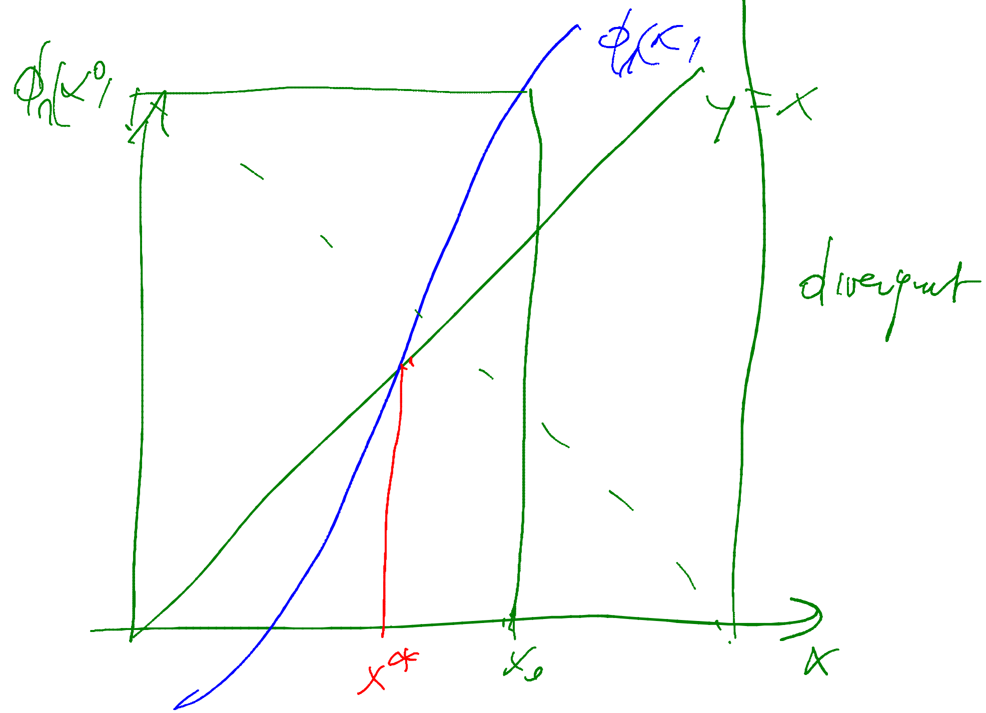
BFS $\quad \| \Phi_1(y) - \Phi_1(x) \| \leq L \| x - y \| \quad \forall x, y \in D$

Lipschitzbedingung $\Longleftarrow$ Φ_1 C^1 Funktion

$L < 1$ (Kontraktion) $\Longrightarrow$ $x^n \xrightarrow{n \to \infty} x^*$



$$|\phi_1'(x^*)| < 1$$



$$|\phi_1'(x^*)| > 1$$

–) konvergiert langsam

+) $x^{k+1} = \phi_1(x^k)$ keine Ableitungen

$\pm$) $x = x + f(x) = \phi_1(x)$

$x = x - f(x) = \phi_2(x)$

$x = \dfrac{x}{1 + f(x)} = \phi_3(x)$

Zur Konstruktion des Newton–Verfahrens.

Ausgangspunkt: Gegeben sei eine $\mathcal{C}^1$–Funktion $\mathbf{f} : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$ offen.

Wir suchen eine Nullstelle von $\mathbf{f}$, d.h ein $\mathbf{x}^* \in D$ mit

$$\mathbf{f}(\mathbf{x}^*) = \mathbf{0}$$

Konstruktion des Newton–Verfahrens:

Die Taylor–Entwicklung von $\mathbf{f}(\mathbf{x})$ um einen Startwert $\mathbf{x}^0$ lautet

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^0) + \mathbf{Jf}(\mathbf{x}^0)(\mathbf{x} - \mathbf{x}^0) + \mathbf{o}(\|\mathbf{x} - \mathbf{x}^0\|)$$

Setzen wir $\mathbf{x} = \mathbf{x}^*$, so folgt

$$\mathbf{Jf}(\mathbf{x}^0)(\mathbf{x}^* - \mathbf{x}^0) \approx -\mathbf{f}(\mathbf{x}^0)$$

Eine Näherungslösung für $\mathbf{x}^*$ ist dann $\mathbf{x}^1$, $\mathbf{x}^1 \approx \mathbf{x}^*$, die Lösung des linearen Gleichungssystems

$$\mathbf{Jf}(\mathbf{x}^0)(\mathbf{x}^1 - \mathbf{x}^0) = -\mathbf{f}(\mathbf{x}^0)$$

Newton $n=1$

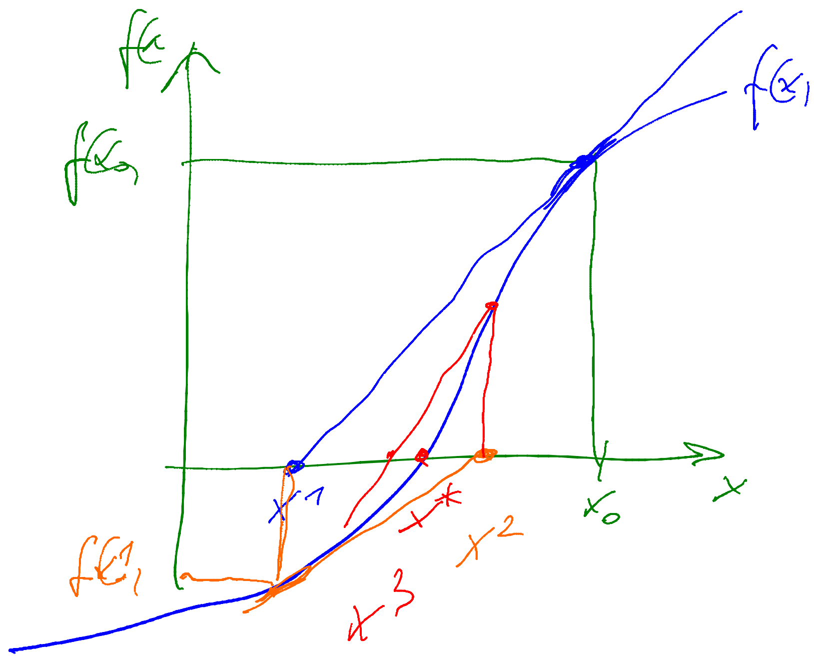
$f(x^*)=0$

$f(x) = f(x_0) + f'(x_0)(x-x_0) + \dots$

$y = f(x_0) + f'(x_0)(x-x_0)$ $\quad y'(x_0) = f'(x_0)$

$y=0 \iff x = x_0 - \frac{f(x_0)}{f'(x_0)}$

$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$



$n>1$ $\quad f(x) = f(x_0) + Jf(x_0)(x-x_0) + \dots$

$y = f(x_0) + Jf(x_0)(x-x_0)$

$y=0 \iff$ i) $x = x_0 - (Jf(x_0))^{-1} f(x_0)$ invertieren einer Matrix

ii) $Jf(x_0)(x-x_0) = -f(x_0)$ lineares Gleichungssystem für $x-x_0$ ✓

Das Newton–Verfahrens als Algorithmus.

Das Newton–Verfahren kann man somit wie folgt als Algorithmus formulieren.

Algorithmus (Newton–Verfahren)**:**

(1) FOR $k = 0, 1, 2, \dots$

(2a) Löse $\boxed{\mathbf{J}f(\mathbf{x}^k) \cdot \underline{\Delta\mathbf{x}^k} = -f(\mathbf{x}^k);}$

(2b) Setze $\mathbf{x}^{k+1} = \mathbf{x}^k + \underline{\Delta\mathbf{x}^k}$**;**

- Man löst in *jedem* Newton–Schritt eine lineares Gleichungssystem.
- Die Lösung $\Delta\mathbf{x}^k$ heißt Newton–Korrektur.
- Das Newton–Verfahren ist skalierungsinvariant.

Skalierungsinvarianz des Newton–Verfahrens.

Satz: Das Newton–Verfahren ist invariant unter linearen Transformationen der Form

$$\mathbf{f}(\mathbf{x}) \to \mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{f}(\mathbf{x}) \qquad \text{für } \mathbf{A} \in \mathbb{R}^{n\times n} \text{ regulär,}$$

d.h. die Iterierten für **f** und **g** sind in diesem Fall identisch.

Beweis: Bildet man das Newton–Verfahren für $\mathbf{g}(\mathbf{x})$, so lautet die Newton–Korrektur

$(*) = (J Af(x^k))^{-1}(x^{k+1}-x^k) = (Jg(x^k))^{-1}(x^{k+1}-x^k) = -g(x^k) = \boxed{-Af(x^k)}$

$\boxed{A\,Jf(x^k)(x^{k+1}-x^k)} = (*)$

$A^{-1} \,|\quad Jf(x^k)^{-1}(x^{k+1}-x^k) = -f(x^k)$

$$\begin{aligned}
\Delta \mathbf{x}^k &= -(\mathbf{Jg}(\mathbf{x}^k))^{-1}\cdot \mathbf{g}(\mathbf{x}^k)\\
&= -(\mathbf{AJf}(\mathbf{x}^k))^{-1}\cdot \mathbf{Af}(\mathbf{x}^k)\\
&= -(\mathbf{Jf}(\mathbf{x}^k))^{-1}\cdot \mathbf{A}^{-1}\mathbf{A}\cdot \mathbf{f}(\mathbf{x}^k)\\
&= -(\mathbf{Jf}(\mathbf{x}^k))^{-1}\cdot \mathbf{f}(\mathbf{x}^k)
\end{aligned}$$

womit die Newton–Korrektur von **f** und **g** übereinstimmen.

Bei gleichem Startwert $\mathbf{x}^0$ stimmen somit auch alle Iterierten $\mathbf{x}^k$ überein.

Zur lokalen Konvergenz des Newton–Verfahrens.

Satz: Sei $\mathbf{f} : D \to \mathbb{R}^n$ eine $\mathcal{C}^1$–Funktion, $D \subset \mathbb{R}^n$ offen und konvex. Sei $\mathbf{x}^* \in D$ eine Nullstelle von $\mathbf{f}$, d.h. $\boxed{\mathbf{f}(\mathbf{x}^*) = 0}$.

Weiterhin sei die Jacobi–Matrix $\boxed{\mathbf{Jf}(\mathbf{x}) \text{ regulär}}$ für $\mathbf{x} \in D$, und es gelte eine Lipschitz–Bedingung für $\phi(z) = (Jf(x))^{-1} Jf(z)$

$$\|(\mathbf{Jf}(\mathbf{x})^{-1}(\mathbf{Jf}(\mathbf{y}) - \mathbf{Jf}(\mathbf{x}))\| \leq L\|\mathbf{y} - \mathbf{x}\| \qquad \text{für alle } \mathbf{x}, \mathbf{y} \in D,$$

mit einem $L > 0$. Dann ist das Newton–Verfahren für alle Startwerte $\mathbf{x}^0 \in D$ mit

$$\boxed{\|\mathbf{x}^0 - \mathbf{x}^*\| < \frac{2}{L} =: r} \quad \text{und} \quad K_r(\mathbf{x}^*) \subset D$$

wohldefiniert mit $\mathbf{x}^k \in K_r(\mathbf{x}^*)$, $k = 0, 1, 2, \ldots$, und die Newton–Iterierten $\mathbf{x}^k$ konvergieren quadratisch gegen $\mathbf{x}^*$, d.h.

$$\boxed{\|\mathbf{x}^{k+1} - \mathbf{x}^*\| \;\leq\; \frac{L}{2}\, \|\mathbf{x}^k - \mathbf{x}^*\|^2}$$

Weiterhin ist $\mathbf{x}^*$ die eindeutige Nullstelle von $\mathbf{f}(\mathbf{x})$ innerhalb der Kugel $K_r(\mathbf{x}^*)$.

$f(x^*) = 0$

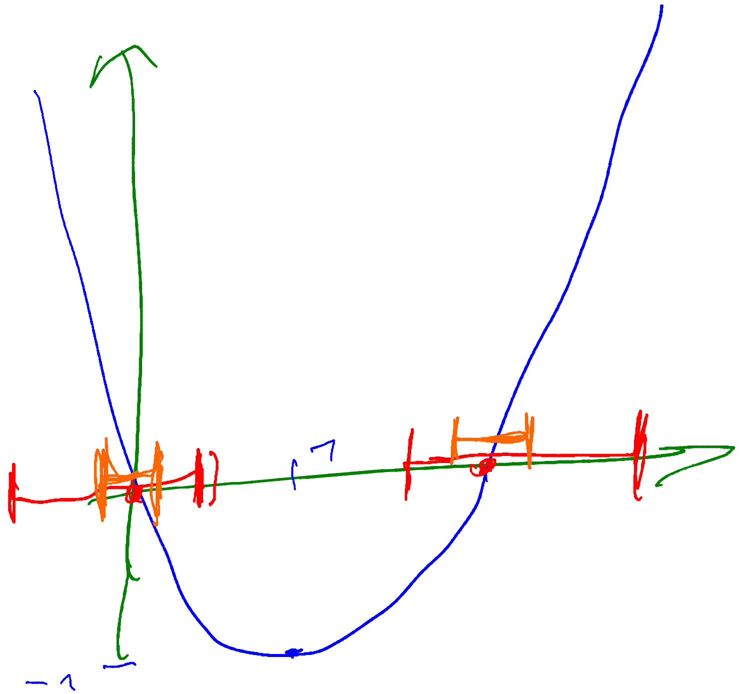
$f(x) = (x-1)^2 - 1$

$Jf(x) = 2(x-1)$

$(Jf(x))^{-1} = \dfrac{1}{2(x-1)}$

$\left|(Jf(x))^{-1}(Jf(y) - Jf(x))\right| = \left|\dfrac{1}{2(x-1)}(2(y-1) - 2(x-1))\right|$

$= \left|\dfrac{1}{x-1}\right| \, |y-x| \leq L\,|x-y|$



$x^* = 0$

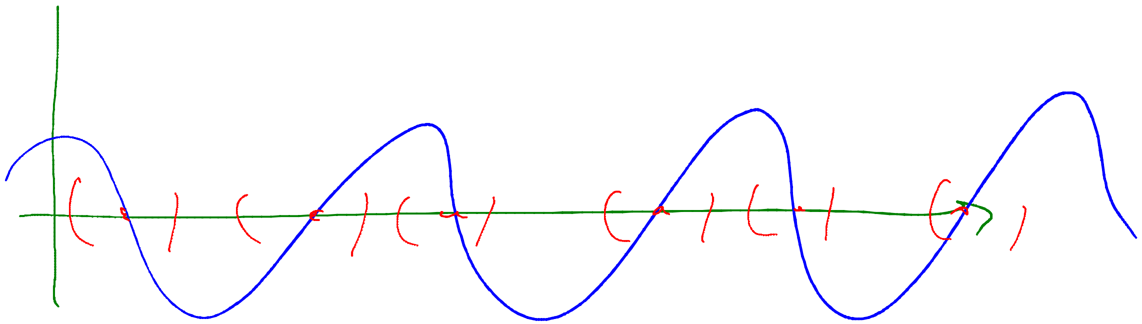
$\left|\dfrac{1}{x-1}\right| < 2$ für $x \in (-\infty, \tfrac{1}{2})$ $L=2$ $|x^0 - x^*| = |x^0| < \tfrac{2}{L} = 1$

$\left|\dfrac{1}{x-1}\right| < 10$ für $x \in (-\infty, \tfrac{9}{10})$ $L = 10$ $|x^0| < \tfrac{2}{10} = \tfrac{1}{5}$

$\left|\dfrac{1}{x-1}\right| < 2$ $x \in (\tfrac{3}{2}, +\infty)$ $|x^0 - 2| < 1$

$x^* = 2$

$\left|\dfrac{1}{x-1}\right| < 10$ $x \in (\tfrac{11}{10}, \infty)$ $|x^0 - 2| < \tfrac{1}{5}$



<u>Beweis:</u>

$$g(t) = (Jf(x))^{-1} f(x+t(y-x)) \qquad g: \mathbb{R}^1 \to \mathbb{R}^n$$

$$g(0) = (Jf(x))^{-1} f(x)$$

$$g(1) = (Jf(x))^{-1} f(y)$$

$$g'(t) = (Jf(x))^{-1} \overbrace{Jf(x+t(y-x)) \cdot (y-x)}^{\text{Kettenregel}}$$

$$\|g'(t) - g'(0)\| = \|(Jf(x))^{-1} Jf(x+t(y-x))(y-x) - (Jf(x))^{-1} Jf(x)(y-x)\|$$

$$= \|(Jf(x))^{-1} [Jf(x+t(y-x)) - Jf(x)](y-x)\|$$

$$\leq \underbrace{\|(Jf(x))^{-1} [Jf(x+t(y-x)) - Jf(x)]\|}_{\leq L\|x+t(y-x) - x\|} \|y-x\|$$

$$\leq L\|x+t(y-x) - x\| \; \|y-x\| = Lt\|y-x\|\|y-x\| \overset{(*)}{=} Lt\|y-x\|^2$$

$$\|Jf(x)^{-1}(f(y) - f(x) - Jf(x)(y-x))\| = \|g(1) - g(0) - g'(0)\| = \left\|\int_0^1 g'(t)\,dt - g'(0)\right\| = \left\|\int_0^1 (g'(t) - g'(0))\,dt\right\|$$

$$\leq \int_0^1 \|g'(t) - g'(0)\|\,dt \overset{(*)}{\leq} \int_0^1 Lt\|y-x\|^2\,dt = L\|y-x\|^2 \int_0^1 t\,dt = \frac{L}{2}\|y-x\|^2$$

$$x^{k+1} - x^* \underset{\text{Newton}}{=} x^k - (Jf(x^k))^{-1} f(x^k) - x^*$$

$$\approx Jf(x^k)^{-1}\left(f(x^*) - f(x^k)\right) + Jf(x^k)^{-1} Jf(x^k)\left(x^k - x^*\right)$$

$$= (Jf(x^k))^{-1}\left[\underbrace{f(x^*)}_{=0} - f(x^k) - Jf(x^k)(x^* - x^k)\right]$$

$$x^* = y \qquad x^k = x$$

$$\boxed{\|x^{k+1} - x^*\|} = \underbrace{\left\| Jf(x^k)^{-1}\left[f(x^*) - f(x^k) - Jf(x^k)(x^* - x^k)\right]\right\|}$$

$$\boxed{\leq \frac{L}{2} \|x^k - x^*\|^2}$$

Das gedämpfte Newton–Verfahren.

Weitere Beobachtungen:

- Das Newton–Verfahren konvergiert zwar quadratisch, aber nur lokal.
- Globale Konvergenz kann ggf. durch einen Dämpfungsterm erreicht werden:

Algorithmus (Gedämpftes Newton–Verfahren)**:**

(1) FOR $k = 0, 1, 2, \ldots$

(2a) Löse $\mathbf{Jf}(\mathbf{x}^k) \cdot \Delta\mathbf{x}^k = -f(\mathbf{x}^k)$**;**

(2b) Setze $\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \Delta\mathbf{x}^k$**;**

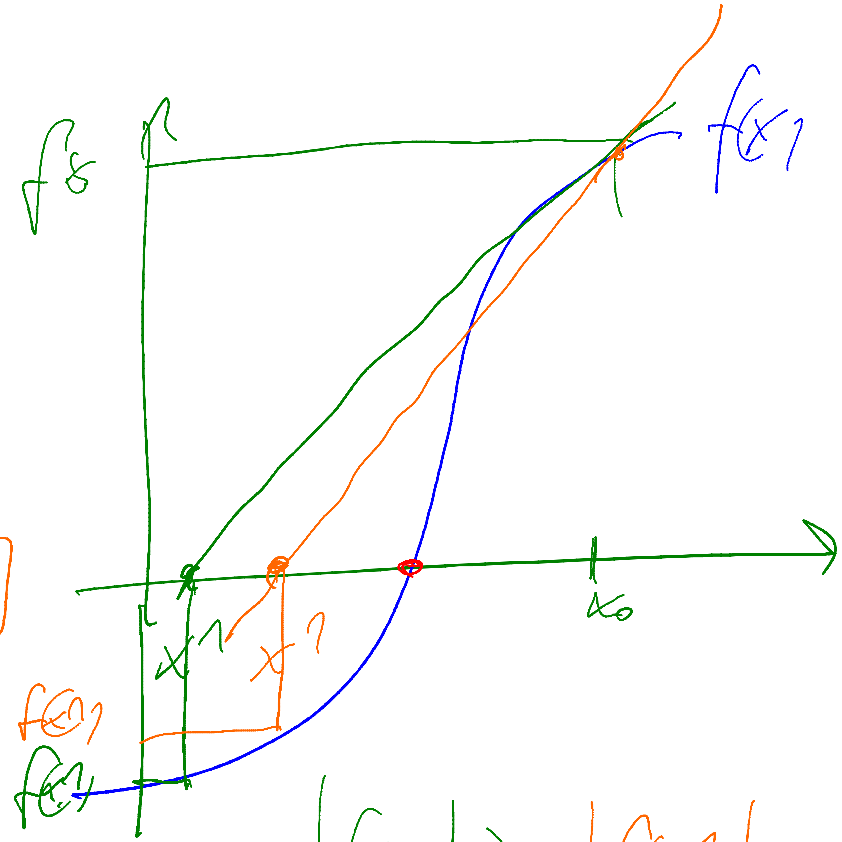
Frage: Wie wählt man die Dämpfungsfaktoren λ_k?

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$ Newton klassisch

$$x_1 = x_0 + \Delta x_0$$

$$x_1 = x_0 + \lambda \Delta x_0 \qquad \lambda \in (0,1]$$

$$\lambda = 1, \tfrac{1}{2}, \tfrac{1}{4}$$



$$|f(x^1)| > |f(x^1)|$$

Wahl des Dämpfungsparameters.

Strategie: Verwende eine Testfunktion $T(\mathbf{x}) = \|\mathbf{f}(\mathbf{x})\|$, womit gilt

$$T(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in D$$
$$T(\mathbf{x}) = 0 \quad \Leftrightarrow \quad f(\mathbf{x}) = \mathbf{0}$$

Wähle nun $\lambda_k \in (0,1)$ so, dass die Folge $T(\mathbf{x}^k)$ streng monoton fällt, d.h.

$$\|\mathbf{f}(\mathbf{x}^{k+1})\| < \|\mathbf{f}(\mathbf{x}^k)\| \qquad \text{für } k \geq 0.$$

In der Nähe der gesuchten Lösung $\mathbf{x}^*$ sollte $\lambda_k = 1$ gewählt werden, um (lokale) quadratische Konvergenz zu sichern.

Der folgende Satz garantiert die Existenz eines Dämpfungsparameters.

Satz: Sei $\mathbf{f}$ eine $\mathcal{C}^1$–Funktion auf der offenen und konvexen Menge $D \subset \mathbb{R}^n$. Für $\mathbf{x}^k \in D$ mit $\mathbf{f}(\mathbf{x}^k) \neq \mathbf{0}$ gibt es dann ein $\mu_k > 0$, sodass

$$\|\mathbf{f}(\mathbf{x}^k + \lambda \Delta x^k)\|_2^2 < \|\mathbf{f}(\mathbf{x}^k)\|_2^2 \qquad \text{für alle } \lambda \in (0, \mu_k).$$

gedämpft Newton klassisch Newton

Dämpfungsstrategie.

Für die **Startiteration** $k = 0$: Wähle $\lambda_0 \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots, \lambda_{min}\}$ möglichst groß, sodass gilt

$$\|\mathbf{f}(\mathbf{x}^0)\|_2 > \|\mathbf{f}(\mathbf{x}^0 + \lambda_0 \Delta\mathbf{x}^0)\|_2$$

Für **nachfolgende Iterationen** $k > 0$: Setze $\lambda_k = \lambda_{k-1}$.

IF $\|\mathbf{f}(\mathbf{x}^k)\|_2 > \|\mathbf{f}(\mathbf{x}^k + \lambda_k \Delta\mathbf{x}^k)\|_2$ **THEN**

- $\mathbf{x}^{k+1} := \mathbf{x}^k + \lambda_k \Delta\mathbf{x}^k$
- $\lambda_k := 2\lambda_k$, falls $\lambda_k < 1$.

ELSE

- Bestimme $\mu = \max\{\lambda_k/2, \lambda_k/4, \ldots, \lambda_{min}\}$ mit

$$\|\mathbf{f}(\mathbf{x}^k)\|_2 > \|\mathbf{f}(\mathbf{x}^k + \lambda_k \Delta\mathbf{x}^k)\|_2$$

- $\lambda_k := \mu$

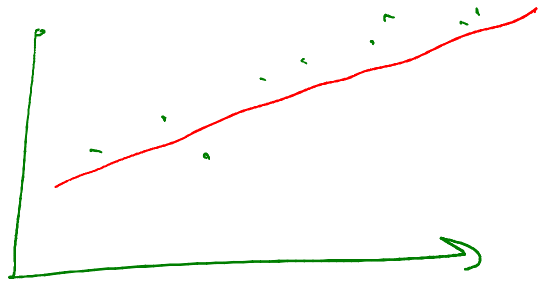
END

Ausgleichsproblem

$(y_i, x_i) \quad i = 1, \dots, m$

Modell $\quad \phi(x_1, \dots, x_n; \alpha_1, \dots, \alpha_n)$

z.B. $\quad \phi = A(x)\alpha$



Gerade $\quad y_i = \alpha_1 + \alpha_2 x_i$

$$A = \underbrace{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{m \times 2} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$\overbrace{2 \times m \;\; m \times 2}^{2 \times 2}$$

$$\|y - A(x)\alpha\|_2^2 \longrightarrow \min (\text{in } \alpha)$$

$$g(\alpha) = \|y - A\alpha\|_2^2 = (y - A\alpha)^T (y - A\alpha) = y^T y - y^T A\alpha - \alpha^T A^T y + \alpha^T A^T A \alpha$$

notwendige Bed. $\quad \text{grad}_\alpha\, g(\alpha) = 0$

$$\text{grad}_\alpha\, g(\alpha) = -y^T A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \underbrace{(1 \dots 1) A^T y}_{\underbrace{y^T A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{y^T A}} + \alpha^T \Big(A^T A + \underbrace{(A^T A)^T}_{A^T A} \Big) = +2\alpha^T A^T A - 2 y^T A = 0$$

$$A^T A \alpha = A y^T$$

$$\boxed{\alpha = (A^T A)^{-1} (A^T y)}$$

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \\ 1 & x_n \end{pmatrix} \qquad A^T y = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = n \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}$$

$$\overline{x} = \frac{1}{n} \sum x_i$$

$$\overline{y} = \frac{1}{n} \sum y_i$$

$$\overline{xy} = \frac{1}{n} \sum x_i y_i$$

$$\overline{x^2} = \frac{1}{n} \sum_i x_i^2$$

$$(A^T A)^{-1} = \frac{1}{n(\overline{x^2} - \overline{x}^2)} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

$$\alpha = \frac{1}{\cancel{n}(\overline{x^2} - \overline{x}^2)} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \begin{pmatrix} \cancel{n}\overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - \overline{x}^2} \begin{pmatrix} \overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy} \\ -\overline{x}\,\overline{y} + \overline{xy} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$