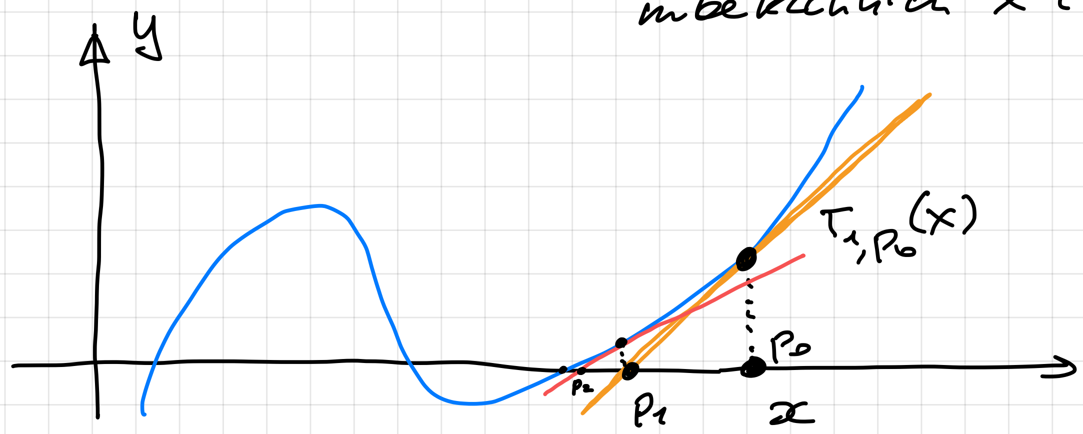
Agenda

+ Newtonverfahren für NGS und Optim.
- Taylorentwicklung für multivariate Funktionen
- Hinreichende und notwendige Bedingungen zweiter Ordnung
- Lineare Algebra reeller symmetrischer Matrizen

Newtonverfahren

$f(x) = 0$ Lösen in einem unbekannten $x \in \mathbb{R}$



- Wir linearisieren (durch das Taylor-Polynom erster Ordnung)
- Wir lösen die lineare Gleichung
- Die Lösung der Gleichung ist unsere nächste Approxim. einer Lösung der Nichtlinearen Gleichung $f(x) = 0$.

Analog, wenn man ein nichtlineares Gleichungssystem lösen will: $F(x) = 0$. $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ Vektor aus n reellwertigen Unbekannten.

$$F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

$\Updownarrow$

$$\begin{cases} f_1(x_1, \dots, x_n) = 0 \\ \vdots \\ f_n(x_1, \dots, x_n) = 0 \end{cases}$$

Linearisierung hier, am Punkt p_k:

$$F(x) \approx \underbrace{F(p_k)}_{\text{vektor}} + \underbrace{F'(p_k)}_{n \times n \text{ Matrix}} \cdot (x - p_k)$$

Komponentenweise ausgeschrieben:

$$f_i(x) \approx f_i(p_k) + \langle \nabla f_i(p_k), x - p_k \rangle$$

(die Linearisierung für die linke Seite der i-ten Gleichung).

Die nächste Approximation der nichtlinearen Gleichung $F(x)=0$ erhält man als die Lösung des LGS

$$F(p_k) + F'(p_k) \cdot (x - p_k) = 0$$

Wir lösen es bzgl. x auf:

$$F'(p_k) \cdot (x - p_k) = -F(p_k)$$

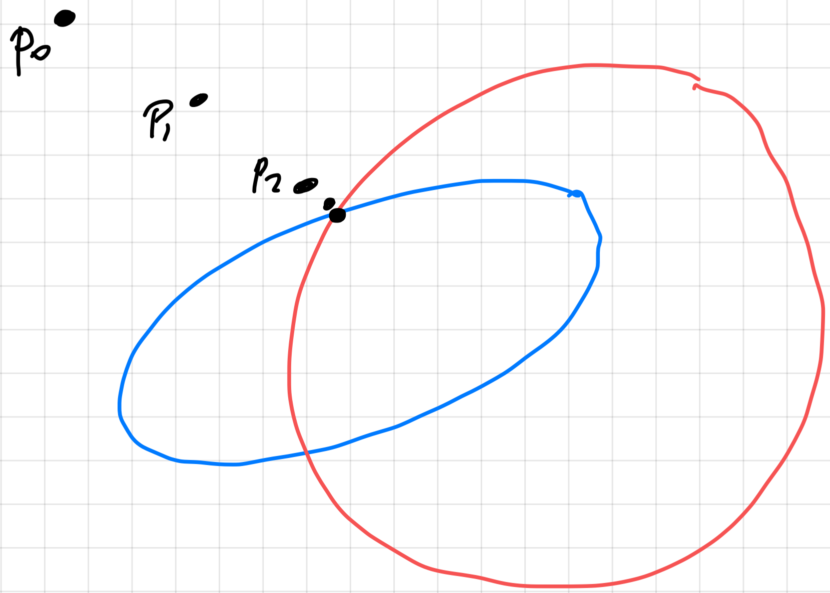
$$x - p_k = -(F'(p_k))^{-1} F(p_k)$$

$$x = p_k - (F'(p_k))^{-1} F(p_k)$$

D.h. $$p_{k+1} := p_k - (F'(p_k))^{-1} F(p_k)$$

$\uparrow$

die nächste Approximation der Lösung von $F(x)=0$



Newtonverfahren zur Optimierung

Zurück zur Dimension 1.

$$\min_{x \in \mathbb{R}} f(x)$$

Wir sind auf der Suche nach den lokalen Optima.

$f'(x) = 0$ ist eine notwendige Bedingung.

D.h. wir sind auf der Suche nach den kritischen Punkten.

Wir können $f'(x)$ numerisch mit dem Newtonverfahren lösen.

p_k – aktuelle Approximation

p_{k+1} ist das x aus der folgenden linearen Gleichung:

$$f'(p_k) + f''(p_k) \cdot (x - p_k) = 0$$

$$f''(p_k) \cdot (x - p_k) = -f'(p_k)$$

$$x - p_k = -f''(p_k)^{-1} \cdot f'(p_k)$$

$$x = p_k - f''(p_k)^{-1} \cdot f'(p_k)$$

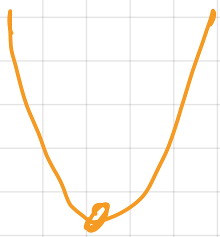
$$p_{k+1} := p_k - f''(p_k)^{-1} \cdot f'(p_k)$$

Das Newtonverfahren zur Optimierung von f.

Wir interpretieren dieses Verfahren wie folgt:

Das Taylorpolynom 2er Ordnung von f an der Stelle p_k:

$$t_k(x) := f(p_k) + f'(p_k) \cdot (x - p_k) + \frac{f''(p_k)}{2}(x - p_k)^2$$



Wo ist das Optimum von $t_k(x)$?

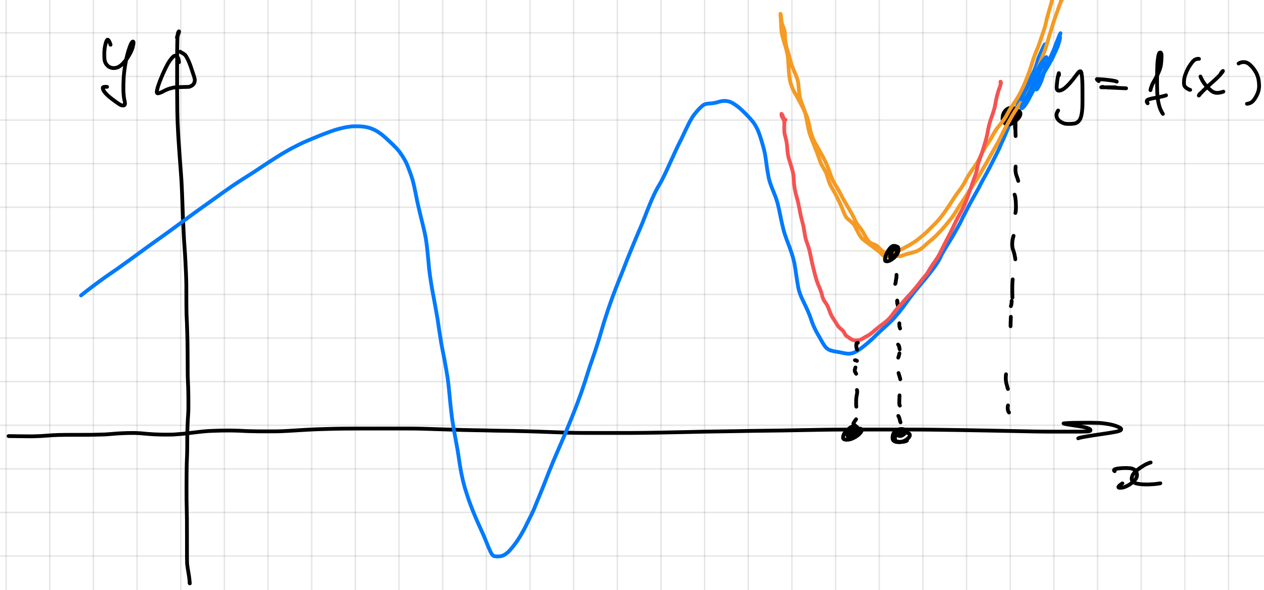
$$t_k'(x) = f'(p_k) + f''(p_k)\cdot(x-p_k)$$

Das Optimum ist die Lösung von $t_k'(x) = 0$

$$f'(p_k) + f''(p_k)\cdot(x-p_k) = 0$$

$$x = p_k - f''(p_k)^{-1}\cdot f'(p_k)$$

$$p_{k+1} = p_k - f''(p_k)^{-1}\cdot f'(p_k)$$

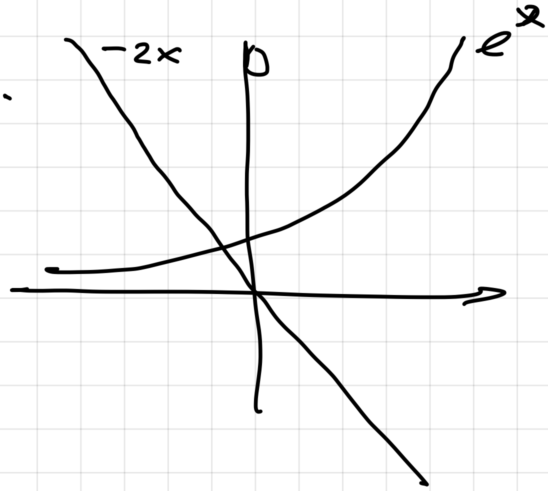


Bsp.

$$f(x) = e^x + x^2$$

$$f'(x) = e^x + 2x.$$

$$e^x = -2x$$



Newtonverfahren zur Optimierung in der Dimension n.

Wir sind auf der Suche nach den kritischen Punkten von $f: \mathbb{R}^n \to \mathbb{R}$, um die Optimierungsaufgabe $\min_{x \in \mathbb{R}^n} f(x)$ zu lösen.

Kritische Punkte sind die $x \in \mathbb{R}^n$, die die Gleichung

$$\nabla f(x) = 0 \quad \text{erfüllen.}$$

Es ist ein nichtlineares System mit n Unbekannten und n Gleichungen.

$$\begin{cases} \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) = 0 \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) = 0 \end{cases}$$

Wir können zu $F(x) = \nabla f(x)$ das Newtonverfahren anwenden:

$$p_{k+1} = p_k - \underbrace{(\nabla f)'}_{= \underbrace{\nabla^2 f}_{}}(p_k) \cdot \nabla f(p_k)$$

Was ist das für eine Matrix?

$$\nabla^2 f = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{i,j = 1, \dots, n}$$

im regulären Fall (wenn die zweiten Ableitungen alle stetig sind) ist das eine symmetrische Matrix.

$\nabla^2 f$ nennt man die Hesse-Matrix von f.

$$p_{k+1} = p_k - (\nabla^2 f)(p_k)^{-1} \cdot \nabla f(p_k).$$

Auch hier hat man eine analoge Interpretation wie in der Dimension 1:

Man approximiert $f(x)$ durch die quadratische Funktion:

$$\underbrace{f(p_k)}_{\text{konstant}} + \underbrace{\langle \nabla f(p_k), x-p_k \rangle}_{\text{linear}} + \frac{1}{2} \underbrace{\langle \nabla^2 f(p_k)(x-p_k), x-p_k \rangle}_{\text{quadratisch.}}$$

Es stellt sich heraus, dass diese quadratische Funktion ihr Optimum im Punkt

$$p_{k+1} = p_k - (\nabla^2 f(p_k))^{-1} \cdot \nabla f(p_k)$$

erreicht.

Taylorentwicklung II

Erinnerung an den univariaten Fall.

Wir brauchen den Entwicklungspunkt $a \in \mathbb{R}$ und die Ordnung der Entwicklung, wir nennen sie diesmal d.

Die allgemeine Formel für das Taylorpolynom:

$$T_d(x) = \sum_{k=0}^{d} \frac{1}{k!} f^{(k)}(a) \cdot (x-a)^k$$

Unter gewissen natürlichen Voraussetzungen an f

ist der Unterschied zwischen $f(x)$ und $T_d(x)$ nicht sehr groß, wenn x nah an a ist.

$$f(x) = T_d(x) + \underbrace{R_d(x)}_{\text{Restglied}}$$

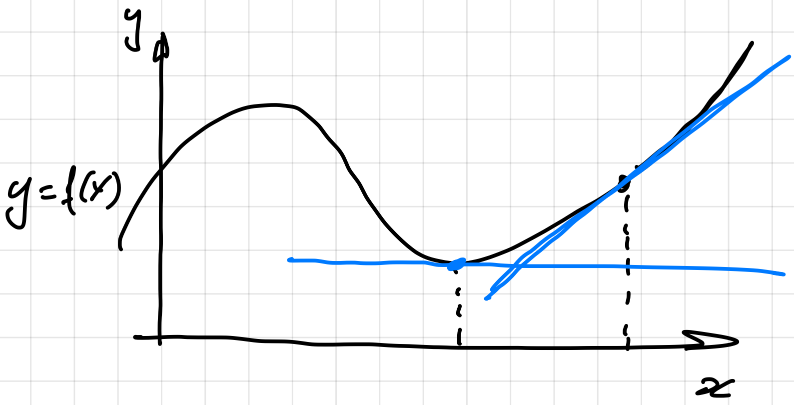
Restglied mit $R_d(x) = \mathcal{O}((x-a)^{d+1})$, $x \to a$

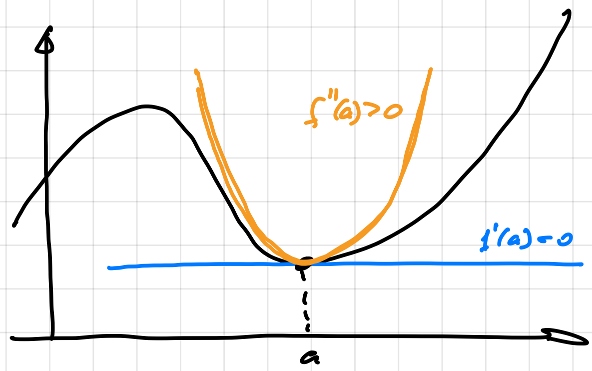
Rest der Ordnung $d+1$

Die Fälle $d=1$ und $d=2$ sind wichtig für das Lösen von Gleichungen (iterativ mit Newton) und für die Optimierung.

$$T_1(x) = f(a) + f'(a) \cdot (x-a)$$

$$T_2(x) = f(a) + f'(a) \cdot (x-a) + \frac{f''(a)}{2} \cdot (x-a)^2$$





Für ein lokales Maximum in a:

- $f'(a) = 0$ - notwendig (1. Ordnung)
- $f'(a) = 0$, $f''(a) < 0$ - hinreichend
- $f'(a) = 0$, $f''(a) \leq 0$ - notwendig (2. Ordnung).

Was sind die Analoga dieser Art Bedingungen in der multivariaten Welt?

Für ein lokales Minimum.

- $f'(a) = 0$ – notwendig (1. Ordnung)
- $f'(a) = 0$, $f''(a) > 0$ – hinreichend.
- $f'(a) = 0$, $f''(a) \geq 0$ – notwendig (2. Ordnung).

Was ist so besonders an $T_d(x)$?

$T_1(a) = f(a)$

$T_1'(a) = f'(a)$

$T_2(a) = f(a)$

$T_2'(a) = f'(a)$

$T_2''(a) = f''(a)$

NR:

$T_2'(x) = f'(a) + f''(a) \cdot (x-a)$

$T_2'(a) = f'(a).$

$T_2''(x) = f''(a)$

$T_2''(a) = f''(a)$

$T_d(x)$ ist das eindeutige Polynom mit $\deg T_d \leq d$ und

$T_d^{(k)}(a) = f^{(k)}(a)$ für alle $k = 0, \dots, d$

Wir sind auf der Suche nach den Polynomen mit derselben Eigenschaft im Fall von Funktionen in n Variablen.

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) = f(x_1, \dots, x_n)$$

$a = (a_1, \dots, a_n)$ – unser Entwicklungspunkt.

$$T_0(x) = f(a). \qquad T_0(a) = f(a).$$

$$T_1(x) = f(a) + \langle \nabla f(a), x - a \rangle$$

$$= f(a) + \frac{\partial f}{\partial x_1}(a) \cdot (x_1 - a_1) + \dots + \frac{\partial f}{\partial x_n}(a) \cdot (x_n - a_n).$$

$$T_1(a) = f(a)$$

$$\frac{\partial T_1}{\partial x_1}(a) = \frac{\partial f}{\partial x_1}(a), \dots\dots, \frac{\partial T_1}{\partial x_n}(a) = \frac{\partial f}{\partial x_n}(a).$$

Was ist mit $T_2(x)$?

$$T_2(x) = f(a) + \langle \nabla f(a), x - a \rangle + \frac{1}{2} \langle \nabla^2 f(a) \cdot (x-a), x - a \rangle$$

$$= f(a) + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)(a) \cdot (x_i - a_i) + \frac{1}{2} \sum_{i,j=1,\dots,n} \frac{\partial^2 f}{\partial x_i \partial x_j} (x_i - a_i)(x_j - a_j)$$

$$T_2(a) = f(a)$$

$$\frac{\partial T_2}{\partial x_1}(a) = \frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial T_2}{\partial x_n}(a) = \frac{\partial f}{\partial x_n}(a)$$

$\boxed{n=2}$

$$\boxed{\begin{aligned} T_2(x) &= f(a) + \frac{\partial f}{\partial x_1}(a)\cdot(x_1-a_1) + \frac{\partial f}{\partial x_2}(a)\cdot(x_2-a_2) \\ &+ \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x_1^2}(a)\cdot(x_1-a_1)^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}(a)\cdot(x_1-a_1)\cdot(x_2-a_2) \\ &+ \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x_2^2}(a)\cdot(x_2-a_2)^2 \end{aligned}}$$

$$\frac{\partial^2 T_2}{\partial x_1^2}(a) = \frac{\partial^2 f}{\partial x_1^2}(a)$$

$$\frac{\partial^2 T_2}{\partial x_1 \partial x_2}(a) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(a).$$

$$T_2(a) = f(a)$$

$$\frac{\partial T_2}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a)$$

$$\frac{\partial^2 T_2}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a).$$

Zurück zur Optimierung:

die Notwendige Bedingung erster Ordnung für die lokale Optimalität an der Stelle a ist, wie bereits diskutiert: $\nabla f(a) = 0$.

In diesem Fall ist

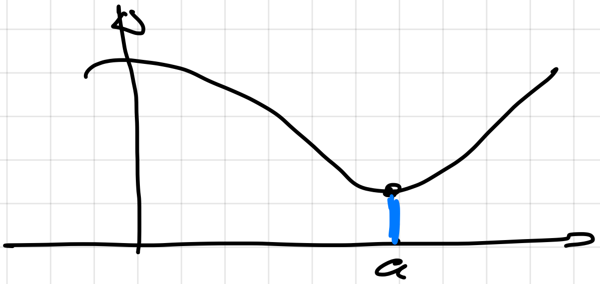
$$T_2(x) = \underbrace{f(a)}_{\uparrow} + \underbrace{\frac{1}{2}\left\langle \nabla^2 f(a)\cdot(x-a), (x-a)\right\rangle}_{\uparrow}$$

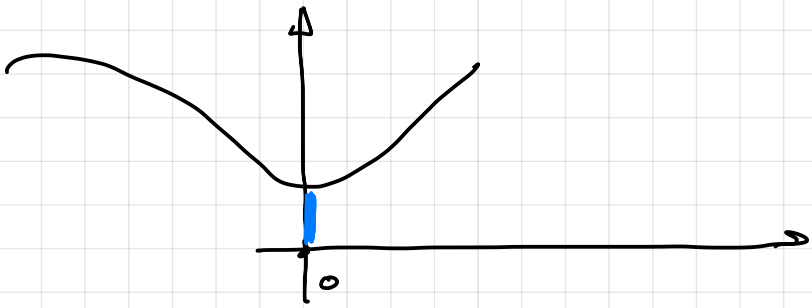
bei der Analyse der Optimalität egal.

hängt von $x-a$ ab,

Setzen wir einfach $a=0$

(das geht immer: die Funktion verschieben).





Wir haben nun mit der Funktion

$$q(x) = \frac{1}{2}\left\langle \nabla^2 f(0)\cdot x, x\right\rangle$$

zu tun.

So eine Funktion q nennt man eine quadratische Form.

Allgemein: ist $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix (d.h. $A^T = A$) so ist

$$q_A(x) = \langle Ax, x \rangle$$ die quadratische Form der Matrix A.