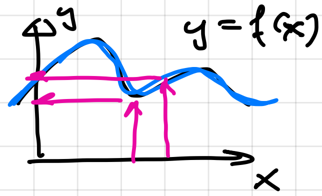
Agenda

+ Keine HS-Aufgaben ab Woche 13.
+ Implizite Funktionen
+ Richtungsableitung
- Optimierung multivariat
- Gradientenabstieg
- Taylorentwicklung multivariat
- Hinreichende Bedingungen zweiter Ordnung
- Lineare Algebra: Eigenwerte

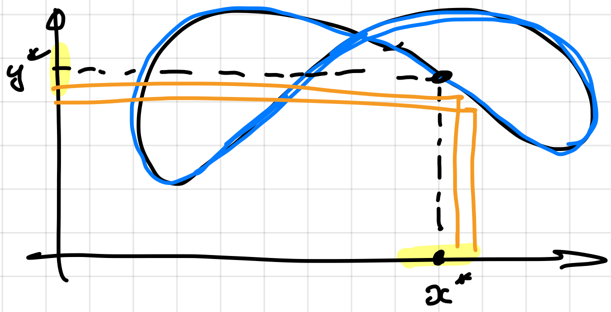
Explizite Funktion: man erhält $f(x)$ direkt aus x anhand eines Ausdruckes für $f(x)$.

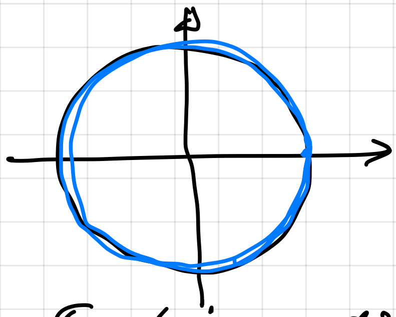
z.B.: $f(x) = 2e^x + \sin x$.



$$\underbrace{e^{f(x)} + e^{-f(x)} + \sin x - \cos f(x) = 0}_{\text{eine implizit gegebene Funktion}}$$

$$\boxed{x^2 + f(x)^2 = 1}$$





Implizit gegebene Funktion, die man auch explizit geben könnte

$f(x) = \pm\sqrt{1-x^2}$

(zwei stetige Funktionen, die aus der Vorschrift entstehen.

$A(x, f(x)) = 0. \leftarrow$ Implizit gegebene Funktion.

Frage: Was ist die Ableitung von $f(x)$.

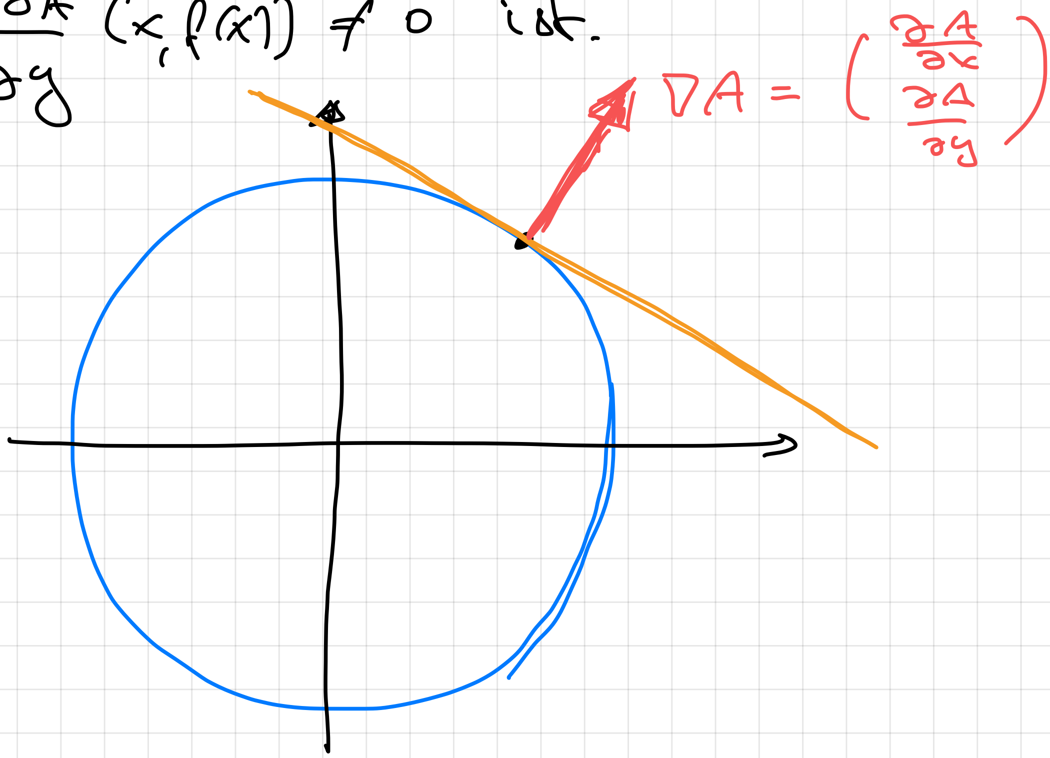
$A(x, y) \leftarrow$ Funktion in x und y.
Sie hat zwei Ableitungen $\frac{\partial A}{\partial x}$ und $\frac{\partial A}{\partial y}$.

$$0 = A(x, f(x))' = \frac{\partial A}{\partial x}(x, f(x)) \cdot x' + \frac{\partial A}{\partial y}(x, f(x)) \cdot f(x)'$$

$$= \frac{\partial A}{\partial x}(x, f(x)) + \frac{\partial A}{\partial y}(x, f(x)) \cdot f'(x)$$

$$\Rightarrow \quad \boxed{f'(x) = - \frac{\frac{\partial A}{\partial x}(x, f(x))}{\frac{\partial A}{\partial y}(x, f(x))}}$$

wenn $\frac{\partial A}{\partial y}(x, f(x)) \neq 0$ ist.



Die Formel für die Richtungsableitung.

Laut der Definition

$$\partial_h f(x) := \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$$

Die Ableitung von f an der Stelle x in Richtung $h \in \mathbb{R}^n$.

Mit anderen Worten:

$$\partial_h f(x) = \left[\frac{\partial}{\partial t} f(x+th)\right]\Big|_{t=0}$$

$$= \left[\frac{\partial}{\partial t} f(x_1 + th_1, \dots, x_n + th_n)\right]\Big|_{t=0} \overset{\text{Kettenregel}}{=}$$

$$\boxed{t \longmapsto (x_1 + th_1, \dots, x_n + th_n) \longmapsto f(x_1+th_1, \dots, x_n+th_n)}$$

$$\left[\frac{\partial f}{\partial x_1}(x+th) \cdot \frac{\partial(x_1+th_1)}{\partial t} + \frac{\partial f}{\partial x_2}(x+th) \cdot \frac{\partial(x_2+th_2)}{\partial t} + \dots + \frac{\partial f}{\partial x_n}(x+th) \cdot \frac{\partial(x_n+th_n)}{\partial t}\right]\Big|_{t=0}$$

$$= \left[\frac{\partial f}{\partial x_1}(x+th) \cdot h_1 + \dots + \frac{\partial f}{\partial x_n}(x+th) \cdot h_n\right]\Big|_{t=0}$$

$$= \frac{\partial f}{\partial x_1}(x)\, h_1 + \frac{\partial f}{\partial x_2}(x)\, h_2 + \dots + \frac{\partial f}{\partial x_n}(x)\, h_n .$$

$$\Longrightarrow \quad \partial_h f = \langle \nabla f, h \rangle$$

Bsp

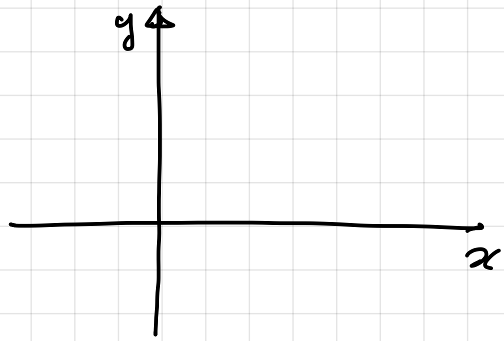
$$f(x,y) = x e^{x+y}$$

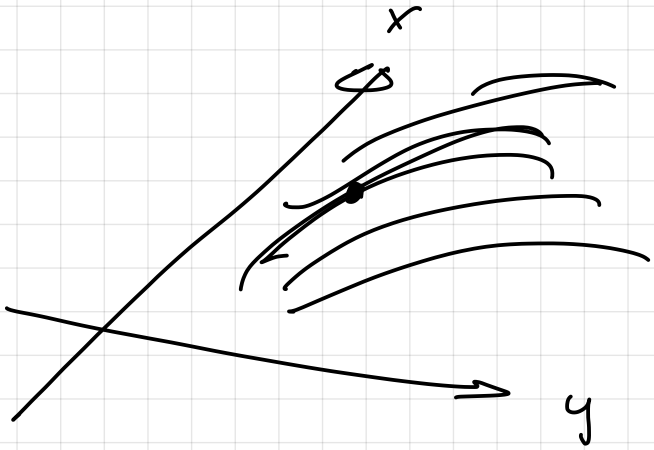
$$\frac{\partial f}{\partial x} = (x+1) e^{x+y}$$

$$\frac{\partial f}{\partial y} = x e^{x+y}$$

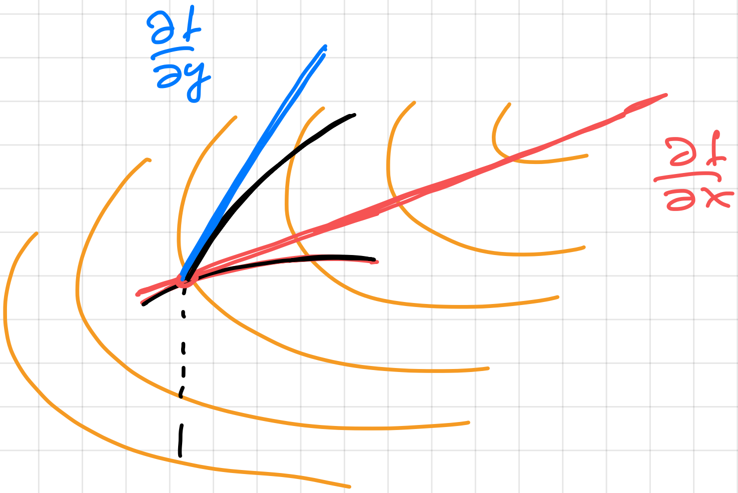
$$\nabla f = \left((x+1) e^{x+y}, \; x e^{x+y} \right)^T$$

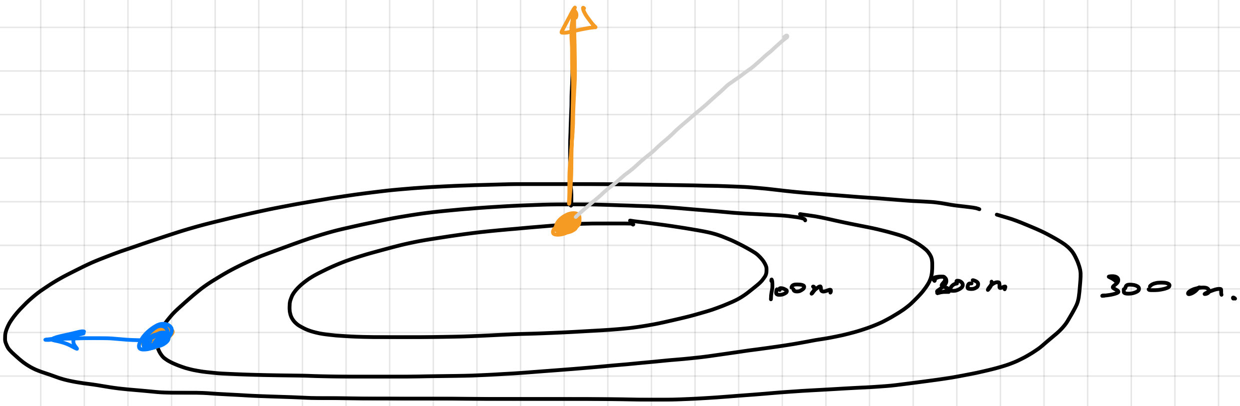
$$h = (-2, 1)$$





$$\partial_h f = -2 \cdot (x+1) \cdot e^{x+y}$$
$$+ 1 \cdot x \cdot e^{x+y}$$
$$= (-x-2) e^{x+y}$$



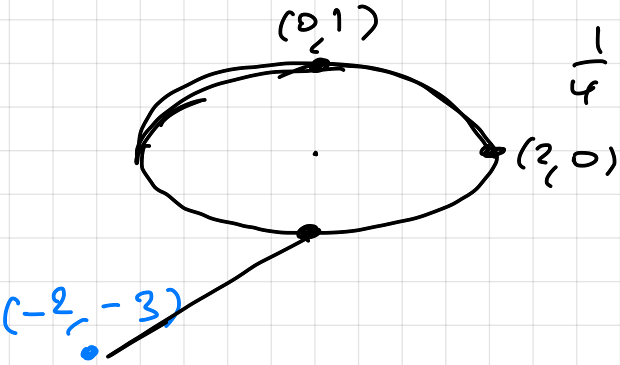


Multivariate Extremwertaufgaben

z.B. wir minimieren $f: \mathbb{R}^n \to \mathbb{R}$, d.h.

$$\min_{x_1,\dots,x_n \in \mathbb{R}} f(x_1,\dots,x_n)$$

Beispiel.



$$\frac{1}{4}x^2 + y^2 = 1$$

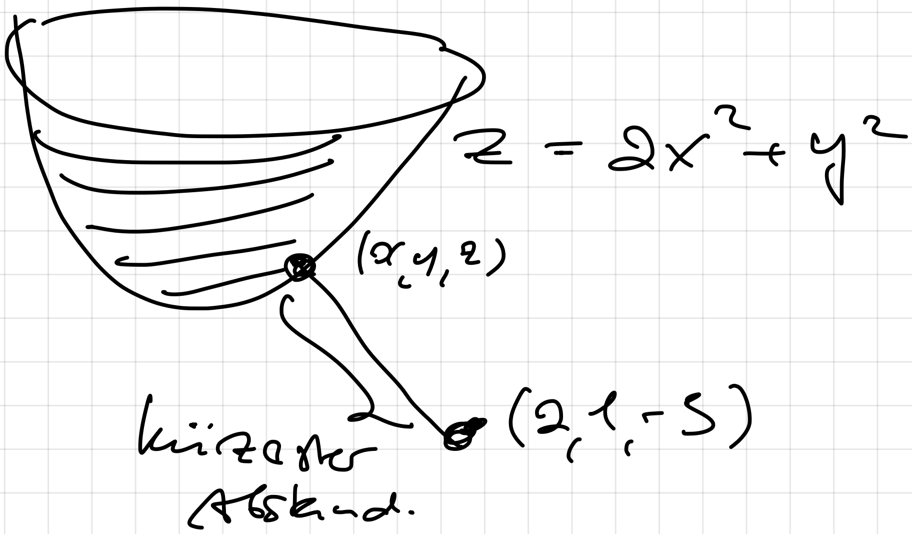
$$\min\left\{ (x+2)^2+(y+3)^2 \;:\; x,y\in\mathbb{R},\ \tfrac{1}{4}x^2+y^2=1 \right\}$$

restringierte Aufgabe.

Man hat eine Bedingung an der Optimierungsvariablen, und zwar hier: $\frac{1}{2}x^2+y^2=1$

Wir behandeln erstmal nichtrestringierte Aufgaben.

Bsp.



$$\min\left\{ \|(x,y,z)-(2,1,-3)\|^2 : z=2x^2+y^2,\ x,y\in\mathbb{R} \right\}$$

$$= \min\left\{ (x-2)^2+(y-1)^2+(2x^2+y^2+3)^2 : x,y\in\mathbb{R} \right\}$$

$$\|(x,y,z)\| = \sqrt{x^2+y^2+z^2}$$

Im univariaten Fall war die Bedingung $f'(x)=0$ notwendig für ein lokales Optimum, für eine differenzierbare Funktion $f: \mathbb{R} \to \mathbb{R}$.

Welche Bedingung ist im multivariaten Fall notwendig für ein lokales Optimum?

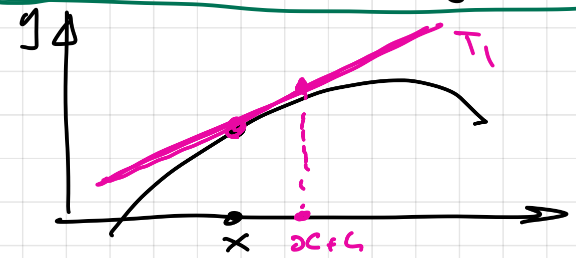
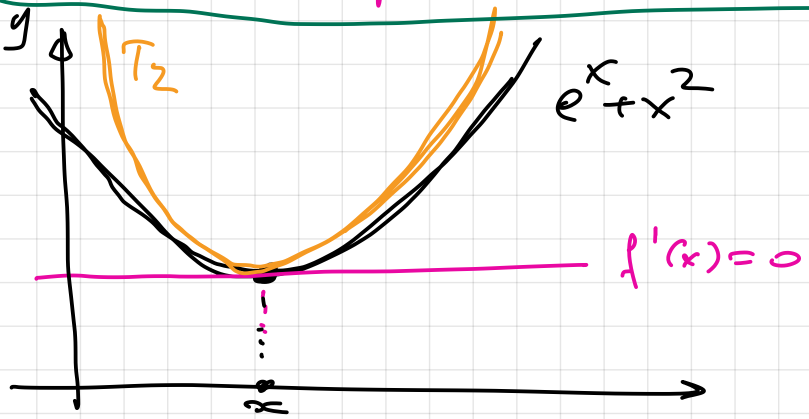
Die n-fache Anwendung der Bedingung im univariaten Fall zeigt, dass im n-variaten Fall die Bedingung $\nabla f = 0$ für lokale Optimalität notwendig ist (bei einer differenzierbaren Funktion $f: \mathbb{R}^n \to \mathbb{R}$!

Bsp $f(x,y) = (x-2)^2 + (y-1)^2 + (2x^2+y^2+3)^2$

$$\frac{\partial f}{\partial x} = 2(x-2) + 2(2x^2+y^2+3)\cdot 4x$$

$$\frac{\partial f}{\partial y} = 2(y-1) + 2(x^2+y^2+3)\cdot 2y$$

Differentialrechnung

Dim. 1	Dim. $n \in \mathbb{N}$:
$f'(x)$	$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$
Taylorentwicklung $$f(x+h) = \underbrace{\left(\sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!} h^k \right)}_{T_n(h) = \text{Taylorentwicklung der Ordnung } n.} + R_n(h)$$	Es gibt Taylor-Reihen für n-variate Funktionen
 $f'(x) = 0$ notwendig für die lokale Optimalität.	$\nabla f = 0$ notwendig für lokale Optimalität
 $\left.\begin{matrix} f''(x) > 0 \\ f'(x) = 0 \end{matrix}\right] \Rightarrow$ lokales Minimum in x.	$\nabla^2 f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)$ Hesse-Matrix ist die Matrix der Ableitungen der zweiten Ordnung.

$$\begin{cases} 2(x-2) + 2(2x^2+y^2+3)\cdot 4x = 0 \\ \\ 2(y-1) + 2(2x^2+y^2+3)\cdot 2y = 0 \end{cases}$$

↑ Das Gleichungssystem zur Bestimmung aller kritischen Punkte.

Wie löst man das?

Das kann man machen (fragen Sie Ihren Mathe-Lehrer :)), aber es ist schon nicht ganz elementar.

Idee ist: Iterativ, das nichtlineare System durch ein Lineares zu approximieren.

Es gibt eine einfachere Methode:

Gradient Descent
= Gradientenabstieg = Gradientenverfahren.

$p_0 \in \mathbb{R}^n$ (Startpunkt)

for $k=1,2,3,\ldots\ldots$

$p_k := p_{k-1} - \gamma \nabla f(p_{k-1})$ ($\gamma > 0$ kleiner Parameter)

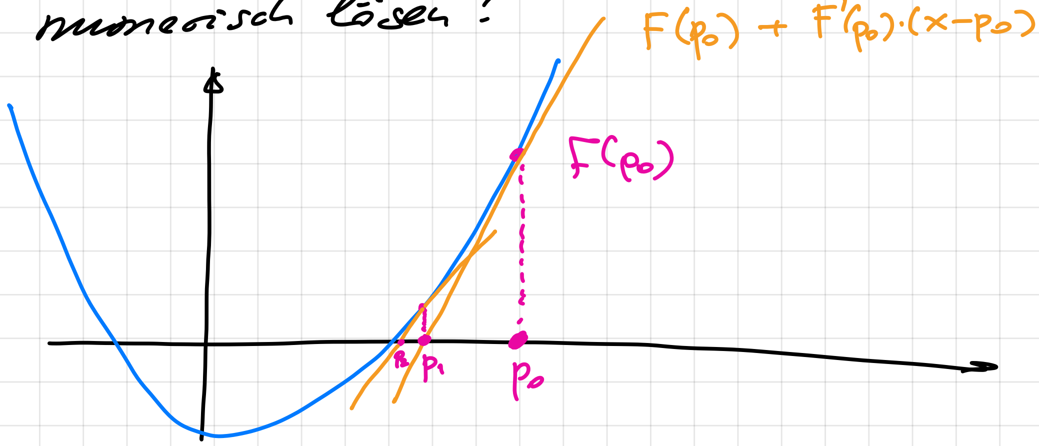
Sonst brauchen wir ein Abbruchkriterium:
z.B. $\|\nabla f\|$ ist klein genug oder
die Schranke an die Anzahl der Iteration oder was anderes.

Somit kann man γ auch adaptiv wählen

(γ_k an der Stelle von γ mit γ_k, das z.B. in Abhängigkeit von $\|\nabla f(p_{k-1})\|$ gewählt wird usw.)

Eine weitere Methode. Zurück zur Dimension 1.

$F: \mathbb{R} \to \mathbb{R}$. Wie können wir $F(x) = 0$ numerisch lösen?



Wir tun so, als ob $F(x)$ die Tangente $F(p_0) + F'(p_0)\cdot(x - p_0)$ wäre.

Wenn wir $F(p_0) + F'(p_0)\cdot(x-p_0) = 0$ nach x auflösen, erhalten wir:

$$F'(p_0)\cdot(x-p_0) = -F(p_0)$$

$$x - p_0 = -(F'(p_0))^{-1}\cdot F(p_0)$$

$$x = p_0 - (F'(p_0))^{-1}\cdot F(p_0)$$

Das ist unser p_1:

$$p_1 = p_0 - F'(p_0)^{-1}\cdot F(p_0).$$

Und so weiter, bis man mit einem p_k zufrieden ist:

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

Newton Verfahren (dim = 1).

Konvergiert nicht immer (je nach Funktion, nicht für jedes p_0) Aber, wenn es konvergiert, dann oft sehr schnell.

Können wir das Gleiche in der Dimension $n \in \mathbb{N}$ machen?

$F: \mathbb{R}^n \to \mathbb{R}^n$. Wir lösen $F(x) = 0$, ein nichtlineares System mit n Unbekannten und n Gleichungen.

Wir hatten: $F(x) = F(p_0) + F'(p_0) \cdot (x - p_0) + \underbrace{\text{Rest}}$

Ordnung höher als $x - p_0$ bei $x \to p_0$

Diese Entwicklung kann man auch in der Dimension n nutzen:

F' ist die Jacobi-Matrix (die Matrix der Ableitung der Komponenten von F).

Wir lösen $\underbrace{F(p_0) + F'(p_0) \cdot (x - p_0) = 0}$ bzgl. x auf.

Das ist ein lineares Gleichungssystem. Das kann man lösen. z.B. mit dem Gauß-Verfahren.

$$F'(p_0) \cdot (x - p_0) = -F(p_0)$$

$$x - p_0 = -F'(p_0)^{-1} \cdot F(p_0)$$

$$x = p_0 - F'(p_0)^{-1} \cdot F(p_0).$$

D.h. $p_1 := p_0 - F'(p_0)^{-1} \cdot F(p_0)$ und allgemein:

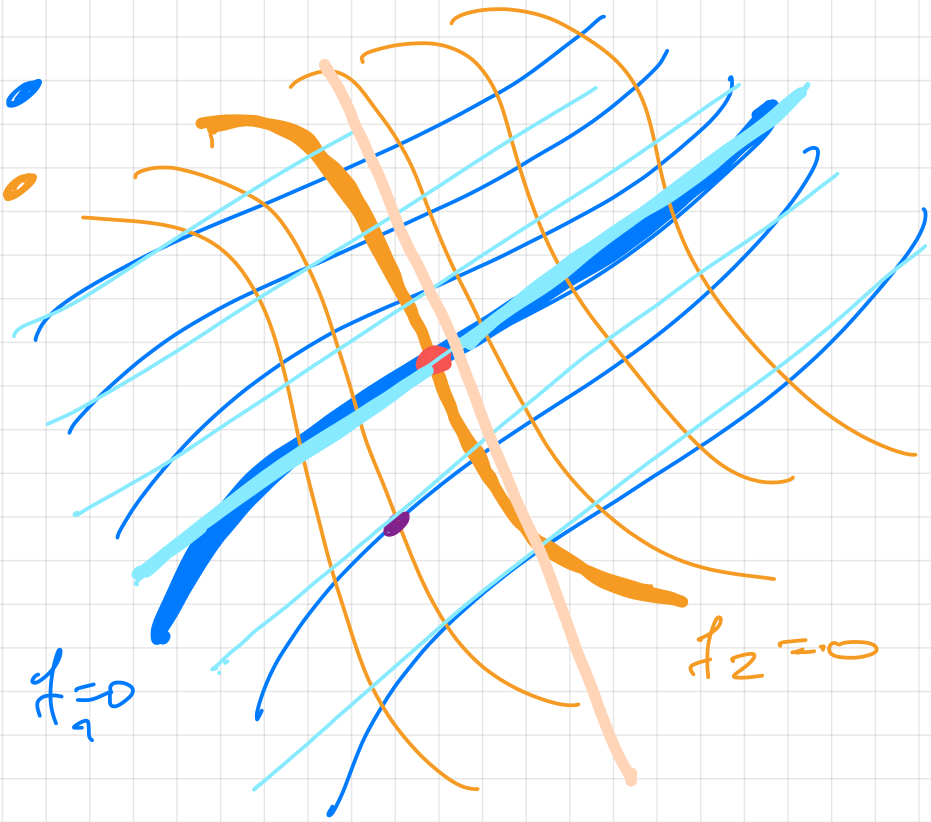
$$p_k := \underbrace{p_{k-1}}_{\text{Vektor}} - \underbrace{F'(p_{k-1})^{-1}}_{\substack{n\times n\\ \text{Matrix}}} \cdot \underbrace{F(p_{k-1})}_{\text{Vektor}}$$

Das berühmte

Newton-Verfahren

(in Dimension n).

$$\begin{cases} f_1(x,y) = 0 \\ f_2(x,y) = 0 \end{cases}$$



$p_0 = (x_0, y_0)$

$$\begin{cases} f_1(x_0,y_0) + \dfrac{\partial f_1}{\partial x}(x_0,y_0)\cdot(x-x_0) + \dfrac{\partial f_1}{\partial y}(x_0,y_0)\cdot(y-y_0) = 0 \\[2ex] f_2(x_0,y_0) + \dfrac{\partial f_2}{\partial x}(x_0,y_0)\cdot(x-x_0) + \dfrac{\partial f_2}{\partial y}(x_0,y_0)\cdot(y-y_0) = 0 \end{cases}$$

Die Lösung (x, y) von diesem 2×2 linearen Gleichungssystem ist $P_1 = (x_1, y_1)$.

$$F(x,y) = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix}$$