2.4 Gradient

The gradient of a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ at $\underline{a} \in \Omega$ is defined as

$$\text{grad}\, f(\underline{a}) = \begin{pmatrix} \partial_{x_1} f(\underline{a}) \\ \vdots \\ \partial_{x_n} f(\underline{a}) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} f(\underline{a})$$

- Using vector notation, the total differential can be rewritten as

$$df(\underline{a}) = \text{grad}\, f(\underline{a}) \underset{\substack{\uparrow \\ \text{scalar} \\ \text{product}}}{\bullet} d\underline{x} \quad \text{with } d\underline{x} = \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\underline{a}) \bullet dx_i$$

- notation: $\text{grad}\, f(\underline{a}) \equiv \underline{\nabla} f(\underline{a})$ with the Nabla operator $\underline{\nabla} = \begin{pmatrix} \partial_{x_1} \\ \vdots \\ \partial_{x_n} \end{pmatrix}$

  ↳ $df(\underline{a}) = \underline{\nabla} f(\underline{a}) \bullet d\underline{x}$

- geometric intuition: (i) $\underline{\nabla} f(\underline{a})$ points in the direction of steepest slope of f at $\underline{a}$.

  (ii) $$\underset{\substack{\uparrow \\ \text{change of } f}}{\Delta f} = \underline{\nabla} f \cdot \Delta \underline{x} = |\underline{\nabla} f| \cdot |\Delta \underline{x}| \cos \overset{\text{angle between } \underline{\nabla} f \text{ and } \Delta \underline{x}}{\theta}$$

- viele Anwendungen:

↳ Physik: $\underset{\underset{\text{E-Feld}}{\uparrow}}{\underline{E}} = -\underline{\nabla}\, \underset{\underset{\text{elektrostatisches Potenzial}}{\uparrow}}{V}$

↳ Physik: Konservative Kräfte: $\underline{F} = -\underline{\nabla} V$

↳ Optimierung (gradient descent)

2.5 Mehrdimensionale Taylor-Entwicklung

Erinnerung: Taylor-Reihe für $f:\mathbb{R}\to\mathbb{R}$: $f(x)=\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)\,\underbrace{(x-x_0)}_{\Delta x}{}^{k}$
(vgl. 1.6)

Δx: Abstand vom Entwicklungspunkt x_0

$$\left(=\sum_{k=0}^{\infty} \frac{1}{k!}\left(\Delta x \frac{d}{dx}\right)^{k} f(x_0)\right)$$

Verallgemeinerung auf $f:\mathbb{R}^n\to\mathbb{R}$:

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \Delta\underline{x} = \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}, \quad \underline{x}_0 = \begin{pmatrix} x_{01} \\ \vdots \\ x_{0n} \end{pmatrix}$$

$$f(\underline{x}) = \sum_{k=0}^{\infty} \frac{1}{k!} \underbrace{\left(\sum_{i=1}^{n} \Delta x_i \frac{\partial}{\partial x_i}\right)}_{(\Delta\underline{x}\, \underset{\underset{\text{"Skalarprodukt"}}{\uparrow}}{\cdot}\, \underline{\nabla})^{k}}{}^{k} f(\underline{x}_0)$$

Bsp.: 0-te Ordnung: $k=0$: $P_0(\underline{x}) = f(\underline{x}_0)$

1. Ordnung: $k=1$: $P_1(\underline{x}) = \sum_{i=1}^{n} \left(\partial_{x_i} f(\underline{x}_0)\right) \Delta x_i$

2. Ordnung: $k=2$: Nebenrechnung: $\left(\sum_{i=1}^{2} a_i\right)\left(\sum_{i=1}^{2} a_i\right) = (a_1+a_2)(a_1+a_2)$

$$= a_1 a_1 + a_1 a_2 + a_2 a_1 + a_2 a_2$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} a_i a_j = \underset{\underset{\text{Doppelsumme}}{\uparrow}}{\sum_{i,j=1}^{2}} a_i a_j$$

Damit: $P_2(\underline{x}) = \frac{1}{2!} \left(\sum_{i=1}^{n} \Delta x_i \, \partial_{x_i} \right)^2 f(\underline{x}_0)$

$$= \frac{1}{2!} \sum_{i,j=1}^{n} \left(\partial_{x_i} \partial_{x_j} f(\underline{x}_0) \right) \Delta x_i \, \Delta x_j$$

$\Rightarrow f(\underline{x}) = P_0(\underline{x}) + P_1(\underline{x}) + P_2(\underline{x}) + \mathcal{O}(|\Delta \underline{x}|^3)$

$$= f(\underline{x}_0) + \sum_{i=1}^{n} \left(\partial_{x_i} f(\underline{x}_0) \right) \Delta x_i + \frac{1}{2} \sum_{i,j=1}^{n} \left(\partial_{x_i} \partial_{x_j} f(\underline{x}_0) \right) \Delta x_i \, \Delta x_j + \mathcal{O}(|\Delta \underline{x}|$$

- Bem: Kurzschreibweise:

$$f(\underline{x}) = f(\underline{x}_0) + \left(\nabla f(\underline{x}_0) \right) \cdot \Delta \underline{x} + \frac{1}{2} \Delta \underline{x}^{\overset{\text{transponiert}}{T}} \underline{\underline{H}}(\underline{x}_0) \, \Delta \underline{x} + \mathcal{O}(|\Delta \underline{x}|^3)$$

mit der Hesse-Matrix

$$\underline{\underline{H}}(\underline{x}_0) = \begin{pmatrix} \partial_{x_1} \partial_{x_1} f(\underline{x}_0) & \cdots & \partial_{x_1 x_n} f(\underline{x}_0) \\ \vdots & & \vdots \\ \partial_{x_n} \partial_{x_1} f(\underline{x}_0) & \cdots & \partial_{x_n} \partial_{x_n} f(\underline{x}_0) \end{pmatrix}$$

Bsp: • Taylorentwicklung für $n = 2$: $f: \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto f(x, y)$

$$f(\underline{x}) = f(\underline{x}_0) + \partial_x f(\underline{x}_0) \Delta x + \partial_y f(\underline{x}_0) \Delta y$$

$$+ \frac{1}{2} \partial_x \partial_x f(\underline{x}_0) \Delta x^2 + \frac{1}{2} \partial_{yy} f(\underline{x}_0) \Delta y^2$$

$$+ \frac{1}{2} \overset{\text{S. v. Schwarz}}{2} \partial_x \partial_y f(\underline{x}_0) \Delta x \Delta y + \mathcal{O}(|\Delta \underline{x}|^3)$$

$$= f(\underline{x}_0) + \begin{pmatrix} \partial_x f(\underline{x}_0) \\ \partial_y f(\underline{x}_0) \end{pmatrix} \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \begin{pmatrix} \partial_x \partial_x f(\underline{x}_0) & \partial_x \partial_y f(\underline{x}_0) \\ \partial_y \partial_x f(\underline{x}_0) & \partial_y \partial_y f(\underline{x}_0) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

$$+ \mathcal{O}(|\Delta \underline{x}|^3)$$

2.6 Extremwerte / -Stellen

- Erinnerung: $f: \mathbb{R} \to \mathbb{R}$, $x \mapsto f(x)$: $f'(a)=0$ notwendige Bedingung für ein Extremum bei a

  ↳ $f''(a) > 0$: positive Krümmung ⇒ Minimum

  $f''(a) < 0$: negative Krümmung ⇒ Maximum

  $f''(a) = 0$: Maximum, Minimum oder Sattelpunkt
  (z.B.: $f(x) = x^4$ bei $a=0$)

- $f'(a)=0$ ⇒ a ist ein stationärer Punkt (Fixpunkt) $\dot{x} = f(x)$

- Verallgemeinerung auf $f: \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \longmapsto f(x,y)$

  ↳ notwendige Bedingung: $df = \frac{\partial f}{\partial x}(\underline{a})\, dx + \frac{\partial f}{\partial y}(\underline{a})\, dy = 0$ für beliebige dx, dy.

$$\Rightarrow \frac{\partial f}{\partial x}(\underline{a}) = 0 \quad \wedge \quad \frac{\partial f}{\partial y}(\underline{a}) = 0$$

Art des stationären Punkts hängt von der Krümmung ab: (Eigenwerte Kap. 6)

(i) Maximum : $\partial_x^2 f(\underline{a}) < 0$, $\partial_y^2 f(\underline{a}) < 0$, $D := (\partial_x^2 f)(\partial_y^2 f) - (\partial_x \partial_y f)^2 > 0$

(ii) Minimum : $\partial_x^2 f(\underline{a}) > 0$, $\partial_y^2 f(\underline{a}) > 0$, $D > 0$

(iii) Sattelpunkt: $D < 0$

(iv) $D = 0$: höhere Ordnungen zu untersuchen

Bsp.: $f(x,y) = xy + x^2 + y^2 - 6y$

↳ notwendig:
$$\left.\begin{aligned} \partial_x f &= y + 2x = 0 \\ \partial_y f &= x + 2y - 6 = 0 \end{aligned}\right\} \quad \begin{aligned} x &= -2 \\ y &= 4 \end{aligned}$$

2. Ableitungen:
$$\left.\begin{aligned} \partial_x^2 f &= 2 \quad > 0 \\ \partial_y^2 f &= 2 \quad > 0 \\ \partial_x \partial_y f &= 1 \Rightarrow D > 0 \end{aligned}\right\} \quad \text{Minimum}$$

• Verallgemeinerung auf $f: \mathbb{R}^n \to \mathbb{R}$:

↳ $df(\underline{a}) = 0 \quad \forall\, dx_1, \ldots, dx_n \Rightarrow \partial_{x_1} f = 0 = \partial_{x_2} f = \ldots = \partial_{x_n} f$

oder kompakt: $\underline{\nabla} f(\underline{a}) = 0 \leftarrow$ Nullvektor

↳ Art des stationären Punkts über Eigenwerte der Hesse-Matrix

↳ alle Eigenwerte $>0 \Rightarrow$ Minimum

↳ alle Eigenwerte $<0 \Rightarrow$ Maximum

2.7 Extremum unter Nebenbedingungen / Methode der Lagrange-Multiplikatoren

hier: $n=2$

Idee: Extrema von $f(x,y)$ unter einer Nebenbedingung $g(x,y)=0$

Bsp.: Maximum von $f(x,y) = -(x^2+y^2)$ unter Nebenbedingung $y = x + c$

$\uparrow$

$\Rightarrow g(x,y) = y - x - c = 0$

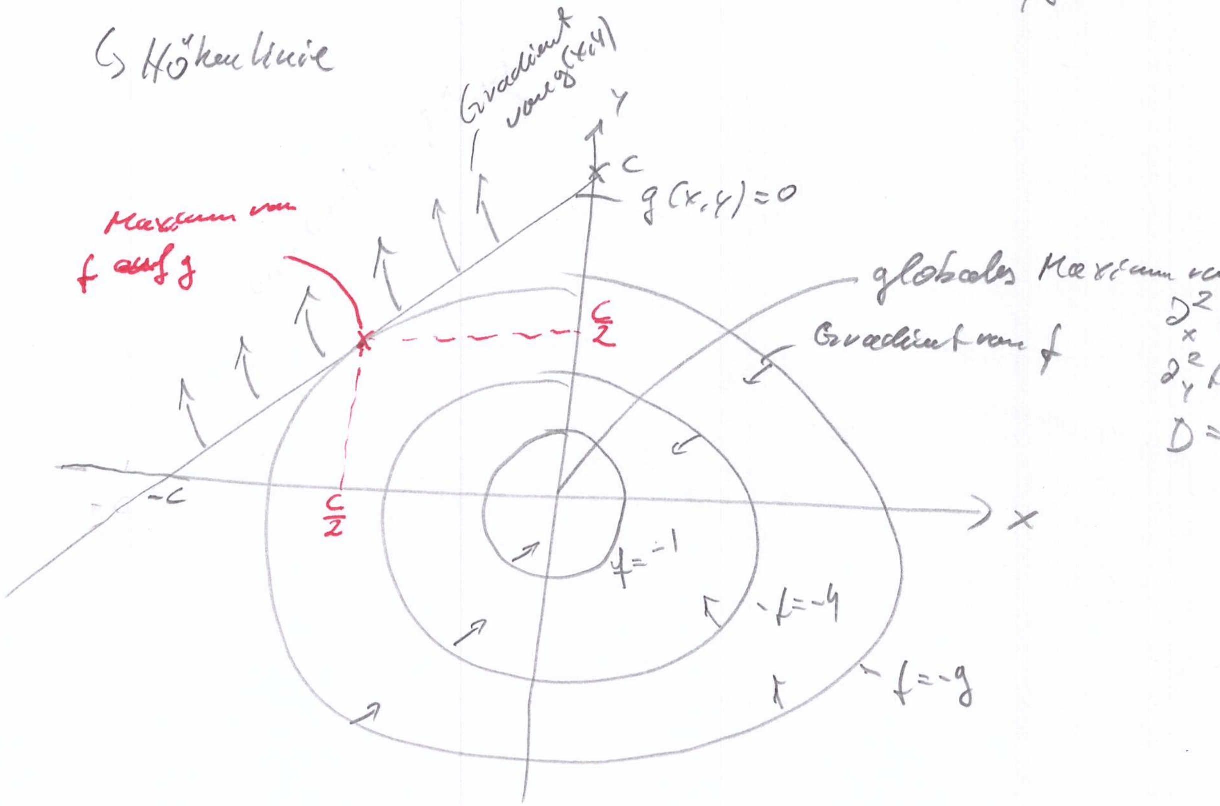
$\partial_x g = -1$

$\partial_y g = 1$

nach unten geöffneter Rotationsparaboloid

↳ s. 2.2 (VL6)

↳ Höhenlinie



$\partial_x^2 f = -2 < 0$

$\partial_y^2 f = -2 < 0$

$D = 4 \quad > 0$

↳ Gradienten stehen senkrecht auf Höhenlinien.

↳ Extremum mit Nebenbedingungen, wo Gradient von f und g parallel sind

$\Rightarrow \underline{\nabla} f(x,y) = -\lambda\, \underline{\nabla} g(x,y) \Leftrightarrow \underline{\nabla}(f + \lambda g) = 0$

$\nabla(\underbrace{f+\lambda g}_{=:h(x,y,\lambda)}) = 0 \quad \wedge \quad g(x,y) = 0$: 3 Gleichungen für x, y, λ

$h(x,y,\lambda) = -(x^2+y^2) + \lambda(y-x-c)$

$$\Rightarrow \quad \nabla h = 0 \quad \Leftrightarrow \quad \left.\begin{aligned} \frac{\partial h}{\partial x} &= -2x - \lambda &= 0 \\ \frac{\partial h}{\partial y} &= -2y + \lambda &= 0 \\ y - x - c &&= 0 \end{aligned}\right\}$$

$\Rightarrow$ Lösung: $\lambda = c$, $x = -\frac{c}{2}$, $y = \frac{c}{2}$

- Verallgemeinerung auf s Nebenbedingungen: $g_i(\underline{x}) = 0$, $i = 1, \dots, s$

$\hookrightarrow h(\underline{x}, \lambda_1, \dots, \lambda_s) := f(\underline{x}) + \sum_{i=1}^{s} \underset{\uparrow}{\lambda_i}\, g_i(\underline{x})$

s Lagrange-Parameter / -Multiplikator

$$\left.\begin{aligned} &\Rightarrow n \text{ Gleichungen: } \nabla h(\underline{x}, \lambda, \dots, \lambda_s) = 0 \\ &s \text{ Gleichungen: } g_i(\underline{x}) = 0 \quad i = 1, \dots, s \end{aligned}\right\} n+s \text{ Unbekannte } \checkmark$$