

6. Normen und Metriken

Die meisten Abschätzungen in der eindimensionalen Analysis (allgemein Ana I genannt) nutzten bzw. beruhten im Wesentlichen auf zwei Eigenschaften des Abstands zwischen den Punkten x_1, x_2 auf der Zahlengerade \mathbb{R} , den wir als Abstand der Differenz $x_1 - x_2$ zur Null durch den Absolutbetrag

$$|x| = \begin{cases} x & \text{wenn } x \geq 0 \\ -x & \text{wenn } x \leq 0 \end{cases}$$

der Differenz $x_1 - x_2$ darstellten. Diese entscheidenden Bedingungen sind

- a) $\forall \lambda \forall x : |\lambda \cdot x| = |\lambda| |x|$ positive 1-Homogenität in λ
- b) $\forall x \forall y : |x + y| \leq |x| + |y|$ Δ -Ungleichung

Die gleichen Eigenschaften hat $|z| = \sqrt{z \cdot \bar{z}}$ auf \mathbb{C} , deshalb galten viele Argumente aus \mathbb{R} auch in \mathbb{C} . Wenn wir ähnliche Ergebnisse in höhere Dimensionen analog folgern wollen, brauchen wir ein Analogum zur Betragsfunktion. Dies wird im Folgenden als "Norm" eingeführt. Zunächst wiederholen wir jedoch kurz, was "·" und "+" in n Dimensionen bedeutet.

DEFINITION 6.1. Der $(\mathbb{R}^n, \cdot, +)$ ist die Menge

$$\mathbb{R}^n = \overbrace{\mathbb{R} \times \dots \times \mathbb{R}}^{n\text{-mal}} = \{(x_1, \dots, x_n) : x_i \in \mathbb{R}\}, \text{ wobei } n \in \mathbb{N}_+ = \{1, 2, \dots\},$$

also der n -dimensionale Standard(vektor)raum, dessen Elemente $x = (x_1, \dots, x_n)$ wir Vektoren nennen, zusammen mit folgenden Operationen:

- i) skalare Multiplikation

$$\text{"} \cdot \text{"} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n; \lambda \cdot (x_1, \dots, x_n) = \lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$$

für $\lambda \in \mathbb{R}$ ein Skalar und $(x_1, \dots, x_n) \in \mathbb{R}^n$ ein Vektor.

- ii) Addition von Vektoren

$$\text{"} + \text{"} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n; (x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n).$$

BEMERKUNG 6.2.

- \mathbb{C}^n ist analog definiert. • "+" ist wie die Addition in \mathbb{R} assoziativ und kommutativ, hat als neutrales Element den (Koordinaten)Ursprung $0 = (0, \dots, 0)$ (somit einziges neutrales Element!) und $(-x_1, \dots, -x_n)$ ist das additive Inverse (Negative) zu $x = (x_1, \dots, x_n)$.
- man überprüft leicht auch die Analogien zu "Assoziativität" der skalaren Multiplikation und die entsprechenden Distributivgesetze:

$$(\lambda \cdot \mu) \cdot x = \lambda \cdot (\mu \cdot x), (\lambda + \mu)x = \lambda x + \mu x, \lambda(x + y) = \lambda x + \lambda y.$$

\mathbb{R}^n ist trotzdem kein Körper!! (Warum ??)

- $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, ..., $e_n = (0, \dots, 0, 1)$ sind die kanonischen Einheitsvektoren im \mathbb{R}^n , wichtig ist, dass dann jeder Vektor $x = (x_1, \dots, x_n)$ die **eindeutige** Darstellung $x = \sum_{i=1}^n x_i e_i$ hat! Wir nennen x_i die i -te Koordinate des Vektors $x = (x_1, \dots, x_n)$

- n ist die Anzahl der unabhängigen Richtungen, der Freiheitsgrade
- Der \mathbb{R}^n ist ein "Spezial"fall des allgemeinen n -dimensionalen reellen Vektorraumes (siehe LA I, warum ist "Spezial" in Anführungsstrichen !?)

DEFINITION 6.3. Eine Abbildung $f : \mathbb{K}^n \rightarrow \mathbb{K}^m$, $m, n \in \mathbb{N}$ ist (\mathbb{K} -) linear wenn

$$\forall x, y \in \mathbb{K} : f(\lambda x + y) = \lambda f(x) + f(y)$$

DEFINITION 6.4. Der n -dimensionale **Euklidische Raum** ist der $\mathbb{R}^n = (\mathbb{R}^n, \cdot, +)$ zusammen mit $|\cdot| : \mathbb{R}^n \rightarrow [0, \infty)$ wobei $|(x_1, \dots, x_n)| = \sqrt{\sum_{i=1}^n x_i^2}$. Wir nennen $|x|$ die **Euklidische Norm** des Vektors $x \in \mathbb{R}^n$.

BEMERKUNG 6.5. $|x|$ entspricht notwendigerweise dem Abstand von x zum Ursprung, wenn der Satz des Pythagoras im Euklidischen Raum gelten soll. Allerdings ist es umgekehrt noch nicht klar, dass mit unserer intuitiven "Definition: Abstand von x nach y ist $|x - y|$ " der Satz des Pythagoras auch wirklich für alle rechtwinkligen (insbesondere nicht achsenparallele) Dreiecke gilt. Dies, sowie die Gültigkeit von obigen Eigenschaften a), b) werden wir zeigen, vorher noch folgende Präzisierungen und Verallgemeinerungen

DEFINITION 6.6. Ein normierter \mathbb{R}^n ist der $\mathbb{R}^n = (\mathbb{R}^n, \cdot, +)$ zusammen mit einer allgemeinen Abbildung $\|\cdot\| : \mathbb{R}^n \rightarrow [0, \infty)$ – die **Norm**, mit folgenden Eigenschaften

- $\forall x \in \mathbb{R}^n : \|x\| \geq 0$ Positivität der Norm,
- $\|x\| = 0 \Rightarrow x = 0$ Definitheit der Norm,
- $\forall \lambda \in \mathbb{R} \forall x \in \mathbb{R}^n : \|\lambda x\| = |\lambda| \|x\|$ (positive 1-) Homogenität der Norm
- $\forall x, y \in \mathbb{R}^n : \|x + y\| \leq \|x\| + \|y\|$ die Δ -Ungleichung

BEMERKUNG 6.7. Wie zuvor, $\|x\|$ präsentiert den (nichteuclidischen) Abstand von x zum Ursprung 0.

BEISPIEL 6.8. andere Normen auf \mathbb{R}^n :

- $\|x\|_1 = \sum_{i=1}^n |x_i|$, die New York-Norm: $\|(k, m)\|$ - Abstand von k th Ave, m th Street zum "Zentrum"
- $\|x\|_\infty = \max_{i \leq n} |x_i|$, die Schachbrettnorm: Zahl der Schritte, die der König von (k, m) bis $(0, 0)$ braucht.
- * $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ mit $p \in [1, \infty)$, dann $|x| = \|x\|_2$, $\|x\|_1 = \|x\|_1$ aus i) und $\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty!$

BEMERKUNG 6.9.

- Es ist einfach zu überprüfen, dass $\|\cdot\|_1, \|\cdot\|_\infty$ Normen sind, z.B.

$$\begin{aligned} \forall i \leq n \|(x + y)_i| &= |x_i + y_i| \leq |x_i| + |y_i| \leq \|x\|_\infty + \|y\|_\infty \\ \Rightarrow \|x + y\|_\infty &= \max_i \|(x + y)_i| \leq \|x\|_\infty + \|y\|_\infty, \text{ Rest ÜZ.} \end{aligned}$$

- Warum benutzen wir nicht **nur** die einfacher zu berechnenden Normen $\|\cdot\|_1$ oder $\|\cdot\|_\infty$? Diese sind weniger geeignet für feinere geometrische oder realistische physikalische Anwendungen.

Zum Beispiel gilt: Wenn $x, y \in \mathbb{R}^n$ und $|x| = |y|$, dann gibt es eine lineare Isometrie Φ des \mathbb{R}^n die x auf y schickt. D.h. $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist eine lineare Bijektion (Isomorphismus), und $|\Phi(x)| = |x| \forall x \in \mathbb{R}^n$ — also

$$\forall x, y \in \mathbb{R}^n : |\Phi(x) - \Phi(y)| = |\Phi(x - y)| = |x - y|$$

und somit erhält Φ alle Abstände - genaue Abstandsdefinition folgt noch.

- auch für $\|\cdot\|_p$ ist i), ii) und iii) einfach, für die Δ -Ungleichung und $p \in (1, \infty)$ muss man die Hölder-Ungleichung (Ana I, Satz 4.32) nutzen, um die jetzt für die Euklidische Norm folgende Idee zu adoptieren.
- Idee des Beweises der Δ -Ungleichung für $|\cdot|$: Seien P und R Punkte mit $0\vec{P} = x$, $\vec{P}\vec{R} = y$, also $0\vec{R} = x + y$. Wir betrachten auf der Strecke $\overline{0\vec{R}}$ den Punkt Q , der die orthogonale Projektion von P ist, dann sollte $|x + y| = |\vec{OQ}| + |\vec{QR}| \leq |\vec{OP}| + |\vec{PR}| = |x| + |y|$ nach Pythagoras gelten. Da wir diesen Satz aber noch nicht allgemein gezeigt haben, brauchen wir eine bessere Methode, um mit orthogonalen Projektionen umzugehen.

DEFINITION 6.10. Für $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ ist

$$\langle x, y \rangle = x \bullet y = \sum_{i=1}^n x_i y_i$$

das Skalarprodukt der Vektoren x und y .

Natürlich gilt $|x|^2 = x \bullet x$, aber auch die Eigenschaften $\langle x, y \rangle = \langle y, x \rangle$ und $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ sind leicht zu nachzurechnen und im Folgenden wichtig.

SATZ 6.11. CAUCHY-SCHWARZSCHE UNGLEICHUNG: $\forall x, y \in \mathbb{R}^n : |\langle x, y \rangle| \leq |x||y|$ mit Gleichheit genau dann wenn $x = 0$ oder $\exists \lambda \in \mathbb{R} : y = \lambda x$.

Also ist für $x, y \neq 0$ der (unorientierte) Winkel zwischen den Vektoren x und y durch

$$\cos(\angle(x, y)) = \frac{\langle x, y \rangle}{|x||y|}, \angle(x, y) \in [0, \pi]$$

wohldefiniert.

Beweis: OBdA. können wir uns auf den Fall $|x||y| > 0$ beschränken. Da $\frac{1}{2}(a^2 + b^2) - ab = \frac{1}{2}(a - b)^2 \geq 0$ mit Gleichheit genau dann wenn $a = b$, haben wir

$$\forall i \leq n : \frac{|x_i||y_i|}{|x||y|} \leq \frac{1}{2} \left(\frac{|x_i|^2}{|x|^2} + \frac{|y_i|^2}{|y|^2} \right) \quad \& \text{ Gleichh. gdw } \exists c_i \in \{-1, 1\} : c_i \frac{x_i}{|x|} = \frac{y_i}{|y|}.$$

Summieren über $i = 1, \dots, n$ ergibt

$$\sum_{i=1}^n \frac{|x_i y_i|}{|x||y|} \leq \frac{1}{2} \left(\sum_{i=1}^n \frac{|x_i|^2}{|x|^2} + \sum_{i=1}^n \frac{|y_i|^2}{|y|^2} \right) = \frac{1}{2}(1 + 1) = 1,$$

d.h. $\sum_i^n |x_i y_i| \leq |x||y|$ mit Gleichheit genau dann, wenn $\forall i \leq n \exists c_i \in \{-1, 1\} : c_i \frac{|y|}{|x|} x_i = y_i$.
 Überdies gibt es ein $c \in \{-1, 1\}$ so dass

$$|\langle x, y \rangle| = c \langle x, y \rangle = \sum_1^n c(x_i y_i) \leq \sum_i^n |x_i y_i|,$$

mit Gleichheit gdw. $\forall i \leq n : 0 \leq c(x_i y_i)$.

Wir haben also

$$\langle x, y \rangle \leq |\langle x, y \rangle| \leq \sum_i^n |x_i y_i| \leq |x||y|.$$

Falls $|\langle x, y \rangle| = |x||y|$, dann $\forall i \leq n : 0 \leq c(x_i y_i) = c c_i x_i^2 |y|/|x|$ und somit (OBdA) $c = c_i$ für alle i bzw. $y = \lambda x$ mit $\lambda = c|y|/|x|$; wenn $\langle x, y \rangle = |x||y|$ dann $c = +1$ und $\lambda > 0$. Klarerweise gilt in diesem letzten Absatz die Umkehrung.

BEMERKUNG 6.12.

- wir sagen, dass $x \perp y$ (x und y stehen senkrecht aufeinander) wenn $x \bullet y = 0$
- der Kosinussatz $|x - y|^2 = |x|^2 + |y|^2 - 2 \cos(\angle(x, y))|x||y|$ folgt nun direkt durch Ausmultiplizieren von $|x - y|^2 = \langle x - y, x - y \rangle$. und Einsetzen der $\cos(\angle(x, y))$ -Definition.

KOROLLAR 6.13. $|\cdot|$ ist eine Norm auf \mathbb{R}^n .

Beweis: Interessant ist nur die Überprüfung der Δ -Ungleichung $\forall x, y \in \mathbb{R}^n : |x + y| \leq |x| + |y|$, wir können uns natürlich auf den Fall $|x + y| > 0$ beschränken. Dann

$$|x + y|^2 = \langle x + y, x + y \rangle = \langle x, x + y \rangle + \langle y, x + y \rangle \leq |x||x + y| + |y||x + y| = (|x| + |y|)|x + y|$$

und nach Division durch $|x + y|$ folgt das Gewünschte.

DEFINITION 6.14. Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf \mathbb{R}^n sind äquivalent, wenn für jede Folge $(x_k)_{k=1}^\infty$ in \mathbb{R}^n gilt

$$\lim_{k \rightarrow \infty} \|x_k\| = 0 \Leftrightarrow \lim_{k \rightarrow \infty} \|x_k\|' = 0.$$

SATZ 6.15. Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf \mathbb{R}^n sind äquivalent, genau dann wenn es eine Konstante $C \geq 1$ gibt, so dass

$$\forall x \in \mathbb{R}^n : \frac{1}{C} \|x\| \leq \|x\|' \leq C \|x\|.$$

PROPOSITION 6.16. Die Normen $\|\cdot\|$, $\|\cdot\|_1$ und $\|\cdot\|_\infty$ sind paarweise äquivalent, genauer gilt

$$\forall x \in \mathbb{R}^n : \|x\|_\infty \leq |x| \leq \|x\|_1 \leq n \|x\|_\infty.$$

BEMERKUNG 6.17.

- aus der Norm werden wir den Abstand von x nach y als $\|x - y\|$ festlegen. Wenn die Normen äquivalent sind, sind die daraus erhaltenen Abstände vergleichbar und somit geben äquivalente Normen die gleichen Grenzwerte, konvergenten Folgen und stetigen Funktionen - alle diese Begriffe lassen sich auf die Definition der Nullfolgen zurückführen, die per Definition für zwei äquivalente Normen übereinstimmen.
- es gilt (Beweis folgt): auf jedem \mathbb{R}^n ist jedes Paar von Normen äquivalent.

Definition 6.6' Sei V ein Vektorraum über dem Körper \mathbb{K} . Eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ – ist eine **Norm** auf V , wenn gilt

- $\forall x \in V : \|x\| \geq 0$ Positivität der Norm,
- $\|x\| = 0 \Rightarrow x = 0$ Definitheit der Norm,
- $\forall \lambda \in \mathbb{K} \forall x \in V : \|\lambda x\| = |\lambda| \|x\|$ (positive 1-) Homogenität der Norm
- $\forall x, y \in V : \|x + y\| \leq \|x\| + \|y\|$ die Δ -Ungleichung

BEISPIEL 6.18. a) Sei $M \neq \emptyset$ eine Menge, dann ist

$$\mathcal{B}(M) = \mathcal{B}_{\mathbb{K}}(M) = \{f : M \rightarrow \mathbb{K} : f(M) \text{ ist beschränkt.}\}$$

ein Vektorraum über dem Körper \mathbb{K} ($= \mathbb{R}, \mathbb{C}, \mathbb{Q}$) und wir können auf $\mathcal{B}(M)$ die (Supremums)Norm

$$\|f\|_{\infty} = \sup\{|f(x)| : x \in M\}$$

definieren. Wenn M endlich ist, dann ist $\mathcal{B}_{\mathbb{K}}(M)$ isomorph zu $\mathbb{K}^{\#(M)}$ und die Norm entspricht der Schachbrettnorm aus Beispiel 6.8.ii). Wenn M unendlich ist, dann hat $\mathcal{B}(M)$ unendliche Dimension, da die Menge $\{\chi_{\{x\}} : x \in M\}$ eine unendliche, linear unabhängige Teilmenge von $\mathcal{B}(M)$ – aber keine Basis! Der Raum $\mathcal{B}_{\mathbb{K}}(M)$ enthält im Allgemeinen viele aus analytischer Sicht unbrauchbare Funktionen. Wir werden deshalb vor allem die Unterräume

$$C([a, b]) \subset \mathfrak{R}([a, b]) \subset \mathcal{B}([a, b]), a < b \text{ reell,}$$

wiederum mit der Supremumsnorm $\|\cdot\|_{\infty}$ betrachten, beide haben auch unendliche Dimension.

- b) Für $p \in [1, \infty]$, $a < b$ reell und $f \in \mathfrak{R}([a, b])$ definieren wir analog zu Beispiel 6.8.iii)

$$\|f\|_p = \sqrt[p]{\int_a^b |f(x)|^p dx}.$$

Dieses Funktional erfüllt klarerweise Positivität i), Homogenität iii) und mit einiger Arbeit (unter Benutzung der Youngschen und Integral-Hölderungleichung, vergleiche 4.31, 4.32) kann man die Dreiecksungleichung iv) aus Def 6.6' zeigen. Allerdings gilt die Definitheit ii) nicht, deshalb wird $\|\cdot\|_p$ nur als Halbnorm bezeichnet. Auf dem Teilraum $C([a, b]) \subset \mathfrak{R}([a, b])$ ist $f \mapsto \|f\|_p$ aber eine (vollwertige) Norm, die sogenannte L^p -Norm.

Analysis II betreiben wir im Prinzip im \mathbb{R}^n , häufig aber auch in dessen Untermengen, die keine "Vektorraum"struktur im Sinne unseres " $\cdot, +$ " tragen. Zum Beispiel, auf der

Sphäre $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$ als Modell der Erdoberfläche ist es nicht möglich die Mitte zwischen 2 Punkten x, y als $z = \frac{1}{2}(x + y)$ (oder überhaupt allgemein) zu definieren, da $|z| < 1$. Trotzdem macht es auch dann Sinn, über den Abstand zwischen zwei beliebigen Punkten zu reden, und es zeigt sich, dass dieser Abstandsbegriff ausreicht, um eine sinnvolle Theorie stetiger Funktionen zu erhalten.

DEFINITION 6.19. *Ein metrischer Raum besteht aus einer Menge X und einer Abbildung $\varrho : X \times X \rightarrow \mathbb{R}$, der sogenannten Metrik, die folgende Eigenschaften hat*

- i) $\forall x \in X \forall y \in X : \varrho(x, y) \geq 0$ *Positivität*
- ii) $\forall x \in X \forall y \in X : \varrho(x, y) = 0 \Rightarrow x = y$ *Definitheit*
- iii) $\forall x \in X \forall y \in X : \varrho(x, y) = \varrho(y, x)$ *Symmetrie*
- iv) $\forall x \in X \forall y \in X \forall z \in X : \varrho(x, z) \leq \varrho(x, y) + \varrho(y, z)$ Δ -*Ungleichung*

$\varrho(x, y)$ stellt also den Abstand zwischen x und y dar.

BEISPIEL 6.20.

- a) Wenn $\|\cdot\|$ eine Norm (auf $X = \mathbb{R}^n$) ist, dann ist $\varrho(x, y) = \|x - y\|$ eine Metrik.
- b) Wenn (X, ϱ) ein metrischer Raum ist und $Y \subset X$, dann ist $(Y, \varrho|_{Y \times Y})$, d.h. Y mit der Einschränkung der Metrik auf Punkte(paare) aus Y , ein metrischer Raum - der "metrische Teilraum".
- c) Technisch nützlich (aber eigentlich ohne jegliche Information die spezifische Situation betreffend) ist manchmal die sogenannte "diskrete Metrik" auf einer (beliebigen) Menge X definiert als

$$\varrho(x, y) = \begin{cases} 0 & \text{wenn } x = y, \\ 1 & \text{sonst.} \end{cases}$$

- d) Auf $X = \mathbb{S}^1 = \{z \in \mathbb{R}^2 : |z| = 1\} = \{(\cos(\phi), \sin(\phi)) : \phi \in [0, 2\pi]\}$ definiert $\varrho((\cos(\phi), \sin(\phi)), (\cos(\psi), \sin(\psi))) = \min\{|\phi - \psi|, 2\pi - |\phi - \psi|\}$ eine (natürliche) Metrik - die "Länge" (wenn auch noch mathematisch zu definieren) des kürzesten der (beiden) Kreisbögen, die die beiden Punkte verbinden.

Wir versuchen nun, mit unserem verallgemeinerten Abstandsbegriff grundlegende geometrische Konzepte nachzubilden.

DEFINITION 6.21. *Sei (X, ϱ) ein metrischer Raum, $x \in X$ und $r > 0$. Dann ist*

$B_r(x) = B(x, r) = \{y \in X : \varrho(x, y) < r\}$ die offene ϱ -Kugel mit Radius r und Mittelpunkt x

$\bar{B}_r(x) = \bar{B}(x, r) = \{y \in X : \varrho(x, y) \leq r\}$ die abgeschlossene ϱ -Kugel mit Rad. r & Mpkt x

DEFINITION 6.22. *Eine Teilmenge M eines metrischen Raumes X ist **offen (in X)** wenn $\forall x \in M \exists r > 0 : B_r(x) \subset M$.*

BEISPIEL 6.23. *Für $X = \mathbb{R}$ mit $\varrho(x, y) = |x - y|$ ist ein beschränktes Intervall offen im Sinne der Definition 2.4 genau dann wenn es offen im Sinne von Analysis I (d.h. vom Typ (a, b)) ist.*

PROPOSITION 6.24. *In einem metrischen Raum ist jede offene Kugel offen in X .*

SATZ 6.25. *Sei X ein metrischer Raum.*

- \emptyset und X sind immer offen in X .
- Wenn I eine beliebige (Index)Menge ist und für jedes $i \in I$ ist M_i offen in X , dann ist auch $\bigcup_{i \in I} M_i$ offen in X .
- Wenn I eine **endliche** (Index)Menge ist und für jedes $i \in I$ ist M_i offen in X , dann ist auch $\bigcap_{i \in I} M_i$ offen in X .

BEISPIEL 6.26. **Warnungen:**

- $M_i = (-\frac{1}{i}, \frac{1}{i})$ ist offen $\forall i \in \mathbb{N}_+$, aber $M = \bigcap_{i=1}^{\infty} M_i = \{0\}$ nicht offen.
- Der Durchschnitt von zwei offenen Kugeln ist natürlich immer offen, kann allerdings recht untypisch aussehen, z.B. für $x = (1, 0)$ und $y = (-1, 0)$ in $X = \mathbb{S}^1$ aus Beispiel 2.2.d sowie $r \in (\pi/2, \pi)$ besteht $B_r(x) \cap B_r(y)$ aus zwei Kreisbögen!

Deshalb darf man sich beim Studium offener Mengen (und damit zusammenhängender weiterer Konzepte) von der geometrischen Intuition nur leiten lassen, diese jedoch nicht als Beweis betrachten. Diese Studien fallen in das mathematische Gebiet der Topologie, die sich nicht auf die klassischen geometrischen Eigenschaften von Mengen konzentriert, sondern den Schwerpunkt auf die robusteren Eigenschaften, welche auch bei einer (stetigen) Deformation (z.B. Euklidische Kugel wird Würfel) ihre Gültigkeit behalten, legt.

Offenheit ist eine Eigenschaft, die eine Menge "überall" haben muss, wenn man nur einen einzelnen Punkt betrachtet, scheint folgendes Konzept natürlich.

DEFINITION 6.27. *Eine Menge U in einem metrischen Raum X ist eine Umgebung eines Punktes $x \in X$ wenn es ein M offen in X gibt, so dass $x \in M \subset U$.*

BEMERKUNG 6.28. *In jedem metrischen Raum X gilt:*

- Sei $x \in X$ und $U \subset X$, dann ist U Umgebung von $x \Leftrightarrow \exists r > 0 : B_r(x) \subset U$.
- Eine Menge ist offen (in X) \Leftrightarrow sie Umgebung jedes ihrer Punkte ist.
- Wenn $U \subset V \subset X$ und U ist Umgebung von $x \in X$, dann ist auch V Umgebung von x .
- Der Durchschnitt von endlichen vielen Umgebungen eines beliebigen Punktes x ist wiederum Umgebung von x .

DEFINITION 6.29. *Sei X ein metrischer Raum und $M \subset X$. Dann ist $x \in X$*

- ein Berührungspunkt von M wenn jede Umgebung von x ein $y \in M$ enthält
- ein Häufungspunkt von M wenn jede Umgebung von x ein $y \in M \setminus \{x\}$ enthält.

BEMERKUNG 6.30. *Sei (X, ρ) ein metrischer Raum, $M \subset X$. Dann*

- x ist Häufungspunkt von M genau dann wenn jede Umgebung von x unendlich viele Punkte aus M enthält
- x Berührungspunkt von $M \Leftrightarrow (x \in M \text{ oder } x \text{ Häufungspunkt von } M)$.

DEFINITION 6.31. Sei (X, ϱ) ein metrischer Raum. Dann ist $M \subset X$ abgeschlossen (in X) wenn $X \setminus M$ offen (in X) ist.

SATZ 6.32. In jedem metrischen Raum (X, ϱ) ist für ein $M \subset X$ Folgendes äquivalent

- M ist abgeschlossen.
- M enthält jeden seiner Berührungspunkte.
- M enthält jeden seiner Häufungspunkte.

BEISPIEL 6.33. $[a, b]$ in \mathbb{R} , $\bar{B}_r(x)$ in X abgeschlossen.

SATZ 6.34. Sei X ein metrischer Raum.

- \emptyset und X sind immer abgeschlossen in X .
- Wenn I eine beliebige (Index)Menge ist und für jedes $i \in I$ ist M_i abgeschlossen in X , dann ist auch $\bigcap_{i \in I} M_i$ abgeschlossen in X .
- Wenn I eine **endliche** (Index)Menge ist und für jedes $i \in I$ ist M_i abgeschlossen in X , dann ist auch $\bigcup_{i \in I} M_i$ abgeschlossen in X .

BEISPIEL 6.35. **Warnungen:**

- $M_l = \{\frac{k}{l} : k \in \mathbb{Z}\} = \mathbb{R} \setminus \bigcup_{k \in \mathbb{Z}} (\frac{k}{l}, \frac{k+1}{l})$ ist abgeschlossen $\forall l \in \mathbb{N}_+$, aber $M = \bigcup_{l=1}^{\infty} M_l = \mathbb{Q}$ nicht abgeschlossen (alle $x \in \mathbb{R}$ sind Berührungspunkte!!).
- Es gilt **NICHT**, dass jede Menge entweder offen oder abgeschlossen ist, zB. ist $[0, 1)$ keines von beiden, während \mathbb{R} beides (in \mathbb{R}) ist.

KOROLLAR 6.36. Jede endliche Menge in einem metrischen Raum ist abgeschlossen.

DEFINITION 6.37. Eine Folge $(x_k)_{k=1}^{\infty}$ in einem metrischen Raum (X, ϱ) hat Grenzwert $x \in X$ wenn $\lim_{k \rightarrow \infty} \varrho(x_k, x) = 0$, wir schreiben dann $\lim_{k \rightarrow \infty} x_k = x$ oder $x_k \rightarrow x$ (wenn $k \rightarrow \infty$).

PROPOSITION 6.38. Für jede Folge $(x_k)_{k=1}^{\infty}$ in einem metrischen Raum (X, ϱ) , $x \in X$ sind äquivalent

- $\lim_{k \rightarrow \infty} x_k = x$
- $\forall \varepsilon > 0 \exists K \in \mathbb{N} \forall k \geq K : x_k \in B_\varepsilon(x)$.
- Für jede Umgebung U von $x \exists K \in \mathbb{N} \forall k \geq K : x_k \in U$. (topologische Definition der Folgenkonvergenz).

SATZ 6.39. In einem beliebigen metrischen Raum kann jede Folge höchstens einen Grenzwert haben.

KOROLLAR 6.40. Seien $\|\cdot\|, \|\cdot\|'$ zwei äquivalent Normen auf \mathbb{R}^n mit den dazugehörigen Metriken ϱ und ϱ' . Dann ist $(x_k)_k \in \mathbb{R}^n$ im Sinne von ϱ konvergent gdw sie auch im Sinne von ϱ' konvergent ist. Die Grenzwerte stimmen in beiden Metriken überein.

BEISPIEL 6.41. Sei $\|\cdot\|$ eine Norm auf \mathbb{R}^n , $(x_k)_k$ Folge in \mathbb{R}^n . Dann $x_k \rightarrow 0$ gdw. $\forall i \leq n : \lim_{k \rightarrow \infty} (x_k)_i = 0$, d.h die x_k konvergieren koordinatenweise. Beweis jetzt nur für $\|\cdot\|_p$ (allgemein folgt aus Normäquivalenz in Kapitel 4).

DEFINITION 6.42. Sei (X, ϱ) ein metrischer Raum und $M \subset X$. Dann

- a) $\text{int}(M) = \{x : M \text{ ist Umgebung von } x\}$ ist das Innere von M , die Elemente von $\text{int}(M)$ sind die inneren Punkte von M .
- b) $\overline{M} = \{x : x \text{ ist Berührungspunkt von } M\}$ ist der Abschluss von M .

SATZ 6.43. $\text{int}(M)$ ist die größte offene Menge, die in M enthalten ist, \overline{M} die kleinste abgeschlossene Menge, die M enthält. Überdies gilt

$$X \setminus \text{int}(M) = \overline{X \setminus M}.$$

BEISPIEL 6.44. Für jede Norm auf \mathbb{R}^n , jedes $r > 0$ und $x \in \mathbb{R}^n$ gilt

$$\text{int}(\overline{B}_r(x)) = B_r(x), \overline{B_r(x)} = \overline{B}_r(x).$$

Dies gilt nicht in allen metrischen Räumen!!

DEFINITION 6.45. Sei (X, ρ) ein metrischer Raum und $M \subset X$. Dann ist x ein Randpunkt von M wenn x Berührungspunkt von M und ein Berührungspunkt von $X \setminus M$ ist. Die Menge aller Randpunkte von M wird mit ∂M bezeichnet und bildet den Rand von M .

PROPOSITION 6.46. Sei (X, ρ) ein metrischer Raum und $M \subset X$. Dann gilt $\partial M = \overline{M} \cap \overline{X \setminus M}$, also ist der Rand von M abgeschlossen, und

$$X = \text{int}(M) \cup \partial M \cup \text{int}(X \setminus M),$$

wobei alle 3 Mengen paarweise disjunkt sind.

BEISPIEL 6.47. **Warnung:** $\partial \mathbb{Q} = \mathbb{R}$, $\text{int}(\mathbb{Q}) = \emptyset$, d.h der Rand einer Menge kann viel größer sein als die Menge selbst.

BEMERKUNG 6.48. Das System aller offenen Mengen hängt in der Tat von der Metrik ab, z.B. für ρ die diskrete Metrik aus 2.2.c) ist jede Menge offen.

DEFINITION 6.49. Der metrische Raum ist vollständig wenn jede Cauchyfolge in X einen Grenzwert hat. Das heißt, falls $(x_k)_{k=1}^{\infty}$ eine Folge in X ist, so dass für jedes $\varepsilon > 0$ ein N mit $n, m \geq N \Rightarrow \rho(x_n, x_m) < \varepsilon$ existiert, dann gibt es ein $x \in X$ mit $\lim_k \rho(x_k, x) = 0$.

SATZ 6.50. Sei (X, ρ) ein metrischer Raum und $Y \subset X$ dessen metrischer Unterraum (mit der Metrik $\rho_{Y \times Y}$). Dann gilt

- a) Wenn Y vollständig ist, dann ist Y abgeschlossen (in X).
- b) Wenn X vollständig ist und Y abgeschlossen in X , dann ist Y vollständig.

BEISPIEL 6.51. **Warnung** Die Vollständigkeit eines metrischen Raumes ist nicht nur durch das System seiner offenen Mengen bestimmt!

Wir betrachten $X = (-\pi/2, \pi/2)$ mit den Metriken $\rho_1(x, y) = |x - y|$ und $\rho_2(x, y) = |\tan(x) - \tan(y)|$, da \tan auf X eine injektive Funktion ist, sieht man leicht, dass auch ρ_2 alle geforderten Eigenschaften einer Metrik aufweist. Ausserdem gilt, für eine beliebige Folge $(x_k)_k$ und jeden Punkt x in X dass für $k \rightarrow \infty$

$$\rho_2(x_k, x) \rightarrow 0 \Leftrightarrow \tan(x_k) \rightarrow \tan(x) \Leftrightarrow x_k \rightarrow x \Leftrightarrow \rho_1(x_k, x) \rightarrow 0,$$

wobei alle Grenzwerte im Sinne von Analysis I genommen wurden und wir nutzen, dass $\tan|_X$ und $\arctan|_{\mathbb{R}}$ stetig sind. Also sind genau die gleichen Folgen konvergent bzgl ρ_1 und

ϱ_2 , somit sind auch die beiden Familien ϱ_1 - bzw. ϱ_2 -abgeschlossener Mengen gleich (siehe ÜZ 7, 2.a)) und dann auch die beiden Familien ϱ_1 - bzw. ϱ_2 -offener Mengen identisch. Andererseits ist der Raum (X, ϱ_1) nicht vollständig, da dieser Unterraum des $(\mathbb{R}, \varrho_{eucl})$ nicht abgeschlossen in \mathbb{R} ist, hingegen ist (X, ϱ_2) vollständig - da dieser Raum isometrisch zu $(\mathbb{R}, \varrho_{eucl})$ ist. Genauer, wenn $(x_k)_k$ eine Cauchyfolge in (X, ϱ) ist, dann ist $(\tan(x_k))_k$ eine Cauchyfolge in der euklidischen reellen Achse. Diese ist, wie in Analysis I gezeigt, vollständig und somit existiert $y \in \mathbb{R}$ mit $\tan(x_k) \rightarrow y$. Wie oben hergeleitet, bedeutet das $\lim_k \varrho_2(x_k, \arctan(y)) = 0$, also ist $\arctan(y)$ der geforderte ϱ_2 -GW von $(x_k)_k$.

Dieses Phänomen beruht auf dem Fakt, dass die Familie der offenen Menge nicht bestimmt, welche Folgen Cauchyfolgen sind, zum Beispiel ist $x_k = (\pi/2) - (1/k)$ eine Cauchyfolge bzgl ϱ_1 aber keinesfalls bzgl ϱ_2 .

7. Stetigkeit und Kompaktheit

Wir wollen nun die Stetigkeit von Funktionen auf eine Art definieren, welche die Analysis I Definition auf natürliche Weise (d.h. $|\cdot|$ als Abstandsmaß wird einfach durch die Metrik ersetzt) verallgemeinert. Es zeigt sich, dass wir Stetigkeit auch alleine mit Hilfe offener oder abgeschlossener Mengen (d.h. in der Sprache der sogenannten "Topologie") charakterisieren können.

DEFINITION 7.1. Seien (X, ϱ) , (Y, σ) metrische Räume und $f : X \rightarrow Y$. Wir sagen, die Funktion f ist im Punkt $x_0 \in X$ stetig, wenn

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in X : \varrho(x, x_0) < \delta \Rightarrow \sigma(f(x), f(x_0)) < \varepsilon.$$

Die Abbildung f heißt stetig (auf X), wenn sie in jedem $x \in X$ stetig ist.

Vollkommen analog zu Analysis I zeigt man, dass $f : (X, \varrho) \rightarrow (Y, \sigma)$ in $x_0 \in X$ stetig ist genau dann wenn $x_k \rightarrow x_0$ in (X, ϱ) impliziert, dass $f(x_k) \rightarrow f(x_0)$ in (Y, σ) .

LEMMA 7.2. Seien (X, ϱ) , (Y, σ) metrische Räume, $f : X \rightarrow Y$ und $x_0 \in X$. Dann ist f stetig in x gdw. für jede Umgebung U von $f(x_0)$ das Urbild $f^{-1}(U)$ eine Umgebung von x_0 ist. Analog ist f stetig (auf X), wenn für jedes $U \subset Y$ offen in Y das Urbild $f^{-1}(U)$ offen in X ist.

KOROLLAR 7.3. Wenn (X_1, ϱ_1) , (X_2, ϱ_2) und (X_3, ϱ) metrische Räume sind, $f : X_1 \rightarrow X_2$, $g : X_2 \rightarrow X_3$ und falls f in $x \in X_1$ und g in $f(x)$ stetig, dann ist $g \circ f$ in x stetig. Wenn f stetig ist (auf X_1) und g stetig ist (auf X_2), dann ist $g \circ f$ stetig (auf X_1).

BEISPIEL 7.4.

- a) Wenn $\forall x \in X f(x) = y_0 \in Y$ fest, dann ist f stetig
- b) wenn $(Y, \sigma) = (X, \varrho)$, dann ist $f(x) = x$ stetig
- c) **wichtig:** Für jedes $x_0 \in X$ ist $d_{x_0}(x) = \varrho(x, x_0)$ stetig, dies ist ein wichtiger (erster) Schritt zur Konstruktion vieler stetiger Funktionen auf einem allgemeinen metrischen Raum. (zB. gilt wie in Analysis I: $f, g : (X, \varrho) \rightarrow \mathbb{R}$ stetig, dann auch

$f + g, f \cdot g$ stetig.) Insbesondere ist d_{x_0} Lipschitz stetig (mit Lipschitz-Konstante L kleiner gleich 1), das heißt

$$\forall x, y \in X : \varrho_{\text{eucl}}(d_{x_0}(x), d_{x_0}(y)) \leq L \varrho(x, y), \text{ für } L = 1.$$

DEFINITION 7.5. Seien $(X, \varrho), (Y, \sigma)$ metrische Räume und $f : X \rightarrow Y$. Wir sagen f ist gleichmäßig stetig wenn es für jedes positive ε ein $\delta > 0$ gibt, so dass aus $\varrho(x, y) < \delta$ immer $\sigma(f(x), f(y)) < \varepsilon$ folgt.

SATZ 7.6. Seien $(X, \varrho), (Y, \sigma)$ metrische Räume. Für $f : X \rightarrow Y$ sind folgende Aussagen äquivalent

- i) f ist stetig
- ii) $G \subset Y$ offen (in Y) $\Rightarrow f^{-1}(G)$ offen (in X)
- iii) $F \subset Y$ abgeschlossen (in Y) $\Rightarrow f^{-1}(F)$ abgeschlossen (in X)
- iv) wenn $\lim_k x_k = x$ in X , dann $\lim_k f(x_k) = f(x)$ in Y (man sagt, f ist folgenstetig).

KOROLLAR 7.7. Seien $(X, \varrho), (Y, \sigma)$ metrische Räume, $f, g : X \rightarrow Y$ stetig. Dann ist $\{x \in X : f(x) = g(x)\}$ abgeschlossen (in X).

Wenn $(Y, \sigma) = (\mathbb{R}, \varrho_{\text{euc}})$, also $\varrho_{\text{euc}}(x, y) = |x - y|$, dann ist auch $\{x \in X : f(x) \leq g(x)\}$ abgeschlossen (in X).

PROPOSITION 7.8. Wir betrachten \mathbb{R}^n mit einer beliebigen Norm $\|\cdot\|$ und $f : (X, \varrho) \rightarrow \mathbb{R}^n$, wobei also $f(x) = (f_1(x), \dots, f_n(x))$. Dann ist f stetig gdw. für jedes $i \leq n$ f_i stetig ist.

Wiederum können wir jetzt den Beweis nur für eine der $\|\cdot\|_p$ -Normen führen. Um die Äquivalenz aller Normen auf dem \mathbb{R}^n zu zeigen benötigen wir noch den Begriff der kompakten Mengen.

Wir untersuchen nun, für welche metrischen Räume der Satz von Bolzano-Weierstrass gilt.

DEFINITION 7.9. Ein metrischer Raum (X, ϱ) ist (folgen)kompakt, wenn jede Folge in X eine konvergente Teilfolge hat.

BEMERKUNG 7.10.

- i) Eine Menge M in einem (X, ϱ) bezeichnen wir als kompakt, wenn der metrische Teilraum $(M, \varrho_{M \times M})$ kompakt ist, also wenn $\forall (x_k)_1^\infty$ in M eine Teilfolge $(k_m)_1^\infty$ und $x \in M$ (!) existieren, so dass $\lim_m x_{k_m} = x$.
- ii) Ein beschränktes oder unbeschränktes Intervall I in \mathbb{R} ist kompakt gdw es $a \leq b \in \mathbb{R}$ mit $I = [a, b]$ gibt.

DEFINITION 7.11. Sei (X, ϱ) ein metrischer Raum

- a) Wir sagen X ist beschränkt, wenn $\text{diam}(X) = \sup\{\varrho(x, y) : x, y \in X\} < \infty$, d.h. der Durchmesser von X endlich ist ($\text{diam}(\emptyset) = 0$). Klarerweise gilt dies (für $X \neq \emptyset$) genau dann wenn es ein $x_0 \in X$ und ein $R \in (0, \infty)$ mit $X = B(x_0, R)$ gibt (und dann gilt $\text{diam}(X) \leq 2R$ sowie $\forall x \in X : X \subset \bar{B}(x, \text{diam}(X))$).

- b) Wir sagen X ist total beschränkt wenn es für jedes positive ε endlich viele Punkte x_1, \dots, x_m , $m = m_\varepsilon$, gibt, so dass $X = \bigcup_{j=1}^m B(x_j, \varepsilon)$. Das ist klarerweise äquivalent zur Forderung, dass es für jedes $\varepsilon > 0$ endlich viele M_1, \dots, M_k , $k = k_\varepsilon$, Teilmengen von X gibt mit $X = \bigcup_{j=1}^k M_j$ und $\forall j : \text{diam}(M_j) < \varepsilon$.

BEMERKUNG 7.12. Wenn (X, ρ) ein total beschränkter metrischer Raum, dann ist jeder metrische Unterraum davon wieder total beschränkt.

SATZ 7.13. Sei (X, ρ) ein metrischer Raum. Dann ist X kompakt genau dann wenn X vollständig und total beschränkt ist.

KOROLLAR 7.14. (Heine-Borel) Sei $\|\cdot\| = \|\cdot\|_p$, $p \in [1, \infty]$, auf \mathbb{R}^n . Dann ist $M \subset \mathbb{R}^n$ kompakt gdw. M abgeschlossen und beschränkt ist.

SATZ 7.15. Wenn $f : (X, \rho) \rightarrow (Y, \sigma)$ stetig ist und X kompakt, dann ist f gleichmäßig stetig und $f(X) \subset Y$ auch kompakt.

Wenn ausserdem f injektiv ist, dann ist die Umkehrfunktion $f^{-1} : (f(X), \sigma) \rightarrow (X, \rho)$ auch stetig.

Insbesondere, wenn $(Y, \sigma) = (\mathbb{R}, \rho_{\text{eucd}})$, dann existieren $\max(f)(X)$, $\min(X)$.

KOROLLAR 7.16. Jede Norm $\|\cdot\|$ auf \mathbb{R}^n ist äquivalent zu $\|\cdot\|_1$ und somit allen Normen auf dem \mathbb{R}^n . Deshalb gilt Korollar ?? für alle Normen auf \mathbb{R}^n !

Nun betrachten wir abschliessend der Raum aller stetigen Funktionen als metrischen Raum und wollen seine kompakten Teilmengen finden.

DEFINITION 7.17. Sei X eine Menge und (Y, σ) ein metrischer Raum. Dann führen wir auf

$$\mathfrak{B}(X, Y) = \{f : X \rightarrow Y : f(X) \subset Y \text{ ist beschränkt}\}$$

die (Supremums)Metrik ein:

$$\sigma_\infty(f, g) = \sup\{\sigma(f(x), g(x)) : x \in X\}.$$

Wenn auch auf X eine Metrik ρ definiert ist, dann betrachten wir den Unterraum

$$C_b(X, Y) = \{f \in \mathfrak{B}(X, Y) : f : (X, \rho) \rightarrow (Y, \sigma) \text{ ist stetig}\}.$$

BEMERKUNG 7.18.

- a) σ_∞ ist eine Metrik, welche die Supremumsnorm aus Beispiel 6.18 in den metrischen Kontext verallgemeinert, d.h. wenn $(Y, \|\cdot\|)$ ein normierter Raum ist und $\sigma(y_1, y_2) = \|y_1 - y_2\|$ die induzierte Metrik ist, dann ist $\mathfrak{B}(X, Y)$ ein linearer Raum und für alle $f, g \in \mathfrak{B}(X, Y)$ ist

$$\sigma_\infty(f, g) = \|f - g\|_\infty = \{\|(f - g)(x)\| : x \in X\},$$

das heißt, σ_∞ ist die von der Norm $\|\cdot\|_\infty$ auf $\mathfrak{B}(X, Y)$ induzierte Metrik.

- b) Für $(f_k)_{k=1}^\infty$ und f aus $\mathfrak{B}(X, \mathbb{K})$ gilt natürlich

$$f_k \rightrightarrows f \text{ für } k \rightarrow \infty \Leftrightarrow \sigma_\infty(f_k, f) \rightarrow 0 \text{ für } k \rightarrow \infty.$$

Man sagt, die Supremumsmetrik “topologisiert” die gleichmäßige Konvergenz – d.h. beschreibt diesen Sachverhalt durch Umgebungen bzw. offene Mengen im Raum $\mathfrak{B}(X, Y)$, so wie ρ_{eucl} die gewöhnliche Konvergenz durch offene Teilmengen der reellen Achse beschreibt.

PROPOSITION 7.19. Sei $X \neq \emptyset$, dann ist der metrische Raum (Y, σ) vollständig genau dann wenn $(\mathfrak{B}(X, Y), \sigma_\infty)$ vollständig ist.

Falls auch auf X eine Metrik ρ definiert ist, dann ist $C_b(X, Y)$ eine abgeschlossene Untermenge von $\mathfrak{B}(X, Y)$ und vollständig genau dann wenn (Y, σ) vollständig.

BEISPIEL 7.20. Sei

$$f_m(x) = \begin{cases} 2mx & x \in [0, \frac{1}{2m}] \\ 1 - 2m(x - \frac{1}{2m}) & x \in [\frac{1}{2m}, \frac{1}{m}] \\ 0 & x \in [\frac{1}{m}, 1] \end{cases}$$



Dann $\forall m \|f_m\|_\infty \leq 1$ d.h. $(f_m)_m \subset \overline{B}_1(0)$ in $C([0, 1], \mathbb{R})$ und klarerweise $f(x) \equiv 0$ ist punktweser Grenzwert (f_m) aber $f_m \not\rightarrow f$. Analog gilt für jede Teilfolge (m_k) , dass $f_{m_k} \rightarrow 0$ aber $\sup_x |f_{m_k}(x) - f(x)| = 1 \forall k$, also gibt es keine in $\|\cdot\|_\infty$ konvergente Teilfolge von (f_m) .

Wir wissen, aufgrund der Vollständigkeit von $C([0, 1])$, dass die bedeutet: aus $(f_m)_m$ kann man keine Cauchfolge auswählen, und der Beweis von Satz 7.13 sagt uns, dass die Menge $\{f_m : m \in \mathbb{N}\}$ nicht total beschränkt sein kann. Man kann dies auch direkt sehen, wenn man bemerkt, dass $\|f_m - f_k\|_\infty \geq |f_m(\frac{1}{2m}) - f_k(\frac{1}{2m})| = |1 - 0| = 1$ wenn $k \geq 2m$.

Da $f_n \xrightarrow{loc} 0$ in $(0, 1]$, muss die mangelnde Kompaktheit durch das Verhalten der Funktionen (f_m) bei 0 entstehen. Dort tritt die maximale Oszillation (“1”) der Funktionen auf beliebig kleinen Intervallen auf.

Folgende Definition schließt das aus:

DEFINITION 7.21. Seien (X, ρ) , (Y, σ) metrische Räume. Eine Menge $\mathcal{F} = \{f_\alpha | \alpha \in A\}$ von Funktionen $f_\alpha : X \rightarrow Y$ ist gleichgradig stetig, wenn

$$\forall \epsilon > 0 \exists \delta > 0 \forall \alpha \in A \forall x_1, x_2 \in X : \rho(x_1, x_2) < \delta \Rightarrow \sigma(f_\alpha(x_1), f_\alpha(x_2)) < \epsilon.$$

BEMERKUNG 7.22. $\mathcal{F} = \{f\}$ ist gleichgradig stetig $\Leftrightarrow f$ gleichmäßig stetig, das heißt gleichgradig stetig bedeutet die gleichmäßige Stetigkeit der f_α – über alle Paare $(x_1, x_2) \in X^2$ – ist auch noch gleichmässig über alle $\alpha \in A$.

LEMMA 7.23. Sei (X, ρ) kompakt und (Y, σ) beliebig.

- a) Sei $f : X \rightarrow Y$ stetig, dann ist $f \in C_b(X, Y)$, also $C(X, Y) = C_b(X, Y)$.
- b) Sei $\mathcal{F} = \{f_1, \dots, f_m\}$ eine endliche Menge stetiger Funktionen, dann ist \mathcal{F} gleichgradig stetig.
- c) $f_m : X \rightarrow Y$ stetig und $f_m \rightarrow f$ dann ist $\{f\} \cup \{f_m | m \in \mathbb{N}\}$ gleichgradig stetig.

SATZ 7.24. (Arzela-Ascoli) Seien (X, ρ) und (Y, σ) kompakt. Dann ist $\mathcal{F} \subset C(X, Y)$ total beschränkt genau dann wenn \mathcal{F} gleichgradig stetig ist.

BEMERKUNG 7.25. Was passiert in den wichtigen Fällen, dass X oder Y gleich $\mathbb{R}^{(n)}$ sind?

$X = \mathbb{R}$ oder \mathbb{R}^n ist recht hoffnungslos, z.B. sind die Funktionen definiert als $f_m(x) = \max\{0, 1 - |m - x|\}$ alle 1-lipschitz, also gleichgradig stetig mit $\delta = \varepsilon$ und $f_m : X \rightarrow Y = [0, 1]$ kompakt. Aber $\|f_k - f_m\|_\infty = 1$ da $(f_k - f_m)(k) = 1 - 0 = 1$ wenn $k \neq m$, also $\{f_m : m \in \mathbb{N}\}$ ist nicht total beschränkt.

(X, ρ) kompakt aber $Y = \mathbb{R}^n$ ist bedeutend besser, dann ist $\mathcal{F} \subset C(X, Y)$ total beschränkt

$\Leftrightarrow \mathcal{F}$ gleichgradig stetig und beschränkt in $C(X, Y)$

$\Leftrightarrow \mathcal{F}$ gleichgradig stetig und für jedes $x \in X$ ist $\{f(x) : f \in \mathcal{F}\}$ beschränkt in \mathbb{R}^n .

Falls (X, ρ) überdies zusammenhängend ist, z.B. $X = \bar{B}(x, r) \subset \mathbb{R}^n$ oder $X = [a, b]^n$ — und allgemeiner wenn es für alle $x, y \in X$ ein stetiges $c : [0, 1] \rightarrow X$ mit $c(0) = x$, $c(1) = y$ gibt —, dann ist “ \mathcal{F} gleichgradig stetig und es gibt ein $x_0 \in X$ mit $\{f(x_0) : f \in \mathcal{F}\}$ beschränkt in \mathbb{R}^n ” eine weitere äquivalente Bedingung.

Das Beispiel $X = \{0, 1\} \subset \mathbb{R}$, auf dem die Funktionsmenge $\mathcal{F} = \{f : X \rightarrow \mathbb{R} : f(0) = 0\}$ gleichgradig stetig ist, (für jedes $\delta = 1/2$ funktioniert für jedes $\varepsilon > 0$ in Definition 7.21) zeigt, dass die Existenz eines x_0 mit $\{f(x_0), f \in \mathcal{F}\}$ beschränkt im Allgemeinen nicht ausreicht um (totale) Beschränktheit in $C(X, Y)$ zu sichern.

BEMERKUNG 7.26. Wir haben bereits in Beispiel 6.51 gesehen, dass weder die totale Beschränktheit noch die Vollständigkeit alleine unter stetigen Abbildungen erhalten bleiben. Das stetige Bild $\tan((-\pi/2, \pi/2))$ des total beschränkten Intervalles $(-\pi/2, \pi/2)$ ist \mathbb{R} und also unbeschränkt, während das stetige Bild $(-\pi/2, \pi/2) = \arctan(\mathbb{R})$ der vollständigen reellen Achse nicht mehr vollständig ist. Aber Satz 7.13 sagt verblüffenderweise: wenn totale Beschränktheit und Vollständigkeit zusammen auftreten (also der Raum kompakt ist), dann bleiben auch beide unter beliebigen stetigen Abbildungen erhalten.

8. Partielle Ableitungen und totales Differential

NOTATION 8.1. Im Weiteren wird, soweit nicht anders erwähnt, durch G, U oder Ω immer eine offene Teilmenge des \mathbb{R}^n bezeichnet.

DEFINITION 8.2. Sei $f : G \rightarrow \mathbb{R}$ eine Funktion und $x = (x_1, \dots, x_n) \in G$ ein beliebig gewählter Punkt. Für $i \leq n$ ist die Menge

$$G_{i,x} = \{t : te_i + \sum_{j \neq i} x_j e_j = (x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) \in G\} \subset \mathbb{R} \text{ offen}$$

und auf ihr die Partialfunktion

$$f_{i,x}(t) = f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n)$$

definiert. Wenn $f'_{i,x}(x_i)$ existiert, nennen wir $f'_{i,x}(x_i)$ die i -te partielle Ableitung von f in x und schreiben $\frac{\partial f}{\partial x_i}(x) = \frac{\partial}{\partial x_i} f(x)$ dafür.

Falls für alle $i \leq n$ die partielle Ableitung $\frac{\partial f}{\partial x_i}(x)$ existiert, sagen wir dass f in x partiell differenzierbar ist. Wir sagen f ist in G partiell differenzierbar (hat in G partielle Ableitungen) wenn f in jedem x in G partiell differenzierbar ist.

BEMERKUNG 8.3.

- Das Konzept der partiellen Ableitung hat Vor- und Nachteile: Einerseits ist $\frac{\partial f}{\partial x_i}(x)$ einfach auszurechnen, da alle $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ Konstanten sind und beim Differenzieren keine Schwierigkeiten bereiten (sollten!). Andererseits erfasst $\frac{\partial f}{\partial x_i}(x)$ nur Informationen über f entlang der "Koordinatenrichtungen", also entlang einer relativ kleinen Menge (bestehend aus n Geraden im Vergleich zum ganzen \mathbb{R}^n). Dies führt zu solchen Problemen, dass partielle Differenzierbarkeit nicht einmal Stetigkeit impliziert! Die Situation in Analysis I ist viel einfacher. Wir werden deshalb bald ein stärkeres Ableitungskonzept einführen, um solche Schwierigkeiten auszuschliessen (so, dass also alle differenzierbaren Funktionen stetig sind).
- in \mathbb{R}^2 oder \mathbb{R}^3 schreiben wir oft (x, y, z) für (x_1, x_2, x_3) und benutzen dann die Notation

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$$

BEISPIEL 8.4.

- a) $f(x, y) = \sin(x^2 + y)$ für alle $(x, y) \in \mathbb{R}^2$. Dann

$$\frac{\partial f}{\partial x} = \cos(x^2 + y)2x, \quad \frac{\partial f}{\partial y} = \cos(x^2 + y) \text{ im Punkt } (x, y).$$

- b) $g(x, y) = \log(x/y)$ auf $G = \{(x, y) \in \mathbb{R}^2 : xy > 0\}$ (Wieso offen?). Dann

$$\frac{\partial g}{\partial x} = \frac{1}{x}, \quad \frac{\partial g}{\partial y} = \frac{-1}{y} \text{ in } (x, y) \in G,$$

da $g(x, y) = \log(|x|) - \log(|y|)$ und $(\frac{d}{dz}|z|)/|z| = 1/z$ wenn $z \in \mathbb{R} \setminus \{0\}$.

- c) Unstetige Funktion mit partiellen Ableitungen: Wir definieren

$$h(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{wenn } (x, y) \neq (0, 0) \\ 0 & \text{wenn } (x, y) = (0, 0) \end{cases}.$$

Dann ist h ausserhalb des Ursprungs wegen $x^2 + y^2 > 0$ partiell differenzierbar, und $h(0, t) = h(t, 0) = 0$ für alle $t \in \mathbb{R}$ sichert $\frac{\partial h}{\partial x}(0, 0) = \frac{\partial h}{\partial y}(0, 0) = 0$. Trotzdem ist h unstetig in 0 da $\forall \lambda \neq 0$ gilt $h(\lambda x, \lambda y) = h(x, y)$, also ist h entlang Geraden zum Ursprung konstant und die Oszillation von h auf dem Einheitskreis ($h(\cos(\phi), \sin(\phi)) = \sin(2\phi)/2$) pflanzt sich bis zu $(0, 0)$ fort. Zum Beispiel $p_n = (1/n, 1/n) \rightarrow (0, 0)$ aber $h(p_n) = 1/2$ für $n \rightarrow \infty$. Wenn wir

$$\frac{\partial h}{\partial x}(x, y) = \frac{y(y^2 - x^2)}{(x^2 + y^2)^2}, \quad \frac{\partial h}{\partial y}(x, y) = \frac{x(x^2 - y^2)}{(x^2 + y^2)^2}$$

berechnen, sehen wir $\lim_{t \searrow 0} \frac{\partial h}{\partial y}(t, 0) = \lim_{t \searrow 0} \frac{1}{t} = \infty$ und analog $\lim_{t \searrow 0} \frac{\partial h}{\partial x}(0, t) = \infty$, also lässt sich $\frac{\partial h}{\partial y}$ in $(0, 0)$ nicht nochmal nach x partiell differenzieren.

Dieser "Schönheitsfehler" der Konstruktion lässt sich leicht beheben unter Benutzung des Beispiels 4.35 in Analysis I. Wir definieren

$$\tilde{h}(x, y) = \begin{cases} 0 & \text{wenn } h(x, y) \leq \frac{1}{3} \\ \exp\left(\frac{1}{\frac{1}{3} - h(x, y)}\right) & \text{wenn } h(x, y) > \frac{1}{3} \end{cases},$$

dann ist für jedes $(x, y) \neq (0, 0)$ mit $xy = 0$ $h = 0$ in einer Umgebung, somit sind alle (auch höhere partielle) Ableitungen auf den Koordinatenachsen identisch null. Ausserdem sieht man, dass in ganz $\mathbb{R}^2 \setminus \{0\}$ alle partiellen Ableitungen existieren, siehe die folgende Definition.

DEFINITION 8.5. Höhere Ableitungen Sei $f : G \rightarrow \mathbb{R}$ in G partiell differenzierbar. Wenn alle partiellen Ableitungen

$$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} : G \rightarrow \mathbb{R}$$

wiederum in G partiell differenzierbar sind, dann heißt f zweimal partiell differenzierbar und wir schreiben $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right)$. Analog definieren wir höhere partielle Ableitungen (f k -mal differenzierbar).

NOTATION 8.6.

- $\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i} f$, die zweite Schreibweise ist vorteilhaft wenn f eine lange Formel darstellt. (Wir erwähnen auch die Analogie zur klassischen "Leibniznotation" in \mathbb{R} : $f'(x) = \frac{df}{dx}$.)
- $x \rightarrow \frac{\partial f}{\partial x_i}(x)$ wird oft abgekürzt zu $D_i f : G \rightarrow \mathbb{R}$, z.B. $D_1 D_2^2 f = \frac{\partial^3 f}{\partial x_1 \partial x_2^2} = \frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_2}$, weitere Abkürzungen werden folgen.

DEFINITION 8.7. Die Funktion $f : G \rightarrow \mathbb{R}$ ist in $C^k(G)$ wenn $\forall l \leq k$: alle l -ten partiellen Ableitungen von f existieren und sind stetig in ganz G .

Analog ist $f : G \rightarrow \mathbb{R}^m$ k -mal partiell differenzierbar bzw. in $C^k(G)$ wenn dies für jede Koordinate $f_i : x \rightarrow \langle f(x), e_i \rangle, i \leq m$, gilt.

Zum praktischen Rechnen benutzen wir Konzepte aus der Linearen Algebra.

DEFINITION 8.8. Sei $f : G \rightarrow \mathbb{R}$ partiell differenzierbar. Wir definieren den Gradienten von f als

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right), \text{ also } \nabla f : G \rightarrow \mathbb{R}^n.$$

Der Ausdruck ∇f wird oft "Nabla f " gelesen.

Analog werden für $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ die einzelnen $\nabla f_1, \nabla f_2, \dots, \nabla f_n$ als Zeilenvektoren aufgefasst und in einer Matrix angeordnet.

DEFINITION 8.9. Sei $f : G \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ partiell differenzierbar in $x \in G$. Dann ist

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix},$$

in Kompaktnotation also

$$(Df(x))_{ij} = \left(\frac{\partial f_i}{\partial x_j} \right) (x) \text{ wenn } i \leq m, j \leq n,$$

die Funktional- oder Jacobi-Matrix von f an der Stelle x .

Wir verstehen Ableitungen, ∇f und Df als (mehrfache) Zeilenvektoren, die auf den durch Spaltenvektoren dargestellten räumlichen Richtungen wie lineare Abbildungen agieren, die Koeffizienten dieser Abbildungen sind die partiellen Ableitungen. Die wesentliche Frage ist im Hinblick auf Beispiel 8.4.c), wann diese linearen Abbildungen die Zuwächse von f nahe x wirklich gut approximieren.

DEFINITION 8.10. Sei $f : G \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $x \in G$. Wir sagen f ist in x total differenzierbar (hat Frechet-Ableitung) wenn

$$\text{eine Matrix } A \text{ existiert, so dass } \lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}^n}} \frac{|f(x+h) - f(x) - A \cdot h|}{|h|} = 0. \quad (*)$$

Wir nennen A das totale Differential von f in x .

BEMERKUNG 8.11.

- Wenn $(*)$ gilt, kann man durch Betrachtung von $h = te_j$ für $t \rightarrow 0$ leicht finden, dass $A_{ij} = \frac{\partial f_i}{\partial x_j}(x)$ (d.h. $A = Df(x)$), also ist dann f partiell differenzierbar und A in der vorherigen Definition eindeutig. Beispiel 8.4.c) zeigt, dass die partielle keinesfalls die totale Differenzierbarkeit impliziert.
- Wenn $f(x) = A \cdot x + v$, $v \in \mathbb{R}^m$, auf G , dann ist für alle $x \in G$ die Matrix A gleich $Df(x)$ und das totale Differential von f in x .
- A ist totales Differential von f in x_0 genau dann wenn die Nullmatrix 0 das totale Differential von $x \mapsto f(x) - A \cdot x$ in x_0 ist.
- Aus c) und der Δ -Ungleichung folgt: wenn A bzw B totales Differential von f bzw g (beide $G \rightarrow \mathbb{R}^m$) in x_0 sind, dann ist $A + B$ totales Differential von $f + g$ in x_0 .

SATZ 8.12. Sei $f : G \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ partiell differenzierbar, und alle (ersten) partiellen Ableitungen in $x_0 \in G$ stetig. Dann ist f in x_0 total differenzierbar mit totalem Differential $DF(x_0)$.

BEMERKUNG 8.13. Also $f \in C^1$ impliziert "f hat totales Differential", die Umkehrung gilt bereits in \mathbb{R} nicht, wie

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

zeigt.

LEMMA 8.14. Sei A eine $m \times n$ Matrix, dann ist $x \rightarrow A \cdot x$ von \mathbb{R}^n nach \mathbb{R}^m (lipschitz) stetig und $\forall x \in \mathbb{R}^n$ gilt $|A \cdot x| \leq \left(\sum_{i=1}^n |A \cdot e_i|\right) |x|$.

KOROLLAR 8.15. (Von der Bedeutung eigentlich ein Satz) Sei f in $x_0 \in G$ total differenzierbar, dann ist f in x_0 stetig.

Insbesondere gilt $|f(x_0 + h) - f(x_0)| \leq (1 + \sum_{i=1}^n |Df \cdot e_i|) |h|$ wenn $|h| (\geq 0)$ genügend klein.

BEMERKUNG 8.16. Da Stetigkeit manchmal schwieriger überprüfbar ist, ist folgendes zwar schwächeres, aber oft einfacher zu überprüfendes Kriterium eventuell nützlich: wenn für alle i, j die partiellen Ableitungen $\frac{\partial f_i}{\partial x_j}$ auf G beschränkt sind, dann ist f in G stetig. Wenn also alle zweiten partiellen Ableitungen beschränkt sind, ist f in $C^1(G)$ und hat in jedem $x \in G$ ein totales Differential.

LEMMA 8.17. Seien $f, g : G \rightarrow \mathbb{R}$ in $C^k(G)$. Dann sind auch $f + g, f \cdot g$ in $C^k(G)$.

SATZ 8.18. Kettenregel Seien $G \subset \mathbb{R}^n, \tilde{G} \subset \mathbb{R}^m$ offen, $f : G \rightarrow \tilde{G}$ und $g : \tilde{G} \rightarrow \mathbb{R}^l$. Wenn f in $x_0 \in G$ total differenzierbar ist und g in $f(x_0) \in \tilde{G}$ total differenzierbar ist, dann ist auch $g \circ f : G \rightarrow \mathbb{R}^l$ total differenzierbar mit

$$D(g \circ f)(x_0) = Dg(f(x_0)) \cdot Df(x_0).$$

Wenn überdies $f \in C^k(G)$ und $g \in C^k(\tilde{G})$, dann $g \circ f$ in $C^k(G)$. (Dies gilt auch für $k = \infty$.)

Man kann leicht sehen, dass die Aussage auch gilt, wenn nur g in $f(x_0)$ total differenzierbar ist, und sowohl für f als auch $g \circ f$ partielle Differenzierbarkeit betrachtet wird.

Andererseits sehen wir, dass mit $g = h$ aus Beispiel ?? und $f(x) = (x_1 - x_2, x_1 + x_2)$, das aus totaler Differenzierbarkeit von f und partieller Differenzierbarkeit von g weder keinesfalls partielle Differenzierbarkeit von $g \circ f$ folgt.

KOROLLAR 8.19. $f : G \rightarrow \tilde{G}, g : \tilde{G} \rightarrow \mathbb{R}$, (also $l = 1$ im vorigen Satz) dann

$$\frac{\partial(g \circ f)}{\partial x_i}(x) = \sum_{j=1}^m \frac{\partial g}{\partial y_j}(f(x)) \frac{\partial f_j}{\partial x_i}(x) \text{ für alle } x \in G,$$

also $\nabla(g \circ f) = \nabla g \cdot Df$ als Zeilenvektoren. (Falls wir ∇ als Spaltenvektoren begreifen wollten, erhielten wir $\nabla(g \circ f) = (Df)^T \nabla g!$)

BEISPIEL 8.20. Seien $c : (0, 1) \rightarrow G$ und $g : G \rightarrow \mathbb{R}$ beide total differenzierbar und sei $t \rightarrow g(c(t))$ konstant, dann $0 = (g \circ c)'(t) = \langle \nabla g(c(t)), c'(t) \rangle$, also sind die Höhenlinien einer Funktion senkrecht zum Gradienten dieser Funktion.

DEFINITION 8.21. Die Abbildung $f : G \rightarrow \mathbb{R}^m$ hat in $x \in G$ die Richtungsableitung $D_v f(x)$ für $v \in \mathbb{R}^n$ falls für $g_{v,x}(t) = f(x + tv)$ die Funktion $g_{v,x}$ in $t = 0$ differenzierbar ist – wir setzen dann $D_v f(x) = g'_{v,x}(0)$.

SATZ 8.22. Sei $f : G \rightarrow \mathbb{R}^m$ in $x_0 \in G$ total differenzierbar. Dann $D_v f(x_0) = Df(x_0) \cdot v$ für alle $v \in \mathbb{R}^n$. Falls $m = 1$, dann ist nach der Cauchy-Schwarzschen Ungleichung $\nabla f(x_0)$ die Richtung des schnellsten Wachstums von f in x_0 .

9. Hesse Matrix, Mittelwertsatz und Taylorformel

DEFINITION 9.1. Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}$ zweimal partiell differenzierbar in x_0 in G . Dann heißt

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix},$$

in Kompaktnotation also

$$(H_f(x))_{ij} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) (x) \text{ wenn } i, j \leq n,$$

die Hesse-Matrix von f an der Stelle x .

SATZ 9.2. (H.A. Schwarz) Sei $f \in C^2(G)$, dann ist $H_f(x)$ symmetrisch für alle $x \in G$.

Genauer gilt, wenn $\frac{\partial f}{\partial x_i}$, $\frac{\partial f}{\partial x_j}$ und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ in G existieren und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ in $x_0 \in G$ stetig ist, dann existiert auch

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \text{ und ist gleich } \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0).$$

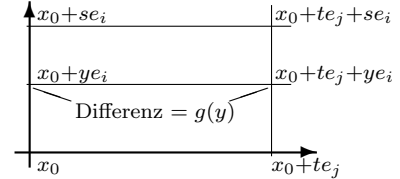
BEMERKUNG 9.3. Die etwas schwächere, aber handlichere Formulierung “wenn $f \in C^2$, dann ist $H_f(x)$ symmetrisch” ist der Satz von Clairaut. Der Satz von W.H. Young sagt: wenn $\frac{\partial f}{\partial x_i}$, $\frac{\partial f}{\partial x_j}$ (existieren und) beide in x_0 ein totales Differential haben, dann ist $\frac{\partial^2 f}{\partial x_j \partial x_i}(x_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$. Somit ist der Satz von Clairaut sowohl eine Folgerung aus dem Satz von Schwarz als auch eine Folgerung aus dem Satz von Young (und 8.12). Die beiden letzteren Sätze sind allerdings voneinander logisch unabhängig und haben verschiedene Beweise!

Beweis Satz 9.2 Mögen also $\frac{\partial f}{\partial x_i}$, $\frac{\partial f}{\partial x_j}$ und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ in G existieren und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ in $x_0 \in G$ stetig sein.

Sei zunächst $\varepsilon > 0$ gegeben, dann gibt es ein $\delta > 0$, so dass

$$h \in (-\delta, \delta)^n \Rightarrow x_0 + h \in G \text{ und } \left| \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0 + h) - \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0) \right| < \varepsilon.$$

Wir wählen nun s, t mit $0 < |s|, |t| < \delta$ und betrachten



$$\Delta(s, t) = \frac{1}{st} (f(x_0 + t e_j + s e_i) - f(x_0 + s e_i) - f(x_0 + t e_j) + f(x_0)).$$

Wir definieren $g(y) = f(x_0 + t e_j + y e_i) - f(x_0 + y e_i)$ auf $(-\delta, \delta)$, und somit erhalten wir, dass $g'(y) = \frac{\partial f}{\partial x_i}(x_0 + t e_j + y e_i) - \frac{\partial f}{\partial x_i}(x_0 + y e_i)$. Nun finden wir ξ zwischen 0 und s mit

$$\Delta(s, t) = \frac{1}{t} \left(\frac{g(s) - g(0)}{s} \right) = \frac{1}{t} g'(\xi) = \frac{1}{t} \left(\frac{\partial f}{\partial x_i}(x_0 + t e_j + \xi e_i) - \frac{\partial f}{\partial x_i}(x_0 + \xi e_i) \right),$$

da wir auf g den Mittelwertsatz aus Analysis 1 anwenden können.

Wieder wählen wir auf $(-\delta, \delta)$ eine Hilfsfunktion $h(z) = \frac{\partial f}{\partial x_i}(x_0 + z e_j + \xi e_i)$, so dass $h'(z) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) (x_0 + z e_j + \xi e_i)$ und finden nach MWS ein η zwischen 0 und t mit

$$\Delta(s, t) = \frac{1}{t} (h(t) - h(0)) = h'(\eta) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) (x_0 + \eta e_j + \xi e_i).$$

Weil $h = \eta e_j + \xi e_i \in (-\delta, \delta)^n$ haben wir also nach Voraussetzung

$$\left| \Delta(s, t) - \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0) \right| < \varepsilon, \text{ wenn } 0 < |s|, |t| < \delta.$$

Nun senden wir t nach 0 und erhalten wegen der Existenz von $\partial f / \partial x_j$ in G , dass

$$\frac{1}{s} \left(\frac{\partial f}{\partial x_j}(x_0 + s e_i) - \frac{\partial f}{\partial x_j}(x_0) \right) = \lim_{t \rightarrow 0} \Delta(s, t),$$

und also

$$\left| \frac{1}{s} \left(\frac{\partial f}{\partial x_j}(x_0 + s e_i) - \frac{\partial f}{\partial x_j}(x_0) \right) - \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0) \right| \leq \varepsilon, \text{ wenn } 0 < |s| < \delta.$$

Da $\varepsilon > 0$ beliebig war, heißt dies

$$\lim_{s \rightarrow 0} \frac{1}{s} \left(\frac{\partial f}{\partial x_j}(x_0 + s e_i) - \frac{\partial f}{\partial x_j}(x_0) \right) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0),$$

bzw. $\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) (x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0)$, was zu zeigen war.

KOROLLAR 9.4. Wenn $f \in C^k(G)$, dann kommt es bei den partiellen Ableitungen der Ordnung $l \leq k$ nicht auf die Reihenfolge an.

Zum Beispiel

$$\begin{aligned} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} &= \frac{\partial}{\partial x_i} \left(\frac{\partial^2 f}{\partial x_j \partial x_k} \right) = \frac{\partial}{\partial x_i} \left(\frac{\partial^2 f}{\partial x_k \partial x_j} \right) = \frac{\partial^2}{\partial x_i \partial x_k} \left(\frac{\partial f}{\partial x_j} \right) \\ &= \frac{\partial^2}{\partial x_k \partial x_i} \left(\frac{\partial f}{\partial x_j} \right) = \frac{\partial}{\partial x_k} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) = \frac{\partial}{\partial x_k} \left(\frac{\partial^2 f}{\partial x_j \partial x_i} \right) = \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}. \end{aligned}$$

Aus Analysis I erhalten wir folgenden eindimensionalen Mittelwertsatz (MWS, vergleiche Ana I 4.15): wenn $g : (a, b) \rightarrow \mathbb{R}$ in (a, b) differenzierbar ist und $x < y$ in (a, b) , dann $\exists \xi \in (x, y) : g(y) - g(x) = g'(\xi)(y - x)$. Eine direkte Verallgemeinerung in höhere Dimensionen ist nicht möglich, siehe folgendes

BEISPIEL 9.5. Sei $c : \mathbb{R} \rightarrow \mathbb{R}^2$ durch $c(t) = (\cos(t), \sin(t))$ definiert. Dann $c \in C^\infty(\mathbb{R})$ und $|c'(\xi)| = 1 > 0$ für jedes $\xi \in \mathbb{R}$. Trotzdem ist $c(2\pi) - c(0) = 0$.

Wir wollen nun zeigen, dass die (totale) Ableitung trotzdem den Zuwachs einer Funktion auch in mehreren Dimensionen kontrolliert, bzw. sogar approximiert.

SATZ 9.6. Mittelwertsatz Sei $G \subset \mathbb{R}^n$ offen, $f : G \rightarrow \mathbb{R}^m$ in C^1 . Wenn $x \in G$ und $\xi \in \mathbb{R}^n$, so dass die Strecke $[x, x + \xi] = \{x + t\xi; t \in [0, 1]\} \subset G$, dann gibt es ein $M \in \mathbb{R}$, mit

$$\forall t \in [0, 1] \forall v \in \mathbb{R}^n : |Df(x + t\xi) \cdot v| \leq M|v|.$$

Jedes solches M erfüllt dann auch

$$|f(x + \xi) - f(x)| \leq M|\xi|.$$

KOROLLAR 9.7. Konstantheitskriterium Sei $G \subset \mathbb{R}^n$ offen, $f : G \rightarrow \mathbb{R}^m$ mit $Df(y) = 0$ für jedes $y \in G$. Wenn $x \in G$ und $\xi \in \mathbb{R}^n$, so dass $[x, x + \xi] = \{x + t\xi; t \in [0, 1]\} \subset G$ dann $f(x) = f(x + \xi)$.

BEMERKUNG 9.8. Die starke Voraussetzung $[x, x + \xi] \subset G$ schließt eigentlich unnötig viele Menge G von der Anwendung des Kriteriums aus. Da das Konstantheitskriterium andererseits aber auf allen Kugeln gilt und wiederholt angewendet werden kann, sehen wir dass $Df \equiv 0$ in G die Konstantheit von f in G impliziert falls es für alle $x, y \in G$ ein $c : [0, 1] \rightarrow G$ stetig gibt, so dass $c(0) = x$ und $c(1) = y$. (Solch G heißt wegzusammenhängend.)

BEMERKUNG 9.9. Man kann auch folgende schwächere (da eigentlich unscharfe) Variante des Mittelwertsatzes zeigen (DIY, analog zum Beweis von 8.12): Wenn $f : B(x_0, r) \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ in $B(x_0, r)$ differenzierbar und

$$\forall x \in B(x_0, r) \forall i \leq n \forall j \leq m : \left| \frac{\partial f_j}{\partial x_i}(x) \right| \leq c,$$

dann

$$\forall x, y \in B(x_0, r) : |f(x) - f(y)| \leq c\sqrt{m}\|x - y\|_1 \leq c\sqrt{mn}|x - y|.$$

KOROLLAR 9.10. Approximationsform des MWS Sei $G \subset \mathbb{R}^n$ offen, $f : G \rightarrow \mathbb{R}^m$ in C^1 . Wenn $x \in G$ und $\xi \in \mathbb{R}^n$, so dass $[x, x + \xi] = \{x + t\xi; t \in [0, 1]\} \subset G$, und wenn für eine $m \times n$ Matrix A und ein $\varepsilon \geq 0$ gilt

$$\forall t \in [0, 1] \forall v \in \mathbb{R}^n : |Df(x + t\xi) \cdot v - A \cdot v| \leq \varepsilon|v|,$$

dann

$$|f(x + \xi) - f(x) - A \cdot \xi| \leq \varepsilon|\xi|.$$

Wir wollen nun höhere Ableitungen und die Taylorsche Formel studieren, und brauchen dazu eine effektive Schreibweise, die nur die Anzahl aber nicht die irrelevante Reihenfolge der Ableitungen notiert.

DEFINITION 9.11. Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}^m$.

i) Für $\alpha = (\alpha_1, \dots, \alpha_n)$ mit $\alpha_i \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ definieren wir

$$|\alpha| = \sum_{i=1}^n \alpha_i \text{ Betrag von } \alpha, \quad \alpha! = \prod_{i=1}^n \alpha_i!,$$

und nennen ein in diesem Zusammenhang verstandenes α einen Multiindex.

ii) Wenn α ein Multiindex ist und $f \in C^{|\alpha|}(G)$, dann setzen wir für $x \in G$

$$D^\alpha f(x) = D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n} f(x) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}(x).$$

Hierbei $D_i^0 f = f$. Ausserdem definieren wir für $x \in \mathbb{R}^n$

$$x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}.$$

LEMMA 9.12. Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}^m$. Falls die "offene" Strecke $(x - \xi, x + \xi)$ in G enthalten ist, betrachten wir die Funktion $g(t) = f(x + t\xi)$ auf $(-1, 1)$. Wenn $f \in C^k(G)$ mit $k \geq 1$, dann gilt

$$\forall t \in (-1, 1) : g^{(k)}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(x + t\xi) \xi^\alpha; \quad (9.1)$$

hierbei ist $|\alpha|!/\alpha!$ also die Anzahl der Möglichkeiten, D^α als Abfolge einzelner Ableitungen zu schreiben, d.h. die Anzahl der verschiedenen $(j_1, j_2, \dots, j_{|\alpha|}) \in \{1, \dots, n\}^{|\alpha|}$ mit $\sum_{l=1}^n e_{j_l} = \alpha$.

Genauso wie (??), aber mit einem etwas einfacheren Induktionsschritt erhält man unter den Voraussetzungen dieses Lemmas auch

$$\forall t \in (-1, 1) : g^{(k)}(t) = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_k=1}^n D_{i_1} D_{i_2} \dots D_{i_k} f(x + t\xi) \xi_{i_1} \xi_{i_2} \dots \xi_{i_k}, \quad (9.2)$$

hier erspart man sich die nach Korollar 9.4 mögliche Zusammenfassung identischer Ableitungen wie $D_1 D_2 = D_2 D_1 = D^{(1,1)}$.

Lohnt sich die Arbeit mit der komplizierten Form (??)? In (??) haben wir klarerweise n^k Terme. Für die Anzahl $c(n, k)$ der Multiindefizes $(\alpha_1, \dots, \alpha_n)$ mit $|\alpha| = k$ leitet man leicht die Rekursionsformel $c(n+1, k+1) = c(n, k+1) + c(n+1, k)$ her und aus dieser wiederum mit Induktion die Formel

$$c(n, k) = \binom{n+k-1}{k}.$$

Für $n = k = 4$ erhalten wir also 256 Terme in (??) aber nur $c(4, 4) = 35$ Terme in (??).

Analog zur und basierend auf der 1-dimensionalen Integraldarstellung der Taylorentwicklung erhalten wir das entsprechende Resultat in \mathbb{R}^n . Nach 5.26 gilt für jede Funktion $g \in C^{k+1}([0, 1])$

$$g(1) = \sum_{l=0}^k \frac{1}{l!} g^{(l)}(0) + \frac{1}{k!} \int_0^1 (1-t)^k g^{(k+1)}(t) dt.$$

Dies liefert, kombiniert mit (??), den folgenden

SATZ 9.13. *Taylorische Formel* Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}^m$ in $C^{k+1}(G)$. Wenn $x + t\xi \in G$ für alle $t \in [0, 1]$, dann

$$f(x + \xi) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x) \xi^\alpha + (k+1) \sum_{|\alpha|=k+1} \int_0^1 (1-t)^k \frac{D^\alpha f(x + t\xi)}{\alpha!} \xi^\alpha dt.$$

DEFINITION 9.14. Sei $f : G \rightarrow \mathbb{R}^m$ in $x \in G$ wenigstens k -mal partiell differenzierbar. Wir definieren

$$T_{f,x}^k(\xi) = \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x) \xi^\alpha, \text{ Taylorentwicklung der Ordnung } k,$$

$$R_{f,x}^k(\xi) = f(x + \xi) - T_{f,x}^k(\xi) \text{ Taylor-Restglied der Ordnung } k.$$

LEMMA 9.15. Sei $f : G \rightarrow \mathbb{R}$ in $x \in G$ zweimal partiell differenzierbar, dann

$$T_{f,x}^2(\xi) = f(x) + \langle (\nabla f(x))^T, \xi \rangle + \frac{1}{2} \langle \xi, H_f(x) \cdot \xi \rangle \quad (9.3)$$

wobei $H_f(x) = D(y \mapsto (\nabla f(y))^T)(x)$ die Hessematrix aus 9.1 ist.

SATZ 9.16. Wir betrachten $f : G = B_R(x_0) \rightarrow \mathbb{R}$.

a) Wenn $f \in C^3(G)$ und $0 < r < R$ dann gibt es ein M , so dass

$$\xi \in \mathbb{R}^n, |\xi| \leq r \Rightarrow |R_{f,x}^k(\xi)| \leq M|\xi|^3.$$

Überdies ist $T_{f,x}^2$ das einzige quadratische Polynom P (d.h. $\exists c_\alpha \in \mathbb{R} \forall \xi : P(\xi) = \sum_{|\alpha| \leq 2} c_\alpha \xi^\alpha$), mit

$$|f(x + \xi) - P(\xi)| = o(|\xi|^2), \text{ das heißt } \lim_{\xi \rightarrow 0} \frac{1}{|\xi|^2} |f(x + \xi) - P(\xi)| = 0. \quad (9.4)$$

- b) Wenn f in G zumindest einmal partiell differenzierbar ist, und $x \mapsto (\nabla f(x))^T$ überdies in x_0 total differenzierbar ist (z.B. wenn $x \mapsto H_f(x)$ stetig in x_0 , vergleiche mit Satz 8.12), dann gilt immer noch (entsprechend zu (??)), dass

$$\lim_{\xi \rightarrow 0} \frac{1}{|\xi|^2} |R_{f,x_0}^2(\xi)| = 0.$$

In diesem Satz können wir die euklidische Norm durch jede andere Norm ersetzen, auch die betrachtete Taylor-Ordnung (hier $k = 2$) kann man erhöhen.

Wir könnten auch den Fall $f : B_R(x) \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $m \geq 1$ ohne wesentliche Änderungen behandeln, allerdings muss das Skalarprodukt $\langle (\nabla f(x))^T, \xi \rangle$ durch das übliche Matrixprodukt $Df(x) \cdot \xi$ ersetzt werden und der quadratische Term $\langle \xi, H_f(x)\xi \rangle$ wird am besten zeilenweise als $(H_{f_i}(x))$ für $i = 1, \dots, m$ verstanden, da die volle zweite Ableitung von f aus mn^2 partiellen Ableitungen gebildet ist, und somit nicht als einzelne Matrix dargestellt werden kann.

10. Extremwerte und kritische Stellen

DEFINITION 10.1. Sei (X, ρ) ein metrischer Raum, $f : X \rightarrow \mathbb{R}$ und $x \in X$. Dann ist

- x lokale Maximalstelle von f (wir sagen auch f hat lokales Maximum in x) wenn $\exists \varepsilon > 0 \forall y \in B_\varepsilon(x) : f(y) \leq f(x)$.
- x strenge lokale Maximalstelle von f (wir sagen auch f hat strenges lokales Maximum in x) wenn $\exists \varepsilon > 0 \forall y \in B_\varepsilon(x) \setminus \{x\} : f(y) < f(x)$.

Analog definieren wir lokale und strenge lokale Minima und Minimalstellen sowie Extrema und Extremalstellen.

SATZ 10.2. Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}$. Wenn f in $x \in G$ ein lokales Extremum hat, und f in x partiell differenzierbar ist, dann $\nabla f(x) = 0$. Wenn überdies eine Richtungsableitung $D_v f(x)$ existiert, dann auch $D_v f(x) = 0$.

DEFINITION 10.3. Sei $G \subset \mathbb{R}^n$ offen und $f : G \rightarrow \mathbb{R}$ partiell differenzierbar in G . Dann hat f in $x \in G$ eine kritische Stelle wenn $\nabla f(x) = 0$.

Gemäß Satz 9.16 gilt für $f \in C^2$ in einem kritischen Punkt $f(x + \xi) - f(x) = \frac{1}{2} \langle \xi, H_f(x)\xi \rangle + \text{err}(\xi)$ wobei $|\text{err}(\xi)| \leq o(|\xi|^2)$ viel kleiner als $\langle \xi, H_f(x)\xi \rangle$ sein sollte, wenn ξ nahe Null ist. Deshalb wird die Frage, ob f in einer kritische Stelle ein lokales Minimum, Maximum oder überhaupt ein Extremum hat im Wesentlichen vom Vorzeichenverhalten der quadratische Form

$$Q_{f,x}(\xi) = Q(\xi) = \langle \xi, H_f(x)\xi \rangle = \xi^T H_f(x)\xi$$

bestimmt. Aber bereits die naheliegendste Frage, wann $Q(\xi) > 0 \forall \xi \neq 0$ ist nicht ganz einfach zu beantworten. Uns hilft hierbei zumindest, dass $H_f(x)$ symmetrisch ist.

BEISPIEL 10.4. Sei $n = 2$ und $H_f(x) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Dann gilt $\forall \xi \in \mathbb{R}^2 \setminus \{(0,0)\} : Q(\xi) > 0$ gdw. $a, c > 0$ und $\det(H_f(x)) = ac - b^2 > 0$ und $\forall \xi \in \mathbb{R}^2 : Q(\xi) \geq 0$ gdw. $a, c \geq 0$ und

$ac \geq b^2$. In allen anderen Fällen existiert $\xi \in \mathbb{R}^2 : Q(\xi) < 0$, aber das bedeutet noch nicht $\forall \xi \neq 0 : Q(\xi) < 0$ bzw. dass f ein Maximum in x hat.

DEFINITION 10.5. Sei A eine symmetrische reelle $n \times n$ Matrix ($A_{ij} = A_{ji}$, also $A^T = A$.)

- A (und die dazugehörige Form $Q_A(\xi) = \langle \xi, A\xi \rangle$) heißt positiv definit wenn $\forall \xi \in \mathbb{R}^n \setminus \{0\} : \langle \xi, A\xi \rangle > 0$ und heißt negativ definit wenn $\forall \xi \in \mathbb{R}^n \setminus \{0\} : \langle \xi, A\xi \rangle < 0$.
- A (und die dazugehörige Form Q_A) heißt positiv(negativ) semidefinit wenn $\forall \xi \in \mathbb{R}^n : \langle \xi, A\xi \rangle \geq 0$ ($\langle \xi, A\xi \rangle \leq 0$).
- A (und die dazugehörige Form Q_A) ist indefinit wenn es $\xi_-, \xi_+ \in \mathbb{R}^n$ mit $\langle \xi_-, A\xi_- \rangle < 0 < \langle \xi_+, A\xi_+ \rangle$ gibt.

Um zu entscheiden, welche der Bedingungen für eine gegeben Matrix gilt, brauchen wir ein paar Begriffe und Fakten aus der linearen Algebra.

- SATZ 10.6. a) Seien A, B $n \times n$ Matrizen, dann gilt $\det(AB) = \det(A) \det(B)$.
 b) Wenn A eine reelle und symmetrische $n \times n$ -Matrix ist, dann existieren $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ Eigenwerte von A und v_1, \dots, v_n Orthogonalbasis aus dazugehörigen Eigenvektoren von A . Das heißt für alle $i, j \leq n$ gilt

$$\langle v_i, v_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (\text{Orthogonalbasis}), Av_i = \lambda_i v_i. (\text{Eigenwert und -vektor.})$$

Die Beweise werden wir hier nicht geben, sie können aber in jedem Lehrbuch der Linearen Algebra gefunden werden. Teil b) lässt sich analytisch beweisen mit einem einfachen Induktionsargument (allerdings muss man wissen, dass eine symmetrische Matrix, die einen linearen Unterraum des \mathbb{R}^n in sich selbst abbildet, auf diesem Unterraum wieder als symmetrische Matrix auf einem \mathbb{R}^m , $m < n$, aufgefasst werden kann). Teil a) ist sehr einleuchtend, wenn man die Determinante von A als Faktor der Volumenänderung bei Multiplikation mit A versteht, ein exakter Beweis (z.B. dass dieses Verständnis korrekt ist) verlangt etwas mehr Vorbereitungen.

KOROLLAR 10.7. Sei A eine reelle und symmetrische $n \times n$ -Matrix und sei v_1, \dots, v_n eine Orthogonalbasis aus Eigenvektoren. Wenn wir die Matrix $P_{ij} = (v_j)_i \forall i, j \leq n$ definieren, dann $Pe_j = v_j$, $P^T P = Id$ und also

$$P^T A P = P^{-1} A P = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \Lambda,$$

und $\det(A) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n$ Produkt der Eigenwerte. (Die zweite Gleichheit in der obigen Zeile gilt wann immer v_1, \dots, v_n eine Basis aus Eigenvektoren ist, man sagt dann aus klar ersichtlichen Gründen, dass A diagonalisierbar ist – Orthogonalität wird hier nicht benötigt.)

LEMMA 10.8. Sei A eine reelle und symmetrische $n \times n$ -Matrix mit den Eigenwerten $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Dann ist

- a) positiv definit $\Leftrightarrow \forall i : \lambda_i > 0$
- a)' negativ definit $\Leftrightarrow \forall i : \lambda_i < 0$
- b) positiv (bzw. negativ) semidefinit $\Leftrightarrow \forall i : \lambda_i \geq 0$ (bzw. $\lambda_i \leq 0$).
- b)' indefinit $\Leftrightarrow \exists i, j \lambda_i < 0 < \lambda_j$.

Dies gilt weil $\forall x \in \mathbb{R}^n$ und $y = P^T x$

$$\langle x, Ax \rangle = x^T Ax = y^T (P^T AP)y = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2.$$

Allerdings ist das Finden der Eigenwerte nicht ganz einfach. Wir können zwar nutzen, dass jeder Eigenwert eine Nullstelle des charakteristischen Polynoms $c_A(\lambda) = \det(\lambda Id - A)$ ist, und es somit immer höchstens n Nullstellen dieses Polynoms vom Grade gibt. Über deren Positivität zu entscheiden ist jedoch schwieriger. Neben der sogenannten "Sturmschen Kette" (und digitalen Rechengehilfen), benutzt man hier oft das

SATZ 10.9. Hurwitz -oder Hauptblock Kriterium Sei A eine reelle und symmetrische $n \times n$ -Matrix. Für $k \leq n$ definieren wir $A_k = (A_{ij})_{i,j \leq k}$, den "ersten $k \times k$ -Hauptblock" und $\Delta_k = \det(A_k)$. Dann gilt

- a) A ist positiv definit gdw. $\forall k \leq n : \Delta_k > 0$,
- a)' A ist negativ definit gdw. $\forall k \leq n : (-1)^k \Delta_k > 0$, und
- b) wenn ein k mit $2k \leq n$ und $\Delta_{2k} < 0$ existiert, dann ist A indefinit.

Die Matrix $\text{diag}(0, -1)$ zeigt, dass $\forall k : \Delta_k \geq 0$ keinesfalls positive Semidefinitheit impliziert, und $\text{diag}(-1, -1, 1, 1)$ zeigt, dass die Bedingung in b) nur hinreichend, aber nicht notwendig ist.

SATZ 10.10. Sei $G \subset \mathbb{R}^n$ offen, $f : G \rightarrow \mathbb{R}$ in $C^2(G)$ und $x \in G$ eine kritische Stelle von f .

- a) Wenn $H_f(x)$ positiv definit ist, dann hat f ein strenges lokales Minimum in x .
- a)' Wenn $H_f(x)$ negativ definit ist, dann hat f ein strenges lokales Maximum in x .
- b) Wenn f in x ein lokales Minimum (bzw. Maximum) hat, dann ist $H_f(x)$ positiv (bzw. negativ) semidefinit.

BEMERKUNG 10.11. i) Im vorigen Satz reicht es natürlich wiederum, dass $y \mapsto (\nabla f(y))^T$ in x total differenzierbar und x eine kritische Stelle von f ist.
 ii) Es ist geometrisch intuitiv (gerade in 2 Dimensionen), einen kritischen Punkt, der keine Extremstelle ist, einen Sattelpunkt - die Motivation mag sowohl von der Kavalerie oder aus der Topographie stammen.

BEISPIEL 10.12. a) $f(x, y) = y^2 - x^2$, dann $\nabla f = 0 \Leftrightarrow (x, y) = (0, 0)$. In diesem Punkt ist

$$H_f(0, 0) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

indefinit, dies ist klarerweise (bis auf Rotation des Koordinatensystems und Änderung der Länge auf der x oder der y -Achse) der typische Sattelpunkt in 2 Dimensionen.

- b) $g(x, y) = x^3 + y^3 - 3xy$ auf \mathbb{R}^2 , dann $\nabla g(x, y) = (3(x^2 - y), 3(y^2 - x))$, somit $\nabla g(x, y) = 0$ gdw. $x^2 = y$ & $y^2 = x$, also $x^4 = x$, d.h. $x(x - 1)(x^2 + x + 1) = 0$ und somit nur $x = y = 0$ und $x = y = 1$ kritische Stellen. Da

$$H_f(x, y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix},$$

haben wir in $(0, 0)$ einen Sattelpunkt, und in $(1, 1)$ ein strenges lokales Minimum.

PROPOSITION 10.13. Fundamentalsatz der Algebra *Jedes auf nicht ganz \mathbb{C} konstante Polynom $p : \mathbb{C} \rightarrow \mathbb{C}$ hat mindestens eine komplexe Nullstelle.*

11. Die Weg- oder Bogenlänge

Wir wollen hier die Länge einer Kurve (auch Weg oder “Bogen” genannt) definieren.

DEFINITION 11.1. *Ein stetiger Weg (eine stetige Kurve) ist eine stetige Abbildung $c : [a, b] \rightarrow \mathbb{R}^n$. Sie hat Startpunkt $c(a)$ und Endpunkt $c(b)$ und ist “geschlossen” wenn $c(a) = c(b)$*

Man kann zu einem gewissen Grad den Weg c mit seinem Bild $c([a, b])$ identifizieren, sollte aber nie vergessen, dass c Werte mehrmals annehmen kann, als sein Bild mehrmals durchlaufen kann.

Wir wiederholen aus Kapitel 5 (Riemann Integral) den Begriff der Partion. Seien $a < b$ reell. Eine Partition des Intervalls $[a, b]$ ist ein endliches $P = \{t_0, t_1, \dots, t_k\} \subset [a, b]$ mit den Teilungspunkten $a = t_0 < t_1 < \dots < t_k = b$.

Feinheit der Partition $P = \{t_0, \dots, t_k\}$ ist $d(P) = \max_{l=0, \dots, k-1} |t_{l+1} - t_l|$. Wir sagen: die Partion Q von $[a, b]$ ist feiner als Partion P von $[a, b]$, wenn $P \subset Q$. (“ P -Intervalle” $[t_0, t_1], [t_1, t_2], \dots, [t_{k-1}, t_k]$ werden von Punkten in Q nochmals partitioniert).
Bezeichnung $\mathfrak{P}([a, b]) = \{P : P \text{ Partition von } [a, b]\}$.

DEFINITION 11.2. *Der Weg $c : [a, b] \rightarrow \mathbb{R}^n$ ist C^1 wenn es ein $\varepsilon > 0$ und $\tilde{c} : (a - \varepsilon, b + \varepsilon) \rightarrow \mathbb{R}^n$ mit $\tilde{c} \in C^1(a - \varepsilon, b + \varepsilon)$ und $\forall t \in [a, b] : \tilde{c}(t) = c(t)$ gibt.*

Ausserdem sagen wir $c : [a, b] \rightarrow \mathbb{R}^n$ ist stückweise- C^1 , wenn es eine Partition $a = t_0 < t_1 < \dots < t_k = b$ gibt, so dass $\forall l < k$ der Weg $c|_{[t_l, t_{l+1}]}$ C^1 ist.

DEFINITION 11.3. *Gegeben einen stetigen Weg $c : [a, b] \rightarrow \mathbb{R}^n$ und eine Partition $P = \{t_0, t_1, \dots, t_k\} \in \mathfrak{P}([a, b])$, setzen wir*

$$L(c, P) = \sum_{l=0}^{k-1} |c(t_l) - c(t_{l+1})|, \text{ und}$$

$$L(c) = \sup_{P \in \mathfrak{P}([a, b])} L(c, P) \text{ die Bogenlänge von } c,$$

wobei das Supremum einer von oben unbeschränkten Menge gleich $+\infty$ ist.

Wir nennen c rektifizierbar, falls $L(c) < \infty$.

DEFINITION 11.4. *Wenn $c : [a, b] \rightarrow \mathbb{R}^n$ ein stetiger und rektifizierbarer Weg ist, dann definieren wir $s(t) = s_c(t) = L(c|_{[a, t]})$.*

LEMMA 11.5. *Sei $c : [a, b] \rightarrow \mathbb{R}^n$ ein stetiger Weg und $P \in \mathfrak{P}([a, b])$, so dass $|x - y| \leq d(P) \Rightarrow |c(x) - c(y)| \leq \varepsilon$. Dann gilt*

$$\forall x \in [a, b] : L(c, P) \leq L(c, P \cup \{x\}) \leq L(c, P) + 2\varepsilon.$$

Also

$$\forall Q \in \mathfrak{P}([a, b]) : P \subset Q \Rightarrow L(c, P) \leq L(c, Q) \leq L(c, P) + 2\varepsilon \text{card}(P \setminus Q).$$

Der Beweis folgt den Ideen des Beweises von Lemma 5.3.

SATZ 11.6. Sei $c : [a, b] \rightarrow \mathbb{R}^n$ ein stetiger Weg und $P_l \in \mathfrak{P}([a, b])$ eine Folge von Partitionen mit $d(P_l) \rightarrow 0$ für $l \rightarrow \infty$. Dann $L(c, P_l) \rightarrow L(c)$ wenn $l \rightarrow \infty$.

Der Beweis nutzt wiederum Ideen aus Kapitel 5, genauer des Beweises von Proposition 5.13

KOROLLAR 11.7. Wenn $c : [a, b] \rightarrow \mathbb{R}^n$ ein stetiger Weg ist, dann

$$\forall x \in [a, b] : L(c) = L(c|_{[a, x]}) + L(c|_{[x, b]}).$$

Insbesondere $s_c(t_2) = s_c(t_1) + L(c|_{[t_1, t_2]})$, wobei wir in beiden Gleichungen $+\infty$ zulassen (mit Rechenregeln $(+\infty) + (+\infty) = (+\infty) + y = +\infty$ für $y \in \mathbb{R}$).

Somit ist ein stetiger Weg $c : [a, b] \rightarrow \mathbb{R}^n$ rektifizierbar gdw. für jede Partition $P \in \mathfrak{P}([a, b])$ und jedes P -Intervall $[t_i, t_{i+1}]$ gilt, dass $c|_{[t_i, t_{i+1}]}$ rektifizierbar ist.

SATZ 11.8. Wenn der Weg $c : [a, b] \rightarrow \mathbb{R}^n$ stückweise C^1 -ist, dann gilt

$$L(c) = \int_a^b |c'(t)| dt < \infty$$

und somit $s'_c(t) = |c'(t)|$ wenn c' in t stetig ist.

BEISPIEL 11.9.

- a) Seien $r, h > 0$ und $c(t) = (r \cos(t), r \sin(t), ht) \in \mathbb{R}^3$ eine Schraubenlinie ("Helix"). Dann ist $|c'(t)| = \sqrt{r^2 + h^2}$ für jedes $t \in \mathbb{R}$ und also zum Beispiel

$$L(c|_{[a, b]}) = (b - a)\sqrt{r^2 + h^2} \text{ für alle } a < b.$$

- b) Wenn $f : [a, b] \rightarrow \mathbb{R}$ eine (stückweise) C^1 -Funktion ist, dann bildet $c : t \mapsto (t, f(t))$ das Intervall $[a, b]$ auf den Graphen von f ab. Dieser Graph hat also Länge

$$\int_a^b \sqrt{1 + (f'(t))^2} dt.$$

12. Lagrange Multiplikatoren

SATZ 12.1. Sei $U \subset \mathbb{R}^n$ offen und seien die Funktionen $f : U \rightarrow \mathbb{R}$ sowie $g_i : U \rightarrow \mathbb{R}$, $i = 1, \dots, m$, alle C^1 .

Wir betrachten die Menge

$$M = \{x \in U : g_1(x) = \dots = g_m(x) = 0\},$$

und möge $f|_M$ in $x_0 \in M$ eine lokale Extremstelle haben, dass heißt, es gibt ein $c \in \{-1, 1\}$ und $\varepsilon > 0$ mit $x \in M \cap B(x_0, \varepsilon) \Rightarrow (f(x) - f(x_0)) \cdot c \geq 0$. Falls

die Vektoren $(\nabla g_1(x_0))^T, \dots, (\nabla g_m(x_0))^T$ linear unabhängig sind,

dann gilt

$$(\nabla f(x_0))^T \in \text{span}\{(\nabla g_1(x_0))^T, \dots, (\nabla g_m(x_0))^T\},$$

dass heißt, es gibt $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, so dass

$$(\nabla f(x_0))^T = \sum_{i=1}^m \lambda_i (\nabla g_i(x_0))^T \quad \text{in } \mathbb{R}^n.$$

Die Zahlen $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ heißen Lagrange Multiplikatoren für dieses Extremproblem $f \rightarrow \min / \max$ mit Nebenbedingungen $g_i(x) = 0$, $i = 1, \dots, m$.

BEMERKUNG Die Schlussfolgerung dieses Satzes läßt sich auch "symmetrischer" und ohne die eingerahmte Voraussetzung formulieren, da die spezielle Bedeutung von f dabei etwas verschwindet, ist die folgende Formulierung aber weniger häufig in der Literatur.

Wenn die Funktion f in $x_0 \in \mathbb{R}^n$ eine lokale Extremstelle bezüglich der Nebenbedingungen $g_1 = \dots = g_m = 0$ hat (wobei $f, g_j : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ in C^1), dann sind die Zeilenvektoren

$$\nabla f(x_0), \nabla g_1(x_0), \dots, \nabla g_m(x_0) \text{ linear abhängig,}$$

d.h. eine geeignete "Lagrange Funktion" $\lambda_0 f + \sum_{j=1}^m \lambda_j g_j$, $\sum_0^m \lambda_j^2 > 0$ hat in x_0 eine kritische Stelle (bezüglich \mathbb{R}^n).

BEISPIEL 12.2. Sei $U = \mathbb{R}^4$ und $f(x) = (x_1 - x_2)^2 + (x_3 - x_4)^2$, sowie $g_1(x) = x_1^2 + x_3^2 - 2$, $g_2(x) = x_4 - x_2 + 4$. Die Menge

$$M = \{x \in \mathbb{R}^4 : x_1^2 + x_3^2 = 2 \text{ und } x_4 = x_2 - 4\}$$

ist zwar abgeschlossen, aber nicht beschränkt, insbesondere $(1, t, 1, t - 4) \in M$ für jedes $t \in \mathbb{R}$ - woraus man leicht sieht, dass $f(M) \subset \mathbb{R}$ keine obere Schranke hat.

DIY: Gleichwohl kann man leicht zeigen, dass $x^k \in M$ und $|x^k| \rightarrow \infty$ auch $f(x^k) \rightarrow \infty$ für $k \rightarrow \infty$ erzwingt, deshalb existiert wegen der üblichen Kompaktheitsargumente im \mathbb{R}^n auch ein $x \in M$ mit $f(x) = \min f(M)$.

Da $(\nabla g_1(x)) = (2x_1, 0, 2x_3, 0)$ und $(\nabla g_2(x)) = (0, -1, 0, 1)$ sind diese beiden Vektoren linear unabhängig wenn $(2x_1, 0, 2x_3, 0) \neq 0$, dies ist aber überall in M der Fall, da dort $x_1^2 + x_3^2 = 2 > 0$. Laut Satz 12.1 gibt es also $\lambda_1, \lambda_2 \in \mathbb{R}$, so dass

$$\begin{pmatrix} 2(x_1 - x_2) \\ 2(x_2 - x_1) \\ 2(x_3 - x_4) \\ 2(x_4 - x_3) \end{pmatrix} = \lambda_1 \begin{pmatrix} 2x_1 \\ 0 \\ 2x_3 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Die in der Definition von g_1, g_2 aber auch f vorhandene Symmetrie bzw. Zuordnung $x_1 \leftrightarrow x_3$, $x_2 \leftrightarrow x_4$ gilt es beim Lösen des Gleichungssystem zu erhalten und zu nutzen. Wenn wir 2. und 4. Zeile vergleichen, erhalten wir $-2(x_2 - x_1) = \lambda_2 = 2(x_4 - x_3)$, d.h. $(x_1 - x_2) = (x_4 - x_3)$. Aus $x_3 \times 1$. versus $x_1 \times 3$. Zeile folgt

$$2(x_1 - x_2)x_3 = \lambda_1 2x_1 x_3 = 2(x_3 - x_4)x_1 = -2(x_1 - x_2)x_1 \Rightarrow (x_3 + x_1)(x_1 - x_2) = 0.$$

Falls $x_1 = x_2$ dann auch $x_3 = x_4$ und aus $x_3 - x_1 + 4 = 0$ folgt $|x_i| \geq 2$ für ein $i \in \{1, 3\}$ und also $x_1^2 + x_3^2 \geq 4$, Widerspruch. Somit $x_3 = -x_1$, was wegen $(x_1 - x_2) = (x_4 - x_3)$ auch $x_2 = -x_4 = 2$ erzwingt. Dann $|x_1| = |x_3| = 1$ und wir müssen das Minimum von $(x_1 - 2)^2 + (-x_1 + 2)^2 = 2(x_1 - 2)^2$ über $x_1 \in \{-1, 1\}$ finden. Dies liefert $x = (1, 2, -1, -2)$ und $\min f(M) = 2$.

BEISPIEL 12.3. /Gegenbeispiel Sei $U = \mathbb{R}^3$, und $f(x) = x_2$, $g_1(x) = x_1^6 - x_3$, $g_2(x) = x_2^3 - x_3$. Dann scheint es keine Extremalstelle von f unter der Nebenbedingung $x \in M$ zu geben, da die entsprechende Vektoridentität aus 12.1 keine Lösung (mit $x \in M$) hat. Gleichwohl ist $0 = f(0, 0, 0) = \min f(M)!!$

Der Beweis des Satzes 12.1 ist eine einfache Konsequenz aus folgender Variante des fundamentalen Satzes über die Umkehrfunktion.

PROPOSITION 12.4. Sei $U \subset \mathbb{R}^n$ offen und sei die Funktion $F : U \rightarrow \mathbb{R}^m$ in $C^1(U)$. Wenn $x \in U$ und das totale Differential $DF(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ surjektiv ist, dann ist für jedes $\varepsilon > 0$ das Bild $F(U \cap B(x, \varepsilon))$ eine Umgebung von $F(x)$ im \mathbb{R}^m .

13. Kontraktionen und Fixpunkte, Newton Iteration, Umkehr- und implizite Funktionen

DEFINITION 13.1. Seien $(X, \varrho), (Y, \sigma)$ metrische Räume und eine Abbildung $f : X \rightarrow Y$ gegeben. Dann ist f Lipschitz stetig wenn es ein $L \in [0, \infty)$ gibt, so dass

$$\forall x, y \in X : \sigma(f(x), f(y)) \leq L\varrho(x, y).$$

Die Lipschitzkonstante $Lip(f)$ ist dann das Infimum aller solcher L , für welche die obige Ungleichung gilt, es ist leicht zu sehen, dass noch immer

$$\forall x, y \in X : \sigma(f(x), f(y)) \leq Lip(f)\varrho(x, y).$$

Wenn $(Y, \sigma) = (X, \varrho)$ und $Lip(f) < 1$ (strikt ! kleiner eins), dann nennen wir f eine Kontraktion, und schreiben oft q oder q_f für $Lip(f)$.

SATZ 13.2. Banachscher Fixpunktsatz Sei (X, ϱ) ein vollständiger(!) metrischer Raum und $f : X \rightarrow X$ eine Kontraktion. Dann gibt es genau ein $x \in X$ mit $f(x) = x$, wir sagen $x = x_f$ ist ein/der Fixpunkt von f . Überdies gilt

- Für jedes $x \in X$ und $x_0 = x$, $x_k = f(x_{k-1})$, $k \in \mathbb{N}$, konvergiert die so definierte Folge gegen den Fixpunkt x_f . (x_f "zieht ganz X an".)
- Für jedes $x \in X$ gilt $\varrho(x, x_f) \leq \frac{\varrho(x, f(x))}{1 - Lip(f)}$.
- Sei $g : X \rightarrow X$ eine beliebige Abbildung und $y = g(y)$ einer derer Fixpunkte. Dann $\varrho(y, x_f) \leq \frac{\varrho_\infty(f, g)}{1 - Lip(f)}$. (Vergleiche Def 7.17, $\varrho_\infty(f, g) = \infty$ hier zugelassen.)

Im Weiteren werden wir die Gleichung $F(x) = y$ im \mathbb{R}^n lösen, um die Existenz der Umkehrfunktion F^{-1} zu zeigen. Wir betrachten zunächst, wie man das für $f : \mathbb{R} \rightarrow \mathbb{R}$ machen kann.

BEISPIEL 13.3. Newtoniteration. Nehmen wir an, dass wir die/eine Nullstelle einer differenzierbaren Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ suchen und bereits einen Näherungswert x_1 haben. Wie können wir den verbessern? Falls f affin (d.h. $f' \equiv \text{const.}$) wäre, dann ist die gesuchte Nullstelle genau in $x_1 - (f(x_1)/f')$. Wir hoffen, dass diese Idee auch funktioniert wenn wir

nahe genug bei der Nullstelle sind, da sich die Ableitung dann in diesen kleinen Intervall nicht mehr viel ändern wird. Deshalb definieren wir

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \text{ wenn } f'(x_k) \neq 0.$$

Da f' im Allgemeinen nicht konstant ist, wird x_{k+1} nicht unbedingt die Nullstelle von f sein, aber oftmals eine besser Näherung als x_k . Statt einer genauen (und etwas verwickelten) Untersuchung, wann $(x_k)_k$ gegen eine Nullstelle konvergiert betrachten wir folgendes Beispiel.

Wir wollen $\sqrt{\alpha}$ (zum Beispiel $\alpha = 2$) bestimmen, d.h. suchen die Nullstelle von

$$f(x) = x^2 - \alpha, \text{ also } x_{k+1} = x_k - \frac{x_k^2 - \alpha}{2x_k} = \frac{x_k^2 + \alpha}{2x_k}.$$

Wie konvergiert das gegen $\sqrt{\alpha}$? Es gilt

$$\Delta_{k+1} = x_{k+1} - \sqrt{\alpha} = \frac{x_k^2 - 2\sqrt{\alpha}x_k + \alpha}{2x_k} = \frac{(x_k - \sqrt{\alpha})^2}{2x_k},$$

also sollte Δ_{k+1} viel kleiner als Δ_k sein. Zum Beispiel mit $\alpha = 2$ und $x_1 = 3/2$ sehen wir $x_3 = 577/408$ und $\Delta_3 < 1/40000$, also $\Delta_4 < 10^{-9}$

In höheren Dimensionen müssen wir in der Newtonschen Iteration das $1/f'(x_k)$ durch $(DF(x_k))^{-1}$ ersetzen. Dazu müssen wir den Inversionsoperator $A \mapsto A^{-1}$ besser verstehen. Wir benutzen hierzu nicht die Cramersche Regel, sondern die Neumannsche Reihe (Carl Gottfried Neumann, Universität Leipzig 1868-1910) welche im Wesentlichen auch in unendlichen Dimensionen (wo es also keine Determinante gibt) genausogut funktioniert.

DEFINITION 13.4. Sei $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ linear, dann setzen wir

$$\|T\| (= \|T\|_{2,2}) = \sup\{\|T(x)\|_2 : x \in \mathbb{R}^n \text{ \& } \|x\|_2 \leq 1\} = \sup\{\|T(x)\| : x \in \mathbb{R}^n \text{ \& } \|x\| \leq 1\}$$

und nennen dies die "Norm" des Operators T .

Analog, setzen wir $\|A\| = \sup\{|A \cdot x| : |x| \leq 1\}$ für jede $(m \times n)$ -Matrix $A \in \mathbb{R}^{m \times n}$.

PROPOSITION 13.5. $A \mapsto \|A\|$ ist eine vollständige Norm auf $\mathbb{R}^{m \times n}$ und ist äquivalent zu $\|A\|_2 = \left(\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2\right)^{1/2}$. Ausserdem ist $\|A\| = \text{Lip}(T_A)$ im Sinne von Definition 13.1 wenn $T_A(x) = A \cdot x$ und es gilt

$$\|A \cdot B\| \leq \|A\| \cdot \|B\| \quad \text{Submultiplikativität der Norm}$$

wenn $B \in \mathbb{R}^{n \times k}$.

SATZ 13.6. Sei $n \geq 1$, dann ist

$$\text{Reg} = \text{Reg}_n = \{A \in \mathbb{R}^{n \times n} : A \text{ ist invertierbar, d.h. } \exists A^{-1}\}$$

eine offene Teilmenge von $\mathbb{R}^{n \times n}$ und

$$\Phi : A \in \text{Reg} \mapsto A^{-1} \in \mathbb{R}^{n \times n} \text{ ist eine } C^\infty\text{-Abbildung.}$$

Insbesondere, wenn $\|A\| < 1$, dann gilt die Reihenentwicklung¹

$$(Id - A)^{-1} = \sum_{k=0}^{\infty} A^k,$$

und wir finden, dass für jedes $A \in Reg$ das totale Differential der Invertierungsabbildung gegeben ist durch

$$D\Phi(A) : B \in \mathbb{R}^{n \times n} \mapsto -A^{-1} \cdot B \cdot A^{-1} = -\Phi(A) \cdot B \cdot \Phi A,$$

analog zu $(x^{-1})' = -x^{-1} \cdot x^{-1}$ auf der reellen Achse, $x \neq 0$.

PROPOSITION 13.7. "Prototyp des Satzes über die Umkehrfunktion" Sei $U \subset \mathbb{R}^n$ offen, $F : U \rightarrow \mathbb{R}^n$ in $C^1(U)$, $0 \in U$, $F(0) = 0$ und $DF(0) = Id$. Dann gibt es ein $\varepsilon > 0$ so dass

$$B(0, \varepsilon) \subset U \text{ und } \forall x \in B(0, \varepsilon) : \|DF(x) - Id\| < \frac{1}{2}. \quad (13.1)$$

Wenn (??) gilt, dann ist $f = F|_{B(0, \varepsilon)}$ injektiv, f^{-1} stetig und für jedes $r \in (0, \varepsilon)$ gilt $F(\overline{B(0, r)}) \supset \overline{B(0, \frac{r}{2})}$ sowie $D(f^{-1})(0) = Id$.

SATZ 13.8. "Von der Umkehrfunktion" Sei $x \in U \subset \mathbb{R}^n$ offen und $F : U \rightarrow \mathbb{R}^n$ in $C^1(U)$.

Wenn $DF(x)$ invertierbar ist, dann gibt es ein $G \subset U$ offen mit $x \in G$, so dass $f = F|_G$ injektiv ist und die Umkehrfunktion f^{-1} in $C^1(\tilde{G})$ ist, wobei $\tilde{G} = f(G)$ in \mathbb{R}^n offen ist. Ausserdem gilt,

$$\forall z \in G : D(f^{-1})(f(z)) = [Df(z)]^{-1}.$$

Falls $F \in C^k(G)$, dann ist auch $f^{-1} \in C^k(\tilde{G})$.

Umgekehrt, wenn es ein offenes $\tilde{G} \subset \mathbb{R}^n$, ein $g \in C^1(\tilde{G})$ und ein ε gibt, so dass für jedes $y \in B(x, \varepsilon)$ die Identität $y = g(F(y))$ gilt, dann ist $DF(x)$ invertierbar.

BEMERKUNG 13.9. Sei $U \subset \mathbb{R}^n$ offen und $F : U \rightarrow \mathbb{R}^m$ in $C^1(U)$. Wenn $x \in U$ und $DF(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ surjektiv ist, dann existieren m kanonische Einheitsvektoren e_{i_1}, \dots, e_{i_m} , so dass die entsprechenden Vektoren $DF(e_{i_j}), j = 1, \dots, m$, den gesamten $\mathbb{R}^m = \text{span}\{DF(e_{i_1}), DF(e_{i_2}), \dots, DF(e_{i_m})\}$ auch aufspannen.

Somit erfüllt $\hat{F}(y) = F(x + \sum_{j=1}^m y_j e_{i_j})$, für $y \in B(0, \varepsilon) \subset \mathbb{R}^m$ und ε positiv aber genügend klein, die Voraussetzungen des Satzes über die Umkehrfunktion und damit ist $\hat{F}(B(0, \varepsilon)) \subset F(B(x, \varepsilon) \cap U)$ also eine Umgebung von $F(x) = \hat{F}(0)$. Proposition 12.4 folgt.

SATZ 13.10. "Über die implizite Funktion"

Seien $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ offen und $f : U \times V \rightarrow \mathbb{R}^m$ in $C^1(U \times V)$. Wenn $(x_0, y_0) \in U \times V$, $f(x_0, y_0) = 0$ und $A = \left(\frac{\partial f_j}{\partial y_i}(x_0, y_0) \right)_{1 \leq i, j \leq m}$ invertierbar ist, dann gibt es $\varepsilon, \delta > 0$ und ein $g : B(x_0, \delta) \subset U \rightarrow B(y_0, \varepsilon) \subset V$, so dass

$$\forall x \in B(x_0, \delta) \forall y \in B(y_0, \varepsilon) : f(x, y) = 0 \Leftrightarrow y = g(x),$$

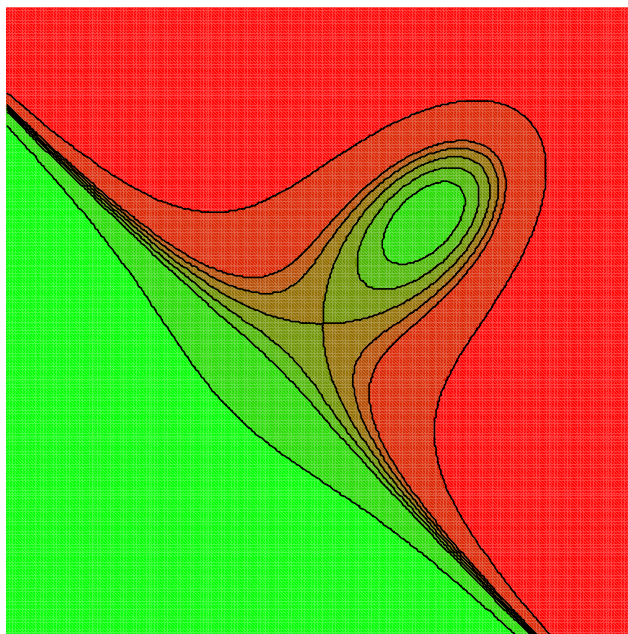
¹vergleiche mit der geometrischen Reihe aus Analysis I – hier sind Frage und Antwort nur vertauscht!

d.h. die Funktion g ist implizit durch die Bedingung $f(x, g(x)) = 0$ gegeben. Ausserdem gilt

$$Dg(x_0) = -A^{-1} \cdot \left(\frac{\partial f_j}{\partial x_i}(x_0, y_0) \right)_{1 \leq i \leq n, 1 \leq j \leq m},$$

also ist $g \in C^1$. Wie zuvor, wenn überdies $F \in C^k$, dann auch $g \in C^k$.

BEISPIEL 13.11. Wenn man die Niveaumengen $F^{-1}(\alpha)$ einer Funktion $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ studieren will, kann man diese als Nullmenge der Funktion $F_\alpha = F - \alpha$ betrachten und versuchen mittels des Satzes über die implizite Funktion eine Darstellung als Graph einer Funktion, d.h. als Menge $\{(x, g(x)) : x \in I\}$ oder $\{(h(y), y) : y \in I\}$ zu finden. Dies ist zumindest lokal möglich, wenn $\partial F/\partial y$ bzw. $\partial F/\partial x$ nicht verschwinden. Insbesondere ist eine Darstellung immer möglich und die Niveaumenge somit nahe (x_0, y_0) zumindest eine C^1 -Kurve, wenn (x_0, y_0) kein kritischer Punkt von F ist.



An der Funktion

$$F(x, y) = x^3 + y^3 - 3xy$$

lässt sich dieser Effekt sehr schön beobachten, die Stelle wo sich die Niveaukurve selbst schneidet, ist der Koordinatenursprung $(0, 0)$. Grün bedeutet negative, rot positive Funktionswerte.

Die Sätze über die Umkehr- und implizite Funktion spielen auch (und gerade) in höheren Dimensionen eine wichtige Rolle, um die geometrische Struktur von Flächen oder anderen Teilmengen des \mathbb{R}^n zu verstehen, auch wenn die Anschauung an ihre Grenzen stößt. Dies werden wir in zum Beispiel in der Maß- und Integrationstheorie nutzen.