Argumente der Analysis II, SS 2000, H. Karcher

1. Überblick

Hauptthema dieses Textes ist die mehrdimensionale Differentialrechnung. Ebenso wie Teil I einerseits zur genaueren Information der Übungsleiter und andererseits als roter Faden für die Studierenden des WS 99/00 gedacht war, so soll dieser Text die Vorlesung des SS unterstützen.

Zunächst werden Hauptsätze über stetige Abbildungen wiederholt und die gleichmäßige Konvergenz besprochen: dabei wird die Dimensions**un**abhängigkeit der Argumentation hervorgehoben. Ebenfalls wiederholt wird Integral- und Differentialrechnung der Kurven im euklidischen Raum $\mathbb{R}^d$. Ausgebaut wird diese Kurventheorie mit der Behandlung gewöhnlicher Differentialgleichungen. Dabei spielt die Veranschaulichung von Abbildungen $\mathbb{R}^d \to \mathbb{R}^d$ als Vektorfelder eine wichtige Rolle. Es folgt der zentrale Begriff der mehrdimensionalen Differenzierbarkeit und dessen Beziehung zu eindimensionalen "partiellen" Ableitungen. Natürlich braucht man wieder Differentiationsregeln, auch ein Schrankensatz gilt. Extremwertbestimmungen bei mehrdimensionalen Funktionen sind naheliegende Verallgemeinerungen der eindimensionalen Situation. Ein mehrdimensionaler Satz mit weitreichenden Konsequenzen verallgemeinert das Lösen von Linearen Gleichungssystemen auf nichtlineare Abbildungen (Umkehrsatz und Satz über implizit definierte Funktionen). An mehreren Stellen können die Voraussetzungen mit reeller Differenzierbarkeit oder mit komplexer Differenzierbarkeit formuliert werden; die Gegenüberstellung bietet Gelegenheit zur Wiederholung von Argumenten, aber sie bietet auch Hilfen für die Anschauung beim Übergang von einer zu mehreren Dimensionen.

2. Vektorräume und Stetigkeit.

Die Länge oder die Norm von Elementen eines Vektorraums V wird durch eine Abbildung

$$|\,.\,| : V \longrightarrow \mathbb{R}_+$$

beschrieben, die folgende Axiome erfüllt:

1.) "positiv definit": $0 \le |v| \quad \text{und} \quad 0 = |v| \Rightarrow v = 0$.

2.) "homogen": $|r \cdot v| = |r| \cdot |v| \quad (v \in V, r \in \mathbb{R})$.

3.) "Dreiecksungleichung": $|v + w| \le |v| + |w|$.

Manche Normen ("Pythagoras Normen") kommen von Skalarprodukten, $|v| := \langle v, v\rangle^{1/2}$. Mit Hilfe von Normen werden **Abstände** definiert:

$$\begin{aligned} d(v,w) &:= |v - w| \\ &= d(v+u, w+u) \quad \text{“translationsinvariant”}. \end{aligned}$$

Grundlegende Definitionen aus der 1-dimensionalen Analysis werden mit diesen Abständen auf höhere Dimensionen verallgemeinert:

beschränkt: Eine Teilmenge $A \subset \mathbb{R}^d$ heißt beschränkt, falls es eine "Schranke" S gibt, so daß für alle $a \in A$ gilt $d(0,a) = |a| \leq S$.

dehnungsbeschränkt: Eine Abbildung $F : A \subset \mathbb{R}^d \to \mathbb{R}^e$ besitzt die "Dehnungsschranke" oder "Lipschitz Schranke" L, falls gilt $x, y \in A \Rightarrow |F(y) - F(x)|_{\mathbb{R}^e} \leq L \cdot |y - x|_{\mathbb{R}^d}$.

konvergent: Eine Folge $\{a_n\} \subset \mathbb{R}^d$ heißt "konvergent" gegen $a \in \mathbb{R}^d$ falls $\{d(a_n, a)\}_{n \in \mathbb{N}}$ eine reelle Nullfolge ist.

stetig, folgenstetig: Eine Funktion $F : A \subset \mathbb{R}^d \to \mathbb{R}^e$ heißt bei $a \in A$ stetig, falls für jede gegen a konvergente Folge $\{a_n\} \subset A$ gilt: $F(a_n)$ konvergiert gegen $F(a)$.

stetig, ε-δ-stetig: Eine Funktion $F : A \subset \mathbb{R}^d \to \mathbb{R}^e$ heißt bei $a \in A$ stetig, falls gilt:
Zu jeder Fehlerschranke $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß

$$x \in A,\ d(x,a) \leq \delta \Rightarrow d(F(x), F(a)) \leq \varepsilon.$$

Cauchyfolge: Eine Folge $\{c_n\} \subset \mathbb{R}^d$ heißt Cauchyfolge, falls es eine reelle Nullfolge $\{r_n\}$ gibt, so daß gilt:

$$m \geq n \Rightarrow d(c_m, c_n) \leq r_n.$$

(oder: Zu jedem $\varepsilon > 0$ gibt es n_ε, so daß $m, n \geq n_\varepsilon \Rightarrow d(c_m, c_n) \leq \varepsilon$.)

vollständig: Eine Teilmenge $A \subset \mathbb{R}^d$ heißt "vollständig", falls jede Cauchyfolge $\{a_n\} \subset A$ einen Grenzwert $a \in A$ hat.

Wegen der Vollständigkeit von $\mathbb{R}$ ist auch $\mathbb{R}^d$ mit der standard Pythagoras-Norm, also $|x| := \left(\sum_{k=1}^n x_k^2\right)^{1/2}$, vollständig.

Die meisten Sätze über stetige Funktionen sind entweder in der einen oder in der anderen Formulierung leicht zu beweisen. Daher ist die Äquivalenz der beiden Stetigkeitsformulierungen ein grundlegender Satz für das Verständnis von Stetigkeit. Ehe wir dazu kommen, muß bemerkt und behandelt werden, daß alle Definitionen von der verwendeten Norm abzuhängen scheinen. Tatsächlich hängen nur die Konstanten in den Definitionen von den Normen ab, nicht dagegen die Existenz solcher Konstanten. Zur Diskussion benötigen wir eine weitere

Definition: Eine Norm $|\ \ |_2 : V \to \mathbb{R}_+$ **dominiert** eine Norm $|\ \ |_1 : V \to \mathbb{R}_+$, falls es eine Konstante C gibt, so daß für alle $v \in V$ gilt

$$|v|_1 \leq C \cdot |v|_2.$$

Satz. Falls die dominierende Norm $|\ \ |_2$ in obigen Definitionen verwendet wird, so ist $|\ \ |_1 : V \to \mathbb{R}_+$ eine dehnungsbeschränkte Funktion.

Beweis. $\big||v|_1 - |w|_1\big|_{\mathbb{R}} \leq |v - w|_1 \leq C \cdot |v - w|_2$.

Satz. Eine beliebige Norm $|\ \ |_1$ auf V wird durch eine Pythagoras-Norm $|v|_2 := \sqrt{\langle v, v\rangle}$ dominiert.

Beweis. Wähle zu dem Skalarprodukt eine ON-Basis $\{e_1, \cdots, e_d\}$ für V und schreibe v als $v = \sum_{k=1} x_k \cdot e_k$. Dann gilt $|v|_2 = \left(\sum_{k=1}^{d}(x_k)^2\right)^{1/2}$. Ferner gilt wegen der Dreiecksungleichung und der Homogenität für $|\ \ |_1$

$$|v|_1 = \Big|\sum_{k=1}^{d} x_k e_k\Big|_1 \leq \sum_{k=1}^{d} |x_k|_{\mathbb{R}} \cdot |e_k|_1$$

und weiter wegen der Schwarzschen Ungleichung

$$\leq \left(\sum_{k=1}^{d}(x_k)^2\right)^{1/2} \cdot \left(\sum_{k=1}^{d}(|e_k|_1)^2\right)^{1/2} = |v|_2 \cdot C.$$

(Hier wird der rechte Faktor als C definiert.)

Folgerung. Zwei verschiedene Pythagoras-Normen dominieren sich gegenseitig. Daher spielt es keine Rolle für unsere Begriffe, welches Skalarprodukt wir verwenden.

Wir wiederholen zunächst zwei Hauptsätze für stetige Abbildungen, wobei wir Pythagoras-Normen zur Definition der Stetigkeit verwenden.

Äquivalenzsatz: Die beiden Stetigkeitsformulierungen sind äquivalent.

Beweis. a) F sei ε-δ-stetig bei a und $\{a_n\}$ sei gegen a konvergent. Zu zeigen ist, daß $\{F(a_n)\}$ gegen $F(a)$ konvergiert. Zu jedem gegebenen $\varepsilon > 0$ gilt es nach Voraussetzung zuerst $\delta > 0$, so daß $x \in a,\ d(x, a) \leq \delta \Rightarrow d(F(x), F(a)) \leq \varepsilon$. Dann gibt es nach Voraussetzung über $\{a_n\}$ zu $\delta > 0$ ein n_δ, so daß

$$n \geq n_\delta \Rightarrow d(a_n, a) \leq \delta.$$

Aus beidem zusammen folgt

$$n \geq n_\delta \Rightarrow d(F(a_n), F(a)) \leq \varepsilon.$$

Das war die einfache Richtung. Für die andere Folgerung benötigen wir einen indirekten Beweis:

b) F sei folgenstetig bei a, aber nicht ε-δ-stetig. Das heißt: Zu einem $\varepsilon^* > 0$ ist weder $\delta = \frac{1}{2}$, noch $\delta = \frac{1}{4}$, noch irgendein $\delta_n = 2^{-n}$ klein genug: Zu jedem dieser δ_n gibt es einen Ausreißer $a_n \in A$ mit $d(a_n, a) \leq \delta_n$ aber $d(F(a_n), F(a)) > \varepsilon^*$. Dann ist $\{a_n\} \subset A$ gegen a konvergent, aber $\{F(a_n)\}$ konvergiert nicht gegen $F(a)$ — im Widerspruch zur vorausgesetzten Folgenstetigkeit.

Beschränktheitssatz. $A \subset \mathbb{R}^d$ sei vollständig und beschränkt; ferner sei die Abbildung $F : A \to \mathbb{R}^e$ stetig. Dann ist $F(A)$ beschränkt.

Indirekter Beweis mit Quaderhalbierung. $F(A)$ sei unbeschränkt. Wähle zu dem Skalarprodukt eine ON-Basis $\{e_1, \cdots, e_d\}$. Die Beschränktheit von A bedeutet, daß es ein Quader $Q_1 := \{v = \sum_{k=1}^d x_k e_k; |x_k| \le l_k\}$ mit Kanten der Länge l_k parallel zu den Vektoren e_k gibt mit $A \subset Q_1$. Wähle $a_1 \in A \cap Q_1$ mit $|F(a_1)| > 1$ — das ist möglich weil 1 keine Schranke für die unbeschränkte Menge $F(a)$ ist. Betrachte die 2^d Teilquader von Q_1, die durch Halbieren der Kantenlängen von Q_1 entstehen. Für mindestens einen, genannt Q_2, ist wahr, daß $F(A \cap Q_2)$ nicht beschränkt ist. Wähle $a_2 \in A \cap Q_2$ mit $|F(a_2)| > 2$. Konstruiere durch Wiederholung dieses Schrittes eine Quaderschachtelung Q_n, so daß F auf $A \cap Q_n$ nicht beschränkt ist, also $a_n \in A \cap Q_n$ gewählt werden kann mit $|F(a_n) \ge n$.

Wegen $a_n \in Q_n$ ist $\{a_n\}$ eine Cauchyfolge, die wegen der Vollständigkeit von A einen Grenzwert $a \in A$ hat. Wegen $|F(a_n)| \ge n$ konvergiert $\{F(a_n)\}$ nicht gegen $F(a)$ — im Widerspruch zur vorausgesetzten (Folgen-) Stetigkeit.

Damit sind wir so weit, daß wir die Unabhängigkeit der aus dem 1-dimensionalen nach $\mathbb{R}^d$ verallgemeinerten Begriffe von der auf $\mathbb{R}^d$ verwendeten Norm zeigen können. Wir wissen schon, daß Pythogoras-Normen andere Normen dominieren, wir benötigen noch die Umkehrung, also den

Satz. Eine beliebige Norm $|\ \ |_1$ auf $\mathbb{R}^d$ dominiert eine beliebige Pythogoras-Norm $|\ \ |_2$ auf $\mathbb{R}^d$.

Beweis. Die Funktion $f = |\ \ |_1 : \mathbb{R}^d \to \mathbb{R}_+$ ist stetig bezüglich $|\ \ |_2$ (sogar dehnungsbeschränkt). Ferner ist f auf der Einheitssphäre $D := \{x \in \mathbb{R}^d, |x|_2 = 1\}$ ungleich 0, weil die Norm $|\ \ |_1$ positiv definit ist. Daher ist die Funktion $\frac{1}{f} : D \to \mathbb{R}_+$ stetig und nach dem Beschränktheitssatz ist $0 < \frac{1}{f}(D) \le S$ für ein geeignetes S. Damit gilt für alle x mit $|x|_2 = 1$ auch $\frac{1}{f}(x) \le S$ oder $\frac{1}{S}|x|_2 \le |x|_1$. Wegen der Homogenität beider Normen gilt dies nicht nur für alle $|\ \ |_2$-Einheitsvektoren sondern für alle $x \in \mathbb{R}^d$. Also haben wir mit $|x|_2 \le S \cdot |x|_1$ eine Dominationskonstante gefunden.

Folgerung. Alle am Anfang zusammengestellten Begriffe hängen tatsächlich nicht von der auf $\mathbb{R}^d$ verwendeten Norm ab, weder die Beschränktheit, noch die Vollständigkeit, weder Dehnungsbeschränktheit noch Stetigkeit. Diese Grundbegriffe der Analysis erfassen also wichtige Eigenschaften von $\mathbb{R}^d$ und von Abbildungen $F : \mathbb{R}^d \to \mathbb{R}^e$.

Besonders hervorheben möchte ich noch, ein wie großer Teil der Argumente ganz unabhängig von der Dimension funktioniert. Nur an zwei Stellen spielt eine Rolle, daß die Dimension endlich ist: Für die Dominierungskonstante einer beliebigen Norm durch eine Pythagoras-Norm und für die endliche Anzahl der Teilquader bei Halbierung der Kantenlängen. Natürlich werden die auf diesen Argumenten beruhenden Sätze mehrfach zitiert, und deswegen bleibt in unendlichdimensionalen Vektorräumen mit Norm von den hier betonten Sätzen nicht viel übrig. Sobald man jedoch die Vollständigkeit auf Grund anderer

Mittel zur Verfügung hat, lassen sich die dimensionsunabhängigen Argumente so bequem wie hier verwenden. Wir kommen bald dazu.

3. Kurven, Ableitungen und Integrale mit Werten in $\mathbb{R}^d$.

Für Kurven $f, g : [a, b] \to \mathbb{R}^d$ verweise ich auf frühere Texte zu den Differentiationsregeln,
Linearität: $(\alpha \cdot f + \beta \cdot g)' = \alpha \cdot f' + \beta \cdot g'$,
Produktregel: $\langle f, g\rangle' = \langle f', g\rangle + \langle f, g'\rangle$,
Schrankensatz: $|f'| \leq L \Rightarrow |f(t_2) - f(t_1)| \leq L \cdot |t_2 - t_1|$,
Kettenregel: $f : [a, b] \to \mathbb{C}$, $F : \mathbb{C} \to \mathbb{C} \;\Rightarrow\; (F \circ f)'(t) = F'(f(t)) \cdot f'(t)$.
Nachzutragen ist der

Mittelwertsatz. $f : [a, b] \to \mathbb{R}$ sei stetig und in (a, b) differenzierbar (keine weiteren Voraussetzungen über f', nur Existenz). Dann gilt:

Es gibt eine Tangente mit derselben Steigung wie die Sehne, d.h.
Es gibt $\tau \in (a, b)$ mit $\qquad f'(\tau) = (f(b) - f(a))/(b - a)$.

Beweis. Es genügt, den Satz für $f(a) = f(b) = 0$ zu beweisen, da wir die Sehne von der Funktion abziehen können ("Satz von Rolle"). Als *stetige* Funktion besitzt f auf $[a, b]$ ein Maximum und ein Minimum. Entweder ist $f = 0$ oder einer der Extremwerte ist $\neq 0$, tritt also *nicht* am Rande auf sondern im Innern, etwa bei $\tau \in (a, b)$. Da f an dieser Extremstelle differenzierbar ist, folgt $f'(\tau) = 0$ und das ist gleich der Sehnensteigung 0.

Bemerkung. Die Standardanwendung des Mittelwertsatzes ist: Hat man eine Schranke $L \geq |f'|$, so ist L Dehnungsschranke für f, also die "Schrankensatzfolgerung". — Unter den bisher besprochenen Sätzen hat der Maximumsatz für stetige Funktionen den längsten Beweis; der Mittelwertsatz ist jetzt Spitzenreiter, da im Beweis der Maximumsatz zitiert wurde.

Der **Integralrechnung** ist mehr als der Differentialrechnung hinzuzufügen. Wir hatten zu Kurven $f : [a, b] \to \mathbb{R}^d$, die gleichmäßig durch stückweise lineare Funktionen approximierbar sind, Stammfunktionen $F : [a, b] \to \mathbb{R}^d$, $F' = f$ konstruiert. Wir hatten ferner gesehen, daß Riemannsummen von f zu immer feineren Einteilungen mit $F(b) - F(a)$ immer besser übereinstimmen. Für dehnungsbeschränkte oder auch nur stetige f ergibt sich aus diesen Vorbereitungen die Riemann Integrierbarkeit zusammen mit dem Hauptsatz der Differential- und Integralrechnung. Als Anwendung hatten wir zunächst für dehnungsbeschränkte $f : [a, b] \to \mathbb{R}^d$ die Kurvenlänge Länge(f) als Supremum der Längen von Sehnenpolygonen *definiert* und dann für stetig differenzierbare f *bewiesen*:

$$\text{Länge}(f) = \int_a^b |f'(t)|dt.$$

Auf der Schule haben Sie weniger die Bogenlänge von Kurven mit dem Integral in Verbindung gebracht, sondern eher den Flächeninhalt unter dem Graphen einer Funktion f :

$[a,b] \to \mathbb{R}_+$. Das Problem dabei ist, daß ich mit den bisher entwickelten Mitteln noch keine *Definition* "des" Flächeninhalts krummlinig berandeter ebener Gebiete angeben kann — anders als bei der Bogenlänge. Trotzdem muß ich sagen, wie unsere Integrale mit Ihrem Schulstoff in Verbindung stehen. Das ist so:

Betrachte eine (z. B. Polynom-) Funktion $f : [a,b] \to \mathbb{R}_+$. Angenommen es gibt einen "Flächeninhalt" A der Menge $G := \{(x,y) \mid a \le x \le b, 0 \le y \le f(x)\}$. Angenommen ferner, derartige "Flächeninhalte" könnten so definiert werden, daß aus $G_1 \subset G_2$ folgt Fläche$(G_1) \le$ Fläche(G_2). Dann können wir zeigen: Die Zahl A kann mit Hilfe einer Stammfunktion F, $F' = f$ so berechnet werden, wie Sie das kennen.

(Erinnerung: $A = F(b) - F(a)$. A liegt unterhalb der Obersummen und oberhalb der Untersummen.)

Außerhalb der Mathematik ist dann klar, daß A der Flächeninhalt "ist" — *wie soll's denn sonst sein?* Innerhalb der Mathematik geht das nicht, weil es zu viele Beispiele gibt, bei denen aus plausibel erscheinenden Annahmen Widersprüche folgen. Immerhin, bis auf "wie soll's denn sonst sein?" können Sie jetzt mit Integralen auch Flächeninhalte berechnen.

Schrankensatz und numerische Integration.

Offenbar gilt für *lineare* Funktionen $f : [a,b] \to \mathbb{R}$

$$\begin{aligned}\int_a^b f(x)dx &= \tfrac{1}{2}(f(a)+f(b)) \cdot (b-a),\\ \int_a^b f(x)dx &= f\left(\tfrac{a+b}{2}\right) \cdot (b-a).\end{aligned}$$

Für allgemeinere f gelten diese Formeln natürlich nicht. Die erste,

$FS(f) := \frac{1}{2}(f(a)+f(b)) \cdot (b-a)$, ist der Flächeninhalt under der Sehne, die zweite

$FT(f) := f\left(\frac{a+b}{2}\right) \cdot (b-a)$, ist der Flächeninhalt unter der Tangente im Mittelpunkt $(a+b)/2$ (und beidemal zwischen $x = a$ und $x = b$).

Hat man nun Schranken für die zweite Ableitung $-B \le f'' \le +B$, so liegt die Differenz von f und der Tangente zwischen $\pm\frac{1}{2}B \cdot (x - \frac{a+b}{2})^2$; die Differenz zwischen f und der Sehne liegt zwischen $\pm\frac{1}{2}B(x-a) \cdot (b-x)$. Daher folgt als Fehlerabschätzung:

$$\begin{aligned}\Big|\int_a^b f(x)dx - FT(f)\Big| &\le \int_a^b \frac{B}{2} \cdot \left(x - \frac{a+b}{2}\right)^2 dx = \frac{B}{24}(b-a)^3,\\ \Big|\int_a^b f(x)dx - FS(f)\Big| &\le \int_a^b \frac{B}{2}(x-a) \cdot (b-x)dx = \frac{B}{12}(b-a)^3.\end{aligned}$$

Kann man dies verbessern? Und woran soll man ein "besseres" Verfahren überhaupt erkennen? Natürlich kann man das Intervall $[a,b]$ erst in zwei gleich große Teile teilen und auf jeder Hälfte die Sehnenfläche bzw. die Tangentenfläche berechnen. Über den kleineren Intervallen liefert unsere Fehlerabschätzung $\frac{1}{8}$ so große Werte, aber da man die Beiträge der beiden Hälften noch addieren muß, nehmen die Fehlerschranken für das Intervall $[a,b]$ nur um $\frac{1}{4}$ ab. Wir verkleinern also den Fehler, aber wir nennen das nicht ein "besseres" Verfahren, sondern "wiederholte Anwendung" desselben Verfahrens. Man würde doch wohl von einem "besseren" Verfahren reden, wenn die Fehlerschranken bei einer solchen "wiederholten Anwendung" stärker als mit dem Faktor $\frac{1}{4}$ verbessert würden. (Oder fällt Ihnen ein einleuchtenderes Kriterium für ein "besseres Verfahren" ein?) Aber, kann man den Verbesserungsfaktor $\frac{1}{4}$ tatsächlich überbieten?

Für quadratische Funktionen f ist der eine Fehler positiv, der andere negativ und dem Betrage nach ist der Fehler von $FT(f)$ *halb so groß* wie der von $FS(f)$. Daher sollte man es mit dem $(2:1)$-Mittelwert

$$SS(f) := \frac{2}{3} \cdot FT(f) + \frac{1}{3} \cdot FS(f) = \frac{1}{6} \left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right) \cdot (b-a)$$

versuchen. Dieses sogenannte Simpson-Verfahren berechnet das Integral desjenigen kubischen Polynoms, das bei $a, \frac{a+b}{2}, b$ dieselben Werte wie f und bei $\frac{a+b}{2}$ auch noch dieselbe Ableitung wie f hat. (Rechnen Sie das bitte nach.) Das läßt sich zwar gut an, aber haben wir wirklich ein besseres Verfahren gefunden? Mehrfache Anwendung des Schrankensatzes zeigt, daß die Differenz aus f und dem eben beschriebenen kubischen Polynom zwischen $\pm\frac{S_4}{24}(x-a)\cdot(x-\frac{a+b}{2})^2\cdot(b-x)$ liegt; dabei ist $S_4 \geq |f^{(4)}|$ eine Schranke der vierten Ableitung von f in $[a,b]$. Daraus folgt:

$$\Big| \int_a^b f(x)dx - SS(f) \Big| \leq \frac{S_4}{24} \int_a^b (x-a)\left(x - \frac{a+b}{2}\right)^2 (b-x)\,dx = \frac{S_4}{2880} \cdot (b-a)^5.$$

In der Tat, wendet man diese Formel auf die beiden halb so langen Teilintervalle an und addiert die Teilergebnisse, so wird die Fehlerschranke um den Faktor $\frac{1}{16}$ kleiner. Das Simpsonverfahren ist im Sinne unseres Kriteriums eine deutliche Verbesserung.

Beachten Sie bitte, daß die besprochenen numerischen Integrationsformeln *lineare* Abbildungen definieren, z. B.

$$SS(\alpha \cdot f + \beta \cdot g) = \alpha \cdot SS(f) + \beta \cdot SS(g).$$

Außerdem sollte man derartige Formeln nie benutzen, ohne zu überprüfen, daß zumindest für konstante und für lineare Funktionen f die Integrale richtig berechnet werden.

4. Gleichmäßige Konvergenz.

Bisher hatte ich bei konvergenten Funktionenfolgen $F_n : \mathbb{R}^d \to \mathbb{R}^e$ gleichmäßige Fehlerabschätzungen und das Archimedes Axiom benutzt, um Aussagen über die Grenzfunktion

F_∞ zu bekommen. Beispiele:

$$|F_n(x)| \leq C \Rightarrow |F_\infty(x)| \leq C$$
$$|F_n(x) - F_n(y)| \leq L \cdot |x-y| \Rightarrow |F_\infty(x) - F_\infty(y)| \leq L \cdot |x-y|.$$

Falls $|x-y| \leq \delta$ (unabhängig von n) $\Rightarrow |F_n(x) - F_n(y)| \leq \varepsilon$ gilt, so folgt auch

$$|x-y| \leq \delta \Rightarrow |F_\infty(x) - F_\infty(y)| \leq \varepsilon.$$

Man kann nun die Gleichmäßigkeitsvoraussetzungen von den Fehlerabschätzungen auf die Konvergenz schieben. Das erfordert dann etwas schwierigere Argumente, führt aber zu einem erstaunlichen Satz über konvergente Folgen stetiger Funktionen. Der zu beweisende Satz steht im Standardaufbau der Analysis viel weiter am Anfang als er hier besprochen wird.

Definition. Eine Folge von Abbildungen $F_n : A \subset \mathbb{R}^d \to \mathbb{R}^e$ heißt **gleichmäßig konvergent**, falls es zu jeder Fehlerschranke $\varepsilon > 0$ einen Garantieindex n_ε gibt — ausdrücklich unabhängig von $a \in A$ — so daß für alle $a \in A$ gilt:

$$m, n \geq n_\varepsilon \Rightarrow |F_m(a) - F_n(a)| \leq \varepsilon;$$

Wegen der Vollständigkeit von $\mathbb{R}^e$ folgt aus dieser Cauchy-Eigenschaft, daß eine Grenzfunktion $F_\infty : A \subset \mathbb{R}^d \to \mathbb{R}^e$ existiert, und daß gilt:

$$n \geq n_\varepsilon \Rightarrow |F_\infty(x) - F_n(x)| \leq \varepsilon.$$

Hauptsatz. Jede gleichmäßig konvergente Folge *stetiger* Funktionen $F_n : A \subset \mathbb{R}^d \to \mathbb{R}^e$ besitzt eine *stetige* Grenzfunktion $F_\infty : A \subset \mathbb{R}^d \to \mathbb{R}^e$.

Beispiel. Die Funktionenfolge $\{f_n\}$ zur Konstruktion des Weierstraß-Oszillators ist gleichmäßig geometrisch majorisiert, also gleichmäßig konvergent. Daher ist die Grenzfunktion stetig. (Erinnerung: Mit der Sägezahnfunktion $\text{säg}\,(x) := \min\{|x-m|; m \in \mathbb{Z}\}$ hatten wir definiert:

$$f_n(x) := \sum_{k=1}^{n} 2^{-k} \cdot \text{säg}\,(4^k \cdot x).)$$

Beweis des Hauptsatzes mit einem $\frac{\varepsilon}{3}$-Argument. Sei $\varepsilon > 0$ und $a \in A$ gegeben. Zu finden ist $\delta > 0$, so daß gilt:

$$x \in A, |x-a| \leq \delta \Rightarrow |F_\infty(x) - F_\infty(a)| \leq \varepsilon.$$

Dazu liefert die gleichmäßige Konvergenz zunächst n_ε, so daß für alle $x \in A$ und $n \geq n_\varepsilon$ gilt

$$|F_\infty(x) - F_n(x)| \leq \frac{\varepsilon}{3}.$$

Nun wird ein $n^* \geq n_\varepsilon$ ausgesucht und mit der Stetigkeit von F_{n^*} läßt sich ein $\delta^* > 0$ finden, so daß
$$x \in A, |x - a| \leq \delta^* \Rightarrow |F_{n^*}(x) - F_{n^*}(a)| \leq \frac{\varepsilon}{3}.$$
(Beachten Sie: In den meisten Fällen, z. B. beim Weierstraß Oszillator, ist es ungünstig, n^* "möglichst groß" zu wählen, weil das zu unnötig kleinen δ^* führt. Die Voraussetzungen erlauben nämlich, daß die Funktionen F_n mit wachsendem n "immer schlechter" stetig sind.)

Nun beendet die Dreiecksungleichung den Beweis:

$$x \in A, |x - a| \leq \delta^* \Rightarrow$$
$$|F_\infty(x) - F_\infty(a)| \leq |F_\infty(x) - F_{n^*}(x)| + |F_{n^*}(x) - F_{n^*}(a)| + |F_{n^*}(a) - F_\infty(a)| \leq 3 \cdot \frac{\varepsilon}{3}.$$

Der Hauptsatz läßt sich zusammen mit der Sprache der Normen auf Vektorräumen zu sehr prägnanten und leicht benutzbaren Aussagen umformulieren.
Sei $V := \{F : A \subset \mathbb{R}^d \to \mathbb{R}^e; F \text{ stetig und beschränkt}\}$ der Vektorraum der stetigen und beschränkten Abbildungen von $A \subset \mathbb{R}^d$ nach $\mathbb{R}^e$. Wir definieren eine Norm, genannt Supremum-Norm auf V durch
$$|F|_V = |F| := \sup_{a \in A} |F(a)| = \sup |F(A)|.$$
Die Axiome für Normen sind erfüllt. Zur Dreiecksungleichung:
$$\begin{aligned} |F + G| &= \sup_{a\in A} |F(a) + G(a)| \leq \sup_{a\in A}(|F(a)| + |G(a)|) \\ &\leq \sup_{a\in A} |F(a)| + \sup_{b\in A} |G(b)| = |F| + |G|. \end{aligned}$$

Umformulierung des Hauptsatzes. Der Vektorraum V mit der eben definierten Supremum-Norm ist **vollständig**.

Beweis. Eine $|\ \ |$-Cauchyfolge $\{F_n\} \subset V$ ist nach Definition eine gleichmäßig konvergente Folge stetiger Abbildungen; diese konvergiert nach dem Hauptsatz gleichmäßig gegen eine Grenzfunktion f_∞. Mit anderen Worten: Zu $\varepsilon > 0$ gibt es n_ε (unabhängig von $x \in A$), so daß für alle $x \in A$ gilt:
$$n \geq n_\varepsilon \Rightarrow |F_\infty(x) - F_n(x)| \leq \varepsilon,$$
also auch $|F_\infty - F_n| = \sup_{x\in A} |F_\infty(x) - F_n(x)| \leq \varepsilon$. Damit konvergiert $\{F_n\}$ auch im Sinne der Norm (also wie im ersten Kapitel erklärt) gegen F_∞, d. h. jede $|\ \ |$-Cauchyfolge in V konvergiert, V ist vollständig. — Beachten Sie wieder die Dimensionsunabhängigkeit der Argumentation.

Bemerkung Wenn die Menge A vollständig und beschränkt ist, so sind stetige Abbildungen $F : A \to \mathbb{R}^e$ immer auch beschränkt. Andernfalls gibt es unbeschränkte stetige Abbildungen; in diesem Fall ist gleichmäßige Konvergenz auf A ein etwas zu starker Begriff. Z.B. sind in der folgenden Anwendung unendliche Intervalle $[a, \infty)$ ungeeignet.

Die Supremum-Norm verträgt sich sehr gut mit der Integration. Sei für $f, g : [a, b] \to \mathbb{R}^e$

$$F(x); = \int_a^x f, \quad G(x) := \int_a^x g.$$

Dann folgt

$$\begin{array}{rcl} |F(x) - G(x)| & = & |\int_a^x (f-g)(t)dt| \le \int_a^x |f-g|(t)dt \\ & \le & |f-g|_V \cdot |x-a|_{\mathbb{R}} \\ |F-G|_V & \le & |f-g|_V \cdot |b-a|_{\mathbb{R}}. \end{array}$$

Insbesondere: Falls stetige $f_n : [a, b] \to \mathbb{R}^e$ gleichmäßig gegen f_∞ konvergieren, so konvergieren die Stammfunktionen $F_n(x) := \int_a^x f_n$ gleichmäßig gegen die Stammfunktion $F_\infty(x) := \int_a^x f_\infty$. Anders ausgedrückt: $(\lim F_n)' = \lim(F_n')$.

5. Parameterabhängige Integrale

Es kommt häufig vor, daß man, ausgehend von einer Funktion "von zwei Variablen", $f : \mathbb{R} \times [a, b] \to \mathbb{R}$, durch Integration eine neue Funktion definiert.

Definition :

$$F(s) := \int_a^b f(s,t)\ dt.$$

Bei genügend harmlosen Funktionen f kann das Integrationsintervall $[a, b]$ sogar unendlich sein:
Beispiel, die Gammafunktion:

$$\Gamma : \mathbb{R}_+ \to \mathbb{R},\ \Gamma(x) := \int_0^\infty t^{x-1} e^{-t}\ dt.$$

Diese Funktion ist berühmt, weil sie die zunächst nur auf $\mathbb{N}$ definierte Funktion $n \mapsto n!$ auf die reelle Halbgrade ausdehnt — und mit wenig zusätzlicher Mühe kann sie als komplexe Funktion auf der rechten Halbeben Re $z > 0$ durch dieselbe Integralformel definiert werden.

Satz. Es gilt $\Gamma(n+1) = n!,\ n \in \mathbb{N}$.

Beweis. $\Gamma(1) = 1$ ist klar. Wir zeigen $x \cdot \Gamma(x) = \Gamma(x+1)$. Wegen der Produktregel gilt $(t^x \cdot e^{-t})^{\cdot} = x \cdot t^{x-1} \cdot e^{-t} - t^x \cdot e^{-t}$. Damit folgt aus dem Hauptsatz der Differential- und Integralrechnung $0 = \int_0^\infty (t^x \cdot e^{-t})^{\cdot}\ dt = x \cdot \Gamma(x) - \Gamma(x+1),\ (x > 0)$.

Eigenschaften der Integralfunktion F
Ziel ist, aus Voraussetzungen über f Aussagen über F herzuleiten.
Beispiel **Dehnungsschranken** (benutze $|\int h(t)\ dt| \le \int |h(t)|\ dt$) :

$$|f(s_1,t) - f(s_2,t)| \le L(t) \cdot |s_1 - s_2| \Rightarrow |F(s_1) - F(s_2)| \le \left(\int_a^b L(t)\ dt\right) \cdot |s_1 - s_2|.$$

Beispiel **Stetigkeit**: Falls die Stetigkeit von $s \to f(s,t)$ "gleichmäßig" in $t \in [a,b]$ ist, d.h. zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ (ausdrücklich unabhängig von t) so daß gilt

$$|s - s_0| \le \delta \Rightarrow |f(s,t) - f(s_0,t)| \le \epsilon,$$

dann folgt sofort auch

$$|s - s_0| \le \delta \Rightarrow |F(s) - F(s_0)| \le \epsilon \cdot |b - a|.$$

Es braucht nicht einmal ganz so gleichmäßig zu sein: Zu jedem $\epsilon > 0$ gebe es ein $\delta > 0$ so daß mit einer geeigneten Funktion $t \to c(t)$ gilt

$$|s - s_0| \le \delta \Rightarrow |f(s,t) - f(s_0,t)| \le c(t) \cdot \epsilon,$$

dann folgt ebenso

$$|s - s_0| \le \delta \Rightarrow |F(s) - F(s_0)| \le \left(\int_a^b c(t)\ dt\right) \cdot \epsilon.$$

In dieser Formulierung sind dann Anwendungen auf unendliche Integrationsintervalle enthalten, etwa auf die Gammafunktion:

Satz $\quad 2 \le x, y \le n \quad \Rightarrow \quad |\Gamma(x) - \Gamma(y)| \ \le \ (1 + (n+1)!) \cdot |x - y|.$

Beweis. Zunächst folgt aus $\log t \le t$ und damit auch $\log t = -\log(1/t) >= -1/t$
die Abschätzung

$$0 < t,\ 2 \le x \le n \Rightarrow |(x \to t^x \cdot e^{-t})'| = |\log t \cdot t^{x-1} \cdot e^{-t}| \le \max(t^n, t^{n-2}) \cdot e^{-t}.$$

Daher folgt aus dem Schrankensatz $|t^{x-1} \cdot e^{-t} - t^{y-1} \cdot e^{-t}| \le \max(t^n, 1) \cdot e^{-t} \cdot |x - y|$,
und Integration ergibt die Behauptung.

Schließlich gibt es Voraussetzungen, so daß die Ableitung von F "durch Differenzieren von f unter dem Integral" berechnet werden kann. Dazu führen wir eine Bezeichnung ein:
Definition. Für festes t heißt die Ableitung der partiellen Funktion $s \to f(s,t)$ **partielle Ableitung von f nach s**, Bezeichnung:

$$\frac{\partial}{\partial s} f(s,t) := \lim_{s \to s_0} \frac{(f(s,t) - f(s_0,t)}{(s - s_0)}.$$

Satz über das Differenzieren unter dem Integral. Voraussetzung. Die zweite partielle Ableitung sei durch eine stetige Funktion beschränkt: $|\frac{\partial}{\partial s}\frac{\partial}{\partial s} f(s,t)| \le B(t)$. Dann gilt

$$\left| F(s) - F(s_0) - (s - s_0) \cdot \int_a^b \frac{\partial}{\partial s} f(s_0,t)\ dt \right| \le \frac{1}{2} \int_a^b B(t)\ dt \cdot |s - s_0|^2.$$

also $$F'(s_0) = \int_a^b \frac{\partial}{\partial s} f(s_0,t)\ dt$$

Variante. Die partielle Ableitung $(s,t) \to \frac{\partial}{\partial s} f(s,t)$ sei stetig. Dann gilt: Zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ so daß

$$|s - s_0| \le \delta \Rightarrow \Big| F(s) - F(s_0) - (s - s_0) \cdot \int_a^b \frac{\partial}{\partial s} f(s_0, t)\; dt \Big| \le \epsilon \cdot |s - s_0|.$$

Beweis. Die erste Version eignet sich auch für unendliche Integrationsintervalle, sie folgt durch Integration der folgenden Schrankensatzabschätzung (wie immer unter Benutzung von $|\int h(t)\; dt| \le \int |h(t)|\; dt$):

$$|f(s,t) - f(s_0,t) - \frac{\partial}{\partial s} f(s_0,t) \cdot (s - s_0)| \le \frac{1}{2} B(t) \cdot |s - s_0|^2.$$

Die zweite Variante benutzt im Beweis die **gleichmäßige Stetigkeit** von $(s,t) \to \frac{\partial}{\partial s} f(s,t)$; diese steht aber nur auf einem vollständigen ($=$ abgeschlossen in $\mathbb{R}^d$) und beschränkten Definitionsbereich zur Verfügung. In unserm Fall: Zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ (ausdrücklich unabhängig von $t \in [a,b]$ und $s_0 \in [c,d]$) so daß gilt:

$$|s - s_0| \le \delta \;\; \Rightarrow \;\; \Big| \frac{\partial}{\partial s} f(s,t) - \frac{\partial}{\partial s} f(s_0,t) \Big| \;\le\; \epsilon.$$

Dies setzen wir im Hauptsatz der Differential- und Integralrechnung ein und haben zunächst

$$\begin{aligned} |s - s_0| \le \delta \;\; \Rightarrow & \Big| f(s) - f(s_0) - (s - s_0) \cdot \frac{\partial}{\partial s} f(s_0,t) \Big| = \\ & \Big| \int_{s_0}^{s} \big(\frac{\partial}{\partial s} f(s,t) - \frac{\partial}{\partial s} f(s_0,t) \big)\; ds \Big| \le \epsilon \cdot |s - s_0|. \end{aligned}$$

Schließlich wird noch einmal über t integriert:

$$|s - s_0| \le \delta \;\; \Rightarrow \;\; \Big| F(s) - F(s_0) - (s - s_0) \cdot \int_a^b \frac{\partial}{\partial s} f(s_0,t)\; dt \Big| \le |b - a| \epsilon \cdot |s - s_0|.$$

Als Anwendung erhalten wir einen Satz über das Vertauschen von zwei Integrationen.

Satz. Die Funktion $h : \mathbb{R} \times [a,b] \to \mathbb{R},\; (s,t) \mapsto h(s,t)$ sei stetig. Dann gilt:

$$\int_{s_0}^{s_1} \left(\int_a^{t_1} h(s,t)\; dt \right)\; ds = \int_a^{t_1} \left(\int_{s_0}^{s_1} h(s,t)\; ds \right)\; dt.$$

Beweis. Definiere

$$\begin{aligned} & f : \mathbb{R} \times [a,b] \to \mathbb{R},\; f(s,t_1) := \int_a^{t_1} h(s,t)\; dt \\ & F : \mathbb{R} \times [a,b] \to \mathbb{R},\; F(s_1,t_1) := \int_{s_0}^{s_1} f(s,t_1)\; ds = \int_{s_0}^{s_1} \left(\int_a^{t_1} h(s,t)\; dt \right)\; ds. \end{aligned}$$

Nach unserem Satz ist die partielle Funktion $t \to F(s_1,t)$ differenzierbar und

$$\frac{\partial}{\partial t}F(s_1,t)\big|_{t=t_1} = \int_{s_0}^{s_1} \frac{\partial}{\partial t}f(s,t)\big|_{t=t_1}\ ds.$$

Die rechte Seite ist stetig in t_1, also folgt aus dem Hauptsatz der Differential- und Integralrechnung, daß sich F durch Integrationen in umgekehrter Reihenfolge als in der Definition von F berechnen läßt, wie behauptet:

$$\begin{aligned} F(s_1,t_1) =& F(s_1,t_1) - F(s_1,a) = \int_a^{t_1} \frac{\partial}{\partial t}F(s_1,t)\ dt = \\ =& \int_a^{t_1} \left(\int_{s_0}^{s_1} \frac{\partial}{\partial s} f(s,t)\ ds \right)\ dt. \end{aligned}$$

6. Vektorfelder auf $\mathbb{R}^d$ und gewöhnliche Differentialgleichungen.

Mit der Veranschaulichung komplexer Funktionen $f : \mathbb{C} \to \mathbb{C}$ hatten wir zum ersten Mal "Bilder" von Funktionen gemacht, die man sich nicht mehr mit Hilfe ihres Graphen, also der Kurve $x \to (x, f(x))$, vorstellen kann. Schon eine Dimension höher, also für Abbildungen $F : \mathbb{R}^3 \to \mathbb{R}^3$, stößt man wieder an anschauliche Grenzen. Stellen Sie sich ungleichmäßige Windverhältnisse in der Atmosphäre vor. Diese Situation kann durch eine solche dreidimensionale Abbildung F beschrieben werden, indem mit $F(p)$ die Windgeschwindigkeit in p gemeint ist. Dieses F wird man dann so veranschaulichen, daß man in jedem Punkt p den Wert $v = F(p)$ als Pfeil abträgt. Man stellt sich also F als "Vektorfeld" vor. (So wie ein Kornfeld ein Feld voller Halme ist, so ist ein Vektorfeld ein Feld voller Vektoren.) Besonders leicht können Sie sich das Vektorfeld vorstellen, das als Geschwindigkeitsfeld zu einer Rotation gehört. Auch eine Beschreibung mit Formeln ist leicht: Betrachte $\mathbb{R}^3$ mit Standardbasis und Standardskalarprodukt. Dann wird die Rotationsbewegung durch eine Drehmatrix beschrieben, z. B.

$$A(t) = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Bahnkurve eines Punktes p ist dann $c(t) = A(t) \cdot p$, und die Geschwindigkeit dieser Bewegung bei $t = 0$ ist

$$\dot{c}(0) = \dot{A}(0) \cdot p = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot p.$$

Unser Vektorfeld ist also

$$F(p) := \dot{A}(0) \cdot p.$$

Wegen $F(\alpha p+\beta q) = \alpha\cdot F(p)+\beta\cdot F(q)$ nennen wir diese Vektorfelder “lineare Vektorfelder”.

Der Begriff Vektorfeld wird sich in vielfältiger Weise als nützlich erweisen. Zunächst kommt es darauf an, diesen Begriff mit einer möglichst unmittelbaren anschaulichen Vorstellung zu verbinden. Ich finde, das Bild der fließenden Wasserteilchen auf der Oberfläche des Rheins zeigt sehr gut, was beabsichtigt ist, wenn man von ihrem Geschwindigkeitsfeld spricht.

Was sind nun, in der beschriebenen Situation, “natürliche Fragen”? Zum Beispiel möchten die Meteorologen wissen, wie viel Luft transportiert wird, wenn sie die Windgeschwindigkeit, als gemessenes Vektorfeld, kennen. Mathematisch einfacher ist die Frage: Können wir die Bahnkurven einzelner Teilchen rekonstruieren, wenn das Geschwindigkeitsfeld bekannt ist?

Zunächst muß diese Frage in einen Kalkül übersetzt werden. Gegeben sei also ein Vektorfeld $F : \mathbb{R}^d \to \mathbb{R}^d$, mit noch nachzutragenden weiteren Voraussetzungen. Gesucht ist zu jedem Punkt $p \in \mathbb{R}^d$ eine Kurve $c : [0,\infty) \to \mathbb{R}^d$, deren Geschwindigkeit $\dot{c}(t)$ für jedes t gegeben ist durch das Vektorfeld an der Stelle $c(t)$, also

$$\dot{c}(t) = F(c(t)).$$

Unsere Frage nach den Bahnkurven c von Geschwindigkeitsfeldern F ist damit als “Differentialgleichung” präzisiert. Für lineare Geschwindigkeitsfelder $F : \mathbb{R}^d \to \mathbb{R}^d$,

$$F(p) := A \cdot p \qquad\qquad (p \in \mathbb{R}^d,\ A \text{ eine } d \times d\text{–Matrix})$$

kennen wir die Antwort schon, sie muß nur noch deutlicher formuliert werden.

Im Falle $d = 1$, also $A = a \in \mathbb{R}$, lautet die Differentialgleichung

$$\dot{c}(t) = a \cdot c(t).$$

Diese wird durch $c(t) = \exp(a \cdot t) \cdot c(0)$ gelöst, und zwar, wie wir wissen, auf eindeutige Weise.

Für $d > 1$ haben wir schon $d \times d$–Matrizen A in die Exponentialreihe eingesetzt. Wenn wir wüßten, daß $M(t) = \exp(t \cdot A)$ die Ableitung $\dot{M}(t) = A \cdot \exp(t \cdot A)$ hätte, dann wäre jedenfalls eine Lösung der Differentialgleichung:

$$c(t) := \exp(t \cdot A) \cdot c(0)\ ,\ \dot{c}(t) = A \cdot \exp(t \cdot A) \cdot c(0) = F(c(t)).$$

Zur genaueren Behandlung definieren wir eine bequeme Norm, einen Spezialfall der Supremum-Norm, auf dem Vektorraum der Endomorphismen $\mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d)$:

$$\|A\| := \sup\{|A \cdot x|_{\mathbb{R}^d}; |x|_{\mathbb{R}^d} = 1\}.$$

Die Axiome für eine Norm sind erfüllt, für die Dreiecksungleichung beachte

$$|(A+B)\cdot x| \le |A\cdot x| + |B\cdot x| \le (\|A\| + \|B\|)\cdot |x|.$$

Außerdem folgt $\|A\cdot B\| \le \|A\|\cdot\|B\|$, $\|A^k\| \le \|A\|^k$. Der Vektorraum $\mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d)$ mit dieser (oder einer anderen) Norm ist vollständig, wir müssen also zur Definition von $\exp A$ nur die Cauchy-Eigenschaft zeigen für die Folge

$$M_n := \mathrm{id} + \sum_{k=1}^{n} \frac{A^k}{k!}.$$

Diese wird aber durch die konvergente Taylorreihe für exp bei $\|A\|$ majorisiert

$$\Big\| M_{n+p} - M_n \Big\| = \Big\| \sum_{k=n+1}^{n+p} \frac{A^k}{k!} \Big\| \le \sum_{k=n+1}^{n+p} \frac{\|A\|^k}{k!} \le \frac{\|A\|^{n+1}}{(n+1)!}\cdot \exp(\|A\|),$$

so daß $\exp A := \lim M_n$ definiert werden kann. Natürlich ist dann

$$M(t) = \mathrm{id} + \sum_{k=1}^{\infty} \frac{A^k \cdot t^k}{k!}.$$

Um die Ableitung an der Stelle t_0 zu bekommen, setzen wir $|t-t_0| \le 1$ voraus und schätzen ab (Abkürzung $A^0 = \mathrm{id}$):

$$\begin{aligned}
&\|M(t) - M(t_0) - A\cdot M(t_0)\cdot (t-t_0)\| = \\
&\Big\| \sum_{k=1}^{\infty} \frac{A^k\cdot (t^k - t_0^k)}{k!} - A\cdot \sum_{k=1}^{\infty} \frac{A^{k-1}\cdot k\cdot t_0^{k-1}}{k!} \cdot (t-t_0) \Big\| \\
&\le \sum_{k=1}^{\infty} \frac{\|A\|^k}{k!} \big| t^k - t_0^k - k\cdot t_0^{k-1}\cdot (t-t_0) \big| \\
&\le \frac{\|A\|^2}{2} \cdot \left(\sum_{k=2}^{\infty} \frac{\|A\|^{k-2}}{(k-2)!} \cdot (|t_0|+1)^{k-2} \right) \cdot |t-t_0|^2 \\
&\le \frac{\|A\|^2}{2} \cdot \exp(\|A\|\cdot (|t_0|+1)) \cdot |t-t_0|^2 = \mathrm{const}\cdot |t-t_0|^2.
\end{aligned}$$

Diese Ungleichungskette zeigt, daß $M(t)$ mit quadratischem Fehler differenzierbar ist und die Ableitung $A\cdot M(t)$ hat. Benutzt wurde die reelle Exponentialreihe und für $f(t) = t^k$ die Abschätzung $|f(t) - f(t_0) - f'(t_0)\cdot (t-t_0)| \le \frac{|\mathrm{Schranke}(f'')|}{2} \cdot |t-t_0|^2$, also schon häufiger gebrauchte Hilfsmittel.

Wir klären noch die **Eindeutigkeitsfrage**, also: kann es zwei verschiedene Lösungen $c_1(t)$, $c_2(t)$ mit $\dot c_j(t) = A\cdot c_j(t)$, $c_1(0) = c_2(0)$ zu gleichen Anfangsbedingungen geben? Wie immer nutzen wir Linearität aus, wo wir sie sehen: $c := c_1 - c_2$ ist Lösung von $\dot{c} = A \cdot c$ mit Anfangswert $c(0) = 0$, und es genügt zu zeigen $c = 0$. — Definiere die Hilfsfunktion

$$h(t) := \langle c(t), c(t) \rangle \cdot \exp(-2 \cdot \|A\| \cdot t).$$

Offenbar gilt $0 \le h(t)$, $h(0) = 0$. Wir zeigen mit der Schwarzschen Ungleichung $\dot{h}(t) \le 0$, dann folgt $h(t) = 0$ für alle t. In der Tat:

$$\begin{array}{rcl} \dot{h}(t) & = & \big(2\langle c(t), \dot{c}(t)\rangle - 2\|A\| \cdot \langle c(t), c(t)\rangle\big) \cdot \exp(-2\|A\| \cdot t) \\ & \le & (2 \cdot |c(t)| \cdot |A \cdot c(t)| - 2\|A\| \cdot |c(t)|^2) \cdot \exp(-2\|A\| \cdot t) \le 0. \end{array}$$

Zusammenfassung. Für lineare Vektorfelder

$$F : \mathbb{R}^d \longrightarrow \mathbb{R}^d, \quad F(p) = A \cdot p$$

geht durch jeden Punkt $p \in \mathbb{R}^d$ genau eine "Integralkurve c des Vektorfeldes" oder auch "Lösungskurve c der Differentialgleichung"

$$\dot{c}(t) = F(x(t)),\ c(0) = p.$$

Diese Lösung wird durch die Exponentialreihe geliefert:

$$c(t) = \exp(t \cdot A) \cdot c(0).$$

Man sagt hierzu auch: Ein homogenes lineares Differentialgleichungssystem erster Ordnung mit konstanten Koeffizienten lösen.

Differentialgleichungen höherer Ordnung.

Es kommt auch vor, daß nicht ein Geschwindigkeitsfeld sondern z. B. ein Kraftfeld gegeben ist, und daß die gesuchten Kurven eine Differentialgleichung zweiter Ordnung erfüllen sollen:

$$\ddot{c}(t) = F(c(t), \dot{c}(t)).$$

Dies wird durch einen einfachen Trick auf den vorherigen Fall zurückgeführt: Wir betrachten das Problem in $\mathbb{R}^d \times \mathbb{R}^d$, bezeichnen die Punkte darin mit (x, v) und sehen nicht nur die Kurve $c(t)$ sondern das Paar $(c(t), \dot{c}(t))$ als Lösung an und zwar als Integralkurve des Vektorfeldes

$$\begin{pmatrix} F^1 \\ F^2 \end{pmatrix} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d,\ \begin{pmatrix} F^1(x,v) \\ F^2(x,v) \end{pmatrix} := \begin{pmatrix} v \\ F(x,v) \end{pmatrix},$$

oder als Lösung der Differentialgleichung

$$\begin{pmatrix} c(t) \\ v(t) \end{pmatrix}^{\bullet} = \begin{pmatrix} v(t) \\ F(c(t), v(t)) \end{pmatrix}.$$

Wir haben das schon bei sin und cos kennen gelernt, statt $f'' + f = 0$ zu lösen, haben wir

$$\begin{pmatrix} \cos \\ \sin \end{pmatrix}' = \begin{pmatrix} -\sin \\ \cos \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \cos \\ \sin \end{pmatrix}$$

als 2-dimensionale lineare Differentialgleichung erster Ordnung betrachtet. Diese Umformulierung wird eigentlich immer leicht verstanden, so daß diese kurzen Bemerkungen genügen sollen.

7. Satz und Verfahren von Picard-Lindelöf.

Wir wenden uns jetzt allgemeinen Vektorfeldern (statt der bisherigen linearen Felder) zu. Zunächst eine kleine Verallgemeinerung: Das Geschwindigkeitsfeld von Wind in der Atmosphäre ändert sich offenbar im Laufe der Zeit, daher betrachten wir Abbildungen

$$F : \mathbb{R}^d \times \mathbb{R} \longrightarrow \mathbb{R}^d \ , \ (x,t) \longrightarrow F(x,t)$$

die wir für jedes feste t als Vektorfelder $x \to F(x,t)$ interpretieren. F selber nennen wir zeitabhängiges Vektorfeld. Die gesuchten Bahnkurven $t \to c(t)$ haben dann zum Zeitpunkt t die Geschwindigkeit $\dot{c}(t)$, die durch F an der Stelle $(c(t), t)$ gegeben sein soll, also als Differentialgleichung so lautet:

$$\dot{c}(t) = F(c(t), t).$$

Wir machen nun Zusatzvoraussetzungen, die zu einer übersichtlichen Lösungstheorie führen und die in den allermeisten Anwendungen erfüllt sind (oder durch Variationen des folgenden erfüllbar sind).

Voraussetzungen. $F : \mathbb{R}^d \times \mathbb{R} \longrightarrow \mathbb{R}^d$ sei stetig und zusätzlich dehnungsbeschränkt im ersten Argument, also

$$|F(x,t) - F(y,t)| \leq L \cdot |x - y|, \ \text{(“Lipschitz Bedingung”)}.$$

Daß wir damit auf dem richtigen Weg sind, zeigen wir dadurch, daß wir den Eindeutigkeitssatz in dieser neuen Situation fast wie im linearen Fall beweisen können.

Satz (Fehlerfortpflanzung und Eindeutigkeit). $c_j(t)$ seien zwei Lösungen der Differentialgleichung, also

$$\dot{c}_j(t) = F(c_j(t), t) \qquad (j = 1, 2).$$

Dann ist die Funktion

$$h(t) := \langle (c_1 - c_2)(t), (c_1 - c_2)(t) \rangle \cdot e^{-2L \cdot t}$$

nicht wachsend, also

$$|(c_1 - c_2)(t)|^2 \leq |(c_1 - c_2)(0)|^2 \cdot e^{2L \cdot t}.$$

Insbesondere stimmen Lösungen mit gleichen Anfangswerten überein.

Beweis. Wir zeigen $\dot{h} \leq 0$ mit der Schwarzschen Ungleichung und zitieren den Monotoniesatz. Zunächst gilt:

$$\begin{aligned}
&\langle (c_1 - c_2)(t), (c_1 - c_2)(t) \rangle^{\bullet} = \\
&2\langle \dot{c}_1(t) - \dot{c}_2(t), (c_1 - c_2)(t) \rangle \leq \\
&2 \cdot |\dot{c}_1(t) - \dot{c}_2(t)| \cdot |(c_1 - c_2)(t)| = \\
&2 \cdot |F(c_1(t), t) - F_2(c_2(t), t)| \cdot |(c_1 - c_2)(t)| \leq \\
&2 \cdot L \cdot |(c_1 - c_2)(t)|^2.
\end{aligned}$$

Differenziert man nun h nach der Produktregel und setzt die letzte Abschätzung ein, so folgt $\dot{h} \leq 0$.

Wir wenden uns nun dem Existenzbeweis zu. Schon die einfache Differentialgleichung $f' = f$ hat unter den rationalen Funktionen keine Lösung; zu ihrer Lösung mußte eine neue Funktion, nämlich exp, konstruiert werden. Daher sollten Sie erwarten, daß Sie i. a. die Lösung einer Differentialgleichung nicht mit Ihnen schon bekannten Funktionen hinschreiben können. Stattdessen brauchen Sie ein *Verfahren*, daß neue Funktionen konstruiert. Wir werden dazu lauter schon bekannte Hilfsmittel zusammensetzen.

Der Existenzbeweis wird auf der **Vollständigkeit** des Vektorraums der stetigen Kurven in $\mathbb{R}^d$ mit der Supremum-Norm beruhen. Die Lösungskurve wird **Fixpunkt einer kontrahierenden Abbildung** dieses Vektorraums in sich sein. Dieser Fixpunkt kann dann wie im schon bekannten Kontraktionslemma durch Iterieren dieser Abbildung gefunden werden. Die Erfindung dieser Abbildung dürfen Sie nicht von sich erwarten, sie stammt von Picard und Lindelöf.

Als erstes wenden wir den Hauptsatz der Differential- und Integralrechnung auf die Differentialgleichung $\dot{c}(t) = F(c(t), t)$ an:

$$\begin{aligned}
&c(t) - c(0) = \int_0^t F(c(\tau), \tau) d\tau, \\
&c(0) = c_0 \text{ ist der gewünschte Anfangswert.}
\end{aligned}$$

Picard und Lindelöf haben das so angesehen: Zu jeder stetigen Kurve c finden sie eine neue stetige Kurve $\mathcal{L}c$ durch die Definition

$$\mathcal{L}c(t) := c_0 + \int_0^t F(c(\tau), \tau) d\tau.$$

Die Bildkurve hat den Anfangswert $\mathcal{L}c(0) = c_0$, und sie ist sogar differenzierbar mit der Ableitung

$$\mathcal{L}c^{\bullet}(t) = F(c(t), t).$$

Daher ist ein Fixpunkt der Lindelöf-Abbildung, also $c = \mathcal{L}c$, eine Lösung der Differentialgleichfung, offenbar zum Anfangswert c_0.

Zweitens erinnern wir uns daran, daß man Fixpunkte von Abbildungen besonders leicht für kontrahierende Abbildungen findet. Wir wissen schon daß $\mathbb{R}^d$ viele verschiedene Normen hat. Auf dem Vektorraum der stetigen Kurven in $\mathbb{R}^d$ gibt es erst recht viele verschiedene Normen (die nicht einmal alle äquivalent zueinander sind). Ich möchte schon bei diesem ersten Beispiel betonen, daß man sich die Normen zu seinem Problem passend aussuchen darf. Wir wissen von dem Fehlerfortpflanzungssatz, daß Ungenauigkeiten in den Startwerten mit dem Exponentialfaktor $e^{L\cdot t}$ anwachsen können. Wir wissen auch, daß wir die gesuchte Lösung durch Näherungen approximieren wollen. Die Fehlerfortpflanzung zeigt, daß wir bei den Näherungen einen Genauigkeitsverlust mit dem Faktor $e^{L\cdot t}$ erwarten. Daher nehmen wir nicht die erstbeste Supremum-Norm, sondern wir berücksichtigen die größere Ungenauigkeit durch ein kleineres Gewicht:

$$\begin{aligned} V &:= \{c : [0,T] \longrightarrow \mathbb{R}^d \ , \ c \text{ stetig}\} \\ \|c\| &:= \sup_{t\in[0,T]} |c(t)| \cdot e^{-L\cdot t}. \end{aligned}$$

Da der Gewichtsfaktor zwischen $e^{-L\cdot T}$ und 1 liegt, ist diese Norm *äquivalent* zu derjenigen "erstbesten", mit der wir die Vollständigkeit von V mit Hilfe des Hauptsatzes über gleichmäßig konvergente Folgen stetiger Abbildungen gezeigt haben.

Satz. Die Lindelöf-Abbildung $\mathcal{L}$ ist bezüglich dieser Norm kontrahierend. (Das beendet den Existenzbeweis.)

Beweis. Es seien $f, g \in V$. Dann gilt zunächst:

$$\begin{aligned} |\mathcal{L}f(t) - \mathcal{L}g(t)| &= \Big| \int_0^t F(f(\tau),\tau) - F(g(\tau),\tau) d\tau \Big| \\ &\le \int_0^t |F(f(\tau),\tau) - F(g(\tau),\tau)| d\tau \\ &\le \int_0^t L \cdot |f(\tau) - g(\tau)| d\tau. \end{aligned}$$

Daher ist

$$\begin{aligned} |\mathcal{L}f(t) - \mathcal{L}g(t)| \cdot e^{-L\cdot t} &\le e^{-L\cdot t} \int_0^t L \cdot e^{+L\cdot\tau} \cdot e^{-L\cdot\tau} |f(\tau) - g(\tau)| d\tau \\ &\le e^{-L\cdot t} \int_0^t L \cdot e^{L\tau} \cdot \|f - g\| d\tau \\ &\le (1 - e^{-L\cdot T}) \cdot \|f - g\|. \end{aligned}$$

Damit schließlich

$$\|\mathcal{L}f - \mathcal{L}g\| \le \big(1 - e^{-L\cdot T}\big) \cdot \|f - g\|,$$

d. h. $\mathcal{L}$ ist kontrahierend.

Ich hoffe, Sie haben den Überblick behalten. Obwohl ich diesen Beweis seit meinem Studium kenne, bin ich noch immer beeindruckt, mit welcher Eleganz und Leichtigkeit hier neue Funktionen so konstruiert werden, daß sie das gestellte Problem lösen.

Kommentar zur Modellierung mit Differentialgleichungen. Differentialgleichungen werden in vielen Wissenschaften als Werkzeug der Modellierung benutzt. Dazu ist eine weitere, bisher nicht angesprochene Eigenschaft wesentlich. In der Regel kennt man nämlich die "richtige" Differentialgleichung gar nicht beliebig genau (z. B. kennt man nicht alle Steinbrocken, die im Planetensystem herumfliegen). Daher ist es nötig, daß kleine Störungen der Differentialgleichung auch nur zu kleinen Änderungen der Lösungen führen. Da wir den Existenzsatz mit kontrahierenden Abbildungen bewiesen haben, können wir leicht ein Beispiel einer solchen für ein Modellierungswerkzeug notwendigen Stabilitätsaussage hinzufügen.

Satz. Zwei kontrahierende Abbildungen $\mathcal{L}_1, \mathcal{L}_2$, die wenig voneinander verschieden sind, haben auch wenig verschiedene Fixpunkte. Genauere Formulierung:

Gegeben seien zwei kontrahierende Abbildungen $\mathcal{L}_j$, also $|\mathcal{L}_j(x) - \mathcal{L}_j(y)| \le q \cdot |x - y|$, $j = 1, 2$. $\mathcal{L}_1$ und $\mathcal{L}_2$ seien "wenig" verschieden, z. B.

$$|\mathcal{L}_1(x) - \mathcal{L}_2(x)| \le \varepsilon \quad \text{für alle} \quad x \in \mathbb{R}^d.$$

Dann gilt für den Abstand der Fixpunkte $x_j = \mathcal{L}_j(x_j)$ die Abschätzung

$$|x_1 - x_2| \le \frac{\varepsilon}{1 - q}.$$

Beweis.

$$\begin{aligned} |x_1 - x_2| &= |\mathcal{L}_1(x_1) - \mathcal{L}_1(x_2) + \mathcal{L}_1(x_2) - \mathcal{L}_2(x_2)| \\ &\le q \cdot |x_1 - x_2| + \varepsilon, \end{aligned}$$

also $(1 - q) \cdot |x_1 - x_2| \le \varepsilon$, wie behauptet.

Dieses Argument werden wir noch öfter benutzen.

8. Differentation in $\mathbb{R}^d$

Für Abbildungen $f : \mathbb{R}^d \to \mathbb{R}$ können wir keine Differenzenquotienten bilden, da wir nicht durch $p_2 - p_1 \in \mathbb{R}^d$ dividieren können. Die Definition: *Die Ableitung ist Grenzwert von Differenzenquotienten* läßt sich daher nicht verallgemeinern. Wir hatten Differenzierbarkeit von Anfang an auch schon als *gute Approximierbarkeit durch lineare Funktionen* angesehen. Da wir alle linearen Abbildungen $l : \mathbb{R}^d \to \mathbb{R}$ schon kennen, nämlich als den Vektorraum $\mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$, können wir das Approximationskonzept leicht verallgemeinern: *Wir wollen f an einer Stelle $p \in \mathbb{R}^d$ differenziebar nennen, wenn $h \to f(p+h) - f(p)$ genügend gut durch eine lineare Abbildung approximierbar ist.*

Definition: $f : \mathbb{R}^d \to \mathbb{R}$ heißt in p differenzierbar, wenn es eine lineare Abbildung $l : \mathbb{R}^d \to \mathbb{R}$ (genauer l_p) gibt, die so gut approximiert, daß zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert und

$$h \in \mathbb{R}^d,\ |h| \leq \delta \ \Rightarrow\ |f(p+h) - f(p) - l_p(h)| \leq \epsilon \cdot |h|.$$

l_p heißt Ableitung von f bei p.

Folgende Bezeichnungen kommen vor:

$$l = f', l = df, l = Df, l = f*, l = Tf.$$

Zur Zeit ist $l = df$ am verbreitesten, $l = Tf$ hat unter Berücksichtigung aller Bezeichnungskollisionen wohl die besten Zukunftsaussichten. Schreibt man mit Hilfe des verwendeten Skalarprodukts auf $\mathbb{R}^d$ die Ableitung als $l(h) = \langle v, h\rangle$, so heißt v Gradient von f bei p , $v = (\mathrm{grad} f)(p)$. Mit den Konventionen der linearen Algebra ist die Ableitung $l \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$, und für den Gradienten gilt $(\mathrm{grad} f)(p) \in \mathbb{R}^d$. Der Gradient einer differenzierbaren Funktion f ist ein Vektorfeld auf dem Definitionsbereich von f.

Genau wie im eindimensionalen Fall braucht man zum Differenzieren der einfachsten Funktionen keine Grenzwerte. Wähle z.B. $v, w \in \mathbb{R}^d$ und definiere die quadratische Funktion

$$f : \mathbb{R}^d \to \mathbb{R}, \quad f(x) := \langle v, x\rangle \cdot \langle w, x\rangle.$$

Wegen $f(x+h) - f(x) = \langle v, x\rangle \cdot \langle w, h\rangle + \langle v, h\rangle \cdot \langle w, x\rangle + \langle v, h\rangle \cdot \langle w, h\rangle$
haben wird $\mathrm{grad}\, f(x) = \langle v, x\rangle \cdot w + \langle w, x\rangle \cdot v$, denn für alle $h \in \mathbb{R}^d$ gilt:

$$|f(x+h) - f(x) - \langle(\mathrm{grad} f)(x), h\rangle| \leq |v| \cdot |w| \cdot |h|^2.$$

Oder, für $f(x) := \langle x, x\rangle$ ist $(\mathrm{grad} f)(x) = 2 \cdot x$, denn

$$|f(x+h) - f(x) - \langle 2x, h\rangle| = 1 \cdot |h|^2.$$

Oder, mit einer linearen Abbildung $A \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d)$ definiere $f(x) := \langle A \cdot x, x\rangle$. Dann haben wir

$$|f(x+h) - f(x) - \langle A \cdot x, h\rangle - \langle A \cdot h, x\rangle| \leq \| A \| \cdot |h|^2,$$

die Ableitung von f bei x ist also die lineare Abbildung

$$df_x(h) = l(h) := \langle A \cdot x, h\rangle + \langle A \cdot h, x\rangle.$$

Falls A symmetrisch ist, ist $\mathrm{grad} f(x) = 2A \cdot x$.

Differentationsregeln

Zum Differenzieren von Funktionen, die aus einfacheren zusammengesetzt sind, beweist man Differentiationsregeln. Die Beweise sind kaum anders als im 1-dimensionalen Fall. Die Dreiecksungleichung liefert die *Linearität*:

$$d(\alpha f + \beta g) = \alpha \cdot df + \beta \cdot dg$$
$$\operatorname{grad}(\alpha f + \beta g) = \alpha \cdot \operatorname{grad} f + \beta \cdot \operatorname{grad} g.$$

Praktisch unverändert ist der Beweis der *Produktregel*:

$$d(f \cdot g) = g \cdot df + f \cdot dg$$
$$\operatorname{grad}(f \cdot g) = g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g.$$

Die Kettenregel hat zwei Versionen. Falls die äußere Funktion von $\mathbb{R}$ nach $\mathbb{R}$ geht, also $f : \mathbb{R}^d \to \mathbb{R}$ und $h : \mathbb{R} \to \mathbb{R}$, so braucht der 1-dimensionale Beweis kaum geändert zu werden, um zu zeigen:

Kettenregel 1:

$$d(h \circ f)_x = h'(f(x)) \cdot df_x$$
$$\operatorname{grad}(h \circ f)(x) = h'(f(x)) \cdot \operatorname{grad} f(x).$$

Die zweite Version der Kettenregel erfordert mehr Aufmerksamkeit, weil sie eine Verbindung zwischen den 1-dimensionalen Ableitungen aus dem letzten Semester mit der Ableitung mehrdimensionaler Funktionen herstellt. Gegeben seien eine Kurve $c : \mathbb{R} \to \mathbb{R}^d$ und eine Funktion $f := \mathbb{R}^d \to \mathbb{R}$, also $h := f \circ c : \mathbb{R} \to \mathbb{R}$.

Kettenregel 2:

$$h'(t) = df \mid_{c(t)} (c'(t)) = \langle \operatorname{grad} f(c(t)), c'(t) \rangle.$$

Die Ableitung h' der eindimensionalen Funktion h ist wie im letzten Semester zu verstehen. Z.B. sei c eine Gerade, $c(t) = x + t \cdot v$, dann kann die Ableitung von $f \circ c$ wie folgt als Grenzwert eines Differenzenquotienten berechnet werden:

$$(f \circ c)'(0) = \lim_{t \to 0} \frac{f(x + t \cdot v) - f(x)}{t}.$$

Dieser Grenzwert heißt *Richtungsableitung* von f in Richtung v. Dieser Begriff ist von sehr großer praktischer Bedeutung, weil sich oft so viele Richtungsableitungen *ohne Kenntnis von df* direkt ausrechnen lassen, daß nun die mehrdimensionale Ableitung df aus 1-dimensionalen Ableitungen bestimmt werden kann. Vor einem Beweis der Kettenregel 2 sehen wir Beispiele zu diesen direkt berechenbaren Richtungsableitungen an.

Koordinaten und Funktionen auf $\mathbb{R}^d$, partielle Ableitungen

Es sei eine *ON*-Basis $\{e_1, \ldots, e_d\}$ für $\mathbb{R}^d$ gewählt. Wir schreiben Elemente $x \in \mathbb{R}^d$ als Linearkombinationen der Basisvektoren:

$$x = \sum_{k=1}^{d} x_k \cdot e_k, \quad x_k = \langle e_k, x \rangle.$$

Die x_k heißen Koordinaten von x. Beachten Sie, daß die x_k Werte linearer Funktionen sind, nämlich $l_k : x \to \langle e_k, x\rangle$. Es ist sehr weit verbreitet, zwischen den Funktionen l_k und den Werten x_k keinen Unterschied zu machen und von den Koordinatenfunktionen x_k zu sprechen.
Außerordentlich häufig werden nun Funktionen $f : \mathbb{R}^d \to \mathbb{R}$ als explizite Ausdrücke in den Koordinaten gegeben. Man schreibt $f(x) = f(x_1, \ldots, x_d)$ mit demselben Symbol f. Hier kann man sich nun leicht vorstellen, daß man *alle Koordinaten außer einer konstant hält, also die* **1-dimensionale** *Funktion*

$$t \to h(t) := f(x_1, \ldots, x_k + t, \ldots, x_d) = f(x + t \cdot e_k)$$

betrachtet.

Da die Funktion f als expliziter Ausdruck in $x_1, \ldots, x_d$ gegeben sein soll, kann man natürlich $\dot{h}(t)$ ohne Kenntis von df ausrechnen. Man schreibt

$$\dot{h}(0) = \lim_{t\to 0} \big(h(t) - h(0)\big)/t = \frac{\partial f}{\partial x_k}(x_1, \ldots, x_d)$$

und nennt dies "die *partielle Ableitung* von f nach x_k". Und aus diesen partiellen Ableitungen kann man nun die Ableitung df rekonstruieren (Voraussetzung siehe unten). Wenn nämlich f differenzierbar ist und wenn wir die Kettenregel 2 bewiesen haben, dann gilt offenbar

$$df\,|_x \cdot e_k = \frac{\partial f}{\partial x_k}(x).$$

Mit anderen Worten:

> *Die partiellen Ableitungen* $\frac{\partial f}{\partial x_k}$ $(k = 1, \ldots, d)$ *sind die Werte der linearen Abbildung df auf den Basisvektoren* $\{e_1, \ldots, e_d\}$. *D.h. bezüglich dieser Basis hat df die Zeilenmatrix* $(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d})$.

Hierdurch wird die Berechnung d-dimensionaler Ableitungen sehr erleichtert. Und das wird auch durch den folgenden Schönheitsfehler nicht aufgehoben.

Beispiel. Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, definiert als $f(x_1, x_2) := x_1 x_2^3/(x_1^2 + x_2^4)$ für $(x_1, x_2) \neq (0, 0)$, $f(0, 0) = 0$, ist bei $(0, 0)$ stetig definiert wegen $|f(x_1, x_2)| \leq |x_2|/2$. Außerdem existieren alle Richtungsableitungen bei 0, denn längs der Geraden $(x_1, x_2) := t \cdot (\alpha, \beta)$ ist $h(t) = f(t \cdot \alpha, t \cdot \beta) = \alpha\beta^3 t^2/(\alpha^2 + \beta^4 t^2)$, also $\dot{h}(0) = 0$, d.h. die Richtungsableitungen sind alle 0. Trotzdem ist f nicht differenzierbar bei 0, denn die Ableitung kann nichts anderes als 0 sein, aber der Fehlerfaktor in der Differenzierbarkeitsdefinition $r(h) := |f(0 + h_1, 0 + h_2) - f(0, 0) - 0|/|h|$ ist nicht stetig bei $(h_1, h_2) = (0, 0)$ mit Wert 0: Zu $h(t) := (t^2, t)$ ergibt sich $\lim_{t\to 0} r(h(t)) = \frac{1}{2}$.

Das Beispiel zeigt, daß nicht einmal aus der Existenz *aller* Richtungsableitungen von f in 0 die Differenzierbarkeit in 0 folgt. Die Tatsache, daß die partiellen Ableitungen von f die

Matrix (bezüglich der Standardbasis) der Ableitung df liefern, *falls f differenzierbar ist*, bedeutet daher **nicht**, daß die Existenz der partiellen Ableitungen die Differenzierbarkeit zur Folge hat. Wir werden jedoch zeigen können: Falls die partiellen Ableitungen von f bei $a \in \mathbb{R}^d$ existieren und stetig sind, so ist f auch differenzierbar bei a.

Da die Bedeutung der Kettenregel gar nicht genug unterstrichen werden kann, verallgemeinere ich mit ihr den eindimensionalen Schrankensatz:

Schrankensatz. Hat man eine Schranke $L \geq \|df\|$ für die Norm der Ableitung von f zumindest längs der Verbindungsstrecke von p nach $q \in \mathbb{R}^d$, so folgt

$$|f(p) - f(q)| \leq L \cdot |p - q|.$$

Beweis. Parametrisiere die Verbindungsstrecke, d.h. $c(t) := p \cdot (1-t) + q \cdot t$ und definiere $h(t) := f(c(t))$. Dann ist erstens $h(1) - h(0) = f(q) - f(p)$. Vor allem aber folgt aus der Kettenregel:

$$\dot{h} = df\ |_c \cdot \dot{c}, \quad \text{also die Schranke} \quad L \cdot |q - p| \geq |\dot{h}(t)|.$$

Daher folgt aus dem 1-dimensioanlen Schrankensatz die Behauptung:

$$|f(q) - f(p)| = |h(1) - h(0)| \leq \text{Schranke } (|\dot{h}|) \cdot (1 - 0) = L \cdot |q - p|.$$

Beweis der Kettenregel

Tatsächlich ist der Beweis der mehrdimensioanlen Kettenregel nicht schwieriger als für eindimensionale Funktionen. Da die Bedeutung der Kettenregel für die mehrdimensionale Differentialrechnung jedoch erheblich größer ist als im eindimensioanlen Fall, habe ich einige Konsequenzen zuerst besprochen, damit die Einfachheit des Beweises nicht zu einer Unterschätzung der Kettenregel führt. Zunächst besagt die Differenzierbarkeit von c, daß man zu jedem $\epsilon_c > 0$ ein $\delta_c > 0$ finden kann so, daß gilt:

$$|t - t_0| \leq \delta_c \ \Rightarrow \ |c(t) - c(t_0) - c'(t_0) \cdot |t - t_0| \leq \epsilon_c \cdot t - t_0|.$$

Mit der Dreiecksungleichung ergibt das

$$|t - t_0| \leq \delta_c \ \Rightarrow \ |c(t) - c(t_0)| \leq (|c'(t_0)| + \epsilon_c) \cdot |t - t_0|.$$

Daher kann δ_c noch so verkleinert werden, daß man $x = c(t)$, $x_0 = c(t_0)$ in die Differenzierbarkeitsdefinition von f (bei gegebenem c_f) einsetzten kann.

Zu jedem $\epsilon_f > 0$ gibt es ein $\delta_f > 0$ so, daß

$$|x - x_0| \leq \delta_f \Rightarrow |f(x) - f(x_0) - df_{x_0} \cdot (x - x_0)| \leq \epsilon_f \cdot |x - x_0|.$$

Also können wir δ_c so wählen daß:

$$|t - t_0| \leq \delta_c \Rightarrow |c(t) - c(t_0)| \leq \delta_f \Rightarrow$$
$$|f(c(t)) - f(c(t_0)) - df|_{c(t_0)} \cdot (c(t) - c(t_0))| \leq \epsilon_f \cdot (|c'(t_0)| + \epsilon_c) \cdot |t - t_0|.$$

Hierin wird $c(t) - c(t_0) = c'(t_0) \cdot (t - t_0) + Fehler$ eingesetzt und $\|df\| \cdot |Fehler| \le \|df\| \cdot \epsilon_c \cdot |t - t_0|$ mit der Dreiecksungleichung zu den übrigen Fehlern geschlagen. Damit schließlich:

$$|t - t_0| \le \delta_c \Rightarrow |f(c(t)) - f(c(t_0)) - df\,|_{c(t_0)} \cdot c'(t_0) \cdot (t - t_0)| \le \\ \Big(\epsilon_f \cdot (|c'(t_0)| + \epsilon_c) + \epsilon_c \cdot \|df_{c(t_0)}\|\Big) \cdot |t - t_0|.$$

Damit ist die Kettenregel bewiesen. Vergleichen Sie bitte die Ähnlichkeit mit den früheren Beweisen.

9. Ausdehnung auf $F : \mathbb{R}^d \to \mathbb{R}^e$

Wir hatten keine Mühe, die Differentialrechnung von Funktionen $f : \mathbb{R} \to \mathbb{R}$ auszudehnen auf Kurven $c : \mathbb{R} \to \mathbb{R}^d$, weil eine Kurve in $\mathbb{R}^d$ einfach durch d Funktionen gegeben ist. Ebenso einfach ist die Ausdehnung der Differentialrechnung von Funktionen $f : \mathbb{R}^d \to \mathbb{R}$ auf Abbildungen $F : \mathbb{R}^d \to \mathbb{R}^e$, denn eine solche Abbildung wird durch e Funktionen $(F^1, \ldots, F^e) : \mathbb{R}^d \to \mathbb{R}$ gegeben.

Definition. $F : \mathbb{R}^d \to \mathbb{R}^e$ ist differenzierbar bei $a \in \mathbb{R}^d$, wenn es eine lineare Abbildung (genannt Ableitung) $TF_a \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$ gibt, die die Differenz $F(a+h) - F(a)$ genügend gut approximiert: Zu jedem $\epsilon > 0$ gibt es $\delta > 0$ so, daß

$$h \in \mathbb{R}^d, |h| \le \delta \Rightarrow |F(a+h) - F(a) - TF_a \cdot h| \le \epsilon \cdot |h|.$$

Bei den Beweisen der Differentationsregeln sind keine Änderungen nötig, außer daß man beim Beweis von Produktregeln für das betrachtete Produkt $\bullet$ wissen muß:

$$v, w, v \bullet w \in \mathbb{R}^d, \mathbb{R}^e, \mathbb{R}^f \Rightarrow |v \bullet w| \le C|v| \cdot |w|.$$

(Die Konstante C ist in den meisten Fällen 1).

Außerdem hat man wieder die partiellen Ableitungen, die bezüglich der Standardbasen von $\mathbb{R}^d$ und $\mathbb{R}^e$ die Matrix von TF_a liefern:

$$\begin{pmatrix} \frac{\partial F^1}{\partial x_1}(a), & \cdots, & \frac{\partial F^1}{\partial x_d}(a) \\ \vdots & & \vdots \\ \frac{\partial F^e}{\partial x^1}(a), & \cdots, & \frac{\partial F^e}{\partial x_d}(a) \end{pmatrix}$$

Diese Matrix heißt **Jacobi-Matrix** von F bei a. Wir haben wieder die *Warnung*: Die Existenz der partiellen Ableitungen genügt nicht, um die Differenzierbarkeit von F zu garantieren. Umgekehrt, falls TF_a existiert, so ist die Jacobi-Matrix eine bequeme Methode, um TF_a auszurechnen.

Abweichende Konventionen. Im Fall $e = 1$ hatten wir zwei verschiedene Kalküle zur Beschreibung der Ableitung von f, die linearen Abbildungen $df_a \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$ und die Gradienten, also die Vektoren $\mathrm{grad} f(a) \in \mathbb{R}^d$. Diese waren durch die einfache Beziehung

$df_a(h) = \langle \mathrm{grad} f(a), h \rangle$ verbunden. Für $e > 1$ ist es nicht üblich, die Gradienten $\mathrm{grad} F^h(a)$ der Koordinatenfunktionen zu betrachten; die Zeilen der Jacobi-Matrix sind die Matrizen der linearen Abbildungen $dF^k \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$, $k = 1, \ldots, e$.
Eine wirkungsvolle Quelle für Verwirrungen ist die Tatsache, daß die partiellen Ableitungen als Komponenten $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}$ der Zeilenmatrix für $df_a \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$ **bei Verwendung einer Orthonormalbasis** dieselben Zahlen sind wie als Komponenten der Spaltenmatrix für $\mathrm{grad} f(a) \in \mathbb{R}^d$. Erst bei Koordinatenwechseln, z.B. Polarkoordinaten, verhalten sich die Komponenten von df und $\mathrm{grad}\, f$ verschieden.

10. Zwei Sätze ohne eindimensionale Vorbilder, über stetige partielle Ableitungen und über Symmetrien höherer Ableitungen.

Zunächst erledigen wir ein offen gebliebenes Problem durch den

Satz. Die partiellen Ableitungen $\frac{\partial f}{\partial x_j}$, $j = 1, \ldots, d$ seien in einer Umgebung von $a \in \mathbb{R}^d$ *stetig.* Dann ist f in a *differenzierbar.*

Beweis. Es sei $h^* = \sum_{j=1}^d h_j^* \cdot e_j$ und $a_n := a + \sum_{j=1}^n h_j^* \cdot e_j$, $n = 0, \ldots, d$. Zu jedem gegebenem $\epsilon > 0$ wähle (wegen der vorausgesetzten Stetigkeit) δ so, daß

$$|h| \leq \delta \Rightarrow \left| \frac{\partial f}{\partial x_j}(a+h) - \frac{\partial f}{\partial x_j}(a) \right| \leq \epsilon, \; j = 1, \ldots, d.$$

also auch

$$\frac{\partial f}{\partial x_j}(a) - \epsilon \leq \frac{\partial f}{\partial x_j}(a+h) \leq \frac{\partial f}{\partial x_j}(a) + \epsilon.$$

Mit dem Schrankensatz gilt daher für $j = 1, \ldots, n$

$$\left| f(a_j) - f(a_{j-1}) - \frac{\partial f}{\partial x_j}(a) \cdot h_j^* \right| \leq \epsilon \cdot |h_j^*|.$$

Diese Ungleichungen können wir mit Hilfe von

$$\sum_{n=1}^d (f(a_n) - f(a_{n-1})) = f(a + h^*) - f(a)$$

zusammensetzen zu

$$|f(a+h^*) - f(a) - \sum_{j=1}^d \frac{\partial f}{\partial x_j}(a) \cdot h_j^*| \leq \epsilon \cdot \sum_{j=1}^d |h_j^*| \leq \epsilon \cdot \sqrt{d} \cdot |h^*|.$$

Die Fehlerabschätzung zeigt, daß f differenzierbar ist und daß gilt

$$df_a(\sum h_j \cdot e_j) = \sum_{j=1}^{d} \frac{\partial f}{\partial x_j}(a) \cdot h_j.$$

Anwendung. *Wenn die partiellen Ableitungen nicht nur existieren sondern auch stetig sind, dann existiert die Ableitung df und ihre Zeilenmatrix ist durch die partiellen Ableitungen gegeben.*

Höhere Ableitungen

Für Funktionen $f : \mathbb{R} \to \mathbb{R}$ ist die Ableitung wieder eine Funktion $f' : \mathbb{R} \to \mathbb{R}$. Daher ist der Begriff der zweiten Ableitung nicht schwieriger als der ersten. Für Abbildungen $F : \mathbb{R}^d \to \mathbb{R}^e$ ist die Ableitung an der Stelle a eine lineare Abbildung $TF_a \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$; daher ist die Ableitung eine Abbildung $TF : \mathbb{R}^d \to \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$, also ist TF eine Abbildung in einen höherdimensionalen Vektorraum als F es war, $F : \mathbb{R}^d \to \mathbb{R}^e$. Dadurch entstehen zwar keine begrifflichen Probleme beim Weiterdifferenzieren, aber wegen der langen eindimensionalen Erfahrung muß man sich daran gewöhnen, daß TF eine höherdimensionale Abbildung ist als F.
Insbesondere ist die zweite Ableitung an einer Stelle a eine lineare Abbildung

$$T^2F_a : \mathbb{R}^d \to \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e).$$

Dies wird zunächst etwas übersichtlicher gemacht, weil jedes $\Phi \in \mathrm{Hom}(\mathbb{R}^d,\ \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e))$ eine **bilineare Abbildung** $\widetilde{\Phi} \in \mathrm{Bilin}(\mathbb{R}^d, \mathbb{R}^e)$ liefert durch die naheliegende Definition

$$\widetilde{\Phi}(v, w) := \Phi(w)(v) \quad (\text{beachte } \Phi(w) \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)).$$

Umgekehrt ist auch Φ durch $\widetilde{\Phi}$ bestimmt, nämlich durch dieselbe Formel. Man faßt meistens die zweite Ableitung T^2F_a als bilineare Abbildung auf; in den Bezeichnungen wird der Unterschied zwischen Φ und $\widetilde{\Phi}$ selten gemacht. Damit sind wir vorbereitet für einen harmlos aussehenden Satz von außerordentlicher Tragweite:

Satz (H.A. Schwarz) über die Symmetrie der zweiten Ableitung

$F : \mathbb{R}^d \to \mathbb{R}^e$ sei zweimal differenzierbar. Dann ist T^2F symmetrisch, d.h.

$$T^2F(v, w) = T^2F(w, v).$$

Beweis. Wir benutzen erstens die Definition der Ableitung von $TF : \mathbb{R}^d \to \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$ bei $a \in \mathbb{R}^d$: Zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ so, daß

$$v \in \mathbb{R}^d, |v| \le \delta \Rightarrow \|TF_{a+v} - TF_a - T^2F_a \cdot v\| \le \epsilon \cdot |v|.$$

Zweitens, die Kurve $\tau(t) := F(a + h + t \cdot k)$, $h, k \in \mathbb{R}^d$, hat nach der Kettenregel die Ableitung $\dot{\tau}(t) = TF_{a+h+t\cdot k} \cdot k$. Für $|h| + |k| \le \delta$ folgt aus der ersten Ungleichung:

$$|\dot{\tau}(t) - TF_a \cdot k - T^2F_a(h + t \cdot k, k)| \le \epsilon \cdot (|h| + |k|) \cdot |k|.$$

Weil die linke Seite offenbar **Ableitung** von $t \to \tau(t) - TF_a \cdot t \cdot k - T^2F_a(t \cdot h + 0.5t^2 \cdot k, k)$ ist, liefert der Schrankensatz ($0 \le t \le 1$):

$$|F(a + h + k) - F(a + h) - TF_a \cdot k - T^2F_a(h, k) - \frac{1}{2}T^2F_a(k, k)| \le \epsilon \cdot (|h| + |k|) \cdot |k|.$$

Dies gilt insbesondere auch für $h = 0$. Aus der Differenz dieser beiden Ungleichungen fallen die Ableitungsterme, die h nicht enthalten, heraus, und aus der Dreiecksungleichung folgt

$$|F(a + h + k) - F(a + h) - F(a + k) + F(a) - T^2F_a(h, k)| \le \epsilon \cdot (|h| + 2|k|) \cdot |k|.$$

Diese Abschätzung kann für $s \cdot h$, $s \cdot k$ und $0 < s \le 1$ benutzt werden. Nach Division durch s^2 haben wir

$$\left| \frac{F(a + sh + s \cdot k) - F(a + s \cdot h) - F(a + s \cdot k) + F(a)}{s^2} - T^2F_a(h, k) \right| \le \epsilon \cdot (|h| + 2|k|) \cdot |k|.$$

Also ist der Grenzwert für $s \to 0$ der linken Seite 0, d.h. wir haben eine in h und k **symmetrische** Formel für $T^2F_a(h, k)$ erhalten.

Anwendung. Wir zeigen mit einer Folgerung aus dem Satz von Schwarz, daß der mehrdimensionale Hauptsatz der Differential-und Integralrechnung auf Überraschungen stößt. Eindimensional hat **jede** stetige Funktion eine Stammfunktion. Wir fragen jetzt, welche Vektorfelder $F : \mathbb{R}^d \to \mathbb{R}^d$ Ableitungen, also Gradienten, sind. Aus der Annahme $F = \mathrm{grad} f$ folgt für die Ableitung TF, weil sie zweite Ableitung von f ist, aus dem Satz von Schwarz : TF_a ist ein *symmetrischer Endomorphismus* für jedes $a \in \mathbb{R}^d$. Nur Vektorfelder mit dieser sehr speziellen Eigenschaft haben eine Chance, Gradientenfelder zu sein. Unter milden Zusatzvoraussetzungen ist das dann tatsächlich der Fall.

11. Extremwerte I

Wir verallgemeinern mit der Kettenregel eine notwendige Bedingung für Extremwerte aus dem 1-dimensionalen Fall.

Voraussetzung: $f : \mathbb{R}^d \to \mathbb{R}$ sei differenzierbar und f habe in a ein lokales Maximum, d.h. es gibt $\delta > 0$ so, daß für $|x - a| \le \delta$ gilt $f(x) \le f(a)$.

Notwendige Bedingung für eine Extremeum bei a: $\mathrm{grad} f(a) = 0$.

Beweis. Sei $v \in \mathbb{R}^d, |v| = 1$, ein beliebiger Einheitsvektor. Betrachte die Funktion f längs der Strecke $c : [-\delta, \delta] \to \mathbb{R}^d$, $c(t) := a + t \cdot v$. Dann hat $h := f \circ c : [-\delta, \delta] \to \mathbb{R}$ in 0 ein Maximum, d.h. $h'(0) = 0$. Aus der Kettenregel folgt $0 = h'(0) = \langle \mathrm{grad} f(a), \dot{c}(0) \rangle =$

$\langle \text{grad} f(a), v\rangle$. Hieraus folgt $\text{grad} f(a) = 0$. (Es genügt, $0 = \langle \text{grad} f(a), e_k\rangle$ für die Vektoren einer ON-Basis zu wissen, $v = e_1, \ldots, e_d$).

Leider ist dieses Kriterium nicht so häufig wie im eindimensionalen anwendbar. Als wir einen Eigenvektor für einen symmetrischen Endomorphismus $A \in \text{Hom}(\mathbb{R}^d, \mathbb{R}^d)$ gesucht haben, haben wir ein Extremum der quadratischen Funktion $f : x \to \langle A \cdot x, x\rangle$ **nicht** auf $\mathbb{R}^d$ sondern auf der Einheitsspäre in $\mathbb{R}^d$ gesucht. Die Hilfskurven c im obigen Beweis müssen also auf der Einheitsshäre liegen. Erinnerung: a und $v \perp a$ waren Einheitsvektoren, dann ist $c(t) := a \cdot \cos t + v \cdot \sin t$ eine Kurve mit $|c(t)| = 1,\ c(0) = a,\ \dot c(0) = v$. Wie oben folgt:

$$0 = \langle \text{grad} f(a), \dot c(0)\rangle = \langle 2 \cdot A \cdot a, v\rangle.$$

Da v ein beliebiger Einheitsvektor $\perp a$ war, folgt $A \cdot a$ proportional zu a, oder etwas anschaulicher: $\text{grad} f(a)$ muß an der Stelle a senkrecht zur Einheitssphäre sein.

Das an diesem Beispiel geschilderte Verfahren heißt “Extremwertbestimmung unter Nebenbedingungen”. In dieser Form, also “mit Nebenbedingungen”, treten Extremwertfragen im höherdimensionalen Fall in der Regel auf. Wir benötigen die Ergebnisse des nächsten Abschnitts, um die eben angefangenen Ideen zu Ende führen.

12. Der Umkehrsatz

Für lineare Gleichungssysteme hat man eine übersichtliche Lösungstheorie und auch Verfahren, die die Lösungen finden. Im nichtlinearen Fall muß man sich mit Teilresultaten zufrieden geben. Betrachten Sie als Beispiel Polynome vom Grad d mit d gegebenen Nullstellen. Offenbar sind die Koeffizienten $a_0, \ldots, a_{d-1}$ leicht zu berechnende Funktionen der Nullstellen. Kann man auch umgekehrt die Nullstellen als Funktionen der Koeffizienten auffassen? Für das Beispiel $x^2 - a_0 = 0$ geht das in einer Umgebung von $a_0 = 0$ offenbar nicht, aber für *einfache* Nullstellen wird es sich als richtig erweisen. Zumindest in kleinen Umgebungen: die Koeffizienten der Polynome $(x - e^{i\varphi}) \cdot (x + e^{i\varphi})$, $0 \le \varphi \le \pi$, hängen offenbar stetig von φ ab, aber da die "erste" und die "zweite" Nullstelle bei $\varphi = 0$ und bei $\varphi = \pi$ vertauscht worden sind, kann man die Nullstellen nicht als Funktionen der Koeffizienten erwarten, jedenfalls nicht auf dem ganzen Intervall $0 \le \varphi \le \pi$.

Den durch diese einfachen Beispiele aufgezeigten Schwierigkeiten trägt der *Umkehrsatz* Rechnung, tatsächlich behauptet er so viel, wie man auf Grund dieser Beispiele noch hoffen kann. Leider kommt noch ein weiteres Problem hinzu. Wenn eine komplizierte Abbildung $F : \mathbb{R}^d \to \mathbb{R}^d$ gegeben ist, dann kennt man den genauen Wertebereich nicht, d.h. man weiß nicht, ob es zu einem "interessanten" $q \in \mathbb{R}^d$ (= Bildraum) ein $p \in \mathbb{R}^d$ mit $F(p) = q$ gibt. Bei linearen Abbildungen wird dies durch die Bedingung $det \neq 0$ positiv entschieden und zwar schon bevor man die Lösung oder das Urbild p tatsächlich gefunden hat. Bei nichtlinearen Abbildungen (auch: Gleichungssystemen) gibt es kein solches Kriterium und die Frage bleibt unbeantwortet. Der *Umkehrsatz* leistet nur das folgende: *Wenn man ein*

Paar (p,q) mit $F(p) = q$ schon hat, dann gibt es ein anwendbares Kriterium, aus dem folgt: Für genügend kleine w gibt es Urbilder von $q+w$. Außerdem gibt es Verfahren, um ein Urbild $p+v$ mit v nahe 0 zu finden, also $F(p+v) = q+w$. Allerdings sieht der Umkehrsatz nur im Vergleich mit der linearen Theorie schwach aus; die Ausgangsfrage ist im nichtlinearen Fall so viel schwieriger, daß die Aussage des Umkehrsatzes ein so nützliches Ergebnis ist, daß der Umkehrsatz zu den meistzitierten Sätzen gehört (jenseits der linearen Theorie oder eindimensionaler Situationen).

Umkehrsatz. Gegeben sei eine stetig differenzierbare Abbildung $F : \mathbb{R}^d \to \mathbb{R}^d$. Ein Paar (p,q) mit $F(p) = q$ sei bekannt. Ferner sei das aus der linearen Theorie bekannte Kriterium erfüllt: $\det(TF_p) \neq 0$ (d.h. TF_p ist invertierbar). Dann gilt:
Es gibt eine differenzierbare Umkehrabbildung G von einer kleinen Umgebung von q auf eine Umgebung von p, d.h. es gibt $\epsilon > 0$, so daß $|w| < \epsilon \Rightarrow F \circ G(q+w) = q+w$. Für die Ableitung gilt : $TG_{q+w} = (TF_{p+v})^{-1}$, wie die Kettenregel erwarten läßt.

Beweis. Wir wollen zu genügend kleinem w eine kontrahierende Abbildung auf einer Cauchy-vollständigen Umgebung von p finden, deren Fixpunkt Urbild von $q+w$ ist. Um die Bezeichnung zu vereinfachen, betrachten wir die Abbildung $\widetilde{F} : x \to (TF_p)^{-1} \cdot (F(x+p) - q)$. Aus den Voraussetzungen folgt $\widetilde{F}(0) = 0$, $T\widetilde{F}_0 = \mathrm{id}$. Nun wird zu gegebenem w eine Abbildung Φ_w definiert durch:

$$\Phi_w(v) := v - \widetilde{F}(v) + w.$$

Offenbar gilt für einen Fixpunkt $v = \Phi_w(v)$ wie gewünscht $\widetilde{F}(v) = w$ (und damit $F(p+v) = q + TF_p w$). Außerdem ist $T\Phi_w|_{v=0} = 0$, d.h. Φ_w ist wegen der Stetigkeit von $T\widetilde{F}$ sicher nahe 0 kontrahierend (Schrankensatz). Die verbleibende Arbeit des Beweises besteht also darin, eine Umgebung von 0 zu beschreiben, die durch Φ_w kontrahierend in sich abgebildet wird. Da TF stetig ist, gibt es ein $r > 0$ so, daß aus $|v| \leq 2r$ folgt $|T\Phi_{w|_v}| = |T\widetilde{F}_v - T\widetilde{F}_0| \leq \frac{1}{2}$. Damit folgt für jedes w mit $|w| \leq r$:

$$|v| \leq 2r \Rightarrow \; |\Phi_w(v)| \leq |\Phi_w(v) - \Phi_w(0)| + |\Phi_w(0)| \leq \frac{1}{2} \cdot |v| + |w| \leq 2r,$$

mit anderen Worten: Φ_w bildet die abgeschlossene Kugel vom Radius $2r$ $\frac{1}{2}$-kontrahierend in sich ab. Das Kontraktionslemma liefert also einen eindeutig bestimmten Fixpunkt von Φ_w in dieser Kugel. Da der Fixpunkt Urbild von w ist, nennen wir ihn $G(w)$:

$$G(w) = \Phi_w(G(w)), \quad \widetilde{F}(G(w)) = w.$$

Welche Eigenschaften hat G? Für die $\frac{1}{2}$-kontrahierenden Abbildungen Φ_w gilt

$$|\Phi_{w_1}(v) - \Phi_{w_2}(v)| \leq |w_1 - w_2|.$$

Wir hatten schon bei den gewöhnlichen Differentialgleichungen gesehen, daß daraus für die Fixpunkte folgt

$$|G(w_1) - G(w_2)| \leq \frac{w_1 - w_2}{1 - \frac{1}{2}} = 2 \cdot |w_1 - w_2|,$$

die Umkehrabbildung ist also *2-dehnungsbeschränkt*. Daher kann man wie im eindimensionalen Fall die Differenzierbarkeit von G "durch Einsetzen" aus der Differenzierbarkeit von $\widetilde{F}$ folgern: Zu jedem $\widetilde{\epsilon} > 0$ gibt es $\widetilde{\delta} > 0$ so, daß

$$|v_1 - v_2| \leq \widetilde{\delta} \Rightarrow |\widetilde{F}(v_2) - \widetilde{F}(v_1) - T\widetilde{F}_{v_1} \cdot (v_2 - v_1)| \leq \widetilde{\epsilon} \cdot |v_2 - v_1|$$

Jetzt wähle $|w_1|, |w_2| \leq r$, $|w_1 - w_2| \leq \frac{1}{2}\widetilde{\delta}$. Dann ist $|G(w_1) - G(w_2)| \leq \widetilde{\delta}$, wir dürfen also einsetzen:

$$\begin{aligned} |w_1 - w_2| \leq \frac{1}{2}\widetilde{\delta} \ \Rightarrow \ & |\widetilde{F}(G(w_2)) - \widetilde{F}(G(w_1)) - T\widetilde{F}_{v_1} \cdot (G(w_2) - G(w_1))| \\ = |w_2 - w_1 - T\widetilde{F}_{v_1} \cdot (G(w_2) - G(w_1))| & \leq \widetilde{\epsilon} \cdot |G(w_2) - G(w_1)| \leq 2\widetilde{\epsilon} \cdot |w_1 - w_2|. \end{aligned}$$

Da die lineare Abbildung $T\widetilde{F}_{v_1}$ invertierbar ist, gilt schließlich:

$$|w_1 - w_2| \leq \frac{1}{2}\widetilde{\delta} \Rightarrow |(T\widetilde{F}_{v_1})^{-1} \cdot (w_2 - w_1) - G(w_2) + G(w_1)| \leq \widetilde{2\epsilon} \cdot \|T\widetilde{F}_{v_1}^{-1}\| \cdot |w_2 - w_1|.$$

Das ist die Differenzierbarkeit von G mit $TG_{w_1} = (T\widetilde{F}_{G(w_1)})^{-1}$.

Zusammenhang mit Differentialgleichungen.

Wenn man für mehr als ein w ein Urbild berechnen will, kann folgender Zusammenhang mit gewöhnlichen Differentialgleichungen helfen. Z.B. sei im Bild die Kurve $q(t) := q + t \cdot w$, $0 \leq t \leq 1$, gegeben. Für die Urbildkurve gilt:

$$F(p(t)) = q + t \cdot w, \quad TF_{p(t)} \cdot \dot{p}(t) = w.$$

Daher ist die Urbildkurve Lösung der gewöhnlichen Differentialgleichung

$$\dot{p}(t) = (TF_{p(t)})^{-1} \cdot w, \ p(0) = p.$$

(Diese Bemerkung gewinnt an Überzeugungskraft, wenn man schnelle Verfahren zum Lösen von Differentialgleichungen kennt.)

Mehr Variable als Gleichungen, der Satz über implizit definierte Abbildungen.

Wir wollen folgende Situation aus der Linearen Algebra auf den nichtlinearen Fall ausdehnen. Gegeben seien lineare Abbildungen

$$A \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d), \ B \in \mathrm{Hom}(\mathbb{R}^e, \mathbb{R}^d), \det A \neq 0.$$

Gelöst werden soll zu gegebenem $b \in \mathbb{R}^d, u \in \mathbb{R}^e$

$$A \cdot x + B \cdot u = b.$$

Das wird durch $x = A^{-1}(b - B \cdot u)$ erreicht. Wir bekommen also eine Abbildung $\Phi : u \to x = A^{-1}(b - B \cdot u)$, die zu jedem u eine Lösung $x = x(u)$ liefert. In der angestrebten nichtlinearen Situation ist auch das '+' in $A \cdot x + B \cdot u$ nichtlinear. Der Zusammenhang mit dem Umkehrsatz wird dann deutlicher, wenn wir schreiben:

$$\begin{pmatrix} A\,B \\ 0\,\mathrm{id} \end{pmatrix} \cdot \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} b \\ u \end{pmatrix}, \quad \begin{pmatrix} A\,B \\ 0\,\mathrm{id} \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} \cdot B \\ 0 & \mathrm{id} \end{pmatrix}.$$

Schließlich wird die erwünschte Ähnlichkeit mit der linearen Algebra noch dadurch vorbereitet, daß wir den Begriff der *partiellen Ableitung* etwas erweitern:

Gegeben sei eine differenzierbare Abbildung

$$F : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^f, \quad (x, y) \to F(x, y).$$

Für jedes feste $y \in \mathbb{R}^e$ haben wir die partielle Abbildung $F_1(x) := F(x, y)$. Wie definieren:

$$T_1 F_{(x,y)} := TF_1|_x \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^f).$$

Entsprechend bei festem $x : F_2(y) := F(x, y), \;\; T_2 F := TF_2$. Die Differenzierbarkeit von F liefert nun für $a = (x, y) \in \mathbb{R}^d \times \mathbb{R}^e$:

Zu jedem $\epsilon > 0$ gibt es $\delta > 0$ so, daß

$$\begin{aligned} |a| \le \delta \Rightarrow & |F(x + h_1, y) - F(x, y) - T_1 F_{(x,y)} \cdot h_1| \le \epsilon \cdot |h_1| \\ & |F(x, y + h_2) - F(x, y) - T_2 F_{(x,y)} \cdot h_2| \le \epsilon \cdot |h_2| \\ & |F(x + h_1, y + h_2) - F(x, y) - TF_{(x,y)} \cdot h| \le \epsilon \cdot |h| \end{aligned}$$

also mit der Dreiecksungleichung wie erwartet

$$TF_{(x,y)} \cdot h = T_1 F_{(x,y)} \cdot h_1 + T_2 F_{(x,y)} \cdot h_2.$$

Satz über implizit definierte Abbildungen.

Gegeben sei die stetig differenzierbare Abbildung $F : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^d$ und ein Tripel x_0, u_0, b_0 mit $F(x_0, u_0) = b_0$. Analog zur linearen Situation sei die partielle Ableitung $T_1 F_{(x_0,u_0)} \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^d)$ invertierbar — Kriterium: $\det(T_1 F_{(x_0,u_0)}) \neq 0$. Dann gilt:

Es gibt eine differenzierbare Abbildung $\Phi : \mathbb{R}^e \to \mathbb{R}^d$ mit $F(\Phi(u), u) = b_0$. Deren Ableitung erfüllt $T_1 F_{(\Phi(u),u)} \cdot T\Phi_u + T_2 F_{(\Phi(u),u)} = 0$ oder

$$T\Phi_u = -T_1 F_{(\Phi(u),u)})^{-1} \cdot T_2 F_{(\Phi(u),u)}.$$

Beweis. Wir können $x_0 = 0, u_0 = 0, b_0 = 0$ annehmen. Wie angekündigt wird F zu einer Abbildung $\widetilde{F} : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^d \times \mathbb{R}^e$ ergänzt durch die Definition:

$$\widetilde{F}(x,u) = \begin{pmatrix} F(x,u) \\ u \end{pmatrix}.$$

Dann ist

$$T\widetilde{F}_{(0,0)} = \begin{pmatrix} T_1F_{(0,0)} & T_2F_{(0,0)} \\ 0 & \mathrm{id} \end{pmatrix}$$

invertierbar, also der Umkehrsatz anwendbar. Wir finden also $\widetilde{G} : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^d \times \mathbb{R}^e$ (in einer Umgebung von $(0,0)$) mit $\widetilde{F}(\widetilde{G}(y,u)) = (y,u) \in \mathbb{R}^d \times \mathbb{R}^e$. Es folgt $\widetilde{F}(\widetilde{G}(0,u)) = (0,u)$. Wir bezeichnen mit G_d, G_e die beiden Komponentenabbildungen von $\widetilde{G}$:

$$\widetilde{G}(y,u) = (G_d(y,u), G_e(y,u)) \in \mathbb{R}^d \times \mathbb{R}^e.$$

Damit kommen wir zu den Ausgangsbezeichnungen zurück:

$$\widetilde{F}(\widetilde{G}(0,u)) = \begin{pmatrix} F(G_d(0,u), G_e(0,u)) \\ G_e(0,u) \end{pmatrix} = \begin{pmatrix} 0 \\ u \end{pmatrix}.$$

Daher ist $\Phi(u) := G_d(0,u)$ die Abbildung mit den Eigenschaften des Satzes. Die Differenzierbarkeit von $\widetilde{G}$ folgt aus dem Umkehrsatz, daher läßt sich die Ableitung von Φ aus der Kettenregel berechnen, wie im Satz angegeben.

Auch hier läßt sich die Abbildung $\Phi \circ u(t)$ längs gegebener Kurven $u(t)$ in $\mathbb{R}^e$ als Lösung einer *Differentialgleichung* gewinnen. Kürze ab $x(t) = \Phi(u(t))$, also $F(x(t), u(t)) = 0$. Dann erfüllt die Urbildkurve $t \to x(t)$ zu der gegebenen Bildkurve $t \to u(t)$ die Differentialgleichung

$$\dot{x}(t) + \left(T_1F_{(x(t),u(t))}\right)^{-1} \cdot T_2F_{(x(t),u(t))} \cdot \dot{u}(t) = 0.$$

Zusammenfassung. Der Satz über implizit definierte Abbildungen sagt insbesondere, daß zwei Sorten Teilmengen euklidischer Räume in der Regel dieselben sind (die Ausnahmen kommen von der Voraussetzung det $\neq 0$ im Satz). Erstens können wir **Graphen von Abbildungen** $\Phi : \mathbb{R}^e \to \mathbb{R}^d$ betrachten, sogenannte parametrisierte Teilmengen $\{(\Phi(u), u) \in \mathbb{R}^d \times \mathbb{R}^e; u \in \mathbb{R}^e\}$. Zweitens können wir **Lösungsmengen von Gleichungen** betrachten, also $\{(x,u) \in \mathbb{R}^d \times \mathbb{R}^e; F(x,u) = 0 \in \mathbb{R}^d\}$. Ist die erste Beschreibung gegeben und die zweite gesucht, so definiere $F(x,u) := x - \Phi(u)$; ist die zweite gegeben, so können wir in einer Umgebung von Punkten (x_0, u_0) mit $F(x_0, u_0) = 0$ **und** $T_1F_{(x_0,u_0)}$ **invertierbar** mit Hilfe des Satzes über implizit definierte Abbildungen eine parametrisierte Beschreibung gewinnen.

Extremwerte II

Diese Teilmengen treten bei der Extremwertbestimmung unter Nebenbedingungen auf. Falls die Nebenbedingung eine parametrisierte Teilmenge M ist und Extremwerte von $f|_M$ gesucht werden, so braucht man nur $h = f \circ \Phi$ im Definitionsbereich der Parametrisierung zu untersuchen, also grad$h = 0$. Dies ist die bequemere Variante, *sofern man eine Parametrisierung kennt.* Im anderen Fall muß man im Prinzip dasselbe tun, nämlich Kurven c durch den hypothetischen Extremalpunkt a betrachten, Kurven, die die Gleichungen erfüllen:

$$F(x(t), u(t)) = 0 \text{ oder } F^j(x(t), u(t)) = 0,\ j = 1, \dots, d.$$

Die Tangentialvektoren $\dot{c}$ dieser Kurven sind wegen der Kettenregel senkrecht zu den d Vektoren grad$F^j(a)$. Gesucht (aber noch nicht gefunden) ist die Stelle a, an der grad$f(a)$ senkrecht zu all den erlaubten Kurven c ist. Das führt nun zu einer brauchbaren Formulierung: An einer Extremstelle a muß grad$f(a)$ Linearkombination der Gradienten grad$F^j(a)$ sein, also ist a zu bestimmen aus

$$F^j(a) = 0,\ (j = 1, \dots, d) \ \text{ und } \ \mathrm{grad}\, f(a) = \sum_{j=1}^{d} \lambda_j \cdot \mathrm{grad} F^j(a) \in \mathbb{R}^d \times \mathbb{R}^e.$$

Dies sind $2d+e$ Gleichungen für die $(d+e)$ Komponenten $a_1, \dots, a_{d+e}$ von a und für die λ_j, $j = 1, \dots, d$, die sogenannten *Lagrangeschen Multiplikatoren.* Diese Formulierung erlaubt also, ein Gleichungssystem für die Extremstelle a aufzustellen, ohne die Nebenbedingungen zu parametrisieren. Das ist manchmal eine große Vereinfachung.

Beispiel. Es sei $\alpha \in (0, 2)$. Wir suchen das Maximum der Funktion $f(x, y) = x^\alpha \cdot y^{2-\alpha}$ unter der Nebenbedingung $F^1(x, y) : x^2 + y^2 = 1$ im ersten Quadranten $0 \le x, 0 \le y$. Sei m dieses Maximum, dann haben wir damit $x^\alpha \cdot y^{2-\alpha} \le m \cdot (x^2 + y^2)$ gezeigt, eine Verallgemeinerung von $x \cdot y \le \frac{1}{2}(x^2 + y^2)$.

1.) Parametrisiere die Nebenbedingung:

$x = \cos t$, $y = \sin t$, $0 \le t \le \frac{\pi}{2}$ (oder auch $x = x$, $y = \sqrt{1 - x^2}$). Suche das Maximum von $h(t) = f(\cos t, \sin t) = (\cos)^\alpha \cdot (\sin)^{2-\alpha}$. Differenzieren gibt

$$\dot{h}(t) = (\cos)^{\alpha-1} \cdot (\sin t)^{1-\alpha} \cdot (-\alpha \sin^2 t + (2 - \alpha) \cos^2 t)$$
$$\dot{h}(t) = 0 \ \Rightarrow \ \cos^2 t = \frac{\alpha}{2}, \sin^2 t = \frac{2 - \alpha}{2}, \quad m = \max f = \frac{1}{2}\sqrt{\alpha^\alpha \cdot (2 - \alpha)^{2-\alpha}}.$$

2) Berechne grad$F^1 = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$, grad$f = \begin{pmatrix} \alpha \cdot x^{\alpha-1} \cdot y^{2-\alpha} \\ (2 - \alpha) \cdot x^\alpha \cdot y^{1-\alpha} \end{pmatrix}$
und löse die drei Gleichungen für x, y, λ:

$$x^2 + y^2 = 1,\ \mathrm{grad}\, f = \lambda \cdot \mathrm{grad} F^1.$$

Also
$$x^2+y^2=1,\ \lambda\cdot 2x=\alpha\cdot x^{\alpha-1}\cdot y^{2-\alpha}, \lambda\cdot 2y=(2-\alpha)\cdot x^{\alpha}\cdot y^{1-\alpha},$$
oder
$$x^2+y^2=1,\ 2\lambda=\alpha\left(\frac{y}{x}\right)^{2-\alpha}, 2\lambda=(2-\alpha)\left(\frac{x}{y}\right)^{\alpha},$$
oder
$$x^2+y^2=1,\ 1=\frac{\alpha}{2-\alpha}\cdot\left(\frac{y}{x}\right)^2, \text{ also } x^2=\frac{\alpha}{2},\ y^2=\frac{2-\alpha}{2}.$$

3) Warum nimmt f an der berechneten Stelle ein Maximum an, obwohl nur *notwendige* Bedingungen benutzt wurden?
f ist am Rande des Definitionsbereichs 0 und sonst > 0. Als stetige Funktion *besitzt* f ein Maximum, notwendig im Innern des Definitionsbereichs. Dort muß notwendig $\dot{h}(t)=0$ bzw. $\text{grad} f=\lambda\cdot\text{grad} F^1$ sein. Da das aus diesen notwendigen Bedingungen gefolgerte Gleichungssystem nur *eine* Lösung hat, muß an dieser Stelle *das* Maximum vorliegen.

13. Kurvenintegrale

Die höherdimensionale Integralrechnung muß bei einem einjährigen Mathematikkurs leider zu kurz kommen. Einen Schritt in dieser Richtung wollen wir jedoch noch machen. Es sei also eine Funktion $f:\mathbb{R}^d\to\mathbb{R}$ gegeben. Wollen wir auf f mit eindimensionalen Integralen losgehen, so bleibt uns nichts anderes übrig, als f längs Kurven $c:[a,b]\to\mathbb{R}^d$ zu betrachten. Ein erster Versuch wäre, $I:=\int_a^b f(c(t))\,dt$ als Integral von f längs c anzusehen. Erfahrung hat gezeigt, daß ein solches Integral erstaunlich wenig nützt. Zunächst sind kaum andere Definitionsmöglichkeiten zu sehen. Experimentieren wir also etwas mit endlichen Summen, statt fertige Formeln raten zu wollen.
Wenn wir an die Definition der Bogenlänge von c mit Hilfe von Sehnenlängen denken, dann werden wir bei einer "Riemann Summe" von f längs c — unter Benutzung einer Einteilung $\mathfrak{T}$, $a=t_0<t_1<\ldots<t_n=b$ und Zwischenstellen $\tau_j\in[t_{j-1},t_j]$ — wohl zuerst an die Summe
$$\mathfrak{RS}_c(f)=\sum_{j=1}^{n} f(c(\tau_j))\cdot\big(c(t_j)-c(t_{j-1})\big)$$
denken. Diese ist nun **nicht** Riemann Summe für das oben versuchsweise hingeschriebene Integral sondern für
$$I_c(f):=\int_a^b f(c(t))\cdot\dot{c}(t)\ dt. \qquad (*)$$
Auf dem Hintergrund Ihrer Vorkenntnisse ist das zusätzliche Auftreten von $\dot{c}(t)$ im Integranden wenig plausibel, aber es verbessert das Verhalten des Integrals in unglaublicher Weise:

Satz Der Wert des Integrals $I_c(f)$ in $(*)$ ändert sich **nicht**, wenn die Kurve c monoton reparametrisiert wird. Genauer:

Sei $\mu : [\alpha, \beta] \to [a, b]$ stetig differenzierbar und monoton wachsend. Sei ferner $\gamma := c \circ \mu : [\alpha, \beta] \to \mathbb{R}^d$, so gilt (mit Definition $(*)$)

$$I_c(f) = I_\gamma(f).$$

Beweis. Der Beweis folgt aus der Kettenregel. Zunächst sei $H : [a, b] \to \mathbb{R}$ eine Stammfunktion der stetigen Funktion $h(t) := f(c(t)) \cdot \dot{c}(t)$. Dann folgt aus dem Hauptsatz der Differential- und Integralrechnung $I_c(f) = H(b) - H(a)$. Wegen der Kettenregel ist nun $H \circ \mu : [\alpha, \beta] \to \mathbb{R}$ eine Stammfunktion des Integranden von $I_\gamma(f)$, also von

$$\tau \to f(c(\mu(\tau))) \cdot \dot{c}(\mu(\tau)) \cdot \dot{\mu}(\tau),$$

also noch einmal mit dem Hauptsatz

$$I_\gamma(f) = H \circ \mu(\beta) - H \circ \mu(\alpha) = H(b) - H(a) = I_c(f).$$

Diese Überlegungen können wir ohne Änderungen auf Funktionen $f : \mathbb{C} \to \mathbb{C}$ und Kurven $c : [a, b] \to \mathbb{C}$ anwenden. Da wir für viele solcher komplexer Funktionen Stammfunktionen $F : \mathbb{C} \to \mathbb{C},\ F' = f$ kennen, wird das Ergebnis noch erstaunlicher:

Satz Es seien $F, f : \mathbb{C} \to \mathbb{C}$ komplex differenzierbar mit $F' = f$. Ferner sei $c : [a, b] \to \mathbb{C}$ stetig differenzierbar. Dann hängt $I_c(f)$ nur von den Endpunkten von c ab.

Beweis. Nach der (gemischten) Kettenregel ist $F \circ c$ eine Stammfunktion des Integranden $h(t) := f(c(t)) \cdot \dot{c}(t)$ von $I_c(f)$ und daher

$$I_c(f) = F(c(b)) - F(c(a)).$$

Wenn man sich die Integrale, entsprechend ihrer Definition, als Grenzwerte von Riemann Summen vorstellt, dann ist schon der erste Satz, also die Unabhängigkeit von der Kurvenparametrisierung, erstaunlich. Daß solche Integrale auch noch vom Verlauf der Kurve unabhängig sind (bei unveränderten Endpunkten), ist ganz und gar unerwartet und bedarf einer Erklärung.

Wir stellen nun eine Frage, die vor den beiden letzten Sätzen unvernünftig optimistisch geklungen hätte: *"Können wir ein Kurvenintegral von Vektorfeldern so definieren, daß eine Verallgemeinerung des Hauptsatzes der Differential- und Integralrechnung gilt?"* Was soll das genauer heißen? Gegeben sei eine stetig differenzierbare Funktion $H : \mathbb{R}^d \to \mathbb{R}$, deren Ableitung wir als das Vektorfeld $\operatorname{grad} H : \mathbb{R}^d \to \mathbb{R}^d$ ansehen. Können wir die Ableitung so integrieren, daß sich das Integral als Differenz von Werten von H ergibt? Können wir das Vektorfeld $\operatorname{grad} H$ so längs beliebiger stetig differenzierbarer Kurven $c : [a, b] \to \mathbb{R}^d$ integrieren, daß sich als Wert des Integrals $H(c(b)) - H(c(a))$ ergibt, also die Differenz der Werte von H in den Endpunkten der Kurve? Aus den vorgehenden Beweisen ensteht folgender Versuch: Wenn es überhaupt gehen soll, dann nur so

$$H(c(b)) - H(c(a)) = \int_a^b H(c(t))^\bullet \; dt = \int_a^b \langle \operatorname{grad} H(c(t)), \dot{c}(t)\rangle dt.$$

Damit sind wir bei folgender

Definition. Gegeben sei ein stetiges Vektorfeld $X : \mathbb{R}^d \to \mathbb{R}^d$ und eine stetig differenzierbare Kurve $c : [a,b] \to \mathbb{R}^d$. Dann wird das **Kurvenintegral von X längs** c so definiert:

$$\int_c X \;:=\; \int_a^b \langle X(c(t)), \dot{c}(t)\rangle \; dt.$$

Natürlich ist das so gemacht, daß der Hauptsatz in der vorher formulierten Version gilt. Aber, die neue Definition ist für beliebige Vektorfelder nicht nur hinschreibbar sondern es gilt auch der

Satz von der Parametrisierungsunabhängigkeit.
Sei $\mu : [\alpha,\beta] \to [a,b]$ stetig differenzierbar und monoton wachsend. Sei ferner $\gamma := c \circ \mu : [\alpha,\beta] \to \mathbb{R}^d$ eine Umparametrisierung von c, so gilt

$$\int_c X \;=\; \int_\gamma X.$$

Beweis. Ist $H : [a,b] \to \mathbb{R}$ eine Stammfunktion des stetigen Integranden von $\int_c X$, also von $t \mapsto \langle X(c(t)), \dot{c}(t)\rangle$, so ist wegen der Kettenregel (!) $H \circ \mu : [\alpha,\beta] \to \mathbb{R}$ Stammfunktion des Integranden von $\int_\gamma X$, also von $\tau \mapsto \langle X(\gamma(\tau)), \dot{\gamma}(\tau)\rangle$. Daher gilt

$$\int_\gamma X = H \circ \mu(\beta) - H \circ \mu(\alpha) = H(b) - H(a) = \int_c X.$$

Damit haben wir Motivation, Definition und Rechtfertigung des Kurvenintegrals von Vektorfeldern erledigt. Wir wenden uns nun der Frage zu, was der Satz von H.A. Schwarz über die Symmetrie der zweiten Ableitung mit dem verallgemeinerten Hauptsatz der Differential- und Integralrechnung zu tun hat.

Satz über Wegunabhängigkeit von Kurvenintegralen
Auf einem konvexen Gebiet $G \subset \mathbb{R}^d$ sei ein stetig differenzierbares Vektorfeld $X : G \to \mathbb{R}^d$ gegeben. Mit $c : [a,b] \to G$ bezeichnen wir stetig differenzierbare Kurven in G. Die Kurvenintegrale $\int_c X$ hängen genau dann nur von den Endpunkten der Kurven ab, wenn an jeder Stelle $x \in G$ die Ableitung $TX_x \in \operatorname{End}(\mathbb{R}^d)$ sogar ein symmetrischer Endomorphismus ist. (Für die Definition von "symmetrisch" verwenden wir dasselbe Skalarprodukt auf $\mathbb{R}^d$,

Bezeichnung $\langle\ ,\ \rangle$, das auch in der Definition des Kurvenintegrals $\int_c X$ benutzt wird.)
In diesem Fall gibt es eine Funktion $h : G \to \mathbb{R}$ mit $\operatorname{grad} h = X$.

Beweis. Zunächst bestimmen wir, wie stark die Kurvenintegrale von X von den Kurven abhängen. Dazu seien $c_j : [a,b] \to G,\ j = 1,2,\ c_1(a) = c_2(a) = p,\ c_1(b) = c_2(b) = q$ zwei stetig differenzierbare Kurven mit gleichen Endpunkten p, q in G. Wir verbinden sie durch Linearkombination zu einer stetigen Familie

$$C : [0,1] \times [a,b] \to G,\ C(s,t) := (1-s) \cdot c_1(t) + s \cdot c_2(t).$$

Die Kurven der Schar nennen wir $c_s,\ c_s(t) := C(s,t)$. Mit deren Kurvenintegralen definieren wir eine Hilfsfunktion

$$h : [0,1] \to \mathbb{R},\ h(s) := \int_{c_s} X = \int_a^b \langle X(c_s(t)), \dot{c}_s(t)\rangle\ dt.$$

Die Funktion h präzisiert, "wie stark die Kurvenintegrale des Vektorfeldes X von den Wegen abhängen". Mit dem Hauptsatz der Differential- und Integralrechnung schreiben wir $h(1) - h(0) = \int_0^1 h'(s)\ ds$. Es ist ein wichtiger (in Abschnitt 5. bewiesener) Satz über Integrale, daß wir hier h' durch Differenzieren des Integranden berechnen können:

$$\begin{aligned}
h'(s) &= \int_a^b \frac{\partial}{\partial s}\langle X(c_s(t)), \dot{c}_s(t)\rangle\ dt \\
&= \int_a^b \left(\langle TX\big|_{C(s,t)} \cdot \frac{\partial}{\partial s}C(s,t), \frac{\partial}{\partial t}C(s,t)\rangle + \langle X(C(s,t)), \frac{\partial^2}{\partial s \partial t}C(s,t)\rangle\right) dt \\
h(1) - h(0) &= \int_0^1 h'(s)\ ds.
\end{aligned}$$

Ehe wir dies weiterbearbeiten können, brauchen wir eine Hilfsformel, die dadurch entsteht, daß wir dieselben Überlegungen auf zwei Kurvenintegrale anwenden, deren Wert wir kennen. Die beiden Kurven $\sigma_a(s) := C(s,0) = p,\ \sigma_b(s) := C(s,1) = q$ sind nach Definition sogenannte "konstante" Kurven, Kurven mit Ableitung 0. Da sie durch die differenzierbare Kurvenfamilie $\sigma_t,\ \sigma_t(s) := C(s,t)$ verbunden sind, haben wir

$$\begin{aligned}
0 =& 0 - 0 = \int_{\sigma_b} X - \int_{\sigma_a} X = \int_a^b \frac{\partial}{\partial t}\left(\int_{\sigma_t} X\right) dt = \\
&\int_a^b \frac{\partial}{\partial t}\left(\int_0^1 \langle X(C(s,t)), \frac{\partial}{\partial s}C(s,t)\rangle\ ds\right) dt.
\end{aligned}$$

Genau wie eben kann unter dem Integral nach der Produktregel differenziert werden:

$$0 = \int_a^b \left(\int_0^1 \left(\langle TX\big|_{C(s,t)} \cdot \frac{\partial}{\partial t}C(s,t), \frac{\partial}{\partial s}C(s,t)\rangle + \langle X(C(s,t)), \frac{\partial^2}{\partial t \partial s}C(s,t)\rangle\right) ds\right) dt.$$

Dies wollen wir von dem vorher bestimmten Ausdruck für $h(1) - h(0)$ abziehen. Dazu zitieren wir: Erstens kann nach Abschnitt 5. die Reihenfolge der beiden Integrationen vertauscht werden. Zweitens sind nach dem Satz von H. A. Schwarz die zweiten partiellen Ableitungen von $C(s,t)$ symmetrisch, oder unabhängig von der Differentiationsreihenfolge. Daher heben sich die Terme, die zweite Ableitungen von $C(s,t)$ enthalten, beim Subtrahieren weg. Übrig bleibt als Integrand

$$\langle TX\big|_{C(s,t)} \cdot \frac{\partial}{\partial s}C(s,t), \frac{\partial}{\partial t}C(s,t)\rangle \; - \; \langle TX\big|_{C(s,t)} \cdot \frac{\partial}{\partial t}C(s,t), \frac{\partial}{\partial s}C(s,t)\rangle.$$

Diese Differenz ist 0, weil wir die Endomorphismen $TX_p :\in \mathrm{End}(\mathbb{R}^d)$ als symmetrisch vorausgesetzt hatten. Damit ist

$$\int_{c_0} X \;=\; \int_{c_1} X$$

gezeigt. Darüber hinaus haben wir sogar eine Formel erhalten, die für Vektorfelder X, für die die TX_p **nicht** symmetrisch sind, ausrechnet, wie die Wegintegrale von X von den Wegen (zwischen festen Endpunkten) abhängen. *Diese Formel öffnet den Weg zum höherdimensionalen Hauptsatz der Differential- und Integralrechnung.* Zunächst zeigt sie sofort: Wenn $p \to TX$ stetig ist und ein TX_p nicht symmetrisch, dann gibt es zunächst eine kleine Umgebung U von p in der $p \to TX_p$ beinahe konstant ist; als nächstes kann man in dieser Umgebung Wege mit gleichen Endpunkten wählen, so daß die Kurvenintegrale von X längs dieser Wege tatsächlich verschieden sind. Ein Blick auf die Behauptungen des Satzes zeigt, daß wir alles gezeigt haben bis auf die Konstruktion einer Stammfunktion h von X.

Wenn wir h mit $X = \mathrm{grad}\, h$ schon hätten, dann würde für **jeden Weg** $c : [0,1] \to G$ mit $c(0) = p,\ c(1) = x \in G$ wegen Kettenregel (für $h \circ c$) und Hauptsatz gelten:

$$h(x) = h(p) + \int_0^1 \langle X(c(t)), \dot{c}(t)\rangle \; dt = h(p) + \int_c X. \qquad (*)$$

Daher nehmen wir $(*)$ als Definition von h. Um für diese Funktion h zu zeigen $\mathrm{grad}\, h = X$, betrachte eine beliebige Kurve $\gamma : [1-\delta, 1+\delta] \to G$ mit $\gamma(1) = x$. Zur Definition von $h(x)$ können wir eine **beliebige** Kurve c von p nach x wählen. Wir machen es so, daß γ das Endstück von c ist. Nun zitieren wir noch einmal die Unabhängigkeit der Kurvenintegrale von der Parametrisierung der Kurve, das zeigt

$$h(c(t)) = h(p) + \int_{c|_{[0,t]}} X = h(p) + \int_0^t \langle X(c(t)), \dot{c}(t)\rangle \; dt,$$

und daher mit dem Hauptsatz der Differential- und Integralrechnung schließlich für beliebige Vektoren $\dot{c}(1)$:

$$(h \circ c)^{\bullet}(1) = \langle X(c(1)), \dot{c}(1)\rangle$$

also
$$\mathrm{grad}\, h(x) = X(x).$$