HUMBOLDT-UNIVERSITÄT ZU BERLIN
Institut für Mathematik
C. Wendl

**Analysis II***
**SoSe 2019**

# Der Mittelwertsatz und die Taylorformel mit Integralrestglied

Dieses Skript betrifft Inhalte der Vorlesung vom 25.6.2019, die zum größten Teil als Standardthemen gelten aber nicht im Skript von Helga Baum vorkommen.

Im Folgenden sei $\mathcal{U} \subset \mathbb{R}^n$ eine offene Teilmenge, $E$ ein Banachraum, $f : \mathcal{U} \to E$ eine Funktion, und $\mathbf{a}$, $\mathbf{x} = \mathbf{a} + \mathbf{h}$ Punkte in $\mathcal{U}$ mit der Eigenschaft, dass die Gerade zwischen $\mathbf{a}$ und $\mathbf{x}$ auch in $\mathcal{U}$ liegt, d.h.

$$\mathbf{a} + t\mathbf{h} \in \mathcal{U} \quad \text{für alle } t \in [0,1].$$

## Der Mittelwertsatz als Integral

Wenn $f$ in $C^1(\mathcal{U}, E)$ ist, dann ist $Df : \mathcal{U} \to \mathscr{L}(\mathbb{R}^n, E)$ stetig, also folgt vom Hauptsatz der Differential- und Integralrechnung durch eine Anwendung der Kettenregel die Relation

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{a}) = f(\mathbf{a} + t\mathbf{h})\big|_{t=0}^{t=1} &= \int_0^1 \frac{d}{dt} f(\mathbf{a} + t\mathbf{h})\, dt = \int_0^1 Df(\mathbf{a} + t\mathbf{h})\mathbf{h}\, dt \\ &= \left( \int_0^1 Df(\mathbf{a} + t\mathbf{h})\, dt \right) \mathbf{h}. \end{aligned} \tag{1}$$

In der zweiten Zeile wird die Funktion $[0,1] \to \mathscr{L}(\mathbb{R}^n, E) : t \mapsto Df(\mathbf{a} + t\mathbf{h})$ integriert, das Ergebnis liegt also in $\mathscr{L}(\mathbb{R}^n, E)$ und wird dann auf $\mathbf{h} \in \mathbb{R}^n$ ausgewertet, um ein Element von $E$ zu erzeugen. Für den Übergang von der ersten Zeile zur Zweiten haben wir ein Lemma angewendet, das als Übungsaufgabe 10.B(b) zu beweisen war:

**Lemma 1** (Aufgabe 10.B(b))**.** *Sei $\mathbf{L} : E \to F$ eine stetige lineare Abbildung zwischen Banachräumen und $f : [a,b] \to E$ eine Riemann-integrierbare Funktion. Dann ist $\mathbf{L} \circ f : [a,b] \to F$ auch Riemann-integrierbar, und es gilt:*

$$\int_a^b \mathbf{L}(f(x))\, dx = \mathbf{L}\left( \int_a^b f(x)\, dx \right).$$

□

In (1) wird das Lemma mit der linearen Abbildung $\mathbf{L} : \mathscr{L}(\mathbb{R}^n, E) \to E : \mathbf{A} \mapsto \mathbf{A}\mathbf{h}$ angewendet. Dies ist stetig, denn von der Definition der Norm auf $\mathscr{L}(\mathbb{R}^n, E)$ folgt

$$\|\mathbf{A}\mathbf{h}\| \le \|\mathbf{A}\| \cdot \|\mathbf{h}\|,$$

also für eine konvergente Folge $\mathbf{A}_k \to \mathbf{A}$ in $\mathscr{L}(\mathbb{R}^n, E)$ gilt

$$\|\mathbf{A}\mathbf{h} - \mathbf{A}_k\mathbf{h}\| \le \|\mathbf{A} - \mathbf{A}_k\| \cdot \|\mathbf{h}\| \to 0, \qquad \Rightarrow \qquad \mathbf{L}(\mathbf{A}_k) \to \mathbf{L}(\mathbf{A}).$$

Wir haben gerade die folgende Aussage bewiesen:

**Satz 1.** *Sei $E$ ein Banachraum und $\mathcal{U} \subset \mathbb{R}^n$ eine offene Teilmenge. Dann gilt für jede Funktion $f \in C^1(\mathcal{U}, E)$ und Punkte $\mathbf{a}, \mathbf{x} = \mathbf{a} + \mathbf{h}$ mit $\mathbf{a} + t\mathbf{h} \in \mathcal{U}$ für alle $t \in [0, 1]$,*

$$f(\mathbf{x}) - f(\mathbf{a}) = \left( \int_0^1 Df(\mathbf{a} + t\mathbf{h})\, dt \right) (\mathbf{x} - \mathbf{a}).$$

□

Die folgende Konsequenz ist durch das Lemma $\left\| \int_a^b f(x)\, dx \right\| \le |b-a| \cdot \max_{x\in[a,b]} \|f(x)\|$ leicht zu beweisen:

**Korollar 1.** *Gilt $\|Df(\mathbf{y})\| \le M$ für alle $\mathbf{y} \in \mathcal{U}$ in der Situation von Satz 1, dann folgt $\|f(\mathbf{x}) - f(\mathbf{a})\| \le M\|\mathbf{x} - \mathbf{a}\|$.* □

Dieses Korollar ist eng verwandt mit einem Resultat, das wir früher in dieser Vorlesung als "Mittelwertsatz für vektorwertige Funktionen" (Satz 5.13 im Skript von Helga Baum) bezeichnet haben. Das frühere Resultat galt für Funktionen $f : [a, b] \to E$ unter etwas lockereren Voraussetzungen, z.B. $E$ durfte ein beliebiger normierter Vektorraum (aber nicht zwingend ein Banachraum) sein, und $f$ musste auf $(a, b)$ differenzierbar aber nicht unbedingt *stetig* differenzierbar sein. In der Praxis sind aber die Anwendungsbereiche für diese zwei Resultate gleich: in allen wichtigen Situationen, wo wir den früheren Mittelwertsatz für vektorwertige Funktionen angewendet haben, könnte man genau so gut Korollar 1 anwenden.

Satz 1 stellt ebenfalls eine etwas gehobene Form des gewöhnlichen Mittelwertsatzes dar. Der gewöhnlicher Mittelwertsatz betrifft stetige Funktionen $f : [a, b] \to \mathbb{R}$, die auf dem Inneren $(a, b)$ differenzierbar sind, und garantiert die Existenz eines Punktes $\xi \in (a, b)$ mit

$$f(b) - f(a) = (b - a) \cdot f'(\xi). \tag{2}$$

Dieses Resultat hat im Vergleich mit Satz 1 den Vorteil, dass $f'$ nicht stetig sein muss, d.h. $f$ muss auf $(a, b)$ differenzierbar sein, aber nicht zwingend *stetig* differenzierbar. Aber dazu gibt es zwei deutliche Nachteile: erstens, wir wissen nicht genau, wo $\xi$ ist, sondern nur, dass $\xi$ existiert—insb. wissen wir nicht, genau wie sich $\xi$ ändern wird, wenn wir $a$ oder $b$ ändern. Zweitens, der Satz gilt wirklich nur für *reellwertige* Funktionen, z.B. haben wir schon mal gesehen, dass die analoge Aussage für Funktionen $f : [a, b] \to \mathbb{R}^m$ falsch ist, geschweige denn Funktionen mit Werten in beliebigen Banachräumen. Durch Satz 1 werden diese zwei Nachteile aufgehoben, solange wir bereit sind, nur Funktionen mit stetigen Ableitungen (s. Bemerkung 1 unten) zu betrachten: dann wird (2) mit

$$f(b) - f(a) = (b - a) \cdot \left( \int_0^1 f'(a + t(b - a))\, dt \right)$$

ersetzt, wobei auf der rechten Seite nichts Unbekanntes steht, und die Funktion auch Werte in einem beliebigen Banachraum haben darf. Anstelle vom Wert der Ableitung in einem unbekannten Punkt haben wir jetzt den *Durchschnittswert* der Ableitung auf dem ganzen Interval.

*Bemerkung* 1. Der Grund, warum Satz 1 eine stetige Ableitung als Voraussetzung braucht, ist dass er wesentlich vom Hauptsatz der Differential- und Integralrechnung abhängt. Insb. besteht der Hauptschritt im Beweis darin, für die stetig differenzierbare Funktion $g : [0, 1] \to E$ definiert durch $g(t) := f(\mathbf{a} + t\mathbf{h})$, die Relation

$$g(1) - g(0) = \int_0^1 g'(t)\, dt \tag{3}$$

hinzuschreiben. Der Hauptsatz sagt: wenn $g'$ stetig ist, dann ist $h(x) := \int_0^x g'(t)\,dt$ eine Stammfunktion von $g'$, also gilt $h = g$ bis auf Addition einer Konstante, und (3) folgt. Aber wenn $Df$ und deswegen auch $g'$ nicht stetig ist, dann kann es passieren, dass $h$ keine Stammfunktion von $g'$ ist, und dann ist theoretisch möglich, dass (3) nicht stimmt. In der Vorlesung haben wir (mit dem Mittelwertsatz) im Fall $E = \mathbb{R}^m$ bewiesen, dass das nicht passiert, falls wir irgendwie auch wissen, dass $g'$ Riemann-integrierbar ist. Der Gebrauch des Mittelwertsatzes hat nun eine Beschränkung der Allgemeinheit verursacht, aber das ist nicht das größte Problem: wenn $g'$ nicht stetig ist, dann kann nicht garantiert werden, dass sie überhaupt Riemann-integrierbar ist,[1] und dann gibt es keine Hoffnung mehr für Satz 1. Fazit: das Leben ist viel schöner, wenn man unstetige Ableitungen einfach meidet!

In dieser Vorlesung haben wir den Mittelwertsatz in Beweisen wichtiger Resultate oft angewendet, immer mit der etwas heiklen Konsequenz, dass der Beweis nur für reellwertige oder höchstens für $\mathbb{R}^m$-wertige Funktionen gültig war. Wichtige Beispiele solcher Resultate sind:

- Eine Funktion ist genau dann von Klasse $C^1$, wenn all ihre partiellen Ableitungen existieren und stetig sind (Satz 6.9 im Skript von Baum).
- Für Funktionen der Klasse $C^2$ gilt $\partial_i \partial_j f = \partial_j \partial_i f$ (Lemma von Schwarz, Satz 6.11 im Skript von Baum).
- Der Satz von Taylor.

Für jedes Resultat dieser Art kann jetzt der Mittelwertsatz mit Satz 1 ersetzt werden, um die Gültigkeit des Resultats zu verallgemeinern. Diese Idee wird für den Satz von Taylor unten im dritten Abschnitt dieses Skripts realisiert, und für die ersten zwei Beispiele im letzten Abschnitt.

## Anwendung: Funktionen, die in einem Punkt verschwinden

Die folgende Anwendung zeigt, dass Satz 1 auch für reellwertige Funktionen einige Vorteile im Vergleich mit dem gewöhnlichen Mittelwertsatz hat.

**Proposition 1.** *Sei $f : \mathbb{R} \to \mathbb{R}$ eine glatte Funktion mit $f(0) = 0$, und definiere*

$$g(x) := \begin{cases} \dfrac{f(x)}{x} & \text{für } x \neq 0, \\ f'(0) & \text{für } x = 0. \end{cases}$$

*Dann ist die Funktion $g : \mathbb{R} \to \mathbb{R}$ auch glatt.*

Dass $g$ auf dem Bereich $\mathbb{R} \setminus \{0\}$ unendlich oft differenzierbar ist, kann man sofort sehen. Die Stetigkeit von $g$ in $x = 0$ ist auch klar als Konsequenz der Definition der Ableitung von $f$. Man könnte dann direkt zeigen, dass $g$ in $x = 0$ auch differenzierbar ist, in dem man die Existenz des Grenzwerts

$$\lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(h)/h - f'(0)}{h} = \lim_{h \to 0} \frac{f(h) - hf'(0)}{h^2}$$

beweist, aber schon diese Existenz ist nicht ganz offensichtlich, und man hat ein erheblich größeres Problem,[2] wenn man, um die Existenz von $g^{(k)}(0)$ für jedes $k \in \mathbb{N}$ zu beweisen,

[1] Hier ein Gegenbeispiel: `https://en.wikipedia.org/wiki/Volterra%27s_function`

[2] Das Problem, auf das ich hier hinweise, ist die Sterblichkeit.

dieses Verfahren unendlich oft wiederholen möchte. Die Rettung kommt von Satz 1.

*Beweis von Proposition 1.* Da $f(0) = 0$ impliziert Satz 1 die Formel

$$f(x) = f(x) - f(0) = \left( \int_0^1 f'(tx)\, dt \right) x,$$

also für alle $x \neq 0$ gilt

$$g(x) = \int_0^1 f'(tx)\, dt. \tag{4}$$

Eigentlich ist diese Formel auch für $x = 0$ richtig, denn die rechte Seite wird dann $\int_0^1 f'(0)\, dt = f'(0) = g(0)$. Die Gleichung (4) identifiziert $g$ mit einem sogenannten *parameterabhängigen Integral*

$$g(x) = \int_0^1 F(t, x)\, dt,$$

wobei wir $F : [0, 1] \times \mathbb{R} \to \mathbb{R}$ durch $F(t, x) := f'(tx)$ definiert haben. Die Funktion $F$ ist glatt, denn sie ist die Verknüpfung der polynomiellen (und daher glatten) Funktion $[0, 1] \times \mathbb{R} \to \mathbb{R} : (t, x) \mapsto tx$ mit $f' : \mathbb{R} \to \mathbb{R}$, die glatt ist, weil $f$ glatt ist. Laut Satz 7.19 im Skript von Helga Baum ist $g$ daher stetig und differenzierbar, mit Ableitung

$$g'(x) = \int_0^1 \partial_x F(t, x)\, dt = \int_0^1 t f''(tx)\, dt.$$

Da $\partial_x F : [0, 1] \times \mathbb{R} \to \mathbb{R}$ auch stetig differenzierbar ist, kann der Satz nun wieder angewendet werden, und impliziert, dass $g'$ auch stetig und differenzierbar ist. Dieses Verfahren lässt sich als Induktionsargument formulieren und führt zur Konsequenz, dass $g$ unendlich oft differenzierbar ist, mit

$$g^{(k)}(x) = \int_0^1 \frac{\partial^k}{\partial x^k} F(t, x)\, dt$$

für alle $k \in \mathbb{N}$. $\square$

*Bemerkung* 2. Der gewöhnliche Mittelwertsatz wäre im obigen Beweis nutzlos gewesen. Er hätte uns eine Formel

$$f(x) = f(x) - f(0) = f'(\xi) x$$

gegeben, wobei $\xi$ ein Punkt zwischen 0 und $x$ ist, den wir nicht genauer bestimmen können. Wir könnten dann $g(x) = f'(\xi)$ schreiben, aber der Punkt $\xi$ hängt von $x$ ab, und wir haben keine Informationen darüber, *wie* $\xi$ von $x$ abhängt, können also nicht sagen, ob diese Abhängigkeit stetig/differenzierbar/glatt ist oder nicht.

## Der Satz von Taylor

Wir haben die Taylorformel bisher nur für reellwertige Funktionen betrachtet, hauptsächlich weil wir für Beweise der Restgliedformeln auf den Mittelwertsatz angewiesen waren. Aber zumindest die Taylorpolynome sind auch für Funktionen mit Werten in beliebigen Vektorräume sinnvoll. Im Folgenden gelten die üblichen Annahmen für die Punkte $\mathbf{a}, \mathbf{x} = \mathbf{a} + \mathbf{h} \in \mathcal{U}$, und wir bezeichnen mit

$$P_k(\mathbf{x}) = \sum_{m=0}^{k} Q_m(\mathbf{x})$$

das $k$-te Taylorpolynom von $f : \mathcal{U} \to E$ um den Entwicklungspunkt $\mathbf{a}$, wobei $Q_m(\mathbf{x})$ der Teil davon bezeichnet, der ein homogenes Polynom vom Grad $m$ in den $n$ Variablen $(h_1, \dots, h_n) = \mathbf{h} = \mathbf{x} - \mathbf{a}$ ist. Per Definition ist $P_k(\mathbf{x})$ das eindeutige Polynom von Grad höchstens $k$ in den Variablen $h_1, \dots, h_n$ mit Koeffizienten in $E$, die eine Funktion $\mathcal{U} \to E$ definiert, die im Punkt $\mathbf{a}$ die gleichen höheren partiellen Ableitungen wie $f$ bis Ordnung $k$ hat. Konkret haben wir zwei explizite Formeln für $Q_m(\mathbf{x})$ gesehen, nämlich

$$Q_m(\mathbf{x}) = \sum_{|\alpha|=m} \mathbf{h}^\alpha \frac{\partial^\alpha f(\mathbf{a})}{\alpha!} = \frac{1}{m!} D^m f(\mathbf{a})(\mathbf{h}, \dots, \mathbf{h}).$$

Die Gleichheit dieser zwei Formeln wurde in Aufgabe 7.Z bewiesen. Das Restglied $R_k : \mathcal{U} \to E$ ist definiert durch die Relation

$$f(\mathbf{x}) = P_k(\mathbf{x}) + R_k(\mathbf{x}).$$

Jede Version des Satzes von Taylor kann im Prinzip als eine unter bestimmten Voraussetzungen gültige Formel für das Restglied aufgefasst werden. Die für uns schon bekannte Version heißt: falls $f \in C^{k+1}(\mathcal{U}, \mathbb{R})$ und die Punkte $\mathbf{a} + t\mathbf{h}$ alle in $\mathcal{U}$ liegen, dann existiert ein Punkt $\boldsymbol{\xi} = \mathbf{a} + \theta\mathbf{h}$ auf der Gerade zwischen $\mathbf{a}$ und $\mathbf{x}$, so dass

$$R_k(\mathbf{x}) = \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(\boldsymbol{\xi})}{\alpha!} h^\alpha = \frac{1}{(k+1)!} D^{k+1} f(\boldsymbol{\xi})(\mathbf{h}, \dots, \mathbf{h}), \tag{5}$$

wobei die Gleichheit dieser zwei Formen wieder von Aufgabe 7.Z kommt. Im Fall $n = 1$ und $k = 0$ haben wir also

$$f(x) = f(a) + f'(\xi)h$$

für $x = a + h$ und ein Punkt $\xi$ zwischen $a$ und $x$. Dies ist wieder der Mittelwertsatz, und er stimmt nicht für vektor- oder komplexwertige Funktionen, z.B. erfüllt die komplexwertige Funktion $f(t) := e^{2\pi i t}$ die Relation $f(1) - f(0) = 0$, obwohl $f'(\xi) = 2\pi i e^{2\pi i \xi}$ nie verschwindet.

Vielleicht können Sie die Lösung zu diesem Problem inzwischen ahnen: die Verallgemeinerung der Taylorformel für vektorwertige Funktionen soll den Mittelwertsatz mit Satz 1 ersetzen. Hier ist die Aussage:

**Satz 2.** *Sei $E$ ein Banachraum, $\mathcal{U} \subset \mathbb{R}^n$ eine offene Teilmenge und $k \geq 0$ eine ganze Zahl. Dann gilt für jede Funktion $f \in C^{k+1}(\mathcal{U}, E)$ und Punkte $\mathbf{a}, \mathbf{x} = \mathbf{a} + \mathbf{h}$ mit $\mathbf{a} + t\mathbf{h} \in \mathcal{U}$ für alle $t \in [0, 1]$,*

$$f(\mathbf{x}) = P_k(\mathbf{x}) + R_k(\mathbf{x}), \qquad \textit{wobei} \qquad R_k(\mathbf{x}) = \frac{1}{k!} \int_0^1 (1-t)^k \, D^{k+1} f(\mathbf{a} + t\mathbf{h})(\mathbf{h}, \dots, \mathbf{h}) \, dt.$$

Wer Formeln mit Multiindizes lieber hat, kann wieder Aufgabe 7.Z einsetzen, um das Integralrestglied in der alternativen Form

$$R_k(\mathbf{x}) = \sum_{|\alpha|=k+1} \frac{k+1}{\alpha!} \mathbf{h}^\alpha \int_0^1 (1-t)^k \partial^\alpha f(\mathbf{a} + t\mathbf{h}) \, dt$$

hinzuschreiben.

Wie beim Mittelwertsatz kann in allen Anwendungen der Taylorformel das Integralrestglied anstelle von (5) eingesetzt werden. Hier ein konkretes Beispiel: für $f \in C^2(\mathcal{U}, \mathbb{R})$ gibt Satz 2 die Formel

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle + \int_0^1 (1-t) \cdot \langle \mathbf{h}, Hf(\mathbf{a}+t\mathbf{h})\mathbf{h} \rangle \, dt,$$

wobei $Hf(\mathbf{y})$ die Hesse-Matrix von $f$ im Punkt $\mathbf{y} \in \mathcal{U}$ bezeichnet. Falls $\mathbf{a}$ ein kritischer Punkt von $f$ ist, gilt dann

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \int_0^1 (1-t) \cdot \langle \mathbf{h}, Hf(\mathbf{a}+t\mathbf{h})\mathbf{h} \rangle \, dt. \tag{6}$$

Nehmen wir jetzt an, $Hf(\mathbf{a})$ sei negativ-definit. Für $t \in [0,1]$ liegt $\mathbf{a} + t\mathbf{h}$ immer in der Kugel von Radius $\|\mathbf{h}\|$ um $\mathbf{a}$, also für $\|\mathbf{h}\|$ hinreichend klein dürfen wir annehmen, dass $Hf(\mathbf{a}+t\mathbf{h})$ für alle $t \in [0,1]$ auch negativ-definit ist.[3] Da $1-t \geq 0$ für alle $t \in [0,1]$ gilt, ist in diesem Fall das Integral in (6) negativ, also impliziert (6) die Ungleichung

$$f(\mathbf{a}+\mathbf{h}) < f(\mathbf{a})$$

für alle $\mathbf{h} \in \mathbb{R}^n$ hinreichend klein; anders gesagt, $f$ hat in $\mathbf{a}$ ein isoliertes lokales Maximum. Das wussten wir schon, aber jetzt haben wir zwei leicht unterschiedliche Beweise dafür.

*Beweis von Satz 2.* Im Fall $k = 0$ ist die Aussage $f(\mathbf{x}) = f(\mathbf{a}) + \int_0^1 Df(\mathbf{a}+t\mathbf{h})\mathbf{h}\, dt$, und dies ist wieder Satz 1. Wir argumentieren jetzt per vollständige Induktion über $k$. Als Induktionsvoraussetzung nehmen wir für ein gegebenes $k \in \mathbb{N}$ Folgendes an: für jede Funktion $f \in C^k(\mathcal{U}, E)$ gilt $f(\mathbf{x}) = P_{k-1}(\mathbf{x}) + R_{k-1}(\mathbf{x})$, wobei

$$R_{k-1}(\mathbf{x}) = \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} \, D^k f(\mathbf{a}+t\mathbf{h})(\mathbf{h}, \ldots, \mathbf{h}) \, dt. \tag{7}$$

Um dieses Integral besser zu verstehen, können wir eine kleine Änderung vornehmen und $D^k f(\mathbf{a}+t\mathbf{h})$ mit der linearen Abbildung $D^k f(\mathbf{a}) \in \mathscr{L}(\mathbb{R}^n, E)$ ersetzen, die nicht von $t$ abhängt und also beim Berechnen dieses Integrals als "Konstante" betrachtet werden kann. Das Integral lässt sich dann explizit ausrechnen, und zwar

$$\begin{aligned} \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} \, D^k f(\mathbf{a})(\mathbf{h}, \ldots, \mathbf{h}) \, dt &= \frac{1}{(k-1)!} \int_0^1 \frac{d}{dt} \left( -\frac{(1-t)^k}{k} \, D^k f(\mathbf{a})(\mathbf{h}, \ldots, \mathbf{h}) \right) dt \\ &= -\frac{1}{k!} (1-t)^k \, D^k f(\mathbf{a})(\mathbf{h}, \ldots, \mathbf{h}) \Big|_{t=0}^{t=1} \\ &= \frac{1}{k!} D^k f(\mathbf{a})(\mathbf{h}, \ldots, \mathbf{h}) = Q_k(\mathbf{x}). \end{aligned} \tag{8}$$

Eine gegebene Funktion $f \in C^{k+1}(\mathcal{U}, E)$ ist auch in $C^k(\mathcal{U}, E)$ und kann daher als $f = P_{k-1} + R_{k-1}$ geschrieben werden, mit $R_{k-1}$ gegeben durch (7). Da $P_k = P_{k-1} + Q_k$, gilt dann $f = P_k + R_{k-1} - Q_k$, also $R_k = R_{k-1} - Q_k$. Wegen (8) kann nun $Q_k(\mathbf{x})$ als Integral

[3] Zur Erinnerung: Negativ-Definitheit ist eine "offene Bedingung," d.h. wenn eine Matrix diese Eigenschaft hat, dann bleibt die Eigenschaft bei allen hinreichend kleinen Störungen der Matrix erhalten.

geschrieben werden, und die Voraussetzung $f \in C^{k+1}$ impliziert $D^k f \in C^1$, damit wir in der dritten Zeile unten partielle Integration anwenden können:

$$\begin{aligned}
R_k(\mathbf{x}) = R_{k-1}(\mathbf{x}) - Q_k(\mathbf{x}) &= \frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1}\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right](\mathbf{h},\dots,\mathbf{h})\,dt \\
&= \frac{1}{(k-1)!}\int_0^1 \frac{d}{dt}\left(-\frac{(1-t)^k}{k}\right)\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right](\mathbf{h},\dots,\mathbf{h})\,dt \\
&= -\frac{1}{(k-1)!}\frac{(1-t)^k}{k}\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right](\mathbf{h},\dots,\mathbf{h})\Big|_{t=0}^{t=1} \\
&\quad + \frac{1}{(k-1)!}\int_0^1 \frac{(1-t)^k}{k}\frac{d}{dt}\left(\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right](\mathbf{h},\dots,\mathbf{h})\right)\,dt \\
&= \frac{1}{k!}\int_0^1 (1-t)^k \left[D(D^k f)(\mathbf{a}+t\mathbf{h})\mathbf{h}\right](\underbrace{\mathbf{h},\dots,\mathbf{h}}_{k})\,dt \\
&= \frac{1}{k!}\int_0^1 (1-t)^k\, D^{k+1} f(\mathbf{a}+t\mathbf{h})(\underbrace{\mathbf{h},\dots,\mathbf{h}}_{k+1})\,dt
\end{aligned}$$

□

*Bemerkung* 3. Die restlichen Inhalte dieses Skripts wurden in der Vorlesung nicht direkt diskutiert und sind nicht Prüfungsrelevant, aber vielleicht trotzdem interessant.

Im Fall $n = 1$ haben wir eine weitere Versionen vom Satz von Taylor gesehen, die auch auf den Mittelwertsatz basierte. Sie war eine direkte Verallgemeinerung der Definition von Differenzierbarkeit: sei $I \subset \mathbb{R}$ ein Intervall und $f : I \to \mathbb{R}$ eine Funktion, die im Punkt $a \in I$ $k$-fach differenzierbar ist, dann folgt

$$f(a+h) = \sum_{m=0}^{k} \frac{f^{(m)}(a)}{m!} h^m + o(h^k).$$

Unter der leicht stärkeren Voraussetzung, dass $f$ $k$-fach *stetig* differenzierbar ist, können wir dies jetzt für Funktionen von $n$ Variablen mit Werten in einem Banachraum verallgemeinern:

**Korollar 2.** *Sei $E$ ein Banachraum, $\mathcal{U} \subset \mathbb{R}^n$ eine offene Teilmenge und $k \in \mathbb{N}$, dann gilt für jede Funktion $f \in C^k(\mathcal{U}, E)$ und jeden Punkt $\mathbf{a} \in \mathcal{U}$,*

$$f(\mathbf{a}+\mathbf{h}) = P_k(\mathbf{a}+\mathbf{h}) + o(\|\mathbf{h}\|^k) \qquad \text{für } \mathbf{h} \in \mathbb{R}^n \text{ hinreichend klein.}$$

*Beweis.* Für $\mathbf{x} = \mathbf{a} + \mathbf{h}$ mit $\mathbf{h}$ hinreichend klein gibt Satz 2 $f(\mathbf{x}) = P_{k-1}(\mathbf{x}) + R_{k-1}(\mathbf{x}) = P_k(\mathbf{x}) + R_{k-1}(\mathbf{x}) - Q_k(\mathbf{x})$ mit einer Integraldarstellung für $R_{k-1}(\mathbf{x})$. Wir haben im Beweis von Satz 2 gesehen, dass $Q_k(\mathbf{x})$ eine ähnliche Integraldarstellung hat, wobei $D^k f(\mathbf{a}+t\mathbf{h})$ mit der Konstante $D^k f(\mathbf{a})$ ersetzt wird. Es folgt,

$$\begin{aligned}
R_k(\mathbf{x}) = R_{k-1}(\mathbf{x}) - Q_k(\mathbf{x}) &= \frac{1}{(k-1)!}\int_0^1 (1-t)^{k-1}\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right](\mathbf{h},\dots,\mathbf{h})\,dt \\
&= \frac{1}{(k-1)!}\left(\int_0^1 (1-t)^{k-1}\left[D^k f(\mathbf{a}+t\mathbf{h}) - D^k f(\mathbf{a})\right]\,dt\right)(\underbrace{\mathbf{h},\dots,\mathbf{h}}_{k})
\end{aligned} \tag{9}$$

Hier wurde beim Übergang von der ersten zur zweiten Zeile Lemma 1 wieder angewendet, diesmal mit der stetigen linearen Abbildung

$$\mathbf{L} : \mathscr{L}^k(\mathbb{R}^n, E) \to E : \mathbf{A} \mapsto \mathbf{A}(\mathbf{h}, \ldots, \mathbf{h}),$$

wobei $\mathscr{L}^k(\mathbb{R}^n, E)$ den Raum von $k$-fach multilinearen Abbildungen $\mathbf{A} : \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to E$ bezeichnet. Das Integral in Klammern in der zweiten Zeile von (9) definiert also ein Element $I(\mathbf{h}) \in \mathscr{L}^k(\mathbb{R}^n, E)$, und wegen der Stetigkeit von $D^k f$ konvergiert der Integrand bei $\mathbf{h} \to 0$ gleichmäßig gegen 0, also folgt

$$\lim_{\mathbf{h}\to 0} \|I(\mathbf{h})\| = 0.$$

Angesichts der Multilinearität gilt dann[4]

$$R_k(\mathbf{x}) = \|\mathbf{h}\|^k \cdot \frac{1}{(k-1)!} I(\mathbf{h}) \left( \frac{\mathbf{h}}{\|\mathbf{h}\|}, \ldots, \frac{\mathbf{h}}{\|\mathbf{h}\|} \right)$$

mit

$$\lim_{\mathbf{h}\to 0} I(\mathbf{h}) \left( \frac{\mathbf{h}}{\|\mathbf{h}\|}, \ldots, \frac{\mathbf{h}}{\|\mathbf{h}\|} \right) = 0,$$

also $R_k(\mathbf{x}) = o(\|\mathbf{h}\|^k)$. □

## Weitere Anwendungen vom Mittelwertsatz

Wo wir schon bei diesem Thema sind, wollte ich jetzt zeigen, wie sich ein paar andere wichtige Resultate durch die Integraldarstellung des Mittelwertsatzes verallgemeinern lassen.

**Satz 3.** *Sei $E$ ein Banachraum, $\mathcal{U} \subset \mathbb{R}^n$ eine offene Teilmenge und $f : \mathcal{U} \to E$ eine stetige Funktion mit der Eigenschaft, dass die partiellen Ableitungen $\partial_i f : \mathcal{U} \to E$ für $i = 1, \ldots, n$ alle existieren und auch stetige Funktionen sind. Dann ist $f$ stetig differenzierbar.*

*Beweis.* Sei $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{U}$ und betrachte $\mathbf{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ klein genug, so dass die Würfel $[a_1 - |h_1|, a_1 + |h_1|] \times \ldots \times [a_n - |h_n|, a_n + |h_n|]$ enthalten in $\mathcal{U}$ ist. Dann gilt:

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = {} & f(a_1 + h_1, a_2, \ldots, a_n) - f(a_1, a_2, \ldots, a_n) \\ & + f(a_1 + h_1, a_2 + h_2, a_3, \ldots, a_n) - f(a_1 + h_1, a_2, a_3, \ldots, a_n) \\ & + \ldots \\ & + f(a_1 + h_1, \ldots, a_{n-1} + h_{n-1}, a_n + h_n) - f(a_1 + h_1, \ldots, a_{n-1} + h_{n-1}, a_n). \end{aligned} \tag{10}$$

Die Existenz und Stetigkeit von $\partial_1 f$ impliziert, dass die Funktion $g_1(x) := f(x, a_2, \ldots, a_n)$ auf $[a_1 - |h_1|, a_1 + |h_1|]$ stetig differenzierbar ist, mit $g_1'(x) = \partial_1 f(x, a_2, \ldots, a_n)$, also kann

[4] Analog zu $\mathscr{L}(\mathbb{R}^n, E)$ wird auf dem Raum $\mathscr{L}^k(\mathbb{R}^n, E)$ die kanonische Norm durch

$$\|\mathbf{A}\| := \sup_{\|\mathbf{v}_1\| = \ldots = \|\mathbf{v}_k\| = 1} \|\mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\|$$

definiert, also gilt $\|\mathbf{A}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\| \le \|\mathbf{A}\| \cdot \|\mathbf{v}_1\| \cdot \ldots \cdot \|\mathbf{v}_k\|$ für alle $\mathbf{A} \in \mathscr{L}^k(\mathbb{R}^n, E)$ und $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$.

die erste Zeile in (10) durch Anwendung von Satz 1 als

$$\begin{aligned} g_1(a_1+h_1)-g_1(a_1) &= \left(\int_0^1 Dg_1(a_1+th_1)\,dt\right)h_1 = h_1\int_0^1 g_1'(a_1+th_1)\,dt \\ &= h_1\int_0^1 \partial_1 f(a_1+th_1,a_2,\dots,a_n)\,dt \end{aligned}$$

geschrieben werden. Bei den weiteren Zeilen in (10) wendet man Satz 1 analog an und erhält dadurch die Darstellung

$$\begin{aligned} f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) &= h_1\int_0^1 \partial_1 f(a_1+th_1,a_2,\dots,a_n)\,dt + h_2\int_0^1 \partial_2 f(a_1+h_1,a_2+th_2,a_3,\dots,a_n)\,dt \\ &\quad + \ldots + h_n\int_0^1 \partial_n f(a_n+h_n,\dots,a_{n-1}+h_{n-1},a_n+th_n)\,dt \\ &= \sum_{i=1}^n h_i\cdot\partial_i f(\mathbf{a}) + h_1\int_0^1 \left[\partial_1 f(a_1+th_1,a_2,\dots,a_n)-\partial_1 f(\mathbf{a})\right]\,dt \\ &\quad + h_2\int_0^1 \left[\partial_2 f(a_1+h_1,a_2+th_2,a_3,\dots,a_n)-\partial_2 f(\mathbf{a})\right]\,dt \\ &\quad + \ldots + h_n\int_0^1 \left[\partial_n f(a_1+h_1,\dots,a_{n-1}+h_{n-1},a_n+th_n)-\partial_n f(\mathbf{a})\right]\,dt \end{aligned}$$

Die Summe $\sum_{i=1}^n h_i\cdot\partial_i f(\mathbf{a})$ definiert eine lineare Funktion von $\mathbf{h}=(h_1,\dots,h_n)\in\mathbb{R}^n$ mit Werten in $E$, die hier als das Differential $Df(\mathbf{a}):\mathbb{R}^n\to E$ zu interpretieren ist. Alles, was danach folgt, ist als Restglied zu verstehen, denn es hat alles die Form $h_iR_i(\mathbf{h})$ für $i=1,\dots,n$, mit $R_i(\mathbf{h})$ gegeben durch Integrale, die wegen der Stetigkeit von $\partial_i f$ beliebig klein sind, wenn $\mathbf{h}$ beliebig klein angenommen wird. Dies beweist,

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^n h_i\cdot\partial_i f(\mathbf{a}) + o(\|\mathbf{h}\|),$$

also ist $f$ in $\mathbf{a}$ differenzierbar, mit Differential gegeben durch $Df(\mathbf{a})\mathbf{h}=\sum_{i=1}^n h_i\cdot\partial_i f(\mathbf{a})$. Die Stetigkeit von $Df:\mathcal{U}\to\mathscr{L}(\mathbb{R}^n,E)$ folgt nun direkt von der Stetigkeit der Funktionen $\partial_i f:\mathcal{U}\to E$. □

*Bemerkung* 4. In unserer früheren Version von Satz 3 für Funktionen $f:\mathcal{U}\to\mathbb{R}$ waren die Voraussetzungen ein kleines bisschen schwächer: die partiellen Ableitungen mussten nicht zwingend überall stetig sein, sondern für die Differenzierbarkeit von $f$ im Punkt $\mathbf{a}$ reichte, dass die partiellen Ableitungen in einer Umgebung davon existieren und in diesem einen Punkt stetig sind. Das wäre für Satz 3 zu schwach, denn wir send jetzt auf den Hauptsatz der Differential- und Integralrechnung angewiesen, also müssen die partiellen Ableitungen in der Umgebung von $\mathbf{a}$ stetig sein, wo sie integriert werden.

Abschließend möchten wir noch das Lemma von Schwarz verallgemeinern:

**Satz 4.** *Sei $E$ ein Banachraum, $\mathcal{U}\subset\mathbb{R}^n$ eine offene Teilmenge und $f\in C^2(\mathcal{U},E)$. Dann gilt für alle $i,j\in\{1,\dots,n\}$,*

$$\partial_i\partial_j f = \partial_j\partial_i f.$$

*Beweis.* Um die Notation in Grenzen zu halten, betrachten wir nur den Fall $n=2$, also zu beweisen ist die Formel

$$\partial_1\partial_2 f(\mathbf{a}) = \partial_2\partial_1 f(\mathbf{a})$$

in einem beliebigen Punkt $\mathbf{a} = (a_1, a_2) \in \mathcal{U}$. Wähle $\mathbf{h} = (h_1, h_2)$ hinreichend klein, so dass die Kiste $[a_1 - |h_1|, a_1 + |h_1|] \times [a_2 - |h_2|, a_2 + |h_2|]$ in $\mathcal{U}$ enthalten ist, und betrachte die Funktionen

$$\varphi(x) := f(x, a_2 + h_2) - f(x, a_2), \qquad \psi(x) := f(a_1 + h_1, x) - f(a_1, x)$$

auf den Intervallen $[a_1 - |h_1|, a_1 + |h_1|]$ bzw. $[a_2 - |h_2|, a_2 + |h_2|]$. Es gilt,

$$\varphi(a_1 + h_1) - \varphi(a_1) = \psi(a_2 + h_2) - \psi(a_2).$$

Durch zwei Anwendungen von Satz 1 lässt sich nun die linke Seite dieser Gleichung als

$$\begin{aligned}
h_1 \int_0^1 \varphi'(a_1 + sh_1)\, ds &= h_1 \int_0^1 [\partial_1 f(a_1 + sh_1, a_2 + h_2) - \partial_1 f(a_1 + sh_1, a_2)]\, ds \\
&= h_1 \int_0^1 \left( h_2 \int_0^1 \partial_2 \partial_1 f(a_1 + sh_1, a_2 + th_2)\, dt \right) ds \\
&= h_1 h_2 \int_0^1 \left( \int_0^1 \partial_2 \partial_1 f(a_1 + sh_1, a_2 + th_2)\, dt \right) ds
\end{aligned}$$

hinschreiben, und die rechte Seite lässt sich analog als

$$\begin{aligned}
h_2 \int_0^1 \psi'(a_2 + th_2)\, dt &= h_2 \int_0^1 [\partial_2 f(a_1 + h_1, a_2 + th_2) - \partial_2 f(a_1, a_2 + th_2)]\, dt \\
&= h_2 \int_0^1 \left( h_1 \int_0^1 \partial_1 \partial_2 f(a_1 + sh_1, a_2 + th_2)\, ds \right) dt \\
&= h_1 h_2 \int_0^1 \left( \int_0^1 \partial_1 \partial_2 f(a_1 + sh_1, a_2 + th_2)\, ds \right) dt
\end{aligned}$$

hinschreiben. Es folgt, für alle $h_1 \neq 0$ und $h_2 \neq 0$ hinreichend klein gilt

$$\int_0^1 \left( \int_0^1 \partial_2 \partial_1 f(a_1 + sh_1, a_2 + th_2)\, dt \right) ds = \int_0^1 \left( \int_0^1 \partial_1 \partial_2 f(a_1 + sh_1, a_2 + th_2)\, ds \right) dt.$$

Da alle partiellen Ableitungen bis Ordnung 2 stetig sind, kann jetzt für $h_1$ und $h_2$ hinreichend klein angenommen werden, dass der Integrand auf beiden Seiten für alle $s, t \in [0, 1]$ beliebig nahe an der konstanten Funktion $\partial_2 \partial_1 f(a_1, a_2)$ bzw. $\partial_1 \partial_2 f(a_1, a_2)$ ist, und folglich konvergiert die linke Seite bei $\mathbf{h} \to 0$ gegen $\partial_2 \partial_1 f(\mathbf{a})$, während die rechte Seite gegen $\partial_1 \partial_2 f(\mathbf{a})$ konvergiert. $\square$