

---

# Kapitel I

## Differentialrechnung mehrerer Variablen

*Ja, saadan var det, saadan vokser Ens Væsen  
med Ens Viden, klares deri, samles igjennem  
den. Det er saa skjønt at lære som at leve.  
Vær ikke bange for at miste dig selv i større  
Aander end din egen.*

*Niels Lyhne.*

Wir erklären die grundlegenden Regeln der Differentialrechnung für Abbildungen zwischen endlichdimensionalen reellen Vektorräumen. Es erweist sich, daß alle eigentlich analytische Arbeit schon im Eindimensionalen getan ist. Wenn die Definition der Ableitung einmal richtig gefaßt ist, kommt es hier vor allem darauf an, den Kalkül der Linearen Algebra für unsere Zwecke zu interpretieren und zu benutzen.

### § 1. Kurven im euklidischen Raum.

Eine Abbildung  $X \rightarrow Y \times Z$  in ein Produkt von topologischen Räumen ist genau dann stetig, wenn die beiden Komponenten  $X \rightarrow Y$  und  $X \rightarrow Z$  stetig sind, und ganz genau so verhält es sich mit differenzierbaren Abbildungen. Daher ist noch nichts Bedenkliches geschehen, wenn wir jetzt als ersten Blick auf das Höherdimensionale Abbildungen  $D \rightarrow \mathbb{R}^n$  betrachten, wo  $D$  ein Intervall und eben nur der Bildraum höherdimensional, nämlich das  $n$ -fache Produkt von  $\mathbb{R}$  mit sich selbst ist. Mit solchen Abbildungen, die wir im stetigen

Fälle Wege genannt haben und in dem hier betrachteten Zusammenhang auch Kurven nennen, hat man es in Anwendungen oft zu tun. Etwa in der Mechanik wird einem System von  $n$  Massenpunkten eine Kurve  $\mathbb{R} \rightarrow \mathbb{R}^{6n}$  zugeordnet, nämlich jedem Zeitpunkt  $t \in \mathbb{R}$  ordnet man drei Ortskoordinaten und drei Geschwindigkeitskoordinaten jedes Massenpunktes zu.

Wenn wir hernach auch zum Beispiel Abbildungen studieren, deren Definitionsgebiet höherdimensional ist, so wird eine wichtige Methode sein, daß wir allerlei Kurven durch das Definitionsgebiet legen, und die Einschränkung der Abbildung auf diese Kurven also auf Eindimensionale betrachten.

Kommen wir nun zu genaueren und auch etwas technischen Erklärungen. Sei also  $D \subset \mathbb{R}$  ein Intervall mit mindestens zwei Punkten, so daß also  $D$  weder leer ist, noch zu einem Punkt degeneriert. Sei

$$\gamma : D \rightarrow \mathbb{R}^n, \quad t \mapsto \gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$$

eine Abbildung. Wie gesagt, ist  $\gamma$  stetig, genau wenn alle Komponenten  $\gamma_i : D \rightarrow \mathbb{R}$  stetig sind; eine stetige solche Abbildung heißt eine **Kurve** in  $\mathbb{R}^n$  oder auch ein **Weg**. Sie heißt **stetig differenzierbar**, wenn jede Komponente  $\gamma_i$  von  $\gamma$  diese Eigenschaft hat. Sie heißt **stückweise stetig differenzierbar**, falls es eine Zerlegung

$$a = t_0 \leq t_1 \leq \dots \leq t_m = b$$

von  $D \subset [a, b]$  gibt, so daß  $\gamma|_{[t_i, t_{i+1}] \cap D}$  stetig differenzierbar ist.

Ganz entsprechend erklärt man komponentenweise, wann die Kurve  $k$ -mal stetig differenzierbar ( $C^k$ ), stückweise  $k$ -mal stetig differenzierbar oder an der Stelle  $\tau \in D$  differenzierbar heißt. Ist letzteres der Fall, so heißt der Vektor

$$\dot{\gamma}(\tau) := (d/dt \gamma_1(\tau), \dots, d/dt \gamma_n(\tau))$$

der **Geschwindigkeitsvektor** oder die **Ableitung** von  $\gamma$  bei  $\tau$  und die reellen Vielfachen von  $\dot{\gamma}(\tau)$  heißen **Tangentialvektoren** von  $\gamma$

in  $\tau$ . Die **Geschwindigkeit** von  $\gamma$  in  $\tau$  ist  $|\dot{\gamma}(\tau)|$ , und wir deuten gern den Parameter  $t$  als Zeit. Statt “an der Stelle” heißt es dann entsprechend “zur Zeit  $\tau$ ”. Hier ist immer die euklidische Norm in  $\mathbb{R}^n$  zum Standard-Skalarprodukt gemeint. Im allgemeinen schreiben wir Geschwindigkeitsvektoren und Tangentialvektoren wie hier als Zeilentupel; nur wenn wir in Rechnungen vom Matrizenkalkül der Linearen Algebra Gebrauch machen, sind die Geschwindigkeits- und Tangentialvektoren als Spalten zu notieren. Sind die Komponenten  $\gamma_i$  von  $\gamma$  integrierbar, so setzen wir für  $[a, b] \subset D$

$$\int_a^b \gamma(t) dt := \left( \int_a^b \gamma_1(t) dt, \dots, \int_a^b \gamma_n(t) dt \right) \in \mathbb{R}^n.$$

Eine stückweise stetig differenzierbare Kurve  $\gamma$  in  $\mathbb{R}^n$  dürfen wir uns durch ihr Bild im  $\mathbb{R}^n$ , also etwa



veranschaulichen. Wir nennen  $\gamma(a)$  den **Anfangs-** und  $\gamma(b)$  den **Endpunkt** der Kurve und sagen: Die Kurve läuft von  $\gamma(a)$  nach  $\gamma(b)$ , oder sie **verbindet** diese Punkte miteinander, wenn  $D = [a, b]$  ist. Das Bild von  $\gamma$  nennt man auch die **Spur**, aber wir nennen es meist einfach wieder die Kurve  $\gamma$ .

### Beispiele.

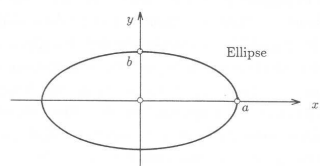
$$\gamma(t) = p + r(\cos t, \sin t), \quad p \in \mathbb{R}^2, \quad r \in \mathbb{R}_+, \quad 0 \leq t \leq 2\pi,$$

beschreibt den Kreis um  $p$  mit Radius  $r$ . Die Tangentialvektoren des Kreises  $(-\sin t, \cos t)$  stehen senkrecht auf dem Ortsvektor  $\gamma(t) - p$ .

Durch affine Verzerrung von  $\mathbb{R}^2$  erhält man aus dem Kreis eine **Ellipse**, z.B. durch eine Gleichung

$$\gamma(t) = (a \cos t, b \sin t)$$

beschrieben.



Die Funktion

$$\gamma(t) = p + t \cdot v, \quad p, v \in \mathbb{R}^n$$

beschreibt eine Gerade durch  $p$  mit Geschwindigkeitsvektor  $v$ .

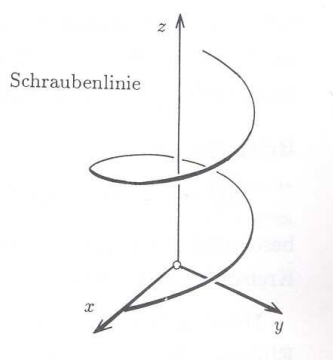
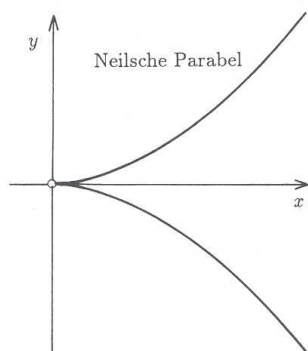
Die Funktion

$$\gamma(t) = (t^2, t^3)$$

beschreibt die **Neilsche Parabel** mit einer Spitze im Ursprung, wo der Geschwindigkeitsvektor verschwindet. Die Funktion

$$\gamma(t) = (\cos t, \sin t, t)$$

ist eine **Schraubenlinie** in  $\mathbb{R}^3$ . Ihre Projektion auf die  $(x, y)$ -Ebene ist der Kreis.



Den Geschwindigkeitsvektor  $\dot{\gamma}(t)$  zeichnen wir gerne an den Punkt  $\gamma(t)$ , obwohl es natürlich ein Vektor in  $\mathbb{R}^n$  ist.

Wir wollen auch von der **Länge** einer Kurve im euklidischen Raum reden. Ist  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  die Kurve und wäre  $s(t)$  die Länge zwischen  $a$  und  $t$ , so wäre  $ds/dt$  die Geschwindigkeit, also  $\dot{s} = ds/dt = |\dot{\gamma}|$ , oder in sinnfälliger Schreibweise:

$$\gamma(t) = (x_1(t), \dots, x_n(t)), \text{ und}$$

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \dots + \left(\frac{dx_n}{dt}\right)^2}, \quad \text{oder } ds = \sqrt{dx_1^2 + \dots + dx_n^2}.$$

Wir erklären daher:

**Definition.** Die **Bogenlänge** einer stückweise stetig differenzierbaren Kurve  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  ist

$$s(\gamma) = \int_a^b |\dot{\gamma}(t)| dt.$$

Weil der Integrand stückweise stetig ist, ist das Integral wohldefiniert, und die Funktion

$$s : [a, b] \rightarrow \mathbb{R}, \quad t \mapsto \int_a^t |\dot{\gamma}(\tau)| d\tau = s(\gamma|[a, t])$$

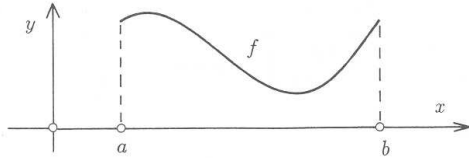
(die wir auch **Bogenlänge** nennen) ist stetig, monoton wachsend, und wenn  $\gamma$  stetig differenzierbar ist, ist auch  $s$  stetig differenzierbar mit der Ableitung  $ds/dt =: \dot{s}(t) = |\dot{\gamma}(t)|$ .

**Beispiel.** Der Kreis mit Radius 1 ist gegeben durch  $\gamma : [0, 2\pi] \rightarrow \mathbb{R}^2$ ,  $t \mapsto (\cos t, \sin t)$ , also

$$\begin{aligned} s(\gamma) &= \int_0^{2\pi} |(\cos t, \sin t)'| dt = \int_0^{2\pi} |(-\sin t, \cos t)| dt \\ &= \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t} dt = \int_0^{2\pi} 1 \cdot dt = 2\pi. \end{aligned}$$

Ist eine Kurve in  $\mathbb{R}^2$  durch die  $x$ -Achse parametrisiert, also ist  $\gamma(x) = (x, f(x))$ , so ist  $d/dx \gamma = (1, f'(x))$ , also  $|\dot{\gamma}| = \sqrt{1 + (f')^2}$ , und damit

$$s(\gamma) = \int_a^b \sqrt{1 + f'(x)^2} dx.$$



**(1.1) Satz.** Die Bogenlänge einer stückweise stetig differenzierbaren Kurve ist unabhängig von der Parametrisierung.

Das heißt genauer folgendes: Gegeben sei die Kurve  $\gamma: [a, b] \rightarrow \mathbb{R}^n$  und eine stetig differenzierbare Parametertransformation

$$\varphi: [\alpha, \beta] \rightarrow [a, b], \quad \varphi(\alpha) = a, \quad \varphi(\beta) = b,$$

so daß stets  $\varphi' \geq 0$ , dann haben  $\gamma$  und  $\gamma \circ \varphi$  gleiche Bogenlänge.

**Beweis:** Nach geeigneter Zerlegung  $\alpha = \tau_0 \leq \tau_1 \leq \dots \leq \tau_n = \beta$  ist  $\gamma$  stetig differenzierbar auf  $[\varphi(\tau_i), \varphi(\tau_{i+1})]$  — hier wurde der Zwischenwertsatz auf  $\varphi$  angewandt und  $\varphi' \geq 0$  benutzt. Man darf also annehmen, daß  $\gamma$  stetig differenzierbar ist, und hat dann

$$\begin{aligned} s(\gamma) &= \int_a^b |\dot{\gamma}(t)| dt = \int_{\varphi(\alpha)}^{\varphi(\beta)} |d/dt \gamma(\varphi(\tau))| d\varphi(\tau) = \\ &= \int_{\alpha}^{\beta} |d/dt \gamma(\varphi(\tau))| \varphi'(\tau) d\tau = \int_{\alpha}^{\beta} |d/d\tau \gamma \circ \varphi(\tau)| d\tau = s(\gamma \circ \varphi). \quad \square \end{aligned}$$

Die Geschwindigkeit und der Geschwindigkeitsvektor sind natürlich abhängig von der Parametrisierung, aber ist  $\varphi'(\tau) \neq 0$ , so besagt

für  $\varphi(\tau) = t$  die Gleichung

$$\dot{\gamma}(t) \cdot \varphi'(\tau) = \frac{d}{d\tau}(\gamma \circ \varphi),$$

daß  $\gamma$  und  $\gamma \circ \varphi$  gleiche Tangentialvektoren in  $t = \varphi(\tau)$  bzw.  $\tau$  haben, der **Tangentialraum**  $\{\lambda \cdot \dot{\gamma}(t) \mid \lambda \in \mathbb{R}\}$  von  $\gamma$  zur Zeit  $t$  ist unabhängig von der Parametrisierung.

Am natürlichsten ist es, eine Kurve durch ihre Bogenlänge zu parametrisieren. Sei also  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  eine stetig differenzierbare Kurve, und sei stets  $\dot{\gamma} \neq 0$  auf  $[a, b]$ . Dann ist die Bogenlänge  $s : [a, b] \rightarrow [0, s(\gamma)]$  stetig differenzierbar mit der Ableitung  $|\dot{\gamma}| > 0$ , also besitzt diese Transformation eine differenzierbare Umkehrung  $s^{-1} : [0, s(\gamma)] \rightarrow [a, b]$ .

$$\begin{array}{ccc} [a, b] & \xrightarrow{\gamma} & \mathbb{R}^n \\ \cong \downarrow s & \nearrow \tilde{\gamma} & \\ [0, s(\gamma)] & & \end{array}$$

Und wenn man  $\tilde{\gamma} = \gamma \circ s^{-1}$  setzt, so folgt  $|\dot{\gamma}| = |d\tilde{\gamma}/ds| \cdot |ds/dt| = |d\tilde{\gamma}/ds| \cdot |\dot{\gamma}|$ , also  $\tilde{\gamma}$  hat Einheitsgeschwindigkeit, der Parameter ist die Bogenlänge. Diese Parametrisierung benutzt man, um das Integral einer Funktion auf einer Kurve zu erklären, genauer:

**Definition.** Sei  $\gamma : [a, b] \rightarrow X \subset \mathbb{R}^n$  eine stückweise stetig differenzierbare Kurve und  $f : X \rightarrow \mathbb{R}^k$  eine Funktion, so daß  $f \circ \gamma$  integrierbar ist, dann ist

$$\int_{\gamma} f := \int_0^{s(\gamma)} f(\tilde{\gamma}(s)) ds := \int_a^b f \circ \gamma(t) |\dot{\gamma}(t)| dt.$$

Jedes Stückchen  $dt$  wird also beim Integrieren mit  $|\dot{\gamma}(t)|$  gewichtet, also soviel gezählt, wie die Länge der Kurve zunimmt. Auch dieses Integral ist unabhängig von der Parametrisierung, und der Wert ist in unserem Falle ein Punkt in  $\mathbb{R}^k$ .

Ist nämlich  $t = \varphi(\tau)$ , so  $dt = d\varphi/d\tau \cdot d\tau$ , also

$$\int (f \circ \gamma) \cdot |\dot{\gamma}| dt = \int (f \circ \gamma \circ \varphi) \cdot |\dot{\gamma} \circ \varphi| \cdot \frac{d\varphi}{d\tau} d\tau = \int (f \circ \gamma \circ \varphi) \cdot \left| \frac{d\gamma \circ \varphi}{d\tau} \right| d\tau.$$

Für das Integral längs  $\gamma$  hat man folgende plausible

**(1.2) Integralabschätzung.** Sei  $\gamma : [a, b] \rightarrow X \subset \mathbb{R}^n$  eine stückweise stetig differenzierbare Kurve und  $f : X \rightarrow \mathbb{R}^k$  eine Funktion, so daß  $|f \circ \gamma(t)| \leq M$  für alle  $t \in [a, b]$ . Dann ist

$$\left| \int_{\gamma} f \right| \leq \int_{\gamma} |f| \leq s(\gamma) \cdot M.$$

**Beweis:** Es bezeichne  $\langle u, v \rangle$  das euklidische Skalarprodukt von  $u$  und  $v$  in  $\mathbb{R}^k$ , dann ist für eine Funktion  $\varphi : [a, b] \rightarrow \mathbb{R}^k$  und einen Vektor  $v \in \mathbb{R}^k$  offenbar

$$\left\langle \int_a^b \varphi(t) dt, v \right\rangle = \int_a^b \langle \varphi(t), v \rangle dt,$$

wegen der Linearität des Integrals. Setzen wir daher  $\int_{\gamma} f = v$ , so ist

$$\begin{aligned} |v|^2 &= \left| \int_{\gamma} f \right|^2 = \left\langle \int_{\gamma} f, v \right\rangle = \int_a^b \langle f \circ \gamma, v \rangle |\dot{\gamma}| dt \leq \int_a^b |f \circ \gamma| \cdot |v| \cdot |\dot{\gamma}| dt \\ &= |v| \cdot \int_a^b |f \circ \gamma| \cdot |\dot{\gamma}| dt = |v| \cdot \int_{\gamma} |f|, \quad \text{also} \quad |v| \leq \int_{\gamma} |f|. \end{aligned}$$

Hier habe ich die Schwarzsche Ungleichung  $|\langle f, v \rangle| \leq |f| \cdot |v|$  benutzt. Ist nun  $|f| \leq M$ , so können wir weiter abschätzen:

$$\int_{\gamma} |f| = \int_a^b |f \circ \gamma(t)| |\dot{\gamma}(t)| dt \leq M \int_a^b |\dot{\gamma}(t)| dt = M \cdot s(\gamma). \quad \square$$

**(1.3) Anwendung.** Ist  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  eine stückweise stetig differenzierbare Kurve, so ist

$$s(\gamma) \geq |\gamma(b) - \gamma(a)|,$$

das heißt, die Gerade ist die kürzeste Verbindung zweier Punkte.

**Beweis:** Wir wenden die Integralabschätzung an und wählen dabei  $f = \dot{\gamma}$  als Funktion auf der Kurve  $\text{id} : [a, b] \rightarrow [a, b]$ , dann folgt:

$$|\gamma(b) - \gamma(a)| = \left| \int_a^b \dot{\gamma}(t) dt \right| \leq \int_a^b |\dot{\gamma}(t)| dt = s(\gamma). \quad \square$$

Daß eine Kurve  $\gamma : D \rightarrow \mathbb{R}^n$  an der Stelle  $\tau \in D$  differenzierbar ist heißt, wenn man alle Komponenten wieder zu einem Vektor zusammenfaßt:

$$\gamma(\tau + h) = \gamma(\tau) + h \cdot \Gamma(h), \quad \Gamma(0) =: \dot{\gamma}(\tau),$$

mit einer bei 0 stetigen Funktion  $\Gamma : D - \tau \rightarrow \mathbb{R}^n$ . Daraus erhält man leicht folgende

**(1.4) Rechenregeln.**

(i) Sind  $\gamma, \varphi : D \rightarrow \mathbb{R}^n$  differenzierbar an der Stelle  $\tau \in D$ , und ist  $\langle \gamma, \varphi \rangle : D \rightarrow \mathbb{R}$  durch  $\langle \gamma, \varphi \rangle(t) = \langle \gamma(t), \varphi(t) \rangle$  erklärt, so ist

$$\left. \frac{d}{dt} \right|_{t=\tau} \langle \gamma(t), \varphi(t) \rangle = \langle \dot{\gamma}(\tau), \varphi(\tau) \rangle + \langle \gamma(\tau), \dot{\varphi}(\tau) \rangle.$$

(ii) Ist  $\gamma$  wie oben und  $\varphi : D \rightarrow \mathbb{R}$  differenzierbar bei  $\tau$ , so ist dort

$$(\varphi \cdot \gamma)' = \dot{\varphi} \gamma + \varphi \dot{\gamma}.$$

**Beweis:** Wir schreiben:

$$\gamma(\tau + h) = \gamma(\tau) + h \Gamma(h) \quad \text{und} \quad \varphi(\tau + h) = \varphi(\tau) + h \Phi(h),$$

wie oben, und erhalten:

$$\langle \gamma, \varphi \rangle(\tau + h) = \langle \gamma, \varphi \rangle(\tau) + h (\langle \gamma(\tau), \Phi(h) \rangle + \langle \Gamma(h), \varphi(\tau) \rangle) + h^2(\dots),$$

und daraus sogleich die erste Behauptung. Ebenso die zweite.  $\square$

## § 2. Differenzierbare Abbildungen

Jetzt werden wir Abbildungen  $\mathbb{R}^n \supset U \rightarrow \mathbb{R}^p$  betrachten, deren Definitionsgebiet auch höherdimensional ist. Wir wollen uns darauf besinnen, was lineare Abbildungen  $\mathbb{R}^n \rightarrow \mathbb{R}^p$  sind, und wollen diese ohne weiteres mit  $(p \times n)$ -Matrizen identifizieren, sodaß also der Matrix  $A = (a_{ij})$  die Abbildung

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^p, \quad x \mapsto y, \quad y_i = \sum_j a_{ij} x_j$$

entspricht.

Um die Ableitung einer Abbildung  $f$  an der Stelle  $x$  zu erklären, konnte man im Eindimensionalen vom Differenzenquotienten  $\Delta(h)$  ausgehen, der durch die Gleichung

$$f(x+h) - f(x) = \Delta(h) \cdot h$$

eindeutig definiert ist. Ist nun aber  $U$  offen in  $\mathbb{R}^n$  und eine Abbildung  $f : U \rightarrow \mathbb{R}^p$  gegeben, so ist für  $h$  entsprechend ein Vektor aus  $\mathbb{R}^n$  zu nehmen, durch den man nicht dividieren kann. Auch ist eine lineare Abbildung  $\Delta : \mathbb{R}^n \rightarrow \mathbb{R}^p$  für  $n > 1$  keineswegs durch *einen* Wert  $\Delta \cdot h$  bestimmt. Es kommt allerdings eben auch gar nicht auf die Eindeutigkeit des  $\Delta(h)$  an, sondern nur auf die Existenz, und wie es schon im Eindimensionalen oft geschickter ist, die Definitionsgleichung so stehen zu lassen, ohne durch  $h$  zu dividieren, so ist es hier unumgänglich notwendig.

**Definition** (Differenzierbarkeit). Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}^p$  eine Abbildung. Sie heißt **differenzierbar** bei  $x \in U$  mit dem **Differential** oder der **Ableitung**  $Df(x)$ , wenn folgendes gilt:

Es gibt eine Abbildung

$$A : U \rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^p) \cong \mathbb{R}^{p \cdot n}$$

von  $U$  in den Raum der linearen Abbildungen  $\mathbb{R}^n \rightarrow \mathbb{R}^p$  (also in den Raum der reellen  $(p \times n)$ -Matrizen), so daß für alle  $h \in \mathbb{R}^n$  mit  $x+h \in U$  gilt:

$$f(x+h) = f(x) + A(x+h) \cdot h,$$

und zwar so, daß  $A$  am Punkt  $x$  stetig ist, mit dem Wert

$$A(x) =: Df(x) \in \text{Hom}(\mathbb{R}^n, \mathbb{R}^p).$$

Das Differential  $Df(x)$  ist also eine lineare Abbildung; in Koordinaten wird sie durch eine  $(p \times n)$ -Matrix gegeben, die wir noch genauer betrachten werden. Wenn wir den Matrizenkalkül der linearen Algebra benutzen, müssen wir die Vektoren, wie hier  $h \in \mathbb{R}^n$ , als Spalten schreiben und entsprechend auch die Komponenten von  $f(x)$ .

Die Definition zeigt unmittelbar, daß eine am Punkt  $x$  differenzierbare Abbildung dort auch stetig ist, die rechte Seite der Formel ist stetig bei  $h = 0$ . Auch ist  $f$  genau dann an der Stelle  $x$  differenzierbar, wenn alle Komponenten  $f_j$ ,  $j = 1, \dots, p$  von  $f$  dort differenzierbar sind: Die Matrix  $A(x+h)$  ist das  $p$ -Tupel der Zeilen von  $A$ , und  $A$  ist genau dann stetig bei  $x$ , wenn dasselbe für jede Zeile gilt.

Oft faßt man die Definition der Differenzierbarkeit etwas anders — etwas weniger geschickt, wie mir scheint, und etwas weniger verallgemeinerungsfähig, dafür vielleicht etwas verständlicher — nämlich wie in der folgenden

**(2.1) Bemerkung.** Die Abbildung  $f : U \rightarrow \mathbb{R}^p$  ist genau dann differenzierbar bei  $x \in U$ , wenn es eine (feste) lineare Abbildung  $A : \mathbb{R}^n \rightarrow \mathbb{R}^p$  gibt, so daß

$$f(x+h) = f(x) + A \cdot h + \varphi(h), \quad \text{mit} \quad \lim_{h \rightarrow 0} \varphi(h)/|h| = 0.$$

**Beweis:** Ist  $f$  differenzierbar bei  $x$  nach unserer Definition, so setze  $A = A(x)$ , dann ist

$$f(x+h) = f(x) + A \cdot h + (A(x+h) - A(x)) \cdot h.$$

Setze  $\varphi(h) = (A(x+h) - A(x)) \cdot h$ , dann ist

$$\lim_{h \rightarrow 0} \varphi(h)/|h| = \lim_{h \rightarrow 0} (A(x+h) - A(x)) \cdot (h/|h|) = 0,$$

weil alle Komponenten der Matrix gegen 0 gehen und die von  $h/|h|$  höchstens 1 sind.

Erfüllt die Funktion umgekehrt die Bedingung von (2.1), so schreibe für  $h \neq 0$  nun:

$$f(x+h) = f(x) + A \cdot h + \langle h, h \rangle \cdot |h|^{-2} \cdot \varphi(h),$$

und erkläre die lineare Abbildung  $A(x+h) : \mathbb{R}^n \rightarrow \mathbb{R}^p$  durch

$$v \mapsto Av + \langle h, v \rangle \cdot |h|^{-2} \cdot \varphi(h) \quad (:= Av \text{ falls } h = 0).$$

Dann ist offenbar  $f(x+h) = f(x) + A(x+h) \cdot h$ , und die Stetigkeit von  $A(x+h)$  bei  $h = 0$  ist zu zeigen; aber für alle  $v \in \mathbb{R}^n$  ist

$$\lim_{h \rightarrow 0} \langle h, v \rangle \cdot |h|^{-2} \varphi(h) = \lim_{h \rightarrow 0} \langle |h|^{-1} h, v \rangle \cdot |h|^{-1} \varphi(h) = 0. \quad \square$$

Häufig fassen wir die Definition auch mit einer Formel

$$f(x+h) = f(x) + A(h) \cdot h,$$

dann ist  $A : U - x \rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^p) = \mathbb{R}^{pn}$  stetig am Nullpunkt.

**Bemerkung.** Das Differential ist durch  $f$  eindeutig bestimmt.

Ist nämlich

$$f(x+h) = f(x) + A(h)h = f(x) + B(h)h,$$

so folgt  $(A(h) - B(h)) \cdot h = 0$  für alle genügend kleinen  $h$ . Ist also  $v$  irgendein Vektor in  $\mathbb{R}^n$ , so gilt für kleine  $t > 0$  demnach  $(A(tv) - B(tv)) \cdot tv = 0$ , also  $(A(tv) - B(tv)) \cdot v = 0$ . Nun bilde den Limes für  $t \rightarrow 0$ , dann ergibt sich  $(A(0) - B(0)) \cdot v = 0$ , also  $A(0) = B(0)$ .  $\square$

Wie in der Dimension eins ergeben sich unmittelbar aus der Definition folgende

**(2.2) Rechenregeln.** Sind  $f, g : U \rightarrow \mathbb{R}^p$  differenzierbar bei  $x \in U \subset \mathbb{R}^n$  und  $\lambda, \mu$  Konstanten, so ist auch  $\lambda f + \mu g : U \rightarrow \mathbb{R}^p$  differenzierbar bei  $x$ , und es gilt:

**Linearität:**  $D(\lambda f + \mu g)(x) = \lambda Df(x) + \mu Dg(x)$ .

Seien  $U$  offen in  $\mathbb{R}^n$  und  $V$  offen in  $\mathbb{R}^m$ , und es seien Abbildungen

$$U \xrightarrow{f} V \xrightarrow{g} \mathbb{R}^p, \quad f(x) = y,$$

gegeben. Ist dann  $f$  differenzierbar an der Stelle  $x$  und  $g$  an der Stelle  $y$ , so ist  $g \circ f$  differenzierbar bei  $x$ , und es gilt die

**Kettenregel:**  $D(g \circ f)(x) = Dg(y) \cdot Df(x)$ .

Die lineare Approximation der Zusammensetzung ist die Zusammensetzung der linearen Approximationen.

**Beweis:** Beides folgt unmittelbar aus der Definition, die Kettenregel so: Es ist

$$\begin{aligned} f(x+h) &= f(x) + A(h)h, & A(0) &= Df(x), \\ g(y+k) &= g(y) + B(k)k, & B(0) &= Dg(y), \end{aligned}$$

wobei  $A$  und  $B$  bei 0 stetig sind. Daher ist (mit  $y = f(x)$  und  $k = A(h) \cdot h$ ):

$$g \circ f(x+h) = g(f(x) + A(h)h) = g \circ f(x) + B(A(h)h) \cdot A(h) \cdot h,$$

und  $\lim_{h \rightarrow 0} B(A(h)h)A(h) = B(0)A(0) = Dg(y) \cdot Df(x)$ .  $\square$

Eine affine Abbildung

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^p, \quad x \mapsto Ax + b,$$

hat konstantes Differential  $Df = A$ , denn

$$f(x+h) = A \cdot (x+h) + b = f(x) + Ah.$$

Das Differential einer Abbildung  $f$ , als lineare Abbildung  $\mathbb{R}^n \rightarrow \mathbb{R}^p$ , wird auf kanonische Weise durch eine Matrix gegeben. Wie berechnet man die Komponenten dieser Matrix? Das wollen wir jetzt beschreiben. Wir betrachten eine offene Menge  $U$  in  $\mathbb{R}^n$  und eine Abbildung

$$f: U \rightarrow \mathbb{R}^p, \quad x \mapsto (f_1(x), \dots, f_p(x))$$

mit den **Komponenten**

$$f_i : U \rightarrow \mathbb{R}, \quad i = 1, \dots, p.$$

Die  $j$ -te **partielle Ableitung** von  $f$  (bzw.  $f_i$ ) in einem Punkte  $u \in U$  ist

$$D_j f(u) := \lim_{h \rightarrow 0} h^{-1} (f(u + h e_j) - f(u)) =: \frac{\partial f}{\partial x_j}(u),$$

beziehungsweise dasselbe für die  $i$ -te Komponente

$$D_j f_i(u) = \lim_{h \rightarrow 0} h^{-1} (f_i(u + h e_j) - f_i(u)) = \frac{\partial f_i}{\partial x_j}(u).$$

Dabei ist  $e_j$  der  $j$ -te **Standard-Basisvektor** von  $\mathbb{R}^n$  mit  $j$ -ter Komponente 1 und Null sonst.

Mit anderen Worten,  $D_j f_i$  erhält man so: Man betrachtet alle Variablen  $x_\nu$  bis auf die  $j$ -te als konstant und nimmt  $f_i$  als Funktion der einen Veränderlichen  $x_j$ , dann ist  $D_j f_i$  die Ableitung. Wenn  $D_j f_i(u)$  existiert, heißt  $f_i$  bei  $u$  **partiell nach  $x_j$**  differenzierbar.

**(2.3) Satz.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}^p$  sei an der Stelle  $u \in U$  differenzierbar. Dann existieren alle partiellen Ableitungen  $D_j f_i(u) = \frac{\partial f_i}{\partial x_j}(u)$ , und die lineare Abbildung  $Df(u)$  ist durch die Matrix

$$Df(u) = (\partial f_i / \partial x_j(u))$$

gegeben. Diese Matrix heißt auch **Jacobi-Matrix** von  $f$  bei  $u$ . Im Matrizenkalkül sind die Vektoren aus  $\mathbb{R}^n$  und  $\mathbb{R}^p$  hier als Spalten zu schreiben.

**Beweis:** Die partiellen Ableitungen kann man auch so beschreiben: Für ein Intervall  $I$  um  $0 \in \mathbb{R}$  existiert die Zusammensetzung von Abbildungen

$$\varphi_{ij} : \begin{array}{ccccc} I & \xrightarrow{\kappa_j} & U & \xrightarrow{f} & \mathbb{R}^p & \xrightarrow{\text{pr}_i} & \mathbb{R}, \\ t & \longmapsto & u + t e_j, & & y & \longmapsto & y_i, \end{array}$$

und es ist  $D_j f_i(u) = D\varphi_{ij}(0) = d/dt \varphi_{ij}(0)$ .

Nun sind die Inklusionen und Projektionen  $\kappa_j$ ,  $\text{pr}_i$  affin, wenn also  $Df(u)$  existiert, so existiert nach der Kettenregel auch  $D\varphi_{ij}(0)$ , und es ist

$$D\varphi_{ij}(0) = D\text{pr}_i \circ Df(u) \circ D\kappa_j.$$

Aber

$$D\kappa_j : \lambda \mapsto \lambda e_j, \quad D\text{pr}_i : (v_1, \dots, v_p) \mapsto v_i,$$

also  $D\varphi_{ij}(0) = i$ -te Komponente von  $Df(u) \cdot e_j = Df(u)_{ij}$ .  $\square$

Eine Formel wie

$$f(x+h) - f(x) = Df(x) \cdot h + \varphi(h)$$

in der Erklärung der Differenzierbarkeit bedeutet also, wenn man sie in Komponenten und Matrizen expliziter aufschreibt:

$$\begin{bmatrix} f_1(x+h) - f_1(x) \\ \vdots \\ f_p(x+h) - f_p(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(x) & \cdots & \frac{\partial f_p}{\partial x_n}(x) \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \begin{bmatrix} \varphi_1(h) \\ \vdots \\ \varphi_p(h) \end{bmatrix}$$

In allgemeinen Überlegungen, mit denen es ja dieser Text meist zu tun hat, soll man sich aber unter einem Symbol wie  $Df(x) \cdot v$  das Bild des Vektors  $v$  unter der linearen Abbildung  $Df(x) \in \text{Hom}(\mathbb{R}^n, \mathbb{R}^p)$  denken, wie auch immer diese Gegenstände notiert sein mögen. Und so notieren wir, wie schon im ersten Abschnitt, meist die Vektoren aus  $\mathbb{R}^n$  als Zeilen, weil die chinesische Notation unbequem zu schreiben und zu lesen ist. In einem Symbol wie  $Df(x) \cdot v$  ist dann

$$v = (v_1, \dots, v_n), \quad Df(x) \cdot v = (Df_1(x) \cdot v, \dots, Df_p(x) \cdot v),$$

$$Df_i(x) \cdot v = \sum_{j=1}^n \partial f_i / \partial x_j(x) \cdot v_j.$$

Wo wir aber den Matrizenkalkül verwenden, sind Vektoren als Spalten-tupel zu schreiben, und das ist insbesondere der Fall, wenn wir das Differential  $Df(x)$  als Matrix auffassen und über Eigenschaften dieser Matrix, ihrer Zeilen und Spalten, reden. Übrigens bezeichne  ${}^tA$  die transponierte der Matrix  $A$ , dann können wir gelegentlich, wo es der Kalkül verlangt, eine Zeile durch Transponieren in eine Spalte  ${}^t(v_1, \dots, v_n)$  verwandeln.

Man hat etwas Mühe mit den Bezeichnungen, denn einerseits ist einem Punkt  $x$  die Matrix  $Df(x)$  zugeordnet, und andererseits ist ja auch dies eine (lineare) Abbildung, die dann einem Vektor  $v$  den Bildvektor  $Df(x) \cdot v$  zuordnet. Wir schreiben zur Unterscheidung diese letztere Abbildung als Multiplikation, weil wir uns  $Df(x)$  als Matrix denken. Andere schreiben  $x$  als Index, also  $Df_x$  oder  $D_x f$  statt  $Df(x)$ , und dann  $Df_x(v)$  oder  $D_x f(v)$  statt  $Df(x) \cdot v$ .

Die Kettenregel zum Beispiel erhält jetzt folgende Gestalt: Hat man eine Zusammensetzung

$$\begin{array}{ccc} \mathbb{R}^n & & \mathbb{R}^m \\ \cup & & \cup \\ U & \xrightarrow{f} & V \xrightarrow{g} \mathbb{R}^p, \quad f(u) = v, \end{array}$$

und existieren  $Df(u)$  und  $Dg(v)$ , so existiert auch  $D(g \circ f)(u)$ , und die Kettenregel besagt explizit:

$$\sum_j \frac{\partial g_i}{\partial y_j}(v) \cdot \frac{\partial f_j}{\partial x_k}(u) = \frac{\partial (g \circ f)_i}{\partial x_k}(u).$$

Dies kann man auch bei der Berechnung der Ableitung von Funktionen einer Variablen mit Gewinn anwenden. Betrachte zum Beispiel eine Zusammensetzung

$$\begin{array}{ccc} \mathbb{R} & & \mathbb{R}^n \\ \cup & & \cup \\ I & \xrightarrow{\gamma} & U \xrightarrow{f} \mathbb{R}. \end{array}$$

Dann ist  $\frac{d}{dt}(f \circ \gamma)(\tau) = Df(\gamma(\tau)) \dot{\gamma}(\tau) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\gamma(\tau)) \frac{d\gamma_j}{dt}(\tau)$ .

Kurz:

$$\frac{df}{dt} = \sum_{j=1}^n \frac{\partial f}{\partial x_j} \frac{dx_j}{dt}.$$

Ist zum Beispiel  $\gamma(t) = u + tv$  für ein  $u \in U$  und einem Vektor  $v \in \mathbb{R}^n$ , so ist

$$\frac{d}{dt}(f \circ \gamma)(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(u) v_i =: D_v f(u).$$

Ist  $f : U \rightarrow \mathbb{R}$  differenzierbar am Punkt  $u$ , so heißt der Vektor

$$\text{grad } f(u) := (D_1 f(u), \dots, D_n f(u))$$

auch der **Gradient** von  $f$  am Punkt  $u$ , und

$$D_v f(u) = \langle \text{grad } f(u), v \rangle$$

heißt die **Richtungsableitung** von  $f$  in Richtung  $v$ .

Ist  $|v| = 1$ , so ist  $|D_v f(u)| \leq |\text{grad } f(u)|$  nach der Schwarzschen Ungleichung, also ist  $|D_v f|$  maximal, falls  $v = \text{grad } f / |\text{grad } f|$ . Der Gradient zeigt die Richtung des stärksten Anstiegs der Funktion.

Die Existenz der Jacobischen  $(\partial f_i / \partial x_j(u))$  ist notwendig aber nicht hinreichend für die Differenzierbarkeit. Zum Beispiel die Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \frac{xy}{x^2 + y^2}, \quad f(0) := 0,$$

ist am Ursprung nicht einmal stetig, also erst recht nicht differenzierbar, sie hat dort aber die Jacobimatrix  $0$ , denn  $f|_{\mathbb{R} \times 0} = f|_{0 \times \mathbb{R}} = 0$ .

Es gilt aber folgender wichtiger

**(2.4) Satz.** Die Funktion  $f : U \rightarrow \mathbb{R}^p$  sei überall partiell differenzierbar, und die partiellen Ableitungen  $D_j f_i : U \rightarrow \mathbb{R}$  seien am

Punkte  $x \in U$  stetig, dann ist  $f$  bei  $x$  differenzierbar mit dem Differential

$$Df(x) = (\partial f_i / \partial x_j(x)).$$

**Beweis:** Wir dürfen  $p = 1$  annehmen. Wir schreiben  $h \in \mathbb{R}^n$  als  $h = h_1 e_1 + \dots + h_n e_n$ , und erhalten

$$\begin{aligned} & f(x+h) - f(x) \\ &= \sum_{k=1}^n (f(x+h_1 e_1 + \dots + h_k e_k) - f(x+h_1 e_1 + \dots + h_{k-1} e_{k-1})). \end{aligned}$$

Auf jeden Summanden können wir den Mittelwertsatz der Differentialrechnung einer Variablen — nämlich der  $k$ -ten Komponente im  $k$ -ten Summanden — anwenden und erhalten:

$$f(x+h) - f(x) = \sum_{k=1}^n h_k D_k f(x+h_1 e_1 + \dots + h_{k-1} e_{k-1} + \vartheta_k h_k e_k),$$

mit  $0 < \vartheta_k < 1$ . Und dies ist schon die Behauptung, denn es ist ja  $\lim_{h \rightarrow 0} D_k f(\dots) = D_k f(x)$ , weil  $D_k f$  bei  $x$  stetig ist.  $\square$

In der Regel werden wir also die Differenzierbarkeit einer Funktion dadurch feststellen, daß wir die partiellen Ableitungen berechnen — das läuft auf den Kalkül der Differentialrechnung einer Variablen hinaus — und feststellen, ob die partiellen Ableitungen stetig sind. Zum Beispiel ein Polynom in mehreren Variablen ist stets differenzierbar.

Mit diesem Satz überträgt man vieles unmittelbar vom Ein- auf Mehrdimensionale. Zum Beispiel die Grenzfunktion einer Folge stetig differenzierbarer Funktionen ist wieder stetig differenzierbar, wenn die Ableitungen gleichmäßig konvergieren.

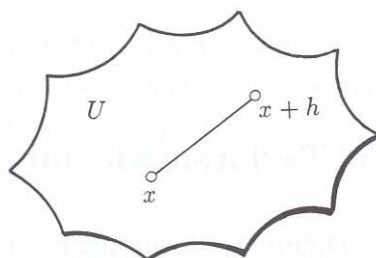
Aus dem Eindimensionalen überträgt sich auch, ähnlich wie im Beweis des Satzes benutzt, ein

**(2.5) Mittelwertsatz.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}$  sei eine differenzierbare Funktion. Es sei  $x + th \in U$  für  $0 \leq t \leq 1$ , dann ist

$$f(x+h) - f(x) = Df(x + \tau h) \cdot h, \quad 0 < \tau < 1.$$

**Beweis:** Setze  $g(t) = f(x + th)$  und wende den Mittelwertsatz für eine Variable an:

$$f(x + h) - f(x) = g(1) - g(0) = \frac{d}{dt}g(\tau) = Df(x + \tau h) \cdot h. \quad \square$$



Beachte jedoch, daß die Dimension des Bildraumes eins ist. Wenn  $f$  mehrere Komponenten hat,  $f = (f_1, \dots, f_p) : U \rightarrow \mathbb{R}^p$ , so muß man im allgemeinen für jede Komponente einen anderen Wert  $\tau$  nehmen. Zum Beispiel bei der Schraubenlinie  $\gamma(t) = (\cos t, \sin t, t)$  zeigt der Vektor  $\dot{\gamma}(t) = (-\sin t, \cos t, 1)$  niemals in Richtung  $\gamma(2\pi) - \gamma(0) = (0, 0, 2\pi)$ . Allgemein muß man, statt  $Df$  an einer Zwischenstelle, vielmehr einen Mittelwert von  $Df$  wählen, nämlich:

**(2.6) Mittelwertsatz.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}^p$  stetig differenzierbar. Auch sei  $x + th \in U$  für  $0 \leq t \leq 1$ , dann ist

$$f(x + h) - f(x) = \int_0^1 Df(x + th) dt \cdot h.$$

Das Integral wird komponentenweise gebildet. Das Ergebnis der Integration ist wieder ein Matrix, der Mittelwert von  $Df$  auf der Strecke von  $x$  nach  $x + h$ .

**Beweis:** Beide Seiten sind gleich  $\int_0^1 \frac{d}{dt} f(x + th) dt$ .  $\square$

Diese Formel hat die Gestalt  $f(x + h) = f(x) + A(x + h) \cdot h$ , mit

$$A(x + h) = \int_0^1 Df(x + th) dt.$$

Sie gibt also eine explizitere Beschreibung einer matrizenwertigen Funktion  $A$ , wie sie in der Definition der Differenzierbarkeit auftritt. Das wird sich noch oft als nützlich erweisen, wenn wir etwa den Zuwachs von  $f$  abschätzen oder die Differenzierbarkeit eines geeigneten  $A$  untersuchen wollen.

### § 3. Taylorentwicklung

Für den Kalkül der Differentialrechnung machen wir nun die wichtige Bemerkung, daß die höheren Ableitungen einer Funktion vertauschbar sind. Die genaue Behauptung ist:

**(3.1) Satz** (über die Vertauschbarkeit der Ableitungen). Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}$  eine Funktion; sie besitze Ableitungen  $D_i f$ ,  $D_j f$  nach der  $i$ -ten und  $j$ -ten Variable auf  $U$ , und es existiere die Ableitung  $D_i D_j f : U \rightarrow \mathbb{R}$  und sei stetig. Dann existiert auch  $D_j D_i f$ , und

$$D_i D_j f = D_j D_i f.$$

**Beweis:** Sei  $u \in U$ . Um das Schreiben zu vereinfachen setzen wir

$$\varphi(s, t) = f(u + se_i + te_j).$$

Dann sagt die Voraussetzung, daß lokal um den Ursprung  $(0, 0)$  die Ableitung  $D_1 D_2 \varphi$  existiert und stetig ist, und es ist zu zeigen:

$$D_1 D_2 \varphi(0, 0) = D_2 D_1 \varphi(0, 0).$$

Nach Definition ist nun

$$\begin{aligned} D_2 D_1 \varphi(0, 0) &= d/dt \Big|_{t=0} \lim_{s \rightarrow 0} \frac{\varphi(s, t) - \varphi(0, t)}{s} \\ &= \lim_{t \rightarrow 0} \lim_{s \rightarrow 0} \frac{1}{s} \cdot \frac{(\varphi(s, t) - \varphi(0, t)) - (\varphi(s, 0) - \varphi(0, 0))}{t}. \end{aligned}$$

Auf den Bruch, einen Differenzenquotienten bezüglich der zweiten Variablen, wenden wir den Mittelwertsatz der Differentialrechnung an und schreiben dafür

$$\frac{1}{s}(D_2\varphi(s, \vartheta_2 t) - D_2\varphi(0, \vartheta_2 t)), \quad 0 < \vartheta_2 < 1,$$

und dies ist wieder ein Differenzenquotient bezüglich der ersten Variablen, auf den wir, weil  $D_1 D_2$  existiert, ebenfalls den Mittelwertsatz anwenden können. Wir erhalten:

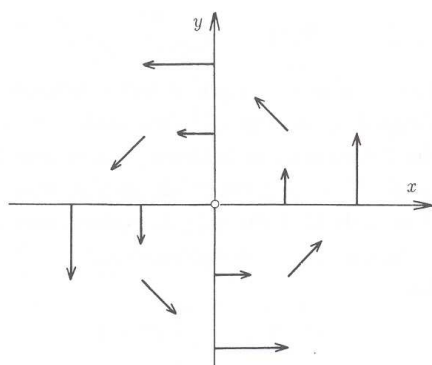
$$D_1 D_2 \varphi(\vartheta_1 s, \vartheta_2 t), \quad 0 < \vartheta_1, \vartheta_2 < 1,$$

und weil  $D_1 D_2 \varphi$  stetig ist, folgt

$$\lim_{t \rightarrow 0} \lim_{s \rightarrow 0} D_1 D_2 \varphi(\vartheta_1 s, \vartheta_2 t) = D_1 D_2 \varphi(0, 0). \quad \square$$

Wir geben sogleich eine Anwendung des Satzes: Gegeben sei eine Abbildung  $v : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , die wir uns als Vektorfeld vorstellen, das heißt, jedem Punkt  $(x, y) \in \mathbb{R}^2$  ist ein Vektor  $(v_1(x, y), v_2(x, y))$  zugeordnet, den wir mit Fußpunkt in  $(x, y)$  abtragen.

Beispiel.  $v(x, y) = \frac{1}{2}(-y, x)$ .



**Frage:** Gibt es zu dem Vektorfeld ein **Potential**, das heißt, gibt es eine Funktion  $P : \mathbb{R}^2 \rightarrow \mathbb{R}$ , so daß  $\partial P / \partial x = v_1$ ,  $\partial P / \partial y = v_2$ ? In

unserem Beispiel ist die Antwort: Nein, denn es würde folgen:

$$-\frac{1}{2} = \frac{\partial v_1}{\partial y} = \frac{\partial v_2}{\partial x} = \frac{1}{2}.$$

Nachdem wir nun wissen, daß die partiellen Ableitungen unter vernünftigen Voraussetzungen vertauschbar sind, machen wir uns folgende bequeme Schreibweise zu eigen: Wir setzen

$$D_i f := \frac{\partial}{\partial x_i} f, \\ D_i^k f := D_i D_i^{k-1} f := \partial^k / \partial x_i^k f, \quad D_i^0 f := f.$$

Dann schreiben wir Zusammensetzungen partieller Ableitungen mit einem **Multiindex**  $\alpha$ , wir benutzen folgende Bezeichnungen:

$$\begin{aligned} \alpha &= (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n \text{ heißt ein } \mathbf{Multiindex}, \\ |\alpha| &:= \alpha_1 + \dots + \alpha_n \quad \text{seine } \mathbf{Ordnung}, \\ \alpha! &:= \alpha_1! \cdot \alpha_2! \cdot \dots \cdot \alpha_n! \quad \mathbf{\alpha\text{-fakultät}}, \\ x^\alpha &:= x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n} \quad \text{heißt das } \mathbf{Monom} \text{ vom } \mathbf{Exponent} \ \alpha, \\ \beta \leq \alpha &: \iff \beta_i \leq \alpha_i \text{ für } i = 1, \dots, n, \quad \mathbf{kleinergleich}, \\ \alpha \pm \beta &:= (\alpha_1 \pm \beta_1, \dots, \alpha_n \pm \beta_n), \\ D^\alpha f &:= D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n} f := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} f, \quad D^0 f := f. \end{aligned}$$

Solange die Zusammensetzungen partieller Ableitungen von  $f$  auf der offenen Menge  $U \subset \mathbb{R}^n$  stetig bleiben, sind sie vertauschbar, und wir können die Reihenfolge so festlegen, wie in dem Symbol  $D^\alpha f$ . Eine Funktion  $f : U \rightarrow \mathbb{R}$  heißt  **$C^k$ -Funktion** oder  **$k$ -mal stetig differenzierbar**, falls  $D^\alpha f$  für  $|\alpha| \leq k$  existiert und stetig ist. Und  $C^k(U)$  sei die Menge der  $C^k$ -Funktionen auf  $U$ . Wir lassen auch  $k = \infty$  zu, also

$$C^\infty(U) = \bigcap_{k=0}^{\infty} C^k(U)$$

ist die Menge der beliebig oft stetig differenzierbaren Funktionen. Dies ist übrigens ein kommutativer Ring mit Eins.

Eine Abbildung  $f : U \rightarrow \mathbb{R}^p$  heißt  **$k$ -mal stetig differenzierbar**, wenn ihre Komponenten  $f_i : U \rightarrow \mathbb{R}$  aus  $C^k(U)$  sind. Die Menge der  $k$ -mal stetig differenzierbaren Abbildungen  $U \rightarrow \mathbb{R}^p$  wird mit

$$C^k(U, \mathbb{R}^p) = C^k(U)^p$$

bezeichnet. Ist  $f \in C^k(U)^p$ , so ist  $D^\alpha f$  das  $p$ -Tupel

$$D^\alpha f = (D^\alpha f_1, \dots, D^\alpha f_p).$$

Man rechnet leicht aus

$$D^\alpha x^\beta = \begin{cases} (\beta! / (\beta - \alpha)!) x^{\beta - \alpha} & \text{falls } \alpha \leq \beta, \\ 0 & \text{sonst,} \end{cases}$$

denn

$$\begin{aligned} D^\alpha x^\beta &= \left( \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} x_1^{\beta_1} \right) \cdot \dots \cdot \left( \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} x_n^{\beta_n} \right) \\ &= (\beta_1! / (\beta_1 - \alpha_1)!) x_1^{\beta_1 - \alpha_1} \cdot \dots \cdot (\beta_n! / (\beta_n - \alpha_n)!) x_n^{\beta_n - \alpha_n} \end{aligned}$$

falls  $\alpha \leq \beta$ , und Null sonst. Insbesondere folgt:

Im Punkte  $x = 0$  ist

$$D^\alpha x^\beta = \begin{cases} \alpha! & \text{falls } \alpha = \beta, \\ 0 & \text{sonst.} \end{cases}$$

Das sind dieselben Formeln, die schon früher am Anfang der Erklärung der Taylorentwicklung standen. Sie bedeuten jetzt etwas mehr als früher, aber der Kalkül macht keinen Unterschied.

**Definition (Jet).** Die Funktion  $f : U \rightarrow \mathbb{R}$  sei  $k$ -mal stetig differenzierbar. Der  **$k$ -Jet** oder das  **$k$ -te Taylorpolynom** von  $f$  bei  $u \in U$  ist das Polynom

$$j_u^k f(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(u)}{\alpha!} x^\alpha.$$

Ist  $k = \infty$ , so ist der **Jet** von  $f$  bei  $u$  die **Potenzreihe**

$$j_u^\infty f(x) = \sum_{|\alpha|=0}^{\infty} \frac{D^\alpha f(u)}{\alpha!} x^\alpha.$$

Ist  $f : U \rightarrow \mathbb{R}^p$  aus  $C^k(U)^p$ , so wird  $j_u^k f$  durch die gleiche Formel definiert. In diesem Fall ist jedes  $D^\alpha f$  ein  $p$ -Tupel, also  $j_u^k f$  das  $p$ -Tupel von Polynomen

$$j_u^k f = (j_u^k f_1, \dots, j_u^k f_p).$$

Wir nennen ein  $p$ -Tupel von Polynomen in  $n$ -Variablen kurzerhand wieder ein Polynom. Entsprechendes gilt für  $j_u^\infty f$  und Potenzreihen. Wie früher können wir den  $k$ -Jet von  $f$  folgendermaßen charakterisieren:

Das Polynom  $j^k f(x)$  hat am Punkte  $0 \in \mathbb{R}^n$  die gleichen Ableitungen der Ordnung  $\leq k$  wie die Abbildung  $f(u+x)$ , oder  $j^k f(x-u)$  hat bei  $u$  gleiche Ableitungen der Ordnung  $\leq k$  wie  $f$ . In der Tat ist nämlich für  $|\beta| \leq k$  am Punkte  $x = 0$

$$D^\beta \sum_{|\alpha| \leq k} \frac{D^\alpha f(u)}{\alpha!} x^\alpha = \sum_{|\alpha| \leq k} \frac{D^\alpha f(u)}{\alpha!} D^\beta x^\alpha = D^\beta f(u).$$

Diese Eigenschaft charakterisiert auch den  $k$ -Jet. Ist

$$p(x) = \sum_{|\alpha| \leq k} a_\alpha x^\alpha$$

ein Polynom vom Grad  $\leq k$ , und  $D^\beta p(0) = D^\beta f(u)$  für  $|\beta| \leq k$ , so ist  $a_\beta = D^\beta p(0)/\beta! = D^\beta f(u)/\beta!$ .

Wie im Fall der Dimension eins finden wir folgende

**(3.2) Rechenregeln.** Sind  $f, g \in C^k(U)$  und sind  $\lambda, \mu$  Konstanten, so ist

$$j_u^k(\lambda f + \mu g) = \lambda j_u^k f + \mu j_u^k g.$$

**Allgemeine Produktregel.**  $j_u^k(f \cdot g) = j_0^k(j_u^k f \cdot j_u^k g)$ .

**Allgemeine Kettenregel.** Hat man eine Zusammensetzung von  $C^k$ -Abbildungen

$$U \xrightarrow[g]{f} V \rightarrow \mathbb{R}^m, \quad g(u) = v,$$

so ist  $j_u^k(f \circ g) = j_0^k(j_u^k f \circ (j_u^k g - v))$ .

**Beweis:** Die erste Regel ist trivial, weil die Abbildungen

$$D^\alpha : C^k(U) \rightarrow C^{k-|\alpha|}(U)$$

linear sind. Die zweite folgt so: Setze

$$\begin{aligned} f(u+x) &= j_u^k f(x) + \varphi(x), \\ g(u+x) &= j_u^k g(x) + \psi(x). \end{aligned}$$

Dann ist

$$D^\alpha \varphi(0) = D^\alpha \psi(0) = 0 \quad \text{für } |\alpha| \leq k,$$

und

$$f \cdot g(u+x) = j_u^k f(x) \cdot j_u^k g(x) + \varphi(x) \cdot (\dots) + \psi(x) \cdot (\dots),$$

und man sieht sofort:  $D^\alpha(\varphi \cdot h)(0) = 0 = D^\alpha(\psi \cdot h)(0)$  für alle  $h \in C^k(U)$ . Also folgt die zweite Regel.

Im Beweis der dritten Regel wollen wir die Koordinaten in  $U$  und  $V$  so verschieben, daß  $u = v = 0$ , dann haben wir  $g(0) = 0$  und müssen zeigen

$$j_0^k(f \circ g) = j_0^k(j_0^k f \circ j_0^k g).$$

Ist  $j_0^k f = 0$ , also  $D^\alpha f(0) = 0$  für  $|\alpha| \leq k$ , so findet man durch Induktion nach  $|\alpha|$ , daß  $j_0^k(f \circ g) = 0$ , das heißt  $D^\alpha(f \circ g)(0) = 0$  für  $|\alpha| \leq k$ . Allgemein setzt man wieder  $f(x) = j_0^k f(x) + \varphi(x)$ , mit  $j_0^k \varphi = 0$ , und erhält

$$j_0^k(f \circ g) = j_0^k(j_0^k f \circ g + \varphi \circ g) = j_0^k(j_0^k f \circ g),$$

aber weil  $j_0^k f$  ein Polynom und der Jet mit Summen und Produkten nach den beiden ersten Regeln verträglich ist, so ist

$$j_0^k(j_0^k f \circ g) = j_0^k(j_0^k f \circ j_0^k g). \quad \square$$

Betrachten wir zum Beispiel die Zusammensetzung

$$D \xrightarrow{g} V \xrightarrow{f} \mathbb{R},$$

wobei  $D$  ein Intervall ist,  $g(s) = x + sh$ , und  $v = g(\tau)$ . Ist dann  $f \in C^k(V)$ , so ist für  $k \geq 1$

$$j_\tau^k g(t) = v + th, \quad \text{also} \quad j_\tau^k g - v = th,$$

also

$$j_\tau^k(f \circ g)(t) = j_v^k f(th) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(v)}{\alpha!} \cdot h^\alpha \cdot t^{|\alpha|},$$

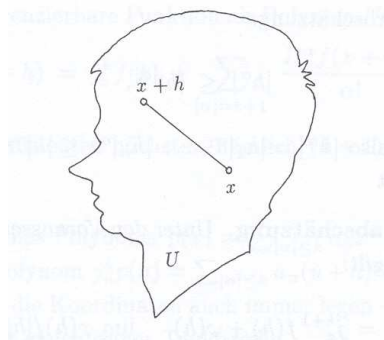
und ein Koeffizientenvergleich mit der Definition von  $j_\tau^k(f \circ g)$  zeigt explizit

$$(3.3) \quad \frac{1}{k!} \frac{d^k}{ds^k} \Big|_{s=\tau} f(x + sh) = \sum_{|\alpha|=k} \frac{D^\alpha f(x + \tau h) \cdot h^\alpha}{\alpha!}.$$

Man kann sich der Mühe unterziehen, dies direkt durch Induktion nach  $k$  zu zeigen.

Nachdem wir uns soweit über die Eigenschaften des Jets unterhalten haben, können wir unmittelbar die Taylorsche Formel auf Funktionen von  $n$  Variablen übertragen, die angibt, wie gut eine Funktion durch ihren  $k$ -Jet approximiert wird. Ist  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}$  eine Funktion, so erhalten wir die Taylorformel für  $f(x + h)$  aus der Taylorformel für Funktionen einer Veränderlichen, indem wir  $f$  auf die Verbindungsstrecke zwischen  $x$  und  $x + h$  ein-

schränken.



**(3.4) Taylorsche Formel.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f \in C^{k+1}(U)$ ; die Verbindungsstrecke

$$\{x + sh \mid 0 \leq s \leq 1\}$$

zwischen  $x$  und  $x+h$  sei ganz in  $U$  enthalten. Dann gilt für eine Zahl  $0 < \tau < 1$ :

$$f(x+h) = j_x^k f(h) + \sum_{|\alpha|=k+1} \frac{D^\alpha f(x+\tau h)}{\alpha!} h^\alpha.$$

**Beweis:** Sei  $\varphi(s) := f(x+sh)$ , dann sagt die Taylorformel mit der Restglieddarstellung von Lagrange

$$\varphi(1) = \sum_{j=0}^k \frac{1}{j!} \frac{d^j}{ds^j} \varphi(0) + \frac{1}{(k+1)!} \frac{d^{k+1}}{ds^{k+1}} \varphi(\tau), \quad 0 < \tau < 1,$$

und setzen wir die Ableitungen von  $\varphi(s) = f(x+sh)$  nach (3.3) mit  $\tau = 0$  bzw.  $\tau = \tau$  ein, so ergibt sich:

$$\varphi(1) = f(x+h) = \sum_{|\alpha|=0}^k \frac{D^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=k+1} \frac{D^\alpha f(x+\tau h)}{\alpha!} h^\alpha,$$

was die Behauptung ist. □

Natürlich überträgt man ähnlich auch andere Restglieddarstellungen auf höhere Dimension. Über die Größenordnung des Restgliedes belehrt uns die Abschätzung:

$$(3.5) \quad |h^\alpha| \leq |h|^{|\alpha|},$$

denn  $|h_i| \leq |h|$ , also  $|h^\alpha| = |h_1|^{\alpha_1} \cdots |h_n|^{\alpha_n} \leq |h|^{\alpha_1 + \cdots + \alpha_n} = |h|^{|\alpha|}$ .  
Wir finden damit

**(3.6) Restgliedabschätzung.** *Unter den Voraussetzungen der Taylorschen Formel gilt:*

$$f(x+h) = j_x^{k+1} f(x) + \varphi(h), \quad \lim_{h \rightarrow 0} \varphi(h)/|h|^{k+1} = 0.$$

**Beweis:** Es ist  $\varphi(h) = \sum_{|\alpha|=k+1} (D^\alpha f(x+\tau h) - D^\alpha f(x)) h^\alpha / \alpha!$ ,  $|h^\alpha / |h|^{k+1}| \leq 1$ , und  $\lim_{h \rightarrow 0} (D^\alpha f(x+\tau h) - D^\alpha f(x)) = 0$ .  $\square$

So ist der  $k$ -Jet in natürlicher Weise eine Verallgemeinerung der Ableitung. Daß  $f$  bei  $x$  differenzierbar ist, heißt:

$$f(x+h) = f(x) + A \cdot h + \varphi(h), \quad \lim_{h \rightarrow 0} \varphi(h)/|h| = 0.$$

Der  $k$ -Jet ist — statt der affinen Abbildung

$$h \mapsto f(x) + Ah = f(x) + a_1 h_1 + \cdots + a_n h_n,$$

was ja ein Polynom erster Ordnung in  $h$  ist — nun ein Polynom  $k$ -ter Ordnung, das auch  $f(x+h)$  von  $k$ -ter Ordnung approximiert, nämlich

$$f(x+h) = j_x^k f(x) + \varphi(h), \quad \lim_{h \rightarrow 0} \varphi(h)/|h|^k = 0.$$

Eine stetig differenzierbare, also differenzierbare Funktion kann man als Polynom erster Ordnung

$$f(x+h) = f(x) + A(h) \cdot h$$

schreiben, bei dem nur die Koeffizienten höchster — nämlich erster — Ordnung von  $h$  abhängen. Entsprechend schreibt sich eine  $(k+1)$ -mal stetig differenzierbare Funktion als Polynom  $(k+1)$ -ter Ordnung

$$f(x+h) = j_x^k f(h) + \sum_{|\alpha|=k+1} \frac{D^\alpha f(x+\tau h)}{\alpha!} h^\alpha,$$

wo nur die Koeffizienten höchster, nämlich  $(k+1)$ -ter Ordnung von  $h$  abhängen.

Der  $k$ -Jet eines Polynoms  $p(x) = \sum_{|\alpha| \leq k} a_\alpha x^\alpha$  im Punkte  $u$  ist natürlich das Polynom  $j_u^k p(h) = \sum_{|\alpha| \leq k} a_\alpha (u+h)^\alpha$ , insbesondere — so kann man ja die Koordinaten auch immer legen —  $j_0^k p(x) = p(x)$ . Ein Polynom ist gleich seiner Taylorreihe.

#### § 4. Das lokale Verhalten einer Funktion

Die Formel  $j_u^k f(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(u)}{\alpha!} x^\alpha$  ist in allgemeinen Überlegungen leicht zu handhaben, weil man wie im eindimensionalen Fall rechnen kann. Man tut aber gut, auch etwas expliziter aufzuschreiben, was sie bedeutet, zum Beispiel:

$$\begin{aligned} j_u^2 f(x) &= f(u) + \sum_{i=1}^n D_i f(u) x_i + \frac{1}{2} \sum_{i,j} D_i D_j f(u) x_i x_j \\ &= f(u) + \sum_{i=1}^n \frac{\partial}{\partial x_i} f(u) x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} (u) x_i x_j. \end{aligned}$$

Die  $\alpha$  mit  $|\alpha| = 2$  sind nämlich  $\alpha = (0, \dots, 0, 2, 0, \dots, 0)$ ,  $\alpha! = 2$ , und die mit zwei Einsen  $\alpha = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$ ,  $\alpha! = 1$ , und letztere sind in der Summe  $\sum_{i,j}$  zweimal aufgeführt, an den Stellen  $(i, j)$  und  $(j, i)$ , treten aber in  $\sum_{|\alpha|=2}$  nur einmal auf.

Das Studium des Zweijets einer Funktion genügt meistens, um das lokale Verhalten der Funktion um einen Punkt aufzuklären.

**Definition.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}$  eine stetig differenzierbare Funktion. Ein Punkt  $u \in U$  heißt **kritisch**, wenn

$Df(u) = 0$ , und in diesem Fall heißt  $f(u)$  ein **kritischer Wert** von  $f$ .

Offenbar ist  $u$  genau dann kritisch, wenn die Richtungsableitungen  $D_v f(u)$  für alle  $v$  verschwinden, d.h. wenn  $\frac{d}{dt} f \circ \gamma(0) = 0$  für jede stetig differenzierbare Kurve  $t \mapsto \gamma(t)$  mit  $\gamma(0) = u$ .

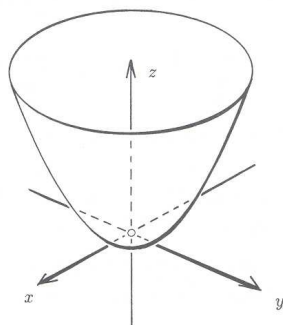
**Definition.** Der Punkt  $u \in U$  heißt ein **lokales Maximum** von  $f$ , wenn es ein  $\varepsilon > 0$  gibt, so daß  $f(u + h) \leq f(u)$  für  $|h| < \varepsilon$ . Ein **lokales Minimum** ist analog mit  $\geq$  statt  $\leq$  definiert, und ein **lokales Extremum** ist ein lokales Maximum oder Minimum.

Hat eine stetig differenzierbare Funktion  $f : U \rightarrow \mathbb{R}$  bei  $u \in U$  ein lokales Extremum, so ist  $u$  kritisch. Offenbar hat dann nämlich insbesondere  $f \circ \gamma$  für jede stetig differenzierbare Kurve  $\gamma$  mit  $\gamma(0) = u$  ein lokales Extremum an der Stelle 0, also  $D(f \circ \gamma)(0) = 0$ .

Umgekehrt braucht ein kritischer Punkt kein Extremum zu sein. Im Gegensatz zum eindimensionalen Fall findet man jetzt eine reichere Geometrie des lokalen Verhaltens einer Funktion, die nicht nur durch die Begriffe von Monotonie und Extremalität beschrieben wird. Typisch sind folgende

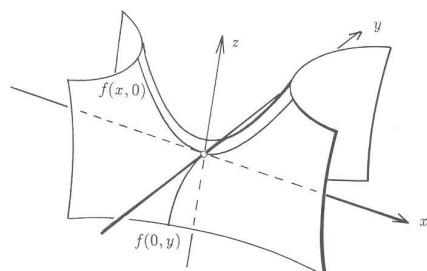
**Beispiele.**

$$f(x, y) = x^2 + y^2$$



Diese Funktion zeigt ein (lokales) Minimum am Nullpunkt.

$$f(x, y) = x^2 - y^2.$$



Hier ist der Ursprung ein kritischer Punkt, aber kein Extremum, sondern ein sogenannter **Sattelpunkt**.

Und schließlich hat die Funktion

$$f(x, y) = -x^2 - y^2,$$

das Negative der ersten, ein (lokales) Maximum am Ursprung, man erhält ihren Graphen, wenn man das erste Bild auf den Kopf stellt.

Zu den Bildern wollen wir uns folgende Begriffe machen, um das lokale Verhalten einer Funktion zu beschreiben:

**Definition.** Sei  $U$  offen in  $\mathbb{R}^n$ , und  $f : U \rightarrow \mathbb{R}$  eine  $C^2$ -Funktion, dann nenne ich  $f$  bei  $u \in U$  lokal **positiv** (negativ) **definit**, falls es ein  $\varepsilon > 0$  und  $\lambda > 0$  gibt, so daß gilt:

$$f(u + h) - f(u) \geq \lambda|h|^2 \quad (\text{bzw. } \leq -\lambda|h|^2), \text{ für } |h| < \varepsilon.$$

Der Punkt  $u$  heie **Sattelpunkt** von  $f$ , falls es eine orthogonale Zerlegung  $\mathbb{R}^n = V \oplus W$  gibt, so da  $\dim V \neq 0, n$ , und so da  $f|_{u+V}$  lokal positiv definit,  $f|_{u+W}$  lokal negativ definit ist.

Eine bei  $u$  lokal positiv definite Funktion hat dort ein lokales Minimum, eine lokal negativ definite Funktion hat ein lokales Maximum,

und ein Sattel hat weder das eine noch das andere. Meistens, wenn auch nicht immer, wird eine Funktion in einem kritischen Punkt eine der drei genannten Verhaltensweisen zeigen, und man erkennt das lokale Verhalten an dem Term zweiter Ordnung der Taylorentwicklung.

**Definition** (Hessesche). Sei  $u$  ein kritischer Punkt der  $C^2$ -Funktion  $f : U \rightarrow \mathbb{R}$ , dann heißt die symmetrische Matrix

$$H_u = (\partial^2 f / \partial x_i \partial x_j(u)), \quad i, j = 1, \dots, n,$$

die **Hessematrix** von  $f$  bei  $u$ , und die Abbildung

$$x \mapsto H_u(x) := {}^t x H_u x := \sum_{i,j} (\partial^2 f / \partial x_i \partial x_j(u)) x_i x_j$$

heißt die **Hesseform** von  $f$  bei  $u$ .

Das  ${}^t$  bedeutet "transponiert". Hier schreiben wir die Vektoren als Spalten- $n$ -tupel.

**(4.1) Satz** (über das lokale Verhalten). Sei  $U$  offen in  $\mathbb{R}^n$  und  $u$  ein kritischer Punkt der  $C^2$ -Funktion  $f : U \rightarrow \mathbb{R}$ . Die Determinante der Hessematrix in  $u$  sei ungleich 0. Dann hat  $f$  bei  $u$  dasselbe lokale Verhalten wie die Hesseform  $H_u$  von  $f$  am Ursprung. Ist also die Hesseform  $x \mapsto {}^t x H x$  positiv (negativ) definit, so gilt dasselbe lokal für  $f$  um  $u$ , und hat die Hesseform einen Sattel, so auch  $f$ , mit derselben Zerlegung  $\mathbb{R}^n = V \oplus W$ .

**Erläuterung.** Wir dürfen nach Verschiebung der Koordinaten annehmen:  $u = 0$ , und  $f(0) = 0$ . Ist  $A$  ein linearer Endomorphismus von  $\mathbb{R}^n$ , so ist

$$j_0^2(f \circ A)(x) = j_0^2 f(Ax) = \frac{1}{2} {}^t(Ax) H A x = \frac{1}{2} {}^t x ({}^t A H A) x,$$

wenn  $H$  die Hessematrix von  $f$  bei 0 ist. Man erkennt, daß  $f \circ A$  die Hessematrix  ${}^t A H A$  hat. Nun lehrt die lineare Algebra: Durch Wahl einer geeigneten orthonormalen Abbildung  $A$  kann man  $H$  auf Diagonalgestalt transformieren:

$$H = \text{Diag}(\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_\ell, 0, \dots, 0)$$

sei die quadratische Matrix mit den genannten Koeffizienten  $\lambda_i > 0$  und  $\mu_j < 0$  und 0 in der Diagonale und verschwindenden Koeffizienten außerhalb der Diagonale. Sei dann  $\lambda = \min\{\lambda_i\}$ ,  $\mu = \min\{-\mu_j\}$ .

Nach orthonormaler Transformation des Koordinatensystems findet man daher die koordinatenweise Zerlegung

$$\mathbb{R}^n = V \oplus W \oplus N = \mathbb{R}^k \oplus \mathbb{R}^\ell \oplus \mathbb{R}^m,$$

so daß  $H|V$  positiv definit und  $H|W$  negativ definit ist, nämlich für  $x \in V$  ist  $H(x) = \sum_{i=1}^k \lambda_i x_i^2 \geq \lambda|x|^2$ , und entsprechend für  $W$ ; und auf  $N$  verschwindet  $H(x)$ . Ist also  $\det(H) \neq 0$ , so ist  $\mathbb{R}^n = V \oplus W$ , und es gilt:

**Ergänzung.** Ist  $V$  der Teilraum, auf dem die Hesseform positiv definit ist (also  $f$  ist auf  $u + V$  positiv definit), so ist  $\dim V$  gleich der Anzahl der positiven Eigenwerte von  $H$ .

Ist  $\det(H) = 0$ , so kann man  $f$  auf  $V \oplus W$  einschränken, und das lokale Verhalten von  $f|V \oplus W$  ist dasselbe wie das der Hesseform auf  $V \oplus W$ .

Für das Verhalten der Funktion auf dem Unterraum  $N$ , wo die Hesseform verschwindet, lehrt der Satz nichts. Beachte auch, da zwar die Dimensionen von  $V$  und  $W$ , nicht aber diese Räume selbst durch die Eigenschaften in der Definition eines Sattelpunktes eindeutig bestimmt sind.

**Beweis** (4.1): Die Taylorformel lehrt, falls  $u = f(u) = Df(u) = 0$ :

$$f(x) = {}^t x A x + |x|^2 \varphi(x), \quad \lim_{x \rightarrow 0} \varphi(x) = 0, \quad \text{mit } A = \frac{1}{2} H.$$

Wir denken uns nun die Zerlegung  $\mathbb{R}^n = V \oplus W$  gegeben, dann existiert ein  $\lambda > 0$ , so daß für alle  $v \in V$  gilt:

$${}^t v A v \geq \lambda |v|^2, \quad \text{also} \quad {}^t v A v \geq \lambda$$

für alle  $v \in V$  mit  $|v| = 1$ . Weil nun  $\lim_{x \rightarrow 0} \varphi(x) = 0$ , so folgern wir, daß es zu jedem  $\lambda_1 < \lambda$  ein  $\varepsilon > 0$  gibt, so daß

$${}^t v A v + \varphi(x) \geq \lambda_1 \quad \text{für} \quad |v| = 1, \quad |x| < \varepsilon.$$

Ersetzen wir  $v$  durch  $x/|x|$  und multiplizieren mit  $|x|^2$ , so erhalten wir

$${}^t x A x + |x|^2 \varphi(x) \geq \lambda_1 |x|^2 \quad \text{für } x \in V \quad \text{mit } 0 < |x| < \varepsilon.$$

Das heißt aber  $f(x) \geq \lambda_1 |x|^2$  für  $x \in V$  mit  $|x| < \varepsilon$ . Entsprechend schließt man für  $W$ .  $\square$

Um also Rechner auf den rechten Weg zu bringen: Will man das lokale Verhalten in einem kritischen Punkt bestimmen, so berechne man die Vorzeichen der Eigenwerte der Hessematrix. Dazu braucht man die Eigenwerte selbst nicht auszurechnen, was auch im allgemeinen sehr aufwendig wäre. Ein gangbarer Weg ist die Diagonalisierung der Hesseform durch simultane Zeilen- und Spaltenumformungen.

Im Falle der Funktionen in 2 Variablen zeigt die Betrachtung einer Hessematrix in Diagonalgestalt unmittelbar:

$\det(H) < 0 \implies$  Der Punkt ist ein Sattelpunkt.

$\det(H) > 0 \implies$  Der Punkt ist ein lokales Extremum.

Ob das Extremum ein Maximum oder Minimum ist, erkennt man dann natürlich am Vorzeichen der Diagonalelemente der Hessematrix. Dazu braucht man nicht zu transformieren.

**Beispiel.** Zu bestimmen ist das lokale Verhalten der Funktion

$$f(x, y) = (3x + 4y + \sin(xy))(2x + y - \cos(x)(1 - \cos y))$$

am Nullpunkt. Wir berechnen den Zwei-Jet:

$$j_0^2 f(x, y) = (3x + 4y)(2x + y),$$

denn  $\sin(xy)$  und  $(1 - \cos y)$  verschwinden schon von mindestens zweiter Ordnung, also haben sie keinen Einfluß auf den 2-Jet des Produkts. Ist nun  $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  durch

$$A: (x, y) \mapsto (\xi, \eta), \quad \xi = 3x + 4y, \quad \eta = 2x + y,$$

gegeben, so ist  $j_0^2 f(x, y) = p(A(x, y))$ , mit  $p(\xi, \eta) = \xi \cdot \eta$ . Also haben wir  $j_0^2(f \circ A^{-1}) = xy$ . Daher hat  $j_0^2 f \circ A^{-1}$  einen Sattel, also auch  $j_0^2 f$ , also auch  $f$ .

Das Polynom  $xy$  wird übrigens durch  $x = \xi - \eta$ ,  $y = \xi + \eta$  in  $\xi^2 - \eta^2$  transformiert.

Die Behandlung des Beispiels zeigt, daß man durch etwas Überlegung viel Rechnung sparen kann.

Wir werden später das sogenannte Morselemma kennenlernen, das in gewissem Sinne eine Verschärfung des hier Bewiesenen enthält: Eine genügend oft differenzierbare Funktion sieht lokal um einen kritischen Punkt mit nicht entarteter Hesseform nach geeigneter Koordinatentransformation ebenso aus, wie die Hesseform für diesen Punkt (V, § 3 und Bd. 3, III, 2.5).

## § 5. Vertauschbarkeit von Ableitung und Integral

Da wir über Integration noch wenig wissen, sagen wir zum Thema erst etwas Vorläufiges, das uns unmittelbar hilfreich für die Differentialrechnung ist (vergl. III, 4.7).

**(5.1) Satz.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $D = [a, b]$  ein kompaktes Intervall. Die Funktion  $f : D \times U \rightarrow \mathbb{R}$  sei stetig, dann ist die Funktion

$$F : U \rightarrow \mathbb{R}, \quad x \mapsto \int_a^b f(t, x) dt$$

stetig. Hat  $f$  stetige Ableitungen  $\partial/\partial x_i f : D \times U \rightarrow \mathbb{R}$ , so ist auch  $F$  stetig nach  $x_i$  differenzierbar, und

$$\frac{\partial}{\partial x_i} \int_a^b f(t, x) dt = \int_a^b \frac{\partial}{\partial x_i} f(t, x) dt.$$

**Beweis:** Die Menge  $D \times \{x\} \subset D \times U$  ist kompakt, und daher ist  $f$  auf  $D \times \{x\}$  gleichmäßig stetig (Bd. 1, VI, 7.12). Also gibt es zu  $\varepsilon > 0$  stets ein  $\delta > 0$ , so daß

$$|f(t, x+h) - f(t, x)| < \varepsilon/|b-a| \quad \text{für } |h| < \delta.$$

Daher  $|F(x+h) - F(x)| = \left| \int_a^b (f(t, x+h) - f(t, x)) dt \right| \leq \int_a^b \varepsilon/|b-a| dt = \varepsilon$  für  $|h| < \delta$ . Das ist die Stetigkeit. Jetzt setze für festes  $x \in U$  und genügend kleine  $h \in \mathbb{R}$

$$g(t, x, h) = \begin{cases} \frac{f(t, x + he_i) - f(t, x)}{h} & \text{falls } h \neq 0 \\ \partial/\partial x_i f(t, x) & \text{falls } h = 0. \end{cases}$$

Dann ist  $g$  stetig in allen Variablen. Das ist nur in den Punkten nicht klar, wo  $h = 0$ , aber nach dem Mittelwertsatz der Differentialrechnung schreibt sich der Differenzenquotient als

$$\frac{\partial}{\partial x_i} f(t, x + \vartheta_h \cdot he_i),$$

was für  $(x, t, h) \rightarrow (x_0, t_0, 0)$  gegen  $\partial/\partial x_i f(x_0, t_0, 0)$  geht. Nach dem ersten Fall ist also

$$\partial/\partial x_i F(x) = \lim_{h \rightarrow 0} \int_a^b g(t, x, h) dt = \int_a^b g(t, x, 0) dt = \int_a^b \partial/\partial x_i f(t, x) dt,$$

und dies hängt stetig von  $x$  ab. □

Beachte, daß in dem Beweis nichts über  $U$  benutzt wird; das hängt an der zitierten angemessen gefaßten Aussage, daß stetige Funktionen auf kompakten Mengen gleichmäßig stetig sind.

Hat die Funktion  $f$  stetige Ableitungen der Ordnung  $\leq k$  nach den  $x_i$ , so sieht man induktiv

$$D^\alpha F(x) = D^\alpha \int_a^b f(t, x) dt = \int_a^b D^{0, \alpha} f(t, x) dt \quad \text{für } |\alpha| \leq k,$$

mit  $0, \alpha = (0, \alpha_1, \dots, \alpha_n)$ .

Wir geben gleich eine Anwendung, die eigentlich eine wesentliche Aussage über die algebraische Struktur des Ringes der differenzierbaren Funktionen macht:

**(5.2) Satz.** Sei  $U$  offen in  $\mathbb{R}^n$  und  $f : U \rightarrow \mathbb{R}$  eine  $C^k$ -Funktion, dann ist für alle  $u$  und  $x$ , so daß  $u + tx \in U$  für  $0 \leq t \leq 1$ ,

$$f(u + x) - f(u) = \sum_{i=1}^n \varphi_i(x) \cdot x_i$$

mit  $C^{k-1}$ -Funktionen  $\varphi_i$ .

**Beweis:** Nach dem Mittelwertsatz (2.6) setze

$$\varphi_i(x) := \int_0^1 D_i f(u + tx) dt. \quad \square$$

Die Aussage geht insoweit über die Definition der Differenzierbarkeit hinaus, als die durch  $(\varphi_1, \dots, \varphi_n)$  definierte lineare Abbildung eben nicht nur an der Stelle  $x = 0$  stetig ist, sondern als Funktion von  $x$  sogar  $(k - 1)$ -mal stetig differenzierbar — für  $k = \infty$  also auch beliebig oft differenzierbar.

In Aufgabe 12 zu diesem Kapitel weisen wir auf eine Verallgemeinerung dieser Aussage für die höheren Restglieder der Taylorentwicklung hin, die man leicht ebenso, oder durch Induktion aus (5.2) erhält.