denn die j-te Spalte von A enthält ja gerade die Koeffizienten des Vektors Ae_j bezüglich der Standardbasis.

Die vom Standardskalarprodukt induzierte Norm ist die euklidische Norm,

$$\|x\|^2 = \langle x, x \rangle = x_1^2 + .. + x_n^2.$$

Falls Verwechslungsgefahr mit anderen Normen und Skalarprodukten besteht, schreiben wir hierfür genauer $\|\cdot\|_e$ und $\langle \cdot, \cdot \rangle_e$.

Diese Situation bezeichnen wir im Folgenden als den *Standardfall*.

14.2 Totale Ableitung

Wir betrachten zunächst weiterhin Abbildungen zwischen beliebigen Banachräumen. Da es vorerst nur um lokale Aspekte geht, verzichten wir auf eine explizite Bezeichnung des Definitionsbereichs und schreiben

$$f : V \hookrightarrow W,$$

wenn es eine nichtleere offene Teilmenge $\Omega \subset V$ gibt, so dass $f : \Omega \to W$. Zu jedem Punkt im Definitionsbereich von f existiert also eine offene Umgebung, auf der f definiert ist. Dies ist die für uns zur Zeit wesentliche Eigenschaft des Definitionsbereiches einer Abbildung.

Die Normen auf V und W bezeichnen wir nun kürzer mit $|\cdot|_V$ und $|\cdot|_W$, oder noch einfacher $|\cdot|$, wenn der Bezug aus dem Zusammenhang klar ist. Die Doppelstriche $\|\cdot\|$ reservieren wir für die Operatornorm.

4 **Definition** *Eine Abbildung $f : V \hookrightarrow W$ heißt im Punkt a total differenzierbar oder kurz differenzierbar, wenn es eine stetige lineare Abbildung $L : V \to W$ gibt, so dass*

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - Lh|_W}{|h|_V} = 0.$$

In diesem Fall heißt L die totale Ableitung von f im Punkt a und wird mit $Df(a)$ bezeichnet. ⋊

Bemerkungen *a.* Die Eindeutigkeit dieser Ableitung und damit die Berechtigung der Bezeichnung $Df(a)$ zeigen wir gleich $_5$.

b. Diese Definition beinhaltet unsere früheren Definitionen der Differenzierbarkeit. Im Fall *einer* unabhängigen Variablen, also $V = \mathbb{R}$, ist f eine Kurve in W. Eine lineare Abbildung $L : \mathbb{R} \to W$ wird durch einen Vektor $v \in W$ dargestellt, und wir erhalten die Charakterisierung der Ableitung $Df(a) = \dot{f}(a)$ entsprechend dem zweiten Differenzierbarkeitssatz 13.1. Ist auch $W = \mathbb{R}$, so wird dieser Vektor durch eine reelle Zahl dargestellt, und wir erhalten die Ableitung einer reellen Funktion, $Df(a) = f'(a)$, einer Variable wie im ersten Differenzierbarkeitssatz 8.1.

c. Ist V von endlicher Dimension, so ist *jede* lineare Abbildung $A\colon V \to W$ stetig und diese Forderung redundant. Dies ist der einzige, aber wesentliche Unterschied zwischen dem endlich- und unendlich-dimensionalen Fall. ⊸

▪ Die Landausymbole

Das Arbeiten mit Grenzwerten wie in der letzten Definition lässt sich mithilfe der *Landausymbole* erheblich vereinfachen.

Definition *Sei $\dot{\Omega}$ eine punktierte Umgebung von $0 \in V$ und $f, g\colon \dot{\Omega} \to W$ zwei Abbildungen mit $g \neq 0$ auf $\dot{\Omega}$. Dann ist*

$$f(h) = O(g(h)) \quad :\Leftrightarrow \quad \sup_{h \in \dot{\Omega}} \frac{|f(h)|}{|g(h)|} < \infty,$$

und man sagt, f ist von der Ordnung groß-O von g auf Ω. Ferner ist

$$f(h) = o(g(h)) \quad :\Leftrightarrow \quad \lim_{h \to 0} \frac{|f(h)|}{|g(h)|} = 0,$$

und man sagt, f ist von der Ordnung klein-o von g für $h \to 0$. ⋊

Es ist also f groß-O von g auf $\dot{\Omega}$, wenn es eine Konstante M gibt, so dass

$$|f(h)| \leqslant M\,|g(h)|, \qquad h \in \dot{\Omega}.$$

Sie ist klein-o von g, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass

$$|f(h)| \leqslant \varepsilon\,|g(h)|, \qquad 0 < |h| < \delta.$$

Die Landausymbole stehen übrigens nicht für eine bestimmte Funktion, sondern für *Klassen* von Funktionen. Eine Formulierung wie $O(h) + O(h) = O(h)$ ist daher zulässig und auch üblich.

Rechenregeln *Für das Rechnen mit den Landausymbolen gilt:*

(i) *Eine Linearkombination von $O(h)$-Abbildungen ist wieder $O(h)$.*
(ii) *Daselbe gilt mit $o(h)$ anstelle von $O(h)$.*
(iii) *Eine $o(h)$-Abbildung ist auch $O(h)$.*
(iv) *Sind f und g beide $O(h)$, so ist auch $(f \circ g)(h) = O(h)$.*
(v) *Ist sogar $f(h) = o(h)$ oder $g(h) = o(h)$, so ist $(f \circ g)(h) = o(h)$.* ⋊

⟪⟪ Der Beweis ist als Übung überlassen. ⟫⟫

A. Es ist $\sin t = O(t)$ und $1 - \cos t = O(t^2)$.
B. Für eine beschränkte lineare Abbildung A gilt $|Ah| = O(h)$.
C. Ist $\langle\cdot,\cdot\rangle$ eine bilineare Form auf V und $A \in L(V)$, so ist

$$\langle Ah, h\rangle = O(\langle h, h\rangle) = o(h).$$

D. Für $f \in C^{n+1}(I)$ ist aufgrund der Restgliedformel von Lagrange

$$f(a+h) = T_a^n f(h) + O(h^{n+1}).$$

Mit den Landausymbolen können wir Differenzierbarkeit in einem Punkt auf folgende handliche Art charakterisieren.

5 **Satz** *Eine Abbildung $f: V \hookrightarrow W$ ist im Punkt a differenzierbar genau dann, wenn es ein $L \in L(V, W)$ gibt, so dass*

$$f(a+h) = f(a) + Lh + o(h).$$

Die Ableitung ist dann $Df(a) = L$, und sie ist eindeutig bestimmt. Außerdem ist f im Punkt a auch stetig. ⋊

⟪⟪ Die erste Behauptung folgt aus den Definitionen der Differenzierbarkeit und des Landausymbols. Ist $\Lambda: V \to W$ eine weitere lineare Abbildung mit der geforderten Eigenschaft, also

$$f(a+h) = f(a) + \Lambda h + o(h),$$

so ergibt die Subtraktion beider Gleichungen $(L - \Lambda)(h) = o(h)$. Dies ist für eine lineare Abbildung nur möglich, wenn $L - \Lambda = 0$ A-5. ⟫⟫

Somit ist eine Abbildung $f: V \hookrightarrow W$ in a *differenzierbar*, wenn sie dort *linear approximierbar* ist. — Nun definieren wir Differenzierbarkeit *in jedem Punkt* wie üblich.

Definition *Eine Abbildung $f: V \hookrightarrow W$ heißt total differenzierbar, wenn sie in jedem Punkt ihres Definitionsbereiches total differenzierbar ist.* ⋊

Die Ableitung von f definiert in diesem Fall eine Abbildung

$$Df: V \hookrightarrow L(V, W), \quad x \mapsto Df(x)$$

in den Vektorraum $L(V, W)$ aller linearen Abbildungen von V nach W. Ist $\dim V = 1$, also f eine Kurve, so ist $L(V, W) \simeq W$ und $Df: V \hookrightarrow W$ wiederum ein Kurve wie f. Ist dagegen $\dim V > 1$, so ist eine solche Identifikation nicht mehr möglich, und Df ist von einem wesentlich anderen Typ als die Abbildung f selbst. Mehr dazu am Ende von Abschnitt 6.

6 ⊳ A. *Affine Abbildungen:* Eine affine Abbildung

$$f: V \to W, \quad f(x) = Ax + b$$

mit $A \in L(V,W)$ und $b \in W$ ist in jedem Punkt differenzierbar. Denn es gilt

$$f(a+h) = A(a+h) = f(a) + Ah,$$

der o-Term verschwindet also, und wir erhalten $Df(a) = A$ in jedem Punkt a, Die Ableitung als *Abbildung*

$$Df: V \to L(V,W), \quad a \mapsto Df(a) = A,$$

ist somit *konstant.*

B. *Quadratische Formen:* Ist $\langle\cdot,\cdot\rangle$ eine bilineare Form auf V und $A \in L(V)$, so nennt man

$$f: V \to \mathbb{R}, \quad f(x) = \langle Ax, x\rangle$$

eine *quadratische Form* auf V. Dabei kann man ohne Beschränkung der Allgemeinheit annehmen, dass A *symmetrisch* ist, also

$$\langle Ax, y\rangle = \langle x, Ay\rangle, \qquad x, y \in V,$$

gilt A-7. Dann folgt

$$\begin{aligned} f(x+h) &= \langle A(x+h), x+h\rangle \\ &= \langle Ax, x\rangle + \langle Ax, h\rangle + \langle Ah, x\rangle + \langle Ah, h\rangle \\ &= f(x) + 2\langle Ax, h\rangle + o(h). \end{aligned}$$

Also ist f differenzierbar, und $Df(x)$ ist die lineare Abbildung $h \mapsto 2\langle Ax, h\rangle$. Die Ableitung selbst ist die Abbildung

$$Df: V \to L(V,\mathbb{R}), \quad x \mapsto Df(x) = 2\langle Ax, \cdot\rangle. \quad ◅$$

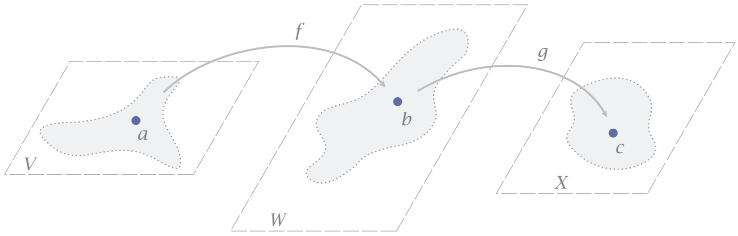
■ Kettenregel

Man sieht sofort, dass Differenziation eine lineare Operation ist. Sind also $f, g: V \hookrightarrow W$ im Punkt a differenzierbar, so auch $\lambda f + \mu g: V \hookrightarrow W$, und es gilt

$$D(\lambda f + \mu g)(a) = \lambda Df(a) + \mu Dg(a).$$

Interessanter ist die Verknüpfung zweier differenzierbarer Abbildungen. Das Ergebnis ist die allgemeine

Abb 1 Zur Kettenregel



7 **Kettenregel** *Ist $f\colon V \hookrightarrow W$ im Punkt a und $g\colon W \hookrightarrow X$ im Punkt $f(a)$ differenzierbar, so ist auch $g \circ f\colon V \hookrightarrow X$ im Punkt a differenzierbar, und es gilt*

$$D(g \circ f)(a) = Dg(f(a))Df(a). \quad ⋊$$

Die Ableitung von $g \circ f$ ist also die Verknüpfung der beiden linearen Abbildungen

$$Dg(b): \ W \to X, \qquad Df(a): \ V \to W,$$

wobei $b = f(a)$. Offensichtlich kommt es hierbei auf die Reihenfolge der Faktoren an, denn für $V \neq X$ wäre eine umgekehrte Verknüpfung der linearen Abbildungen erst gar nicht definiert.

⟨⟨⟨⟨⟨ Der Beweis ist praktisch identisch zum eindimensionalen Fall $_{8.5}$. Aufgrund der Voraussetzungen ist f in einer Umgebung von a und g in einer Umgebung von $b = f(a)$ definiert. Für hinreichend kleine h und k gilt somit

$$\begin{aligned} f(a+h) &= f(a) + Ah + o(h), \qquad & A &= Df(a), \\ g(b+k) &= g(b) + Bk + o(k), \qquad & B &= Dg(b). \end{aligned}$$

Ist h hinreichend klein, so ist auch $Ah + o(h) = O(h)$ hinreichend klein und g für $k = Ah + o(h)$ wohldefiniert. Mit dieser Wahl von k erhalten wir

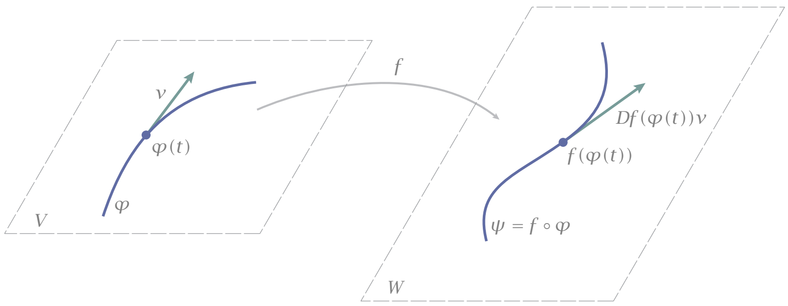
$$\begin{aligned} (g \circ f)(a+h)) &= g(f(a) + Ah + o(h)) \\ &= g(f(a)) + B(Ah + o(h)) + o(Ah + o(h)). \end{aligned}$$

Nun zeigt man noch, dass $Bo(h) = o(h)$ und $o(Ah + o(h)) = o(h)$. Also ist

$$(g \circ f)(a+h)) = (g \circ f)(a) + BAh + o(h).$$

Somit $_5$ ist $g \circ f$ im Punkt a differenzierbar mit Ableitung BA, was der Behauptung entspricht. ⟩⟩⟩⟩⟩

Abb 2 Die Kurve $f \circ \varphi$ und ihre Ableitung



8 ⊳ A. *Lineare Abbildung:* Ist $f : V \hookrightarrow W$ differenzierbar und $A : W \to X$ eine stetige lineare Abbildung, so ist auch

$$Af : V \hookrightarrow X$$

differenzierbar, und für jeden Punkt a im Definitionsbereich ist [6]

$$D(Af)(a) = (DA)(f(a))\, Df(a) = ADf(a).$$

B. *Abbildung einer Kurve:* Ist $\varphi : I \to V$ eine differenzierbare Kurve in V und $f : V \to W$ differenzierbar, so ist

$$f \circ \varphi : I \to W$$

eine differenzierbare Kurve in W mit Ableitung

$$(f \circ \varphi)^{\cdot}(t) = Df(\varphi(t))\dot{\varphi}(t), \qquad t \in I.$$

Im Falle einer Geraden φ, mit $\varphi(t) = a + tv$, ist insbesondere

$$(f \circ \varphi)(t)^{\cdot}\Big|_{t=0} = f(a + tv)^{\cdot}\Big|_{t=0} = Df(a)v.$$

C. *Quadratische Form entlang einer Kurve:* Ist $\varphi : I \to V$ eine differenzierbare Kurve in V und

$$f : V \to R, \quad f(x) = \langle Ax, x \rangle$$

eine quadratische Form auf V [6] mit symmetrischem, beschränktem A, so ist

$$\Phi = f \circ \varphi : I \to \mathbb{R}, \quad \Phi(t) = \langle A\varphi(t), \varphi(t) \rangle$$

eine differenzierbare reelle Funktion einer Variablen. Ihre Ableitung ist

$$\Phi'(t) = Df(\varphi(t))\dot{\varphi}(t) = 2\,\langle A\varphi(t), \dot{\varphi}(t) \rangle.$$

Speziell für eine Gerade $\varphi(t) = a + tv$ ist $\Phi'(0) = 2\,\langle Aa, v \rangle$. ⊲