1. Smooth Mappings Between Normed Spaces

This introductory chapter reviews properties of smooth functions $f : V \to W$ between normed spaces. Even though for our purposes the study of functions $f : \mathbb{R}^n \to \mathbb{R}^m$ would suffice, virtually no extra effort is required to deal with the more general situation. Moreover, even in the finite-dimensional setting a coordinate-free approach is conceptually clearer than writing down everything in terms of coordinates. Let us begin with the concept of differentiability. The usual definition of differentiability for functions $f : \mathbb{R} \to \mathbb{R}$ carries over *verbatim* to functions $f : \mathbb{R} \to W$ with values in a normed space. (Here and in the sequel a normed space always means a vector space over the field of real numbers.)

(1.1) Definition. *Let W be a normed space and let $I \subseteq \mathbb{R}$ be an open set (typically an interval). A function $f : I \to W$ is called* **differentiable** *at a point $x_0 \in I$ if the* **derivative** $f'(x_0) := \lim_{x \to x_0} \big(f(x) - f(x_0)\big)/(x - x_0)$ *exists. Note that this means*

$$\left\| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right\| \;\to\; 0 \quad \text{as } x \to x_0.$$

(1.2) Examples. (a) It is readily checked that if $W = \mathbb{R}^n$ (so that $f(x) = \big(f_1(x), \ldots, f_n(x)\big)^T$ with component functions $f_i : I \to \mathbb{R}$) then $f : I \to \mathbb{R}^n$ is differentiable if and only if each of the component functions f_i is differentiable, and $f'(x) = \big(f_1'(x), \ldots, f_n'(x)\big)^T$. For example, if we define $f : \mathbb{R} \to \mathbb{R}^4$ by $f(x) := (x^2, x^3, x^5, \sin x)^T$ then f is differentiable at each point $x \in \mathbb{R}$ with $f'(x) = (2x, 3x^2, 5x^4, \cos x)^T$.

(b) To see an example with W infinite-dimensional, we choose $W = C[a, b]$ (with given real numbers $a < b$), equipped with the maximum norm. We consider the mapping $f : \mathbb{R} \to C[a, b]$ given by $f(x)(u) := \sin(xu)$ and write f_x instead of $f(x)$ to psychologically ease the fact that each value $f(x) = f_x$ of f is itself a function which can be applied to an argument $u \in [a, b]$. Given a fixed number $x_0 \in \mathbb{R}$, we let $g_{x_0}(u) := (\partial/\partial x)|_{x=x_0} f_x(u) = (\partial/\partial x)|_{x=x_0} \sin(xu) = u\cos(x_0 u)$. For all $x \neq x_0$ in $\mathbb{R}$ and all $u \in [a, b]$, the Mean Value Theorem assures us that there are numbers ξ between x_0 and x and η between x_0 and ξ such that

$$\begin{aligned} \left| \frac{f_x(u) - f_{x_0}(u)}{x - x_0} - g_{x_0}(u) \right| \;&=\; \left| \frac{\sin(xu) - \sin(x_0 u)}{x - x_0} - u\cos(x_0 u) \right| \\ =\; |\, u\cos(\xi u) - u\cos(x_0 u)\,| \;&=\; |-u^2 \sin(\eta u)(\xi - x_0)| \;\leq\; u^2 |x - x_0| \end{aligned}$$

which, since this estimate is valid uniformly for all $u \in [a, b]$, implies that

$$\left\| \frac{f_x - f_{x_0}}{x - x_0} - g_{x_0} \right\| \;\leq\; \max(a^2, b^2) \cdot |x - x_0| \to 0 \text{ as } x \to x_0.$$

This shows that f is differentiable at x_0 with $f'(x_0) = g_{x_0}$. ■

Definition (1.1) does not carry over to functions $f : V \to W$ where $\dim(V) > 1$ because then the expression $\big(f(x) - f(x_0)\big)/(x - x_0)$ does not make sense. Hence we give a characterization of differentiability which lends itself to generalization. Namely, we will show that a mapping $f : \mathbb{R} \to W$ is differentiable at a point x_0 if and only if f can be "well approximated" by a continuous affine function $\alpha : \mathbb{R} \to W$ in a neighborhood of x_0. Recall that an affine function $\alpha : V \to W$ is a mapping of the form $\alpha(v) = A(v) + b$ with a linear mapping $A : V \to W$ and a fixed vector $b \in W$.

(1.3) Theorem. *Consider a mapping $f : I \to W$ where I is an open subset of $\mathbb{R}$ (typically an interval) and where W is a normed space. Then f is differentiable at $x_0 \in V$ if and only if there is a continuous affine mapping $\alpha : \mathbb{R} \to W$ with $\alpha(x_0) = f(x_0)$ such that $\|f(x) - \alpha(x)\|/|x - x_0| \to 0$ as $x \to x_0$.*

Proof. If f is differentiable at $x = x_0$, define $\alpha : \mathbb{R} \to W$ by $\alpha(x) := f(x_0) + (x - x_0)f'(x_0)$. This affine mapping is obviously continuous (because $\|\alpha(x_1) - \alpha(x_2)\| = |x_1 - x_2|\,\|f'(x_0)\|$ for all $x_1, x_2 \in \mathbb{R}$) and satisfies

$$\frac{f(x) - \alpha(x)}{x - x_0} \;=\; \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \;\to\; 0 \quad \text{for } x \to x_0.$$

Conversely, assume that there is an affine mapping α with the above properties. Such a mapping is necessarily of the form $\alpha(x) = f(x_0) + (x - x_0)w$ with a fixed vector $w \in W$. We then have

$$\frac{f(x) - f(x_0)}{x - x_0} - w \;=\; \frac{f(x) - \alpha(x)}{x - x_0} \;\to\; 0 \quad \text{for } x \to x_0,$$

which shows that f is differentiable at x_0 with $f'(x_0) = w$. ■

Somewhat sloppily, we say that f is "linearizable" around x_0 if there is an affine mapping with the indicated properties, i.e., if f can be "well approximated" by a continuous affine mapping in a neighborhood of x_0.

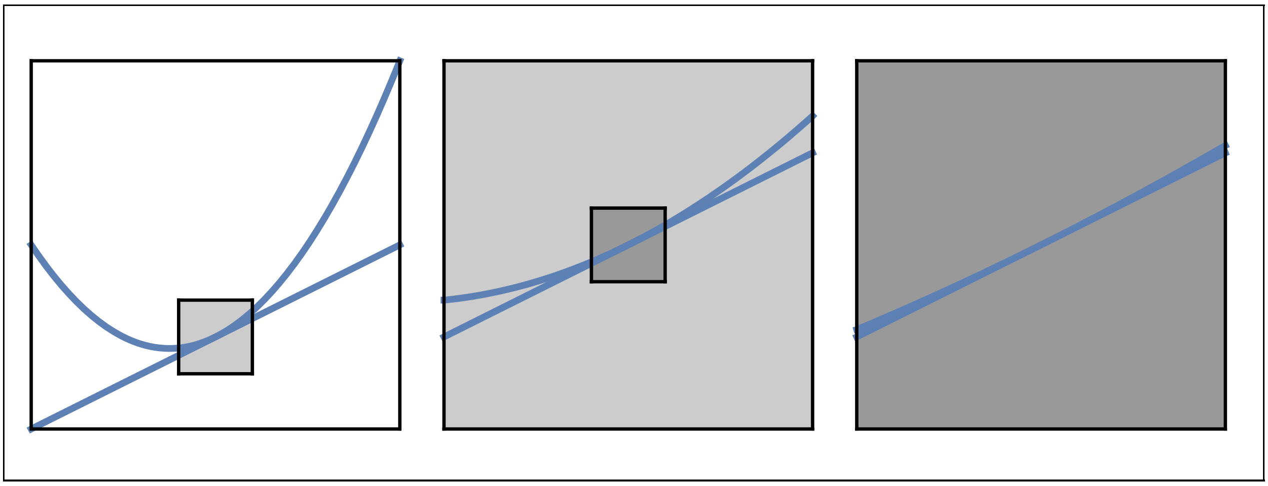

Figure 1.1: Differentiability as local linearizability in the case of a mapping $\mathbb{R} \to \mathbb{R}$.

The characterization of differentiability in Theorem (1.3) makes sense for mappings between arbitrary normed spaces and will be made into a definition in this general setting. Thus we *define* the property of differentiability as linearizability.

(1.4) Definition. *Let V and W be normed spaces and let $\Omega \subseteq V$ be an open subset of V. A mapping $f : \Omega \to W$ is called* **differentiable** *at a point $p \in \Omega$ if there is a continuous affine mapping $\alpha : V \to W$ with $\alpha(p) = f(p)$ which approximates f so well in a neighborhood of p that for $x \to p$ the difference $r(x) := f(x) - \alpha(x)$ tends faster to zero than $x - p$ in the sense that $\|r(x)\|/\|x-p\| \to 0$ as $x \to p$. The mapping f is simply called differentiable if it is differentiable at each point $p \in \Omega$.*

The fact that for $x \to p$ the difference $r(x) := f(x) - \alpha(x)$ (which is the remainder term if we approximately replace f by α) tends faster to zero than the argument difference $x - p$ means geometrically that f can be the less distinguished from the affine mapping α the smaller neighborhoods of p we consider.

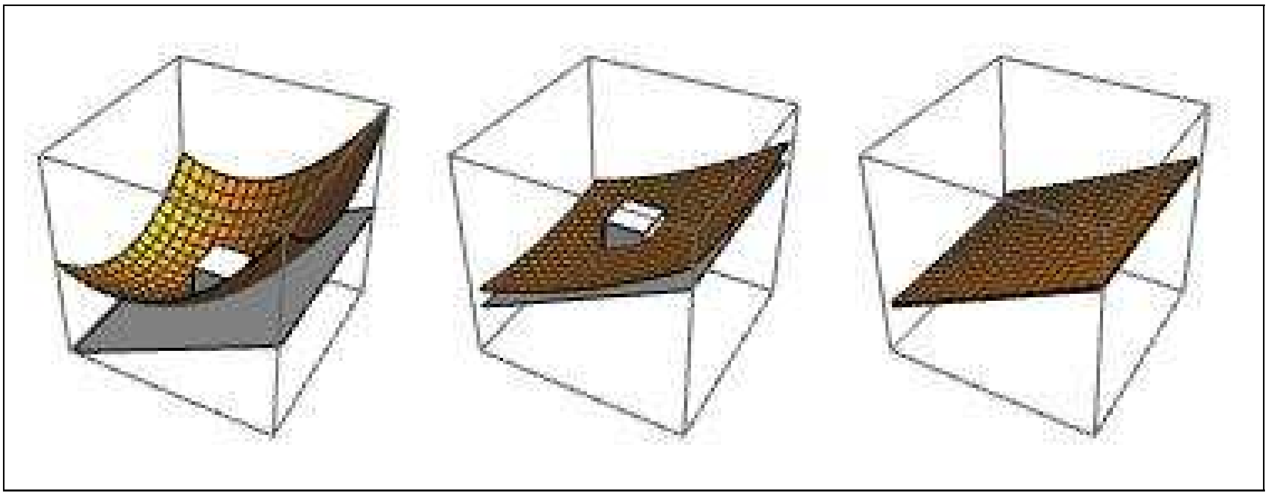

Figure 1.2: Differentiability as linearizability in the case of a mapping $\mathbb{R}^2 \to \mathbb{R}$.

We note that if an affine mapping with the indicated properties exists, then it is uniquely determined.

(1.5) Remark. *Assume that $\alpha : V \to W$ is a continuous affine mapping such that $\alpha(p) = f(p)$ and $\|f(x) - \alpha(x)\|/\|x-p\| \to 0$ as $x \to p$. Then $\alpha(x) = f(p) + A(x-p)$ where A is a continuous linear mapping which satisfies*

$$(\star) \qquad A(v) \;=\; \lim_{t\to 0} \frac{f(p+tv) - f(p)}{t} \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f(p+tv) \;=:\; (\partial_v f)(p).$$

Clearly, this condition determines A uniquely, since it specifies the effect of A on any vector v. We will write $f'(p)$ for this linear mapping A. Note that $(\star)$ shows that applying $f'(p)$ to a vector v yields the directional derivative of f at the point p in the direction v.

Proof. Write $\alpha(x) = f(p) + A(x-p)$. Fixing $v \neq 0$ and applying the condition $\|f(x) - \alpha(x)\|/\|x-p\| \to 0$ as $x \to p$ with $x = p + tv$ and $t \to 0$, we see that

$$\frac{\|f(p+tv) - f(p) - t\,A(v)\|}{t\|v\|} \to 0 \text{ as } t \to 0$$

which implies that $\big(f(p+tv) - f(p)\big)/t - A(v) \to 0$ as $t \to 0$, which is the claim. ∎

This shows that if f is differentiable at p then the directional derivative $(\partial_v f)(p)$ exists for each $v \in V$, and $(\partial_v f)(p)$ depends linearly and continuously on v.

(1.6) Counterexample. We define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) \;:=\; \frac{x^2 y}{x^2+y^2} \quad \text{if } (x,y) \neq (0,0) \quad \text{and} \quad f(0,0) := 0.$$

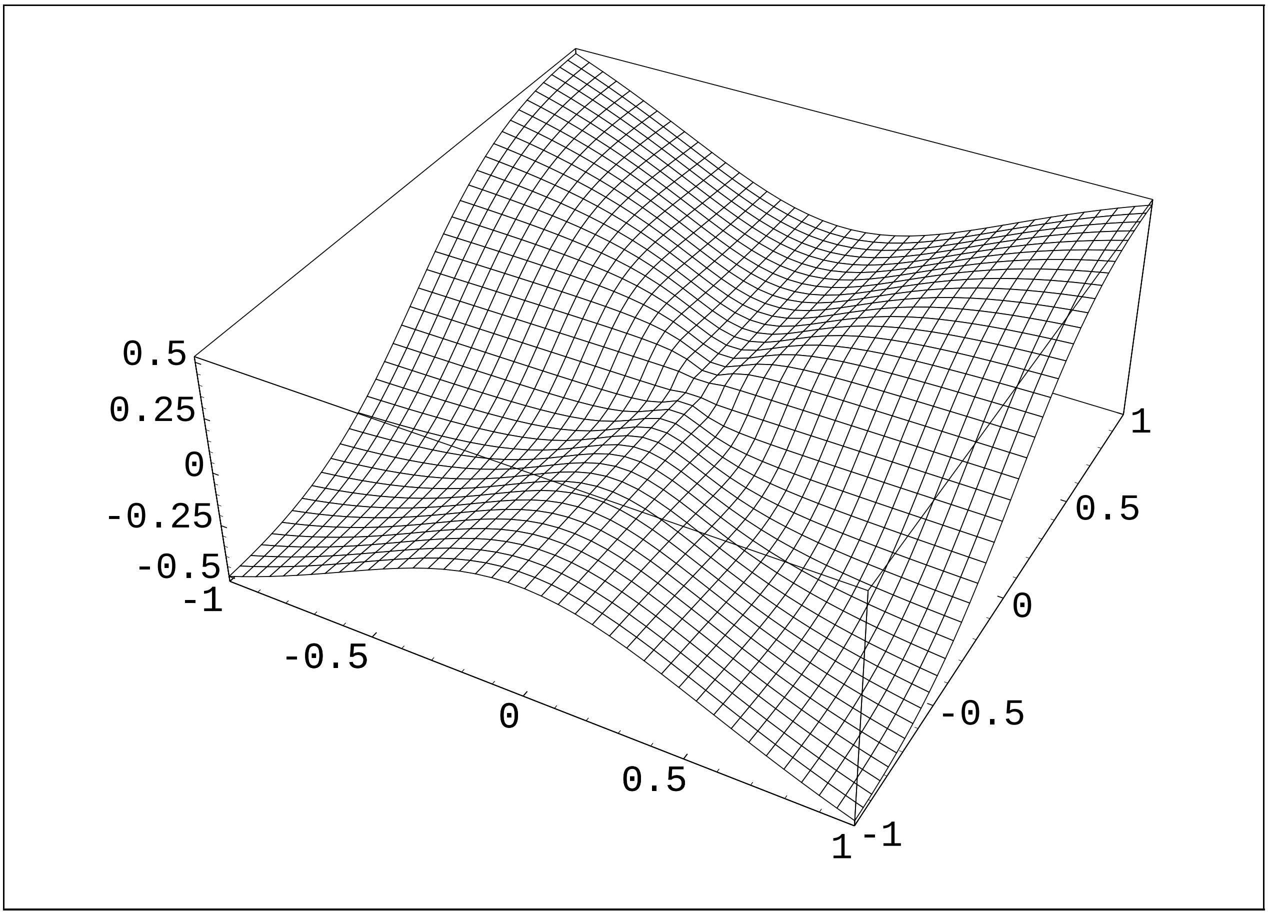

Figure 1.3: Graph of the function $f(x,y) = x^2y/(x^2+y^2)$.

The fact that f is continuous at the origin $(0,0)$ follows immediately from the estimate $|f(x,y)| = \big(x^2/(x^2+y^2)\big) \cdot |y| \leq 1 \cdot |y| = |y|$. We claim that if $t \mapsto x(t)$ and $t \mapsto y(t)$ are differentiable at $t_0 = 0$ with $\big(x(0), y(0)\big) = (0,0)$, then $\varphi(t) := f\big(x(t), y(t)\big)$ is differentiable at $t_0 = 0$. (Thus f is differentiable along each differentiable curve through $(0,0)$; in particular, all directional derivatives $(\partial_v f)(0,0)$ exist.) To see this, let us write $\xi(t) := x(t)/t$ and $\eta(t) := y(t)/t$. For all t with $\big(x(t),y(t)\big) \neq (0,0)$ we then have $\varphi(t)/t = \xi(t)^2\eta(t)/\big(\xi(t)^2+\eta(t)^2\big)$, which, because of $\xi(t) \to \dot{x}(0) =: v_1$ and $\eta(t) \to \dot{y}(0) =: v_2$, implies that

$$\dot{\varphi}(0) \;=\; \left\{\begin{array}{ll} v_1^2 v_2/(v_1^2+v_2^2) & \text{if } (v_1,v_2)\neq(0,0) \\ 0 & \text{if } (v_1,v_2)=(0,0) \end{array}\right\} \;=\; f(v_1,v_2).$$

Writing $p := (0,0)$ and $\alpha(t) := \big(x(t),y(t)\big)$, this shows that $(f\circ\alpha)'(0) = f\big(\alpha'(0)\big)$.) In particular, letting $\alpha(t) := p + tv = tv$, we see that $(\partial_v f)(p) = f(v)$ for all $v \in \mathbb{R}^2$. Thus all directional derivatives exist, but they do not depend linearly on the direction. Hence f is not differentiable at p. (The slight "crinkles" in the graph of f which can be seen in Figure 1.3 indicate a lack of smoothness of f at $p = (0,0)$, which, however, cannot be detected by following individual smooth curves through p.) ■

(1.7) Remark. An affine mapping $\alpha : V \to W$ is automatically continuous if V is finite-dimensional. Hence explicitly requiring the continuity of α in Definition (1.4) is relevant only if V is of infinite dimension. Moreover, in the case of a mapping $f : V \to W$ between finite-dimensional vector spaces (which we can always identify with a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ after choosing bases of V and W) it does not matter which norms we choose for V and W, because any two norms on a finite-dimensional vector space are equivalent.

(1.8) Examples. (a) Let V and W be real vector spaces of finite dimensions n and m and let $f : V \to W$ be differentiable at a point $p \in V$. If $\widehat{f} : \mathbb{R}^n \to \mathbb{R}^m$ is the coordinate representation of f with respect to bases $(v_1,\ldots,v_n)$ of V and $(w_1,\ldots,w_m)$ of W and if $p = p_1v_1 + \cdots + p_nv_n$, then the matrix representation of $f'(p) : V \to W$ with respect to these bases is the Jacobi matrix

$$\begin{bmatrix} \dfrac{\partial\widehat{f}_1}{\partial x_1}(p_1,\ldots,p_n) & \cdots & \dfrac{\partial\widehat{f}_1}{\partial x_n}(p_1,\ldots,p_n) \\ \vdots & & \vdots \\ \dfrac{\partial\widehat{f}_m}{\partial x_1}(p_1,\ldots,p_n) & \cdots & \dfrac{\partial\widehat{f}_m}{\partial x_n}(p_1,\ldots,p_n) \end{bmatrix}$$

whose entries are the partial derivatives of the component functions of f with respect to the coordinate variables used for V. (The existence of all of these partial derivatives is necessary, but not sufficient, for the differentiability of f at p. On the other hand, as we shall see later, the existence and continuity of all of these partial derivatives in a neighborhood of p is sufficient, but not necessary, for the differentiability of f at p.)

(b) Let $a < b$ be fixed real numbers and consider the mapping $\Phi : C[a,b] \to C[a,b]$ given by $\Phi(f) = f^2$. Clearly, $\Phi(f+h) - \Phi(f) = (f+h)^2 - f^2 = 2fh + h^2$. Since $\|h^2\| = \max_{a\le x\le b} |h(x)^2| = \big(\max_{a\le x\le b}|h(x)|\big)^2 = \|h\|^2$, this implies that

$$\frac{\|\Phi(f+h)-\Phi(f)-2fh\|}{\|h\|} \;=\; \frac{\|h^2\|}{\|h\|} \;=\; \|h\| \;\to\; 0 \text{ as } h \to 0.$$

Thus Φ is differentiable at each "point" $f \in C[a,b]$ with $\Phi'(f)h = 2fh$ for all $h \in C[a,b]$.

(c) Let $V = C^1[a,b]$ with the norm $\|y\| := \|y\|_\infty + \|y'\|_\infty$. We define $I : V \to \mathbb{R}$ by

$$I[y] \;:=\; \int_a^b 2\pi y(x)\sqrt{1+y'(x)^2}\,\mathrm{d}x.$$

(Thus I assigns to each C^1-function $y : [a,b] \to \mathbb{R}$ the area of the surface of revolution which is obtained by rotating the graph of f about the x-axis.) We want to show that I is differentiable at each "point" $y \in V$. To do so, we invoke the Taylor expansion of the function $u \mapsto \sqrt{1+u^2}$ to find a representation

$$I[y+h] = 2\pi \int_a^b \big(y(x)+h(x)\big) \left[\sqrt{1+y'(x)^2} + \frac{y'(x)h'(x)}{\sqrt{1+y'(x)^2}} + \frac{h'(x)^2/2}{\big(1+(y'(x)+\theta_x h'(x))^2\big)^{3/2}} \right] \mathrm{d}x$$

where θ_x is a number (depending on x) between 0 and 1. With y being fixed, this reads $I[y+h] = I[y] + A(h) + R(h)$ where

$$A(h) \;:=\; 2\pi \int_a^b \left(\sqrt{1+y'(x)^2}\cdot h(x) + \frac{y(x)y'(x)}{\sqrt{1+y'(x)^2}} \cdot h'(x) \right) \mathrm{d}x$$

and

$$R(h) \;:=\; 2\pi \int_a^b \left(\frac{y'(x)h(x)h'(x)}{\sqrt{1+y'(x)^2}} + \frac{\big(y(x)+h(x)\big)h'(x)^2/2}{\big(1+(y'(x)+\theta_x h'(x))^2\big)^{3/2}} \right) \mathrm{d}x.$$

We note that the mapping $h \mapsto A(h)$ is linear and continuous, whereas $|R(h)|/\|h\| \to 0$ as $\|h\| \to 0$. Since $I[y+h] = I[y] + A(h) + R(h)$, this shows that I is differentiable at y with $I'(y)h = A(h)$ for all $h \in V$. (Those who are acquainted with the calculus of variations will recognize the steps towards deriving the Euler-Lagrange equation which y must satisfy to minimize $I[y]$ subject to fixed boundary conditions $y(a) = y_a$ and $y(b) = y_b$.)

(d) Let $V := C^1_{2\pi}$ the space of all 2π-periodic C^1-functions $y : \mathbb{R} \to \mathbb{R}$, equipped with the norm $\|y\| := \|y\|_\infty + \|y'\|_\infty$, and let $W := C^0_{2\pi}$ be the space of all 2π-periodic continuous functions $y : \mathbb{R} \to \mathbb{R}$, equipped with the maximum norm. Let $p : \mathbb{R} \to \mathbb{R}$ be a C^2-function. We claim that the differential operator $D : V \to W$ given by $D(y) := \dot{y} + p \circ y$ (so that $(Dy)(t) = \dot{y}(t) + p\big(y(t)\big)$ for all $t \in \mathbb{R}$) is differentiable at each "point" $y \in V$. Thus let $y \in V$ be fixed. Then for any function $h \in V$, we invoke the Mean Value Theorem for the function p to obtain a representation

$$\begin{aligned} D[y+h](t) &= \dot{y}(t) + \dot{h}(t) + p\big(y(t)+h(t)\big) \\ &= \dot{y}(t) + \dot{h}(t) + p\big(y(t)\big) + p'\big(y(t)\big)h(t) + (1/2)\,p''\big(\xi(t)\big)h(t)^2 \\ &= \dot{y}(t) + p\big(y(t)\big) + \dot{h}(t) + p'\big(y(t)\big)h(t) + (1/2)\,p''\big(\xi(t)\big)h(t)^2 \end{aligned}$$

where $\xi(t)$ lies between $y(t)$ and $y(t)+h(t)$. Thus letting $A[h] := \dot{h} - (p' \circ y)h$, we see that $\|D[y+h] - D[y] - A[h]\|_\infty \le C\|h\|^2$ with a suitable constant C (because p'' is continuous and hence bounded on compact intervals). Since the assignment $h \mapsto A[h]$ is linear and continuous, we see that D is differentiable at y with $D'(y)[h] = \dot{h} + (p' \circ y)h$ for all $h \in V$.

(e) A continuous affine mapping $\alpha : V \to W$, say $\alpha(v) := w_0 + f(v)$ with a fixed vector $w_0 \in W$ and a continuous linear mapping $f : V \to W$, is everywhere differentiable with $\alpha'(p) = f$ for all $p \in V$. In particular, each constant mapping $\alpha : V \to W$ is differentiable at each point $p \in V$, with $\alpha'(p) = \mathbf{0}$ being the zero map from V to W for each $p \in V$, and each continuous linear mapping $f : V \to W$ is everywhere differentiable with $f'(p) = f$ for all $p \in V$.

(f) As a consequence of (1.3), a function $f : \mathbb{R} \to W$ with values in a normed space W ist differentiable at a point $x_0 \in \mathbb{R}$ in the sense of (1.1) if and only if it is differentiable at x_0 in the sense of (1.4). The definitions (1.1) and (1.4) for the derivative $\alpha'(t_0) =: w$ coincide, if we identify a vector $w \in W$ with the linear mapping $\lambda \mapsto \lambda w$. (For $W = \mathbb{R}^n$ this simply amounts to identifying a column vector of length n with an $(n \times 1)$-Matrix.)

(g) If $p \in \Omega_1 \subseteq \Omega_2 \subseteq V$ with open subsets Ω_i of a normed space V and if $f : \Omega_2 \to W$ is differentiable at p, then the restriction g of f to Ω_1 is also differentiable at p, with $g'(p) = f'(p)$. ■

The next two examples are so important that they deserve getting their own numbers and being upgraded to the status of "theorems".

(1.9) Theorem. *Let $V_1, \ldots, V_n, W$ be normed spaces and let $\beta : V_1 \times \cdots \times V_n \to W$ be a continuous n-linear mapping (where $V := V_1 \times \cdots \times V_n$ is equipped with the norm $\|(v_1, \ldots, v_n)\|_V := \sum_{i=1}^n \|v_i\|_{V_i}$). Then β is differentiable at each point $v = (v_1, \ldots, v_n) \in V$, and for all $h = (h_1, \ldots, h_n)$ we have*

$$\beta'(v)(h) \;=\; \sum_{i=1}^n \beta(v_1, \ldots, v_{i-1}, h_i, v_{i+1}, \ldots, v_n)\,.$$

Proof. Since β is n-linear we have $\beta(v+h) = \sum_{k=0}^n \beta_k(v,h)$ where

$$\begin{aligned}
\beta_0(v,h) &= \beta(v_1, \ldots, v_n) \;=\; \beta(v),\\
\beta_1(v,h) &= \sum_{i=1}^n \beta(v_1, \ldots, v_{i-1}, h_i, v_{i+1}, \ldots, v_n),\\
\beta_2(v,h) &= \sum_{i<j} \beta(v_1, \ldots, h_i, \ldots, h_j, \ldots, v_n),\\
\beta_3(v,h) &= \sum_{i<j<k} \beta(v_1, \ldots, h_i, \ldots, h_j, \ldots, h_k, \ldots, v_n),
\end{aligned}$$

and so on. Since β is continuous, there is a constant $C > 0$ such that

$$\|\beta(x_1, \ldots, x_n)\| \le C\|x_1\| \cdots \|x_n\| \quad \text{for all } x_i \in V_i\,.$$

This estimate implies that for all $k \ge 2$ the condition $\|\beta_k(v,h)\|_W / \|h\|_V \to 0$ holds for $h \to 0$, whence $\beta(v+h) = \beta(v) + \beta_1(v,h) + r$ where $\|r\|/\|h\| \to 0$ for $h \to 0$. Since $h \mapsto \beta_1(v,h)$ is a linear mapping, this shows that β is differentiable at v with $\beta'(v)h = \beta_1(v,h)$. ■

(1.10) Theorem. *Let X be a Banach space (i.e., a complete normed space) and let $\mathbb{B}(X,X)$ be the space of all bounded linear mappings $X \to X$, equipped with the operator norm $\|\!|T|\!\| := \|T\|_{\text{op}} = \sup\{\|Tx\| \mid \|x\| \leq 1\}$. We denote by $\text{GL}(X)$ the set of all continuous linear mappings $T : X \to X$ which possess a continuous inverse T^{-1}. (A classical theorem of functional analysis states that if $T : X \to X$ is a continuous linear bijection then T^{-1} is automatically continuous, but we do not need to use this fact here.) Then $\text{GL}(X)$ is an open subset of $\mathbb{B}(X,X)$, and the inversion map*

$$I : \begin{array}{ccc} \text{GL}(X) & \to & \text{GL}(X) \\ A & \mapsto & A^{-1} \end{array}$$

is differentiable with $I'(A)B = -A^{-1}BA$ for all $A \in \text{GL}(X)$ and all $B \in \mathbb{B}(X,X)$.

Proof. We fix $A \in \text{GL}(X)$. For any element $B \in \mathbb{B}(X,X)$ with $\|\!|B|\!\| < 1/\|\!|A^{-1}|\!\|$ we have $\|\!|A^{-1}B|\!\| \leq \|\!|A^{-1}|\!\| \cdot \|\!|B|\!\| < 1$ which implies that the Neumann series $\sum_{k=0}^{\infty}(-A^{-1}B)^k$ converges absolutely and yields the inverse of $\mathbf{1} + A^{-1}B$. (This is the step where the completeness of X is used.) Now since both A and $\mathbf{1} + A^{-1}B$ are invertible then so is $A + B = A(\mathbf{1} + A^{-1}B)$ (being the product of two invertible operators). We have shown that if $A \in \text{GL}(X)$ and $\|B\| < \|A^{-1}\|$ then $A + B \in \text{GL}(X)$, which implies that $\text{GL}(X)$ is an open subset of $\mathbb{B}(\mathbb{X},\mathbb{X})$. Moreover, we have

$$\begin{aligned} (A+B)^{-1} &= \big(\mathbf{1} + A^{-1}B\big)^{-1}A^{-1} = \left(\sum_{k=0}^{\infty}(-1)^k(A^{-1}B)^k\right)A^{-1} \\ &= A^{-1} - A^{-1}BA^{-1} + \sum_{k=2}^{\infty}(-1)^k(A^{-1}B)^kA^{-1} \end{aligned}$$

so that $I(A+B) = I(A) - A^{-1}BA^{-1} + R$ where $R := \sum_{k=2}^{\infty}(-1)^k(A^{-1}B)^kA^{-1}$. Now

$$\begin{aligned} \|R\| &\leq \sum_{k=2}^{\infty}\|A^{-1}B\|^k\|A^{-1}\| = \frac{\|A^{-1}B\|^2}{1 - \|A^{-1}B\|}\|A^{-1}\| \\ &\leq \frac{\|A^{-1}B\|^2\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|B\|} \leq \frac{\|A^{-1}\|^3\|B\|^2}{1 - \|A^{-1}\| \cdot \|B\|} \end{aligned}$$

implies that $\|R\|/\|B\| \to 0$ for $B \to 0$, and since $B \mapsto -A^{-1}BA^{-1}$ is a linear mapping $\mathbb{B}(X,X) \to \mathbb{B}(X,X)$ this implies the differentiability of I at A with $I'(A)B = -A^{-1}BA^{-1}$ for all $B \in \mathbb{B}(X,X)$. ■

Next, we note that differentiability is a stronger property than continuity.

(1.11) Remark. *Differentiability at a point p implies continuity at p.*

Proof. Let α be the affine approximation of f at p. For $x \to p$ we have $\alpha(x) \to \alpha(p) = f(p)$ (because α is continuous by hypothesis), and then

$$\begin{aligned} \|f(x) - f(p)\| &= \|f(x) - \alpha(p)\| \leq \|f(x) - \alpha(x)\| + \|\alpha(x) - \alpha(p)\| \\ &= \frac{\|f(x) - \alpha(x)\|}{\|x - p\|} \cdot \|x - p\| + \|\alpha(x) - \alpha(p)\| \to 0 \cdot 0 + 0 = 0 \quad \text{as } x \to p. \end{aligned}$$

■

We now derive some elementary rules for differentiability. We begin by showing that the formation of the derivative is a linear operation.

(1.12) Theorem. *Let V and W be normed spaces and let Ω be an open subset of V. If $f, g : \Omega \to W$ are both differentiable at p and if $\lambda, \mu \in \mathbb{R}$ are arbitrary real numbers, then $\lambda f + \mu g$ is also differentiable at p, and*

$$(\lambda f + \mu g)'(p) \;=\; \lambda\, f'(p) + \mu\, g'(p).$$

Proof. We have $f(p+v) = f(p)+f'(p)v+r_1(v;p)$ and $g(p+v) = g(p)+g'(p)v+r_2(v;p)$ where $\|r_1\|/\|v\| \to 0$ and $\|r_2\|/\|v\| \to 0$ as $v \to 0$. Letting $r := \lambda r_1 + \mu r_2$, this yields $(\lambda f + \mu g)(p + v) = (\lambda f + \mu g)(p) + \big(\lambda f'(p) + \mu g'(p)\big)(v) + r(v;p)$ and $\|r(v;p)\|/\|v\| \leq (\|r_1\| + \|r_2\|)/\|v\| = (\|r_1\|/\|v\|) + (\|r_2\|/\|v\|) \to 0 + 0 = 0$ as $\|v\| \to 0$; whence the claim. ■

The next result is of fundamental importance. It shows that the concatenation of two differentiable maps is again differentiable and that the linearization of this concatenation is the concatenation of the individual linearizations.

(1.13) Chain Rule. *Let X, Y, Z be normed spaces and let $\Omega_X \subseteq X$ and $\Omega_Y \subseteq Y$ be open subsets. If $f : \Omega_X \to Y$ is differentiable at the point $p \in \Omega_X$ and if $g : \Omega_Y \to Z$ is differentiable at the point $f(p) \in \Omega_Y$, then $g \circ f$ is differentiable at p, and we have*

$$(g \circ f)'(p) \;=\; g'\big(f(p)\big) \circ f'(p).$$

Proof. Let $q := f(p)$. By hypothesis, we have $f(p + v) = f(p) + f'(p)v + r_1(v;p)$ and $g(q + w) = g(q) + g'(q)w + r_2(w;q)$ where $\|r_1\|/\|v\| \to 0$ as $\|v\| \to 0$ and where $\|r_2\|/\|w\| \to 0$ as $\|w\| \to 0$. In particular we can choose $w := f'(p)v + r_1$ (which tends to zero as $v \to 0$ due to the continuity of $f'(p)$). Then

$$\begin{aligned}(g \circ f)(p+v) &= g\big(f(p+v)\big) = g\left(f(p) + \big[f'(p)v+r_1\big]\right) \\ &= g\big(f(p)\big) + g'\big(f(p)\big)\left[f'(p)v+r_1\right] + r_2 \\ &= (g \circ f)(p) + \big[g'\big(f(p)\big) \circ f'(p)\big]\, v + r_3\end{aligned}$$

where $r_3 := g'\big(f(p)\big)r_1 + r_2$. Due to the hypotheses on r_1 and r_2 and the continuity of $g'(q)$, we have $\|r_3\|/\|v\| \to 0$ as $\|v\| \to 0$; whence the claim. ■

Next, we derive a product rule for functions which depend linearly on each of their arguments. This rule includes the classical product rule $(uv)' = u'v + uv'$ and Leibniz' rule $(u_1 \cdots u_n)' = \sum_{i=1}^{n} u_1 \cdots u_{i-1} u_i' u_{i+1} \cdots u_n$ for real functions as special cases.

(1.14) Product Rule. *Let $V_1, \ldots, V_n, W$ be normed spaces and let Ω_i be an open subset of V_i for $1 \leq i \leq n$. If $f_i : \Omega_i \to V_i$ is differentiable at $p_i \in \Omega_i$ and if $\beta : V_1 \times \cdots \times V_n \to W$ is a continuous n-linear mapping, then $f := \beta(f_1, \ldots, f_n)$ is differentiable at the point $p := (p_1, \ldots, p_n)$, and, writing $q_i := f_i(p_i)$ for $1 \leq i \leq n$, we have*

$$f'(p)(u_1, \ldots, u_n) \;=\; \sum_{k=1}^{n} \beta\bigl(q_1, \ldots, q_{i-1}, f_i'(p_i)u_i, q_{i+1}, \ldots, q_n\bigr).$$

Proof. Since f_i is differentiable at p_i for $1 \leq i \leq n$, the mapping $(u_1, \ldots, u_n) \mapsto \bigl(f_1(u_1), \ldots, f_n(u_n)\bigr)$ is differentiable at $(p_1, \ldots, p_n)$. Moreover, β is differentiable at the point $\bigl(f_1(p_1), \ldots, f_n(p_n)\bigr)$ because of (1.9). Hence the chain rule implies that $f = \beta \circ (f_1, \ldots, f_n)$ is differentiable at $(p_1, \ldots, p_n)$ with $f'(p)$ as above. ■

The following theorem shows how a vector-valued real function can be estimated against a real-valued comparison function. This theorem has a simple physical interpretation: We consider a particle moving freely in space and a second particle moving along a straight line. If the speed of the first particle never exceeds that of the second one, then the first particle cannot move further away from its initial position than the second one during the same time interval.

(1.15) Comparison Theorem. *Let W be a normed space and $a, b \in \mathbb{R}$ real numbers. Consider two differentiable functions $g : [a,b] \to W$ and $\varphi : [a,b] \to \mathbb{R}$ such that $\|g'(t)\| \leq \varphi'(t)$ for all $t \in [a,b]$. Then $\|g(b) - g(a)\| \leq \varphi(b) - \varphi(a)$.*

Proof. Pick $\varepsilon > 0$ arbitrarily and let J_ε the set of all $x \in [a,b]$ such that

$$\|g(\xi) - g(a)\| \leq \varphi(\xi) - \varphi(a) + \varepsilon(\xi - a) \text{ for } 0 \leq \xi < x.$$

Clearly, J_ε is an interval whose left endpoint is a. Due to the continuity of g and φ, this interval is closed so that $J_\varepsilon = [a, \beta]$ where $\beta := \sup J_\varepsilon$. Assume that $\beta < b$ to obtain a contradiction. Since g and φ are differentiable at β, there would be a number $\delta > 0$ such that

$$\frac{\|g(\beta+h) - g(\beta) - g'(\beta)h\|}{h} \leq \frac{\varepsilon}{2} \quad \text{and} \quad \frac{|\varphi(\beta+h) - \varphi(\beta) - \varphi'(\beta)h|}{h} \leq \frac{\varepsilon}{2}$$

whenever $0 < h < \delta$. For each such number h this would imply that

$$\begin{aligned}
&\|g(\beta+h) - g(a)\| \;\leq\; \|g(\beta+h) - g(\beta)\| + \|g(\beta) - g(a)\| \\
&\leq\; \left(\frac{\varepsilon h}{2} + \|g'(\beta)\| \cdot h\right) + (\varphi(\beta) - \varphi(a) + \varepsilon(\beta - a)) \\
&\leq\; \frac{\varepsilon h}{2} + \varphi'(\beta) \cdot h + \varphi(\beta) - \varphi(a) + \varepsilon(\beta - a) \\
&\leq\; \frac{\varepsilon h}{2} + \left(\varphi(\beta+h) + \frac{\varepsilon h}{2}\right) - \varphi(a) + \varepsilon(\beta - a) \\
&=\; \varphi(\beta+h) - \varphi(a) + \varepsilon(\beta + h - a)
\end{aligned}$$

and hence $\beta + h \in J_\varepsilon$, contradicting the definition of β. Thus the assumption $\beta < b$ was wrong; so we have $\beta = b$ and hence $\|g(b) - g(a)\| \le \varphi(b) - \varphi(a) + \varepsilon(b-a)$. Since $\varepsilon > 0$ was chosen arbitrarily, the claim follows. ■

For a differentiable function $f : \mathbb{R} \to \mathbb{R}$, the Mean Value Theorem states that between any two points a and b there is a point ξ such that $f(b) - f(a) = f'(\xi)(b-a)$. This is no longer true for vector-valued functions, as can be seen by the example $f(t) := (\cos t, \sin t)$ with $a = 0$ and $b = 2\pi$). However, the crucial point (namely the possibility to estimate the value difference $f(b)-f(a)$ in terms of the argument difference $b-a$ using the values of the derivative of f) can be carried over to the more general setting of functions $f : V \to W$.

(1.16) Mean Value Theorem. *Let V and W be finite-dimensional real vector spaces, let $\Omega \subseteq V$ be a convex open set and let $f : \Omega \to W$ be a differentiable mapping. If $[x, y]$ denotes the line segment between any two points $x, y \in \Omega$, we have*

$$\|f(x) - f(y)\| \;\le\; M \cdot \|x - y\| \quad \textit{where} \quad M := \sup_{\xi \in [x,y]} \|f'(\xi)\|_{\rm op}\,.$$

Proof. The functions $g(t) := f\big(x+t(y-x)\big)$ and $\varphi(t) := M\|x-y\|\cdot t$ are differentiable on $[0,1]$, and the chain rule yields $g'(t) = f'\big(x + t(y-x)\big)(y-x)$ for all t, hence $\|g'(t)\| \le M\|x-y\| = \varphi'(t)$. The comparison theorem (1.15) yields $\|g(1) - g(0)\| \le \varphi(1) - \varphi(0)$, which is the claim. ■

(1.17) Remark. We can also argue as follows. If $M = \infty$, the claim holds trivially. If not then g' is bounded, and we have $\|f(y) - f(x)\| = \|g(1) - g(0)\| = \left\|\int_0^1 g'(t)\,\mathrm{d}t\right\| \le \int_0^1 \|g'(t)\|\,\mathrm{d}t = \int_0^1 \|f'\big(x + t(y-x)\big)\|\,\mathrm{d}t \cdot \|y - x\| \le M \cdot \|y - x\|$. However, the function $t \mapsto g'(t)$ (not assumed to be continuous) need not be integrable in the sense of Riemann. Hence this argument requires properties of the Lebesgue integral for general vector-valued functions and hence is less elementary then applying the comparison theorem (1.15).

We now generalize the concept of partial derivatives, which is well known for functions $f : \mathbb{R}^n \to \mathbb{R}$.

(1.18) Definition. *Let $V_1, \ldots, V_n, W$ be normed spaces, let $V := V_1 \times \cdots \times V_n$ (equipped with the norm $\|(v_1, \ldots, v_n)\|_V := \sum_{i=1}^n \|v_i\|_{V_i}$ or any equivalent norm) and let $\Omega \subseteq V$ be an open subset. A mapping $f : \Omega \to W$ is called* **partially differentiable** *at a point $p = (p_1, \ldots, p_n)$ with respect to its i-th argument, if the mapping*

$$f_i : x \mapsto f(p_1, \ldots, p_{i-1}, x, p_{i+1}, \ldots, p_n)$$

is differentiable at $x = p_i$. (Note that f_i does not just depend on the index i, but also on the point p.) In this case we call $(\partial_i f)(p) := f_i'(p_i)$ the **partial derivative** *of f at p with respect to the i-th argument.*

(1.19) Example. Consider $V = \mathbb{R}^n$ with the canonical basis $(e_1, \ldots, e_n)$. Writing $V_i := \mathbb{R}e_i$, the partial derivatives of a function $f : V \to \mathbb{R}$ with respect to the decomposition $V = V_1 \times \cdots \times V_n$ coincide with the classical partial derivatives $\partial f / \partial x_i$ known from calculus.

As is to be expected, differentiability implies partial differentiability with respect to all arguments.

(1.20) Theorem. *Let $V_1, \ldots, V_n, W$ be normed spaces, let $V := V_1 \times \cdots \times V_n$ be the product of $V_1, \ldots, V_n$ and let $f : V \to W$ be differentiable at the point $p = (p_1, \ldots, p_n)$. Then f possesses partial derivatives at p with respect to all arguments, and if $\pi_i : V \to V_i$ denotes the projection onto the i-th factor, we have $f'(p) = \sum_{i=1}^n (\partial_i f)(p) \circ \pi_i$, which means that*

$$f'(p)(v_1, \ldots, v_n) \;=\; \sum_{i=1}^n (\partial_i f)(p) v_i \quad \textit{for all } v_i \in V_i.$$

Proof. Denote by $I_i : V_i \to V$ the inclusion map of V_i into V and write $\widehat{v}_i := I_i(v_i) = (0, \ldots, v_i, \ldots, 0)$ for $v_i \in V_i$. Then $\|\widehat{v}_i\|_V = \|v_i\|_{V_i}$ and hence

$$\frac{\|f_i(p_i + v_i) - f_i(p_i) - (f'(p) \circ I_i)(v_i)\|}{\|v_i\|} \;=\; \frac{\|f(p + \widehat{v}_i) - f(p) - f'(p)\widehat{v}_i\|}{\|\widehat{v}_i\|}.$$

Due to the differentiability of f at p, the last expression tends to zero as $v_i \to 0$; this shows that f is partially differentiable at p with respect to the i-th argument with $(\partial_i f)(p) = f'(p) \circ I_i$. Thus for $v = (v_1, \ldots, v_n)$ we have $\sum_{i=1}^n \big((\partial_i f)(p) \circ \pi_i\big)(v) = \sum_{i=1}^n (\partial_i f)(p)(v_i) = \sum_{i=1}^n \big(f'(p) \circ I_i\big)(v_i) = \sum_{i=1}^n f'(p)\widehat{v}_i = f'(p)\big(\sum_{i=1}^n \widehat{v}_i\big) = f'(p)v$. ■

We now turn to the definition of higher derivatives. Note that if a function $f : \Omega \to W$ (where $\Omega \subseteq V$) is differentiable at each point $x \in \Omega$ then for each $x \in \Omega$ the derivative $f'(x)$ is a continuous linear mapping from V to W, i.e., an element of the vector space $\mathbb{B}(V, W)$ of all bounded linear mappings from V to W. Equipped with the operator norm, this is again a normed space, so that we can consider the assignment $x \mapsto f'(x)$ as a mapping $f' : \Omega \to \mathbb{B}(V, W)$, which may again be continuous or even differentiable. Thus we can inductively define the higher derivatives of $f : \Omega \to W$, as follows.

(1.21) Definition. *Let V and W be normed spaces and let $\Omega \subseteq V$ be an open subset. Given a (continuous) function $f : \Omega \to W$, we call $f^{(0)} := f$ the zero-th derivative of f. If $f^{(k-1)}$ is defined on Ω and is differentiable at $p \in \Omega$, we say that f is k times differentiable at p and call $f^{(k)}(p) := (f^{(k-1)})'(p)$ the k-th derivative of f at p. If $f^{(k)}(x)$ exists for all $x \in \Omega$ and if $x \mapsto f^{(k)}(x)$ is continuous, we say that f is k times continuously differentiable, that f is smooth of class C^k, or simply that f is a C^k-function. We denote by $C^k(\Omega; W)$ the set of all C^k-functions $f : \Omega \to W$. Moreover, we denote by $C^\infty(\Omega; W) := \bigcap_{k \geq 1} C^k(\Omega; W)$ the set of all functions $f : \Omega \to W$ which possess derivatives of all orders.*

From (1.12) we conclude inductively that linear combinations of C^k-functions are again C^k-functions, i.e., that $C^k(\Omega; W)$ is a vector space. Moreover, as a consequence of the chain rule, the concatenation of two C^k-functions is again a C^k-function. We want to describe the higher derivatives of a function as concretely as possible. If $f : \Omega \to W$ is differentiable, then f' is a mapping $f' : \Omega \to \mathbb{B}(V,W)$. If f' is again differentiable, then f'' is a mapping $f'' : \Omega \to \mathbb{B}\big(V, \mathbb{B}(V,W)\big)$. Thus for any two elements $v_1, v_2 \in V$ the expression $\big(f''(p)v_1\big)(v_2)$ is an element of W which depends linearly and continuously on both v_1 and v_2. Hence writing $f''(p)[v_1, v_2]$ instead of $\big(f''(p)(v_1)\big)(v_2)$, we can interpret $f''(p)$ as *bilinear* continuous mapping $V \times V \to W$. We will always make this identification. Continuing in this way, we find that the k-th derivative $f^{(k)}(p)$ at a point $p \in \Omega$ is a continuous k-linear mapping $f^{(k)}(p) : V \times \cdots \times V \to W$. The definining property of $f^{(k)}(p)$ is

$$f^{(k-1)}(p+h)\,[v_1, \ldots, v_{k-1}] \;=\; f^{(k-1)}(p)\,[v_1, \ldots, v_{k-1}] + f^{(k)}(p)\,[v_1, \ldots, v_{k-1}, h] + \mathrm{o}\big(\|h\|\big)\,,$$

where $\mathrm{o}\big(\|h\|\big)$ denotes a quantity which, after being divided by $\|h\|$, still tends to zero as $h \to 0$ uniformly on the set of all $(v_1, \ldots, v_{k-1})$ satisfying $\|v_i\| \leq 1$ for $1 \leq i \leq k-1$. To evaluate $f^{(k)}(p)$, we can use the formula

$$f^{(k)}(p)[v_1, \ldots, v_k] \;=\; \frac{\partial}{\partial_{t_k}} \cdots \frac{\partial}{\partial_{t_1}} f(p + t_1 v_1 + \cdots + t_k v_k)\Big|_{t_1 = \cdots = t_k = 0}\,.$$

(1.22) Examples. (a) Any continuous n-linear mapping $\beta : V_1 \times \cdots \times V_n \to W$ between normed spaces is of class C^∞, as can be shown by induction on n. If $n = 1$ then f is a continuous linear mapping; such a mapping is everywhere differentiable with $f'(p) = f$ for all p. If $n \geq 2$, then (1.9) shows that β is differentiable at any point $(v_1, \ldots, v_n)$ with $\beta' = \sum_{i=1}^n \beta_i$ where $\beta_i : V_1 \times \cdots \times V_{i-1} \times V_{i+1} \times \cdots \times V_n \to \mathbb{B}(V_i, W)$ is a continuous $(n-1)$-linear mapping for each i. By induction hypothesis each of the mappings β_i is of class C^∞. But then so is β as a sum of C^∞-mappings.

(b) Let $V := \mathbb{B}(X,X)$ where X is a Banach space, and let $\mathrm{GL}(X)$ be the open subset of all invertible operators in $\mathbb{B}(X,X)$. In (1.10) we showed that the inversion map $I : \mathrm{GL}(X) \to \mathrm{GL}(X)$ is differentiable. We claim that, in fact, I is of class C^∞. To prove this claim we introduce the continuous bilinear mapping $\beta : V \times V \to \mathbb{B}(V,V)$ given by $\beta(S,T)[B] := -SBT$. In (1.10) we showed that $I'(A)B = -A^{-1}BA^{-1}$ for all $A \in \mathrm{GL}(X)$ and all $B \in \mathbb{B}(X,X)$, which may be restated by saying that $I'(A) = \beta\big(I(A), I(A)\big)$ for all $A \in \mathrm{GL}(X)$. Hence $I' = \beta \circ (I, I)$. Since β is of class C^∞ by part (a), this last equation implies that if I is of class C^k, then so is I', so that I is, in fact, of class C^{k+1}. This clearly implies that I is of class C^∞.

(c) Slightly more generally, assume that X and Y are Banach spaces for which the set $\mathrm{GL}(X,Y)$ of all invertible bounded linear mappings $T : X \to Y$ is not empty (so that X and Y are isomorphic not only algebraically, but also topologically). Fix an element $T_0 \in \mathrm{GL}(X,Y)$. The inversion maps $I : \mathrm{GL}(X,Y) \to \mathrm{GL}(Y,X)$ and $\widehat{I} : \mathrm{GL}(X,X) \to \mathrm{GL}(X,X)$ are then related by the equation $I = R \circ \widehat{I} \circ L$ where

$$L : \begin{array}{ccc} \mathbb{B}(X,Y) & \to & \mathbb{B}(X,X) \\ T & \mapsto & T_0^{-1}T \end{array} \qquad \text{and} \qquad R : \begin{array}{ccc} \mathbb{B}(X,X) & \to & \mathbb{B}(Y,X) \\ S & \mapsto & ST_0^{-1} \end{array}$$

because for all $T \in \mathrm{GL}(X,Y)$ the equation

$$I(T) \;=\; T^{-1} \;=\; (T_0^{-1}T)^{-1}T_0^{-1} \;=\; (R \circ \widehat{I} \circ L)(T)$$

holds. Since $\widehat{I}$ is of class C^∞ by part (b) and since L and R are obviously of class C^∞, we see that $\widehat{I}$ is also a C^∞-mapping. ■

We now want to investigate what can be said about the differentiability of the inverse function of a differentiable bijection. By considering the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$, we note that even if a bijective function is of class C^∞, the inverse function f^{-1} need not be differentiable. On the other hand, if $f : \Omega_1 \to \Omega_2$ is a bijection which is differentiable at $p \in \Omega_1$ and if, in addition, $g := f^{-1}$ is differentiable at $f(p)$, then applying the chain rule to the equation $g\big(f(x)\big) = x$ at $x = p$ implies that $g'\big(f(p)\big)f'(p) = \mathbf{1}$, which shows that $f'(p)$ and $g'\big(f(p)\big)$ are necessarily inverses of each other. Conversely, we can ask whether or not the invertibility of $f'(p)$ already implies the differentiability of f at p. A first result in this direction is given by the following theorem.

(1.23) Inversion Rule. *Let V and W be normed spaces, let $A \subseteq V$ and $B \subseteq W$ be open sets and let $f : A \to B$ be a homeomorphism (i.e., a bijection such that f and f^{-1} are both continuous). Assume that f is differentiable at $x_0 \in A$ and that $f'(x_0) : V \to W$ is invertible with continuous inverse $f'(x_0)^{-1} : W \to V$. Then f^{-1} is differentiable at $y_0 := f(x_0)$, and*

$$(f^{-1})'(y_0) \;=\; f'(x_0)^{-1}\,.$$

Proof. By hypothesis, $T := f'(x_0)^{-1}$ is a continuous linear mapping. Hence there is a constant $C > 0$ such that $\|Tw\| \le C\|w\|$ for all $w \in W$. Pick a number $\varepsilon > 0$ with $\varepsilon \le 1/(2C)$. Since f is differentiable at x_0, there is a number $r > 0$ such that

$$(1) \qquad \frac{\|f(x) - f(x_0) - f'(x_0)(x - x_0)\|}{\|x - x_0\|} \;\le\; \varepsilon$$

for all $x \in A$ with $\|x - x_0\| \le r$. Without loss of generality, we may choose r so small that $B_r(x_0) \subseteq A$ so that (1) holds for all x with $\|x - x_0\| \le r$. Since f^{-1} is continuous at y_0, there is a number $\rho > 0$ such that $\|y - y_0\| \le \rho$ implies $\|f^{-1}(y) - f^{-1}(y_0)\| \le r$, i.e., $\|f^{-1}(y) - x_0\| \le r$. We pick such an element y with $\|y - y_0\| \le \rho$ and let $x := f^{-1}(y)$. Then $\|x - x_0\| \le r$, hence due to (1) also $y = y_0 + f'(x_0)(x - x_0) + R$ with a remainder R satisfying $\|R\| \le \varepsilon \cdot \|x - x_0\|$. Applying the linear mapping T to this equation results in $T(y) = T(y_0) + T\big(f'(x_0)(x - x_0)\big) + T(R) = T(y_0) + (x - x_0) + T(R)$. Since

$$(2) \qquad \|T(R)\| \;\le\; C\|R\| \;\le\; C\varepsilon\|x - x_0\| \;\le\; \frac{\|x - x_0\|}{2}$$

this implies that $\|T(y) - T(y_0)\| \ge \|x - x_0\| - \|T(R)\| \ge \|x - x_0\|/2$ and therefore $\|x - x_0\| \le 2\|T(y - y_0)\| \le 2C\|y - y_0\|$. Plugging this into (2) results in $\|T(R)\| \le C\varepsilon\|x - x_0\| \le 2C^2\varepsilon\|y - y_0\|$. Hence from $\|y - y_0\| \le \rho$ it follows that

$$\|f^{-1}(y) - f^{-1}(y_0) - T(y - y_0)\| \;=\; \|T(y) - T(y_0) - (x - x_0)\| \;=\; \|T(R)\| \;\le\; 2C^2\varepsilon\|y - y_0\|$$

and hence
$$\frac{\|f^{-1}(y) - f^{-1}(y_0) - T(y - y_0)\|}{\|y - y_0\|} \leq 2C^2\varepsilon.$$
Since $\varepsilon > 0$ was arbitrary, this shows that f^{-1} is differentiable at y_0 with $(f^{-1})'(y_0) = T = f'(x_0)^{-1}$. ■

As a consequence of (1.22)(b), we will show that, under the hypotheses of (1.23), the inverse f^{-1} is as smooth as f itself, i.e., that if f is of class C^k then so is f^{-1}.

(1.24) Theorem. *Let V_1 and V_2 be Banach spaces and let $\Omega_1 \subseteq V_1$ and $\Omega_2 \subseteq V_2$ be open subsets. Let $f : \Omega_1 \to \Omega_2$ be a homeomorphism. If f is of class C^k with $k \geq 1$, then so is f^{-1}.*

Proof. We use induction on k, starting with $k = 1$. Since f is differentiable at every point $p \in \Omega_1$ by hypothesis, the inversion rule (1.23) implies that the inverse map f^{-1} is differentiable at every point $q \in \Omega_2$ and satisfies

$$(\star) \qquad (f^{-1})' = \big(f' \circ f^{-1}(\bullet)\big)^{-1}.$$

Since f^{-1} and f' are continuous by hypothesis and since the inversion map $\mathrm{GL}(V_1, V_2) \to \mathrm{GL}(V_2, V_1)$ is also continuous, we conclude from $(\star)$ that also $(f^{-1})'$ is continuous, which means that f^{-1} is of class C^1. For the induction step we assume that f is of class C^k where $k \geq 2$. Then f is, *a fortiori*, of class C^{k-1}. Hence by induction hypothesis f^{-1} is also of class C^{k-1}. Buth then so is the concatenation $f' \circ f^{-1}$. Since the inversion map is of class C^∞ by (1.22)(b), hence also of class C^{k-1}, equation $(\star)$ then implies that $(f^{-1})'$ is of class C^{k-1}, which means that f^{-1} is of class C^k. This concludes the induction step. ■

(1.25) Definition. *Let V_1 and V_2 be normed spaces and let $\Omega_1 \subseteq V_1$ and $\Omega_2 \subseteq V_2$ be open subsets. A mapping $f : \Omega_1 \to \Omega_2$ is called a* **diffeomorphism** *of class C^k if f is a bijection such that both f and f^{-1} are of class C^k.*

The inversion rule is of limited usefulness, because it assumes the existence of an inverse function. It would be much more desirable to assume only the invertibility of the linearization $f'(x_0)$ (which is usually easy to verify) and then conclude the invertibility of f itself (which is usually much more difficult, as f may be given by a very complicated formula). If we assume not just differentiability, but *continuous* differentiability of f in a neighborhood of x_0, we can indeed prove such a theorem.

(1.26) Inverse Function Theorem. *Consider Banach spaces X and Y, an open subset $\Omega \subseteq X$, a C^k-mapping $f : \Omega \to Y$ where $k \geq 1$ and a point $x_0 \in \Omega$ such that $f'(x_0)$ is invertible. Then there exist an open neighborhood U of x_0 and an open neighborhood V of $y_0 := f(x_0)$ such that $f : U \to V$ is a C^k-diffeomorphism.*

Proof. We denote by $\|\!|\cdot\|\!|$ the operator norm on $\mathbb{B}(X,Y)$. Let $\delta > 0$ be so small that $\overline{B_\delta(x_0)} \subseteq \Omega$ and $\|\!|\mathbf{1} - f'(x_0)^{-1}f'(x)\|\!| \le 1/2$ for all $x \in \overline{B_\delta(x_0)}$. Let $\varepsilon := \delta/\big(2\|\!|f'(x_0)^{-1}\|\!|\big)$. We claim that for each $y \in B_\varepsilon(y_0)$ there is a unique element $x \in \overline{B_\delta(x_0)}$ such that $f(x) = y$. To prove this statement, let us fix $y \in B_\varepsilon(y_0)$. The element x we want to find must satisfy

$$y \;=\; f(x) \;=\; f(x_0) + f'(x_0)(x - x_0) + r(x_0;x) \tag{1}$$

where $\|r(x_0;x)\|/\|x - x_0\| \to 0$ as $x \to x_0$. Equation (1) can be rewritten as

$$x \;=\; x_0 + f'(x_0)^{-1}\big(y - f(x_0) - r(x_0;x)\big) \;=\; x_0 + f'(x_0)^{-1}\big(y - y_0 - r(x_0;x)\big). \tag{2}$$

The strategy to find a solution x of (2) is to determine a sequence $\xi_0, \xi_1, \xi_2, \ldots$ of approximate solutions which will be shown to converge to an element x which is then an exact solution. The initial approximant is obtained by simply ignoring the remainder term r in (2), i.e., by letting

$$\xi_0 \;:=\; x_0 + f'(x_0)^{-1}\big(y - f(x_0)\big) \;=\; x_0 + f'(x_0)^{-1}\big(y - y_0\big). \tag{3}$$

Each subsequent approximant is obtained by plugging in the previously obtained approximant into the right-hand side of (2) to define the left-hand side, i.e., by letting

$$\begin{aligned} \xi_{n+1} \;:=&\; x_0 + f'(x_0)^{-1}\big(y - f(x_0) - r(x_0;\xi_n)\big) \\ =&\; x_0 + f'(x_0)^{-1}\big(y - f(\xi_n) + f'(x_0)(\xi_n - x_0)\big) \\ =&\; \xi_n + f'(x_0)^{-1}\big(y - f(\xi_n)\big) \end{aligned} \tag{4}$$

where we used the fact that $f(\xi_n) = f(x_0) + f'(x_0)(\xi_n - x_0) + r(x_0;\xi_n)$ in making the transition from the first to the second line. For this sequence to be well-defined we must be sure that all elements ξ_n lie in the domain Ω of f (so that $f(\xi_n)$ is defined). First, $\|\xi_0 - x_0\| = \|f'(x_0)^{-1}(y - y_0)\| \le \|\!|f'(x_0)^{-1}\|\!|\,\|y - y_0\| < \varepsilon \cdot \|\!|f'(x_0)^{-1}\|\!| = \delta/2$ and thus $\xi_0 \in B_\delta(x_0) \subseteq \Omega$. Next we show that if $\xi_n \in \overline{B_\delta(x_0)}$ then $\xi_{n+1} \in B_\delta(x_0)$. To do so, let us define the auxiliary function $F_y : \Omega \to \mathbb{R}^n$ via

$$F_y(x) \;:=\; x + f'(x_0)^{-1}\big(y - f(x)\big). \tag{5}$$

We note that F_y is differentiable on Ω with $F_y'(x) = \mathbf{1} - f'(x_0)^{-1}f'(x)$ for all $x \in \Omega$ $\big($because $f \in C^1(\Omega)\big)$ and hence $\|\!|F_y'(x)\|\!| \le 1/2$ for all $x \in \overline{B_\delta(x_0)}$ (by the choice of δ). The Mean Value Theorem then implies that $\|F_y(x_1) - F_y(x_2)\| \le (1/2)\|x_1 - x_2\|$ for all $x_1, x_2 \in \overline{B_\delta(x_0)}$. As a consequence, we see that F_y maps $\overline{B_\delta(x_0)}$ into $B_\delta(x_0)$ because if $\|x - x_0\| \le \delta$ then

$$\begin{aligned} \|F_y(x) - x_0\| \;\le&\; \|F_y(x) - F_y(x_0)\| \;+\; \|F_y(x_0) - x_0\| \\ =&\; \|F_y(x) - F_y(x_0)\| + \|f'(x_0)^{-1}(y - y_0)\| \\ \le&\; (1/2)\cdot\|x - x_0\| + \|\!|f'(x_0)^{-1}\|\!| \cdot \|y - y_0\| \\ <&\; (1/2)\cdot\|x - x_0\| + \varepsilon \cdot \|\!|f'(x_0)^{-1}\|\!| \\ <&\; (\delta/2) + (\delta/2) \;=\; \delta. \end{aligned} \tag{6}$$

This shows in particular that if $\xi_n \in \overline{B_\delta(x_0)}$ then $\xi_{n+1} = F_y(\xi_n) \in B_\delta(x_0)$; thus the sequence (ξ_n) is well-defined. We have established that F_y maps $\overline{B_\delta(x_0)}$ into itself and is a contraction. Since $\overline{B_\delta(x_0)}$, being a closed subset of a Banach space, is a complete metric space with the induced metric, the Banach Fixed Point Theorem is applicable and shows that F_y has a unique fixed point x in $\overline{B_\delta(x_0)}$ and that the sequence (ξ_n) converges to x. In fact, $x \in B_\delta(x_0)$ because F_y maps $\overline{B_\delta(x_0)}$ into $B_\delta(x_0)$. This fixed point satisfies the desired equation because

$$F_y(x) = x \iff f'(x_0)^{-1}\big(y - f(x)\big) = 0 \iff y - f(x) = 0 \iff y = f(x), \tag{7}$$

and the uniqueness of the fixed point implies the uniqueness of the solution. Thus we can define a map $g : B_\varepsilon(y_0) \to B_\delta(x_0)$ by letting $g(y)$ be the unique solution of the equation $y = f(x)$. Hence $f \circ g = \mathrm{id}$ by the very definition of g, so that g is automatically one-to-one. Let $V := B_\varepsilon(y_0)$ and $U := B_\delta(x_0) \cap f^{-1}(V)$; then $g : U \to V$ is a bijection. Note that (7) then shows that $x = g(y)$ if and only if $F_y(x) = x$ for $y \in U$. Thus if $y_1, y_2 \in U$ then

$$\begin{aligned}
&\|g(y_1) - g(y_2)\| \;=\; \|F_{y_1}\big(g(y_1)\big) - F_{y_2}\big(g(y_2)\big)\| \\
&\le\; \|F_{y_1}\big(g(y_1)\big) - F_{y_1}\big(g(y_2)\big)\| \;+\; \|F_{y_1}\big(g(y_2)\big) - F_{y_2}\big(g(y_2)\big)\| \\
&\le (1/2)\cdot\|g(y_1) - g(y_2)\| + \|F_{y_1}\big(g(y_2)\big) - F_{y_2}\big(g(y_2)\big)\| \\
&\le (1/2)\cdot\|g(y_1) - g(y_2)\| + \|f'(x_0)^{-1}(y_1 - y_2)\| \\
&\le (1/2)\cdot\|g(y_1) - g(y_2)\| + \|\!|f'(x_0)^{-1}|\!\|\cdot\|y_1 - y_2\| \\
&= (1/2)\cdot\|g(y_1) - g(y_2)\| + \big(1/(2\varepsilon)\big)\cdot\|y_1 - y_2\|
\end{aligned} \tag{8}$$

so that

$$\|g(y_1) - g(y_2)\| \;\le\; (1/\varepsilon)\cdot\|y_1 - y_2\|. \tag{9}$$

This shows that g is Lipschitz-continuous with Lipschitz constant $1/\varepsilon$. Theorem (1.24) now implies that g is of class C^k. ■

(1.27) Examples. (a) Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x,y) := \big(x^3e^y + y - 2x,\, 2xy + 2x\big)$. We have $f(1,0) = (-1,2)$. Since

$$f'(1,0) \;=\; \begin{bmatrix} 3x^2e^y - 2 & x^3e^y + 1 \\ 2y+2 & 2x \end{bmatrix}\Bigg|_{(x,y)=(0,0)} \;=\; \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$

is invertible, the Inverse Function Theorem assures us that there are neighborhoods U of $(1,0)$ and V of $(-1,2)$ such that $f : U \to V$ is a C^∞-diffeomorphism (even though we cannot explicitly write down the local inverse f^{-1}).

(b) We consider the differential operator $D : C^1_{2\pi} \to C^0_{2\pi}$ given by $Dy = \dot{y} + p \circ y$ as in (1.8)(d), where we now assume $p' > 0$ in addition. We claim that $D'(y)$ is invertible for each fixed $y \in C^1_{2\pi}$. To show this, we must verify that the equation $D'(y)[h] = \kappa$, i.e., the differential equation $\dot{h} + (p' \circ y)\, h = \kappa$, has a unique solution $h \in C^1_{2\pi}$ for each $\kappa \in C^0_{2\pi}$. Letting $\Phi(t) := \int_0^t p'\big(y(\tau)\big) \mathrm{d}\tau$, this differential equation has the general solution

$$h(t) \;=\; e^{-\Phi(t)} \left(h(0) + \int_0^t e^{\Phi(\tau)} \kappa(\tau)\, \mathrm{d}\tau \right).$$

Since a 2π-periodic solution exists if and only if $h(2\pi) = h(0)$, we must have

$$h(0)\big(1 - e^{-\Phi(2\pi)}\big) \;=\; e^{-\Phi(2\pi)} \int_0^{2\pi} e^{\Phi(\tau)} \kappa(\tau)\, \mathrm{d}\tau.$$

Since $\Phi(2\pi) = \int_0^{2\pi} p'\big(y(\tau)\big)\, \mathrm{d}\tau > 0$, this equation has a unique solution $h(0)$ (and hence the differential equation $\dot{h} + p'(y)h = \kappa$ has a unique 2π-periodic solution) for each given right-hand side κ. This shows that $D'(y)$ is invertible. The Inverse Mapping Theorem then shows that if $y \in C^1_{2\pi}$ is a solution of the differential equation $\dot{y} + p \circ y = g$ with $g \in C^0_{2\pi}$, then there are neighborhoods U of y in $C^1_{2\pi}$ and V of g in $C^0_{2\pi}$ such that the differential equation $\dot{\eta} + p \circ \eta = \gamma$ possesses a unique solution $\eta \in U$ for each $\gamma \in V$ and that η varies smoothly with γ. ∎

The following theorem deals with the question when an equation $f(x, y) = c$ can be solved for y in terms of x, i.e., rewritten in the form $y = g(x)$.

(1.28) Implicit Function Theorem. *Consider Banach spaces X, Y, Z, open sets $U \subseteq X$ and $V \subseteq Y$, a C^k-mapping $f : U \times V \to Z$ with $k \geq 1$ and a point $(x_0, y_0) \in U \times V$ with $f(x_0, y_0) = z_0$. Assume that*

$$(\partial_2 f)(x_0, y_0) : Y \to Z$$

is an isomorphism. Then there are a neighborhood U_0 of x_0 in X, a neighborhood V_0 of y_0 in Y and a C^k-mapping $g : U_0 \to Y$ such that

$$\{(x, y) \in U_0 \times V_0 \mid f(x, y) = z_0\} \;=\; \big\{\big(x, g(x)\big) \mid x \in U_0\big\}.$$

(Hence the equation $f(x, y) = z_0$ can be locally solved for y, and the function which expresses y in terms of x is again of class C^k.)

Proof. The mapping

$$\Phi : \begin{array}{ccc} U \times V & \to & X \times Z \\ (x, y) & \mapsto & \big(x, f(x, y)\big) \end{array}$$

is differentiable at (x_0, y_0) with derivative

$$\Phi'(x_0, y_0) \;=\; \begin{bmatrix} \mathrm{id}_X & \mathbf{0} \\ (\partial_1 f)(x_0, y_0) & (\partial_2 f)(x_0, y_0) \end{bmatrix}.$$

Due to our assumption, $\Phi'(x_0, y_0) : X \times Y \to X \times Z$ is an isomorphism. By the Inverse Function Theorem, Φ restricts to a C^k-diffeomorphism $\Phi : \widehat{U} \times \widehat{V} \to \widehat{W}$ with suitable neighborhoods $\widehat{U} \subseteq U \subseteq X$ of x_0 and $\widehat{V} \subseteq V \subseteq Y$ of y_0 and finally $\widehat{W} \subseteq Z$ of $f(x_0, y_0) = z_0$. The inverse function Φ^{-1} is necessarily of the form $\Phi^{-1}(x,z) = \big(x, h(x,z)\big)$ with a C^k-function h. Letting $g(x) := h(x, z_0)$, this yields

$$\begin{aligned} & \{(x,y) \in \widehat{U} \times \widehat{V} \mid f(x,y) = z_0\} \\ = \; & \{(x,y) \in \widehat{U} \times \widehat{V} \mid \Phi(x,y) = (x, z_0)\} \\ = \; & \{(x,y) \in \widehat{U} \times \widehat{V} \mid (x,y) = \Phi^{-1}(x,z_0) = \big(x, h(x,z_0)\big)\} \\ = \; & \{(x,y) \in \widehat{U} \times \widehat{V} \mid y = g(x)\} \\ = \; & \{\big(x, g(x)\big) \mid x \in \widehat{U} \cap g^{-1}(\widehat{V})\}. \end{aligned}$$

Letting $V_0 := \widehat{V}$ and $U_0 := \widehat{U} \cap g^{-1}(\widehat{V})$, this yields the claim. ∎

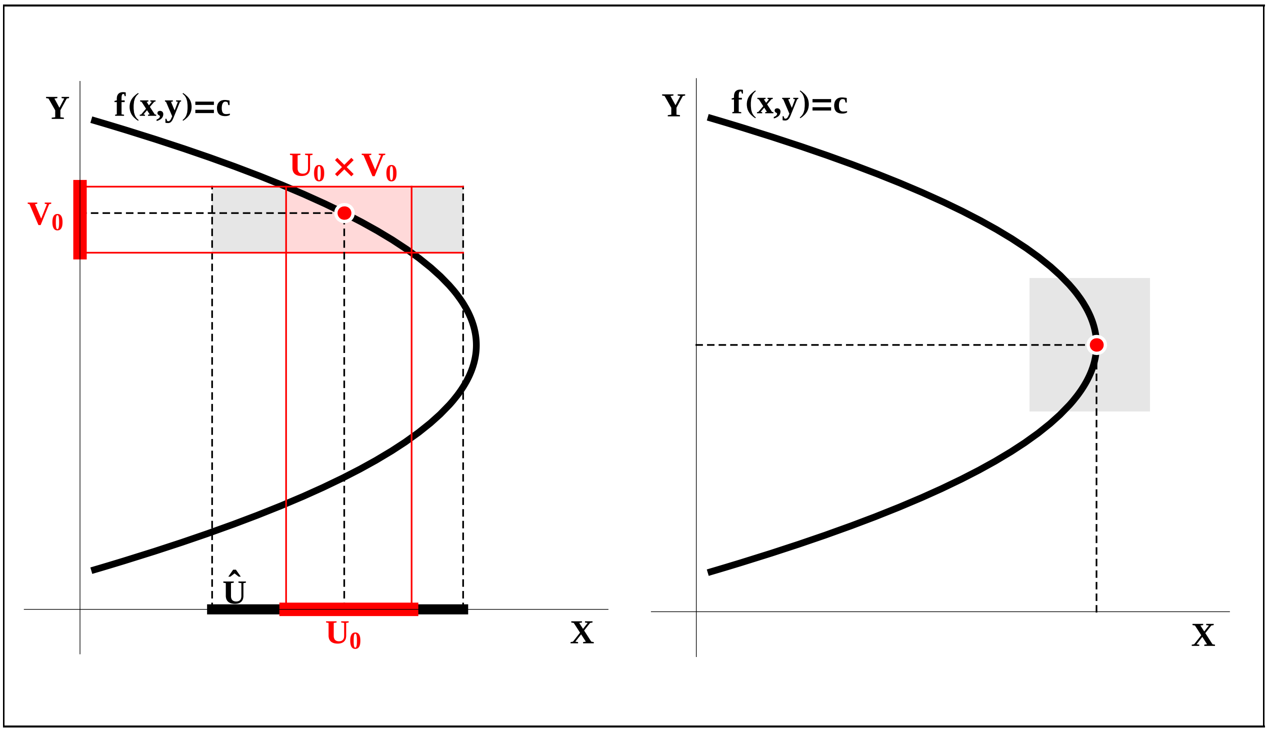

Figure 1.4: Illustration of the Implicit Function Theorem (left) and counterexample (right). In the counterexample, the equation $f(x,y) = c$ cannot be rewritten as $y = g(x)$ in a neigborhood of the indicated point, no matter how small the neighborhood is chosen.

(1.29) Remark. If a function $x \mapsto y(x)$ is implicitly given by an equation $f(x,y) = 0$ in a neighborhood of a point (x_0, y_0), then the derivatives of y at x_0 can be found by **implicit differentiation**, i.e., by taking derivatives on both sides of the identity $f\big(x, y(x)\big) = 0$ using the chain rule, without the need to explicitly write down y as a function of x (which is often not even possible). Taking derivatives on both sides of the equation $f(x,y)=0$ with respect to x yields

$$(1) \qquad 0 \;=\; f_x + f_y\, y'$$

(or, more explicitly, $f_x\big(x,y(x)\big)+f_y\big(x,y(x)\big)y'(x)=0$) and thus $f_x(x_0,y_0)+f_y(x_0,y_0)y'(x_0) = 0$, which can be solved for $y'(x_0)$. Taking again derivatives in (1), we find that

$$(2) \quad 0 \;=\; f_{xx}+f_{xy}y'+\big(f_{yx}+f_{yy}y'\big)y'+f_yy'' \;=\; f_{xx}+2f_{xy}y'+f_{yy}(y')^2+f_yy''.$$

Plugging in x_0 for x and y_0 for y in (2) and using the value $y'(x_0)$ known from (1), we can determine $y''(x_0)$, and so on. To solve (1), (2) and the subsequent equations in terms of the derivatives of y at the point x_0 requires in each step the invertibility of $f_y(x_0,y_0)$. ■

(1.30) Example. Consider the function $f:\mathbb{R}^2\to\mathbb{R}$ given by

$$f(x,y) \;=\; \sin(x^2{+}y)+\cos(x{+}y^2)-e^x\,.$$

Note that $f(0,0)=0$. Hence we ask if it is possible to solve the equation $f(x,y)=0$ in a neighborhood of $(0,0)$ for y in terms of x or for x in terms of y. Since

$$\begin{aligned} f_x &= 2x\cos(x^2{+}y)-\sin(x{+}y^2)-e^x \quad \text{and}\\ f_y &= \cos(x^2{+}y)-2y\sin(x{+}y^2), \end{aligned}$$

we have $f_x(0,0)=-1\neq 0$ and $f_y(0,0)=1\neq 0$. Hence solving for either x or y is possible by the Implicit Function Theorem. Let $x\mapsto y(x)$ be a local solution for y as a function of x. Plugging in $(x,y)=(0,0)$ into the equation $f_x+f_yy'=0$, we find that $-1+y'(0)=0$ so that $y'(0)=1$. The second-order partial derivatives are

$$\begin{aligned} f_{xx} &= \cos(x^2{+}y)-4x^2\sin(x^2{+}y)-e^x,\\ f_{xy} &= -2x\sin(x^2{+}y)-2y\cos(x{+}y^2) \;=\; f_{yx},\\ f_{yy} &= -\sin(x^2{+}y)-2\sin(x{+}y^2)-4y^2\cos(x{+}y^2); \end{aligned}$$

Plugging in $f_{xx}(0,0)=f_{xy}(0,0)=f_{yy}(0,0)=0$ and $y'(0)=1$ into the equation $f_{xx}+2f_{xy}y'+f_{yy}(y')^2+f_yy''=0$, we find that $y''(0)=0$. Continuing in this way, we can in principle compute all derivatives $y^{(k)}(0)$. ■

The Inverse Function Theorem states that a smooth mapping f whose linearization $f'(p)$ at a point p is invertible is itself invertible in a neighborhood of p. We now show that a similar property holds for surjectivity instead of bijectivity: If $f'(p)$ is surjective, then f is locally surjective.

(1.31) Theorem on Local Surjectivity. *Consider Banach spaces X and Z, an open subset $\Omega \subseteq X$, a point $p \in \Omega$ and a C^1-mapping $f : \Omega \to Z$. Let X_0 be a closed subspace of X such that*

$$(\star) \qquad X = \ker\big(f'(p)\big) \oplus X_0.$$

If $f'(p)$ is surjective then $f(\Omega)$ contains a neighborhood of $f(p)$; in other words, there is a neighborhood W of $f(p)$ such that for each element $\eta \in W$ the equation $f(\xi) = \eta$ possesses a solution $\xi \in \Omega$.

Proof. We identify X with $\ker\big(f'(p)\big) \times X_0$ and choose open sets $\Omega_1 \subseteq \ker f'(p)$ and $\Omega_2 \subseteq X_0$ with $p \in \Omega_1 \times \Omega_2 \subseteq \Omega$. Next, we form the partial derivative $\partial_2 f$ with respect to the decomposition $(\star)$. (Note that X_0, being a closed subspace of a Banach space, is a Banach space itself.) Then $(\partial_2 f)(p) : X_0 \to Z$ is an isomorphism. Write $p = (p_1, p_2)$ according to the decomposition $(\star)$ and define $g : \Omega_1 \times \Omega_2 \times Z \to Z$ by $g(x_1, x_2, z) := f(x_1, x_2) - z$; then $(\partial_2 g)(x_1, x_2, z) = (\partial_2 f)(x_1, x_2)$. Now $g\big(p_1, p_2, f(p)\big) = 0$, and $(\partial_2 g)\big(p_1, p_2, f(p)\big) = (\partial_2 f)(p)$ is invertible. Hence the Implicit Function Theorem (1.30) shows that the equation $g(x_1, x_2, z) = 0$ can be locally solved for x_2 in terms of x_1 and z in a neighborhood of $\big(p_1, p_2, f(p)\big)$, say $x_2 = \varphi(x_1, z)$ and hence $0 = g\big(x_1, \varphi(x_1, z), z\big) = f\big(x_1, \varphi(x_1, z)\big) - z$ for all x_1 in a neighborhood U of p_1 and all z in a neighborhood W of $f(p)$. Thus if $z \in W$ then $z = f\big(x_1, \varphi(x_1, z)\big) \in \mathrm{im}(f)$ so that $W \subseteq \mathrm{im}(f)$. ■