CHAPTER 12

Infinite-Dimensional Analysis

In this chapter we discuss foundations of differential calculus in infinite-dimensional spaces and some related questions. In the finite-dimensional case there are two different types of differentiability: differentiability at a point based on the consideration of increments of the function and also a global differentiability based on the consideration of the derivative as some independent object (as this is done in the theory of distributions and in the theory of Sobolev spaces). A similar, but more complicated, picture is observed in the infinite-dimensional case. Here we consider only the first type of differentiability, although now the second one plays an increasingly notable role in research and applications. At present foundations of differentiable calculus in infinite-dimensional spaces are usually not included in courses of functional analysis and are studied in courses of optimization or calculus of variations. However, we have decided to include this short chapter for a more complete representation of main directions of functional analysis as well as due to its conceptual connections with the linear theory.

12.1. Differentiability and Derivatives

Similarly to the case of functions of two real variables one can consider partial derivatives of functions on a linear space and one can also define derivatives in the spirit of the classical "main linear part of the increment of the function". We start with differentiability along directions. Suppose we are given two real linear spaces X and Y and a mapping $F\colon X \to Y$. Suppose that Y is equipped with some convergence (for example, is a locally convex space, as it will be the case everywhere below). Let $x_0 \in X$ and $h \in X$. We shall say that F has a *partial derivative* along h (or in the direction of h) if in Y there exists a limit

$$\partial_h F(x_0) := \lim_{t\to 0} \frac{F(x_0 + th) - F(x_0)}{t}.$$

Even if for every $h \in X$ the partial derivative $\partial_h F(x_0)$ exists, the mapping $h \mapsto \partial_h F(x_0)$ can fail to be linear (see Example 12.1.3). It is often useful to have not only the linearity of this mapping, but also the possibility to approximate F by it "up to an infinitely small quantity of higher order".

Various types of differentiability can be described by the following simple scheme of differentiability with respect to a class of sets $\mathcal{M}$. Let X, Y be locally convex spaces and let $\mathcal{M}$ be some class of nonempty subsets of X.

V. I. Bogachev, O. G. Smolyanov, *Real and Functional Analysis*,
Moscow Lectures 4, https://doi.org/10.1007/978-3-030-38219-3_12

12.1.1. Definition. *The mapping $F\colon X \to Y$ is called differentiable with respect to $\mathcal{M}$ at the point x if there exists a sequentially continuous linear mapping from X to Y, denoted by $DF(x)$ or $F'(x)$ and called the derivative of F at the point x, such that uniformly in h from every fixed set $M \in \mathcal{M}$ we have*

$$\lim_{t\to 0} \frac{F(x+th)-F(x)}{t} = DF(x)h. \tag{12.1.1}$$

Taking for $\mathcal{M}$ the collection of all finite sets we obtain the *Gateaux differentiability* (according to Mazliak [**696**], the correct spelling of Gateaux is without the circumflex, unlike the French word "gâteaux"). Obviously, the Gateaux derivative is unique if it exists. Thus, the Gateaux differentiability differs from the existence of derivatives $\partial_h F(x)$ by the property that we require in addition the linearity of the mapping $h \mapsto \partial_h F(x)$ and also its sequential continuity.

If $\mathcal{M}$ is the class of all compact subsets, we arrive at the *differentiability with respect to the system of compact sets*, which for normed spaces is called the *Hadamard differentiability*. One can also use the *differentiability with respect to the system of sequentially compact sets* (for normed spaces it coincides with the Hadamard differentiability); it is equivalent to the property that, as $t_n \to 0$, $t_n \neq 0$ and $h_n \to h$ in X, we have $[F(x+t_n h_n) - F(x)]/t_n \to DF(x)h$ in Y.

Finally, if X, Y are normed spaces and $\mathcal{M}$ consists of all bounded sets, then we obtain the definition of the *Fréchet differentiability* (of course, such definition can be also considered for locally convex spaces; then this differentiability is called the *bounded differentiability*).

The main idea of differentiability is a local approximation of the mapping F by a linear mapping, i.e., the representation

$$F(x+h) = F(x) + DF(x)h + r(x,h),$$

where the mapping $h \mapsto r(x,h)$ is in a sense an "infinitely small quantity of higher order" as compared to h. In the case of normed spaces the Fréchet differentiability gives the following meaning to this concept of smallness:

$$\lim_{\|h\|\to 0} \frac{\|r(x,h)\|}{\|h\|} = 0.$$

Symbolically this is denoted by $r(x,h) = o(h)$. In the more general case of differentiability with respect to $\mathcal{M}$ the smallness means the uniform (in h from every fixed $M \in \mathcal{M}$) relation

$$\lim_{t\to 0} \frac{r(x,th)}{t} = 0. \tag{12.1.2}$$

It is easy to observe that this condition is equivalent to equality (12.1.1) if we set $r(x,h) := F(x+h) - F(x) - DF(x)h$. Of course, the concept of smallness can be given another sense, which will lead to another type of differentiability. Thus, as for functions on $\mathbb{R}$, the derivative plays the role of some tangent mapping.

12.1.2. Example. Let X be a Hilbert space and let $f(x) = (x,x)$. Then we have $f'(x) = 2x$. Indeed,

$$(x+h, x+h) - (x,x) = 2(x,h) + (h,h) \quad \text{and} \quad (h,h) = o(h).$$

Clearly, for mappings on the real line the differentiabilities of Gateaux, Hadamard and Fréchet coincide. In the space $\mathbb{R}^n$ with $n > 1$ the Hadamard definition is equivalent to the Fréchet definition and is strictly stronger than the Gateaux definition.

12.1.3. Example. (i) Let a function $f\colon \mathbb{R}^2 \to \mathbb{R}^1$ be defined by the formula

$$f(x) = r\cos 3\varphi, \quad x = (r\cos\varphi, r\sin\varphi), \quad f(0) = 0$$

in polar coordinates. At the point $x_0 = 0$ the partial derivatives

$$\partial_h f(x_0) = \lim_{t\to 0} t^{-1} f(th) = \lambda\cos 3\alpha$$

exist for all $h = (\lambda\cos\alpha, \lambda\sin\alpha) \in \mathbb{R}^2$, but the mapping $h \mapsto \partial_h f(x_0)$ is not linear. To see this, it suffices to take the vectors $(1,0)$ and $(0,1)$.

(ii) Let us define a function $f\colon \mathbb{R}^2 \to \mathbb{R}^1$ as follows:

$$f(x) = \begin{cases} 1 & \text{if } x = (x_1, x_2), \text{ where } x_2 = x_1^2 \text{ and } x_1 > 0, \\ 0 & \text{else.} \end{cases}$$

At the point $x = 0$ the Gateaux derivative exists and equals zero, since for every $h \in \mathbb{R}^2$ we have $\lim\limits_{t\to 0} t^{-1} f(th) = 0$ because $f(th) = 0$ if $|t| \leqslant \delta(h)$, where $\delta(h) > 0$. There is no Fréchet differentiability at zero, because $f(h) = 1$ if we take $h = (t, t^2)$.

For a locally Lipschitz (i.e., Lipschitz in a neighborhood of every point) mapping of normed spaces the Gateaux and Hadamard differentiabilities coincide.

12.1.4. Theorem. *Let X and Y be normed spaces and let $F\colon X \to Y$ be a locally Lipschitz mapping. If F is Gateaux differentiable at the point x, then at this point F is also Hadamard differentiable and the corresponding derivatives are equal.*

PROOF. Let K be compact in X and $\varepsilon > 0$. Let F satisfy the Lipschitz condition with constant L on the ball $B(x,r)$ with $r > 0$, let K be contained in the ball $B(0,R)$, and let $M := \max(L, R, \|DF(x)\|)$. We find a finite ε-net $h_1, \ldots, h_m$ in K. There is a number $\delta \in (0, r/R)$ such that, whenever $|t| < \delta$, for every $i = 1, \ldots, m$ we have

$$\|F(x + th_i) - F(x) - tDF(x)(h_i)\| \leqslant \varepsilon|t|.$$

Then, as $|t| < \delta$, for every $h \in K$ we obtain

$$\|F(x + th) - F(x) - tDF(x)(h)\| \leqslant \varepsilon|t| + 2M\varepsilon|t|,$$

since there exists h_i with $\|h - h_i\| \leqslant \varepsilon$, whence

$$\|F(x + th) - F(x + th_i)\| \leqslant M\|th - th_i\| \leqslant M\varepsilon|t|$$

and $\|tDF(x)h - tDF(x)h_i\| \leqslant M\varepsilon|t|$. Thus, F is Hadamard differentiable at x. It is clear that the Hadamard derivative serves as the Gateaux derivative, since the latter is unique. $\square$

In infinite-dimensional Banach spaces the Fréchet differentiability is strictly stronger than the Hadamard differentiability.

12.1.5. Example. The function

$$f\colon\ L^1[0,1] \to \mathbb{R}^1, \quad f(x) = \int_0^1 \sin x(s)\, ds \tag{12.1.3}$$

is everywhere Hadamard differentiable, but nowhere Fréchet differentiable. The same is true for the mapping

$$F\colon\ L^2[0,1] \to L^2[0,1], \quad F(x)(s) = \sin x(s). \tag{12.1.4}$$

PROOF. For the proof of differentiability of a mapping it is often useful to find a candidate for the derivative, which is done by calculating partial derivatives. For the function f we have the following equality:

$$f(x+th) = \int_0^1 \sin[x(s) + th(s)]\, ds.$$

Differentiating in t by the Lebesgue dominated convergence theorem we obtain

$$\partial_h f(x) = \int_0^1 h(s) \cos x(s)\, ds.$$

It is clear that the Gateaux derivative exists and is given by the functional

$$Df(x)h = \int_0^1 h(s) \cos x(s)\, ds.$$

Since $\|Df(x)\| \leqslant 1$, with the aid of the mean value theorem for functions on the real line we conclude that the function f is Lipschitzian (of course, this can be verified directly). By Theorem 12.1.4 we obtain the Hadamard differentiability.

For the mapping F the reasoning is similar. Here we have the operators $DF(x)$ on $L^2[0,1]$ and

$$\bigl(DF(x)h\bigr)(s) = \bigl(\cos x(s)\bigr)h(s).$$

Let us see whether f and F are Fréchet differentiable. Let $x = 0$. Then $f(x) = 0$. We have to check whether the relation $f(h) - Df(0)h = o(\|h\|)$ holds. The left-hand side equals

$$\int_0^1 [\sin h(s) - h(s)]\, ds.$$

Since the Taylor expansion of $\sin h(s) - h(s)$ begins with h^3 and our space is L^1, we may suspect that the Fréchet differentiability fails here. In order to make sure that this is true, we take for h elements of the unit ball for which $f(th) - tDf(0)h$ will not be uniformly $o(t)$. Namely, let $h_k(s) = k$ if $0 \leqslant s \leqslant 1/k$ and $h_k(s) = 0$ if $s > 1/k$. Then

$$f(th_k) - tDf(0)h_k = k^{-1} \sin kt - t.$$

This quantity is not $o(t)$ uniformly in k: it suffices to take $t = k^{-1}$, which gives $t(\sin 1 - 1)$. For an arbitrary point x the reasoning is similar. Let us fix a version of x. We consider the expression

$$\int_0^1 \Big(\sin\big(x(s)+th(s)\big) - \sin x(s) - th(s)\cos x(s)\Big)\, ds.$$

The functions $\cos x(s)$ and $\sin x(s)$ have a common Lebesgue point $s_0 \in (0,1)$. For every ε this point is a Lebesgue point for the function

$$\sin\big(x(s)+\varepsilon\big) - \sin x(s) - \varepsilon\cos x(s).$$

Pick $\varepsilon \in (0,1)$ such that $\sin\big(x(s_0)+\varepsilon\big) - \sin x(s_0) - \varepsilon\cos x(s_0) \neq 0$. Set $h_k = kI_{E_k}$, $E_k = (s_0 - k^{-1}, s_0 + k^{-1})$. For $t = \varepsilon k^{-1}$ we obtain the quantity

$$\int_{E_k} \Big(\sin\big(x(s)+\varepsilon\big) - \sin x(s) - \varepsilon\cos x(s)\Big)\, ds$$

of order of smallness $Lk^{-1} = L\varepsilon^{-1}t$, where $L \neq 0$ is some number, since the limit of this quantity multiplied by $k/2$ is $\sin\big(x(s_0)+\varepsilon\big) - \sin x(s_0) - \varepsilon\cos x(s_0) \neq 0$ as $k \to \infty$. Similar estimates work in the case of F. □

It is worth noting that if the function f is considered not on L^1, but on L^2, then it becomes Fréchet differentiable.

12.1.6. Example. The function f given by formula (12.1.3) on the space $L^2[0,1]$ is everywhere Fréchet differentiable. The mapping F given by (12.1.4) on $C[0,1]$ is everywhere Fréchet differentiable.

PROOF. A nuance making a difference in the properties of f on L^1 and L^2 is that the quantity $|f(x+h) - f(x) - Df(x)h|$ with the aid of the inequality $|\sin(x+h) - \sin x - h\cos x| \leqslant h^2$ is estimated by the integral of h^2, which is the square of the L^2-norm (infinite for some h in L^1). A similar reasoning applies to the mapping F on the space $C[0,1]$. Here $\|F(x+h) - F(x) - DF(x)h\|$ is estimated by $\|h\|^2$ in the case of the sup-norm, but not in the case of L^2-norm, when the indicated estimate leads to the integral of h^4. □

Let us consider one more instructive infinite-dimensional example. It employs the function that is frequently encountered in applications: the distance to a set.

12.1.7. Example. Let X be an infinite-dimensional normed space and let K be a compact set. Set

$$f(x) = \operatorname{dist}(x,K) = \inf\{\|x-y\|:\ y \in K\}.$$

Then the function f satisfies the Lipschitz condition, but is not Fréchet differentiable at points of K. If K is such that $\alpha K \subset K$ whenever $|\alpha| \leqslant 1$ and the set $\bigcup_{n=1}^{\infty} nK$ is everywhere dense in X, then f has the zero Gateaux derivative at the point $0 \in K$. For example, one can take for K the compact ellipsoid

$$K = \Big\{(x_n) \in l^2:\ \sum_{n=1}^{\infty} n^2 x_n^2 \leqslant 1\Big\}$$

in the Hilbert space l^2.

PROOF. Let $x \in K$. Then $f(x) = 0$. Suppose that at x there exists the Fréchet derivative $f'(x)$. This derivative can be only zero, since for every nonzero vector h the function $t \mapsto f(x+th)$ attains its minimum for $t = 0$. We shall arrive at a contradiction if we show that $f(x+h)-f(x)-f'(x)h=f(x+h)$ is not $o(\|h\|)$. For every $n\in\mathbb{N}$ we find a vector h_n such that $\|h_n\|\leqslant 1/n$ and the ball of radius $\|h_n\|/4$ centered at $x+h_n$ does not intersect K. This will give the estimate $f(x+h_n)\geqslant\|h_n\|/4$. The compact set K is covered by finitely many balls of radius $(4n)^{-1}$ centered at points $a_1,\ldots,a_k$. Let L be the finite-dimensional linear space generated by these centers. There is a vector h_n with $\|h_n\|=1/n$ and $\operatorname{dist}(h_n,L)=1/n$. This vector is what we need. Indeed, if there exists a vector $y\in K\cap B(x+h_n,\|h_n\|/4)$, then we obtain the following decomposition: $x=u+l_1$, $y=v+l_2$, where $l_1,l_2\in L$, $\|u\|\leqslant(4n)^{-1}$, $\|v\|\leqslant(4n)^{-1}$ and $\|x+h_n-y\|\leqslant\|h_n\|/4$. Therefore,

$$\|h_n-(l_2-l_1)+u-v\|\leqslant(4n)^{-1},$$

and hence $\|h_n-(l_2-l_1)\|\leqslant 3(4n)^{-1}$ contrary to our choice of h_n, because we have $l_2-l_1\in L$.

Suppose now that K satisfies the indicated additional conditions. We show that at the point $0\in K$ the Gateaux derivative exists and equals zero. For this we have to verify that for each fixed $h\in X$ we have $\lim\limits_{t\to 0}t^{-1}f(th)=0$. Let $\varepsilon>0$. By our condition there exists a vector $v\in nK$ such that $\|h-v\|\leqslant\varepsilon$. Since $tv\in K$ whenever $|t|\leqslant n^{-1}$ by our condition, we have $f(tv)=0$ for such t. Hence $|t^{-1}f(th)|\leqslant\varepsilon$ by the estimate $|f(th)-f(tv)|\leqslant\|th-tv\|\leqslant|t|\varepsilon$, which holds by the Lipschitzness of f. □

In Exercise 12.5.20 it is suggested to verify that if, in addition, the set K is convex, then f has the zero Gateaux derivative at all points of $\bigcup_{0\leqslant t<1}tK$.

For normed spaces one can consider the strict differentiability, which is even stronger than the Fréchet differentiability.

12.1.8. Definition. *Let X and Y be normed spaces, let U be a neighborhood of a point $x_0\in X$, and let $f\colon\ U\to Y$ be a mapping Fréchet differentiable at x_0. If for every $\varepsilon>0$ there exists $\delta>0$ such that, whenever $\|x_1-x_0\|_X\leqslant\delta$ and $\|x_2-x_0\|_X\leqslant\delta$, we have*

$$\|f(x_1)-f(x_2)-f'(x_0)(x_1-x_2)\|_Y\leqslant\varepsilon\|x_1-x_2\|_X,$$

then f is called strictly differentiable at the point x_0.

We observe that a mapping strictly differentiable at x_0 is continuous not only at the point x_0, but also in some ball centered at x_0, since the definition and the triangle inequality yield that in the ball of radius δ centered at x_0 the mapping f satisfies the Lipschitz condition with constant $\|f'(x_0)\|+\varepsilon$. Hence even for scalar functions on the real line the strict differentiability does not reduce to Fréchet differentiability.

If E is a linear subspace in X equipped with some stronger locally convex topology, then one can define the *differentiability along* E (in the respective sense) at the point x as the differentiability at $h=0$ of the mapping $h\mapsto F(x+h)$ from E to Y in the corresponding sense. The derivative along the subspace E is denoted by the symbol $D_{\scriptscriptstyle E}F$. When E is one-dimensional, this gives the usual partial derivative $\partial_h F$.

12.2. Properties of Differentiable Mappings

The most important properties of differentiable mappings include the mean value theorem and the chain rule, i.e., the rule of differentiating the composition. The main role in obtaining multidimensional or infinite-dimensional versions of classical results is played by the corresponding assertions for the real line. However, here there are some subtleties, especially in the infinite-dimensional case, requiring some precautions. First we discuss the differentiability of compositions. Suppose that X, Y and Z are locally convex spaces and mappings

$$F\colon X \to Y \quad \text{and} \quad G\colon Y \to Z$$

are differentiable in a certain sense. Is the mapping $G\circ F\colon X \to Z$ differentiable in the same sense? The answer depends on the type of differentiability. For example, the composition of Gateaux differentiable mappings need not be Gateaux differentiable.

12.2.1. Example. Let us define a mapping $g\colon \mathbb{R}^2 \to \mathbb{R}^2$ by the formula $g\colon (x_1, x_2) \mapsto (x_1, x_2^2)$ and take the function $f\colon \mathbb{R}^2 \to \mathbb{R}^1$ from Example 12.1.3(ii). Then the composition $f\circ g\colon \mathbb{R}^2 \to \mathbb{R}^1$ is not Gateaux differentiable at the point $x = 0$. Moreover, this composition has no partial derivatives at zero along the vectors $(1, 1)$ and $(1, -1)$. Indeed,

$$f\bigl(g(x)\bigr) = \begin{cases} 1 & \text{if } x_1 = |x_2| > 0, \\ 0 & \text{in all other cases.} \end{cases}$$

Note that in this example the inner function is even Fréchet differentiable. It turns out that if the outer function is Fréchet (or Hadamard) differentiable, then the situation becomes better.

12.2.2. Theorem. *Let X, Y and Z be normed spaces, let $\Psi\colon X \to Z$ be the composition of two mappings $F\colon X \to Y$ and $G\colon Y \to Z$, and let $x_0 \in X$ and $y_0 = F(x_0)$. Suppose that the mapping G is Hadamard differentiable at the point y_0. If the mapping F is differentiable at the point x_0 either in the sense of Gateaux or in the sense of Hadamard, then the mapping Ψ is differentiable at x_0 in the same sense and*

$$\Psi'(x_0) = G'(y_0)F'(x_0). \tag{12.2.1}$$

If at the point x_0 the mapping F is Fréchet differentiable and G is differentiable at y_0 also in the Fréchet sense, then Ψ is differentiable at x_0 in the Fréchet sense and (12.2.1) *is fulfilled.*

PROOF. First we observe that if F has the partial derivative $\partial_h F$ at x_0, then there exists the partial derivative

$$\partial_h \Psi(x_0) = G'(y_0)\partial_h F(x_0).$$

Indeed, $F(x_0 + th) = F(x_0) + t\partial_h F(x_0) + r(t) = y_0 + t\partial_h F(x_0) + r(t)$, where $\|r(t)/t\| \to 0$ as $t \to 0$. In addition,

$$G(y_0 + u) - G(y_0) = G'(y_0)u + s(u),$$

where $\lim\limits_{t\to 0} t^{-1}s(tu) = 0$ uniformly in u in every fixed compact set. Hence

$$\Psi(x_0 + th) - \Psi(x_0) = G'(y_0)[t\partial_h F(x_0) + r(t)] + s\bigl(t\partial_h F(x_0) + r(t)\bigr),$$

where

$$\lim_{t\to 0} t^{-1}\Bigl[G'(y_0)r(t) + s\bigl(t\partial_h F(x_0) + r(t)\bigr)\Bigr] = 0$$

in the sense of convergence with respect to the norm in Z.

This yields Gateaux differentiability of the mapping Ψ in the case where the mapping F is Gateaux differentiable.

Suppose that the mapping F is Hadamard differentiable at x_0 and the mapping G is Hadamard differentiable at y_0. For every fixed compact set $K \subset X$ we have

$$F(x_0 + h) = F(x_0) + F'(x_0)h + r(h),$$

where $\lim\limits_{t\to 0} \sup\limits_{h\in K} \|t^{-1}r(th)\| = 0$. Therefore,

$$\begin{aligned}
&\Psi(x_0 + h) - \Psi(x_0) - G'(y_0)F'(x_0)h \\
&\quad = G'(y_0)[F'(x_0)h + r(h)] - G'(y_0)F'(x_0)h + s\bigl(F'(x_0)h + r(h)\bigr) \\
&\quad = G'(y_0)r(h) + s\bigl(F'(x_0)h + r(h)\bigr),
\end{aligned}$$

where $\lim\limits_{t\to 0} \sup\limits_{h\in K} \|t^{-1}G'(y_0)r(th)\| = 0$ and

$$\lim_{t\to 0} \sup_{h\in K} \Bigl\|t^{-1}s\bigl(tF'(x_0)h + r(th)\bigr)\Bigr\| = 0.$$

The last equality follows from the fact that for every sequence $t_n \to 0$ and every sequence $\{h_n\} \subset K$, the sequence of vectors $F'(x_0)h_n + t_n^{-1}r(t_nh_n)$ is contained in the compact set $F'(x_0)(K) + \bigl(\{t_n^{-1}r(t_nh_n)\} \cup \{0\}\bigr)$, because $t_n^{-1}r(t_nh_n) \to 0$ by the definition of Hadamard differentiability.

Finally, in the case of Fréchet differentiability the same reasoning applies with balls in place of compacta. $\square$

A closer look at the proof enables us to modify it for mappings of locally convex spaces.

12.2.3. Theorem. *Let X, Y and Z be locally convex spaces, let a mapping $\Psi\colon X \to Z$ be the composition of two mappings $F\colon X \to Y$ and $G\colon Y \to Z$, and let $x_0 \in X$ and $y_0 = F(x_0)$. Suppose that the mappings F and G are differentiable with respect to the system of compact sets at points x_0 and y_0, respectively, and assume also that the operator $F'(x_0)$ takes compact sets to compact sets. Then Ψ is differentiable at x_0 with respect to the system of compact sets and*

$$\Psi'(x_0) = G'(y_0)F'(x_0).$$

An analogous assertion is true if both mappings are differentiable with respect to the system of bounded sets and the operator $F'(x_0)$ takes bounded sets to bounded sets.

Finally, if the mapping F is Gateaux differentiable at x_0 and the mapping G is differentiable at y_0 with respect to the system of compact sets, then Ψ is Gateaux differentiable at x_0 and the chain rule for $\Psi'(x_0)$ indicated above is true.

PROOF. The reasoning is analogous to the one given above. We only explain the role of additional restrictions on the mapping $F'(x_0)$ that are automatically fulfilled in the case of normed spaces. We have the following representation:

$$\Psi(x_0+h)-\Psi(x_0)-G'(y_0)F'(x_0)h=G'(y_0)r(h)+s\bigl(F'(x_0)h+r(h)\bigr),$$

and we have to show that, for every fixed set K from the class with respect to which F is differentiable, for every sequence $\{h_n\}\subset K$ and every sequence numbers $t_n\to 0$, we have

$$\lim_{n\to\infty} t_n^{-1}G'(y_0)r(t_nh_n)+t_n^{-1}s\bigl(t_nF'(x_0)h_n+t_nt_n^{-1}r(t_nh_n)\bigr)=0.$$

Hence we have to ensure the inclusion of the sequence $F'(x_0)h_n+t_n^{-1}r(t_nh_n)$ to some set from the class $\mathcal{K}$ with respect to which the outer mapping G is differentiable. For this the sequence of vectors $F'(x_0)h_n$ must belong to the class $\mathcal{K}$, since the sequence $t_n^{-1}r(t_nh_n)$ converges to zero and its addition does not influence containment in the classes of bounded or compact sets. □

An analogous assertion is true in the case of differentiability with respect to the system of sequentially compact sets, since $F'(x_0)$ is sequentially continuous (here it is not required that the operator $F'(x_0)$ take compact sets to compact sets).

12.2.4. Example. Suppose that a mapping $F\colon X\to Y$ between normed spaces is differentiable at a point x_0 in the sense of Gateaux, Hadamard or Fréchet and that $G\colon Y\to Z$ is a continuous linear operator with values in a normed space Z. Then the composition $G\circ F$ is differentiable at the point x_0 in the same sense as F.

The situation is similar for differentiable mappings of locally convex spaces in the sense of Gateaux or with respect to the systems of compact sets, sequentially compact sets or bounded sets if the operator G is linear and sequentially continuous.

The next theorem characterizing Hadamard differentiability is also connected with the theorem on the derivative of the composition.

12.2.5. Theorem. *A mapping $F\colon X\to Y$ between normed spaces is Hadamard differentiable at $x_0\in X$ precisely when there exists a continuous linear mapping $L\colon X\to Y$ such that for every mapping $\varphi\colon \mathbb{R}^1\to X$ differentiable at zero with $\varphi(0)=x_0$ the composition $F\circ\varphi\colon \mathbb{R}^1\to Y$ is differentiable at the point 0 and we have $(F\circ\varphi)'(0)=L\varphi'(0)$.*

PROOF. If F is differentiable, then the composition is differentiable too as shown above. Suppose that we have numbers $t_n\to 0$ and vectors $h_n\to h$. A mapping $\varphi\colon \mathbb{R}^1\to X$ will be defined as follows: $\varphi(t_n)=x_0+h_nt_n$ and $\varphi(t)=x_0+th$ if $t\notin\{t_n\}$. Then $\varphi(0)=x_0$ and $t^{-1}[\varphi(t)-\varphi(0)]\to h$ as $t\to 0$, since this difference quotient equals h_n if $t=t_n$ and equals h for all other t. We have

$$\frac{F(x_0+t_nh_n)-F(x_0)}{t_n}=\frac{F\bigl(\varphi(t_n)\bigr)-F\bigl(\varphi(0)\bigr)}{t_n}\to L\varphi'(0)=Lh$$

as $n\to\infty$, which proves Hadamard differentiability of F at the point x_0. □

Note that if F and G are Gateaux, Hadamard or Fréchet differentiable at x, then, as is easily seen, $F+G$ is differentiable at x in the same sense and the equality $(F+G)'(x) = F'(x) + G'(x)$ holds.

Let us proceed to the mean value theorem. We recall that if a function f is differentiable in a neighborhood of the interval $[a,b]$, then

$$f(b) - f(a) = f'(c)(b-a)$$

for some point $c \in (a,b)$. This assertion does not extend in the same form to multidimensional mappings (even to mappings from $\mathbb{R}^1$ to $\mathbb{R}^2$). For example, let

$$f(x) = (\sin x, \cos x), \quad x \in \mathbb{R}^1.$$

Then $f(2\pi) = f(0)$, although the derivative of f does not vanish at any point. The right multidimensional analog of the mean value theorem deals with either inequalities or convex envelopes of the sets of values. We recall that the symbols $\mathrm{conv}\, A$ and $\overline{\mathrm{conv}}\, A$ denote, respectively, the convex envelope and the closed convex envelope of the set A in a locally convex space. The symbol $[a,b]$ denotes the line segment (closed interval) with the endpoints a and b of a linear space, i.e., the set of all vectors of the form $a + t(b-a)$, $t \in [0,1]$. Similarly we define (a,b).

12.2.6. Theorem. *Let X and Y be locally convex spaces, let U be an open convex set in X, and let a mapping $F\colon U \to Y$ be Gateaux differentiable at every point in U. Then for every $a, b \in U$ one has the inclusion*

$$F(b) - F(a) \in \overline{\mathrm{conv}}\,\{F'(c)(a-b)\colon\ c \in (a,b)\}. \tag{12.2.2}$$

PROOF. Let E denote the set in the right-hand side. Let $l \in Y^*$. The function $\varphi\colon\ t \mapsto l\bigl(F(a+tb-ta)\bigr)$ is defined in a neighborhood of $[0,1]$ by our assumption. In addition, it is differentiable in a neighborhood of $[0,1]$. By the classical mean value theorem there exists a point $t \in (0,1)$ such that

$$l\bigl(F(b)\bigr) - l\bigl(F(a)\bigr) = \varphi(1) - \varphi(0) = \varphi'(t) = l\bigl(F'(c)(b-a)\bigr) \leqslant \sup_{y \in E} |l(y)|,$$

where $c = a + t(b-a) \in U$. By a corollary of the Hahn–Banach theorem we conclude that $F(b) - F(a) \in E$. $\square$

Let us give a number of important corollaries.

12.2.7. Corollary. *Suppose that in the situation of the previous theorem we are given a sequentially continuous linear mapping $\Lambda\colon\ X \to Y$ (for example, $\Lambda = F'(a)$). Then for every $a, b \in U$ one has the inclusion*

$$F(b) - F(a) - \Lambda(b-a) \in \overline{\mathrm{conv}}\,\{[F'(c) - \Lambda](a-b)\colon\ c \in (a,b)\}. \tag{12.2.3}$$

PROOF. It suffices to apply the theorem above to the mapping $F - \Lambda$. $\square$

12.2.8. Corollary. *Let X and Y be normed space, let U be an open convex set in X, and let a mapping $F\colon\ U \to Y$ be Gateaux differentiable at every point in U. Then, for all $a, b \in U$, we have*

$$\|F(b) - F(a)\| \leqslant \sup_{c \in (a,b)} \|F'(c)\|\, \|a-b\|. \tag{12.2.4}$$

PROOF. For every $c \in U$ we have $\|F'(c)(b-a)\| \leqslant \|F'(c)\|\,\|b-a\|$, which by (12.2.2) gives (12.2.4). □

12.2.9. Corollary. *Suppose that in the previous corollary we are given a continuous linear mapping $\Lambda\colon X \to Y$. Then, for all $a, b \in U$, we have*

$$\|F(b) - F(a) - \Lambda(b-a)\| \leqslant \sup_{c \in (a,b)} \|F'(c) - \Lambda\|\,\|a-b\|. \tag{12.2.5}$$

12.2.10. Corollary. *Let X and Y be normed spaces, let U be an open convex set in X, and let a mapping $F\colon U \to Y$ be Gateaux differentiable at every point in U. Suppose that the mapping $x \mapsto F'(x)$ from U to the space of operators $\mathcal{L}(X,Y)$ with the operator norm is continuous at some point $x_0 \in U$. Then the mapping F is Fréchet differentiable at the point x_0. Moreover, F is strictly differentiable at x_0.*

PROOF. Let $\varepsilon > 0$ be such that the ball $B(x_0, \varepsilon) = \{x\colon \|x - x_0\| < \varepsilon\}$ is contained in U. As shown above, for all h with $\|h\| < \varepsilon$ we have

$$\|F(x_0 + h) - F(x_0) - F'(x_0)(h)\| \leqslant \sup_{c \in B(x_0,\varepsilon)} \|F'(c) - F'(x_0)\|\,\|h\|.$$

By the continuity of F' at x_0 we have

$$\lim_{\varepsilon \to 0} \sup_{c \in B(x_0,\varepsilon)} \|F'(c) - F'(x_0)\| = 0,$$

which gives Fréchet differentiability. The strict differentiability follows easily from estimate (12.2.5). □

If the derivative of the mapping F is continuous with respect to the operator norm in a domain Ω, then F is called a C^1-mapping in Ω.

12.3. Inverse and Implicit Functions

In this section we consider the local invertibility of nonlinear mappings and existence of a functional dependence $y = y(x)$ between solutions of equations of the type $F(x,y) = 0$. Results of this kind are called, respectively, inverse function theorems and implicit function theorems. At present there exists a well developed theory covering such problems, but here we confine ourselves to some simplest infinite-dimensional analogs of the facts known from finite-dimensional calculus. Even these simplest theorems have many interesting and important applications. The results presented here do not employ refined tools and techniques and are applications of the contracting mapping theorem or some analogous reasoning. However, the method of application is elegant and instructive. The idea is that a given mapping is locally approximated by a simpler (in some sense) mapping. For such a simpler mapping in this section we take the identity mapping and linear invertible operators.

We start with the consideration of Lipschitzian homeomorphisms.

12.3.1. Theorem. *Let $U = B(a,r)$ be an open ball of radius $r > 0$ centered at a point a in a Banach space X and let $F\colon U \to X$ be a mapping such that*

$$\|F(x) - F(y)\| \leqslant \lambda \|x - y\| \quad \forall\, x, y \in U,$$

where $\lambda \in [0,1)$ *is a constant. Then there exists an open neighborhood* V *of* a *such that the mapping* $\Psi\colon\ x \mapsto x + F(x)$ *is a homeomorphism of* V *and the open ball* $W := B\bigl(\Psi(a), r(1-\lambda)\bigr)$. *The inverse mapping* $\Psi^{-1}\colon\ W \to V$ *satisfies the Lipschitz condition with the constant* $(1-\lambda)^{-1}$.

If, in addition, $\|F(a)\| < r(1-\lambda)/2$, *then* Ψ *is a homeomorphism of some neighborhood of the point* a *and the open ball* $B\bigl(a, r(1-\lambda)/2\bigr)$.

PROOF. Replacing F by $F - F(a)$, we can assume that $F(a) = 0$. Let us apply Theorem 1.4.4 to the space $T = W$ and the mapping $(w,x) \mapsto w - F(x)$ from $W \times U$ to X. All necessary conditions are fulfilled. In particular, the inequality $\|w - F(a) - a\| = \|w - a\| < r(1-\lambda)$ holds for all $w \in W$. Hence there exists a continuous mapping $f\colon\ W \to U$ such that we have $f(w) = w - F\bigl(f(w)\bigr)$ for all $w \in W$, i.e., $\Psi\bigl(f(w)\bigr) = w$.

Let us show that f is a homeomorphism from W onto $f(W)$. It is clear that f is an injective mapping. The mapping Ψ on U is also injective, since the equality $\Psi(u_1) = \Psi(u_2)$ yields that

$$\|u_1 - u_2\| = \|F(u_1) - F(u_2)\| \leqslant \lambda \|u_1 - u_2\|.$$

This is only possible if $u_1 = u_2$. Thus, the mapping Ψ is a homeomorphism from the set $V := f(W)$ onto W, inverse to f. The set $V = \Psi^{-1}(W)$ is open by the continuity of Ψ. In addition, $a \in V$, since $\Psi(a)$ is the center of W.

The fact that f is Lipschitz with constant $(1-\lambda)^{-1}$ is clear from the equality

$$f(w) - f(w') = w - w' + F\bigl(f(w')\bigr) - F\bigl(f(w)\bigr),$$

which gives the estimate $\|f(w) - f(w')\| \leqslant \|w - w'\| + \lambda\|f(w) - f(w')\|$, i.e., $\|f(w) - f(w')\| \leqslant (1-\lambda)^{-1}\|w - w'\|$. If, in addition, $\|F(a)\| < r(1-\lambda)/2$, then the ball $B\bigl(a, r(1-\lambda)/2\bigr)$ belongs to W. $\square$

Now we apply the facts established above to differentiable mappings. First we clarify the condition for the differentiability of the inverse mapping for a differentiable homeomorphism.

12.3.2. Proposition. *Let* X *and* Y *be Banach spaces, let* $U \subset X$ *and* $V \subset Y$ *be open sets, and let* $F\colon\ U \to V$ *be a homeomorphism that is Fréchet differentiable at some point* $a \in U$. *For Fréchet differentiability of the mapping* $G = F^{-1}\colon\ V \to U$ *at the point* $b = F(a)$ *it is necessary and sufficient that the operator* $F'(a)$ *map* X *one-to-one onto* Y.

PROOF. If the mapping G is Fréchet differentiable at the point b, then by the chain rule we have $G'(b)F'(a) = I_X$ and $F'(a)G'(b) = I_Y$.

Suppose now that the operator $\Lambda := F'(a)$ is invertible. Passing to the mapping $\Lambda^{-1}F$, we reduce our assertion to the case $X = Y$ and $F'(a) = I$. In addition, we can assume that $a = 0$ and $F(a) = 0$. We have to show that $G'(0) = I$, i.e., that

$$G(y) - y = o(\|y\|) \quad \text{as } \|y\| \to 0.$$

We have

$$F(x) = x + \|x\|\varphi(x), \qquad \lim_{\|x\|\to 0} \|\varphi(x)\| = 0. \tag{12.3.1}$$

Since F is a homeomorphism, we have $y = F(x)$, where $x = G(y)$, for all y in some neighborhood of zero, and for the corresponding x one has (12.3.1). Thus,

$$G(y) = x = y - \|x\|\varphi(x).$$

We have to show that $\|x\|\varphi(x) = o(\|y\|)$. Let us find a neighborhood of zero U_0 such that $\|\varphi(x)\| < 1/2$ for all $x \in U_0$. Then $\|x\| \leqslant 2\|y\|$ for such x and $y = F(x)$, i.e., $\|x\| \, \|\varphi(x)\| \leqslant 2\|y\| \, \|\varphi(x)\|$. It remains to observe that $\|x\| \to 0$ as $\|y\| \to 0$ by the continuity of G and hence $\varphi(x) \to 0$. $\square$

12.3.3. Theorem. *Let F be a continuously differentiable mapping from the open ball $U = B(x_0, r)$ in a Banach space X to a Banach space Y. Suppose that the operator $\Lambda := F'(x_0)$ maps X one-to-one onto Y. Then F maps some neighborhood V of the point x_0 one-to-one onto some neighborhood W of $F(x_0)$, moreover, the mapping $G := F^{-1}\colon W \to V$ is continuously differentiable and*

$$G'(y) = \bigl[F'\bigl(F^{-1}(y)\bigr)\bigr]^{-1}, \quad y \in W.$$

PROOF. The assertion reduces to the case $Y = X$ and $F'(x_0) = I$ if we pass to the mapping $\Lambda^{-1}F$. In addition, we can assume that $x_0 = 0$ and $F(x_0) = 0$. Having in mind to apply Theorem 12.3.1, we write F as $F(x) = x + F_0(x)$, where $F_0(x) = F(x) - x$. By the continuity of the derivative at x_0, there exists a ball $B(0, r)$ such that $\|F'(x) - I\|_{\mathcal{L}(X)} \leqslant 1/2$ for all $x \in B(0, r)$. By the mean value theorem the mapping F_0 is Lipschitz on $B(0, r)$ with constant $1/2$ (any constant less than 1 would suit for us). In addition, $F_0(0) = 0$. By the theorem cited, in the ball $W = B(0, r/2)$ we have a continuous inverse mapping G for F that takes this ball to some neighborhood of zero V. Proposition 12.3.2 ensures Fréchet differentiability of G at zero and the equality $G'(0) = I$. Note that so far we have used only the continuity of the derivative at zero. Since we assumed the continuity of the derivative at all points of U, in a neighborhood of x_0 this derivative is invertible, which shows the differentiability of the inverse mapping in a neighborhood of the point $F(x_0)$. Finally, the continuity of the mapping $y \mapsto G'(y)$ in this neighborhood follows from the formula $G'(y) = \bigl[F'\bigl(G(y)\bigr)\bigr]^{-1}$ on account of the continuity of G and the continuity of the mapping $A \mapsto A^{-1}$ on the set of invertible operators. $\square$

We now obtain a criterion for F to be a diffeomorprhism.

12.3.4. Corollary. *Let F be a C^1-mapping from an open set U in a Banach space X to a Banach space Y. In order F be a C^1-diffeomorphism of the set U onto an open set in Y, it is necessary and sufficient that F be injective and the derivative $F'(x)$ be an invertible operator from X onto Y for all $x \in U$.*

PROOF. The necessity of these conditions is obvious. Let us prove their sufficiency. We already know that every point x in U possesses a neighborhood $U_x \subset U$ such that F is a C^1-diffeomorphism of U_x onto some open ball V_x centered at $F(x)$. Hence the set $F(U) = \bigcup_{x \in U} V_x$ is open. It follows from the results above that the inverse mapping $F^{-1}\colon V \to U$ is continuously differentiable. $\square$

We have already noted that in Theorem 12.3.3 the existence and differentiability of G at zero has been actually obtained under weaker conditions. Namely, we only needed (under the assumption that $x_0 = 0$ and $F(x_0) = 0$) that the auxiliary mapping $F_0(x) = F'(0)^{-1}F(x) - x$ be contracting in a neighborhood of zero. This condition is certainly weaker than the continuity of the derivative at zero and obviously follows from the strict differentiability of F at zero, which gives

$$F'(0)^{-1}F(x) - F'(0)^{-1}F(y) - x + y = \|x - y\|\psi(x, y)$$

with $\lim\limits_{x,y\to 0} \psi(x, y) = 0$. Hence F_0 is Lipschitz with an arbitrarily small constant in a suitable neighborhood of zero. This gives the following result.

12.3.5. Theorem. *Suppose that a mapping F from an open ball with center x_0 in a Banach space X to a Banach space Y is strictly differentiable at the point x_0 and the operator $D := F'(x_0)$ maps X one-to-one onto Y. Then F homeomorphically maps some neighborhood V of the point x_0 onto a neighborhood W of the point $F(x_0)$ and the mapping $G := F^{-1}\colon W \to V$ is Fréchet differentiable at the point $y_0 = F(x_0)$ and $G'(y_0) = D^{-1}$.*

12.3.6. Example. Let X be a Hilbert space and let $F(x) = Ax + B(x, x)$, where A is an invertible linear operator on X and $B\colon X\times X \to X$ is a continuous mapping linear in every argument. Then $F'(x)h = Ah + B(x, h) + B(h, x)$, since $\|B(h, h)\| \leqslant C|h|^2$, which is easily derived from the boundedness of B on some ball centered at zero. Hence $F'(0) = A$. Therefore, F maps diffeomorphically some neighborhood of zero onto a neighborhood of zero.

Let us proceed to the implicit function theorem. Suppose we are given three Banach spaces X, Y and Z and a continuously differentiable mapping F from an open set $U \subset X \times Y$ to Z. Suppose that for some point $(a, b) \in U$ we have $F(a, b) = 0$. We are interested in other solutions (x, y) to the equation $F(x, y) = 0$, sufficiently close to (a, b). Moreover, we would like to represent (locally) the set of solutions as a surface $y = f(x)$. It turns out that for this we need only one condition.

12.3.7. Theorem. *Suppose that in the situation described above the derivative $F'_Y(a, b)$ of the mapping F along Y at the point (a, b) is a linear isomorphism between Y and Z. Then there exists a neighborhood V_a of the point a in X, a neighborhood W_b of the point b in Y and a continuously differentiable mapping $f\colon V_a \to Y$ with the following properties: $V_a\times W_b \subset U$ and the conditions*

$$(x, y) \in V_a\times W_b \quad \textit{and} \quad F(x, y) = 0$$

are equivalent to the conditions

$$x \in V_a \quad \textit{and} \quad y = f(x).$$

In particular, $f(a) = b$. Thus, in the domain $V_a\times W_b$ all solutions to the equation $F(x, y) = 0$ are given by the formula $y = f(x)$.

PROOF. We shall reduce this theorem to the inverse mapping theorem. For this we consider an auxiliary mapping

$$F_1\colon\ U \to X\times Z, \quad (x,y)\mapsto \bigl(x, F(x,y)\bigr).$$

It is readily seen that this mapping is continuously differentiable and

$$F_1'(x,y)(u,v) = \bigl(u, F_X'(x,y)u + F_Y'(x,y)v\bigr), \quad u\in X, v\in Y.$$

Since $F_Y'(a,b)$ is a linear isomorphism between Y and Z, the mapping $F_1'(a,b)$ is an isomorphism between $X\times Y$ and $X\times Z$. Indeed, for any given $h\in X$ and $k\in Z$, there is a uniquely determined $u=h$ and then we take

$$v = [F_Y'(a,b)]^{-1}\bigl(k - F_X'(a,b)h\bigr)$$

with $F_1'(a,b)(u,v) = (h,k)$. As shown above, there exists a neighborhood U_0 of the point (a,b) that is mapped by F_1 diffeomorphically onto some neighborhood W of the point $\bigl(a, F(a,b)\bigr) = (a,0)$. In U_0 we can take a smaller neighborhood of the form $V_a\times W_b$, where V_a is a neighborhood of a and W_b is a neighborhood of b. Let $W_0 = F_1(V_a\times W_b)$. The diffeomorphism inverse to F_1 can be written in the form $(x,z)\mapsto \bigl(x,\varphi(x,z)\bigr)$, where the mapping $(x,z)\mapsto\varphi(x,z)$ with values in Y is defined in a neighborhood of the point $(a,0)$ in $X\times Z$. We let $f(x)=\varphi(x,0)$ and show that we have obtained a desired mapping. Indeed, this mapping is continuously differentiable. In addition, the conditions $(x,y)\in V_a{\times}W_b$ and $F(x,y)=0$ are equivalent to the conditions $(x,0)\in W_0$ and $\varphi(x,0)=y$ by the injectivity of the mapping F_1 on $V_a\times W_b$. The set $V_a' := \{x\colon\ (x,0)\in W_0\}$ is an open neighborhood of the point a. Replacing V_a by a smaller neighborhood $V_a\cap V_a'$, we obtain the desired properties for f. We have actually proved even more: for all z in some neighborhood of zero the solutions (x,y) to the equation $F(x,y)=z$ belonging to a sufficiently small neighborhood of (a,b) have a parametric representation $y=\varphi(x,z)$. □

Differentiating the identity $F\bigl(x,f(x)\bigr)=0$, we find that

$$f'(x) = -\bigl[F_Y'\bigl(x,f(x)\bigr)\bigr]^{-1}F_X'\bigl(x,f(x)\bigr)$$

in a neighborhood of a. In particular,

$$f'(a) = -[F_Y'(a,b)]^{-1}F_X'(a,b).$$

It is clear from the proof that the obtained representation gives all solutions (x,y) from a sufficiently small neighborhood of (a,b). But this assertion should not be understood in the sense that in a sufficiently small neighborhood of a there are no other differentiable mappings f with $F\bigl(x,f(x)\bigr)=0$: we are talking only about f such that $y=f(x)$ belongs to a small neighborhood of b. For example, for the equation $x^2+y^2=1$ on the plane in every neighborhood of the point $a=0$ on the first coordinate line there are two differentiable functions $f_1(x)=\sqrt{1-x^2}$ and $f_2(x)=-\sqrt{1-x^2}$ for which $x^2+f_1(x)^2=1$ and $x^2+f_2(x)^2=1$. Here f_1 uniquely defines solutions from a neighborhood of the point $(0,1)$ and f_2 uniquely describes solutions from a neighborhood of the point $(0,-1)$.

It follows from our discussion that the mapping f is unique in the following sense: if a mapping f_0 from a neighborhood of the point a to the set W_b satisfies the equality $F\bigl(x, f_0(x)\bigr) = 0$, then f_0 coincides with f in some neighborhood of the point a.

If we compare the theorem on the implicit function with the theorem on the inverse function, then it is easy to observe that it corresponds to Theorem 12.3.3, but not to Theorem 12.3.5. It is also possible to obtain an analog of the latter.

12.3.8. Theorem. *Let F be a mapping from a neighborhood $U \subset X\times Y$ of the point (a,b) to the space Z such that $F(a,b) = 0$ and $\lim\limits_{x\to a} F(x,b) = 0$. Suppose that F is differentiable along the space Y at (a,b) and for every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$\|F(x,y_1) - F(x,y_2) - F'_Y(a,b)(y_1 - y_2)\| \leqslant \varepsilon\|y_1 - y_2\|$$

if $\|x - a\| \leqslant \delta$, $\|y_1 - b\| \leqslant \delta$, $\|y_2 - b\| \leqslant \delta$ (for example, the derivative F'_Y exists and is continuous in U). Suppose that the operator $F'_Y(a,b)$ from Y onto Z is invertible. Then there exists a mapping f with values in Z defined in some neighborhood V_a of the point a and possessing the following properties: $f(a) = b$ and $F\bigl(x, f(x)\bigr) = 0$ for all $x \in V_a$.

PROOF. Without loss of generality we can assume that $a = 0$ and $b = 0$. Passing to the mapping $F'_Y(a,b)^{-1}F$, we can also assume that $Z = Y$ and $F'_Y(a,b) = I$. Let us find $r_0 > 0$ such that U contains the ball of radius r_0 centered at zero. For every $x \in B(0,r_0)$ we consider the mapping $\Psi_x\colon\ y \mapsto y - F(x,y)$. By assumption there exists $r \in (0,r_0)$ such that for all $y_1, y_2 \in B(0,r)$, $x \in B(0,r)$ we have

$$\|F(x,y_1) - F(x,y_2) - (y_1 - y_2)\| \leqslant \tfrac{1}{2}\|y_1 - y_2\|.$$

Hence for all $x \in B(0,r)$ the mapping Ψ_x on the ball $B(0,r)$ satisfies the Lipschitz condition with constant $1/2$. By assumption there exists $\delta \in (0,r)$ such that $\|F(x,0)\| < r/2$ if $\|x\| < \delta$. For such x we have $\|\Psi_x(0)\| = \|F(x,0)\| < r/2$. Thus, we are under the assumptions of Theorem 1.4.4, which gives a mapping $f\colon\ B(0,\delta) \to Y$ with $f(x) = f(x) - F\bigl(x, f(x)\bigr)$ for all $x \in B(0,\delta)$. Hence for such points x we have $F\bigl(x, f(x)\bigr) = 0$. $\square$

12.3.9. Remark. If in the situation of this theorem the mapping F is Fréchet differentiable and its derivative is continuous at the point (a,b), then the mapping f constructed above is differentiable at the point a and

$$f'(a) = -[F'_Y(a,b)]^{-1}F'_X(a,b).$$

This is proved in the same way as in Theorem 12.3.7.

12.4. Higher Order Derivatives

If a mapping $F\colon\ X \to Y$ between locally convex spaces is differentiable, then a new mapping $F'\colon\ x \mapsto F'(x)$ arises with values in the space of linear mappings from X to Y. We can equipped this space with some locally convex topology and study the differentiability of the mapping F', denoting its derivative (in a suitable sense) by F''.

In the case of Fréchet differentiability in normed spaces this leads to considering the derivative $F'\colon X \to \mathcal{L}(X,Y)$, $x \mapsto F'(x)$ of a Fréchet differentiable mapping $F\colon X \to Y$ of normed spaces. The space of operators $\mathcal{L}(X,Y)$ is equipped with the operator norm, so F' becomes a mapping with values in a normed space and we can speak of its Fréchet differentiability at the point x. If F' has a Fréchet derivative at x, then it is denoted by $F''(x)$ or $D^2F(x)$. In this case $F''(x) \in \mathcal{L}\bigl(X, \mathcal{L}(X,Y)\bigr)$.

We observe (see §12.5) that the space of operators $\mathcal{L}\bigl(X, \mathcal{L}(X,Y)\bigr)$ can be canonically identified with the space $\mathcal{L}_2(X,X,Y)$ of bilinear continuous mappings from $X \times X$ to Y. To this end, to every operator $\Lambda \in \mathcal{L}\bigl(X, \mathcal{L}(X,Y)\bigr)$ we set in correspondence a bilinear continuous mapping $\widehat{\Lambda}$ by the formula

$$\widehat{\Lambda}(u,v) = [\Lambda(u)](v).$$

If $\Lambda = F''(x)$, then the value of $\widehat{\Lambda}$ on the pair (u,v) is evaluated by the formula

$$\widehat{\Lambda}(u,v) = \partial_v \partial_u F(x).$$

Conversely, to every continuous bilinear mapping B from $X\times X$ to Y we associate a linear mapping Λ from X to the space of linear mappings from X to Y by the formula

$$[\Lambda(u)](v) := B(u,v).$$

We have $\Lambda \in \mathcal{L}\bigl(X, \mathcal{L}(X,Y)\bigr)$ and $\widehat{\Lambda} = B$.

12.4.1. Example. Let X be a Hilbert space. Set $f(x) = (x,x)$. Then $f'(x) = 2x$ and $f''(x) = 2I$.

The higher order differentiability is defined by induction: a mapping F at a point x_0 has a derivative of order $k > 1$ if in some neighborhood of x_0 there exist derivatives $DF(x), \ldots, D^{k-1}F(x)$ and the mapping $x \mapsto D^{k-1}F(x)$ is differentiable at x_0.

The Fréchet derivative $D^kF(x_0)$ of order k is often identified with a continuous k-linear mapping from the space X^k to Y (see §12.5(ii)), i.e., the convention is that $D^kF(x_0) \in \mathcal{L}_k(X^k, Y)$.

12.4.2. Example. Let X and Y be normed space and let $\psi\colon X \to Y$ be the continuous homogeneous polynomial of degree k generated by a continuous symmetric k-linear mapping Ψ_k. Then the mapping ψ is infinitely Fréchet differentiable and

$$D\psi(x)h = k\Psi_k(x, \ldots, x, h).$$

For the proof we write $\Psi_k(x+h, \ldots, x+h) - \Psi_k(x, \ldots, x)$ using the additivity of Ψ_k in every argument. We obtain the term $k\Psi_k(x, \ldots, x, h)$ and a sum of terms in which h enters as an argument of Ψ_k at least twice. So this sum is $o(\|h\|)$, which proves our claim.

12.4.3. Theorem. *Let a mapping F between normed spaces X and Y have a derivative of order k at a point x_0. Then the multilinear mapping $D^kF(x_0)$ is symmetric. In particular, for $k = 2$ we have*

$$D^2F(x_0)(u,v) = D^2F(x_0)(v,u).$$

PROOF. It suffices to verify this assertion for scalar functions considering compositions $l \circ F$, where l is an element of Y^*. Then everything reduces to the case $X = \mathbb{R}^k$ known from calculus. For example, for $k = 2$ we have to take the function $f(t_1, t_2) := F(x_0 + t_1 u + t_2 v)$ on $\mathbb{R}^2$. Then

$$\frac{\partial}{\partial t_1} f(t_1, t_2) = DF(x_0 + t_1 u + t_2 v)(u),$$

whence $D^2F(x_0)(u, v) = \frac{\partial}{\partial t_2}\frac{\partial}{\partial t_1} f|_{t_1=t_2=0}$. A similar equality with the interchanged partial derivatives in t_1 and t_2 is true for $D^2F(x_0)(v, u)$. □

12.4.4. Theorem. (TAYLOR'S FORMULA) *Let U be an open set in a normed space X and let a mapping $F\colon X \to Y$ with values in a normed space Y have Fréchet derivatives up to order $n-1$ in U.*

(i) *If at a point $x_0 \in U$ there exists the Fréchet derivative $D^nF(x_0)$, then, as $\|h\| \to 0$, we have*

$$\|F(x_0 + h) - F(x_0) - DF(x_0)h - \cdots - \frac{1}{n!}D^nF(x_0)(h, \ldots, h)\| = o(\|h\|^n).$$

(ii) *If in U there exists the Fréchet derivative $D^nF(x)$ and $\|D^nF(x)\| \leqslant M$, then, for every $x \in U$ and every $h \in X$ such that the interval with endpoints in x and $x + h$ belongs to U, we have*

$$F(x+h) = F(x) + DF(x)h + \cdots + \frac{1}{(n-1)!}D^{n-1}F(x)(h, \ldots, h) + \frac{1}{(n-1)!}\int_0^1 (1-t)^{n-1}D^nF(x+th)(h, \ldots, h)\,dt. \quad (12.4.1)$$

In addition, the following estimate holds:

$$\|F(x+h) - F(x) - DF(x)h - \cdots - \frac{1}{(n-1)!}D^{n-1}F(x)(h, \ldots, h)\| \leqslant M\frac{\|h\|^n}{n!}.$$

PROOF. (i) Let us apply induction on n. For $n = 1$ our assertion is true by the definition of the Fréchet derivative. Suppose it is true for some $n = k$. If the mapping F has the derivative $D^{k+1}F(x_0)$ at x_0, then the mapping

$$G\colon\ h \mapsto F(x_0 + h) - F(x_0) - DF(x_0)h - \cdots - \frac{1}{(k+1)!}D^{k+1}F(x_0)(h, \ldots, h)$$

is differentiable in a neighborhood of zero. According to Example 12.4.2 we have

$$DG(h)v = DF(x_0 + h)v - DF(x_0)v - \cdots - \frac{1}{k!}D^{k+1}F(x_0)(h, \ldots, h, v),$$

which can be written as

$$DG(h) = \Psi(x_0 + h) - \Psi(x_0) - D\Psi(x_0)v - \cdots - \frac{1}{k!}D^k\Psi(x_0)(h \ldots, h),$$

where $\Psi(x) := DF(x)$. Applying the inductive assumption to the mapping Ψ, we obtain the relation $\|DG(h)\| = o(\|h\|^k)$. Since $G(0) = 0$, we have the estimate $\|G(h)\| \leqslant \sup_{0\leqslant t\leqslant 1} \|DG(th)\| \cdot \|h\| = o(\|h\|^{k+1})$ by the mean value theorem, which proves our assertion for $n = k + 1$.

(ii) In the proof of (12.4.1) we can assume that the space Y is separable, since on the interval with endpoints at x and $x + h$ the mapping F is continuous and hence takes values in a separable subspace of Y. In addition, we can pass to the completion of the space Y and assume that it is Banach (then the integral in (12.4.1) will be an element of the original space). The Y-valued mapping $t \mapsto (1-t)^{n-1} D^n F(x + th)(h, \dots, h)$ is measurable and bounded by our condition, hence it is Lebesgue integrable (see Example 6.10.66). It suffices to verify the equality of both parts of the regarded equality under the action of elements of Y^*. This reduces our assertion to scalar functions on the interval if we pass to the mapping $\varphi(t) = F(x + th)$. The second assertion in (ii) is a simple corollary of the first one. □

As in the case of real functions on $\mathbb{R}$, in applications the first two derivatives turn out to be the most useful and frequently used. These derivatives describe the behavior of the function in a neighborhood of a point of local extremum.

12.4.5. Theorem. *Let U be an open set in a normed space X and let $f\colon X \to \mathbb{R}$ be a Gateaux differentiable function.*

(i) *If f has a minimum at the point $x_0 \in U$, then $Df(x_0) = 0$.*

(ii) *If the function f is twice Fréchet differentiable at the point $x_0 \in U$ and has a minimum at this point, then $D^2 f(x_0)(h, h) \geqslant 0$ for all $h \in X$.*

(iii) *Let the function f be twice Fréchet differentiable in U and let $x_0 \in U$ be a point such that $Df(x_0) = 0$ and for some $\lambda > 0$ we have $D^2 f(x_0)(h, h) \geqslant \lambda \|h\|^2$ for all $h \in X$. Then x_0 is a point of a strict local minimum.*

PROOF. Assertions (i) and (ii) follow from the one-dimensional case applied to the function $t \mapsto f(x + th)$, and (iii) follows from assertion (i) of the previous theorem, giving the estimate $f(x_0 + h) - f(x_0) \geqslant \lambda \|h\|^2/2$ for small $\|h\|$. □

The condition $D^2 f(x_0)(h, h) > 0 \ \forall\, h \neq 0$ is not sufficient here. For example, for the function $f(x) = \sum_{n=1}^\infty (n^{-3} x_n^2 - x_n^4)$ on the space l^2 it holds for $x_0 = 0$, but x_0 is not a point of local minimum.

12.5. Complements and Exercises

12.5(i). Newton's method

The known Newton method of solving the equation $f(x) = 0$ with a smooth function f on the interval $[a, b]$ employs recurrent approximations

$$x_{n+1} := x_n - f(x_n)/f'(x_n), \quad x_0 = a.$$

One can show that if $f'(x) \neq 0$ on $[a, b]$ and x^* is the unique root of this equation, then $x_n \to x^*$. In the infinite-dimensional case there is a modification of Newton's method, the so-called Newton–Kantorovich method, worked out by L.V. Kantorovich. We shall give a typical result, referring for a more detailed discussion to [**312**, Chapter XVIII]. Let X and Y be Banach spaces and let $F\colon X \to Y$ be a Fréchet differentiable mapping in the ball $B = B(x_0, r)$ such that its derivative F' satisfies the Lipschitz condition with constant L, i.e., $\|F'(x) - F'(y)\| \leqslant L\|x-y\|$. Suppose that the operator $F'(x_0)$ is invertible . Then we can set

$$x_{n+1} := x_n - [F'(x_0)]^{-1}\big[F(x_n)\big].$$

These modified approximations possess worse convergence properties, but use the derivative only at the point x_0. Let $M = \big\|[F'(x_0)]^{-1}\big\|$, $k = \big\|[F'(x_0)]^{-1}\big(F(x_0)\big)\big\|$, $h = MkL$, let t_0 be the least root of the equation $ht^2 - t + 1 = 0$, and let $r \geqslant kt_0$.

12.5.1. Theorem. *If $h < 1/4$, then in the ball $\{x\colon \|x - x_0\| \leqslant kt_0\}$ the equation $F(x) = 0$ has a unique solution x^*, $x^* = \lim\limits_{n\to\infty} x_n$, moreover, the bound $\|x^* - x_n\| \leqslant kq^n/(1-q)$ holds, where $q = ht_0 < 1/2$.*

PROOF. Passing to the mapping $[F'(x_0)]^{-1}F$, we can assume that $X = Y$, $F'(x_0) = I$, $M = 1$. In addition, we can assume that $x_0 = 0$. Let us set $\Psi(x) = x - F(x)$. Since the derivative of the mapping $x \mapsto x - F(x) + F(0)$ equals $I - F'(x)$, by the mean value theorem

$$\begin{aligned}\|\Psi(x)\| &= \|x - F(x) + F(0) - F(0)\| \leqslant \|x - F(x) + F(0)\| + k\\ &\leqslant \sup_{\|y\|\leqslant\|x\|} \|I - F'(y)\|\,\|x\| + k \leqslant L\|x\|^2 + k \leqslant Lk^2t_0^2 + k = kt_0\end{aligned}$$

whenever $\|x\| \leqslant kt_0$, so Ψ takes the ball $\{x\colon \|x\| \leqslant kt_0\}$ into the same ball. In this ball Ψ is a contraction, since

$$\|\Psi'(x)\| = \|I - F'(x)\| \leqslant L\|x\| \leqslant Lkt_0 = (1 - \sqrt{1-4h})/2 = q < 1/2$$

by the equalities $h = Lk$ and $t_0 = (2h)^{-1}(1 - \sqrt{1-4h})$ in our case. It remains to observe that $x_{n+1} = \Psi(x_n)$ and apply Theorem 1.4.1 with the estimate of the rate of convergence obtained in the cited theorem. $\square$

12.5(ii). Multilinear mappings

We recall that a mapping

$$\Psi\colon\ X_1\times\cdots\times X_k \to Y$$

from the product of linear spaces X_i to a linear space Y is called *k-linear* if for any fixed $k-1$ arguments Ψ is linear in the remaining argument. A mapping that is k-linear for some k is called multilinear. If $X_1 = \cdots = X_k = X$, then such a mapping generates the mapping $\psi_k(x) := \Psi(x, \ldots, x)$ from X to Y, called homogeneous of order k. It is clear that ψ_k is also generated by any k-linear mapping obtained from Ψ by an arbitrary permutation of variables. Hence different k-linear mappings can generate the same homogeneous mapping. However, if Ψ is symmetric, i.e., invariant with respect to all permutations of arguments, then it is uniquely recovered by ψ_k, which is shown below. The symmetrization of the mapping Ψ on X^k is defined by

$$\widetilde{\Psi}\colon\ x \mapsto \frac{1}{k!}\sum_{\sigma\in S_k}\Psi(x_{\sigma(1)},\ldots,x_{\sigma(k)}),$$

where summation is taken over all permutations σ of the collection $\{1,\ldots,k\}$. It is clear that the symmetrization of a k-linear mapping Ψ on X^k generates the same homogeneous of order k mapping as Ψ itself. Thus, a homogeneous mapping can be always generated by a symmetric multilinear mapping.

A mapping $\psi\colon X\to Y$ is called *polynomial* or a *polynom* if

$$\psi(x) = \psi_n(x) + \cdots + \psi_1(x) + \psi_0,$$

where ψ_0 is a constant element of Y and ψ_k is a homogeneous of order k mapping from X to Y, $1 \leqslant k \leqslant n$. In other words, there exist k-linear mappings Ψ_k such that $\psi(x) = \Psi_n(x,\ldots,x) + \cdots + \Psi_1(x) + \psi_0$. We say that ψ has a degree at most n; the exact degree of ψ is the minimal k for which ψ has a degree at most k. The homogeneous polynomials ψ_k are called homogeneous components of ψ.

The next result shows that the homogeneous components are uniquely determined by the polynomial and any homogeneous polynomial uniquely determines the symmetric multilinear mapping generating it. For an arbitrary mapping $\varphi\colon\ X \to Y$ and every $h \in X$ we define a mapping $\Delta_h\varphi\colon\ X \to Y$ by the formula

$$(\Delta_h\varphi)(x) := \varphi(x+h) - \varphi(x).$$

If $h_1, h_2 \in X$, then we set $\Delta_{h_2}\Delta_{h_1}\varphi := \Delta_{h_2}(\Delta_{h_1}\varphi)$. Since

$$\Delta_{h_2}\Delta_{h_1}\varphi(x) = \varphi(x+h_1+h_2) - \varphi(x+h_1) - \varphi(x+h_2) + \varphi(x),$$

we have $\Delta_{h_2}\Delta_{h_1}\varphi = \Delta_{h_1}\Delta_{h_2}\varphi$. By induction let

$$\Delta_{h_n}\Delta_{h_{n-1}}\cdots\Delta_{h_1}\varphi := \Delta_{h_n}(\Delta_{h_{n-1}}\cdots\Delta_{h_1}\varphi).$$

One can observe that $\Delta_{h_n}\Delta_{h_{n-1}}\cdots\Delta_{h_1}\varphi$ is the sum of 2^n functions of the form $x \mapsto (-1)^{n-p}\varphi(x + h_{i_1} + \cdots + h_{i_p})$, where $i_1 < \cdots < i_p$ are indices from $\{0,\ldots,n\}$, and for $i_p = 0$ we set $h_{i_p} = 0$. It is easy to verify by induction that $\Delta_{h_n}\Delta_{h_{n-1}}\cdots\Delta_{h_1}\varphi(x)$ is a symmetric function of the arguments $h_1,\ldots,h_n$.

12.5.2. Theorem. *Let $\psi = \psi_n + \cdots + \psi_1 + \psi_0\colon\ X \to Y$ be a polynomial of degree at most $n \geqslant 1$. Let Ψ_n be an n-linear symmetric mapping such that we have $\Psi_n(x,\ldots,x) = \psi_n(x)$. Then*

(i) *for every $h \in X$ the mapping $\Delta_h\psi$ is a polynomial of degree at most $n-1$;*

(ii) *for any fixed $h_1,\ldots,h_n \in X$, the mapping $\Delta_{h_n}\cdots\Delta_{h_1}\psi$ is constant and equals $n!\Psi_n(h_1,\ldots,h_n)$.*

The proof is delegated to Exercise 12.5.25.

12.5.3. Corollary. *The homogeneous components of any polynomial ψ are unique. If ψ is a homogeneous polynomial of degree k, then the generating symmetric k-linear mapping is uniquely determined by the relation*

$$\Psi_k(x_1,\ldots,x_k) = \frac{1}{k!}\Delta_{x_1}\cdots\Delta_{x_k}\psi(a),$$

where the right-hand side is constant in a.

We now consider polynomial mappings of normed spaces. As in the case of linear mappings, of particular importance are continuous polynomials. The next assertion shows the equivalence of several natural conditions for the continuity. Below the product $X_1 \times \cdots \times X_k$ of normed spaces is equipped with the norm $(x_1, \ldots, x_k) \mapsto \|x_1\|_{X_1} + \cdots + \|x_k\|_{X_k}$.

The proofs of the following two theorems are delegated to Exercise 12.5.26.

12.5.4. Theorem. *Let X and Y be normed spaces and let $\psi = \psi_n + \cdots + \psi_0$ be a polynomial mapping, where each homogeneous component ψ_k is generated by a symmetric k-linear mapping Ψ_k. The following conditions are equivalent*:

(i) *all Ψ_k are continuous*;

(ii) *all homogeneous components ψ_k are continuous*;

(iii) *ψ is continuous*;

(iv) *ψ is continuous at some point*;

(v) *ψ is bounded on some ball of positive radius*;

(vi) *ψ is bounded on every ball.*

12.5.5. Theorem. *Let $X_1, \ldots, X_k, Y$ be normed spaces, where $k \in \mathbb{N}$, and let $\Psi\colon\ X_1 \times \cdots \times X_k \to Y$ be a k-linear mapping. The following conditions are equivalent*:

(i) *Ψ is continuous*;

(ii) *Ψ is continuous at some point*;

(iii) *Ψ is bounded on a neighborhood of some point*;

(iv) *Ψ is bounded on every bounded set in $X_1 \times \cdots \times X_k$*;

(v) *there exists $M \geqslant 0$ such that $\|\Psi(x_1, \ldots, x_k)\| \leqslant M\|x_1\|_{X_1} \cdots \|x_k\|_{X_k}$.*

Estimate (v) suggests the definition of the norm of a continuous multilinear mapping Ψ by the formula

$$\|\Psi\| := \sup\bigl\{\|\Psi(x_1, \ldots, x_k)\|\colon\ \|x_i\|_{X_i} \leqslant 1\bigr\}.$$

The linear space $\mathcal{L}_k(X_1, \ldots, X_k, Y)$ of all continuous k-linear mappings from $X_1 \times \cdots \times X_k$ to Y is a normed space with this norm (and is Banach if so is Y).

Continuous multilinear mappings can be identified with operators in the following way. If Ψ is a bilinear continuous mapping from $X_1 \times X_2$ to Y, then we associate to it the operator

$$\Lambda_\Psi \in \mathcal{L}\bigl(X_2, \mathcal{L}(X_1, Y)\bigr), \quad \Lambda_\Psi(x_2)(x_1) := \Psi(x_1, x_2).$$

It is not hard to verify that $\|\Lambda_\Psi\| = \|\Psi\|$. Conversely, every operator Λ_Ψ of class $\mathcal{L}\bigl(X_2, \mathcal{L}(X_1, Y)\bigr)$ generates an element $\mathcal{L}_2(X_1, X_2, Y)$ by the indicated formula. The described correspondence is one-to-one and is a linear isometry. Similarly, to every element $\Psi \in \mathcal{L}_k(X_1, \ldots, X_k, Y)$ we associate the operator

$$\Lambda_\Psi \in \mathcal{L}^k := \mathcal{L}\bigl(X_k, \mathcal{L}^{k-1})\bigr), \quad \Bigl(\bigl(\Lambda_\Psi(x_k)\bigr) \cdots\Bigr)(x_1) := \Psi(x_1, \ldots, x_k),$$

where $\mathcal{L}^1 := \mathcal{L}(X_1, Y)$. This correspondence is a linear isometry, i.e., one has

$$\|\Lambda_\Psi\| = \sup_{\|x_1\| \leqslant 1, \ldots, \|x_k\| \leqslant 1} \|\Psi(x_1, \ldots, x_k)\|.$$

This connection was already used in §12.4 when we discussed k-fold differentiable mappings. Polynomial mappings of Banach spaces possess additional properties.

12.5.6. Theorem. *If the spaces $X_1, \dots, X_k$ are Banach, then the continuity of the multilinear mapping $\Psi\colon X_1 \times \cdots \times X_k \to Y$ is equivalent to its separate continuity in every argument.*

PROOF. Let $x_1 \in X_1$ and $\|x_1\|_{X_1} \leqslant 1$. We consider the multilinear mapping $\Psi_{x_1}\colon X_2 \times \cdots \times X_k \to Y$, $\Psi_{x_1}(x_2, \dots, x_k) = \Psi(x_1, x_2, \dots, x_k)$. If $k \geqslant 2$, then by induction we can assume that the continuity of every mapping Ψ_{x_1} is known. The family of these mappings is pointwise bounded: if $x_2, \dots, x_k$ are fixed, then the quantity $\sup\bigl\{\|\Psi_{x_1}(x_2, \dots, x_k)\|\colon \|x_1\| \leqslant 1\bigr\}$ is finite by the continuity and linearity in x_1. Hence it suffices to have the following analog of the Banach–Steinhaus theorem for multilinear mappings: if $X_1, \dots, X_k, Y$ are Banach spaces and a family of continuous multilinear mappings $T_\alpha\colon X_1 \times \cdots \times X_k \to Y$ is pointwise bounded, then it is uniformly bounded on the product of the unit balls. This analog follows at once from the isometric identification of multilinear mappings and operators described above (and also is easily proved by the same reasoning as in the case of operators). $\square$

For incomplete spaces this theorem can fail. For example, if the space X of all polynomials on $[0, 1]$ is equipped with the norm from $L^1[0, 1]$, then the bilinear function
$$\Psi(x, y) = \int_0^1 x(t)y(t)\, dt$$
is discontinuous, although it is continuous in every argument separately.

We have defined above polynomials with the aid of homogeneous components. It is clear that the restriction of such a polynomial to every straight line is a usual polynomial, i.e., for every $a, b \in X$ the function $\psi_{a,b}\colon t \mapsto \psi(a + tb)$ is a polynomial in t of the form $c_n t^n + \cdots + c_0$, where $c_k \in Y$. This gives one more option to define polynomials.

12.5.7. Theorem. (i) *Let $\psi\colon X \to Y$ be a mapping between linear spaces such that for every $a, b \in X$ the mapping $\psi_{a,b}$ of a real variable is a polynomial of degree at most n. Then ψ is a polynomial of degree at most n.*

(ii) *Let X and Y be Banach spaces and let $\psi\colon X \to Y$ be a continuous mapping such that for every $a, b \in X$ the mapping $\psi_{a,b}$ is a polynomial. Then the degrees of such polynomials are uniformly bounded, hence ψ is a polynomial.*

The proof is delegated to Exercise 12.5.27 (assertion (ii) is nontrivial!).

12.5.8. Corollary. *A mapping $\psi\colon X \to Y$ between linear spaces is a polynomial of degree at most n precisely when for every linear function l on Y the composition $l \circ \psi$ is a polynomial of degree at most n. If Y is a normed space, then it suffices to take continuous linear functions l.*

12.5.9. Corollary. *Let X and Y be Banach spaces and let $\psi\colon X \to Y$ be a continuous mapping such that for all elements $a, b \in X$ and every continuous linear functional l on Y the function $l \circ \psi_{a,b}$ is a polynomial. Then ψ is a polynomial.*

PROOF. It is clear from the previous theorem that the real function on the space on $X \times Y^*$ defined by $(x,l) \mapsto l\bigl(\psi(x)\bigr)$ is a polynomial, so the degrees of the polynomials $l\circ\psi$ are uniformly bounded. $\square$

It is readily verified that without the assumption of continuity of ψ the theorem can fail: it suffices to take a Hamel basis $\{v_\alpha\}$ in X, pick a countable part $\{v_{\alpha_n}\}$ and set $\psi(x) = \sum_{n=1}^\infty x_n^n$, where $x = \sum_n x_n v_{\alpha_n} + \sum_{\alpha\notin\{\alpha_n\}} x_\alpha v_\alpha$. It can also fail in the case of an incomplete space, although, as one can show, in place of completeness it suffices to require that $X\times X$ be a Baire space. On the incomplete space of all finite sequences $x = (x_1,\ldots,x_n,0,0,\ldots)$ with the norm from l^2 the function $\psi(x) = \sum_{n=1}^\infty \frac{1}{n!}x_n^n$ is continuous, is polynomial on every finite-dimensional subspace, but is not a polynomial on the whole space.

12.5(iii). Subdifferentials and monotone mappings

The geometric interpretation of the derivative of a function as a tangent line to the graph leads to a very important concept of subdifferential of a convex function. Let E be a real locally convex space, $U \subset E$ a nonempty open convex set, and let $f\colon U \to \mathbb{R}^1$ be a continuous convex function. A linear functional $l \in X^*$ is called a *subgradient* of the function f at the point $x_0 \in U$ if

$$f(x) - f(x_0) \geqslant l(x - x_0)$$

for all $x \in U$. The *subdifferential* of f at x_0 is defined to be the set $\partial f(x_0)$ of all subgradients of f at x_0.

It is clear that the set $\partial f(x_0)$ is convex. We show that, in addition, it is nonempty. Indeed, since the supergraph $G_f := \{(x,t)\colon\ t > f(x)\}$ is convex and open, the point $\bigl(x_0, f(x_0)\bigr)$ can be separated from G_f by a closed hyperplane in $E\times\mathbb{R}^1$ of the form

$$\{(x,t)\colon\ l(x) + \alpha t = c\}, \quad l \in E^*, \alpha, c \in \mathbb{R}^1.$$

Note that $\alpha \neq 0$, so we can assume that $\alpha = 1$ and that $l(x) + c < t$ for all points $(x,t) \in G_f$, i.e., $l(x_0) + t > l(x) + f(x_0)$ for all $t > f(x)$, whence we obtain $l(x_0) + f(x) \geqslant l(x) + f(x_0)$, as required. If the function f is Gateaux differentiable at x_0, then the Gateaux derivative will be the unique element of the set $\partial f(x_0)$.

Among the most important results about subdifferentials we should mention in the first place the Moro–Rockafellar theorem and the Dubovitskii–Milyutin theorem. For continuous convex functions on locally convex spaces, the former gives the equality $\partial(f+g)(x) = \partial f(x) + \partial g(x)$, and the latter gives the equality $\partial \max(f,g)(x) = \operatorname{conv}\bigl(\partial f(x) \cup \partial g(x)\bigr)$. For simplification of technical details we formulate these definitions and results not in their full generality.

The subdifferential of a convex function delivers the most important example of a multivalued monotone mapping. In order not to leave the framework of single-valued mappings, we mention a particular case of the general definition. Let X be a locally convex space. A mapping $F\colon\ X \to X^*$ is called *monotone* if

$$\langle F(x) - F(y), x - y\rangle \geqslant 0 \quad \forall\, x, y \in X.$$

For example, if the convex function f is Gateaux differentiable, then the mapping $x \mapsto f'(x)$ is monotone. Monotone mappings play an important role in applications. Mathematical models of many real world phenomena lead to differential and integral equations with monotone mappings. Among the best known results about monotone mappings we mention only the following two (see [**93**]).

12.5.10. Theorem. *If H is a Hilbert space, $F\colon H \to H$ is monotone and continuous on finite-dimensional subspaces and $\lim\limits_{\|x\|\to\infty} \|F(x)\| = \infty$, then $F(H) = H$.*

We observe that for every monotone mapping $F\colon H \to H$ that is continuous on finite-dimensional subspaces, the mapping $F + \lambda I$ with $\lambda > 0$ satisfies the conditions of the stated theorem and hence is surjective. Indeed,

$$\begin{aligned}\|F(x) + \lambda x\| &\geqslant \|F(x) - F(0) + \lambda x\| - \|F(0)\| \\ &\geqslant \langle F(x) - F(0) + \lambda x, x\|x\|^{-1}\rangle - \|F(0)\| \geqslant \lambda\|x\| - \|F(0)\|.\end{aligned}$$

The mapping $F + \lambda I$ is injective by the inequality

$$\langle F(x) - F(y) + \lambda x - \lambda y, x - y\rangle \geqslant \lambda\|x - y\|^2.$$

Therefore, we obtain everywhere defined inverse mappings

$$J_\lambda := (\lambda F + I)^{-1}, \quad \lambda > 0.$$

The mappings $F_\lambda := \lambda^{-1}(I - J_\lambda)$ are called the *Yosida approximations* of F.

12.5.11. Theorem. *Let $F\colon H \to H$ be a monotone mapping continuous on finite-dimensional subspaces. The mappings F_λ possess the following properties*:

(i) *F_λ is monotone and Lipschitz with constant λ^{-1};*

(ii) *for all $\lambda, \mu > 0$ we have $(F_\lambda)_\mu = F_{\lambda+\mu}$;*

(iii) *for all x, as $\lambda \to 0$ we have*

$$\|F_\lambda(x)\| \uparrow \|F(x)\| \quad \textit{and} \quad \|F_\lambda(x) - F(x)\| \to 0.$$

The proof of a more general theorem can be found in [**39**, §3.5.3], [**93**, Chapter 2]. Let us also note the following fact.

12.5.12. Theorem. *Let a mapping $F\colon H \to H$ of a Hilbert space H be continuous on finite-dimensional subspaces. The following conditions are equivalent*:

(i) *F is monotone,*

(ii) *F is monotone and $(F + I)(H) = H$,*

(iii) *for every number $\lambda > 0$, the mapping $I + \lambda^{-1}F$ is a bijection of H and $(I + \lambda^{-1}F)^{-1}$ is a contraction.*

12.5(iv). Approximations in Banach spaces

Let us make several remarks about approximations by differentiable mappings in infinite-dimensional spaces. Let X be a separable Banach space with the closed unit ball U. The proofs of the following facts can be found in the literature cited below.

Every uniformly continuous real function f on U is uniformly approximated by Lipschitz functions that are Hadamard differentiable. However, on the space

$C[0,1]$ even the norm is not approximated uniformly on U by Fréchet differentiable functions. On a Hilbert space, uniformly continuous functions are uniformly approximated by functions with bounded and continuous second Fréchet derivatives, however, on l^2 there is a Lipschitz function that is not approximated uniformly on U by functions with uniformly continuous second derivatives. Thus, even in the case of a Hilbert space the border between positive and negative results is passing between the continuity and uniform continuity of bounded second derivatives of approximating functions.

The situation with approximation of infinite-dimensional mappings is even more complicated. There exist uniformly continuous mappings from separable Banach spaces to l^2 which cannot be uniformly approximated by Lipschitz mappings. However, uniformly continuous mappings between Hilbert spaces possess uniform Lipschitz approximations with bounded Fréchet derivatives. The problems of constructing smooth approximations are also connected with the existence of smooth functions with bounded support on Banach spaces. Let us mention some interesting facts.

12.5.13. Theorem. (i) *On $C[0,1]$ there are no nonzero Fréchet differentiable functions with bounded support.*

(ii) *If on a Banach space X and on its dual there are nonzero functions with bounded support and locally Lipschitz derivatives, then X is a Hilbert space* (*up to a renorming*).

(iii) *The existence of nonzero functions with bounded support and Lipschitz derivatives is equivalent to the existence of an equivalent norm with a Lipschitz derivative on the unit sphere.*

(iv) *On c_0 there is a nonzero C^∞-function with bounded support (on c_0 there is an equivalent norm that is real-analytic outside the origin).*

(v) *If X possesses a nonzero C^k-function with bounded support, then X contains an isomorphic copy of either c_0 or l^k.*

For a deeper acquaintance with this direction, see the books Benyamini, Lindenstrauss [**56**], Deville, Godefroy, Zizler [**144**] and the papers Nemirovskii, Semenov [**699**], Tsar'kov [**710**], and Bogachev [**673**].

12.5(v). Covering mappings

Here we discuss a number of interesting concepts and results connected with elementary objects such as Lipschitz mappings and fixed points, but discovered very recently. These concepts and results could be discussed already in the first sections of the first chapter, but we have kept them for completing our considerations, believing that for the reader, as for ourselves, it will be pleasant to learn in the end that until the present discoveries of something simple and bright still occur in the very basic things. The main object will be the concept of an α-covering mapping introduced by A. A. Milyutin. A number of fundamental results have been obtained in the papers Levitin, Milyutin, Osmolovskii [**694**], Dmitruk, Milyutin, Osmolovskii [**685**] and effectively reinforced in the recent paper Arutyunov [**668**], which we follow in our presentation. Closed balls in metric spaces X and Y will be denoted by the symbols $B_X(x,r)$ and $B_Y(x,r)$; in the case where $X=Y$ we omit indication of the space.

12.5.14. Definition. *Let X and Y be metric spaces. A mapping $\Psi\colon X\to Y$ is called covering with constant $\alpha>0$, or α-covering, if for every ball $B_X(x,r)$ in X we have $B_Y\bigl(\Psi(x),\alpha r\bigr)\subset\Psi\bigl(B_X(x,r)\bigr)$.*

Note that any α-covering mapping is surjective, since the union of balls $B_Y\bigl(\Psi(x_0),\alpha n\bigr)$ with a fixed center x_0 is the whole space Y.

The main result is the following remarkable theorem from [**668**].

12.5.15. Theorem. *Let $\Psi\colon X\to Y$ be continuous and α-covering with some $\alpha>0$ and let $\Phi\colon X\to Y$ satisfy the Lipschitz condition with constant $\beta<\alpha$. Suppose that X is complete. Then there exists a point $\xi\in X$ such that $\Psi(\xi)=\Phi(\xi)$. Moreover, for every $x_0\in X$, there exists a point $\xi=\xi(x_0)$ such that $\Psi(\xi)=\Phi(\xi)$ and $d_X(x_0,\xi)\leqslant(\alpha-\beta)^{-1}d_Y\bigl(\Psi(x_0),\Phi(x_0)\bigr)$.*

PROOF. Replacing d_Y by d_Y/α, we can pass to the case $\beta<\alpha=1$. There is a point $x_1\in X$ with $\Psi(x_1)=\Phi(x_0)$ and

$$d_X(x_0,x_1)\leqslant d_Y\bigl(\Psi(x_0),\Phi(x_0)\bigr).$$

By induction we can construct points x_n such that

$$\Psi(x_n)=\Phi(x_{n-1}),\quad d_X(x_n,x_{n-1})\leqslant\beta d_X(x_{n-1},x_{n-2}). \tag{12.5.1}$$

Indeed, if points $x_0,\ldots,x_n$ are already found, there is a point x_{n+1} for which $\Psi(x_{n+1})=\Phi(x_{n-1})$ and

$$d_X(x_{n+1},x_n)\leqslant d_Y\bigl(\Psi(x_n),\Phi(x_n)\bigr)=d_Y\bigl(\Phi(x_{n-1}),\Phi(x_n)\bigr)\leqslant\beta d_X(x_{n-1},x_n).$$

Using (12.5.1) and the formula for the sum of a progression it is easy to obtain the estimate $d_X(x_n,x_0)\leqslant(1-\beta)^{-1}d_Y\bigl(\Psi(x_0),\Phi(x_0)\bigr)$. The completeness of X gives a limit ξ of the sequence $\{x_n\}$ and this limit satisfies the desired conditions. □

Let us show how from this theorem one can obtain some known results.

12.5.16. Example. (i) Let $X=Y$ be complete, $f(x)=x$, and let g be a contracting mapping. Then we obtain a fixed point of g, taking $\alpha=1$ and $\beta<1$ equal to the Lipschitz constant for g.

(ii) (Milyutin's theorem) Let X be a complete metric space, Y a normed space, $\Psi\colon X\to Y$ an α-covering continuous mapping, and let $\Phi\colon X\to Y$ satisfy the Lipschitz condition with constant $\beta<\alpha$. Then $\Psi+\Phi$ is $(\alpha-\beta)$-covering. Indeed, we have to show that if $x_0\in X$ and $y_0\in Y$, then there exists $\xi\in X$ with $d_X(x_0,\xi)\leqslant(\alpha-\beta)^{-1}\|y_0\|$ and $\Psi(\xi)+\Phi(\xi)=\Psi(x_0)+\Phi(x_0)+y_0$. To this end, let $\Phi_0(x):=\Psi(x_0)+\Phi(x_0)+y_0-\Phi(x)$. It is clear that Φ_0 is Lipschitz with constant β. The theorem above gives a point ξ with $\Psi(\xi)=\Phi_0(\xi)$, as required.

Exercises

12.5.17° Let $\mathcal{A}$ be a Banach algebra (for example, the algebra of all bounded operators on a Banach space). Prove that the mapping $a\mapsto e^a$, $\mathcal{A}\to\mathcal{A}$ is Fréchet differentiable and find its derivative at zero.

12.5.18.° Prove that the norm on l^1 is nowhere differentiable.
HINT: given $x \in l^1$, consider the maximal $|x_i|$.

12.5.19.° Prove that the following mapping from $L^2[0,1]$ to $C[0,1]$ is Fréchet differentiable and evaluate its derivative:

$$F(x)(t) = \int_0^t \Psi\bigl(x(s), s, t\bigr)\, ds, \quad \text{where } \Psi \in C_b^1(\mathbb{R}^3).$$

12.5.20. Suppose that in the situation of Example 12.1.7 the set K is compact, convex and balanced and that $\bigcup_{n=1}^{\infty} nK$ is dense in X. Show that f has the zero Gateaux derivative at all points of $\bigcup_{0\leqslant t<1} tK$.

12.5.21. Let μ be a nonnegative measure and $p \in (1, +\infty)$. Prove that the mapping $F\colon\ x \mapsto |x|^p$ from $L^p(\mu)$ to $L^1(\mu)$ is Fréchet differentiable and $DF(x)h = p|x|^{p-1}h$.

12.5.22.° For $z \in \mathbb{C}$ let $f(z) = \exp(-z^{-4})$ if $z \neq 0$ and $f(0) = 0$. Show that f as a function on $\mathbb{R}^2$ has all partial derivatives $\partial_x^n \partial_y^k f$ on the whole plane, but is discontinuous at the origin, hence is not differentiable.

12.5.23.° Let H be the Hilbert space l^2 of two-sided sequences $x = (x_n)$, $n \in \mathbb{Z}$, let $\{e_n\}$ be its natural basis, and let T be the linear isometry such that $Te_n = e_{n-1}$. Let us consider the second order polynomial mapping

$$f(x) = T\bigl(x + \varepsilon\bigl(1 - (x,x)\bigr)e_0\bigr), \quad \varepsilon \in (0, 1/2).$$

Show that f takes the closed unit ball U to U and is a diffeomorphism of some neighborhoods of U and a homeomorphism of U, but has no fixed points.

12.5.24. Prove that a function ψ on $\mathbb{R}^d$ is a polynomial if ψ is a polynomial in every argument.

HINT: apply the induction on d. For this assume that $d > 1$ and that for $d-1$ the assertion is true. The elements of $\mathbb{R}^d$ will be written in the form (x, y), where $x \in \mathbb{R}^{d-1}$ and $y \in \mathbb{R}^1$. For every $x \in \mathbb{R}^{d-1}$ the function $y \mapsto \psi(x,y)$ is a polynomial of some degree. By the Baire theorem there is $n \in \mathbb{N}$ and a set $M \subset \mathbb{R}^{d-1}$ dense in some ball B such that for every $x \in M$ and all y we have $\psi(x,y) = a_n(x)y^n + \cdots + a_1(x)y + a_0(x)$, where a_j is some function on $\mathbb{R}^{d-1}$. Substituting $y = 0, 1, 2, \ldots, n$ we obtain $n+1$ polynomials $\psi(x,k)$, $k = 0, 1, \ldots, n$, on $\mathbb{R}^{d-1}$. For every $x \in M$ we have $n+1$ equalities $\psi(x,k) = a_n(x)k^n + \cdots + a_1(x)k + a_0(x)$. There are numbers c_{ij}, where $i, j = 0, 1, \ldots, n$, independent of x, for which

$$a_k(x) = c_{k0}\psi(x,0) + c_{k1}\psi(x,1) + \cdots + c_{kn}\psi(x,n),$$

since the determinant of this system with respect to $a_k(x)$ (the Vandermonde determinant) is not zero. Let us consider the polynomials $p_k(x) := \sum_{j=0}^n c_{kj}\psi(x,j)$ and the polynomial $\varphi(x,y) := \sum_{k=0}^n p_k(x)y^k$. For any $x \in M$ the functions $y \mapsto \psi(x,y)$ and $y \mapsto \varphi(x,y)$ are polynomials of degree n and coincide at $n+1$ points $y = 0, 1, \ldots, n$. Hence they coincide for all y. Thus, the function ψ coincides with a polynomial on the set $M \times \mathbb{R}^1$, i.e., $\psi - \varphi = 0$ on this set. By the inductive assumption for every fixed y the function $x \mapsto \psi(x,y) - \varphi(x,y)$ is a polynomial. Since this polynomial vanishes on the set M dense in the ball, it is identically zero. Hence $\psi(x,y) = \varphi(x,y)$ for all (x,y).

12.5.25. Prove Theorem 12.5.2 by induction on n.

HINT: for $n = 1$ the assertion is verified directly. Suppose that it is true for $n - 1 \geqslant 1$. We have $\Delta_h\psi = \Delta_h\psi_n + \Delta_h(\psi_{n-1} + \cdots + \psi_0) = \Delta_h\psi_n + \varphi$, where φ is a polynomial of degree at most $n-2$ by the inductive assumption. In addition, by using the multilinearity and symmetry of Ψ_n we obtain

$$\Delta_h\psi_n = \Psi_n(x+h,\ldots,x+h) - \Psi_n(x,\ldots,x) = n\Psi_n(x,\ldots,x,h) + \eta(x),$$

where η is a polynomial of degree at most $n-2$ and the argument x enters $\Psi_n(x,\ldots,x,h)$ $n-1$ times. Thus, $\Delta_h\psi$ is the sum of the homogeneous polynomial $n\Psi_n(x,\ldots,x,h)$ of degree $n-1$ and the polynomial $\varphi+\eta$ of degree at most $n-2$, which completes the proof of assertion (i). We now make an inductive step to justify (ii). As already shown, we have $\Delta_{h_n}\psi = \varphi_{n-1}+\cdots+\varphi_0$, where φ_k are homogeneous polynomials of degree k and

$$\varphi_{n-1}(x) = g_{n-1}(x,\ldots,x), \quad g_{n-1}(x_1,\ldots,x_{n-1}) = n\Psi_n(x_1,\ldots,x_{n-1},h_n).$$

It remains to use the inductive assumption for the polynomial $\Delta_{h_n}\psi$ of degree at most $n-1$, which shows that $\Delta_{h_1}\cdots\Delta_{h_{n-1}}(\Delta_{h_n}\psi)$ is the number $(n-1)!g_{n-1}(h_1,\ldots,h_{n-1})$.

12.5.26. Prove Theorems 12.5.4 and 12.5.5.

HINT: each of the assertions (i)–(iv) implies the next one and (vi) implies (v). Show that (v) implies (i) (then also (vi)) by induction on n. For $n=1$ this is true, since $\psi_1=\Psi_1$ is a linear mapping. Suppose that the implication in question is true for some $n-1\geqslant 1$. By assumption there is a ball $B(a,r)$ with $r>0$ on which $\|\psi(x)\|$ does not exceed some M. We know that $n!\Psi_n(h_1,\ldots,h_n) = \Delta_{h_1}\cdots\Delta_{h_n}\psi(a)$ is the sum of 2^n terms of the form $(-1)^{n-p}\psi(a+h_{i_1}+\cdots+h_{i_p})$, where $i_1<\cdots<i_p$ are indices from $\{1,\ldots,n\}$ and $h_{i_p}=0$ for $p=0$. If we take vectors h_i with $\|h_i\|\leqslant r/n$, then $a+h_{i_1}+\cdots+h_{i_p}\in B(a,r)$. Hence $\|\Psi_n(h_1,\ldots,h_n)\|\leqslant 2^nM$ whenever $\|h_i\|\leqslant r/n$. This yields the estimate $\|\Psi_n(x_1,\ldots,x_n)\|\leqslant 2^nM(nR/r)^n$ whenever $\|x_i\|\leqslant R$, which gives the continuity of Ψ_n at zero. By the additivity of Ψ_n in every argument we obtain the continuity of Ψ_n everywhere. The proof of Theorem 12.5.5 is similar.

12.5.27. Prove Theorem 12.5.7 and Corollary 12.5.8.

HINT: (i) If $X=\mathbb{R}^k$, then the assertion is easily verified by induction on k with the aid of the fact that ψ is a polynomial in every variable. Hence the restriction of ψ to every finite-dimensional space is a polynomial of degree at most n. It follows that the mapping $\Psi_n\colon (x_1,\ldots,x_n)\mapsto \Delta_{x_1}\cdots\Delta_{x_n}\psi(0)$ is multilinear and the restriction of the mapping $\psi(x)-\Psi_n(x,\ldots,x)$ to every straight line is a polynomial of degree at most $n-1$. Using induction on n, we obtain our claim. (ii) Show that the degrees of the polynomials $\psi_{a,b}$ are uniformly bounded. In the case of $\mathbb{R}^d$ this is true even without the assumption about the continuity of ψ, which can be derived from Baire's theorem (Exercise 12.5.24). In the infinite-dimensional case we also use Baire's theorem, but in another way. For every $n\in\mathbb{N}$ let us consider the set

$$M_n := \bigl\{(a,b)\in X\times X\colon\ \psi_{a,b} \text{ has a degree at most } n\bigr\}.$$

By our condition, the union of all M_n is $X\times X$. By Baire's theorem some M_{n_1} is dense in some ball U of radius $r>0$ centered at (u_0,v_0). Then $U\subset M_{n_1}$. Indeed, let $(u_n,v_n)\to(u,v)$ in U and $(u_n,v_n)\in M_{n_1}$. By the continuity of ψ for all $t\in\mathbb{R}^1$ we have $\psi(u+tv)=\lim\limits_{n\to\infty}\psi(u_n+tv_n)$. It is not hard to verify that the pointwise limit of polynomials of degree at most n_1 on the real line is also a polynomial of degree at most n_1. Thus, $(u,v)\in M_{n_1}$. Passing to the function $x\mapsto\psi(x-a)$, we can assume that $a=0$. Thus, whenever $\|u\|\leqslant r$ and $\|v-v_0\|\leqslant r$, all functions $\psi_{u,v}$ are polynomials of degree at most d. It follows that all functions $\psi_{u,v}$ possess this property. Indeed, let us fix u_1,v_1. Let us consider the restriction of ψ to the three-dimensional space E generated by u_1, v_1 and v_0. We can assume that $E=\mathbb{R}^3$, $v_1=e_1$, $u_1=e_2$, $v_0=e_3$, where e_i are the vectors of the standard basis (if the vector v_0 belongs to the two-dimensional space generated by u_1 and v_1, then the problem becomes even simpler). We are given the function ψ that is a polynomial on every straight line and for all u from some ball centered at zero and all v from a ball centered at e_3 the polynomials $\psi_{u,v}$ have a degree at most n. We already know that such a function is a polynomial. Hence it is clear that its degree does not exceed n. For the proof of Corollary 12.5.8 observe that if for every linear function l the composition $l\circ\psi$ is a polynomial of degree at most n, then for every $a, b \in X$ we have $l\bigl(\psi(a+tb)\bigr) = c_n(a,b,l)t^n + \cdots + c_0(a,b,l)$, where the functions $l \mapsto c_k(a,b,l)$ are linear. Hence there exist elements $v_k(a,b)$ in the second algebraic dual to Y such that $\psi(a+tb) = v_n(a,b)t^n + \cdots + v_0(a,b)$ for all $t \in \mathbb{R}^1$. Substituting $t = 0, 1, \ldots, n$, we obtain that the elements $v_k(a,b)$ are linear combinations of the elements $\psi(a+kb) \in Y$, since the determinant of the corresponding linear system is nonzero (this is the so-called Vandermonde determinant). Thus, $v_k(a,b) \in Y$, so Theorem 12.5.7 applies. In the case of a normed space Y the same reasoning works with continuous functionals on Y.

12.5.28. Let H be a Hilbert space and $F\colon H \to H$. Prove that F is monotone precisely when $\|x-y\| \leqslant \|x-y+\lambda F(x) - \lambda F(y)\|$ for all $\lambda > 0$ and $x, y \in H$.

12.5.29. Let H be a Hilbert space and let $F\colon H \to H$ be a monotone mapping continuous on finite-dimensional subspaces. Prove that for every $v \in H$ the function $x \mapsto \bigl(F(x), v\bigr)$ is continuous. Give an example where F itself is not continuous.

12.5.30. Let X be a Hilbert space and let $\{T_t\}$ be a one-parameter group of linear operators on X with $\|T_t\| = 1$ such that for some $h \in X$ the bound $\|T_t h - h\| \leqslant C|t|$ holds whenever $t \in [-1, 1]$. Prove that the mapping $F\colon t \mapsto T_t h$ is continuously differentiable on $\mathbb{R}^1$ and Lipschitz with the constant $\|F'(0)\| \leqslant C$.

12.5.31. Let X, Y be metric spaces, let X be complete, and let $F\colon X \to Y$ be continuous and satisfy the following condition: for every $\varepsilon > 0$ there is $\delta > 0$ such that the closure of $F\bigl(U(x,\varepsilon)\bigr)$ contains $U\bigl(F(x),\delta\bigr)$ for all $x \in X$, where $U(a,r)$ is the open ball of radius r centered at a. Prove that F takes open sets to open sets.

HINT: see [**418**, p. 17, Lemma 3.9].

12.5.32. (J. Vanderwerff) Let X be an infinite-dimensional Banach space. Show that on X there is a continuous convex function that is not bounded on the closed unit ball.

HINT: take a sequence of functionals $f_n \in X^*$ with $\|f_n\| = 1$ that is weak-$*$ convergent to zero and set $f := \sum_{n=1}^{\infty} \varphi_n(f_n)$, where φ_n are even continuous convex functions on the real line with $\varphi_n(t) = 0$ for all $t \in [0, 1/2]$ and $\varphi_n(1) = n$.

12.5.33.* (Shkarin [**705**]) Prove that on an infinite-dimensional separable Hilbert space there is an infinitely Fréchet differentiable function f such that $f(x) = 0$ if $\|x\| \geqslant 1$, but $f'(x) \neq 0$ if $\|x\| < 1$. Moreover, there is a polynomial of degree 4 with this property.

HINT: consider $H = L^2[0,1]$, $f(x) = \bigl(1-(x,x)\bigr)\bigl((Ax,x)+2(\varphi,x)+4/27\bigr)$, where $Ax(t) = tx(t)$, $\varphi(t) = t(1-t)$.

12.5.34.* Prove that there exists an infinitely Fréchet differentiable function f on l^2 that is bounded on the closed unit ball, but does not attain its maximum on it.

HINT: use the previous exercise.