Notes and exercises from
Foundations of Modern Analysis

John Peloquin

Introduction

This document contains notes and exercises from [1].

Chapter IV

Section 3

Remark. In the proof of (4.3.1), intuitively we know that $f = \log_a$. Recall by the density of the rationals in the reals (2.2.16) that *a real number is the supremum of the set of rationals less than it.* Therefore $\log_a x$ is just the supremum of the set of rationals m/n with $m/n \le \log_a x$, or equivalently $a^{m/n} \le x$, or equivalently $a^m \le x^n$; that is, $\log_a x = \sup A_x$. Since neither $\log_a x$ nor $a^{m/n}$ are defined yet, but a^m and x^n are, it is natural to show that $f(x) = \sup A_x$ for uniqueness.

For existence, we know the set A_x is well defined. Fix $k \ge 1$ and choose p, q by (4.3.1.1) with $a^p \le x^k \le a^q$. Then $p/k \in A_x$, so A_x is nonempty. If $m/n \in A_x$ ($n \ge 1$), then $a^m \le x^n$, so $a^{mk} \le x^{nk} \le a^{qn}$, so $mk \le qn$, and $m/n \le q/k$; that is, A_x is bounded above by q/k. Therefore $f(x) = \sup A_x$ is defined, and $p/k \le f(x) \le q/k$. Taking $q = p + 1$ yields $p/k \le f(x) \le (p+1)/k$, which is used in the proof that $f(xy) = f(x) + f(y)$. Taking $x = a$ and $k = p = q = 1$ yields $f(a) = 1$.

We provide an alternative proof of (4.3.7):

(4.3.7). *Any continuous mapping g of $\mathbf{R}_+^*$ into itself such that $g(xy) = g(x)g(y)$ has the form $x \mapsto x^a$, with $a \in \mathbf{R}$.*

Proof. If $b > 1$, then $f = \log_b \circ g : \mathbf{R}_+^* \to \mathbf{R}$ is continuous and

$$f(xy) = \log_b g(xy) = \log_b[g(x)g(y)] = \log_b g(x) + \log_b g(y) = f(x) + f(y)$$

If $g \neq 1$, there is $x > 0$ with $g(x) > 1$ (if $g(x) < 1$, then $g(x^{-1}) = g(x)^{-1} > 1$), so $y_n = g(x^n) = g(x)^n \to \infty$ as $n \to \infty$ and $y_n \to 0$ as $n \to -\infty$. It follows from (3.19.1) and (3.19.7) that g is surjective, so there is $c > 0$ with $c \neq 1$ such that $g(c) = b$ and hence $f(c) = 1$. By (4.3.2), $f = \log_c$, so

$$g(x) = b^{\log_c x} = b^{a \log_b x} = x^a$$

where $a = \log_c b$, by (4.3.3) and (4.3.4). $\qquad\square$

Remark. This proof is longer than Dieudonné's, but avoids direct use of (4.1.3), which is already used in the proof of (4.3.2).

Section 5

Remark. In the proof of the Tietze-Urysohn extension theorem (4.5.1), the idea behind the formula defining $g(x)$ for $x \in E - A$ is roughly: *for x very near the boundary (frontier) of A, take the limit of $f(y)$ as $y \in A$ approaches points near x:*

$$g(x) = \inf_{y \in A} \left[f(y) \frac{d(x,y)}{d(x,A)} \right] \qquad (x \in E - A)$$

Write $\rho_x(y) = d(x,y)/d(x,A)$. Since $1 \leq f(y) \leq 2$ by the reduction, $\rho_x(y) \geq 1$, $\rho_x(y) > 2$ for y sufficiently far from x, and $\rho_x(y) \to 1$ as $y \to x$, it follows that $\inf_{y \in A} f(y) \rho_x(y)$ reflects values of $f(y)$ at y nearest to x. This ensures continuity of g on the boundary of A. Since g is also continuous in the interior of A (by continuity of f), and in the exterior of A (by continuity of the metric d), g is continuous on E.

Chapter V

Section 2

Remark. (5.2.4) is an associativity property for convergent series in a normed space. For *absolutely* convergent series in a *Banach* space, (5.3.6) is a stronger associativity property. For example, if $\sum_{n=0}^{\infty} x_n$ is absolutely convergent in $\mathbf{C}$, it follows from the remark below (5.3.6), but not from (5.2.4), that $\sum_{n=0}^{\infty} x_n = \sum_{n=0}^{\infty} x_{2n} + \sum_{n=0}^{\infty} x_{2n+1}$.

Section 3

Remark. In the proof of (5.3.2), to see how the inequality $\|\sum_{n=0}^{\infty} x_n\| \leq \sum_{n=0}^{\infty}\|x_n\|$ follows from the principle of extension of inequalities (3.15.4), recall from the definition of sequential limit in (3.13) that the functions given by

$$s(n) = \sum_{k=0}^{n} x_k \quad s(\infty) = \sum_{k=0}^{\infty} x_k \qquad \text{and} \qquad t(n) = \sum_{k=0}^{n} \|x_k\| \quad t(\infty) = \sum_{k=0}^{\infty} \|x_k\|$$

are *continuous* on $\mathbf{N} \cup \{\infty\} = \overline{\mathbf{N}} \subseteq \overline{\mathbf{R}}$. By continuity of the norm, the composite function $n \mapsto \|s(n)\|$ is also continuous on $\overline{\mathbf{N}}$. Now $\|s(n)\| \leq t(n)$ for all $n \in \mathbf{N}$ and $\mathbf{N}$ is dense in $\overline{\mathbf{N}}$, hence $\|s(\infty)\| \leq t(\infty)$ by (3.15.4), as desired. Other applications of the principle of extension are similar.

Remark. In (5.3.4), $\sum_{\alpha \in A}\|x_\alpha\| = \sup\{\sum_{\alpha \in J}\|x_\alpha\| \mid J \subseteq A \text{ finite}\}$ by (5.3.1). This is implicit in the proof of (5.3.5).

Remark. In (5.3.4), 2ϵ can be replaced by ϵ, since H can be chosen for $\epsilon/2$.

Remark. In the proof of (5.3.4), to construct H, first fix a bijection $\varphi : \mathbf{N} \to A$. Then $\sum_{n=0}^{\infty}\|x_{\varphi(n)}\|$ converges, so there exists M such that for all $n \geq m \geq M$, $\sum_{k=m}^{n}\|x_{\varphi(k)}\| \leq \epsilon$ (5.2.1). Let $H = \varphi[0, M)$. Then if $K \subseteq A$ is finite with $H \cap K = \emptyset$, $\varphi^{-1}[K] \subseteq [M, \infty)$, so if $N = \max\varphi^{-1}[K]$, then $\sum_{\alpha \in K}\|x_\alpha\| \leq \sum_{k=M}^{N}\|x_k\| \leq \epsilon$. Also $\|\sum_{\alpha \in A} x_\alpha - \sum_{\alpha \in H} x_\alpha\| = \|\sum_{n=M}^{\infty} x_{\varphi(n)}\| \leq \sum_{n=M}^{\infty}\|x_{\varphi(n)}\| \leq \epsilon$ by (5.3.1) and (5.3.2). Finally, if L is finite with $H \subseteq L \subseteq A$, then $L - H$ is finite with $(L - H) \cap H = \emptyset$, so $\|\sum_{\alpha \in H} x_\alpha - \sum_{\alpha \in L} x_\alpha\| = \|\sum_{\alpha \in L - H} x_\alpha\| \leq \sum_{\alpha \in L - H}\|x_\alpha\| \leq \epsilon$, and it follows that $\|\sum_{\alpha \in A} x_\alpha - \sum_{\alpha \in L} x_\alpha\| \leq 2\epsilon$.

Remark. In the proof of (5.3.6), the idea is to use finite approximations of finitely many of the z_n to obtain a finite approximation of $\sum_{\alpha \in A} x_\alpha$.

Section 5

Remark. In a Banach space, $\lambda \sum_{n=0}^{\infty} x_n = \sum_{n=0}^{\infty} \lambda x_n$ for any scalar λ, when either series converges, by (5.1.5) and (5.5.2).

Remark. Let $L_1(I)$ be the space of Lebesgue integrable real-valued functions[1] on the interval $I \subseteq \mathbf{R}$ with norm $\|f\|_1 = \int_I |f|$. Then $L_1(I)$ is a Banach space (Riesz-Fischer). The integral operator $f \mapsto \int_I f$ is linear and continuous since $|\int_I f| \leq \int_I |f|$ (5.5.1). Therefore if $f = \sum_{n=0}^{\infty} f_n$ under convergence in the norm, then $\int_I f = \sum_{n=0}^{\infty} \int_I f_n$ (5.5.2).

---

[1]Technically this space consists of equivalence classes of functions.

Remark. The multiplication operation $(x, y) \mapsto xy$ from $\mathbf{C}^2$ to $\mathbf{C}$ is bilinear and continuous by (4.4.1), hence if $\sum_{n=0}^{\infty} x_n$ and $\sum_{n=0}^{\infty} y_n$ are absolutely convergent series of complex numbers, then $(\sum_{n=0}^{\infty} x_n)(\sum_{n=0}^{\infty} y_n) = \sum_{m,n} x_m y_n$, where the latter sum may be taken *in any order*, by (5.5.3).

Remark. In the proof of (5.5.4), to see how the linearity of $\overline{f}$ follows from the principle of extension of identities (3.15.2), observe that the mappings

$$(x, y) \mapsto x + y \mapsto \overline{f}(x + y) \qquad \text{and} \qquad (x, y) \mapsto (\overline{f}(x), \overline{f}(y)) \mapsto \overline{f}(x) + \overline{f}(y)$$

are continuous by (3.11.5), (3.20.15), and (5.1.5), and agree on G^2, which is dense in E^2 by (3.20.3), and hence agree on E^2 by (3.15.2). Other applications of the principle of extension are similar.

Section 8

Remark. Intuitively, a *closed* hyperplane separates a vector space into the two half-spaces on either side of the hyperplane, but a *dense* hyperplane does not have this property.

Chapter VI

Section 3

Remark. (6.3.1.1) is the *parallelogram law*, which states that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lengths of its sides.

Remark. In the proof of (6.3.1),

$$\begin{aligned} \|x - (y + \lambda z)\|^2 &= (x - y - \lambda z | x - y - \lambda z) \\ &= \|x - y\|^2 - \lambda[(x - y|z) + (z|x - y)] + \lambda^2 \|z\|^2 \\ &= \alpha^2 - 2\lambda \mathscr{R}(x - y|z) + \lambda^2 \|z\|^2 > \alpha^2 \end{aligned}$$

so

$$-2\lambda \mathscr{R}(x - y|z) + \lambda^2 \|z\|^2 > 0 \tag{1}$$

Substituting $-\lambda$ for λ in (1) yields

$$2\lambda \mathscr{R}(x - y|z) + \lambda^2 \|z\|^2 > 0 \tag{2}$$

Now (1) and (2) imply

$$|\mathscr{R}(x - y|z)| < \lambda \frac{\|z\|^2}{2} \qquad (\lambda > 0)$$

Letting $\lambda \downarrow 0$ shows $\mathscr{R}(x - y|z) = 0$.

Section 5

Remark. If E is the Hilbert sum of a normal system (a_n) and $x = (x_n) \in E$, then $x_n = \lambda_n a_n$ with

$$(x|j_n(a_n)) = \sum_{m=1}^{\infty} (\lambda_m a_m | \delta_{mn} a_m) = \lambda_n \|a_n\|^2 = \lambda_n$$

It follows that $x = \sum_{n=1}^{\infty} (x|j_n(a_n)) j_n(a_n)$ and this representation in $(j_n(a_n))$ is unique. If $y \in E$,

$$(x|y) = \sum_{n=1}^{\infty} (x|j_n(a_n)) \overline{(y|j_n(a_n))}$$

In particular $\|x\|^2 = \sum_{n=1}^{\infty} |(x|j_n(a_n))|^2$.

Remark. In the proof of (6.5.2), to see how the result follows from (6.4.2), let φ be the isomorphism of F onto the Hilbert sum E of the a_n with $\varphi(a_n) = j_n(a_n)$. If $x \in F$, by the previous remark and properties of φ,

$$\varphi(x) = \sum_{n=1}^{\infty} (\varphi(x)|\varphi(a_n))\varphi(a_n) = \varphi\left[\sum_{n=1}^{\infty} (x|a_n)a_n\right]$$

so $x = \sum_{n=1}^{\infty} (x|a_n)a_n$. If $y \in F$, then

$$(x|y) = (\varphi(x)|\varphi(y)) = \sum_{n=1}^{\infty} (\varphi(x)|\varphi(a_n))\overline{(\varphi(y)|\varphi(a_n))} = \sum_{n=1}^{\infty} (x|a_n)\overline{(y|a_n)}$$

In particular $\|x\|^2 = \sum_{n=1}^{\infty} |(x|a_n)|^2$. If (λ_n) is a sequence of scalars with $\sum_{n=1}^{\infty} |\lambda_n|^2$ convergent, then $z = (\lambda_n a_n) \in E$ and $x = \varphi^{-1}(z)$ is unique such that $(x|a_n) = (z|\varphi(a_n)) = \lambda_n$.

Chapter VII

Section 1

Remark. In (7.1.2.1), the mappings f_i are the *scalar component functions* of f, and the mappings $t \mapsto f_i(t)a_i$ are the *vector component functions* of f.

(7.1.2) shows that if F is finite-dimensional, then $\mathscr{B}_F(A)$ decomposes into subspaces of vector component functions, each of which is isomorphic to a space of scalar component functions. By (5.9.1), this result can be expressed in the equivalent form

$$\mathscr{B}_{K^n}(A) \cong \mathscr{B}_K(A)^n$$

where $K = \mathbf{R}$ or $K = \mathbf{C}$.

Section 2

Remark. If E is a metric space and $K = \mathbf{R}$ or $K = \mathbf{C}$, then $\mathscr{C}_{K^n}^{\infty}(E) \cong \mathscr{C}_K^{\infty}(E)^n$.

Remark. If E is a metric space and F is a Banach space, then $\mathscr{C}_F^{\infty}(E)$ is a Banach space, by (7.2.1), (7.1.3), and (3.14.5).

Remark. The proof of Dini's theorem (7.2.2) is like the proof of (7.2.1) "turned on its side."

Section 5

Remark. (7.5.6) is like Dini's theorem (7.2.2), but with equicontinuity replacing monotonicity and continuity.

Section 6

Remark. If $f : I \to F$ is regulated and $g : F \to G$ is continuous, then $g \circ f : I \to G$ is regulated, by (3.13.6). In particular, $\xi \mapsto \|f(\xi)\|$ is regulated.

Remark. If $f : I \to \mathbf{R}$ is monotone and $g : J \to F$ is regulated with $f(I) \subseteq J$, then $g \circ f : I \to F$ is regulated, by (4.2.1) and (3.13.6).

Remark. If $f_i : I \to F_i$ is regulated for all $1 \leq i \leq n$, then $f = (f_i) : I \to \prod F_i$ is regulated, by (3.20.5). It follows that $\sum f_i$ is regulated, by (5.1.5) and the first remark above.

Remark. The uniform limit of a sequence of regulated functions on a compact interval is regulated, by (7.6.1).

6

Chapter VIII

Section 1

Remark. It is possible to extend the definition of differentiability, and most of the results in this chapter, to functions between arbitrary *normed* spaces.

Remark. It is possible instead to have continuity of derivatives by definition, and continuity of differentiable functions by theorem.

Remark. In the proof of (8.1.4), note that $\|s\| \leq \epsilon/c$ and $\|t\| \leq \|(s,t)\|$, so $\|[s \cdot t]\| \leq c(\epsilon/c)\|(s,t)\| = \epsilon\|(s,t)\|$.

Remark. In (8.1.5), the identification $\mathscr{L}(E; \prod F_i) \cong \prod \mathscr{L}(E; F_i)$ is like (7.1.2).

Section 2

Remark. (8.2.3) is a weak form of the inverse function theorem (10.2.5). Under appropriate conditions, it establishes *differentiability* of an inverse function, *assuming existence of the inverse function*, while the inverse function theorem establishes local *existence and differentiability* of an inverse function.

In the proof, the change in g at y_0 is inverse to the change in f at x_0. Since the latter is well approximated by $f'(x_0)$ in a sufficiently small neighborhood, it is natural that the former is well approximated by $f'(x_0)^{-1}$. More specifically, if $z = g(y_0 + t) - g(y_0)$, then $t = f(x_0 + z) - f(x_0) \approx f'(x_0) \cdot z$, so $f'(x_0)^{-1} \cdot t \approx z$.

Section 4

Remark. If E and F are one-dimensional (identified to **R** or **C**) and $f : E \to F$ is differentiable, then

$$(fg)' = f'g + fg'$$

by (8.1.4), (8.1.5), and (8.2.1). This is just the *product rule*.

Remark. The notation $f'_d(\alpha)$ (resp. $f'_g(\beta)$) for derivative on the right (resp. left) at the point α (resp. β) is inspired by the French word "droite" (resp. "gauche") for right (resp. left).

Section 5

Remark. In the proof of the mean value theorem (8.5.1), the idea is to prove arbitrarily close approximations to the desired inequality by "induction" on subintervals $[\alpha, \gamma]$ of $[\alpha, \beta]$.

Remark. Results about differentiable functions defined on intervals in $\mathbf{R}$ can be extended to functions defined on (neighborhoods of) segments in Banach spaces using the chain rule, as seen in the proof of (8.5.4) and also the proof of Taylor's theorem (8.14.3).

Section 6

Remark. If a function f is differentiable at a point x_0, then change in f *at x_0* is well approximated by $f'(x_0)$. If f is *continuously* differentiable at x_0, then change in f *between any two points sufficiently close to x_0* is well approximated by $f'(x_0)$, by (8.6.2).[2]

Section 7

Remark. By (7.6.1), a regulated function $f : I \to \mathbf{R}$, where $I \subseteq \mathbf{R}$ is a compact interval, is uniformly approximated by real-valued step functions, and hence intuitively has a well defined "area under its curve", namely the limit of the areas under its approximating step functions.

On the other hand, in the proof of (8.7.2), we see that a primitive of a step function gives the area under its curve, and since the primitive of f is just the limit of the primitives of its approximating step functions (8.6.4), the primitive of f gives the area under f. This motivates the definition of the integral in terms of primitives (the Cauchy integral).

Remark. The Cauchy integral makes one half of the fundamental theorem of calculus true by definition; the other half is (8.7.3).

Remark. In the proof of (8.7.3), to see how the inequality follows from (8.5.2), let $h(x) = g(x) - f(\xi)x$. Since g is a primitive of f, there is a denumerable set $D \subseteq I$ such that for all $x \in I - D$, $h'(x) = f(x) - f(\xi)$. It follows from (8.5.2) that for any $0 \le \zeta \le \lambda$,

$$\|h(\xi + \zeta) - h(\xi)\| \le \zeta \sup_{0 \le \eta \le \lambda} \|f(\xi + \eta) - f(\xi)\|$$

---

[2]See also Problem 3.

as desired. This is like the proof of (8.6.2) from (8.5.4).

Remark. In the proof of (8.7.4), to see that $\xi \mapsto f(\varphi(\xi))\varphi'(\xi)$ is regulated, first note that if f is continuous, then $f \circ \varphi$ is continuous by (3.11.5), and hence regulated, by (7.6.2); on the other hand if φ is monotone, then $f \circ \varphi$ is regulated, by a remark from (7.6) above. Since φ' is also regulated,

$$\xi \mapsto (f(\varphi(\xi)), \varphi'(\xi)) \mapsto f(\varphi(\xi))\varphi'(\xi)$$

is regulated, by (5.1.5) and remarks from (7.6) above.

Remark. In the proof of (8.7.5), note for example

$$\xi \mapsto (f(\xi), g'(\xi)) \mapsto [f(\xi) \cdot g'(\xi)]$$

is regulated, by (7.6.2) and remarks from (7.6) above. Also note use of additivity of the integral in the proof, which follows from (8.2.2).

Remark. In the proof of (8.7.6), $u \circ f$ is regulated, by a remark from (7.6) above.

Remark. In the proof of the mean value theorem for integrals (8.7.7), $\xi \mapsto \|f(\xi)\|$ is regulated, by a remark from (7.6) above. If g is a primitive of f and h is a primitive of $\xi \mapsto \|f(\xi)\|$ (8.7.2), then for all ξ not in some denumerable set D, $\|g'(\xi)\| \le h'(\xi)$, hence

$$\|g(\beta) - g(\alpha)\| \le h(\beta) - h(\alpha) \le (\beta - \alpha) \sup_{\xi \in I - D} h'(\xi)$$

by (8.5.1) and (8.5.2), as desired.

Remark. If $f : [\alpha, \beta] \to \mathbf{R}$ is regulated and $m \le f \le M$, then

$$m(\beta - \alpha) \le \int_\alpha^\beta f \le M(\beta - \alpha)$$

by the mean value theorem (8.5.3).

Section 8

Remark. The function x^{-1} is regulated on $[1, a]$ by (4.1.4) and (7.6.2), and

$$\int_1^a x^{-1}\,dx = \log a - \log 1 = \log a$$

since $\log x$ is a primitive of x^{-1}. This integral can serve as a definition of $\log a$.

Section 9

Remark. In the remarks above (8.9.1), the partial mappings $x_1 \mapsto f(x_1, a_2)$ and $x_2 \mapsto f(a_1, x_2)$ are continuous on open subsets of E_1 and E_2, respectively, by (3.20.12) and (3.20.14), so it makes sense to talk about their differentiability.

Remark. In the proof of (8.9.1), we see that *differentiability* alone implies partial differentiability (but the converse is not true!) and (8.9.1.1) holds as long as f is differentiable at (x_1, x_2); it implies

$$Df(x_1, x_2) = D_1 f(x_1, x_2) \circ p_1 + D_2 f(x_1, x_2) \circ p_2$$

which implies[3]

$$Df = P_1 \circ D_1 f + P_2 \circ D_2 f$$

where $P_k(\varphi) = \varphi \circ p_k$. Now $P_k \in \mathscr{L}(\mathscr{L}(E_k, F), \mathscr{L}(E_1 \times E_2, F))$ by (5.7.5) and (5.5.1), so Df is continuous if $D_1 f$ and $D_2 f$ are. Conversely, $D_k f = I_k \circ Df$, where $I_k(\psi) = \psi \circ i_k$, so $D_k f$ is continuous if Df is.

In the proof that continuous partial differentiability implies (continuous) differentiability, it is not sufficient to appeal to the existence of $D_2 f(a_1 + t_1, a_2)$ to establish the third displayed inequality, because then t_2 would depend on t_1; it is necessary to appeal to the continuity of $D_2 f$ to ensure the inequality holds for all sufficiently small t_1 and t_2.

Remark. In (8.9.2), let $g = (g_i)$ and $h = f \circ g$, so g is differentiable by (8.1.5) with $Dg(x) = (Dg_i(x))$ and h is differentiable by the chain rule (8.2.1) with $Dh(x) = Df(g(x)) \circ Dg(x)$. It follows from (8.9.1) that[4]

$$Dh(x) = \sum_{i=1}^{n} D_i f(g(x)) \circ Dg_i(x)$$

This holds as long as f and the g_i are *differentiable* (by the previous remark), but if in addition Df and the Dg_i are continuous, then Dh is continuous (by reasoning as in the previous remark).

Section 10

Remark. In (A), if F is $\mathbf{R}$ or $\mathbf{C}$, then $D_k f(\alpha_1, \ldots, \alpha_n)$ is identified to a *scalar.* For $\nabla f(\alpha_1, \ldots, \alpha_n) = (D_k f(\alpha_1, \ldots, \alpha_n)) \in E$,

$$f'(\alpha_1, \ldots, \alpha_n)(x) = \nabla f(\alpha_1, \ldots, \alpha_n) \cdot \overline{x}$$

---

[3] The equation in the book is incorrect.
[4] The equation in the book is incorrect.

where the product on the right is the dot product in E. The vector $\nabla f(\alpha_1, \ldots, \alpha_n)$ is the *gradient* of f at $(\alpha_1, \ldots, \alpha_n)$ and has important geometric properties.

Remark. In (8.10.1), let K be **R** or **C**, $\varphi = (\varphi_i) : K^n \to K^m$, $\psi = (\psi_i) : K^m \to K^p$, and $\theta = (\theta_i) = \psi \circ \varphi : K^n \to K^p$. Then by the chain rule (8.2.1), θ is differentiable and $D\theta(x) = D\psi(\varphi(x)) \circ D\varphi(x)$. It follows from the remarks above (8.10.1) that[5]

$$(D_k \theta_i(x)) = (D_j \psi_i(\varphi(x)))(D_k \varphi_j(x))$$

and similarly for the determinants when $m = n = p$. This holds as long as the functions φ_i and ψ_i are *differentiable*, by a remark from (8.9) above.

Section 12

Remark. In (8.12.2), note $D^2 f(x_0) \cdot (s, t)$ and $D^2 f(x_0) \cdot (t, s)$ both approximate the change in the change in f across opposing sides of the parallelogram in A with vertices x_0, $x_0 + s$, $x_0 + t$, and $x_0 + s + t$:

$$
\begin{aligned}
D^2 f(x_0) \cdot (s, t) = (D^2 f(x_0) \cdot s) \cdot t &\approx Df(x_0 + s) \cdot t - Df(x_0) \cdot t \\
&\approx [f(x_0 + s + t) - f(x_0 + s)] - [f(x_0 + t) - f(x_0)] \\
&= [f(x_0 + t + s) - f(x_0 + t)] - [f(x_0 + s) - f(x_0)] \\
&\approx Df(x_0 + t) \cdot s - Df(x_0) \cdot s \\
&\approx (D^2 f(x_0) \cdot t) \cdot s = D^2 f(x_0) \cdot (t, s)
\end{aligned}
$$

It is therefore plausible that $D^2 f(x_0) \cdot (s, t) = D^2 f(x_0) \cdot (t, s)$, in other words that $D^2 f(x_0)$ is symmetric. The proof uses the mean value theorem (8.6.2) to bound the distances between approximately equal quantities above, and ultimately to establish symmetry.

In the proof, let $\beta = 2\epsilon \|s\| (\|s\| + \|t\|)$ and observe

$$
\begin{aligned}
\|g(1) - g(0) - (f''(x_0) \cdot t) \cdot s\| &\leq \|g(1) - g(0) - g'(0)\| + \|g'(0) - (f''(x_0) \cdot t) \cdot s\| \\
&\leq \sup_{0 \leq \xi \leq 1} \|g'(\xi) - g'(0)\| + \beta \\
&\leq \sup_{0 \leq \xi \leq 1} \|g'(\xi) - (f''(x_0) \cdot t) \cdot s\| \\
&\quad + \|g'(0) - (f''(x_0) \cdot t) \cdot s\| + \beta \\
&\leq 3\beta
\end{aligned}
$$

---

[5] The equation in the book is incorrect.

Remark. If $E = E_1 \times \cdots \times E_n$ and F are Banach spaces and $f : E \to F$ is twice differentiable at $x \in E$, then each $D_i f$ is differentiable at x and

$$D^2 f(x) \cdot (s, t) = \sum_{i,j} D_i D_j f(x) \cdot (s_i, t_j)$$

where $s = (s_i)$, $t = (t_i)$, and the sum is taken over all pairs (i, j) with $1 \le i, j \le n$.

Proof. By a remark from (8.9) above, $D_i f = I_i \circ Df$, so $D_i f$ is differentiable at x by the chain rule (8.2.1). Also $Df = \sum P_i \circ D_i f$, so $D^2 f(x) = \sum P_i \circ DD_i f(x)$ by additivity (8.2.2) and the chain rule, and

$$\begin{aligned}
D^2 f(x) \cdot (s, t) = (D^2 f(x) \cdot s) \cdot t &= \sum_{j=1}^{n} (DD_j f(x) \cdot s) \cdot t_j \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} D_i D_j f(x) \cdot s_i \right) \cdot t_j \\
&= \sum_{i,j} D_i D_j f(x) \cdot (s_i, t_j) \qquad \Box
\end{aligned}$$

Remark. In the previous remark, if E is $\mathbf{R}^n$ (resp. $\mathbf{C}^n$), then each $D_i D_j f(x)$ can be identified to a vector in F (8.10) and

$$D^2 f(x) \cdot (s, t) = \sum_{i,j} D_i D_j f(x) \xi_i \eta_j$$

where $s = (\xi_i)$ and $t = (\eta_i)$. Since $D^2 f(x)$ is a bilinear mapping, it is uniquely determined by its values $D^2 f(x) \cdot (e_i, e_j) = D_i D_j f(x)$ on pairs of standard basis vectors, and since $D^2 f(x)$ is symmetric (8.12.1), it follows that (8.12.2)

$$D_i D_j f(x) = D^2 f(x) \cdot (e_i, e_j) = D^2 f(x) \cdot (e_j, e_i) = D_j D_i f(x)$$

If F is $\mathbf{R}$ (resp. $\mathbf{C}$), these values can be identified to scalars in the *hessian matrix* $Hf(x) = (D_i D_j f(x))$, which is symmetric. If s and t are identified to column vectors, we have $D^2 f(x) \cdot (s, t) = s^T Hf(x) t$.

Remark. In the proof of (8.12.4), for $g(x) = D^{p-2} f(x) \cdot (t_3, \dots, t_p)$,

$$Dg(x) \cdot t_2 = D^{p-1} f(x) \cdot (t_2, \dots, t_p)$$

by the remark above (8.12.4), and hence

$$D^2 g(x) \cdot (t_1, t_2) = D^p f(x) \cdot (t_1, \dots, t_p)$$

by (8.12.1) and the remark above (8.12.4) again.

12

Remark. In the proof of (8.12.5), if $D^m f$ is $(n+1)$-times differentiable, then $D^m f$ is n-times differentiable and $D^n(D^m f)$ is differentiable, by definition. By induction, f is $(m+n)$-times differentiable and $D^{m+n} f = D^n(D^m f)$. Therefore f is $(m+n+1)$-times differentiable and

$$D^{m+n+1} f = D(D^{m+n} f) = D(D^n(D^m f)) = D^{n+1}(D^m f)$$

Remark. The converse of (8.12.5) holds, by induction on n. The result yields an exponentiation law for the differential operator:

$$D^m D^n = D^{m+n} = D^n D^m$$

Remark. If $E = E_1 \times \cdots \times E_n$ and F are Banach spaces and $f : E \to F$ is p-times differentiable at $x \in E$, then

$$D^p f(x) \cdot (t_1, \ldots, t_p) = \sum_{j_1, \ldots, j_p} D_{j_1} \cdots D_{j_p} f(x) \cdot (t_{1,j_1}, \ldots, t_{p,j_p}) \qquad (1)$$

where $t_i = (t_{i,j})$ for $1 \le i \le p$ and $1 \le j \le n$, and the sum is taken over all n^p p-tuples $(j_1, \ldots, j_p)$ with $1 \le j_i \le n$ for $1 \le i \le p$.

Proof. By induction on p. For $p = 1$, the result holds by (8.9). If the result holds for p and f is $(p+1)$-times differentiable at x, then each partial $D_{j_1} \cdots D_{j_p} f$ is differentiable at x since

$$D_{j_1} \cdots D_{j_p} f = I_{j_1, \ldots, j_p} \circ D^p f$$

where $I_{j_1, \ldots, j_p}(\psi) = \psi \circ (i_{j_1} \times \cdots \times i_{j_p})$. Here $i_{j_k} : E_{j_k} \to E$ is just the canonical injection, and $i_{j_1} \times \cdots \times i_{j_p} : E_{j_1} \times \cdots \times E_{j_p} \to E^p$ maps (x_k) to $(i_{j_k}(x_k))$. $I_{j_1, \ldots, j_p}$ is a continuous linear mapping from $\mathscr{L}_p(E, F)$ to the space of continuous p-linear mappings from $E_{j_1} \times \cdots \times E_{j_p}$ to F. Differentiating both sides of (1) with respect to x (keeping the t_i fixed) yields the result for $p+1$. $\qquad \square$

Remark. If E is $\mathbf{R}^n$ or $\mathbf{C}^n$, the previous remark yields (8.12.7). In this case each $D_{j_1} \cdots D_{j_p} f(x)$ can be identified to a vector in F and it follows from (8.12.4) that

$$D_{j_1} \cdots D_{j_p} f(x) = D^p f(x) \cdot (e_{j_1}, \ldots, e_{j_p})$$
$$= D^p f(x) \cdot (e_{j_{\sigma(1)}}, \ldots, e_{j_{\sigma(p)}}) = D_{j_{\sigma(1)}} \cdots D_{j_{\sigma(p)}} f(x)$$

for any permutation σ of $1, \ldots, p$.

Remark. In the remark above,

$$D^p f = \sum P_{j_1,\dots,j_p} \circ D_{j_1} \cdots D_{j_p} f$$

where $P_{j_1,\dots,j_p}(\varphi) = \varphi \circ (p_{j_1} \times \cdots \times p_{j_p})$. It follows from this and the converse relation above, with (8.9.1), that $D^p f$ exists and is continuous if and only if each $D_{j_1} \cdots D_{j_p} f$ exists and is continuous. If E is $\mathbf{R}^n$ or $\mathbf{C}^n$, this yields (8.12.8).

Remark. In the proof of (8.12.10), the key observation is that

$$h' : x \mapsto ((g' \circ f)(x), f'(x)) \mapsto (g' \circ f)(x) \circ f'(x)$$

Section 13

Remark. In the proof of (8.13.1), we only have the *real* exponential function from (8.8), although the same proof does work with the complex exponential function by (9.5.3). Note $D_i f(x) = \lambda_i f(x)$, hence $D^\alpha f(x) = M_\alpha(\lambda_1, \dots, \lambda_n) f(x)$.

Remark. In (8.13.2), taking $P(X) = X^p$ and $A = E = F = G = \mathbf{R}$ (resp. $= \mathbf{C}$) yields

$$(fg)^{(p)} = \sum_{k=0}^{p} \binom{p}{k} f^{(p-k)} g^{(k)}$$

by the binomial theorem.

Section 14

Remark. In Taylor's theorem (8.14.2), let

$$P(\alpha) = \sum_{k=0}^{p-1} \frac{(\xi - \alpha)^k}{k!} f^{(k)}(\alpha) \qquad \text{and} \qquad R(\alpha) = \int_\alpha^\xi \frac{(\xi - \zeta)^{p-1}}{(p-1)!} f^{(p)}(\zeta) \, d\zeta$$

Then $P(\alpha)$ is the $(p-1)$-th Taylor polynomial for f about α evaluated at ξ, and $R(\alpha)$ is the corresponding integral remainder, so $f(\xi) = P(\alpha) + R(\alpha)$.

In the proof, for $g(\zeta) = (\xi - \zeta)^{p-1}/(p-1)!$,

$$g^{(k)}(\zeta) = (-1)^k \frac{(\xi - \zeta)^{p-1-k}}{(p-1-k)!} \qquad (0 \le k \le p-1)$$

and $g^{(p)}(\zeta) = 0$. Substituting $p - 1 - k$ for k yields

$$\frac{(\xi - \zeta)^k}{k!} f^{(k)}(\zeta) = (-1)^{p-1-k} g^{(p-1-k)}(\zeta) f^{(k)}(\zeta) \qquad (0 \le k \le p-1)$$

14

It follows from (8.14.1) that $DP(\zeta) = g(\zeta)f^{(p)}(\zeta)$, so

$$f(\xi) - P(\alpha) = P(\xi) - P(\alpha) = \int_\alpha^\xi g(\zeta)f^{(p)}(\zeta)\,d\zeta = R(\alpha)$$

It is more natural to run this proof in reverse, first defining the remainder $S(\alpha) = f(\xi) - P(\alpha)$, then computing $DS(\zeta) = g(\zeta)f^{(p)}(\zeta)$, so $S(\alpha) = S(\alpha) - S(\xi) = R(\alpha)$.

Remark. In the proof of Taylor's theorem (8.14.3), $\int_0^1 (1-\zeta)^{p-1}/(p-1)!\,d\zeta = 1/p!$, so the mean value theorem (8.7.7) yields

$$\left\| \int_0^1 \frac{(1-\zeta)^{p-1}}{(p-1)!} f^{(p)}(x+\zeta t)\,d\zeta - \frac{1}{p!}f^{(p)}(x) \right\|$$
$$= \left\| \int_0^1 \frac{(1-\zeta)^{p-1}}{(p-1)!} \left[f^{(p)}(x+\zeta t) - f^{(p)}(x) \right] d\zeta \right\|$$
$$\le \int_0^1 \frac{(1-\zeta)^{p-1}}{(p-1)!} \left\| f^{(p)}(x+\zeta t) - f^{(p)}(x) \right\| d\zeta$$
$$\le p!\epsilon \int_0^1 \frac{(1-\zeta)^{p-1}}{(p-1)!}\,d\zeta = \epsilon$$

References

[1] Dieudonné, J. *Foundations of Modern Analysis.* Academic Press, 1960.