J. Aust. Math. Soc. **111** (2021), 202–220
doi:10.1017/S1446788720000166

AN ALTERNATIVE APPROACH TO FRÉCHET DERIVATIVES

SHANE ARORA, HAZEL BROWNE and DANIEL DANERS

(Received 14 March 2019; accepted 20 February 2020; first published online 14 May 2020)

Communicated by W. Moors

Abstract

We discuss an alternative approach to Fréchet derivatives on Banach spaces inspired by a characterisation of derivatives due to Carathéodory. The approach allows many questions of differentiability to be reduced to questions of continuity. We demonstrate how that simplifies the theory of differentiation, including the rules of differentiation and the Schwarz lemma on the symmetry of second-order derivatives. We also provide a short proof of the differentiable dependence of fixed points in the Banach fixed point theorem.

2010 *Mathematics subject classification*: primary 46G05; secondary 26B05.

Keywords and phrases: Fréchet derivative, Schwarz lemma, Banach fixed point theorem, mean value theorems, Carathéodory derivative.

1. Introduction

The aim of this paper is to promote an alternative approach to Fréchet derivatives of functions defined on open subsets of a real or complex Banach space. The main feature is a simplification of many proofs by reducing questions of differentiability to a question of continuity. The approach is inspired by Carathéodory's characterisation of differentiability of functions on the complex plane from [9] and its extension to vector-valued functions in [1, 8].

To motivate our approach let us start with the notion of a tangent to the graph of a function $f: J \to \mathbb{R}$, where $J \subseteq \mathbb{R}$ is an open interval. Given $x \in J$, the tangent to the graph of f at $(x, f(x))$ is by definition the limit of secants through the points $(x, f(x))$ and $(y, f(y))$ as $y \to x$. The slope of that secant is given by

$$\varphi_x(y) := \frac{f(y) - f(x)}{y - x},$$

and we say that f is differentiable at x if $\varphi_x(y)$ has a limit as $y \to x$. In other words, φ_x has an extension from $J \setminus \{x\}$ to J that is continuous at x. Hence, f is differentiable at

$x \in J$ if and only if there exists a function $\varphi_x \colon J \to \mathbb{R}$, continuous at $y = x$, such that

$$f(y) = f(x) + \varphi_x(y)(y - x) \tag{1-1}$$

for all $y \in J$. By design, the derivative at x is given by $f'(x) := \varphi_x(x)$. We call φ_x the *slope function* of f at x. The continuity of φ_x at x built into the definition offers the biggest advantage over a traditional approach.

For functions between Banach spaces we can apply an idea similar to that in (1-1).

Definition 1.1. Let E, F be real or complex Banach spaces and $U \subseteq E$ open. Suppose that $f \colon U \to F$ and let $x \in U$. We say that f is *Carathéodory differentiable* at x if there exists a map $\Phi_x \colon U \to \mathcal{L}(E, F)$, continuous at x, such that

$$f(y) = f(x) + \Phi_x(y)(y - x) \tag{1-2}$$

for all $y \in U$. Here, $\mathcal{L}(E, F)$ is the space of bounded linear operators from E to F and continuity is with respect to the operator norm in $\mathcal{L}(E, F)$. We call Φ_x a *slope function* of f at x and

$$Df(x) := \Phi_x(x) \in \mathcal{L}(E, F)$$

the *derivative* of f at x.

As Φ_x is continuous at x, it is a direct consequence of (1-2) that f is continuous at every point at which it is differentiable.

We show in Section 2 that the above notion of derivative is equivalent to the usual notion of *Fréchet derivative*. Adding to the exposition in [1, 8], we provide some geometric insight and allow for any real or complex Banach space. As a demonstration of the simplicity of the approach, we then establish the standard rules of differentiation in Section 3.

To further support the case for our alternative approach to derivatives, we provide short and conceptually simple proofs of two further results. First, in Section 5, we establish the Schwarz lemma about the symmetry of second-order derivatives. Second, in Section 7, we provide a simple proof of the differentiable dependence of fixed points in the Banach fixed point theorem. That theorem can be applied directly to prove the inverse function theorem or the differentiable dependence on parameters of solutions to ordinary differential equations; see [7] for many such applications.

If f is differentiable at every point $x \in U$, then it is convenient to view the slope function as a function of two variables and write

$$\Phi(x, y) := \Phi_x(y),$$

where x is the point where we differentiate. By definition, the map $y \mapsto \Phi(x, y)$ is continuous at $y = x$ and $Df(x) = \Phi(x, x)$. We show that in general, the map $x \mapsto \Phi(x, y)$ cannot be expected to be continuous at $x = y$, not even if f is very smooth. In contrast to that, we show that if E is finite-dimensional, then there always exists a slope function that is *separately* continuous on the diagonal $x = y$ as a function of x and y. Such examples are discussed in Section 6.

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

While an arbitrary slope function can behave badly as a function of x regardless of the smoothness of f, we show that f is continuously differentiable at x if and only if there exists a slope function that is *jointly* continuous on the diagonal as a function of both variables. The proof, given in Section 4, requires a mean value inequality which, unlike most references, we prove for functions between complex Banach spaces.

We conclude this introduction by providing some historical comments. The core idea goes back to the definition of the derivative given by Carathéodory in [9]. However, Carathéodory does not really make use of his definition, but instead reverts to a standard approach. Others much later observed the the usefulness of his definition. In the single-variable case, the most complete discussion appears in [15]. In [12, Section III.6], a comparison of the definitions of the derivative due to Cauchy, Weierstrass and Carathéodory is given, and Carathéodory's definition is used to prove the standard rules of differentiation. The text [4] uses Carathéodory's approach to prove some rules of differentiation, but not beyond that. The approach is used quite consistently in the calculus textbook [16].

The first time the definition seems to appear in the multi-variable case is in [5]. The most comprehensive exposition is given in [1]. There is a generalisation to functions on Banach spaces in [8], and [17] focuses on the two-variable case, providing comparisons with other notions of differentiability. The definition also appears in [12, Section IV.3].

2. Equivalence with Fréchet derivatives

Before we start our discussion of differentiability we need some notation. The norm of $B \in \mathcal{L}(E, F)$ is the operator norm given by

$$\|B\|_{\mathcal{L}(E,F)} := \sup_{x \in E \setminus \{0\}} \frac{\|Bx\|_F}{\|x\|_E} = \sup_{\|x\|_E \leq 1} \|Bx\|_F = \sup_{\|x\|_E = 1} \|Bx\|_F;$$

see, for instance, [20, Section II.1]. A special case is the dual space $E' := \mathcal{L}(E, \mathbb{K})$ of E, where $\mathbb{K} = \mathbb{R}$ if E is a real Banach space and $\mathbb{K} = \mathbb{C}$ if E is complex. The dual norm $\|\cdot\|_{E'}$ is just the operator norm in $\mathcal{L}(E, \mathbb{K})$. When no confusion is likely we denote the norms on E and F simply by $\|\cdot\|$.

Let $f \colon U \to F$, where $U \subseteq E$ is open. The usual definition of the derivative at $x \in U$ is the Fréchet derivative. The idea is to find a linear operator $A \in \mathcal{L}(E, F)$ providing the *best linear approximation* of f near $x \in U$ in the sense that

$$\lim_{y \to x} \frac{f(y) - f(x) - A(y - x)}{\|y - x\|} = 0 \tag{2-1}$$

in F. The map A is called the derivative of f at x and is denoted by $Df(x)$. The name goes back to Fréchet [11], but Fréchet attributes the definition to Stolz [19].

We now show that Fréchet's and Carathéodory's notions of derivatives are equivalent. This is shown in [1, 8], but unlike these references we include a proof emphasising the geometric significance of the constructions and allow for complex Banach spaces.

Assume that f is Carathéodory differentiable in the sense of Definition 1.1 and set $A := \Phi_x(x)$. Then

$$\frac{\|f(y) - f(x) - A(y-x)\|}{\|y-x\|} = \frac{\|(\Phi_x(y) - \Phi_x(x))(y-x)\|}{\|y-x\|}$$
$$\leq \|\Phi_x(y) - \Phi_x(x)\|_{\mathcal{L}(E,F)} \frac{\|y-x\|}{\|y-x\|} = \|\Phi_x(y) - \Phi_x(x)\|_{\mathcal{L}(E,F)}.$$

Due to the continuity of $y \mapsto \Phi_x(y)$ at x, we know that

$$\|\Phi_x(y) - \Phi_x(x)\|_{\mathcal{L}(E,F)} \to 0$$

as $y \to x$ and hence (2-1) holds, showing that f is Fréchet differentiable.

Assuming that f is Fréchet differentiable at x, we need to construct a slope function Φ_x at x. Given $y \in U$ with $y \neq x$, that slope function is uniquely defined in the direction of $y - x$ by (1-2), namely by

$$\Phi_x(y)(y-x) = f(y) - f(x).$$

We then need to define Φ_x on a subspace complementary to the line $\{t(y-x) : t \in \mathbb{K}\}$. Such a complement is given by the kernel of a linear functional $\ell(x,y) \in E'$ with $\langle \ell(x,y), y-x\rangle \neq 0$. For $z \in \ker(\ell(x,y))$ we define $\Phi_x(y)z = Df(x)z$. That construction is possible by the Hahn–Banach theorem which guarantees the existence of a bounded linear functional $\ell(x,y) \in E'$ such that $\langle \ell(x,y), y-x\rangle = \|y-x\|$ and $\|\ell(x,y)\|_{E'} = 1$; see [6, Corollary 1.3]. Geometrically this means that, in the direction of $\ker(\ell(x,y))$, the slope function Φ_x is determined by the tangent of f at $(x, f(x))$; see Figure 2.1. We can write

$$\Phi_x(y)z := \begin{cases} \dfrac{f(y) - f(x) - Df(x)(y-x)}{\|y-x\|}\langle \ell(x,y), z\rangle + Df(x)z & \text{if } x \neq y, \\ Df(x)z & \text{if } x = y, \end{cases} \tag{2-2}$$

for all $z \in E$. This is a slope function since $\langle \ell(x,y), y-x\rangle = \|y-x\|$, and so by construction $f(y) = f(x) + \Phi_x(y)(y-x)$. Moreover, since $\|\ell(x,y)\|_{E'} = 1$,

$$\|\Phi_x(y)z - \Phi_x(x)z\| = \frac{\|f(y) - f(x) - Df(x)(y-x)\|}{\|y-x\|}|\langle \ell(x,y), z\rangle|$$
$$\leq \frac{\|f(y) - f(x) - Df(x)(y-x)\|}{\|y-x\|}\|z\|$$

for all $z \in E$. By definition of the operator norm and since f is Fréchet differentiable,

$$\|\Phi_x(y) - \Phi_x(x)\|_{\mathcal{L}(E,F)} \leq \frac{\|f(y) - f(x) - Df(x)(y-x)\|}{\|y-x\|} \xrightarrow{y \to x} 0.$$

Hence f is Carathéodory differentiable. Note that if the dual norm on E' is strictly convex, then the functional $\ell(x,y)$ given by the duality map is uniquely determined; see [6, Exercise 1.1]. For this reason we call (2-2) the *canonical slope function*. We note that it is sufficient for $\ell(x,y) \in E'$ to have a bound independent of y in a neighbourhood of x for the above arguments to work.

We next look at some cases where it is possible to make a natural choice for $\ell(x,y)$.

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

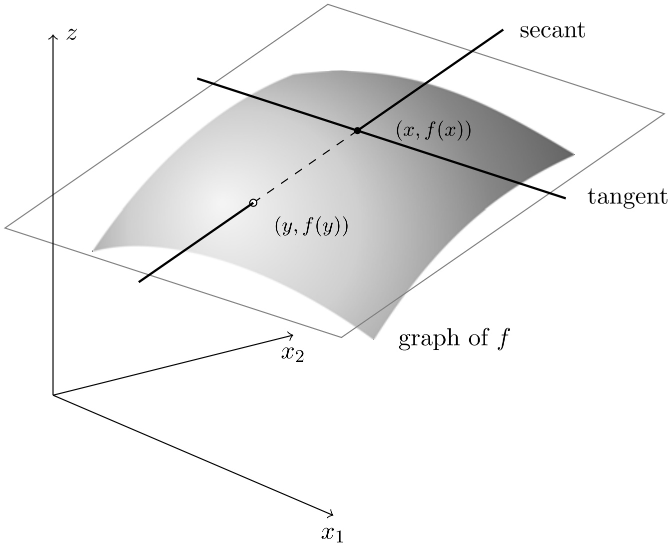

FIGURE 2.1. Plane spanned by secant and tangent to define the canonical slope function.

EXAMPLE 2.1. (a) If $E = H$ is a finite- or infinite-dimensional Hilbert space with inner product $\langle\cdot\,,\cdot\rangle_H$, then $\ell(x, y)$ is the orthogonal projection onto the subspace spanned by $y - x$, or more precisely the component in that direction. This is given by

$$\langle \ell(x, y), z\rangle := \Big\langle \frac{y - x}{\|y - x\|_H}, z\Big\rangle_H \tag{2-3}$$

and illustrated in Figure 2.1. For a complex Hilbert space we take the inner product conjugate linear in the first argument.

(b) If $E = L^p(\Omega)$ for some measure space (Ω, μ) with $1 < p < \infty$, then

$$\langle \ell(u, v), w\rangle := \frac{1}{\|v - u\|_p^{p-1}} \int_\Omega |v - u|^{p-2}(\overline{v - u})w\, d\mu.$$

Clearly $\langle \ell(u, v), v - u\rangle = \|v - u\|_p$ and by Hölder's inequality $|\langle \ell(u, v), w\rangle| \leq \|w\|_p$, so $\|\ell(u, v)\|_{(L^p)'} = 1$. In the Hilbert space case $p = 2$ this coincides with (2-3).

(c) If the norm $\|\cdot\|_E$ on E is equivalent to a norm $\|\cdot\|_H$ induced by an inner product $\langle\cdot\,,\cdot\rangle_H$, then we can choose

$$\langle \ell(x, y), z\rangle := \frac{\|y - x\|_E}{\|y - x\|_H}\Big\langle \frac{y - x}{\|y - x\|_H}, z\Big\rangle_H.$$

In particular, this is the case when working on any finite-dimensional space such as $\mathbb{R}^n$ or $\mathbb{C}^n$, where every norm is equivalent to the Euclidean norm. We do not necessarily have $\|\ell(x, y)\|_{E'} = 1$, but we still maintain the required uniform bound.

Given the nonuniqueness of complements of the space spanned by $y - x$ used to construct the slope function (2-2), it is clear that the slope function cannot be unique

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

unless $\dim(E) = 1$. Also, the slope function does not need to be of the form (2-2). For examples we refer to [1, Section 2] and to our more comprehensive discussion of slope functions in Section 6. However, the derivative is in fact unique. We provide a proof simpler than that given in [1, Section 2].

Proposition 2.2 (Uniqueness of derivative). *Let E, F be Banach spaces, $U \subseteq E$ open and $f\colon U \to F$ Carathéodory differentiable at $x \in U$. Then the derivative at x is unique.*

Proof. Let $\Phi_x\colon U \to \mathcal{L}(E, F)$ be an arbitrary slope function. Fix $z \in E$ and suppose that $t_0 > 0$ is small enough so that $x + tz \in U$ for all $t \in (0, t_0]$. This is possible since U is open. By definition of Φ_x we have $f(x + tz) - f(x) = \Phi_x(x + tz)tz$ for all $t \in (0, t_0]$, and so, by the continuity of Φ_x at x,

$$\lim_{t\to 0+} \frac{f(x + tz) - f(x)}{t} = \lim_{t\to 0+} \Phi_x(x + tz)z = \Phi_x(x)z. \tag{2-4}$$

As the left-hand side of (2-4) is independent of the particular slope function Φ_x, it follows that $\Phi_x(x)z$ is uniquely determined by f, x and z. As this is true for every $z \in E$, the derivative is unique. □

Remark 2.3. If $f\colon \mathbb{R}^n \to \mathbb{R}^m$ (or $f\colon \mathbb{C}^n \to \mathbb{C}^m$), then the identity (2-4) also shows that the matrix representation of $Df(x)$ with respect to the standard basis is given by the Jacobian matrix. Indeed, if we choose $z = e_k$ to be the kth standard basis vector of $\mathbb{R}^n$ (or $\mathbb{C}^n$), then the left-hand side of (2-4) by definition is the partial derivative of f with respect to x_k. Hence,

$$Df(x)e_k = \frac{\partial f}{\partial x_k}(x) := \begin{bmatrix} \dfrac{\partial f_1}{\partial x_k}(x) \\ \vdots \\ \dfrac{\partial f_m}{\partial x_k}(x) \end{bmatrix}$$

for $k = 1, \dots, n$, giving the kth column of the Jacobian matrix.

3. The rules of differentiation

The proofs of the standard rules of differentiation provide a convincing case for the simplicity of Carathéodory's characterisation of derivatives. The idea is always the same: through simple algebraic manipulations we identify a slope function and exploit its continuity at the point at which the derivative is taken. Unlike the traditional approach, no 'ε–δ' or 'little o' arguments are needed, only clean and transparent arguments involving continuity properties of the slope function and the function itself.

Proposition 3.1 (Linearity). *Suppose that E, F are real or complex Banach spaces, that $U \subseteq E$ is open and that $f, g\colon U \to F$ are differentiable at $x \in U$. If $\lambda, \mu \in \mathbb{R}$ (or $\mathbb{C}$), then $D(\lambda f + \mu g)(x) = \lambda Df(x) + \mu Dg(x)$.*

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

Proof. Take slope functions Φ_x and Ψ_x at x for f and g, respectively. Then

$$\lambda f(y) + \mu g(y) = \lambda f(x) + \mu g(x) + (\lambda\Phi_x(y) + \mu\Psi_x(y))(y - x).$$

Clearly, $\lambda\Phi_x(y) + \mu\Psi_x(y) \in \mathcal{L}(E, F)$ is continuous at $y = x$ and hence

$$D(\lambda f + \mu g)(x) = \lambda\Phi_x(x) + \mu\Psi_x(x) = \lambda Df(x) + \mu Dg(x),$$

as claimed. □

We next prove the chain rule, which is a good example of how our approach reduces questions about differentiability to questions of continuity by identifying an appropriate slope function. Compare, for instance, with the traditional proof of the chain rule in [18, Theorem 9.15]. The proof below is given in [1] for functions defined on Euclidean space, but translates without change to real and complex Banach spaces.

Theorem 3.2 (Chain rule). *Suppose that E, F, G are Banach spaces and that $U \subseteq E$ and $V \subseteq F$ are open sets. Assume that $g\colon U \to F$ is differentiable at $x \in U$ and that $g(x) \in V$. Further, assume that $f\colon V \to G$ is differentiable at $g(x)$. Then $f \circ g$ is differentiable at x and $D(f \circ g)(x) = Df(g(x))Dg(x)$.*

Proof. Suppose that $\Phi\colon V \to \mathcal{L}(F, G)$ is a slope function of f at $g(x)$ and that $\Psi\colon U \to \mathcal{L}(E, F)$ is a slope function of g at x, that is,

$$\begin{aligned} f(z) &= f(g(x)) + \Phi(z)(z - g(x)) && \text{for all } z \in V, \\ g(y) &= g(x) + \Psi(y)(y - x) && \text{for all } y \in U. \end{aligned}$$

In particular, f and g are continuous at $g(x)$ and x, respectively. Using the two identities we can write

$$\begin{aligned} (f \circ g)(y) &= f(g(x)) + \Phi(g(y))(g(y) - g(x)) \\ &= (f \circ g)(x) + \Phi(g(y))\Psi(y)(y - x). \end{aligned}$$

Hence, $y \mapsto \Lambda(y) := \Phi(g(y))\Psi(y)$ is a slope function for $f \circ g$ at x. Using that the composition of continuous functions is continuous, Λ is continuous at $y = x$ and thus

$$D(f \circ g)(x) = \Lambda(x) = \Phi(g(x))\Psi(x) = Df(g(x))Dg(x),$$

as claimed. □

We next prove a product rule. Products are not generally defined on Banach spaces, but the main feature of products is that they are bilinear. We let E, F_1 and F_2 be Banach spaces and $U \subseteq E$ an open set. Let G be another Banach space and assume that $b\colon F_1 \times F_2 \to G$ is bounded and bilinear. 'Bounded' means that there exists $M > 0$ such that

$$\|b(y_1, y_2)\|_G \leq M\|y_1\|_{F_1}\|y_2\|_{F_2}$$

for all $y_1 \in F_1$ and $y_2 \in F_2$. Given functions $f_k\colon U \to F_k$, $k = 1, 2$, we consider $g\colon U \to G$ given by

$$g(x) := b(f_1(x), f_2(x))$$

for all $x \in U$. The following proposition applies to pointwise products of functions, the cross product, inner products and other bilinear operations.

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

PROPOSITION 3.3 (Product rule). *Let the above assumptions be satisfied and assume that f_1, f_2 are differentiable at $x \in U$. Then g is differentiable with derivative given by*

$$Dg(x)z = b(Df_1(x)z, f_2(x)) + b(f_1(x), Df_2(x)z) \tag{3-1}$$

for all $z \in E$.

PROOF. Let Φ_1, Φ_2 be slope functions for f_1 and f_2 at x, respectively. Then, using that b is bilinear, we obtain

$$\begin{aligned} g(y) &= b(f_1(y), f_2(y)) = b(f_1(x), f_2(y)) + b(\Phi_1(y)(y-x), f_2(y)) \\ &= b(f_1(x), f_2(x)) + b(\Phi_1(y)(y-x), f_2(y)) + b(f_1(x), \Phi_2(y)(y-x)) \\ &= g(x) + \Psi(y)(y-x), \end{aligned}$$

where we have set

$$\Psi(y)z := b(\Phi_1(y)z, f_2(y)) + b(f_1(x), \Phi_2(y)z)$$

for all $z \in E$. As b is bounded and bilinear we deduce that $\Psi(y) \in \mathcal{L}(E, G)$ is continuous at $y = x$, implying (3-1) since $\Phi_k(x) = Df_k(x)$ by definition. □

Another common rule of differentiation is the quotient rule, but like the usual product rule it does not directly apply in Banach spaces. Note, however, that the quotient rule is really a composition of a function with inversion $t \mapsto 1/t = t^{-1}$ on $\mathbb{R}$ or $\mathbb{C}$. Hence the natural generalisation of the quotient rule is the derivative of the map $B \mapsto B^{-1}$ on the set of bounded invertible linear operators. It is known that this set is open in $\mathcal{L}(E)$ and that the map $B \mapsto B^{-1}$ is continuous; see, for instance, [20, Theorem IV.1.5]. Based on this fact, we show that this map is also differentiable at every invertible $A \in \mathcal{L}(E)$.

THEOREM 3.4 (Inversion). *Let $A \in \mathcal{L}(E)$ be invertible. Then the map f given by $f(B) := B^{-1}$ is differentiable at A, and, for $Z \in \mathcal{L}(E)$,*

$$Df(A)Z = -A^{-1}ZA^{-1}. \tag{3-2}$$

PROOF. If $A, B \in \mathcal{L}(E)$ are invertible, then

$$\begin{aligned} B^{-1} &= A^{-1} - A^{-1} + B^{-1} \\ &= A^{-1} - A^{-1}(B-A)B^{-1} = A^{-1} + \Phi(A,B)(B-A), \end{aligned}$$

where $\Phi(A,B)Z := -A^{-1}ZB^{-1}$ for all $Z \in \mathcal{L}(E)$. Then $\Phi(A,B) \in \mathcal{L}(\mathcal{L}(E))$ and

$$\|\Phi(A,B)Z - \Phi(A,A)Z\| = \|A^{-1}Z(A^{-1} - B^{-1})\| \le \|A^{-1}\| \, \|A^{-1} - B^{-1}\| \, \|Z\|.$$

Here, $\|\cdot\|$ is the norm in $\mathcal{L}(E)$. By definition of the operator norm, continuity of inversion and since the set of invertible operators is open,

$$\|\Phi(A,B) - \Phi(A,A)\|_{\mathcal{L}(\mathcal{L}(E))} \le \|A^{-1}\| \, \|A^{-1} - B^{-1}\| \to 0$$

as $B \to A$ in $\mathcal{L}(E)$. Hence Φ is a slope function for f and $\Phi(A,B) \to \Phi(A,A)$ in $\mathcal{L}(\mathcal{L}(E))$, proving (3-2). □

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

Finally, we look at functions on a product space and partial derivatives.

Proposition 3.5 (Partial derivatives). *Let E_1, E_2 and F be Banach spaces and let $U \subseteq E_1 \times E_2$ be open. Assume that $f\colon U \to F$ is differentiable at $x = (x_1, x_2) \in U$ with slope function Φ. For $z_1 \in E_1$ we define the partial slope function Φ_1 by*

$$\Phi_1(x, y_1)z_1 := \Phi(x, (y_1, x_2))(z_1, 0). \tag{3-3}$$

Then the function $y_1 \mapsto f(y_1, x_2)$ defined on $U_{x_2} := \{y_1 \in E_1 : (y_1, x_2) \in U\}$ is differentiable with slope function $\Phi_1(x, \cdot)\colon U_{x_2} \to \mathcal{L}(E_1)$ and derivative given by $D_1 f(x_1, x_2)z_1 = Df(x)(z_1, 0)$ for all $z_1 \in E_1$.

Proof. By definition of a slope function and (3-3),

$$f(y_1, x_2) = f(x) + \Phi(x, (y_1, x_2))(y_1 - x_1, 0) = f(x) + \Phi_1(x, y_1)(y_1 - x_1).$$

From properties of Φ we have that $\Phi_1(x, y_1) \to \Phi_1(x, x_1)$ in $\mathcal{L}(E_1, F)$ as $y_1 \to x_1$. Hence, $y_1 \mapsto f(y_1, x_2)$ is differentiable at x_1 with slope function $\Phi_1(x, \cdot)$ at x_1. □

Note that the slope function of $y_1 \mapsto f(y_1, x_2)$ depends on x_2. For that reason we have kept $x = (x_1, x_2)$ as the first argument of Φ_1 and not just x_1. As usual, we sometimes write $D_{x_1} f(x_1, x_2)$ or $D_1 f(x_1, x_2)$ for the partial derivative. In a similar fashion we obtain the partial derivative with respect to x_2. A similar approach works for products of more than two spaces.

4. Characterisation of continuous differentiability

Assume that $U \subseteq E$ is open and that $f\colon U \to F$ is differentiable. For every slope function $\Phi(x, y)$ we require the continuity of $y \mapsto \Phi(x, y)$ at x by definition. We do not say anything about continuity as a function of x, let alone joint continuity as a function of (x, y). It turns out that continuous differentiability can be characterised by means of such a joint continuity property. Such a characterisation appears in [1, Section 5], but apart from a generalisation to the Banach space case we also provide details on how exactly the mean value theorem is used.

Theorem 4.1. *Suppose that E, F are Banach spaces, that $U \subseteq E$ is open and that $f\colon U \to F$ is differentiable. Then Df is continuous at x_0 if and only if there exists a slope function $\Phi(\cdot\,, \cdot)$ for f that is (jointly) continuous at (x_0, x_0) as a function of (x, y). In that case, the canonical slope function given by (2-2) is jointly continuous at (x_0, x_0).*

It is tempting to believe that every slope function has the above joint continuity property if Df is continuous at x_0. However, as we show in Section 6, one can always construct a slope function that is not even separately continuous. This is not bad because in practice we only need to know that a jointly continuous slope function exists. Note also that we make no claim on the continuity of Φ at points other than (x_0, x_0).

The main tool to prove the above theorem is a mean value inequality. To simplify the statement, we denote the line segment connecting x and y in E by

$$[\![x, y]\!] := \{x + t(y - x) : t \in [0, 1]\}.$$

The idea is taken from [18, Theorem 5.19], but instead of inner products in $\mathbb{R}^n$ we use duality in Banach spaces. We also deal with the case of complex Banach spaces.

Theorem 4.2 (Mean value inequality). *Assume that E, F are Banach spaces, that $U \subseteq E$ is open and that $f : U \to F$ is differentiable. Let $A \in \mathcal{L}(E, F)$ and let $x, y \in U$ be distinct points such that $[\![x, y]\!] \subseteq U$. Then there exists $c \in [\![x, y]\!]$, $c \neq x, y$, such that*

$$\|f(y) - f(x) - A(y - x)\| \le \|Df(c)(y - x) - A(y - x)\|. \tag{4-1}$$

Proof. By the Hahn–Banach theorem there exists $\varphi \in F'$ with $\|\varphi\|_{F'} = 1$ such that

$$\langle \varphi, f(y) - f(x) - A(y - x)\rangle = \|f(y) - f(x) - A(y - x)\|_F; \tag{4-2}$$

see [6, Corollary 1.3]. We next define the function $g : [0, 1] \to \mathbb{C}$ by

$$g(t) := \langle \varphi, f(x + t(y - x)) - f(x) - tA(y - x)\rangle. \tag{4-3}$$

It is well defined since $[\![x, y]\!] \subseteq U$ by assumption. It is real-valued if E, F are real Banach spaces. To allow for complex Banach spaces, we define the function $H : [0, 1] \to \mathbb{R}$ by

$$H(t) = \operatorname{Re}(\overline{g(1)}g(t))$$

for all $t \in [0, 1]$. We note that a complex-valued function of $t \in \mathbb{R}$ is differentiable if and only if its real and imaginary parts are differentiable. As $g(0) = 0$ and hence $H(0) = 0$, by the classical mean value theorem there exists $t_0 \in (0, 1)$ such that

$$|g(1)|^2 = H(1) - H(0) = H'(t_0) = \operatorname{Re}(\overline{g(1)}g'(t_0)) \le |g(1)||g'(t_0)|.$$

Hence, $|g(1)| \le |g'(t_0)|$. Using (4-2), (4-3) and the chain rule we deduce that

$$\begin{aligned}
&\|f(y) - f(x) - A(y - x)\| \\
&\quad = |g(1)| \le |g'(t_0)| \\
&\quad = |\langle \varphi, Df(x + t_0(y - x))(y - x) - A(y - x)\rangle| \\
&\quad \le \|Df(x + t_0(y - x))(y - x) - A(y - x)\|.
\end{aligned}$$

In the last step we used that $\|\varphi\|_{F'} = 1$. To complete the proof of (4-1) we finally set $c := x + t_0(y - x)$. Clearly $c \in [\![x, y]\!]$, $c \neq x, y$, since $t_0 \in (0, 1)$. □

From the above mean value inequality we can derive an inequality involving the canonical slope function (2-2).

Corollary 4.3. *Suppose that the assumptions of Theorem 4.2 are satisfied, and that $\Phi(x, y)$ is the canonical slope function of f given by (2-2). If $x, y \in U$ are distinct points such that $[\![x, y]\!] \subseteq U$, then there exists $c \in [\![x, y]\!]$, $c \neq x, y$, such that*

$$\|\Phi(x, y) - A\|_{\mathcal{L}(E,F)} \le \|Df(c) - Df(x)\|_{\mathcal{L}(E,F)} + \|Df(x) - A\|_{\mathcal{L}(E,F)}. \tag{4-4}$$

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

Proof. We start by noting that, for all $z \in E$,

$$\begin{aligned}
&\|\Phi(x, y)z - Az\| \\
&\quad= \left\| \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|} \langle \ell(x, y), z \rangle + Df(x)z - Az \right\| \\
&\quad\leq \left\| \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|} \right\| \|\ell(x, y)\|_{E'} \|z\| \\
&\qquad + \|Df(x) - A\|_{\mathcal{L}(E,F)} \|z\| \\
&\quad= \left\| \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|} \right\| \|z\| + \|Df(x) - A\|_{\mathcal{L}(E,F)} \|z\|,
\end{aligned}$$

where we used that $\|\ell(x, y)\|_{E'} = 1$. Hence, by definition of the operator norm,

$$\|\Phi(x, y) - A\|_{\mathcal{L}(E,F)} \leq \left\| \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|} \right\| + \|Df(x) - A\|_{\mathcal{L}(E,F)}.$$

Applying Theorem 4.2, there exists $c \in [\![x, y]\!]$ with

$$\begin{aligned}
&\left\| \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|} \right\| \\
&\qquad\leq \frac{1}{\|y - x\|} \|Df(c)(y - x) - Df(x)(y - x)\| \\
&\qquad\leq \|Df(c) - Df(x)\|_{\mathcal{L}(E,F)} \frac{\|y - x\|}{\|y - x\|} \\
&\qquad= \|Df(c) - Df(x)\|_{\mathcal{L}(E,F)}.
\end{aligned}$$

Combining the above, (4-4) follows. □

Remark 4.4. As seen from the proof of Theorem 4.2 and Corollary 4.3, it is sufficient to assume that f is continuous at the endpoints of $[\![x, y]\!]$ and differentiable inside.

We are now in a position to prove Theorem 4.1.

Proof of Theorem 4.1. First assume that there exists a slope function Φ that is continuous at (x_0, x_0). Then in particular the function $x \mapsto \Phi(x, x) = Df(x)$ is continuous at x_0, that is, Df is continuous at x_0.

Assume now that Df is continuous at x_0. We choose the slope function $\Phi(x, y)$ of f given by (2-2). As U is open we can find $r > 0$ such that $B(x_0, r) \subseteq U$. If we fix $x, y \in B(x_0, r)$, then, applying (4-4) with $A = Df(x_0)$, there exists $c_{x,y} \in [\![x, y]\!]$ with

$$\begin{aligned}
&\|\Phi(x, y) - Df(x_0)\|_{\mathcal{L}(E,F)} \\
&\qquad\leq \|Df(c_{x,y}) - Df(x)\|_{\mathcal{L}(E,F)} + \|Df(x) - Df(x_0)\|_{\mathcal{L}(E,F)}.
\end{aligned} \tag{4-5}$$

As $c_{x,y}$ is a convex combination of x and y, it follows that $c_{x,y} \in B(x_0, r)$ and that $c_{x,y} \to x_0$ as $(x, y) \to (x_0, x_0)$. By the continuity of Df at x_0, we deduce from (4-5) that

$$\lim_{(x,y)\to(x_0,x_0)} \|\Phi(x, y) - Df(x_0)\|_{\mathcal{L}(E,F)} = 0,$$

proving the joint continuity of Φ at (x_0, x_0). □

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

5. The symmetry of second-order derivatives

If $f\colon U \to E$ is differentiable, then it makes sense to consider the second-order derivative. As $Df\colon U \to \mathcal{L}(E, F)$, the second-order derivative $D^2 f(x)$ is a linear operator from E into $\mathcal{L}(E, F)$, that is, $D^2 f(x) \in \mathcal{L}(E, \mathcal{L}(E, F))$. As commonly done, we identify $\mathcal{L}(E, \mathcal{L}(E, F))$ with the space $\mathcal{L}^2(E \times E; F)$ of bounded bilinear maps from $E \times E$ to F; see, for instance, [3, Theorem 4.3]. With that identification $D^2 f(x) \in \mathcal{L}^2(E \times E; F)$. We use the theory developed so far to provide a simple proof of the well-known fact that $D^2 f(x)$ is symmetric, named after Schwarz, Young or Clairaut depending on local tradition. Most references provide a proof if the second-order derivative is continuous. We only assume that it exists at one point.

THEOREM 5.1 (Symmetry of second-order derivatives). *Assume that $f\colon U \to F$ is such that $D^2 f(x)$ exists at the point $x \in U$. Then $D^2 f(x)$ is symmetric, that is, $D^2 f(x)[u, v] = D^2 f(x)[v, u]$ for all $u, v \in E$.*

PROOF. We first note that for $D^2 f(x)$ to exist, f needs to be differentiable in a neighbourhood of x. We fix $u, v \in E$. As f is differentiable in a neighbourhood of x, for fixed $s > 0$ small enough, the function $g\colon [0, s] \to F$ given by

$$g(t) := f(x + su + tv) - f(x + tv) \tag{5-1}$$

is well defined and differentiable. Thus the mean value inequality from Theorem 4.2 implies the existence of $\theta \in (0, 1)$ such that

$$\|g(s) - g(0) - s^2 D^2 f(x)[u, v]\| \le \|g'(\theta s)s - s^2 D^2 f(x)[u, v]\|, \tag{5-2}$$

where we have set $At := tsD^2 f(x)[u, v]$ for the linear map $A\colon \mathbb{R} \to F$. As Df is differentiable at x there exists a slope function $\Phi\colon U \to \mathcal{L}^2(E \times E; F)$ for Df at x. Using the chain rule to compute g', we see that

$$\begin{aligned} g'(\theta s) &= Df(x + su + \theta sv)v - Df(x + \theta sv)v \\ &= (Df(x + su + \theta sv) - Df(x))v - (Df(x + \theta sv) - Df(x))v \\ &= \Phi(x + su + \theta sv)[su + \theta sv, v] - \Phi(x + \theta sv)[\theta sv, v] \\ &= s(\Phi(x + su + \theta sv) - \Phi(x + \theta sv))[\theta v, v] + s\Phi(x + su + \theta sv)[u, v]. \end{aligned}$$

Combining the above identity with (5-2) and using that $\theta \in (0, 1)$, we arrive at

$$\begin{aligned} &\left\| \frac{g(s) - g(0)}{s^2} - D^2 f(x)[u, v] \right\|_F \\ &\quad \le \left\| \frac{g'(\theta s)}{s} - D^2 f(x)[u, v] \right\|_F \\ &\quad \le \|\Phi(x + su + \theta sv) - \Phi(x + \theta sv)\|_{\mathcal{L}^2(E\times E;F)} \|v\|_E^2 \\ &\qquad + \|\Phi(x + su + \theta sv) - D^2 f(x)\|_{\mathcal{L}^2(E\times E;F)} \|u\|_E \|v\|_E. \end{aligned} \tag{5-3}$$

By definition of differentiability, Φ is continuous at x and hence

$$\lim_{s\to 0+} \Phi(x + su + \theta sv) = \lim_{s\to 0+} \Phi(x + \theta sv) = D^2 f(x)$$

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

in $\mathcal{L}^2(E \times E; F)$. Hence, the right-hand side of (5-3) goes to zero as $s \to 0+$, that is,

$$\lim_{s\to 0+} \frac{g(s) - g(0)}{s^2} = D^2 f(x)[u, v].$$

Looking at the definition of g given in (5-1), we see that $g(s) - g(0)$ is symmetric as a function of (u, v), so by interchanging the roles of u and v we also have

$$\lim_{s\to 0+} \frac{g(s) - g(0)}{s^2} = D^2 f(x)[v, u],$$

proving that $D^2 f(x)[u, v] = D^2 f(x)[v, u]$. □

Remark 5.2. By an induction argument, the above theorem implies the symmetry of all higher-order derivatives. The induction argument used in [3, Corollary VII.4.7] can be adapted for that purpose. In the case of a function $f : \mathbb{R}^n \to \mathbb{R}$, symmetry means that the Hessian matrix is symmetric, and more generally that partial derivatives can be taken in any order to yield the same result.

6. Further discussion of slope functions

In this section we provide a further discussion of slope functions. In particular, we discuss joint and separate continuity, symmetry, and derivatives of Lipschitz functions.

6.1. Joint and separate continuity. If $g : \mathbb{R} \to \mathbb{R}$ is differentiable, then the slope function is uniquely determined and given by

$$\varphi(s, t) := \begin{cases} \dfrac{g(t) - g(s)}{t - s} & \text{if } t \neq s, \\ g'(s) & \text{if } t = s. \end{cases} \tag{6-1}$$

Clearly $\varphi(s, t) = \varphi(t, s)$ and hence φ is separately continuous at (s, s), that is, $t \mapsto \varphi(s, t)$ is continuous at s and $t \mapsto \varphi(t, s)$ is continuous at s. We show that this is not necessarily the case for functions of two or more variables.

Example 6.1. For $s \in \mathbb{R}$ define $g(s) := s^2 \cos(1/s)$ if $s \neq 0$ and $g(0) := 0$. We can define a function of two variables by setting

$$f(x) := g(x_1)$$

for all $x = (x_1, x_2) \in \mathbb{R}^2$. If $x_1 = 0$, then the canonical slope function (2-2) is the 1×2 matrix given by

$$\Phi(x, y) = \frac{f(y)}{\|y - x\|^2} [y_1 \quad y_2 - x_2]$$

for all $y \neq x$. If $x_1 \neq 0$, then it is given by

$$\begin{aligned} \Phi(x, y) &= \frac{f(y) - f(x) - Df(x)(y - x)}{\|y - x\|^2} [y_1 - x_1 \quad y_2 - x_2] + Df(x) \\ &= \frac{y_1 - x_1}{\|y - x\|^2} (\varphi(x_1, y_1) - g'(x_1)) [y_1 - x_1 \quad y_2 - x_2] + [g'(x_1) \quad 0], \end{aligned}$$

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

where φ is the slope function of g given by (6-1). Obviously $\Phi(0, x) \neq \Phi(x, 0)$, which is not surprising given the geometric interpretation of slope functions from Section 2. What is more interesting is that Φ is not separately continuous at $(0, 0)$. In particular, $\lim_{x\to 0} \Phi(x, 0)$ does not exist. Indeed, since $|g'(x_1)| \leq 2$ for $|x_1| \leq 1$ and $\varphi(x_1, 0) \to g'(0) = 0$ as $x_1 \to 0$ it follows that

$$\lim_{x\to 0} \frac{x_1}{\|x\|^2}(\varphi(x_1, 0) - g'(x_1))[x_1 \quad x_2] = 0.$$

However, the second term $[g'(x_1) \;\; 0]$ does not converge as $x_1 \to 0$.

The above example also shows that the canonical slope function is not always the best one to use. Here, there is a much simpler one with much better properties, namely

$$\Psi(x, y) := [\varphi(x_1, y_1) \quad 0].$$

Inheriting the properties of φ, it follows that Ψ is separately continuous and symmetric.

At every point $(0, x_2)$, the function f in the above example is not continuously differentiable. We now show that separate continuity of the slope function can fail regardless of how smooth the function is. This makes it clear that Theorem 4.1 is optimal in the sense that it can only ever assert the existence of a jointly continuous slope function, but nothing can be said about an arbitrary slope function.

Example 6.2. Consider the zero function $f(x) := 0$ for all $x \in \mathbb{R}^2$, whose derivative is given by $Df(x) = [0 \;\; 0]$ for all $x \in \mathbb{R}^2$. Suppose that $g\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is such that

$$\lim_{y\to x} g(x, y) = 0. \tag{6-2}$$

Then the linear operator $\Phi(x, y) \in \mathcal{L}(\mathbb{R}^2, \mathbb{R})$ given by

$$\Phi(x, y) := g(x, y)\left[-\frac{y_2 - x_2}{\|y - x\|} \quad \frac{y_1 - x_1}{\|y - x\|}\right]$$

if $x \neq y$ and $\Phi(x, x) := [0 \;\; 0]$ defines a slope function for f at x. Indeed, note that $\Phi(x, y)(y - x) = 0$ and that $\|\Phi(x, y)\| \leq |g(x, y)|$ for all $x, y \in \mathbb{R}^2$. Therefore, by (6-2), for every $x \in \mathbb{R}^2$ we have $\Phi(x, y) \to [0 \;\; 0]$ as $y \to x$. We choose g to be given by

$$g(x, y) := \begin{cases} 1 & \text{if } y = 0 \text{ and } x \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then (6-2) holds for all $x \in \mathbb{R}^2$, but $g(x, 0) = 1 \to 1 \neq 0 = g(0, 0)$ as $x \to 0$. In particular,

$$\lim_{y\to 0} \Phi(0, y) = [0 \quad 0] \quad \text{but } \lim_{x\to 0} \Phi(x, 0) \text{ does not exist.}$$

This means that Φ is not separately continuous as a function of x and y at $(0, 0)$ even though f is as smooth as we like. Given an arbitrary smooth function from $\mathbb{R}^2$ to $\mathbb{R}$, we can always add Φ to the corresponding slope function and get a badly behaved one. Likewise, we can do that at any point in the domain by translation.

The example can be modified to work on any Banach space E by looking at a pair of nontrivial complemented subspaces $E = E_1 \oplus E_2$ and choosing $x \in E_1$ and $y \in E_2$.

In contrast to the above examples we show that at least in finite dimensions, for any differentiable function (not necessarily continuously differentiable) one can always construct a separately continuous and symmetric slope function.

6.2. Symmetry and separate continuity of the slope function. We know that the slope function φ of a differentiable function of one variable is symmetric, that is, $\varphi(x, y) = \varphi(y, x)$. We also know from previous discussions and Example 6.1 that this is not necessarily the case for any given slope function Φ of a function of several variables. If Φ is separately continuous, then the symmetric part

$$\Psi(x, y) := \tfrac{1}{2}(\Phi(x, y) + \Phi(y, x))$$

is a slope function. Hence, there is a symmetric slope function if and only if there exists a separately continuous slope function. If f is continuously differentiable, then, by Theorem 4.1, we have such a slope function. We could ask whether it is possible to construct a separately continuous slope function for a function that is just differentiable. It turns out that this is the case for a function of finitely many variables.

Given a differentiable function $f \colon U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$ open, we now construct a separately continuous slope function. The construction comes closest to the definition of a derivative for a function of one variable as a limit of secants. The idea is to consider a secant plane and pass to the limit to obtain the tangent plane.

For each pair of points x, y set $v_1 = (y - x)/\|y - x\|$ and choose vectors v_k, $k = 2, \dots, n$, so that $(v_1, v_2, \dots, v_n)$ forms an orthonormal basis of $\mathbb{R}^n$. In what follows we should keep in mind that the vectors v_k depend on the direction of $y - x$, but in order to keep the notation simple we do not indicate that dependence explicitly. We now define a linear operator $\Phi(x, y) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ by defining it on the basis $(v_1, \dots, v_n)$ by

$$\Phi(x, y)v_k := \frac{f(x + \|y - x\|v_k) - f(x)}{\|y - x\|} \tag{6-3}$$

for $k = 1, \dots, n$. We claim that Φ is a slope function. By (6-3) and the definition of v_1,

$$\Phi(x, y)(y - x) = \|y - x\|\Phi(x, y)v_1 = f(y) - f(x).$$

To check continuity at x as a function of y, write $z \in \mathbb{R}^n$ in the form $z = \sum_{k=1}^n \alpha_k v_k$, where $\alpha_k := \langle v_k, z\rangle$. As the basis $(v_1, \dots, v_n)$ is orthonormal we have $\|z\|^2 = \sum_{k=1}^n |\alpha_k|^2$ and thus, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\|\Phi(x, y)z - Df(x)z\| \\
&\quad = \Big\| \sum_{k=1}^n \alpha_k \frac{f(x + \|y - x\|v_k) - f(x) - Df(x)\|y - x\|v_k}{\|y - x\|} \Big\| \\
&\quad \le \|z\| \sqrt{\sum_{k=1}^n \Big[\frac{\|f(x + \|y - x\|v_k) - f(x) - Df(x)\|y - x\|v_k\|}{\|y - x\|} \Big]^2} \\
&\quad \longrightarrow 0
\end{aligned}$$

as $y \to x$ by differentiability of f at x. We next show that $\Phi(x, y)$ is continuous as a function of x as $x \to y$. The trick is to rewrite $\Phi(x, y)$ with respect to the basis

$$(w_1, \dots, w_n) := (-v_1, v_2 - v_1, \dots, v_n - v_1).$$

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

As $x = y - (y - x) = y - \|x - y\|v_1 = y + \|x - y\|w_1$ we conclude that for $k = 2, \ldots, n$,

$$x + \|y - x\|v_k = y + \|x - y\|(v_k - v_1) = y + \|x - y\|w_k.$$

Hence, by using (6-3), we obtain for $k = 2, \ldots, n$,

$$\Phi(x, y)w_k = \Phi(x, y)v_k - \Phi(x, y)v_1 = \frac{f(y + \|x - y\|w_k) - f(y)}{\|x - y\|}.$$

Note that the final formula also applies to $k = 1$. Expressing z in terms of the basis $(w_1, \ldots, w_n)$, it turns out that

$$z = \sum_{k=1}^{n} \alpha_k v_k = \sum_{k=2}^{n} \alpha_k w_k - \Big(\sum_{k=1}^{n} \alpha_k\Big) w_1.$$

If we set $\beta_1 := -\sum_{k=1}^{n} \alpha_k$ and $\beta_k := \alpha_k$ for $k = 2, \ldots, n$, we see that

$$\sum_{k=1}^{n} |\beta_k|^2 \le \sum_{k=2}^{n} |\alpha_k|^2 + \Big(\sum_{k=1}^{n} |\alpha_k|\Big)^2 \le (1 + n)\|z\|^2.$$

Hence, applying the Cauchy–Schwarz inequality as before, we have

$$\begin{aligned}
&\|\Phi(x, y)z - Df(y)z\| \\
&\quad = \Big\| \sum_{k=1}^{n} \beta_k \frac{f(y + \|x - y\|w_k) - f(y) - Df(y)\|x - y\|w_k}{\|x - y\|} \Big\| \\
&\quad \le \sqrt{1 + n}\|z\| \sqrt{\sum_{k=1}^{n} \Big[\frac{\|f(y + \|x - y\|w_k) - f(y) - Df(y)\|x - y\|w_k\|}{\|x - y\|} \Big]^2} \\
&\quad \longrightarrow 0
\end{aligned}$$

as $x \to y$ by differentiability of f at y. We conclude that $\Phi(x, y)$ is separately continuous at every point (x, x).

REMARK 6.3. (a) If $n = 2$ there is a natural choice for (v_1, v_2). We choose v_2 to be the rotation of v_1 by $\pi/2$. More precisely, if $v_1 = (z_1, z_2)$ we let $v_2 = (-z_2, z_1)$. However, for $n > 2$ there is no such natural choice.

(b) The slope function $\Phi(x, y)$ constructed above is separately continuous at every point (x, x). One could ask whether or not it is possible to choose it to be continuous at every (x, y) with $x \neq y$. In our particular construction continuity is guaranteed if $(v_2, \ldots, v_n)$ is continuous as a function of $v_1 = (y - x)/\|y - x\|$. This is equivalent to finding $n - 1$ linearly independent solutions to the equation $\langle v_1, w \rangle = 0$ depending continuously on v_1. Sufficient conditions for that are established in [10], and explicit orthonormal bases are given for dimensions $n = 2$, 4 and 8. As shown in [2], these are the only possibilities! If $n \le 8$ we can construct a slope function Φ that is globally separately continuous if we artificially look at f as a function of eight variables by making it constant in $8 - n$ variables, and then restrict the constructed slope function to n variables just like a partial derivative; see Proposition 3.5. We do not claim that the construction of a globally separately continuous slope function is impossible for $n > 8$, but only that some other method is required if it can be done.

6.3. Lipschitz continuous functions. Let E, F be Banach spaces and $U \subseteq E$ open. Recall that a function $f\colon U \to F$ is called Lipschitz continuous if there exists $L > 0$ such that

$$\|f(x) - f(y)\|_F \leq L\|x - y\|_E$$

for all $x, y \in U$. We call L a Lipschitz constant of f.

Proposition 6.4 (Derivatives of Lipschitz functions). *Let E, F be Banach spaces and $U \subseteq E$ open. Assume that $f\colon U \to F$ is differentiable at $x \in U$. If f is Lipschitz continuous with Lipschitz constant L, then $\|Df(x)\|_{\mathcal{L}(E,F)} \leq L$.*

Proof. Assume that f is Lipschitz continuous with Lipschitz constant L. Let Φ be a slope function for f at x. Then, for $z \in E$, we have

$$\|\Phi(x + tz)tz\| = \|f(x + tz) - f(x)\| \leq L\|tz\|$$

whenever $t > 0$ is small enough. Dividing by t and then letting $t \to 0+$, we obtain

$$\|Df(x)z\| = \lim_{t\to 0+} \|\Phi(x + tz)z\| \leq L\|z\|$$

for all $z \in E$. By definition of the operator norm, $\|Df(x)\|_{\mathcal{L}(E,F)} \leq L$. □

Note that the converse is true when U is convex. Indeed, by the mean value inequality in Theorem 4.2, for every $x, y \in U$ there exists $c \in [\![x, y]\!]$ such that

$$\|f(y) - f(x)\|_F \leq \|Df(c)(y - x)\|_F \leq L\|y - x\|_E.$$

7. Application: differentiable dependence of fixed points

The aim of this section is to use our approach to derivatives to give a conceptually simple proof of the differentiable dependence of fixed points in the Banach fixed point theorem. The theorem is known; see, for instance, [14, Section 1.2.6] or [7, 13].

Let E, F be Banach spaces and let $U \subseteq E$ and $\Lambda \subseteq F$ be nonempty open sets. Let $f\colon \bar{U} \times \Lambda \to \bar{U}$ be a *uniform contraction* in $x \in U$. More precisely, assume that there exists $L \in (0, 1)$ such that

$$\|f(x, \lambda) - f(y, \lambda)\|_E \leq L\|x - y\|_E \tag{7-1}$$

for all $x, y \in \bar{U}$ and all $\lambda \in \Lambda$. By the Banach fixed point theorem, for every $\lambda \in \Lambda$ there exists a unique fixed point $x_\lambda \in \bar{U}$.

Proposition 7.1 (Continuous dependence of fixed points). *Assume that $f\colon \bar{U} \times \Lambda \to E$ satisfies (7-1) with $L < 1$. For every $\mu \in \Lambda$, let $x_\mu \in \bar{U}$ be the unique fixed point of $f(\cdot\,, \mu)$. If $\lambda \in \Lambda$ is such that $\mu \mapsto f(x_\lambda, \mu)$ is continuous at λ, then the map $\Lambda \to \bar{U}$, $\mu \mapsto x_\mu$ is continuous at λ.*

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

PROOF. Using the assumption that f is a uniform contraction, we have

$$\begin{aligned}\|x_\mu - x_\lambda\| &= \|f(x_\mu,\mu) - f(x_\lambda,\lambda)\|\\ &\le \|f(x_\mu,\mu) - f(x_\lambda,\mu)\| + \|f(x_\lambda,\mu) - f(x_\lambda,\lambda)\|\\ &\le L\|x_\mu - x_\lambda\| + \|f(x_\lambda,\mu) - f(x_\lambda,\lambda)\|.\end{aligned}$$

As $0 < L < 1$, by the continuity of $\mu \mapsto f(x_\lambda,\mu)$ at λ,

$$\|x_\mu - x_\lambda\| \le \frac{1}{1-L}\|f(x_\lambda,\mu) - f(x_\lambda,\lambda)\| \to 0$$

as $\mu \to \lambda$. □

We next show that the fixed points x_λ depend differentiably on λ. The reader is invited to compare our proof to a proof based on Fréchet derivatives given, for instance, in [14, Section 1.2.6]. By exploiting continuity properties of the slope function, we can avoid all ε–δ arguments and provide a conceptually cleaner proof.

THEOREM 7.2 (Differentiable dependence of fixed points). *Assume that a function $f \in C^1(\bar{U} \times \Lambda, E)$ satisfies (7-1) with $L < 1$. For every $\mu \in \Lambda$, let $x_\mu \in \bar{U}$ be the unique fixed point of $f(\cdot\,,\mu)$. Then the map $\Lambda \to \bar{U}$, $\mu \mapsto x_\mu$ is continuously differentiable.*

PROOF. The idea is to use algebraic manipulations to find a slope function for the fixed points. If Φ is a slope function for f and x_λ, x_μ are fixed points, then

$$\begin{aligned}x_\mu - x_\lambda &= f(x_\mu,\mu) - f(x_\lambda,\lambda)\\ &= f(x_\mu,\mu) - f(x_\lambda,\mu) + f(x_\lambda,\mu) - f(x_\lambda,\lambda)\\ &= \Phi((x_\lambda,\mu),(x_\mu,\mu))(x_\mu - x_\lambda, 0) + \Phi((x_\lambda,\lambda),(x_\lambda,\mu))(0,\mu-\lambda).\\ &= \Phi_1((x_\lambda,\mu),x_\mu)(x_\mu - x_\lambda) + \Phi_2((x_\lambda,\lambda),\mu)(\mu-\lambda),\end{aligned}$$

where Φ_1 and Φ_2 are the partial slope functions for the functions $x \mapsto f(x,\lambda)$ and $\lambda \mapsto f(x,\lambda)$ respectively, as introduced in Proposition 3.5. Rearranging, we see that

$$[I - \Phi_1((x_\lambda,\mu),x_\mu)](x_\mu - x_\lambda) = \Phi_2((x_\lambda,\lambda),\mu)(\mu-\lambda).$$

Since f is continuously differentiable on $\bar{U} \times \Lambda$, Theorem 4.1 allows us to choose Φ to be jointly continuous at $((\lambda,x_\lambda),(\lambda,x_\lambda))$. Hence, as $L \in (0,1)$ and $\mu \mapsto x_\mu$ is continuous, Proposition 6.4 implies the existence of $\delta > 0$ such that $\|\Phi_1((x_\lambda,\mu),x_\mu)\|_{\mathcal{L}(E)} < 1$ whenever $\|\lambda - \mu\| < \delta$. Thus $[I - \Phi_1((x_\lambda,\mu),x_\mu)]^{-1}$ exists by a Neumann series expansion; see, for instance, [20, Theorem IV.1.4]. Hence, if $\|\mu - \lambda\| < \delta$, then

$$x_\mu = x_\lambda + [I - \Phi_1((x_\lambda,\mu),x_\mu)]^{-1}\Phi_2((x_\lambda,\mu),\lambda)(\mu-\lambda).$$

Due to the joint continuity of Φ at $((\lambda,x_\lambda),(\lambda,x_\lambda))$ and the continuity of inversion, we conclude that $\mu \mapsto x_\mu$ is differentiable at λ with slope function given by

$$\Psi(\lambda,\mu)\gamma := [I - \Phi_1((x_\lambda,\mu),x_\mu)]^{-1}\Phi_2((x_\lambda,\mu),\lambda)\gamma$$

for all $\gamma \in F$ and derivative

$$\Psi(\lambda,\lambda) = [I - D_x f(x_\lambda,\lambda)]^{-1}D_\lambda f(x_\lambda,\lambda) \in \mathcal{L}(F,E).$$ □

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press

References

[1] E. G. Acosta and C. G. Delgado, 'Fréchet vs. Carathéodory', *Amer. Math. Monthly (4)* **101** (1994), 332–338.

[2] J. F. Adams, 'Vector fields on spheres', *Ann. of Math. (2)* **75** (1962), 603–632.

[3] H. Amann and J. Escher, *Analysis. II* (Birkhäuser, Basel, 2008).

[4] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, 4th edn (John Wiley & Sons, Hoboken, NJ, 2011).

[5] M. W. Botsko and R. A. Gosser, 'On the differentiability of functions of several variables', *Amer. Math. Monthly (9)* **92** (1985), 663–665.

[6] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext (Springer, New York, 2011).

[7] R. M. Brooks and K. Schmitt, *The Contraction Mapping Principle and Some Applications*, Electronic Journal of Differential Equations Monograph, 9 (Texas State University – San Marcos, Department of Mathematics, San Marcos, TX, 2009).

[8] R. C. Cabrales and M. A. Rojas-Medar, 'Sobre la diferenciabilidad de funciones en espacios de Banach', *Rev. Integr. Temas Mat. (2)* **24** (2006), 87–100.

[9] C. Carathéodory, *Funktionentheorie. Band I* (Birkhäuser, Basel, 1950).

[10] B. Eckmann, 'Stetige Lösungen linearer Gleichungssysteme', *Comment. Math. Helv.* **15** (1943), 318–339.

[11] M. Fréchet, 'Sur la notion de différentielle totale', *Nouv. Ann. (4)* **12** (1912), 385–403.

[12] E. Hairer and G. Wanner, *Analysis by its History* (Springer, New York, 2008).

[13] J. K. Hale, *Ordinary Differential Equations*, Pure and Applied Mathematics, XXI (Wiley-Interscience, New York, 1969).

[14] D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics, 840 (Springer, Berlin, 1981).

[15] S. Kuhn, 'The derivative à la Carathéodory', *Amer. Math. Monthly (1)* **98** (1991), 40–44.

[16] F. Martínez de la Rosa, *Cálculo Diferencial: Consideraciones Teóricas y Metodológicas* (Universidad de Cádiz, Servicio de Publicaciones, Cádiz, 1998).

[17] S. Pinzón and M. Paredes, 'La derivada de Carathéodory en $\mathbb{R}^2$', *Rev. Integr. Temas Mat. (2)* **17** (2003), 65–98.

[18] W. Rudin, *Principles of Mathematical Analysis*, 3rd edn, International Series in Pure and Applied Mathematics (McGraw-Hill, New York, 1976).

[19] O. Stolz, *Grundzüge der Differential- und Integralrechnung. Erster Teil: Reelle Veränderliche und Functionen* (B. G. Teubner, Leipzig, 1893).

[20] A. E. Taylor and D. C. Lay, *Introduction to Functional Analysis*, 2nd edn (John Wiley & Sons, New York, 1980).

SHANE ARORA, School of Mathematics & Statistics,
University of Sydney, NSW 2006, Australia
e-mail: saro0188@uni.sydney.edu.au

HAZEL BROWNE, School of Mathematics & Statistics,
University of Sydney, NSW 2006, Australia
e-mail: hbro4811@uni.sydney.edu.au

DANIEL DANERS, School of Mathematics & Statistics,
University of Sydney, NSW 2006, Australia
e-mail: daniel.daners@sydney.edu.au

https://doi.org/10.1017/S1446788720000166 Published online by Cambridge University Press