Chapter 2
Differentiation

In this chapter we will be primarily concerned with extending the derivative defined for real-valued functions defined on an interval of $\mathbb{R}$. We will also consider minima and maxima of real-valued functions defined on a normed vector space.

2.1 Directional Derivatives

Let O be an open subset of a normed vector space E, f a real-valued function defined on O, $a \in O$ and u a nonzero element of E. The function $f_u : t \longrightarrow f(a + tu)$ is defined on an open interval containing 0. If the derivative $\frac{\mathrm{d}f_u}{\mathrm{d}t}(0)$ is defined, i.e., if the limit

$$\lim_{t \longrightarrow 0} \frac{f(a+tu) - f(a)}{t}$$

exists, then we note this derivative $\partial_u f(a)$. It is called the derivative of f at a in the direction u. We refer to such derivatives as *directional derivatives*. Notice that, if $\partial_u f(a)$ is defined and $\lambda \in \mathbb{R}^*$, then $\partial_{\lambda u} f(a)$ is defined and

$$\partial_{\lambda u} f(a) = \lambda \partial_u f(a).$$

If $E = \mathbb{R}^n$ and (e_i) is its standard basis, then the directional derivative $\partial_{e_i} f(a)$ is called the *ith partial derivative* of f at a, or the derivative of f with respect to x_i at a. In this case we write $\partial_i f(a)$ or $\frac{\partial f}{\partial x_i}(a)$. If $a = (a_1, \ldots, a_n)$, then

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \to 0} \frac{f(a_1, \ldots, a_i + t, \ldots, a_n) - f(a_1, \ldots, a_n)}{t}.$$

If for every point $x \in O$, the partial derivative $\frac{\partial f}{\partial x_i}(x)$ is defined, then we obtain the function ith partial derivative $\frac{\partial f}{\partial x_i}$ defined on O. If these functions are defined and continuous for all i, then we say that the function f is of *class* C^1.

R. Coleman, *Calculus on Normed Vector Spaces*, Universitext,
DOI 10.1007/978-1-4614-3894-6_2, © Springer Science+Business Media New York 2012

Example. If f is the function defined on $\mathbb{R}^2$ by $f(x, y) = x\mathrm{e}^{xy}$, then the partial derivatives with respect to x and y are defined at all points $(x, y) \in \mathbb{R}^2$ and

$$\frac{\partial f}{\partial x}(x, y) = (1 + xy)\mathrm{e}^{xy} \qquad \text{and} \qquad \frac{\partial f}{\partial y}(x, y) = x^2\mathrm{e}^{xy}.$$

As the functions $(x, y) \longmapsto (1 + xy)\mathrm{e}^{xy}$ and $(x, y) \longmapsto x^2\mathrm{e}^{xy}$ are continuous, f is of class C^1.

Remark. If I is an open interval of $\mathbb{R}$, $a \in I$ and $f : I \longrightarrow \mathbb{R}$ has a derivative at a, then f is continuous at a. We have

$$\frac{f(a + t) - f(a)}{t} = \frac{\mathrm{d}f}{\mathrm{d}t}(a) + \epsilon(t),$$

where $\lim_{t\to 0} \epsilon(t) = 0$. This implies that

$$f(a + t) - f(a) = t\frac{\mathrm{d}f}{\mathrm{d}t}(a) + t\epsilon(t),$$

and the continuity of f at a follows. However, a function of two or more variables may have all its partial derivatives defined at a given point without being continuous there. Here is an example. Consider the function f defined on $\mathbb{R}^2$ by

$$f(x, y) = \begin{cases} \frac{x^6}{x^8+(y-x^2)^2} & \text{if} \quad (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}.$$

We have

$$\lim_{t\to 0} \frac{t^6}{t^8 + t^4}/t = 0 \qquad \text{and} \qquad \lim_{t\to 0} \frac{0}{t^2}/t = 0$$

and so

$$\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0.$$

However, $\lim_{x\to 0} f(x, x^2) = \infty$, which implies that f is not continuous at 0.

The next result needs no proof. It is simply an application of the definition of a partial derivative.

Proposition 2.1. *Let O be an open subset of $\mathbb{R}^n$, $a \in O$ and f and g real-valued functions defined on O having partial derivatives with respect to x_i at a. Then*

$$\frac{\partial(f + g)}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a) + \frac{\partial g}{\partial x_i}(a) \quad \textit{and} \quad \frac{\partial(fg)}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a)g(a) + f(a)\frac{\partial g}{\partial x_i}(a).$$

In addition, if $\lambda \in \mathbb{R}$ then

$$\frac{\partial(\lambda f)}{\partial x_i}(a) = \lambda \frac{\partial(f)}{\partial x_i}(a).$$

Suppose now that O is an open subset of $\mathbb{R}^n$ and f a mapping defined on O with image in $\mathbb{R}^m$. f has m coordinate mappings $f_1, \ldots, f_m$. If $a \in O$ and the partial derivatives $\frac{\partial f_i}{x_j}(a)$, for $1 \leq i \leq m$ and $1 \leq j \leq n$, are all defined, then the $m \times n$ matrix

$$J_f(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \ldots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

is called the *Jacobian matrix* of f at a.

Example. If the mapping f of $\mathbb{R}^3$ into $\mathbb{R}^2$ is defined by $f(x, y, z) = (xy, ze^{xy})$, then all partial derivatives are defined at any point $(x, y, z) \in \mathbb{R}^3$ and

$$J_f(x, y, z) = \begin{pmatrix} y & x & 0 \\ yze^{xy} & xze^{xy} & e^{xy} \end{pmatrix}.$$

It is easy to generalize the definition of class C^1 to a mapping having its image in $\mathbb{R}^m$. We say that such a function is of class C^1 if its coordinate mappings are all of class C^1.

Remark. We do not need to restrict the directional derivative to functions defined on open sets of a normed vector space. Let A be a nonempty subset of a vector space E, f a real-valued function defined on A, and $a \in A$. If u is a nonzero element of E and there exists $\epsilon > 0$ such that $a + tu \in A$, when $|t| < \epsilon$, then the function $f_u : t \longrightarrow f(a + tu)$ is defined on the open interval $(-\epsilon, \epsilon)$. If the derivative $\frac{\mathrm{d}f_u}{\mathrm{d}t}(0)$ is defined, then as above we note this derivative $\partial f_u(a)$ and call it the derivative of f at a in the direction u.

2.2 The Differential

Let E and F be normed vector spaces, O an open subset of E containing 0, and g a mapping from O into F such that $g(0) = 0$. If there exists a mapping ϵ, defined on a neighbourhood of $0 \in E$ and with image in F, such that $\lim_{h\to 0} \epsilon(h) = 0$ and

$$g(h) = \|h\|_E \epsilon(h),$$

then we write $g(h) = o(h)$ and say that g is a "small o of h". If $\|\cdot\|_E^\times \sim \|\cdot\|_E$ and $\|\cdot\|_F^\times \sim \|\cdot\|_F$, then $g(h) = o(h)$ for the norms $\|\cdot\|_E$ and $\|\cdot\|_F$ if and only if $g(h) = o(h)$ for the norms $\|\cdot\|_E^\times$ and $\|\cdot\|_F^\times$. In particular, if $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$, then the condition $g(h) = o(h)$ is independent of the norms we choose for the two spaces.

Let O be an open subset of a normed vector space E and f a mapping from O into a normed vector space F. If $a \in O$ and there is a continuous linear mapping ϕ from E into F such that

$$f(a+h) = f(a) + \phi(h) + o(h)$$

when h is close to 0, then we say that f is *differentiable* at a.

Proposition 2.2. *If f is differentiable at a, then*

(a) *f is continuous at a;*
(b) *ϕ is unique.*

Proof. (a) As ϕ is continuous at 0, $\lim_{h\to 0}\phi(h) = 0$ and so $\lim_{h\to 0} f(a+h) = f(a)$.

(b) Suppose that

$$f(a+h) = f(a) + \phi_1(h) + o(h) = f(a) + \phi_2(h) + o(h)$$

and let $x \in E$. For $t > 0$ small we have

$$f(a+tx) - f(a) = t\phi_1(x) + t\|x\|_E\epsilon_1(tx) = t\phi_2(x) + t\|x\|_E\epsilon_2(tx),$$

where $\lim_{t\to 0}\epsilon_i(tx) = 0$. This implies that

$$\phi_2(x) - \phi_1(x) = \|x\|_E(\epsilon_1(tx) - \epsilon_2(tx))$$

Letting t go to 0, we obtain $\phi_1(x) - \phi_2(x) = 0$. □

This unique continuous linear mapping ϕ is called the *differential* of f at a, written $f'(a)$, $\mathrm{d}f(a)$ or Df(a). If f is differentiable at every point $a \in O$, then we say that f is differentiable on O. If in addition f is a bijection onto an open subset U of F and the inverse mapping f^{-1} is also differentiable, then we say that f a *diffeomorphism*. Clearly a diffeomorphism is a homeomorphism.

Notation. We will use the notation f' for differentials. If considering the derivative of a real-valued function f defined on an open interval of $\mathbb{R}$, then we will use the notation $\frac{\mathrm{d}f}{\mathrm{d}t}$ or $\dot{f}$. To simplify the notation, we will usually write $f'(a)h$ for $f'(a)(h)$.

Examples. If E and F are normed vector spaces and $f : E \longrightarrow F$ is constant, then $f'(a)$ is the zero mapping at any point $a \in E$. If $f : E \longrightarrow F$ is linear and continuous, then $f'(a) = f$ at any point $a \in E$.

Exercise 2.1. Let E be a normed vector space and f a real-valued function defined on E such that $|f(x)| \leq \|x\|_E^2$. Show that f is differentiable at 0.

Proposition 2.3. *If we replace the norms on the spaces E and F by equivalent norms, then the differentiability at $a \in O$ and the differential are unaffected. In particular, if E and F are finite-dimensional, then we may choose any pair of norms.*

Proof. If f is differentiable at a and

$$g(h) = f(a+h) - f(a) - f'(a)h,$$

then $g(h) = o(h)$. If we replace one or both the norms of E and F by an equivalent norm, then with respect to the new pair of norms, we have $g(h) = o(h)$. It follows that f is differentiable with respect to the second pair of norms and that the differential at a is the same. □

Example. Let $A \in \mathcal{M}_n(\mathbb{R})$ be symmetric, $b \in \mathbb{R}^n$ and $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$f(x) = \frac{1}{2}x^t Ax - b^t x.$$

(Here, as elsewhere, when employing matrices, we identify elements of $\mathbb{R}^n$ with n-coordinate column vectors.) Let $a \in \mathbb{R}^n$. A simple calculation shows that

$$f(a+h) = f(a) + (a^t A - b^t)h + \frac{1}{2}h^t Ah.$$

The function $\phi : h \longmapsto (a^t A - b^t)h$ is linear. As $\mathbb{R}^n$ is finite-dimensional, ϕ is also continuous. We also have

$$|h^t Ah| \leq \|Ah\|_2 \|h\|_2 \leq |A|_2 \|h\|_2^2,$$

where $|\cdot|_2$ is the matrix norm subordinate to the norm $\|\cdot\|_2$. Hence $f'(a) = \phi$.

Exercise 2.2. Show that the mapping

$$f : \mathcal{M}_n(\mathbb{R}) \longrightarrow \mathcal{M}_n(\mathbb{R}), X \longmapsto X^t X$$

is differentiable at any point $A \in \mathcal{M}_n(\mathbb{R})$ and determine $f'(A)$.

Proposition 2.4. *Let f be a mapping defined on an open subset O of a normed vector space E with image in the cartesian product $F = F_1 \times \cdots \times F_p$. Then f is differentiable at $a \in O$ if and only if the coordinate mappings f_i, for $i = 1, \ldots, p$, are differentiable at a.*

Proof. Suppose first that the coordinate mappings are differentiable at a:

$$f_i(a+h) = f_i(a) + f_i'(a)h + \|h\|_E \epsilon_i(h),$$

where $\lim_{h\to 0} \epsilon_i(h) = 0$. The mapping

$$\phi : E \longrightarrow F, h \longmapsto (f_1'(a)h, \dots, f_p'(a)h)$$

is a continuous linear mapping. If we set $\epsilon(h) = (\epsilon_1(h), \dots, \epsilon_p(h))$, then $\lim_{h\to 0} \epsilon(h) = 0$ and

$$f(a + h) = f(a) + \phi(h) + \|h\|_E \epsilon(h).$$

Therefore f is differentiable at a.

Suppose now that f is differentiable at a:

$$f(a + h) = f(a) + f'(a)h + \|h\|_E \epsilon(h),$$

where $\lim_{h\to 0} \epsilon(h) = 0$. Then

$$f_i(a + h) = f_i(a) + L_i(h) + \|h\|_E \epsilon_i(h),$$

where $f'(a)h = (L_1(h), \dots, L_p(h))$. For each i the mapping L_i is linear and continuous and $\lim_{h\to 0} \epsilon_i(h) = 0$; hence f_i is differentiable at a. □

Remark. The differential $f'(a)$ is a continuous linear mapping from E into F. In the above proof we have shown that the coordinate mappings of $f'(a)$ at a are the differentials at a of the coordinate mappings of f, i.e.,

$$f'(a) = (f_1'(a), \dots, f_p'(a)).$$

Proposition 2.5. *If O is an open subset of a normed vector space E and $f : E \longrightarrow \mathbb{R}$ is differentiable at $a \in O$, then the directional derivative $\partial f_u(a)$ is defined for any nonzero vector $u \in E$ and $\partial f_u(a) = f'(a)u$. In particular, if $E = \mathbb{R}^n$, then the partial derivatives $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ are defined.*

Proof. We have

$$f(a + tu) = f(a) + tf'(a)u + o(tu)$$

and so

$$\lim_{t\to 0} \frac{f(a + tu) - f(a)}{t} = f'(a)u.$$

This ends the proof. □

Suppose that $\dim E = n < \infty$ and that (e_i) is a basis of E. If $x = \sum_{i=1}^n x_i e_i$, then

$$f'(a)x = \sum_{i=1}^n x_i f'(a)e_i = \sum_{i=1}^n \partial_{e_i} f(a) e_i^*(x),$$

where (e_i^*) is the dual basis of (e_i). We thus obtain the expression

$$f'(a) = \sum_{i=1}^{n} \partial_{e_i} f(a) e_i^*.$$

If $E = \mathbb{R}^n$ and (e_i) its standard basis, then we usually write $\mathrm{d}x_i$ for e_i^*. This gives us the expression

$$f'(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\mathrm{d}x_i.$$

What we have just seen has practical importance. If we wish to determine whether a real-valued function f defined on an open subset of $\mathbb{R}^n$ is differentiable at a given point a, then first we determine whether all its partial derivatives at a exist. If this is not the case, then f is not differentiable at a. If all the partial derivatives exist, then we know that the only possibility for $f'(a)$ is the linear function $\phi = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)\mathrm{d}x_i$. To conclude, we consider the expression

$$\frac{f(a+h) - f(a) - \phi(h)}{\|h\|} = \epsilon(h).$$

If $\lim_{h \to 0} \epsilon(h) = 0$, then f is differentiable at a, otherwise it is not.

Example. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by

$$f(x,y) = \begin{cases} \frac{x^3 - y^3}{x^2 + y^2} & \text{if} \quad (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}.$$

A simple calculation shows that

$$\frac{\partial f}{\partial x}(0,0) = 1 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(0,0) = -1.$$

Therefore, if $f'(0,0)$ exists, then $f'(0,0) = \mathrm{d}x - \mathrm{d}y$. However,

$$\left| \frac{h^3 - k^3}{h^2 + k^2} - (h-k) \right| / \|(h,k)\|_2 = \frac{|hk(h-k)|}{(h^2+k^2)^{\frac{3}{2}}}.$$

Setting $k = -h$ in the expression on the right-hand side of the equation, we obtain $\frac{1}{\sqrt{2}}$. Therefore we do not have the necessary convergence and so f is not differentiable at $(0,0)$.

The previous example shows that a function may have partial derivatives at a point without being differentiable at that point. This however is not the case for functions of a single variable.

Proposition 2.6. *Let I be an open interval of $\mathbb{R}$. Then $f : I \longrightarrow \mathbb{R}$ is differentiable at $a \in I$ if and only if f has a derivative at a.*

Proof. From Proposition 2.5, if f is differentiable at a, then f has a derivative at a.
Now suppose that $\frac{\mathrm{d}f}{\mathrm{d}x}(a)$ exists. Then

$$\frac{f(a+h) - f(a)}{h} - \frac{\mathrm{d}f}{\mathrm{d}x}(a) = \epsilon(h),$$

where $\lim_{h \to 0} \epsilon(h) = 0$. Multiplying by h we obtain

$$f(a+h) = f(a) + \frac{\mathrm{d}f}{\mathrm{d}x}(a)h + h\epsilon(h).$$

It follows that f is differentiable at a and that $f'(a)h = \frac{\mathrm{d}f}{\mathrm{d}x}(a)h$. □

If O is an open subset of $\mathbb{R}^n$ and $f : O \longrightarrow \mathbb{R}^m$ is differentiable at $a \in O$, then $f'(a)$ is a linear mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$. Let us write (e_j^n) (resp.(e_i^m)) for the standard basis of $\mathbb{R}^n$ (resp. $\mathbb{R}^m$). We have $f'(a) = (f_1'(a), \ldots, f_m'(a))$. However, $f_i'(a)e_j^n = \frac{\partial f_i}{\partial x_j}(a)$ and so

$$f'(a)e_j^n = \left(\frac{\partial f_1}{\partial x_j}(a), \ldots, \frac{\partial f_m}{\partial x_j}(a)\right) = \sum_{i=1}^m \frac{\partial f_i}{\partial x_j}(a)e_i^m.$$

Therefore the jth column of the matrix of $f'(a)$ with respect to the bases (e_j^n) and (e_i^m) has for elements $\frac{\partial f_1}{\partial x_j}(a), \ldots, \frac{\partial f_m}{\partial x_j}(a)$. It follows that this matrix is the Jacobian matrix $\mathrm{J}_f(a)$.

Notation. If E and F are finite-dimensional vector spaces, $\mathcal{B}_E$ and $\mathcal{B}_F$ bases of these spaces and $L : E \longrightarrow F$ a linear mapping, then we will write $\mathrm{mat}_{\mathcal{B}_E \mathcal{B}_F} L$ for the matrix of L with respect to the bases $\mathcal{B}_E$ and $\mathcal{B}_F$. So we could write

$$\mathrm{mat}_{(e_i^n)(e_i^m)} f'(a) = \mathrm{J}_f(a).$$

The proof of the next result is elementary.

Proposition 2.7. *Let E and F be normed vector spaces, O an open subset of E and a an element of O. If f and g are differentiable at a, then $f + g$ is differentiable at a, as is λf, for any $\lambda \in \mathbb{R}$, and*

$$(f+g)'(a) = f'(a) + g'(a) \qquad \text{and} \qquad (\lambda f)'(a) = \lambda(f'(a)).$$

If F is a commutative normed algebra, then fg is differentiable at a and

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

Suppose that E and F are normed vector spaces, O an open subset of E and $f : O \longrightarrow F$ differentiable at $a \in O$. If $\tilde{F}$ is a vector subspace of F and the image of f lies in $\tilde{F}$, then f is differentiable at a as a mapping from O into $\tilde{F}$ if the image of $f'(a)$ lies in $\tilde{F}$. The following result gives us a sufficient condition for this to be so.

Proposition 2.8. *If $\tilde{F}$ is closed, then f is differentiable at a as a mapping from O into $\tilde{F}$.*

Proof. For any $h \in E$

$$\lim_{t \to 0} \frac{f(a+th) - f(a)}{t} = f'(a)h.$$

Let (t_n) be a sequence in $\mathbb{R}^*$ with limit 0 and such that $a + t_n h \in O$. If we set $u_n = \frac{f(a+t_n h) - f(a)}{t_n}$, then the sequence (u_n) is a convergent sequence contained in $\tilde{F}$. As $\tilde{F}$ is closed, its limit $f'(a)h$ is an element this subspace. This ends the proof. □

2.3 Differentials of Compositions

In this section we consider the differentiability of mappings which are compositions. Let E, F and G be normed vector spaces, O an open subset of E, U an open subset of F and $f : O \longrightarrow F$, $g : U \longrightarrow G$ be such that $f(O) \subset U$. Then the mapping $g \circ f$ is defined on O.

Theorem 2.1. *If f is differentiable at a and g is differentiable at $f(a)$, then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

Proof. To simplify the notation let us write $b = f(a)$. We have

$$f(a+h) = f(a) + f'(a)h + \|h\|_E \epsilon_1(h)$$

and

$$g(b+k) = g(b) + g'(b)k + \|k\|_F \epsilon_2(k),$$

with $\lim_{h \to 0} \epsilon_1(h) = 0$ and $\lim_{k \to 0} \epsilon_2(k) = 0$. For h sufficiently small we may write

$$\begin{aligned} g(f(a+h)) &= g(b + f'(a)h + \|h\|_E \epsilon_1(h)) \\ &= g(b) + g'(b)(f'(a)h + \|h\|_E \epsilon_1(h)) \\ &\quad + \|f'(a)h + \|h\|_E \epsilon_1(h)\|_F \epsilon_2\big(f'(a)h + \|h\|_E \epsilon_1(h)\big) \\ &= g(b) + g'(b) \circ f'(a)h + \|h\|_E \tilde{\epsilon}(h), \end{aligned}$$

where

$$\tilde{\epsilon}(h) = g'(b)\epsilon_1(h) + \| f'(a)\frac{h}{\|h\|_E} + \epsilon_1(h)\|_F \epsilon_2\big(f'(a)h + \|h\|_E \epsilon_1(h)\big).$$

To finish, we only need to show that $\lim_{h\to 0} \tilde{\epsilon}(h) = 0$. However,

- $\lim_{h\to 0} \epsilon_1(h) = 0 \Longrightarrow \lim_{h\to 0} g'(b)\epsilon_1(h)$, because $g'(b)$ is continuous;
- $\| f'(a)\frac{h}{\|h\|_E} + \epsilon_1(h)\|_F$ is bounded for small values of h, because $f'(a)$ is continuous and so bounded on the unit sphere;
- $\lim_{h\to 0} \epsilon_2(f'(a)h + \|h\|_E\epsilon_1(h)) = 0$, because $f'(a)$ is continuous.

Therefore $\lim_{h\to 0} \tilde{\epsilon}(h) = 0$. □

Corollary 2.1. *If in the above theorem the normed vector spaces are euclidian spaces, then*

$$J_{g\circ f}(a) = J_g(f(a))J_f(a).$$

Proof. If $E = \mathbb{R}^n$, $F = \mathbb{R}^m$, $G = \mathbb{R}^s$ and $\mathcal{B}^n$, $\mathcal{B}^m$, $\mathcal{B}^s$ their standard bases, then we have

$$\begin{aligned} \mathrm{J}_{g\circ f}(a) &= \mathrm{mat}_{\mathcal{B}^n\mathcal{B}^s}(g\circ f)'(a) \\ &= \mathrm{mat}_{\mathcal{B}^n\mathcal{B}^s} g'(f(a)) \circ f'(a) \\ &= \mathrm{mat}_{\mathcal{B}^m\mathcal{B}^s} g'(f(a))\mathrm{mat}_{\mathcal{B}^n\mathcal{B}^m} f'(a) \\ &= \mathrm{J}_g(f(a))\mathrm{J}_f(a), \end{aligned}$$

which is the result we were looking for. □

Remark. The expression

$$(g\circ f)'(a) = g'(f(a)) \circ f'(a).$$

is often referred to as the *chain rule*.

Example. Let $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ and $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by

$$f(x,y,z) = (xy, \mathrm{e}^{xz}) \qquad \text{and} \qquad g(u,v) = u^2 v.$$

Then

$$\mathrm{J}_f(x,y,z) = \begin{pmatrix} y & x & 0 \\ z\mathrm{e}^{xz} & 0 & x\mathrm{e}^{xz} \end{pmatrix} \qquad \text{and} \qquad \mathrm{J}_g(u,v) = \begin{pmatrix} 2uv & u^2 \end{pmatrix}.$$

Setting $u = xy$ and $v = \mathrm{e}^{xz}$ in the second matrix, we obtain

$$\mathrm{J}_g(f(x,y,z)) = \begin{pmatrix} 2xy\mathrm{e}^{xz} & x^2y^2 \end{pmatrix}.$$

Multiplying the matrices $\mathrm{J}_g(f(x,y,z))$ and $\mathrm{J}_f(x,y,z)$, we obtain

$$\mathrm{J}_{g\circ f}(x,y,z) = \left((2xy^2 + x^2y^2z)\mathrm{e}^{xz} \; 2x^2y\mathrm{e}^{xz} \; x^3y^2\mathrm{e}^{xz} \right).$$

Exercise 2.3. Let I be an open interval of $\mathbb{R}$, O an open subset of $\mathbb{R}^m$, f a mapping of I into $\mathbb{R}^m$ and g a real-valued function defined on O. We suppose that $f(I) \subset O$. If $a \in I$ and $f'(a)$ and $g'(f(a))$ exist, show that

$$\frac{\mathrm{d}}{\mathrm{d}t} g \circ f(t) = \sum_{i=1}^{m} \frac{\partial g}{\partial x_i}(f(t)) \frac{\mathrm{d} f_i}{\mathrm{d}t}(t).$$

For the functions

$$f : \mathbb{R} \longrightarrow \mathbb{R}^2, t \longmapsto (t^2, t^3) \qquad \text{and} \qquad g : \mathbb{R}^2 \longrightarrow \mathbb{R}, (x,y) \longmapsto xy$$

calculate $\frac{\mathrm{d}}{\mathrm{d}t} g \circ f(t)$ using the formula. Find an expression for the function $g \circ f$ and then confirm the result.

Exercise 2.4. Let E and F be normed vector spaces, O an open subset of E, U an open subset of F and $f : O \longrightarrow U$ a diffeomorphism. Show that, for any point $x \in O$, $f'(x)$ is a normed vector space isomorphism from E into F and that $(f^{-1})'(f(x)) = f'(x)^{-1}$.

2.4 Mappings of Class C^1

If $O \subset \mathbb{R}^n$ is open and f a real-valued function whose partial derivatives are defined and continuous on O, then we say that f is of class C^1. More generally, if $f : O \longrightarrow \mathbb{R}^m$ is such that the mn functions $\frac{\partial f_i}{\partial x_j}$ are defined and continuous on O, then we say that f is of *class* C^1. In this section we will obtain a result giving necessary and sufficient conditions for a mapping to be of class C^1. This will enable us to generalize this notion to mappings between any pair of normed vector spaces.

Consider a real-valued function f defined on an open subset O of $\mathbb{R}^n$. We have seen that if f has partial derivatives at a point $a \in O$, then this does not imply that f is differentiable at a. However, if we add a condition, then we do obtain differentiability.

Theorem 2.2. *If the functions $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ are defined on a neighbourhood V of a and continuous at a, then f is differentiable at a.*

Proof. As V is a neighbourhood of a, there is an open cube

$$C(a,\epsilon) = \{x \in \mathbb{R}^n : |x_i - a_i| < \epsilon\}$$

contained in V. If h lies in the open cube $C(0,\epsilon)$, then $a+h \in C(a,\epsilon)$ and

$$\begin{aligned} f(a+h)-f(a) &= f(a_1+h_1,\dots,a_n+h_n)-f(a_1,\dots,a_n) \\ &= f(a_1+h_1,\dots,a_n+h_n)-f(a_1+h_1,\dots,a_{n-1}+h_{n-1},a_n) \\ &\quad +f(a_1+h_1,\dots,a_{n-1}+h_{n-1},a_n) \\ &\quad -f(a_1+h_1,\dots,a_{n-2}+h_{n-2},a_{n-1},a_n) \\ &\qquad\qquad \vdots \\ &\quad +f(a_1+h_1,a_2,\dots,a_n)-f(a_1,\dots,a_n). \end{aligned}$$

Suppose that $h_n \neq 0$ and consider the function

$$g_n : [0,h_n] \longrightarrow \mathbb{R}, t \longmapsto f(a_1+h_1,\dots,a_{n-1}+h_{n-1},a_n+t).$$

From the mean value theorem there exists $\theta \in (0,1)$ such that

$$g_n(h_n)-g_n(0)=\frac{\mathrm{d}g_n}{\mathrm{d}t}(\theta h_n)h_n.$$

It follows that there exists $y^n \in C(a,\epsilon)$ such that

$$f(a_1+h_1,\dots,a_n+h_n)-f(a_1+h_1,\dots,a_{n-1}+h_{n-1},a_n)=\frac{\partial f}{\partial x_n}(y^n)h_n.$$

(If $h_n=0$, then we may take $y^n=a$.) We proceed in the same way for each line on the right-hand side of the above expression to obtain $y^1,\dots,y^n \in C(a,\epsilon)$ such that

$$f(a+h)-f(a)=\sum_{i=1}^n h_i\frac{\partial f}{\partial x_i}(y^i).$$

Therefore

$$f(a+h)-f(a)-\sum_{i=1}^n h_i\frac{\partial f}{\partial x_i}(a)=\sum_{i=1}^n h_i\left(\frac{\partial f}{\partial x_i}(y^i)-\frac{\partial f}{\partial x_i}(a)\right).$$

Now,

$$\left|\sum_{i=1}^n h_i\left(\frac{\partial f}{\partial x_i}(y^i)-\frac{\partial f}{\partial x_i}(a)\right)\right| \leq \|h\|_\infty \sum_{i=1}^n \left|\frac{\partial f}{\partial x_i}(y^i)-\frac{\partial f}{\partial x_i}(a)\right|$$

and $\lim_{h\to 0} y^i = a$ for each i. As the functions $\frac{\partial f}{\partial x_i}$ are continuous at a, we have

$$\lim_{h\to 0}\sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}(y^i) - \frac{\partial f}{\partial x_i}(a)\right| = 0.$$

It follows that

$$f(a+h) - f(a) - \sum_{i=1}^{n} h_i\frac{\partial f}{\partial x_i}(a) = o(h)$$

and so f is differentiable at a. □

Corollary 2.2. *Let O be an open subset of $\mathbb{R}^n$ and $f : O \longrightarrow \mathbb{R}^m$ such that the functions $\frac{\partial f_i}{\partial x_j}$, for $1 \le i \le m$ and $1 \le j \le n$, are defined on a neighbourhood V of $a \in O$ and continuous at a. Then f is differentiable at a.*

Proof. From the theorem above each coordinate function is differentiable at a and so f is differentiable at a. □

Examples. 1. If $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is defined by

$$f(r,\theta) = (r\cos\theta, r\sin\theta),$$

then we have

$$\frac{\partial f_1}{\partial r} = \cos\theta, \quad \frac{\partial f_1}{\partial\theta} = -r\sin\theta, \quad \frac{\partial f_2}{\partial r} = \sin\theta \quad \text{and} \quad \frac{\partial f_2}{\partial\theta} = r\cos\theta.$$

The four partial derivatives are clearly continuous; therefore f is differentiable at any point (r,θ) and

$$f'(r,\theta) = (\cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta, \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta).$$

2. If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is defined by

$$f(x_1,\dots,x_n) = x_1x_2\cdots x_n,$$

then

$$\frac{\partial f}{\partial x_i}(x) = x_1\cdots x_{i-1}\hat{x}_i x_{i+1}\cdots x_n,$$

where the "hat" indicates that the variable is absent. The n partial derivatives are clearly continuous; hence f is differentiable at any point x and

$$f'(x) = \sum_{i=1}^{n} x_1\cdots x_{i-1}\hat{x}_i x_{i+1}\cdots x_n\,\mathrm{d}x_i.$$

It is relatively easy to extend Theorem 2.2 to general finite-dimensional normed vector spaces. Let E be an n-dimensional normed vector space, (v_i) a basis of E, (e_i) the standard basis of $\mathbb{R}^n$ and L the linear mapping which sends v_i to e_i. Consider a real-valued function f defined on an open subset O of E and suppose that the directional derivatives $\partial_{v_i} f(x)$ are defined and continuous on a neighbourhood V of a point $a \in O$. If we set $\tilde{f} = f \circ L^{-1}$, then $\tilde{f}$ is defined on the open subset $L(O)$ of $\mathbb{R}^n$ and $L(V)$ is a neighbourhood of $L(a)$. A simple calculation shows that $\frac{\partial \tilde{f}}{\partial x_i}(L(x)) = \partial_{v_i} f(x)$ and it follows that $\tilde{f}$ is differentiable at $L(a)$. As L is differentiable at a and $f = \tilde{f} \circ L$, f is differentiable at a and

$$f'(a) = \tilde{f}'(L(a)) \circ L'(a) = \tilde{f}'(L(a)) \circ L.$$

Example. Suppose that $n \geq 1$ and let us consider the mapping $\det : \mathcal{M}_n(\mathbb{R}) \longrightarrow \mathbb{R}$, where $\det(X)$ is the determinant of X. We will write $E(i,j)$, with $1 \leq i, j \leq n$, for the elements of the standard basis of $\mathcal{M}_n(\mathbb{R})$, i.e.,

$$E(i,j)_{k,l} = \begin{cases} 1 & \text{if} \quad i = k, j = l \\ 0 & \text{otherwise} \end{cases}.$$

For $X \in \mathcal{M}_n(\mathbb{R})$ we have

$$\det(X + tE(i,j)) - \det(X) = t\gamma_{ij}(X),$$

where $\gamma_{ij}(X)$ is the (i,j)-cofactor of X. It follows that

$$\partial_{E(i,j)} \det X = \gamma_{ij}(X).$$

As $\gamma_{ij}(X)$ is a polynomial in the entries X_{ij} of X, $\partial_{E(i,j)} X$ is a continuous function of X and so the mapping det is differentiable at any point $X \in \mathcal{M}_n(\mathbb{R})$.

Proposition 2.9. *Let O be an open subset of $\mathbb{R}^n$, f and g real-valued functions of class C^1 defined on O and $\lambda \in \mathbb{R}$. Then $f + g$, λf and fg are of class C^1. If g does not vanish on O, then $\frac{f}{g}$ is of class C^1.*

Proof. We have the relations

$$\frac{\partial(f+g)}{\partial x_i}(x) = \frac{\partial f}{\partial x_i}(x) + \frac{\partial g}{\partial x_i}(x) \qquad \text{and} \qquad \frac{\partial(\lambda g)}{\partial x_i}(x) = \lambda \frac{\partial f}{\partial x_i}(x).$$

Also,

$$\frac{\partial(fg)}{\partial x_i}(x) = \frac{\partial f}{\partial x_i}(x) g(x) + f(x) \frac{\partial g}{\partial x_i}(x).$$

As f and g are of class C^1, the functions f, g, $\frac{\partial f}{\partial x_i}$ and $\frac{\partial g}{\partial x_i}$ are continuous; hence the right-hand sides of the above expressions are continuous and so the functions $f + g$, λf and fg are of class C^1. If g does not vanish and we set $h = \frac{f}{g}$, then

$$\frac{\partial h}{\partial x_i}(x) = \frac{g(x)\frac{\partial f}{\partial x_i}(x) - f(x)\frac{\partial g}{\partial x_i}(x)}{(g(x))^2},$$

and so $\frac{\partial h}{\partial x_i}$ is continuous. Thus h is of class C^1. □

Corollary 2.3. *Let O be an open subset of $\mathbb{R}^n$, f and g mappings of class C^1 defined on O with image in $\mathbb{R}^m$ and $\lambda \in \mathbb{R}$. Then $f + g$ and λf are of class C^1.*

Proof. As f and g are of class C^1, so are their coordinate mappings. Now

$$(f+g)_i = f_i + g_i \qquad \text{and} \qquad (\lambda f)_i = \lambda f_i$$

and so the coordinate mappings of $f + g$ and λf are of class C^1. It follows that $f + g$ and λf are C^1-mappings. □

Remark. From the above proposition and corollary, we see that the C^1-mappings defined on O form a vector space if the image space is $\mathbb{R}^m$ and, in the case where $m = 1$, this is an algebra.

Proposition 2.10. *Let O be an open subset of $\mathbb{R}^n$, U an open subset of $\mathbb{R}^m$, f a mapping from O into $\mathbb{R}^m$ and g a mapping from U into $\mathbb{R}^s$. If $f(O) \subset U$ and f and g are of class C^1, then $g \circ f$ is of class C^1.*

Proof. For $x \in O$ we have

$$\mathrm{J}_{g\circ f}(x) = \mathrm{J}_g(f(x)) \circ \mathrm{J}_f(x),$$

therefore

$$\frac{\partial (g \circ f)_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(f(x))\frac{\partial f_k}{\partial x_j}(x).$$

As f and the partial derivatives $\frac{\partial g_i}{\partial y_k}$ and $\frac{\partial f_k}{\partial x_j}$ are continuous, the partial derivative $\frac{\partial (g\circ f)_i}{\partial x_j}$ is continuous. It follows that $g \circ f$ is of class C^1. □

We now prove the characterization of C^1-mappings referred to at the beginning of the section.

Theorem 2.3. *Let O be an open subset of $\mathbb{R}^n$ and f a mapping from O into $\mathbb{R}^m$. Then f is of class C^1 if and only if f is differentiable on O and the mapping f' from O into $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.*

Proof. Let us set $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ and fix norms on these spaces. Suppose that f is of class C^1. From Theorem 2.2, we know that f is differentiable on O. Also, for $x \in O$ and $u \in \mathbb{R}^n$ we have

$$f'(x)u = \mathrm{J}_f(x)u.$$

Now

$$|f'(x)|_{\mathcal{L}(E,F)} = |\mathrm{J}_f(x)|,$$

where $|\cdot|$ is the norm defined on $\mathcal{M}_{mn}(\mathbb{R})$ by

$$|A| = \sup_{\|x\|_E \le 1} \|Ax\|_F.$$

If we let

$$|A|_M = \max |a_{ij}|,$$

where $A = (a_{ij})$, then $|\cdot|_M$ also defines a norm on $\mathcal{M}_{mn}(\mathbb{R})$. As f is of class C^1, we have

$$\lim_{h \to 0} |\mathrm{J}_f(x+h) - \mathrm{J}_f(x)|_M = 0.$$

However, the norm $|\cdot|_M$ is equivalent to the norm $|\cdot|$ and so

$$\lim_{h \to 0} |\mathrm{J}_f(x+h) - \mathrm{J}_f(x)| = 0.$$

It follows that f' is continuous at x.

Now suppose that f' is defined and continuous on O. From Proposition 2.5 we know that the partial derivatives of f are defined on O. Also, for $x \in O$ and $u \in \mathbb{R}^n$ we have

$$f'(x)u = \mathrm{J}_f(x)u.$$

As f' is continuous, we have

$$\lim_{h \to 0} |\mathrm{J}_f(x+h) - \mathrm{J}_f(x)| = 0 \Longrightarrow \lim_{h \to 0} |\mathrm{J}_f(x+h) - \mathrm{J}_f(x)|_M = 0,$$

which implies that the partial derivatives are continuous at x. □

The preceding theorem suggests the following generalization of the notion of class C^1. If E and F are normed vector spaces and O is open in E, then $f : O \longrightarrow F$ is of *class* C^1 if f is differentiable at every point $x \in O$ and the mapping $f' : O \longrightarrow \mathcal{L}(E, F)$ is continuous.

In closing this section we mention that the exponential mapping is of class C^1. This can be shown directly. However, the calculations are rather long. Later on we will prove this result in a simple way, avoiding arduous calculations.

2.5 Extrema

Let us recall the notions of minimum and maximum. Suppose that f is a real-valued function defined on a set X. We say that $a \in X$ is a *(global) minimum* if, for all $x \in X$,

$$f(a) \le f(x);$$

$a \in X$ is *(global) maximum* if, for all $x \in X$,

$$f(a) \geq f(x).$$

If the inequality is strict when $x \neq a$, then we speak of a *strict minimum* or *strict maximum*. A point which is either a minimum or a maximum is called an *extremum*. A function may have no minimum, a single minimum or several minima. The same is true for maxima.

Examples. 1. $f : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto x$ has neither a minimum nor a maximum.
2. $f : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto x^2$ has a minimum but no maximum.
3. $f : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto -x^2$ has a maximum but no minimum.
4. $f : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto \cos x$ has an infinite number of minima and maxima.

Let us now turn to normed vector spaces. We have already seen that if E is a normed vector space, $X \subset E$ compact and $f : X \longrightarrow \mathbb{R}$ continuous, then f has a minimum and a maximum (Theorem 1.3). This result is one of existence: it does not tell us how to find the extrema. We now introduce a related notion which can often help us in this direction. Let X be a subset of a normed vector space E and f a real-valued function defined on X. We say that $a \in X$ is a *local minimum* if a has a neighbourhood N such that

$$f(a) \leq f(x)$$

for all $x \in N \cap X$. We define a *local maximum* in an analogous way (reversing the direction of the inequality). As above, if the inequality is strict when $x \neq a$, then we speak of a strict local minimum or maximum. A point which is either a local minimum or maximum is called a local extremum . Clearly a global minimum (resp. maximum) is a local minimum (resp. maximum); however, the converse is not true. As a first step in looking for an extremum, it can be useful to look for local extrema. We will present a fundamental result which helps us to do so.

Let a and b be elements of a vector space E. We call the set

$$[a, b] = \{x \in E : x = \lambda a + (1 - \lambda)b \ , \lambda \in [0, 1]\}$$

the *segment* joining a to b. We write (a, b) for $[a, b] - \{a, b\}$. If $X \subset E$ is such that the segment $[a, b]$ always lies in E when $a, b \in E$, then we say that X is *convex*.

Exercise 2.5. Show that segments and affine subspaces are convex and that in a normed vector space closed and open balls are convex.

Theorem 2.4. *Let E be a normed vector space, $O \subset E$ open and $X \subset O$ convex and suppose that f is a real-valued function defined on O. If f restricted to X has a local minimum at x and f is differentiable at x, then*

$$f'(x)(y - x) \geq 0$$

for all $y \in X$.

Proof. Let $\lambda \in (0,1)$. As X is convex, $x + \lambda(y-x) \in X$ and for λ sufficiently small we have

$$\begin{aligned} f(x+\lambda(y-x)) - f(x) &= f'(x)(\lambda(y-x)) + o(\lambda(y-x)) \\ &= \lambda f'(x)(y-x) + |\lambda|\|y-x\|\epsilon(\lambda(y-x)), \end{aligned}$$

where $\lim_{h\to 0}\epsilon(h) = 0$. Dividing by λ, we obtain

$$f'(x)(y-x) + \frac{|\lambda|}{\lambda}\|y-x\|\epsilon(\lambda(y-x)) \geq 0,$$

because x is a local minimum. Letting λ go to 0, we obtain the result. □

Remark. The above inequality is called *Euler's inequality* . If x is a maximum, then x is a minimum of the function $-f$ restricted to X and so we obtain the inequality in the opposite direction.

Corollary 2.4. *Let O be an open subset of a normed vector space E and X an affine subspace of E: $X = a + V$, where V is a vector subspace of E and $a \in X$. We suppose that $O \cap X \neq \emptyset$. If f is a real-valued function defined on O such that f restricted to X has a local extremum at x and f is differentiable at x, then*

$$f'(x)v = 0$$

for all $v \in V$.

Proof. We will prove the result for a local minimum; the other case can be proved by considering $-f$. Suppose that $f'(x)v \neq 0$ for some $v \in V$. Replacing v by $-v$ if necessary, we may suppose that $f'(x)v > 0$. Let B be an open ball centred on x and lying in O. If we set $X_1 = B \cap X$, then X_1 is a convex subset of O. There exists $\lambda > 0$ such that $y = x - \lambda v \in X_1$ and

$$f'(x)(y-x) = -\lambda f'(x)v \geq 0 \Longrightarrow f'(x)v \leq 0,$$

a contradiction. Hence $f'(x)v = 0$ for all $v \in V$. □

Remarks. 1. A special case of the above result is when the function f is defined on E: if $f_{|X}$ has a local extremum at x, then

$$f'(x)v = 0$$

for all $v \in V$.

2. If $E = \mathbb{R}^n$, then we have

$$\nabla f(x)v = 0$$

for all $v \in V$, i.e., $\nabla f(x) \in V^{\perp}$.

Corollary 2.5. *Suppose that O is an open subset of a normed vector space E. If $f : O \longrightarrow \mathbb{R}$ has a local extremum at $x \in O$ and f is differentiable at x, then $f'(x) = 0$.*

Proof. This result follows immediately from Corollary 2.4: it is sufficient to set $X = E$. □

Suppose that O is an open subset of a normed vector space E and that the mapping $f : O \longrightarrow \mathbb{R}$ is differentiable at a point $x \in O$. If $f'(x) = 0$, then we say that x is a *critical point* of f. From what we have just seen, if x is a local extremum, then x is a critical point; however, the converse is false. For example, 0 is a critical point of the mapping $f : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto x^3$, but 0 is not a local extremum.

Example. Consider the function

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}, (x, y) \longmapsto x^3 y^2 (1 - x - y)$$

defined on the set

$$X = \left\{(x, y \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 1\right\}.$$

As X is closed and bounded, X is compact. Therefore f has a minimum and a maximum on X. Clearly $f(x, y) \geq 0$ and $f(x, y) = 0$ if and only if $x = 0, y = 0$ or $x + y = 1$, i.e., if and only if (x, y) lies on the boundary of X. Hence the minima of f are those points lying on the boundary of X. As there are points (x, y) such that $f(x, y) > 0$, any maximum (x, y) of f must lie in the interior O of X, an open set. Such a point is a maximum of the function restricted to O and so a critical point of this function. We have

$$\frac{\partial f}{\partial x}(x, y) = x^2 y^2 (3 - 4x - 3y) \qquad \text{and} \qquad \frac{\partial f}{\partial y}(x, y) = x^3 y (2 - 2x - 3y).$$

Setting the partial derivatives equal to 0, we find a unique solution to these equations in the interior of X, namely the point $A = (\frac{1}{2}, \frac{1}{3})$, so this must be the unique maximum of f. (If we had found more than one critical point, then it would have been necessary to calculate the value of f at each one of them and then choose the point(s) giving the maximum value.)

Exercise 2.6. (An extension of Rolle's Theorem) Let E be a normed vector space, X a compact subset, whose interior is not empty, and f a real-valued continuous function defined on X, which is differentiable on the interior of X. Show that if f is constant on the boundary of X, then f has a critical point in the interior of X.

2.6 Differentiability of the Norm

If E is a normed vector space with norm $\|\cdot\|$, then $\|\cdot\|$ is itself a mapping from E into $\mathbb{R}$ and we may consider its differentiability. We will write $\|x\|'$ for the differential of the norm at x (if it exists). We should first notice that the norm is never differentiable at the origin. Suppose that $\|0\|'$ exists. Then for small nonzero values of h, we have

$$\|h\| = \|0\|'h + o(h) \Longrightarrow \lim_{h\to 0}\left(1 - \|0\|'\frac{h}{\|h\|}\right) = 0$$

and

$$\|h\| = \|-h\| = -\|0\|'h + o(h) \Longrightarrow \lim_{h\to 0}\left(1 + \|0\|'\frac{h}{\|h\|}\right) = 0.$$

Summing the two limits we obtain $2 = 0$, which is clearly a contradiction. Hence $\|0\|'$ does not exist.

The norm may or may not be differentiable at a point other than the origin. For example, the norm $\|\cdot\|_p$ defined on $\mathbb{R}^n$ is differentiable at all points other than the origin if $p \in (1, \infty)$; however, this is not the case if $p = 1$ or $p = \infty$. We will study these norms more in detail at the end of the section. At points where the differential exists, we have the following interesting result:

Proposition 2.11. *Let E be a normed vector space and $\|\cdot\|$ its norm. If $\|\cdot\|$ is differentiable at $a \neq 0$ and $\lambda > 0$, then $\|\cdot\|$ is differentiable at λa and $\|\lambda a\|' = \|a\|'$. In addition, $|\|a\|'|_{E^*} = 1$.*

Proof. If $\|\cdot\|$ is differentiable at a, $\lambda > 0$ and $h \in E \setminus \{0\}$ small, then we have

$$\begin{aligned}\|\lambda a + h\| &= \lambda\|a + \frac{h}{\lambda}\| \\ &= \lambda\left(\|a\| + \|a\|'\frac{h}{\lambda} + o\left(\frac{h}{\lambda}\right)\right) \\ &= \|\lambda a\| + \|a\|'h + o(h).\end{aligned}$$

It follows that $\|\lambda a\|'$ exists and $\|\lambda a\|' = \|a\|'$.

Now let us show that $|\|a\|'|_{E^*} = 1$. Consider the function

$$f : \mathbb{R}_+^* \longrightarrow \mathbb{R}, \lambda \longmapsto \|\lambda a\|.$$

For a given $\lambda \in \mathbb{R}_+^*$ and $h \in \mathbb{R}$ sufficiently small, we have

$$\|(\lambda + h)a\| = (\lambda + h)\|a\|$$

and so

$$\lim_{h\to 0}\frac{\|(\lambda + h)a\| - \|\lambda a\|}{h} = \lim_{h\to 0}\frac{h\|a\|}{h} = \|a\|.$$

Therefore $\dot{f}(\lambda) = \|a\|$ for all values of λ. On the other hand, $f = \|\cdot\| \circ \phi$, where $\phi(\lambda) = \lambda a$, and so

$$f'(\lambda)s = \|\lambda a\|' sa = s\|a\|' a.$$

This implies that $\dot{f}(\lambda) = \|a\|' a$ and hence $\|a\|' a = \|a\|$. It follows that $|\|a\|'|_{E^*} \geq 1$.

Now let us show that $|\|a\|'|_{E^*} \leq 1$. As $\|\cdot\|$ is differentiable at a, we may write

$$\|a + \lambda x\| - \|a\| - \|a\|' \lambda x = \|\lambda x\| \epsilon(\lambda x),$$

where $\lim_{h \to 0} \epsilon(h) = 0$. This implies that

$$|\|a\|' \lambda x| \leq |\|a + \lambda x\| - \|a\|| + |\|\lambda x\| \epsilon(\lambda x)|,$$

hence

$$|\|a\|' x| \leq \|x\|(1 + |\epsilon(\lambda x)|).$$

Letting λ go to 0 we obtain

$$|\|a\|' x| \leq \|x\|,$$

which implies that $|\|a\|'|_{E^*} \leq 1$. This finishes the proof. □

Corollary 2.6. *The norm is differentiable on $E \setminus \{0\}$ if and only if it is differentiable on the unit sphere.*

We have already briefly spoken of the differentiability of the norms $\|\cdot\|_p$ defined on $\mathbb{R}^n$. We will now consider these norms in more detail. We will use the notation $\partial_i \|x\|_p$ for the ith partial derivative at x, which is simpler than $\frac{\partial \|\cdot\|_p}{\partial x_i}(x)$. There are three cases to consider, namely $\|\cdot\|_p$, with $1 < p < \infty$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

Case 1. $\|\cdot\|_p$**,** $1 < p < \infty$: all partial derivatives exist and are continuous on the open set $\mathbb{R}^n \setminus \{0\}$.
If $x_i > 0$, then

$$\partial_i \|x\|_p = \frac{1}{p}(|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}-1} p x_i^{p-1} = \|x\|_p^{1-p} x_i^{p-1}$$

and, for $x_i < 0$, we have

$$\partial_i \|x\|_p = -\frac{1}{p}(|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}-1} p(-x_i)^{p-1} = \|x\|_p^{1-p}(-x_i)^{p-1}.$$

The case where $x_i = 0$ is a little more delicate. We are interested in the following limit (if it exists):

$$\lim_{t \to 0} \frac{1}{t}\left(\|x + te_i\|_p - \|x\|_p\right) = \lim_{t \to 0} \frac{1}{t}\left(\left(|t|^p + \|x\|_p^p\right)^{\frac{1}{p}} - \|x\|_p\right),$$

where e_i is the ith element of the standard basis of $\mathbb{R}^n$. If we divide the numerator of the expression on the right-hand side by $\|x\|_p$ we obtain

$$\left(\left(\frac{|t|}{\|x\|_p}\right)^p + 1\right)^{\frac{1}{p}} - 1 = \left(1 + \frac{1}{p}\left(\frac{|t|}{\|x\|_p}\right)^p + o\left(\left(\frac{|t|}{\|x\|_p}\right)^p\right)\right) - 1$$
$$= \frac{1}{p}\left(\frac{|t|}{\|x\|_p}\right)^p + o(|t|^p).$$

As

$$\lim_{t \to 0} \frac{\|x\|_p}{t}\left(\frac{1}{p}\left(\frac{|t|}{\|x\|_p}\right)^p + o(|t|^p)\right) = 0,$$

we have

$$\partial_i \|x\|_p = 0.$$

Hence the partial derivatives are defined and continuous on $\mathbb{R}^n \setminus \{0\}$ and so the norm $\| \cdot \|_p$ is of class C^1 on this set.

Case 2. $\| \cdot \|_1$: all partial derivatives are defined and continuous on the open set $S = \{x \in \mathbb{R}^n : x_i \neq 0 \ for\ all\ i\}$.
We have

$$\|x\|_1 = |x_1| + \cdots + |x_n|.$$

If $x_i = 0$, then

$$\frac{1}{t}(\|x + te_i\|_1 - \|x\|_1) = \frac{|t|}{t}$$

and so $\partial_i \|x\|_1$ does not exist, which implies that the differential is not defined at a point x with a coordinate whose value is 0. Now suppose that this not the case and that $t \in \mathbb{R}$ is such that $|t| < |x_i|$. For $x_i > 0$

$$\frac{1}{t}(\|x + te_i\|_1 - \|x\|_1) = 1$$

and for $x_i < 0$

$$\frac{1}{t}(\|x + te_i\|_1 - \|x\|_1) = -1.$$

It follows that in the first case $\partial_i \|x\|_1 = 1$ and in the second $\partial_i \|x\|_1 = -1$. If all the coordinates of a point x are nonzero, then we may find a neighbourhood N of x such that, if $y \in N$, then each coordinate y_i of y is nonzero and has the same sign as that of x_i. Hence all partial derivatives of the norm are defined and continuous on N. It follows that $\| \cdot \|_1$ is of class C^1 on S.

Case 3. $\| \cdot \|_\infty$: all partial derivatives are defined and continuous on the open set $T = \{x \in \mathbb{R}^n : |x_i| = \|x\|_\infty \ for\ a\ unique\ i\}$.

Let $x \neq 0$ with $|x_k| = \|x\|_\infty$. First let us suppose that x_k is not unique, i.e., there is an index l, with $l \neq k$, such that $|x_l| = |x_k|$. We now take $h_k, h_l \in \mathbb{R}^*$ with the same absolute value and such that h_k has the same sign as x_k and h_l the sign opposite to that of x_l. We set $h = (h_1, \ldots, h_n)$, where $h_i = 0$ for $i \neq k, l$. Then we have

$$\|x + h\|_\infty = |x_k + h_k| = |x_k| + |h_k| = \|x\|_\infty + \|h\|_\infty$$

and

$$\|x - h\|_\infty = |x_l - h_l| = |x_l| + |h_l| = \|x\|_\infty + \|h\|_\infty.$$

By addition we obtain

$$\|x + h\|_\infty + \|x - h\|_\infty - 2\|x\|_\infty = 2\|h\|_\infty. \tag{2.1}$$

If we suppose that the differential exists at x, then we have

$$\|x + h\|_\infty - \|x\|_\infty - \|x\|'_\infty h = o(h)$$

and

$$\|x - h\|_\infty - \|x\|_\infty + \|x\|'_\infty h = o(h).$$

An addition of the two expressions gives us

$$\|x + h\|_\infty + \|x - h\|_\infty - 2\|x\|_\infty = o(h). \tag{2.2}$$

Now, from the (2.1) and (2.2) we obtain

$$2\|h\|_\infty = o(h) \Longrightarrow \lim_{h \to 0} \frac{\|h\|_\infty}{\|h\|_\infty} = 0,$$

which is clearly false. It follows that the norm is not differentiable at x.

Now let us suppose that x_k is unique. Let e_i be the ith member of the standard basis of $\mathbb{R}^n$ and $t \in \mathbb{R}^*$ small. Then

$$\|x + te_i\|_\infty = \begin{cases} |x_k| & \text{for } i \neq k \\ |x_k + t| & \text{for } i = k \end{cases}.$$

Therefore

$$\partial_i \|x\|_\infty = \begin{cases} 0 & \text{if } i \neq k \\ 1 & \text{if } i = k \text{ and } x_k > 0 \\ -1 & \text{if } i = k \text{ and } x_k < 0 \end{cases}.$$

Also, there is a neighbourhood N of x such that, if $y \in N$, then y_k is nonzero, has the same sign as x_k and $\|y\|_\infty = |y_k|$ for a unique k. It follows that $\partial_i \|y\|_\infty = \partial_i \|x\|_\infty$ for all i. Thus the partial derivatives are defined and continuous on N and so the norm is of class C^1 on T.

Remark. We have seen above that certain norms on $\mathbb{R}^n$ are differentiable at all points other than the origin and others not. Thus equivalent norms on a vector space may have different differentiability properties.

Appendix: Gâteaux Differentiability

There is another differential which is often used. We have already defined the directional derivative at a point with respect to a nonzero vector for a real-valued function. In fact, we can extend this definition to mappings whose image lies in any normed vector space. Let O be an open subset of a normed vector space E, f a mapping defined on O whose image lies in a normed vector space F, $a \in O$ and u a nonzero element of E. If the limit

$$\lim_{t \longrightarrow 0} \frac{f(a+tu) - f(a)}{t}$$

exists, then we note this derivative $\partial_u f(a)$. It is called the *directional derivative of* f *at* a *in the direction* u. If the directional derivative is defined in all directions and there is a continuous linear mapping ϕ from E into F such that for all $u \in E$

$$\partial_u f(a) = \phi(u),$$

then we say that f is Gâteaux-differentiable at a and that ϕ is the *Gâteaux differential* of f at a. If a mapping f is differentiable at a point a, then clearly all its directional derivatives exist and we have $\partial_u f(a) = f'(a)u$. Thus f is Gâteaux-differentiable at a. However, the Gâteaux differential may exist without the differential existing. Here is an example. If the mapping $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is defined by

$$f(x,y) = \begin{cases} \frac{x^6}{x^8+(y-x^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases},$$

then $\partial_{(u,v)} f(0,0) = 0$ for all $(u,v) \in \mathbb{R}^2$ and so the Gâteaux differential exists at the origin. However, $f(x,x^2) = x^{-2}$ and so f is not continuous at the origin and *a fortiori* not differentiable.

Another point should also be made, namely that the existence of directional derivatives at a point does not imply that the mapping is Gâteaux differentiable. Let us consider the mapping $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$g(x,y) = \begin{cases} \frac{x(x^2-3y^2)}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}.$$

Then $\partial_{(u,v)} g(0,0) = g(u,v)$ for all $(u,v) \in \mathbb{R}^2$. As g is not linear, g is not Gâteaux differentiable at $(0,0)$.

To distinguish the differential from the Gâteaux differential, the differential is often referred to as the Fréchet differential. From what we have seen, we have the implications:

$$\text{Fréchet differentiable} \implies \text{Gâteaux differentiable} \implies \text{existence of directional derivatives}$$

and the implications are not reversible.

In fact, the Gâteaux differentiability of the norm is closely related to the geometry of the unit sphere. The book by Beauzamy [3] handles this subject in some detail.

Calculus on Normed Vector Spaces
Coleman, R.
2012, XI, 249 p., Softcover
ISBN: 978-1-4614-3893-9