

Contents

Part A. Mathematical Preliminaries

Chapter A1. Some Basic Concepts

A1.1. Sets and Lists

A1.2. Functions

A1.3. Groups

Chapter A2. Linear Spaces

A2.1. Definition of a Linear Space

A2.2. Finite-Dimensional Linear Spaces

A2.3. Inner Product Spaces. Normed Linear Spaces

A2.4. Linear Transformations

A2.5. Linear Functionals. Dual Bases. Covariant and Contravariant Components

A2.6. Transformations of Bases

A2.7. Volume Orientation Functionals. Oriented Spaces

A2.8. The Vector Product

Chapter A3. Second-Order Tensors

A3.1. Definition of a Second-Order Tensor. The Tensor Product

- A3.2. Second-Order Tensors on Finite-Dimensional Linear Spaces
- A3.3. Transformations of Bases
- A3.4. Boundedness, Continuity, Operator Norm
- A3.5. Multiplication of Second-Order Tensors
- A3.6. Transposition, Symmetric Tensors
- A3.7. The Trace of a Second-Order Tensor. The Trace Inner Product
- A3.8. The Determinant of a Second-Order Tensor
- A3.9. The Inverse of a Second-Order Tensor
- A3.10. Orthogonal Second-Order Tensors
- A3.11. Eigenvalues, Eigenvectors, and Eigenspaces of a Second-Order Tensor
- A3.12. Direct Sum Decompositions, Invariant Linear Subspaces
- A3.13. The Spectral Theorem for Symmetric Tensors on Finite-Dimensional Spaces
- A3.14. The Hamilton-Cayley Theorem
- A3.15. Positive Definite Tensors. Tensor Square Roots. The Polar Decomposition Theorem

Chapter A4. Higher-Order Tensors

Chapter A5. Differentiation on Finite-Dimensional Linear Spaces

- A5.1. Metric Spaces. Open Sets
- A5.2. The Derivative
- A5.3. Functions of a Real Variable. The Ordinary Derivative
- A5.4. Real-Valued Functions. The Gradient
- A5.5. Elementary Rules of Differentiation
- A5.6. Derivatives of Products
- A5.7. Derivatives of Composite Functions. The Chain Rule
- A5.8. Some Additional Derivatives. The Invariants and the Inverse of a Tensor
- A5.9. The Mean-Value Theorem for Derivatives
- A5.10. Higher Derivatives. The CP Spaces
- A5.11. The Inverse Function Theorem. The Tensor Square Root
- A5.12. Functions of Several Variables. Partial Derivatives
- A5.13. The Codomain a Product Space
- A5.14. Identification with the Component Approach

Chapter A6. Point Spaces

- A6.1. Definition of a Point Space
- A6.2. Finite-Dimensional Point Spaces. Cartesian Coordinate Systems
- A6.3. Euclidean Point Spaces. Point Set Topology
- A6.4. Fields
- A6.5. Scalar Fields
- A6.6. Vector Fields
- A6.7. Second-Order Tensor Fields

- A6.8. Point Fields
- A6.9. Some Identities
- A6.10. Some Integral Theorems
- A6.11. General Coordinate Systems
- A6.12. Some Particular Coordinate Systems
- A6.13. Coordinate Transformations
- A6.14. Differentiation of Scalar Fields
- A6.15. Differentiation of Vector Fields. CHRISTOFFEL Symbols
- A6.16. Differentiation of Second-Order Tensor Fields
- A6.17. Differentiation of Point Fields

Chapter A7. The Tensor Exponential Function

v

Chapter AB. Isotropic Tensors and Isotropic Tensor Functions

AB.1. Isotropic Tensors

AB.2. Invariant Sets and Functions

AB.3. Isotropic Scalar-Valued Functions of a Vector

AB.4. Isotropic Scalar-Valued Functions of a Tensor

AB.5. Isotropic Tensor-Valued Functions of a Tensor

Part B. General Continuum Mechanics

Chapter B1. Kinematics

B1.1. Deformations

B1.2. Rigid Deformations. The Extension Theorem

B1.3. Proper Rotations

B1.4. Homogeneous Deformations

B1.5. The Geometry of General Deformations

B1.6. Isochoric Deformations. The Localization Theorem

B1.7. Bodies and Configurations

B1.8. Motions. Material and Spatial Descriptions. Functions of Position and Time

B1.9. Path Lines, Stream Lines, and Streak Lines

B1.10. Rigid Motions

- B1.11. The Geometry of Motions
- B1.12. Isochoric Motions. Parts of a Body
- B1.13. Reference Configurations. The Transport Theorem
- B1.14. The Referential Description of Rigid and Isochoric Motions
- B1.15. Coordinate Analysis in the Referential Description
- B1.16. The Present Configuration as the Reference Configuration.
The Relative Description
- B1.17. The Relative Description of Isochoric and Rigid Motions
- B1.18. Coordinate Analysis in the Relative Description

Chapter B2. Mass. Momentum. Kinetic Energy

- B2.1. Mass Density. Conservation of Mass
- B2.2. Mass Conservation in Isochoric and Rigid Motions
- B2.3. Center of Mass
- B2.4. Momentum, Kinetic Energy
- B2.5. Momentum and Kinetic Energy for Rigid Motions

Chapter B3. Force. Stress. Momentum Balance

- B3.1. The Stress Principle and Balance of Momentum

- B3.2. Momentum Balance for Rigid Bodies
- B3.3. The Theorem of Virtual Power
- B3.4. The Stress Tensor and the Stress Equations of Motion
- B3.5. Additional Properties of the Stress Tensor
- B3.6. Engineering Stress

Chapter B4. A Theory of Constitutive Equations

- B4.1. Simple Materials. The Response Function
- B4.2. The Principle of Material Indifference
- B4.3. Observer Transformations. Another Approach to Material Indifference
- B4.4. Material Indifference in the Relative Description
- B4.5. Material Symmetry
- B4.6. NOLL's Classification. Isotropic, Fluid, and Solid Materials
- B4.7. Material Symmetry in the Relative Description
- B4.8. Internal Constraints
- B4.9. Incompressible, Inextensible, and Rigid Bodies
- B4.10. Material Symmetry of Constrained Bodies

Part C. Examples of Simple Materials

Chapter C1. Some Fluids

C1.1. REINER-RIVLIN Fluids

C1.2. NEWTONIAN Fluids

C1.3. RIVLIN-ERICKSEN Fluids

C1.4. Viscometric Flows of Incompressible Simple Fluids

Chapter C2. Elasticity

C2.1. The Elastic Response Function

C2.2. Material Symmetry

C2.3. Some Equilibrium Solutions for Compressible,
Isotropic, Homogeneous, Elastic Solids

C2.4. Some Equilibrium Solutions for Incompressible,
Isotropic, Homogeneous, Elastic Solids

Chapter A5

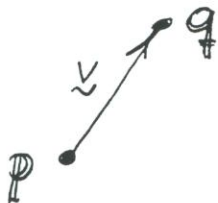
Differentiation on Finite-Dimensional Linear Spaces

In this chapter we give a modern development of the differentiation of a function defined on a subset of a finite-dimensional linear space with values in a (generally-different) finite-dimensional linear space. This gives us a single theory that is sufficiently general for and especially well-suited to the situations commonly encountered in continuum mechanics.

A5.1. Metric Spaces . Open Sets

Roughly speaking, the modern approach to differentiation is to consider the linear approximation to the difference in the values of a function at neighboring arguments. This program requires the notion of distance between objects.

For motivation here, we return to our experience with the Euclidean plane. Any two points in the plane, say P and Q , are connected by a vector, say $\underline{v}$, and it is natural to define the distance between the two points, $d(P, Q)$, to be the length of $\underline{v}$.



But the length of $\underline{v}$ is simply its magnitude or inner

product norm

$$\|\underline{v}\| = |\underline{v}| = \sqrt{\underline{v} \cdot \underline{v}};$$

thus

$$d(p, q) = \|\underline{v}\|.$$

The main properties of the concept of norm were summed up in Definition A2.3.3. We interpret them next in the present context.

Positive homogeneity requires

$$\|\alpha \underline{v}\| = |\alpha| \|\underline{v}\| \quad \forall \alpha \in \mathbb{R}.$$

In particular, the choice $\alpha = -1 \Rightarrow$

$$\|-\underline{v}\| = \|\underline{v}\|.$$

Now, if $\underline{v}$ is the vector from p to q , then $-\underline{v}$ is the vector from q to p ; so $\|-\underline{v}\| = \|\underline{v}\|$ means

$$d(q, p) = d(p, q).$$

Hence, the distance from q to p is the same as the distance from p to q , as it should be.

Positive definiteness requires

$$d(p, q) = \|\underline{v}\| \geq 0$$

with

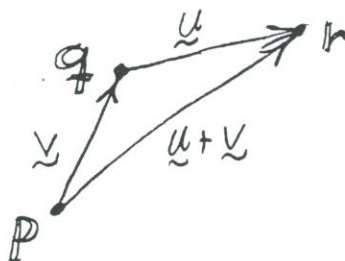
$$d(p, q) = \|z\| = 0 \text{ iff } z = 0.$$

Of course, $z = 0$ means that the points p and q have coalesced.

To interpret the triangle inequality,

$$\|u + v\| \leq \|u\| + \|v\|,$$

consider the figure below.



In terms of distance, we have

$$d(p, r) \leq d(p, q) + d(q, r);$$

i.e., the distance from p to r is less than or equal to the distance from p to q plus the distance from q to r — again a result with which we are very comfortable.

These remarks lead us to

Definition A5.1.1. A metric space is a set, say M , equipped with a function

$$d: M \times M \rightarrow \mathbb{R},$$

called a distance function¹ with the value $d(x, y)$ called the distance between x and y , which satisfies the following axioms:

(M1) Symmetry. $\forall x, y \in M$

$$d(y, x) = d(x, y);$$

(M2) Positive Definiteness. $\forall x, y \in M$

$$d(x, y) \geq 0 \text{ with } d(x, y) = 0 \text{ iff } x = y;$$

(M3) Triangle Inequality. $\forall x, y, z \in M$

$$d(x, z) \leq d(x, y) + d(y, z).$$

In view of our motivation for the above definition, the next result is not surprising.

Theorem A5.1.1. A normed linear space V is a metric space with respect to the distance function

¹Often, metric.

$$d(\underline{u}, \underline{v}) := \|\underline{u} - \underline{v}\|.$$

Moreover, this norm induced distance function is

(i) translation invariant: $\forall \underline{u}, \underline{v}, \underline{a} \in V$

$$d(\underline{u} + \underline{a}, \underline{v} + \underline{a}) = d(\underline{u}, \underline{v});$$

and

(ii) positive homogeneous: $\forall \alpha \in \mathbb{R}$ and $\forall \underline{u}, \underline{v} \in V$

$$d(\alpha \underline{u}, \alpha \underline{v}) = |\alpha| d(\underline{u}, \underline{v}).$$

Proof. Exercise A5.1.1. $\square$

The topology generated by this choice of metric for a normed linear space is called the norm topology or strong topology, and it will always be the topology understood for metrical considerations in normed linear spaces, unless explicitly stated to the contrary.

Because a metric space need not be equipped with the operations of addition and scalar multiplication, the converse of Theorem A5.1.1 is false; i.e., not all metric spaces are normed linear spaces. The following theorem shows that any set, not just normed linear spaces, can be equipped with a suitable distance function and hence can be viewed as a metric space.

Theorem A5.1.2. Any set M is a metric space w.r.t.
the discrete metric

$$d(x,y) := \begin{cases} 0, & x=y \\ 1, & x \neq y \end{cases}.$$

Proof, Exercise A5.1.2. $\square$

In working with metric spaces, it is often helpful to think geometrically as we did in motivating the concept. Then it is natural to refer to the elements of a metric space as points.

Definition A5.1.2. Let M be a metric space with distance function d . Let $x_0 \in M$ and let $r \in \mathbb{R}$ with $r > 0$. Then the subset

$$B_r(x_0) = \{x \in M : d(x, x_0) < r\}$$

is the open ball of radius r centered at point x_0 .

An open ball is not necessarily "solid" or "full". It may miss some points

because they are not in M . However, by definition the center point is always included,

Definition A5.1.3. A subset S of a metric space M is open (in M) if each point of S is the center of an open ball contained in S .

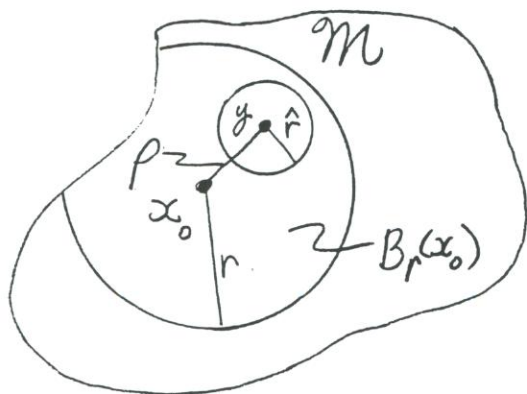
If the following were not true, we would have to adjust our terminology.

Theorem A5.1.3. Let $x_0 \in M$, where M is a metric space. Then the open ball $B_r(x_0)$ centered at x_0 is open.

Proof. Let $y \in B_r(x_0)$. Then if d is the distance function for M ,

$$\rho := d(y, x_0) < r.$$

This situation is depicted in the following figure.



We need $\hat{r} > 0 \ni \hat{r} + \rho < r$. Take $\hat{r} = \frac{1}{2}(r - \rho)$. Then the open ball $B_{\hat{r}}(y)$ should be contained in $B_r(x_0)$. Indeed, let $z \in B_{\hat{r}}(y)$. Then $d(z, y) < \hat{r}$. By

the triangle inequality,

$$d(z, x_0) \leq \underbrace{d(z, y)}_{< \hat{r}} + \underbrace{d(y, x_0)}_{= \rho} < \hat{r} + \rho = \underbrace{\frac{1}{2}(r + \rho)}_{< r} < r.$$

$\therefore z \in B_r(x_0) \Rightarrow B_{\hat{r}}(y) \subset B_r(x_0)$, and $B_r(x_0)$ is open. $\square$

We also have

Theorem A5.1.4. Let $\mathcal{M}$ be a metric space. Then $\mathcal{M}$ is open.

Proof. Let $y \in \mathcal{M}$. Then for any $r > 0$,

$$B_r(y) = \{x \in \mathcal{M} : d(y, x) < r\},$$

where d is the distance function for $\mathcal{M}$. Trivially, $B_r(y) \subset \mathcal{M}$, so $\mathcal{M}$ is open. $\square$

It is obvious that any subset of a metric space is itself a metric space w.r.t. the same distance function. Consequently, subsets of metric spaces are often called subspaces, when this fact is coupled with our earlier observation that open balls need not be "completely developed" because they may miss points not in the underlying metric space, we can get into some subtle situations. The following exercise is an indication that

Despite the simplicity of the last two proofs, we should not get too cavalier about this business.

Exercise A5.1.3, In view of Exercise A2.3.7 and Theorems A2.3.9 and A5.1.1, the real line $\mathbb{R}$ is a metric space w.r.t. the absolute value distance function

$$d(x, y) = |x - y|.$$

For $a < b < c$, consider the "intervals"

$$[a, b) := \{x \in \mathbb{R} : a \leq x < b\}$$

and

$$[a, c] := \{x \in \mathbb{R} : a \leq x \leq c\}.$$



Show that

(i) $[a, b)$ is open in $[a, c]$

and

(ii) $[a, b)$ is not open in $\mathbb{R}$.

The following definition is a rather natural use of the language.

Definition A5.1.4. Let S be a subset of a metric space. Then

a point $x \in S$ is an interior point of S if it is the center of an open ball contained in S . The interior of S , denoted by $\overset{\circ}{S}$, is the set of all interior points of S .

It is not difficult to prove

Theorem A5.1.5. Let S be a subset of a metric space M . Then

(i) if S is open, each point of S is an interior point of S ;

(ii) the interior $\overset{\circ}{S}$ is open;

and

(iii) S is open iff $\overset{\circ}{S} = S$.

Proof. Exercises A5.1.4 - 6. $\square$

Supplementary Reading

DIEUDONNÉ, Foundations of Modern Analysis

KOLMOGOROV and FOMIN, Introductory Real Analysis

MICHEL and HERGET, Mathematical Foundations in Engineering and Science

NAYLOR and SELL, Linear Operator Theory in Engineering and Science

ODEN, Applied Functional Analysis

SIMMONS, Introduction to Topology and Modern Analysis

A5.2. The Derivative

Throughout this section, we consider a function f whose domain D is a subset of a finite-dimensional linear space U , called the domain space and whose co-domain is a (generally different) finite-dimensional space V , called the range space. Thus,

$$f: D \rightarrow V \text{ with } D \subset U.$$

Since U and V are finite-dimensional linear spaces, we take them to be inner product spaces.¹ Then they are normed linear spaces w.r.t. the inner product norm

$$|u| = \sqrt{u \cdot u}, \quad)^2$$

which we shall always use. Finally, they are metric spaces w.r.t. the norm induced distance function

$$d(u, v) = |u - v|, \quad)^3$$

which we shall always use.⁴ Of course, when U and V are different spaces, then they have different inner products, norms, and distances. In the case of the linear space of second-order

¹ Cf. Theorem A2.3.2.

² Cf. Theorem A2.3.9.

³ Cf. Theorem A5.1.1.

tensors $\text{Lin } V_n$, where V_n is an n -dimensional inner product space, we shall always use the inner product

$$\underline{S} \cdot \underline{T} = \text{tr} (\underline{S} \underline{T}^T) \quad . \quad)^1$$

In the case of the real line $\mathbb{R}$, we shall always use the multiplication inner product.² Then the inner product norm of $x \in \mathbb{R}$ is $|x|$,³ and the distance between x and $y \in \mathbb{R}$ is $|x-y|$.⁴

In our general discussions, we shall use our general linear space notation ($\underline{u}, \underline{v}$, etc.) for elements of U and V ; but we keep in mind that finite-dimensional linear spaces with special structures such as $\mathbb{R}$ and $\text{Lin } V_n$ are allowed even though the notation does not suggest this. When elements of $\mathbb{R}$ are viewed as scalars, then we shall generally denote them by Greek letters.

Finally, the domain D is required to be open in the domain space U .

¹ Cf. Theorem A3.7.3.

² Cf. Exercise A2.3.3.

³ Cf. Exercise A2.3.7.

⁴ Cf. Theorem A5.1.1.

As was mentioned at the beginning of the previous section, the basic idea of differentiation is to look at the linear approximation to the difference in the values of a function at neighbouring arguments. In order to be precise about the approximation, we need

Definition A5.2.1. Let $g: D' \rightarrow V$, where $D' \subset U$. Furthermore, let $0 \in U$ be an interior point of D' , and let $g(0) = 0$.¹ Then we say that $g(u)$ approaches 0 (in V) faster than u (approaches 0 in U), and we write

$$g(u) = o(u) \text{ as } u \rightarrow 0, \quad)^2$$

if

$$\lim_{\substack{u \rightarrow 0 \\ u \neq 0}} \frac{|g(u)|}{|u|} = 0. \quad)^3$$

Often it is convenient to write the statement

¹ Of course, the 0 on the r.h.s. refers to the zero element of V .

² Read g is "little oh of u ". There is also a "big oh" symbol, $\sim$ which has a different meaning; be careful with your handwriting.

³ Because of the $|u|$ in the denominator, we must do a "deleted limit" here, as our notation emphasizes. See BARTLE's The Elements of Real Analysis for elucidation of this point. (cont.)

as

$$\underset{\sim}{g}(\underline{u}) - \underset{\sim}{h}(\underline{u}) = o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}$$

$$\underset{\sim}{g}(\underline{u}) = \underset{\sim}{h}(\underline{u}) + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

The next theorem provides some rather transparent properties of little oh functions, which we shall use repeatedly.

Theorem A5.2.1. In the context of Definition A5.2.1, let

$$\underset{\sim}{g}(\underline{u}) = o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

Then

$$(i) \lim_{\underline{u} \rightarrow \underline{0}} \underset{\sim}{g}(\underline{u}) = \underline{0} \text{ ; i.e., } \underset{\sim}{g} \text{ is continuous at } \underline{0} \text{ ;}$$

$$(ii) \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \underset{\sim}{g}(\alpha \underline{u}) = \underline{0} \text{ ;}$$

Insert material on p. A5.2.5 a

Proof. (i): Exercise A5.2.1.

(ii): First, suppose $\underline{u} \neq \underline{0}$. Then

$$\frac{\underset{\sim}{g}(\alpha \underline{u})}{\alpha} = \frac{|\underline{u}| \underset{\sim}{g}(\alpha \underline{u})}{|\underline{u}| \alpha},$$

which by the positive homogeneity of the magnitude function $\Rightarrow$

(cont. from p. A5.2.3) Actually, all limits in these notes are to be viewed as deleted limits. This agrees with the more common practice.

$$\left| \frac{g(\alpha \underline{u})}{\alpha} \right| = \frac{|\underline{u}| |g(\alpha \underline{u})|}{|\alpha \underline{u}|} ,$$

$$\therefore \lim_{\alpha \rightarrow 0} \left| \frac{g(\alpha \underline{u})}{\alpha} \right| = |\underline{u}| \lim_{\alpha \rightarrow 0} \frac{|g(\alpha \underline{u})|}{|\alpha \underline{u}|} \quad (\text{Exercise A5.2.2})$$

$$= |\underline{u}| \lim_{\underline{v} \rightarrow 0} \frac{|g(\underline{v})|}{|\underline{v}|} \quad (\text{Exercise A5.2.3})$$

$$= 0 \quad (g(\underline{v}) = o(\underline{v}) \text{ as } \underline{v} \rightarrow 0)$$

$$\Rightarrow \lim_{\alpha \rightarrow 0} \frac{g(\alpha \underline{u})}{\alpha} = 0 \quad (\text{Exercise A5.2.4}).$$

In the case $\underline{u} = 0$, we note that by hypothesis

$$g(\alpha \underline{u}) = 0 \quad \forall \alpha \in \mathbb{R}.$$

Let $\varepsilon > 0$ be given. Then for any choice of $\delta > 0$,

$$0 < |\alpha - 0| < \delta \Rightarrow \alpha \neq 0 \Rightarrow \left| \frac{g(\alpha \underline{u})}{\alpha} - 0 \right| = |0| = 0 < \varepsilon ,$$

Thus,

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} g(\alpha \underline{u}) = 0 .$$

(iii) Exercise A5.2.5.

(iv) Exercise A5.2.6. $\square$

INSERT for p. A5.2.4

A5.2.5a

(iii) let also $\underline{h}(\underline{u}) = o(\underline{u})$ as $\underline{u} \rightarrow \underline{Q}$, then

$$\underline{g}(\underline{u}) + \underline{h}(\underline{u}) = o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{Q};$$

(iv) let also $\underline{u}(\underline{z}) = o(\underline{z})$ as $\underline{z} \rightarrow \underline{Q}$, then

$$\underline{g}(\underline{u}(\underline{z})) = o(\underline{z}) \text{ as } \underline{z} \rightarrow \underline{Q}.$$

return to p. A5.2.4

With the reminder that throughout this section

$$\underline{f} : \underline{D} \rightarrow \underline{V}$$

where $\underline{D}$ is an open subset of $\underline{U}$, $\underline{U}$ and $\underline{V}$ are finite-dimensional linear spaces, and the agreements w.r.t. inner products, norms, metrics, etc. outlined at the beginning of the section are in force, we finally have.

Definition A5.2.2. The function $\underline{f}$ is differentiable at $\underline{x} \in \underline{D}$ if $\exists$ a linear transformation

$$D\underline{f}(\underline{x}) : \underline{U} \rightarrow \underline{V}$$

$\Rightarrow$ for sufficiently small $|\underline{u}|$

$$\underline{f}(\underline{x} + \underline{u}) - \underline{f}(\underline{x}) = D\underline{f}(\underline{x})[\underline{u}] + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.)^1$$

The linear transformation $D\underline{f}(\underline{x})$ is called the derivative² of $\underline{f}$ at $\underline{x}$.

Here, $D\underline{f}(\underline{x})[\underline{u}]$ denotes the value of the linear

¹Since $\underline{D}$ is open, $\underline{x} + \underline{u}$ will be in $\underline{D}$ for sufficiently small $|\underline{u}|$.

²Sometimes Fréchet derivative, especially when $\underline{U}$ and/or $\underline{V}$ are infinite-dimensional. In this case, the linear transformation $D\underline{f}(\underline{x})$ is required to be bounded. In the finite-dimensional case, all linear transformations are bounded (see Theorem A3.4.2).

transformation $Df(\underline{x})$ at $\underline{x}$. The square bracket is traditionally used to emphasize the linearity. Sometimes $Df(\underline{x})[\underline{u}]$ is called the differential of f at $\underline{x}$ corresponding to the increment $\underline{u}$.

Theorem A5.2.2. If f is differentiable, then its derivative is unique. In fact, its values are given by

$$Df(\underline{x})[\underline{u}] = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [f(\underline{x} + \alpha \underline{u}) - f(\underline{x})] = \left. \frac{d}{d\alpha} f(\underline{x} + \alpha \underline{u}) \right|_{\alpha=0}.$$

Proof. Suppose f is differentiable at $\underline{x}$. Then

$$f(\underline{x} + \alpha \underline{u}) - f(\underline{x}) = Df(\underline{x})[\alpha \underline{u}] + o(\alpha \underline{u}) \text{ as } (\alpha \underline{u}) \rightarrow \underline{0};$$

and by the linearity of $Df(\underline{x})$,

$$f(\underline{x} + \alpha \underline{u}) - f(\underline{x}) = \alpha Df(\underline{x})[\underline{u}] + o(\alpha \underline{u}) \text{ as } (\alpha \underline{u}) \rightarrow \underline{0}.$$

$$\therefore \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [f(\underline{x} + \alpha \underline{u}) - f(\underline{x})] = \lim_{\alpha \rightarrow 0} Df(\underline{x})[\underline{u}] + \lim_{\alpha \rightarrow 0} \frac{o(\alpha \underline{u})}{\alpha}$$

provided the two limits on the right hand side each exist. Clearly,

$$\lim_{\alpha \rightarrow 0} Df(\underline{x})[\underline{u}] = Df(\underline{x})[\underline{u}]$$

since $Df(\underline{x})[\underline{u}]$ is independent of α . By Theorem A5.2.1(ii),

$$\lim_{\alpha \rightarrow 0} \frac{o(\alpha \underline{u})}{\alpha} = \underline{0}.$$

Hence,

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [f(\underline{x} + \alpha \underline{u}) - f(\underline{x})] = Df(\underline{x})[\underline{u}]. \quad \square$$

The right hand side in Theorem A5.2.2 is just the directional derivative. Formally, we make

Definition A5.2.3. The function f is said to have
directional derivative'

$$\delta f(\underline{x}; \underline{u}) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [f(\underline{x} + \alpha \underline{u}) - f(\underline{x})] = \left. \frac{d}{d\alpha} f(\underline{x} + \alpha \underline{u}) \right|_{\alpha=0}$$

at $\underline{x}$ w.r.t. $\underline{u}$ if the above limit exists.

There are at least two important distinctions to draw between directional derivatives and derivatives. First, the directional derivative $\delta f(\underline{x}; \underline{u})$ is not required to be linear in $\underline{u}$; second, the definition of $\delta f(\underline{x}; \underline{u})$ makes sense even if the domain space $\mathcal{U}$ is not normed.

Exercise A5.2.7. Even though the directional derivative $\delta f(\underline{x}; \underline{u})$ is not required to be linear in $\underline{u}$, it is automatically homogeneous. Show that if $\delta f(\underline{x}; \underline{u})$ exists, then

$$\delta f(\underline{x}; \alpha \underline{u}) = \alpha \delta f(\underline{x}; \underline{u}) \quad \forall \alpha \in \mathbb{R}.$$

'Often Gateaux derivative or first variation, especially when $\mathcal{U}$ and/or $\mathcal{V}$ are infinite-dimensional.

On combining Theorem A5.2.2 and Definition A5.2.3, we get

Theorem A5.2.3. If f is differentiable at x , then f has a directional derivative at x w.r.t. all $u \in U$. Moreover,

$$\delta f(x; u) = Df(x)[u].$$

In particular, when f is differentiable at x , $\delta f(x; u)$ is linear in u .

Another consequence of differentiability is continuity. Formally, we have

Theorem A5.2.4. If f is differentiable at x , then it is continuous at x .

Proof. Suppose that f is differentiable at x . Then on taking $u = y - x$ in Definition A5.2.2, we get

$$f(y) = f(x) + Df(x)[y-x] + o(y-x) \text{ as } (y-x) \rightarrow 0.$$

Then

$$\lim_{y \rightarrow x} f(y) = \lim_{y \rightarrow x} f(x) + \lim_{y \rightarrow x} Df(x)[y-x] + \lim_{y \rightarrow x} o(y-x)$$

provided the three limits on the right-hand side exist. Obviously,

$$\lim_{y \rightarrow x} f(x) = f(x)$$

since $f(x)$ is independent of y . By Theorem A5.2.1(i),

$$\lim_{\substack{y \rightarrow x \\ \sim}} o(y-x) = \lim_{\substack{u \rightarrow 0 \\ \sim}} o(u) = 0.$$

Since $Df(x)$ is a linear transformation on the finite-dimensional space U into the finite-dimensional space V , it is bounded. Then by Theorem A3.4.1, $Df(x)$ is continuous; whence

$$\lim_{y \rightarrow x} Df(x)[y-x] = \lim_{u \rightarrow 0} Df(x)[u] = Df(x)[0] = 0.$$

$$(\text{Alternatively, } \lim_{y \rightarrow x} Df(x)[y-x] = \lim_{y \rightarrow x} \{Df(x)[y] - Df(x)[x]\})$$

$$= \lim_{y \rightarrow x} Df(x)[y] - \lim_{y \rightarrow x} Df(x)[x] = Df(x)[x] - Df(x)[x] = 0.)$$

Thus, we are left with

$$\lim_{y \rightarrow x} f(y) = f(x);$$

i.e., f is continuous at x . $\square$

Supplementary Reading

BARTLE, The Elements of Real Analysis

DIEUDONNÉ, Foundations of Modern Analysis

GURTIN, An Introduction to Continuum Mechanics

LANG, Analysis II

LOOMIS and STERNBERG, Advanced Calculus

LUENBERGER, Optimization by Vector Space Methods

LUSTERNIK and SOBOLEV, Elements of Functional Analysis

NICKERSON, SPENCER, and STEENROD, Advanced Calculus

NOLL, Finite-Dimensional Spaces

RUDIN, Principles of Mathematical Analysis

SPIVAK, Calculus on Manifolds

TRUESDELL and NOLL, The Non-Linear Field Theories of Mechanics

A5.3. Functions of a Real Variable. The Ordinary Derivative

We continue with the terminology of the previous section. When the domain-space U is the real line $\mathbb{R}$, then the ordinary derivative of elementary calculus defined by

$$\underline{f}'(x) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [\underline{f}(x+\alpha) - \underline{f}(x)] \quad (*)$$

makes sense; and it is natural to ask how it is related to the derivative of Definition A5.2.2. Accordingly, suppose that $\underline{f}$ is defined on an open subset of $\mathbb{R}$ and that it has an ordinary derivative at x , which means that the limit in $(*)$ exists. The statement $(*)$ is equivalent to

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [\underline{f}(x+\alpha) - \underline{f}(x) - \alpha \underline{f}'(x)] = 0,$$

which $\Rightarrow$

$$\lim_{\alpha \rightarrow 0} \frac{1}{|\alpha|} |\underline{f}(x+\alpha) - \underline{f}(x) - \alpha \underline{f}'(x)| = 0.$$

We leave it as Exercise A5.3.1 to formulate and prove this implication in a reasonably general setting. By Definition A5.2.1, the last limit above tells us that

$$\underline{f}(x+\alpha) - \underline{f}(x) = \alpha \underline{f}'(x) + o(\alpha) \text{ as } \alpha \rightarrow 0.$$

¹ often, denoted by $\frac{df}{dx}(x)$.

Now $\alpha \mapsto \alpha f'(x)$ obviously defines a linear mapping of $\mathbb{R}$ into V . $\therefore$ by Definition A5.2.2, f is differentiable at x and

$$Df(x)[\alpha] = \alpha f'(x).$$

Conversely, suppose that f is differentiable at x . Then by Definition A5.2.2, $\exists$ a linear transformation $Df(x): \mathbb{R} \rightarrow V$

$$f(x+\alpha) - f(x) = Df(x)[\alpha] + o(\alpha) \text{ as } \alpha \rightarrow 0.$$

$\therefore$

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [f(x+\alpha) - f(x)] = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} Df(x)[\alpha] + \lim_{\alpha \rightarrow 0} \frac{o(\alpha)}{\alpha}$$

provided the limits on the right hand side exist. Since $Df(x)$ is linear,

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} Df(x)[\alpha] = \lim_{\alpha \rightarrow 0} Df(x)[1] = Df(x)[1],$$

and by Definition A5.2.1,

$$\lim_{\alpha \rightarrow 0} \frac{o(\alpha)}{\alpha} = 0.$$

Thus,

$$f'(x) = Df(x)[1].$$

In summary, we have proven

Theorem A5.3.1. Let f be defined on an open subset
of $\mathbb{R}$. If f has an ordinary derivative $f'(x)$ at x , then
 f is differentiable at x and

$$Df(x)[\alpha] = \alpha f'(x).$$

Conversely, if f is differentiable at x , then f has an
ordinary derivative at x and

$$f'(x) = Df(x)[1].$$

Supplementary Reading

Same as for § 5.2

A5.4. Real-Valued Functions. The Gradient

Again, we continue with the terminology of § A5.2. When the range space or codomain V is the real line $\mathbb{R}$, then the derivative $Df(\underline{x})$ is a linear transformation on $\mathcal{U}$ into $\mathbb{R}$; i.e., it is a linear functional, and by the representation Theorem A2.5.1 $\exists$ a unique element of $\mathcal{U}$, say $\nabla f(\underline{x})$, $\exists$

$$Df(\underline{x})[\underline{u}] = \nabla f(\underline{x}) \cdot \underline{u}.$$

Formally, we have

Theorem A5.4.1. Let f be real-valued. Then if f is differentiable at $\underline{x} \in \mathcal{D} \subset \mathcal{U}$, $\exists$ a unique element of $\mathcal{U}$ called the gradient of f at $\underline{x}$ and denoted by $\nabla f(\underline{x})$ $\forall \underline{u} \in \mathcal{U}$

$$Df(\underline{x})[\underline{u}] = \nabla f(\underline{x}) \cdot \underline{u}.$$

When both the domain space and the range space are the real line, then both the ordinary derivative and the gradient are defined, and they turn out to be the same thing.

¹Often, $\frac{\partial f}{\partial \underline{x}}(\underline{x})$

Theorem A5.4.2. Let $U = V = \mathbb{R}$. If f is differentiable
at $x \in D \subset \mathbb{R}$, equivalently¹ if f has an ordinary derivative
at x , then

$$\nabla f(x) = f'(x);$$

i.e., the gradient of f at x is the ordinary derivative of
 f at x .

Proof. Exercise A5.4.1. $\square$

Supplementary Reading

Same as for § 5.2

¹cf. Theorem A5.3.1.

A5.5. Elementary Rules of Differentiation

Throughout, we study $\underline{f}: D \rightarrow V$, where D is an open subset of U ; U and V are finite-dimensional linear spaces; and all of the agreements set out in §5.2 are in force.

The rules for differentiating sums and constant scalar multiples that we expect from our experience in elementary calculus do indeed hold here by straight forward application of the definition of differentiability.

Theorem A5.5.1. Let $\underline{f}: D \rightarrow V$ and $\underline{g}: D \rightarrow V$ both be differentiable at $\underline{x} \in D$. Then so are the sum $\underline{f} + \underline{g}$, defined on D into V by

$$(\underline{f} + \underline{g})(\underline{x}) = \underline{f}(\underline{x}) + \underline{g}(\underline{x}),$$

and the constant scalar multiple $\alpha \underline{f}$, defined on D into V by

$$(\alpha \underline{f})(\underline{x}) = \alpha \underline{f}(\underline{x}) \quad (\alpha \in \mathbb{R} \text{ and constant}).$$

Moreover,

$$D(\underline{f} + \underline{g})(\underline{x}) = D\underline{f}(\underline{x}) + D\underline{g}(\underline{x})$$

and

$$D(\alpha \underline{f})(\underline{x}) = \alpha D\underline{f}(\underline{x}).$$

When $V = \mathbb{R}$, then

$$\nabla(f+g)(\underline{x}) = \nabla f(\underline{x}) + \nabla g(\underline{x}) \text{ and } \nabla(\alpha f)(\underline{x}) = \alpha \nabla f(\underline{x}).$$

Proof. Exercise A5.5.1. $\square$

We also expect the derivative of a constant-valued function to be zero.

Theorem A5.5.2. Let $\underline{f}$ be the constant function defined
by $\underline{f}(\underline{x}) = \underline{f}_0 \quad \forall \underline{x} \in \mathcal{D}$, Then $\underline{f}$ is differentiable everywhere
on $\mathcal{D}$, and

$$D\underline{f}(\underline{x})[\underline{u}] = \underline{0}. \quad)^1$$

When $V = \mathbb{R}$, then

$$\nabla f(\underline{x}) = \underline{0}. \quad)^2$$

Proof. Exercise A5.5.2. $\square$

A slightly less trivial example is provided by

Theorem A5.5.3. Let $\underline{f}$ be a linear function on $\mathcal{D} = \mathcal{U}$,³

¹Here, $\underline{0}$ denotes the zero element of V .

²Here, $\underline{0}$ denotes the zero element of $\mathcal{U}$.

³We use the fact that the whole space $\mathcal{U}$ is open (Theorem A5.1.4).

and denote it by $f(x) = L[x]$. Then f is differentiable everywhere on U , and

$$Df(x)[u] = L[u].$$

When $V = \mathbb{R}$, then f is a linear functional, and by the representation Theorem A2.5.1 it has the form $f(x) = L \cdot x$.

Then

$$\nabla f(x) = L.$$

Proof. By the linearity,

$$f(x+u) - f(x) = L[x+u] - L[x] = L[u] = L[u] + o.$$

Since L is linear on U into V and since the function with constant value o goes to o faster than u ,¹ f is differentiable at x and

$$Df(x)[u] = L[u].$$

The proof of the assertion for $V = \mathbb{R}$ is left as Exercise A5.5.3. $\square$

There are a number of special cases of Theorem A5.5.3 that will be important to us in our study of continuum mechanics. They are collected together in the next theorem. No additional proof is needed — apart from the recognition in each case that the function is linear.

¹cf. the argument at the end of the proof of Theorem A5.2.1 (ii).

Theorem A5.5.4. In all cases below, the function is
differentiable everywhere on its domain, and its derivative
is given by the stated formula:

(i) tensor. $\underline{f}: \mathcal{V}_n \rightarrow \mathcal{V}_n$ defined by $\underline{f}(\underline{x}) = \underline{T}\underline{x}$, $\underline{T} \in \text{Lin } \mathcal{V}_n$,

$$D\underline{f}(\underline{x})[\underline{u}] = \underline{T}\underline{u}, \quad D\underline{f}(\underline{x}) = \underline{T};$$

(ii) identity map. $\underline{i}: \mathcal{V}_n \rightarrow \mathcal{V}_n$ defined by $\underline{i}(\underline{x}) = \underline{x}$,

$$D\underline{i}(\underline{x})[\underline{u}] = \underline{u}, \quad D\underline{i}(\underline{x}) = \underline{I};$$

(iii) constant tensor multiples. $\underline{F}, \underline{G}: \text{Lin } \mathcal{V}_n \rightarrow \text{Lin } \mathcal{V}_n$ defined by
 $\underline{F}(\underline{x}) = \underline{C}\underline{x}$, $\underline{G}(\underline{x}) = \underline{x}\underline{C}$,

$$D\underline{F}(\underline{x})[\underline{U}] = \underline{C}\underline{U}, \quad D\underline{G}(\underline{x})[\underline{U}] = \underline{U}\underline{C};$$

(iv) transpose. $\underline{H}: \text{Lin } \mathcal{V}_n \rightarrow \text{Lin } \mathcal{V}_n$ defined by $\underline{H}(\underline{x}) = \underline{x}^T$,

$$D\underline{H}(\underline{x})[\underline{U}] = \underline{U}^T;$$

(v) trace. $\underline{f}: \text{Lin } \mathcal{V}_n \rightarrow \mathbb{R}$ defined by $\underline{f}(\underline{x}) = \text{tr } \underline{x}$,

$$D\underline{f}(\underline{x})[\underline{U}] = \text{tr } \underline{U}, \quad \nabla \underline{f}(\underline{x}) = \underline{I}. \quad)^1$$

¹In order to pick off the gradient, we recognize that
 $\text{tr } \underline{x} = \underline{I} \cdot \underline{x}$ (see Theorem A3.7.6 (ii)).

A5.5.5

Supplementary Reading

Same as for $\S A5.2$

A5.6. Derivatives of Products

A5.6.1

Linear functions were especially easy to differentiate because the linear approximation to $\underline{L}[\underline{x} + \underline{u}] - \underline{L}[\underline{x}]$ was exact; i.e., there was no remainder term. As an important but still tractable example which is not so easy consider the square function on $\text{Lin } V$ into $\text{Lin } V$ defined by $\underline{F}(\underline{x}) = \underline{x}^2$. Then

$$\underline{F}(\underline{x} + \underline{u}) - \underline{F}(\underline{x}) = \underline{x}\underline{u} + \underline{u}\underline{x} + \underline{u}^2,$$

Now for any $\underline{x}$, $\underline{x}\underline{u} + \underline{u}\underline{x}$ is the value at $\underline{u}$ of a linear transformation from $\text{Lin } V$ to $\text{Lin } V$; i.e., the mapping

$$\underline{u} \mapsto \underline{x}\underline{u} + \underline{u}\underline{x}$$

is linear. By Definition A5.2.2, this linear transformation is then the derivative of $\underline{F}$ at $\underline{x}$ provided that $\underline{u}^2 = o(\underline{u})$. By Theorem A3.7.9,

$$\frac{|\underline{u}^2|}{|\underline{u}|} \leq \frac{|\underline{u}|^2}{|\underline{u}|} = |\underline{u}|$$

and $\therefore$

$$\lim_{\substack{\underline{u} \rightarrow \underline{0} \\ \underline{u} \neq \underline{0}}} \frac{|\underline{u}^2|}{|\underline{u}|} = 0, \quad)^1$$

Hence, $\underline{u}^2 = o(\underline{u})$ as $\underline{u} \rightarrow \underline{0}$ in accordance with Definition A5.2.1, and

¹Use the above estimate to give an ε - δ proof of this limit as Exercise A5.6.1.

we have established

Theorem A5.6.1. ^(tensor) The square function $\tilde{F}: \text{Lin } V_n \rightarrow \text{Lin } V_n$
defined by $\tilde{F}(\tilde{x}) = \tilde{x}^2$

is differentiable everywhere on $\text{Lin } V_n$, and

$$D\tilde{F}(\tilde{x})[\tilde{U}] = \tilde{x}\tilde{U} + \tilde{U}\tilde{x},$$

In the same way we are lead to

Theorem A5.6.2. ^(tensor) The cube function $\tilde{F}: \text{Lin } V_n \rightarrow \text{Lin } V_n$
defined by $\tilde{F}(\tilde{x}) = \tilde{x}^3$

is differentiable everywhere on $\text{Lin } V_n$, and

$$D\tilde{F}(\tilde{x})[\tilde{U}] = \tilde{x}^2\tilde{U} + \tilde{x}\tilde{U}\tilde{x} + \tilde{U}\tilde{x}^2.$$

Proof. Exercise A5.6.2. $\square$

Often the functions that arise naturally in continuum mechanics are one sort or another of product. Suppose, e.g., that $f: D \rightarrow \mathbb{R}$ ¹ is defined by

$$f(\tilde{x}) = \alpha(\tilde{x})\beta(\tilde{x})$$

¹If course, the arguments of § A5.2 still apply.

where $\alpha, \beta: D \rightarrow \mathbb{R}$ are differentiable at $\underline{x}$. Then by Definition A5.2.2,

$$\begin{aligned} f(\underline{x} + \underline{u}) - f(\underline{x}) &= \{\alpha(\underline{x}) + D\alpha(\underline{x})[\underline{u}] + o(\underline{u})\} \{\beta(\underline{x}) + D\beta(\underline{x})[\underline{u}] + o(\underline{u})\} - \alpha(\underline{x})\beta(\underline{x}) \\ &= \alpha(\underline{x})\{D\beta(\underline{x})[\underline{u}]\} + \{\alpha(\underline{x})\}D\beta(\underline{x})[\underline{u}] + \{D\alpha(\underline{x})[\underline{u}]\}\{\beta(\underline{x})\} \\ &\quad + o(\underline{u})[\alpha(\underline{x}) + \beta(\underline{x})] + o(\underline{u})\{D\alpha(\underline{x})[\underline{u}] + D\beta(\underline{x})[\underline{u}]\} + \alpha(\underline{u})\alpha(\underline{u}) \\ &\hspace{15em} \text{as } \underline{u} \rightarrow \underline{0}. \end{aligned}$$

The first two terms obviously define a linear transformation from $\mathcal{U}$ into $\mathbb{R}$. The remaining terms each go to zero faster than $\underline{u}$.⁰ The least obvious of these is

$$\{D\alpha(\underline{x})[\underline{u}]\}\{D\beta(\underline{x})[\underline{u}]\}.$$

However, by the boundedness of linear transformations on finite-dimensional spaces into finite-dimensional spaces¹, $\exists$ constants² A and B $\ni$

$$|D\alpha(\underline{x})[\underline{u}]| \leq A|\underline{u}| \text{ and } |D\beta(\underline{x})[\underline{u}]| \leq B|\underline{u}| \quad \forall \underline{u} \in \mathcal{U}.$$

$\therefore$

$$\frac{|D\alpha(\underline{x})[\underline{u}] D\beta(\underline{x})[\underline{u}]|}{|\underline{u}|} \leq AB|\underline{u}| \rightarrow 0 \text{ as } \underline{u} \rightarrow \underline{0}.$$

¹ See Theorem A3.4.2.

² Of course, A and B generally will depend on $\underline{x}$.

⁰ And, obviously, if $g(\underline{u}) = o(\underline{u})$ and $h(\underline{u}) = o(\underline{u})$ as $\underline{u} \rightarrow \underline{0}$, then $(f+g)(\underline{u}) = o(\underline{u})$ as $\underline{u} \rightarrow \underline{0}$ (See Theorem A5.2.1(iii)).

Thus, f is differentiable at $\underline{x}$ and its derivative is given by

$$D\alpha(\underline{x})[\underline{u}]\beta(\underline{x}) + \alpha(\underline{x})D\beta(\underline{x})[\underline{u}].$$

Since the functions involved here are real-valued, the notion of gradient makes sense, and

$$D\alpha(\underline{x})[\underline{u}]\beta(\underline{x}) + \alpha(\underline{x})D\beta(\underline{x})[\underline{u}]$$

$$= \nabla\alpha(\underline{x}) \cdot \underline{u} \beta(\underline{x}) + \alpha(\underline{x}) \nabla\beta(\underline{x}) \cdot \underline{u}$$

$$= [\beta(\underline{x}) \nabla\alpha(\underline{x}) + \alpha(\underline{x}) \nabla\beta(\underline{x})] \cdot \underline{u}.$$

In summary, we have

Theorem A5.6.3. Let $\alpha: D \rightarrow \mathbb{R}$ and $\beta: D \rightarrow \mathbb{R}$ both be
differentiable at $\underline{x} \in D$, Then the product function $f: D \rightarrow \mathbb{R}$
defined by

$$f(\underline{x}) = \alpha(\underline{x})\beta(\underline{x})$$

is differentiable at $\underline{x}$, and

$$Df(\underline{x})[\underline{u}] = D\alpha(\underline{x})[\underline{u}]\beta(\underline{x}) + \alpha(\underline{x})D\beta(\underline{x})[\underline{u}]$$

and

$$\nabla f(\underline{x}) = \beta(\underline{x}) \nabla\alpha(\underline{x}) + \alpha(\underline{x}) \nabla\beta(\underline{x}).$$

Do not miss the fact that the domain space $\mathcal{U}$ in the

above theorem is an arbitrary finite-dimensional vector space. In particular, if $\mathcal{U}$ should be $\text{Lin } V_n$ so that we are considering real-valued functions of a second-order tensor, the result is identical, except that we would probably write $\underline{\underline{X}}$ instead of $\underline{x}$. When $\mathcal{U} = \mathbb{R}$, we get the product rule of elementary calculus.

The above approach works on all products; e.g., see the text by Gurtin which dispatches all of the standard products in one fell swoop. We shall practice by considering a few more of them individually.

Theorem A5.6.4. Let $\alpha: D \rightarrow \mathbb{R}$ and $\underline{v}: D \rightarrow V$ both
be differentiable at $\underline{x} \in D$. Then the product function $\underline{f}: D \rightarrow V$
defined by

$$\underline{f}(\underline{x}) = \alpha(\underline{x}) \underline{v}(\underline{x})$$

is differentiable at $\underline{x}$, and

$$\begin{aligned} D\underline{f}(\underline{x})[\underline{u}] &= D\alpha(\underline{x})[\underline{u}] \underline{v}(\underline{x}) + \alpha(\underline{x}) D\underline{v}(\underline{x})[\underline{u}] \\ &= \{\nabla \alpha(\underline{x}) \cdot \underline{u}\} \underline{v}(\underline{x}) + \alpha(\underline{x}) D\underline{v}(\underline{x})[\underline{u}]. \end{aligned}$$

Proof, Exercise A5.6.3. $\square$

Exercise A5.6.4. Show that if $\mathcal{U} = V$, then result of the
above theorem can be expressed as

$$D\underline{f}(\underline{x}) = \underline{v}(\underline{x}) \otimes \nabla \alpha(\underline{x}) + \alpha(\underline{x}) D\underline{v}(\underline{x}),$$

Again we note that by the arbitrariness of the domain space $\mathcal{U}$, the functions $\underline{f}$, α , and $\underline{v}$ above could just as well depend on a second-order tensor with no formal change in the result. Similarly, the range space $\mathcal{V}$ could be $\text{Lin } \mathcal{V}_n$ (with $\mathcal{V}_n$ not necessarily $\mathcal{U}$), which would lead to

$$\underline{F}(\underline{x}) = \alpha(\underline{x}) \underline{V}(\underline{x})$$

with

$$D\underline{F}(\underline{x})[\underline{u}] = D\alpha(\underline{x})[\underline{u}] \underline{V}(\underline{x}) + \alpha(\underline{x}) D\underline{V}(\underline{x})[\underline{u}].$$

The derivatives of some more complicated products are given in

Theorem A5.6.5. Let $\underline{v}: \mathcal{O} \rightarrow \mathcal{V}$ and $\underline{w}: \mathcal{O} \rightarrow \mathcal{V}$ both be differentiable at $\underline{x}$. Then the product functions $\underline{f}$, $\underline{F}$, and $\underline{f}$ defined below are each differentiable at $\underline{x}$, and their derivatives are given by the stated formulas.

(i) scalar product, $\underline{f}: \mathcal{O} \rightarrow \mathbb{R}$ defined by $\underline{f}(\underline{x}) = \underline{v}(\underline{x}) \cdot \underline{w}(\underline{x})$,

$$D\underline{f}(\underline{x})[\underline{u}] = D\underline{v}(\underline{x})[\underline{u}] \cdot \underline{w}(\underline{x}) + \underline{v}(\underline{x}) \cdot D\underline{w}(\underline{x})[\underline{u}]$$

and when $\mathcal{U} = \mathcal{V}$

$$\nabla \underline{f}(\underline{x}) = \{D\underline{v}(\underline{x})\}^T [\underline{w}(\underline{x})] + \{D\underline{w}(\underline{x})\}^T [\underline{v}(\underline{x})];$$

(ii) tensor product, $\underline{F}: \mathcal{O} \rightarrow \text{Lin } \mathcal{V}$ defined by $\underline{F}(\underline{x}) = \underline{v}(\underline{x}) \otimes \underline{w}(\underline{x})$,

$$D\underline{F}(\underline{x})[\underline{u}] = D\underline{v}(\underline{x})[\underline{u}] \otimes \underline{w}(\underline{x}) + \underline{v}(\underline{x}) \otimes D\underline{w}(\underline{x})[\underline{u}];$$

(iii) vector product, $\underline{f}: \mathcal{O} \rightarrow \mathcal{V}_3$ defined by $\underline{f}(\underline{x}) = \underline{v}(\underline{x}) \times \underline{w}(\underline{x})$,¹

¹ Here, $\mathcal{V} = \mathcal{V}_3$, and $\mathcal{V}_3$ must be oriented,

$$Df(\underline{x})[\underline{u}] = D\underline{v}(\underline{x})[\underline{u}] \times \underline{w}(\underline{x}) + \underline{v}(\underline{x}) \times D\underline{w}(\underline{x})[\underline{u}].$$

Proof. (i) By the differentiability of $\underline{v}$ and $\underline{w}$ at $\underline{x}$,

$$\begin{aligned} f(\underline{x} + \underline{u}) - f(\underline{x}) &= \{ \underline{v}(\underline{x}) + D\underline{v}(\underline{x})[\underline{u}] + o(\underline{u}) \} \cdot \{ \underline{w}(\underline{x}) + D\underline{w}(\underline{x})[\underline{u}] + o(\underline{u}) \} - \underline{v}(\underline{x}) \cdot \underline{w}(\underline{x}) \\ &= D\underline{v}(\underline{x})[\underline{u}] \cdot \underline{w}(\underline{x}) + \underline{v}(\underline{x}) \cdot D\underline{w}(\underline{x})[\underline{u}] + D\underline{v}(\underline{x})[\underline{u}] \cdot D\underline{w}(\underline{x})[\underline{u}] \\ &\quad + o(\underline{u}) \cdot \{ \underline{v}(\underline{x}) + \underline{w}(\underline{x}) \} + o(\underline{u}) \cdot \{ D\underline{v}(\underline{x})[\underline{u}] + D\underline{w}(\underline{x})[\underline{u}] \} \\ &\quad + o(\underline{u}) \cdot o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}. \end{aligned}$$

The first two terms define a linear transformation from $\mathcal{U}$ into $\mathbb{R}$. The remaining terms each go to zero faster than $\underline{u}$. The least obvious of these is

$$D\underline{v}(\underline{x})[\underline{u}] \cdot D\underline{w}(\underline{x})[\underline{u}],$$

However, by the inner product-^(magnitude) inequality² and then the boundedness of linear transformations in the finite-dimensional case³,

$$|D\underline{v}(\underline{x})[\underline{u}] \cdot D\underline{w}(\underline{x})[\underline{u}]| \leq |D\underline{v}(\underline{x})[\underline{u}]| |D\underline{w}(\underline{x})[\underline{u}]| \leq (V|\underline{u}|)(W|\underline{u}|)$$

¹ Of course, the little o terms here are vector-valued.

² Theorem A2.3.7.

³ See Theorem A3.4.2.

for some numbers V and W which are independent of $\underline{u}$. Thus,

$$\frac{|D\underline{V}(\underline{x})[\underline{u}] \cdot D\underline{W}(\underline{x})[\underline{u}]|}{|\underline{u}|} \leq VW|\underline{u}| \rightarrow 0 \text{ as } \underline{u} \rightarrow 0.$$

Hence, f is differentiable at $\underline{x}$, and its derivative is given by

$$Df(\underline{x})[\underline{u}] = D\underline{V}(\underline{x})[\underline{u}] \cdot \underline{w}(\underline{x}) + \underline{v}(\underline{x}) \cdot D\underline{W}(\underline{x})[\underline{u}].$$

Since f is real-valued,

$$Df(\underline{x})[\underline{u}] = \nabla f(\underline{x}) \cdot \underline{u}.$$

If $\mathcal{U} = \mathcal{V}$, then $D\underline{V}(\underline{x})$ is a linear transformation on $\mathcal{U}$ into itself; i.e., $D\underline{V}(\underline{x})$ is a second-order tensor. Since $\mathcal{U}$ is finite-dimensional, $D\underline{V}(\underline{x})$ has a transpose; and

$$D\underline{V}(\underline{x})[\underline{u}] \cdot \underline{w}(\underline{x}) = \{D\underline{V}(\underline{x})\}^T [\underline{w}(\underline{x})] \cdot \underline{u},$$

Similarly, for $D\underline{W}(\underline{x})$, and $\therefore$

$$\nabla f(\underline{x}) \cdot \underline{u} = \left(\{D\underline{V}(\underline{x})\}^T [\underline{w}(\underline{x})] + \{D\underline{W}(\underline{x})\}^T [\underline{v}(\underline{x})] \right) \cdot \underline{u}.$$

Since $\underline{u}$ is arbitrary, we get the claimed expression for the gradient.

(ii) Exercise A5.6.5. (The identity $|\underline{a} \otimes \underline{b}| = |\underline{a}| |\underline{b}|$ (see

Theorem A3.7.4(v) will help you with the estimates.)

(iii) Exercise A5.6.6. $\square$

Again the underlying spaces U and V in the above theorem are arbitrary finite-dimensional spaces (except in part (iii)). If, e.g., in (i) $U = V = \text{Lin } V_n$, we would have

$$f(\underline{x}) = \underline{V}(\underline{x}) \cdot \underline{W}(\underline{x})$$

with

$$Df(\underline{x})[\underline{u}] = D\underline{V}(\underline{x})[\underline{u}] \cdot \underline{W}(\underline{x}) + \underline{V}(\underline{x}) \cdot D\underline{W}(\underline{x})[\underline{u}],$$

Insert material on p. A5.6.9a

Of course, a tensor operating on a vector can be treated as a product, and this leads us to

Theorem A5.6.6. Let $\underline{T}: D \rightarrow \text{Lin } V_n$ and $\underline{v}: D \rightarrow V_n$ both be
differentiable at $\underline{x}$. Then the product function $\underline{f}: D \rightarrow V_n$
defined by

$$\underline{f}(\underline{x}) = \underline{T}(\underline{x}) \underline{v}(\underline{x})$$

is differentiable at $\underline{x}$, and

$$D\underline{f}(\underline{x})[\underline{u}] = D\underline{T}(\underline{x})[\underline{u}] \underline{v}(\underline{x}) + \underline{T}(\underline{x}) D\underline{v}(\underline{x})[\underline{u}].$$

Proof. By the differentiability of $\underline{T}$ and $\underline{v}$ at $\underline{x}$,

¹Of course, the domain space U ($D \subset U$) need not be V_n .

INSERT for p.A5.6.9

As a slightly more involved special case of Theorem A5.6.5, suppose that the domain space $\mathcal{U} = \mathbb{R}$ in (i). Then

$$f(x) = \underline{v}(x) \cdot \underline{w}(x),$$

and

$$Df(x)[u] = D\underline{v}(x)[u] \cdot \underline{w}(x) + \underline{v}(x) \cdot D\underline{w}(x)[u].$$

But by Theorem A5.3.1, in the real variable case,

$$Df(x)[u] = u f'(x), \text{ etc. ;}$$

and we recover the familiar elementary result that

$$f'(x) = \underline{v}'(x) \cdot \underline{w}(x) + \underline{v}(x) \cdot \underline{w}'(x).$$

Similarly, if $\mathcal{U} = \mathbb{R}$ in (iii), so that $\underline{f}: \mathcal{D} \rightarrow \mathcal{V}_3$ is defined by

$$\underline{f}(x) = \underline{v}(x) \times \underline{w}(x),$$

then

$$\underline{f}'(x) = \underline{v}'(x) \times \underline{w}(x) + \underline{v}(x) \times \underline{w}'(x).$$

return to p. A5.6.9

$$\begin{aligned}
\underline{f}(\underline{x} + \underline{u}) - \underline{f}(\underline{x}) &= \{ \underline{T}(\underline{x}) + D\underline{T}(\underline{x})[\underline{u}] + \hat{\underline{T}}(\underline{u}) \} \{ \underline{v}(\underline{x}) + D\underline{v}(\underline{x})[\underline{u}] + \hat{\underline{v}}(\underline{u}) \} - \underline{T}(\underline{x}) \underline{v}(\underline{x}) \\
&= D\underline{T}(\underline{x})[\underline{u}] \underline{v}(\underline{x}) + \underline{T}(\underline{x}) D\underline{v}(\underline{x})[\underline{u}] + D\underline{T}(\underline{x})[\underline{u}] D\underline{v}(\underline{x})[\underline{u}] \\
&\quad + \underline{T}(\underline{x}) \hat{\underline{v}}(\underline{u}) + D\underline{T}(\underline{x})[\underline{u}] \hat{\underline{v}}(\underline{u}) + \hat{\underline{T}}(\underline{u}) \underline{v}(\underline{x}) + \hat{\underline{T}}(\underline{u}) D\underline{v}(\underline{x})[\underline{u}] \\
&\quad + \hat{\underline{T}}(\underline{u}) \hat{\underline{v}}(\underline{u}),
\end{aligned}$$

where

$$\hat{\underline{T}}(\underline{u}) = o(\underline{u}), \quad \hat{\underline{v}}(\underline{u}) = o(\underline{u}) \quad \text{as } \underline{u} \rightarrow \underline{0}.$$

We are using this slightly different style in handling the two little o terms because one is tensor-valued while the other is vector-valued.

The first two terms in the above expression for $\underline{f}(\underline{x} + \underline{u}) - \underline{f}(\underline{x})$ define a linear transformation from $\mathcal{U}$ into $\mathcal{V}_n$. The remaining terms each go to zero faster than $\underline{u}$. The least obvious of these is

$$D\underline{T}(\underline{x})[\underline{u}] D\underline{v}(\underline{x})[\underline{u}],$$

This quantity is not as formidable as it appears at first glance. $D\underline{T}(\underline{x})[\underline{u}]$ is an element of $\text{Lin } \mathcal{V}_n$, and $D\underline{v}(\underline{x})[\underline{u}]$ is an element of $\mathcal{V}_n$; i.e., $D\underline{T}(\underline{x})[\underline{u}]$ is simply a second-order tensor which maps $D\underline{v}(\underline{x})[\underline{u}]$ into some other element of $\mathcal{V}_n$. Because of the finite-dimensionality, $D\underline{T}(\underline{x})[\underline{u}]$ is necessarily bounded; and by the definition of the operator norm (Definition A3.4.5),

$$|D\underline{T}(\underline{x})[\underline{u}] D\underline{v}(\underline{x})[\underline{u}]| \leq \|D\underline{T}(\underline{x})[\underline{u}]\| |D\underline{v}(\underline{x})[\underline{u}]|.$$

Similarly, $D_V(x) \in \text{Lin } V_n$ is also bounded. $\therefore$

$$|D_V(x)[u]| \leq \|D_V(x)\| |u|,$$

and the previous inequality becomes

$$|D_T(x)[u] D_V(x)[u]| \leq \|D_T(x)[u]\| \|D_V(x)\| |u|.$$

We need to examine the dependence of $\|D_T(x)[u]\|$ on u . By the definition of derivative $D_T(x) \in \text{Lin}(V_n, \text{Lin } V_n)$. In fact,

$$D_T(x) \in \text{Bd}(V_n, \text{Bd}(V_n, V_n))$$

by the finite-dimensionality. If we endow the linear space $\text{Bd}(V_n, V_n)$ with the operator norm¹, then application of the definition of operator norm to $D_T(x)$ yields

$$\|D_T(x)[u]\| \leq \|D_T(x)\| |u|;$$

whence,

$$|D_T(x)[u] D_V(x)[u]| \leq \|D_T(x)\| \|D_V(x)\| |u|^2.$$

Then

$$\frac{|D_T(x)[u] D_V(x)[u]|}{|u|} \leq \|D_T(x)\| \|D_V(x)\| |u| \rightarrow 0 \text{ as } u \rightarrow 0.$$

Thus, f is differentiable at x , and its derivative is given by

¹cf. Theorem A3.4.7.

the stated formula. $\square$

As our last formal example of the differentiation of a product, we have

Theorem A5.6.7. Let $\underline{S}: \mathcal{D} \rightarrow \text{Lin } V_n$ and $\underline{T}: \mathcal{D} \rightarrow \text{Lin } V_n$
both be differentiable at $\underline{x}$. Then the product function
 $\underline{F}: \mathcal{D} \rightarrow \text{Lin } V_n$ defined by

$$\underline{F}(\underline{x}) = \underline{S}(\underline{x}) \underline{T}(\underline{x})$$

is differentiable at $\underline{x}$, and

$$D\underline{F}(\underline{x})[\underline{u}] = D\underline{S}(\underline{x})[\underline{u}] \underline{T}(\underline{x}) + \underline{S}(\underline{x}) D\underline{T}(\underline{x})[\underline{u}].$$

Proof. Exercise A5.6.7. (The inequality $|\underline{A}\underline{B}| \leq |\underline{A}||\underline{B}|$
 (see Theorem A3.7.9) will be helpful with the estimates.) $\square$

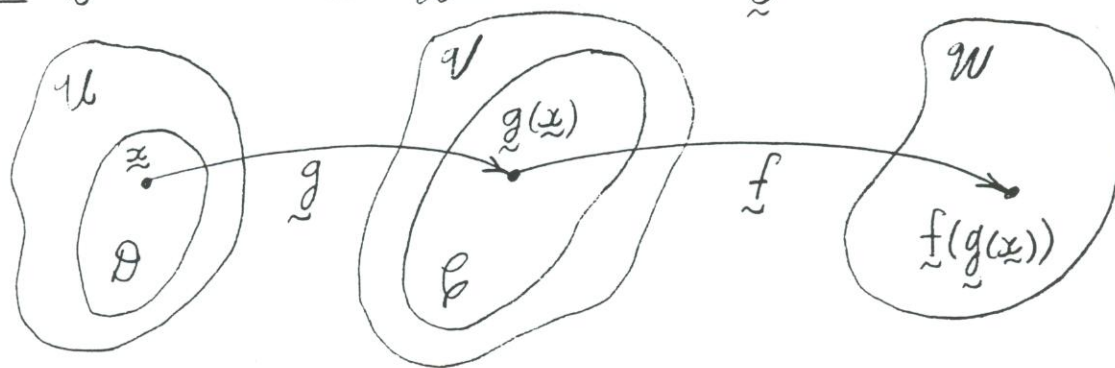
Supplementary Reading

Same as for §A5.2

A5.7. Derivatives of Composite Functions. The Chain Rule

Often in continuum mechanics, we need to differentiate the composition of two functions. This is accomplished via

Theorem A5.7.1. (Chain Rule) Let $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be finite-dimensional linear spaces. Let $\mathcal{D}$ be an open subset of $\mathcal{U}$ and $\mathcal{C}$ an open subset of $\mathcal{V}$. Let $g: \mathcal{D} \rightarrow \mathcal{C}$ be differentiable at $\underline{x} \in \mathcal{D}$ and $f: \mathcal{C} \rightarrow \mathcal{W}$ be differentiable at $g(\underline{x})$.



Then the composite function $\underline{h} = f \circ g: \mathcal{D} \rightarrow \mathcal{W}$ defined by

$$\underline{h}(\underline{x}) = f(g(\underline{x})) \quad , \quad \underline{x} \in \mathcal{D}$$

is differentiable at $\underline{x}$, and

$$D\underline{h}(\underline{x}) = Df(g(\underline{x})) \circ Dg(\underline{x})$$

or

$$D\underline{h}(\underline{x})[\underline{u}] = Df(g(\underline{x}))[Dg(\underline{x})[\underline{u}]] \quad ,$$

Proof. In accordance with Definition A5.2.2, consider

$$\begin{aligned}
h(\underline{x} + \underline{u}) - h(\underline{x}) &= f(g(\underline{x} + \underline{u})) - f(g(\underline{x})) \\
&= f(g(\underline{x}) + \underbrace{Dg(\underline{x})[\underline{u}] + o(\underline{u})}_{\underline{v}}) - f(g(\underline{x})) \\
&= f(g(\underline{x})) + Df(g(\underline{x}))[\underline{v}] + o(\underline{v}) - f(g(\underline{x})) \\
&= Df(g(\underline{x}))[Dg(\underline{x})[\underline{u}] + o(\underline{u})] + o(\underline{v}) \\
&= Df(g(\underline{x}))[Dg(\underline{x})[\underline{u}]] + Df(g(\underline{x}))[o(\underline{u})] + o(\underline{v}) \\
&\hspace{15em} \text{as } \underline{u} \rightarrow \underline{0}.
\end{aligned}$$

The first term defines a linear transformation on $\mathcal{U}$ into $\mathcal{W}$, which in fact is the claimed derivative of h . Thus, all that remains is to show that

$$Df(g(\underline{x}))[o(\underline{u})] + o(\underline{v}) = o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

This constitutes Exercise A5.7.1. $\square$

Supplementary Reading

Same as for §A5.2

A5.8. Same Additional Derivatives. The Invariants and the Inverse of a Tensor

We now have assembled sufficient machinery to enable us to differentiate most of the functions that will arise in our introduction to continuum mechanics. As a first example of how this machinery can be used, we consider the derivative of the determinant of a second-order tensor. This, of course, entails thinking of the determinant as a scalar-valued function of a tensor argument. The natural place to start is the definition of the determinant in terms of a volume orientation function, but this is rather messy. Instead, we look to other settings in which the determinant makes an appearance. One such setting is the Hamilton-Cayley theorem (Theorem A3.14.3),

which states that a tensor $\underline{X} \in \text{Lin } V_3$ satisfies its own characteristic equation:

$$-\underline{X}^3 + \ell_1(\underline{X})\underline{X}^2 - \ell_2(\underline{X})\underline{X} + \ell_3(\underline{X})\underline{I} = \underline{0},$$

where the coefficients

$$\ell_1(\underline{X}) = \text{tr } \underline{X},$$

$$\ell_2(\underline{X}) = \frac{1}{2}[(\text{tr } \underline{X})^2 - \text{tr } (\underline{X}^2)] = \frac{1}{2}[\ell_1(\underline{X})]^2 - \frac{1}{2} \ell_1(\underline{X}^2),$$

and

$$\ell_3(\underline{X}) = \det \underline{X}$$

are the fundamental invariants of $\underline{X}$. Taking the trace yields

$$3 \ell_3(\underline{X}) = \ell_1(\underline{X}^3) - \ell_1(\underline{X}) \ell_1(\underline{X}^2) + \ell_1(\underline{X}) \ell_2(\underline{X}). \quad (\#)$$

Now each of the terms on the right-hand side of (#) can be differentiated in terms of known derivatives by the chain rule or the chain rule combined with a product rule. We go through this step by step.

Since $\ell_1: \text{Lin } V_3 \rightarrow \mathbb{R}$ is just the trace function, we have

$$D\ell_1(\underline{X})[\underline{U}] = \text{tr } \underline{U} = \ell_1(\underline{U})$$

¹As usual, V_3 is a 3-dimensional inner product space, and we use the inner product norm topology.

by Theorem A5.5.4 (v).

Consider the first term in (#):

$$\varphi(\underline{x}^3) = h(\underline{x}) = \varphi(\underline{G}(\underline{x})), \quad (h = \varphi \circ \underline{G})$$

where

$$\underline{G}: \text{Lin } \mathcal{V}_3 \rightarrow \text{Lin } \mathcal{V}_3, \quad \underline{G}(\underline{x}) = \underline{x}^3.$$

By Theorems A5.7.1, A5.5.4 (v), and A5.6.2, h is differentiable and

$$Dh(\underline{x})[\underline{U}] = D\varphi(\underline{G}(\underline{x}))[D\underline{G}(\underline{x})[\underline{U}]]$$

$$= \text{tr} \{ D\underline{G}(\underline{x})[\underline{U}] \}$$

$$= \text{tr} (\underline{x}^2 \underline{U} + \underline{x} \underline{U} \underline{x} + \underline{U} \underline{x}^2) = 3 \text{tr} (\underline{x}^2 \underline{U}) = 3 \varphi(\underline{x}^2 \underline{U});$$

the last step follows from the linearity of the trace function and Theorem A3.7.2 (ii).

The second term in (#) is the product $f: \text{Lin } \mathcal{V}_3 \rightarrow \mathbb{R}$ defined by

$$f(\underline{x}) = \varphi(\underline{x}) \varphi(\underline{x}^2).$$

The first factor is just the trace function, whose derivative we know. The second factor $\varphi(\underline{x}^2)$ can be handled as was $\varphi(\underline{x}^3)$ above. Thus, we write

$$\varphi(\underline{x}^2) = g(\underline{x}) = \varphi(\underline{H}(\underline{x})), \quad (g = \varphi \circ \underline{H})$$

where

$$H(\underline{x}) = \underline{x}^2.$$

Then, since H and ι , are differentiable, it follows from the chain rule that so is g , and

$$\begin{aligned} Dg(\underline{x})[\underline{u}] &= D\iota(H(\underline{x})) [DH(\underline{x})[\underline{u}]] \\ &= \text{tr} \{ DH(\underline{x})[\underline{u}] \} \\ &= \text{tr} (\underline{x}\underline{u} + \underline{u}\underline{x}) = 2\text{tr}(\underline{x}\underline{u}) = 2I(\underline{x}\underline{u}) \end{aligned}$$

Then by the product rule (Theorem A5.6.3), f is differentiable; and

$$\begin{aligned} Df(\underline{x})[\underline{u}] &= D\iota(\underline{x})[\underline{u}] g(\underline{x}) + \iota(\underline{x}) Dg(\underline{x})[\underline{u}] \\ &= (\text{tr } \underline{u}) \text{tr}(\underline{x}^2) + 2(\text{tr } \underline{x}) \text{tr}(\underline{x}\underline{u}) \\ &= \iota_1(\underline{x}^2) \iota_1(\underline{u}) + 2\iota_1(\underline{x}) \iota_1(\underline{x}\underline{u}). \end{aligned}$$

The third term in (#) is the product $j: \text{Lin } V_3 \rightarrow \mathbb{R}$ defined by

$$j(\underline{x}) = \iota_1(\underline{x}) \iota_2(\underline{x}),$$

where in terms of the above notation

$$\iota_2(\underline{x}) = \frac{1}{2} [\iota_1(\underline{x})]^2 - \frac{1}{2} g(\underline{x}).$$

The second term in the above expression for $\ell_2(\underline{x})$ has a known derivative. The first term can be handled by either the product rule or the chain rule. Use of the product rule is easier, but use of the chain rule is more instructive, and so we use the chain rule. Write

$$[\ell_1(\underline{x})]^2 = k(\underline{x}) = l(\ell_1(\underline{x})) \quad , \quad (k = l \circ I)$$

where

$$l: \mathbb{R} \rightarrow \mathbb{R} \quad , \quad l(x) = x^2.$$

By Theorem A5.3.1,

$$Dl(x)[u] = l'(x)u = 2xu.$$

$\therefore$ by the chain rule, k is differentiable; and

$$\begin{aligned} Dk(\underline{x})[\underline{u}] &= Dl(\ell_1(\underline{x}))[D\ell_1(\underline{x})[\underline{u}]] \\ &= 2\ell_1(\underline{x})D\ell_1(\underline{x})[\underline{u}] \\ &= 2\ell_1(\underline{x})\ell_1(\underline{u}), \end{aligned}$$

which is quite obviously the result that the product rule would have given. Thus, ℓ_2 is differentiable and

$$\begin{aligned} D\ell_2(\underline{x})[\underline{u}] &= \frac{1}{2} Dk(\underline{x})[\underline{u}] - \frac{1}{2} Dg(\underline{x})[\underline{u}] \\ &= \ell_1(\underline{x})\ell_1(\underline{u}) - \ell_1(\underline{x}\underline{u}). \end{aligned}$$

Finally, the product j is differentiable; and

$$Dj(\underline{x})[\underline{u}] = \ell_1(\underline{u}) \ell_2(\underline{x}) + \ell_1(\underline{x}) \{ \ell_1(\underline{x}) \ell_1(\underline{u}) - \ell_1(\underline{xu}) \}$$

Returning to (#) and assembling the above results via Theorem A5.5.1, we have

$$D \ell_3(\underline{x})[\underline{u}] = \ell_1(\underline{x}^2 \underline{u}) - \ell_1(\underline{x}) \ell_1(\underline{xu}) + \ell_2(\underline{x}) \ell_1(\underline{u}).$$

But ℓ_1 is just the trace function, and $\therefore$ we can alternatively write

$$\begin{aligned} D \ell_3(\underline{x})[\underline{u}] &= \text{tr} \{ [\underline{x}^2 - \ell_1(\underline{x}) \underline{x} + \ell_2(\underline{x}) \underline{1}] \underline{u} \} \\ &= [\underline{x}^2 - \ell_1(\underline{x}) \underline{x} + \ell_2(\underline{x}) \underline{1}]^T \cdot \underline{u}. \end{aligned}$$

The last expression makes it clear that

$$\nabla \ell_3(\underline{x}) = [\underline{x}^2 - \ell_1(\underline{x}) \underline{x} + \ell_2(\underline{x}) \underline{1}]^T,$$

In view of the Hamilton-Cayley theorem, the term in brackets above is just $\ell_3(\underline{x}) \underline{x}^{-1}$, provided, of course, that $\underline{x}$ is invertible.¹

In summary, we have established

¹Cf. Theorem A3.14.4.

Theorem A5.8.1. Let $\underline{X} \in \text{Lin } \mathcal{V}_3$.)¹ Then its fundamental
invariants,

$$I_1(\underline{X}) = \text{tr } \underline{X},$$

$$I_2(\underline{X}) = \frac{1}{2} [(\text{tr } \underline{X})^2 - \text{tr } (\underline{X}^2)],$$

and

$$I_3(\underline{X}) = \det \underline{X},$$

are each differentiable everywhere on $\text{Lin } \mathcal{V}_3$; and their
derivatives and gradients are given by

$$D I_1(\underline{X})[\underline{U}] = I_1(\underline{U})$$

$$\nabla I_1(\underline{X}) = \underline{I},$$

$$D I_2(\underline{X})[\underline{U}] = I_1(\underline{X}) I_1(\underline{U}) - I_1(\underline{X} \underline{U}),$$

$$\nabla I_2(\underline{X}) = [I_1(\underline{X}) \underline{I} - \underline{X}]^T,$$

$$D I_3(\underline{X})[\underline{U}] = I_3(\underline{X}^2 \underline{U}) - I_1(\underline{X}) I_3(\underline{X} \underline{U}) + I_2(\underline{X}) I_1(\underline{U}),$$

$$\nabla I_3(\underline{X}) = [\underline{X}^2 - I_1(\underline{X}) \underline{X} + I_2(\underline{X}) \underline{I}]^T.$$

When $\underline{X}$ is invertible, then the last two expressions simplify to

¹Of course, there is a corresponding result for the elements of $\text{Lin } \mathcal{V}_n$
see D. E. Carlson and A. Hoger, "On the derivatives of the principal
invariants of a second-order tensor," J. Elasticity 16 (1986) 221-224.

$$D\ell_3(\underline{X})[\underline{U}] = \ell_3(\underline{X})\ell_3(\underline{X}^{-1}\underline{U}),$$

$$\nabla \ell_3(\underline{X}) = \ell_3(\underline{X})(\underline{X}^{-1})^T.$$

Exercise A5.8.1. Actually, we did not work out the expression for $\nabla \ell_2(\underline{X})$ in the proof. Do so.

The Hamilton-Cayley theorem also can be used to differentiate the tensor inverse function. When $\underline{X} \in \text{Lin } V_3$ is invertible, this theorem implies that

$$\underline{X}^{-1} = [\ell_3(\underline{X})]^{-1} [\underline{X}^2 - \ell_1(\underline{X})\underline{X} + \ell_2(\underline{X})\underline{I}];$$

and we know how to differentiate the expression on the right-hand side. Before carrying this any further, there is a technical detail to consider. The domain of the function $\underline{X} \mapsto \underline{X}^{-1}$ is not all of $\text{Lin } V_3$, but rather it is just the subset of invertible tensors. Our concern here is that the definition of differentiability applies to a function whose domain is an open subset of the domain space. We need

Theorem A5.8.2. Let

$$\text{Inv } V_n = \{ \underline{T} \in \text{Lin } V_n : \underline{T} \text{ is invertible} \},$$

Then $\text{Inv } V_n$ is an open subset of $\text{Lin } V_n$.

Proof. In view of Definition A5.1.3, we must show that every element of $\text{Inv } V_n$ is the center of an open ball which is contained in $\text{Inv } V_n$. I.e., we must show that for each $T_0 \in \text{Inv } V_n$ there is an $r_0 > 0 \Rightarrow$

$$\{T \in \text{Lin } V_n : |T - T_0| < r_0\} \subset \text{Inv } V_n.$$

Let $T_0 \in \text{Inv } V_n$. By Theorem A3.9.9, this is equivalent to

$$\det T_0 \neq 0.$$

Now we saw in the previous theorem that the determinant function is differentiable (at least for $n=3$). $\therefore$ by Theorem A5.2.4, the determinant function is continuous. We expect by the continuity that if $\det T_0 \neq 0$, then $\det T \neq 0$ throughout some sufficiently small neighborhood of T_0 . To actually see this, we recall that continuity of $\det T$ at T_0 means; Given any $\varepsilon > 0 \exists \delta(\varepsilon) > 0 \Rightarrow$

$$|T - T_0| < \delta \Rightarrow |\det T - \det T_0| < \varepsilon$$

$$\Rightarrow -\varepsilon < \det T - \det T_0 < \varepsilon$$

$$\Rightarrow \det T_0 - \varepsilon < \det T < \det T_0 + \varepsilon \quad (i)$$

$$\Rightarrow -\det T_0 + \varepsilon > -\det T > -\det T_0 - \varepsilon \quad (ii)$$

Since $\det T_0 \neq 0$, either $\det T_0 > 0$ or $\det T_0 < 0$. If

Let $\underline{T}_0 > 0$, choose $\varepsilon = \det \underline{T}_0$, call $0 < \delta(\det \underline{T}_0) = r_0$, and note by (i) that

$$|\underline{T} - \underline{T}_0| < r_0 \Rightarrow 0 < \det \underline{T} \Rightarrow \underline{T} \in \text{Inv } \mathcal{V}_n.$$

If $\det \underline{T}_0 < 0$, choose $\varepsilon = -\det \underline{T}_0$, call $0 < \delta(-\det \underline{T}_0) = r_0$, and note by (ii) that

$$|\underline{T} - \underline{T}_0| < r_0 \Rightarrow -\det \underline{T} > 0 \Rightarrow \underline{T} \in \text{Inv } \mathcal{V}_n.$$

Thus, for each $\underline{T}_0 \in \text{Inv } \mathcal{V}_n \exists r_0 > 0 \ni$

$$\{\underline{T} \in \text{Lin } \mathcal{V}_n : |\underline{T} - \underline{T}_0| < r_0\} \subset \text{Inv } \mathcal{V}_n;$$

i.e., $\text{Inv } \mathcal{V}_n$ is open in $\text{Lin } \mathcal{V}_n$. $\square$

Now we are ready for

Theorem A5.8.3. The inverse tensor function $\underline{F}: \text{Inv } \mathcal{V}_3 \rightarrow \text{Inv } \mathcal{V}_3$ defined by

$$\underline{F}(\underline{X}) = \underline{X}^{-1}$$

is differentiable everywhere on $\text{Inv } \mathcal{V}_3$, and

$$D\underline{F}(\underline{X})[\underline{U}] = -\underline{X}^{-1}\underline{U}\underline{X}^{-1}.$$

Proof. As suggested on p. A5.8.8, given our recent experience with the determinant function, a natural approach

is to use the Cayley-Hamilton theorem to write

$$\underline{F}(\underline{X}) = \underline{X}^{-1} = [\underline{\zeta}_3(\underline{X})]^{-1} [\underline{X}^2 - \underline{\zeta}_1(\underline{X})\underline{X} + \underline{\zeta}_2(\underline{X})\underline{I}] \quad (*)$$

and then use some of our previous results to differentiate this expression. The result of this calculation is

$$\begin{aligned} D\underline{F}(\underline{X})[\underline{U}] = & -\underline{\zeta}_1(\underline{X}^{-1}\underline{U})\underline{X}^{-1} + [\underline{\zeta}_3(\underline{X})]^{-1} [\underline{X}\underline{U} + \underline{U}\underline{X} - \underline{\zeta}_1(\underline{U})\underline{X} \\ & - \underline{\zeta}_1(\underline{X})\underline{U} + \underline{\zeta}_1(\underline{X})\underline{\zeta}_1(\underline{U})\underline{I} - \underline{\zeta}_1(\underline{X}\underline{U})\underline{I}], \quad (\text{Exercise AS.0.2}) \end{aligned}$$

which does not appear to be the claimed result. It can probably be manipulated into the claimed result by use of the Cayley-Hamilton theorem and Rivlin's identity¹

$$\begin{aligned} \underline{A}\underline{B}\underline{A} + \underline{A}^2\underline{B} + \underline{B}\underline{A}^2 - \underline{\zeta}_1(\underline{A})(\underline{A}\underline{B} + \underline{B}\underline{A}) - \underline{\zeta}_1(\underline{B})\underline{A}^2 \\ - [\underline{\zeta}_1(\underline{A}\underline{B}) - \underline{\zeta}_1(\underline{A})\underline{\zeta}_1(\underline{B})]\underline{A} + \underline{\zeta}_2(\underline{A})\underline{B} \\ - [\underline{\zeta}_1(\underline{A}^2\underline{B}) - \underline{\zeta}_1(\underline{A})\underline{\zeta}_1(\underline{A}\underline{B}) + \underline{\zeta}_1(\underline{B})\underline{\zeta}_1(\underline{A})]\underline{I} = \underline{0}, \end{aligned}$$

which holds for any $\underline{A}, \underline{B} \in \text{Lin } \mathcal{V}_3$ and is itself a consequence of the Cayley-Hamilton theorem. We can avoid this kind of effort by thinking a little more before we start calculating.

¹R.S. Rivlin, "Further remarks on the stress-separation relations for isotropic materials", J. Rational Mech. Anal. 4, 681-702 (1955).

In view of previous results, the formula (*) for $\underline{F}(\underline{X})$ makes it clear that $\underline{F}$ is differentiable. Next we write $\underline{X}^{-1}\underline{X} = \underline{I}$ as

$$\underline{F}(\underline{X})\underline{X} = \underline{I}.$$

Since both $\underline{F}$ and the identity map $\underline{X} \mapsto \underline{X}$ are differentiable, this equation can be differentiated via the product rule. The immediate result is

$$D\underline{F}(\underline{X})[\underline{U}] = -\underline{X}^{-1}\underline{U}\underline{X}^{-1}.$$

The details are left as Exercise A5.8.3. $\square$

The above formula for the derivative of the inverse has nothing in it to suggest that the underlying vector space is three-dimensional. In fact, the formula is the same when the underlying space is V_n . The same is true for the derivative of the determinant when the argument is invertible. In that case we had

$$D \iota_3(\underline{X})[\underline{U}] = \iota_3(\underline{X}) \iota_1(\underline{X}^{-1}\underline{U}).$$

Now the invariants $\iota_1(\underline{X})$ and $\iota_3(\underline{X})$ do suggest three-dimensions. However, we could just as well have written

$$D \det(\underline{X})[\underline{U}] = \det \underline{X} \operatorname{tr}(\underline{X}^{-1}\underline{U}),$$

Of course, the functions $\det$ and tr have meanings in V_n . Actually, if we consider the characteristic polynomial of $\underline{I} \in \mathcal{T}(V_n)$, we see that

$$\det(\underline{I} - \lambda \underline{I}) = (-\lambda)^n - \operatorname{tr} \underline{I} (-\lambda)^{n-1} + \dots + (-1)^n \det \underline{I}.$$

So that, independent of the dimension n , the first invariant is always the trace and the n^{th} invariant is always the determinant. For more on this see the paper cited in the footnote on p. A5.B.7 and also the encyclopedia article by Turesell and Noll listed in the references.

Unlike the determinant, the inverse of an operator is a meaningful notion in the infinite-dimensional case, where we do not have the Hamilton-Cayley theorem. The following exercise gives an alternative method of finding the derivative of the inverse which can be carried out for a bounded linear operator on an infinite-dimensional Banach space. We, of course, restrict our considerations to the finite-dimensional case.

Exercise A5.8.4. Define the inverse tensor function
 $\underline{F}: \text{Inv } V_n \rightarrow \text{Inv } V_n$ by
 $\underline{F}(\underline{X}) = \underline{X}^{-1}.$

Assuming that $\underline{F}$ is differentiable, we know by Exercise A5.8.3 that its derivative is given by

$$D\underline{F}(\underline{X})(\underline{U}) = -\underline{X}^{-1} \underline{U} \underline{X}^{-1}.$$

Forearmed with this knowledge, prove that $\underline{F}$ is, indeed, differentiable by showing that

$$\underline{F}(\underline{X} + \underline{U}) - \underline{F}(\underline{X}) + \underline{X}^{-1} \underline{U} \underline{X}^{-1} = o(\underline{U}) \text{ as } \underline{U} \rightarrow \underline{0}.$$

Supplementary Reading

Same as for §A5.2

A5.9. The Mean-Value Theorem for Derivatives

Here, we return to the abstract setting of $f: D \rightarrow V$ with D open in U , and with the finite-dimensional linear spaces U and V subject to the agreements of §A5.2.

A major tool for studying some of the less transparent properties of the derivative is provided by

Theorem A5.9.1. (Mean-Value Theorem for Derivatives) Let $f: D \rightarrow V$ be differentiable on D with the function $y \mapsto Df(y)$ continuous at each $y \in D$. For $x \in D$ and $u \in U$, let the line segment

$$\{x + \tau u : \tau \in [0, 1]\} \subset D,$$

Then

$$f(x+u) - f(x) = \int_0^1 Df(x+\tau u)[u] d\tau,$$

Proof. Consider the function $g: [0, 1] \rightarrow V$ defined by the composition

$$g(\tau) = f(h(\tau)),$$

where $h: [0, 1] \rightarrow D$ is defined by

$$h(\tau) = x + \tau u, \text{ for fixed } x \text{ and } u.$$

Note that

$$g(1) - g(0) = f(x+u) - f(x).$$

The idea of the proof is to integrate $g'(\tau)$ from 0 to 1

to get $\underline{g}(1) - \underline{g}(0)$ from the fundamental theorem of integral calculus.

Thus, we need $\underline{g}'(\tau)$. Immediately there is a slight problem with our $\sim$ differentiation machinery because the domain $[0,1]$ for $\underline{g}$ and $\underline{h}$ is not open. Let us then temporarily restrict our consideration to the interior $(0,1)$. We shall be sophisticated (or sloppy depending on your point of view) and not introduce new names for the restrictions¹ of the functions $\underline{g}$ and $\underline{h}$ to $(0,1)$. Since $\underline{h}(\tau) = \underline{x} + \tau \underline{u}$, $\underline{h}$ has an ordinary derivative on $(0,1)$ which is

$$\underline{h}'(\tau) = \underline{u},$$

Then by Theorem A5.3.1, $\underline{h}$ is differentiable on $(0,1)$, and

$$D\underline{h}(\tau)[\alpha] = \alpha \underline{h}'(\tau) = \alpha \underline{u}$$

Then since $\underline{f}$ was given as differentiable, it follows from the chain rule (Theorem A5.7.1) that $\underline{g}$ is differentiable on $(0,1)$, and

$$\begin{aligned} D\underline{g}(\tau)[\alpha] &= D\underline{f}(\underline{h}(\tau))[D\underline{h}(\tau)[\alpha]] \\ &= D\underline{f}(\underline{h}(\tau))[\alpha \underline{u}] = \alpha D\underline{f}(\underline{h}(\tau))[\underline{u}]. \end{aligned}$$

¹Of course, the restriction of $\underline{g}$ to $(0,1)$ is the function $\hat{\underline{g}}: (0,1) \rightarrow V$ defined by $\hat{\underline{g}}(\tau) = \underline{g}(\tau)$ for $\tau \in (0,1)$.

Then by Theorem A5.3.1, $\tilde{g}$ has an ordinary derivative on $(0,1)$, which is given by

$$\tilde{g}'(\tau) = D\tilde{f}(\tilde{x} + \tau\tilde{u})[\tilde{u}].$$

Now, what about the boundary points 0 and 1? There is no real difficulty with the ordinary derivative at boundary points; we just use one-sided limits. E.g., at 0 the derivative of $\tilde{g}$ is defined to be

$$\tilde{g}'(0) = \lim_{\substack{\alpha \rightarrow 0 \\ \alpha > 0}} \frac{1}{\alpha} [\tilde{g}(\alpha) - \tilde{g}(0)].$$

If the limit does not exist, then $\tilde{g}$ does not have a derivative at 0. From the definition of $\tilde{g}$ as $\tilde{f} \circ \tilde{h}$, we have that

$$\lim_{\substack{\alpha \rightarrow 0 \\ \alpha > 0}} \frac{1}{\alpha} [\tilde{g}(\alpha) - \tilde{g}(0)] = \lim_{\substack{x \rightarrow 0 \\ x > 0}} \frac{1}{\alpha} [\tilde{f}(\tilde{x} + \alpha\tilde{u}) - \tilde{f}(\tilde{x})].$$

But since $\tilde{f}$ is differentiable at $\tilde{x}$, it follows from Definition A5.2.3 and Theorem A5.2.3 that

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [\tilde{f}(\tilde{x} + \alpha\tilde{u}) - \tilde{f}(\tilde{x})] = \mathcal{D}\tilde{f}(\tilde{x}; \tilde{u}) = D\tilde{f}(\tilde{x})[\tilde{u}].$$

In particular,

$$\lim_{\substack{\alpha \rightarrow 0 \\ \alpha > 0}} \frac{1}{\alpha} [\tilde{f}(\tilde{x} + \alpha\tilde{u}) - \tilde{f}(\tilde{x})] = D\tilde{f}(\tilde{x})[\tilde{u}].$$

Thus, $\tilde{g}$ has an ordinary derivative at 0, and

$$\tilde{g}'(0) = D\tilde{f}(\tilde{x})[\tilde{u}].$$

Similarly,

$$\tilde{g}'(1) = \lim_{\substack{\alpha \rightarrow 0 \\ \alpha < 0}} \frac{1}{\alpha} [\tilde{f}(\tilde{x} + \tilde{u} + \alpha\tilde{u}) - \tilde{f}(\tilde{x} + \tilde{u})];$$

and since $\tilde{f}$ is differentiable at $\tilde{x} + \tilde{u} \in \mathcal{D}$, $\tilde{g}$ has an ordinary derivative at 1, and

$$\tilde{g}'(1) = D\tilde{f}(\tilde{x} + \tilde{u})[\tilde{u}],$$

Hence,

$$\text{at } \tau = 0, \tilde{g}'(0) = D\tilde{f}(\tilde{x})[\tilde{u}],$$

$$\text{for } \tau \in (0, 1), \tilde{g}'(\tau) = D\tilde{f}(\tilde{x} + \tau\tilde{u})[\tilde{u}],$$

$$\text{at } \tau = 1, \tilde{g}'(1) = D\tilde{f}(\tilde{x} + \tilde{u})[\tilde{u}].$$

By hypothesis,

the mapping $y \mapsto D\tilde{f}(y)$ is continuous on $\mathcal{D}$; $\therefore$

$$\tilde{g}'(\tau) = D\tilde{f}(\tilde{x} + \tau\tilde{u})[\tilde{u}] \text{ on } [0, 1],$$

moreover, the function $\tilde{g}' : [0, 1] \rightarrow V$ is continuous.

Then the fundamental theorem of integral calculus applies and

yields

$$\underset{\sim}{g}(1) - \underset{\sim}{g}(0) = \int_0^1 \underset{\sim}{g}'(\tau) d\tau = \int_0^1 D\underset{\sim}{f}(\underset{\sim}{x} + \tau \underset{\sim}{u})[\underset{\sim}{u}] d\tau. \quad \square$$

Of course, we would expect Theorem AS.9.1 to include the ordinary mean-value theorem of elementary differential calculus. To see that this is the case, let $\mathcal{U} = \mathcal{V} = \mathbb{R}$. Then by Theorem AS.3.1,

$$Df(x)[\alpha] = \alpha f'(x);$$

and Theorem AS.9.1 becomes

$$f(x+u) - f(x) = u \int_0^1 f'(x+\tau u) d\tau.$$

By the mean-value theorem for integrals, $\exists \bar{t} \in [0, 1] \ni$

$$\begin{aligned} \int_0^1 f'(x+\tau u) d\tau &= f'(x+\bar{t}u)(1-0) \\ &= f'(\bar{x}), \text{ where } \bar{x} := x + \bar{t}u. \end{aligned}$$

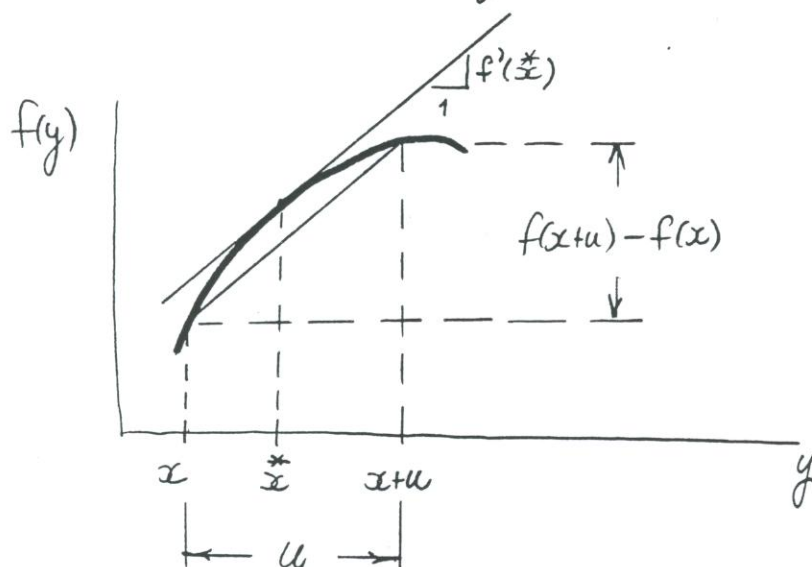
Thus, we have the special case

Theorem AS.9.2. (Mean-Value Theorem for Ordinary Derivatives)
Let $\mathcal{D} \subset \mathbb{R}$ be open. Let $f: \mathcal{D} \rightarrow \mathbb{R}$ have a continuous ordinary
derivative f' on $\mathcal{D}$. Let the closed interval $[x, x+u] \subset \mathcal{D}$, Then

$$\exists \xi^* \in [x, x+u] \Rightarrow$$

$$\frac{f(x+u) - f(x)}{u} = f'(\xi^*)$$

This result has the geometrical interpretation depicted below



We have already pointed out the obvious result that the derivative of a constant function is zero (Theorem A2.5.2). With suitable restrictions on the domain, the mean-value theorem for derivatives can be used to establish the converse. The restriction is that the domain be connected. The concept of connectivity has several formulations; we develop just one of them here.

First, thinking geometrically, we formalize some easy notions that we have used occasionally already.

Definition AS.9.1. Let V be a linear space. Let $u \in V$ with $u \neq 0$. Then the line spanned by u is

$$L_{sp}\{\underline{u}\} = \{\underline{v} : \underline{v} = \alpha \underline{u}, \alpha \in \mathbb{R}\}.$$

Let $\underline{x}_1, \underline{x}_2 \in V$, and write $\underline{u} = \underline{x}_2 - \underline{x}_1$. Then the line segment between $\underline{x}_1$ and $\underline{x}_2$ is the set

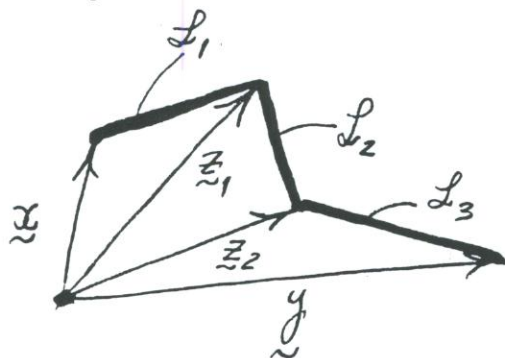
$$\{\underline{x} : \underline{x} = \underline{x}_1 + \tau \underline{u}, \tau \in [0, 1]\}.$$

Let $\underline{x}, \underline{z}_1, \underline{z}_2, \dots, \underline{z}_n, \underline{y} \in V$. Denote by L_1 the line segment between $\underline{x}$ and $\underline{z}_1$, by L_2 the line segment between $\underline{z}_1$ and $\underline{z}_2$, $\dots$, by L_{n+1} the line segment between $\underline{z}_n$ and $\underline{y}$. Then the set

$$L_1 \cup L_2 \cup \dots \cup L_{n+1}$$

is a polygonal line joining $\underline{x}$ and $\underline{y}$.

As an example of a polygonal line in V_2 , consider the following sketch.



Definition A5.9.2. A subset S of a linear space V is polygonally connected if any two elements of S can be joined by a polygonal line contained in S .

Now, we are prepared for

Theorem A5.9.3. Let f be differentiable on D and have zero derivative on D ; i.e., $\forall x \in D$

$$Df(x)[u] = 0 \quad \forall u \in U.$$

Then if D is polygonally connected, f is constant on D .

Proof. Let x denote an arbitrarily fixed element of D . By the polygonal connectivity, any other element $y \in D$ can be joined to x by a polygonal line which lies in D . Label the "vertices" by $z_1, z_2, \dots, z_n$ as in Definition A5.9.1.

Since the mapping $z \mapsto Df(z) = 0$ is obviously continuous, the mean-value theorem applies to f . Since the line segment from x to z_1 is contained in D ,

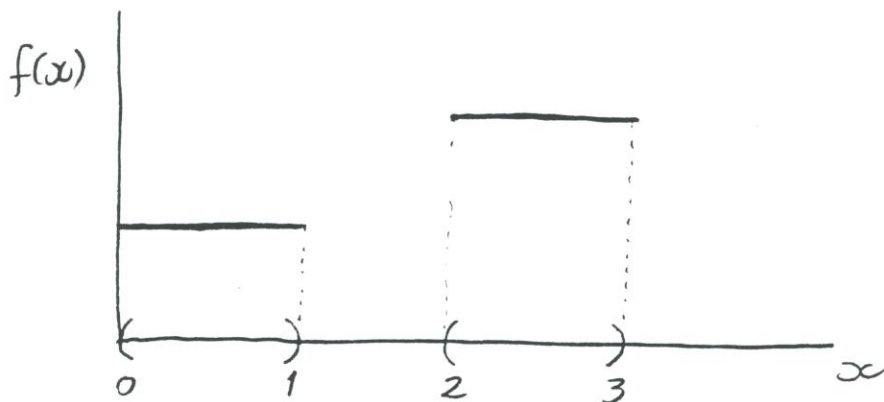
$$f(z_1) - f(x) = \int_0^1 Df(x + \tau(z_1 - x))[z_1 - x] d\tau = 0.$$

Thus, $f(z_1) = f(x)$. Similarly, $f(z_2) = f(z_1)$, etc. After a finite number of steps, we reach

$$f(y) = f(x)$$

By the arbitrariness of $y \in D$, f is constant on D . $\square$

The connectivity requirement in the above theorem sometimes is missed in applications. The 1-dimensional example depicted in the graph below demonstrates that the connectivity is, indeed, necessary.



Here, $f: (0, 1) \cup (2, 3) \rightarrow \mathbb{R}$ has a zero derivative everywhere, but it is not constant.

Supplementary Reading

Same as for §5.2

A 5.10. Higher Derivatives. The C^p Spaces

Suppose that $f: D \rightarrow V$ is differentiable on some open ball $B_r(\underline{x}) \subset D \subset U$. Then at each y in this ball, f has a derivative

$$Df(y) \in \text{Lin}(U, V).$$

Thus, we have a function

$$Df: B_r(\underline{x}) \rightarrow \text{Lin}(U, V).$$

Since $B_r(\underline{x})$ is open (Theorem A5.1.3) and since, as we shall see momentarily, $\text{Lin}(U, V)$ is a finite-dimensional linear space, we have the proper framework for considering the differentiability of Df . This leads us to

Definition A5.10.1. Let $f: D \rightarrow V$ be differentiable on some open ball $B_r(\underline{x}) \subset D \subset U$. If the function $Df: B_r(\underline{x}) \rightarrow \text{Lin}(U, V)$ is differentiable at $\underline{x}$, we denote its derivative there by

$$DDf(\underline{x}) = D^2f(\underline{x})$$

and call it the second derivative of f at $\underline{x}$.

The finite-dimensionality of $\text{Lin}(U, V)$ is taken care of in the next three exercises. We could have developed this much earlier, but

Of course, we continue to use the agreements of §5.2.

There was no need for such results then. Recall from Theorem A2.4.5 that as long as $\mathcal{U}$ and $\mathcal{V}$ are linear spaces, not necessarily finite-dimensional, then $\text{Lin}(\mathcal{U}, \mathcal{V})$ is a linear space.

Exercise A5.10.1. Let $\mathcal{U}$ be an inner product space and $\mathcal{V}$ a linear space. For $\underline{u} \in \mathcal{U}$ and $\underline{v} \in \mathcal{V}$, define the function $\underline{v} \otimes \underline{u} : \mathcal{U} \rightarrow \mathcal{V}$ by

$$(\underline{v} \otimes \underline{u})(\underline{w}) = (\underline{u} \cdot \underline{w}) \underline{v}.$$

Prove that $\underline{v} \otimes \underline{u} \in \text{Lin}(\mathcal{U}, \mathcal{V})$. (Cf. Theorem A3.1.6.)

Exercise A5.10.2. Let $\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_m\}$ and $\{\underline{u}_1^*, \underline{u}_2^*, \dots, \underline{u}_m^*\}$ be a basis and its dual for an m -dimensional inner product space $\mathcal{U}$, and let $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n\}$ and $\{\underline{v}_1^*, \underline{v}_2^*, \dots, \underline{v}_n^*\}$ be a basis and its dual for an n -dimensional inner product space $\mathcal{V}$. Prove that any $\underline{L} \in \text{Lin}(\mathcal{U}, \mathcal{V})$ admits the representation

$$\underline{L} = \sum_{i=1}^n \sum_{j=1}^m L_{ij} \underline{v}_i^* \otimes \underline{u}_j, \quad L_{ij} = \underline{v}_i^* \cdot \underline{L} \underline{u}_j.$$

Thus, $\{\underline{v}_i^* \otimes \underline{u}_j\} (i \in \mathbb{I}_n, j \in \mathbb{I}_m)$ spans $\text{Lin}(\mathcal{U}, \mathcal{V})$, (Cf. Theorem A3.2.2.)

Exercise A5.10.3. Refer to Exercise A5.10.2. Show that the set $\{\underline{v}_i^* \otimes \underline{u}_j\}$ is linearly independent. Hence, conclude that it is a basis for $\text{Lin}(\mathcal{U}, \mathcal{V})$ and that $\text{Lin}(\mathcal{U}, \mathcal{V})$ is $m \times n$ -dimensional. (Cf. Theorem A3.2.3.)

Suppose that $f: D \rightarrow V$ has a second derivative $D^2f(x)$ at $x \in D \subset U$. Since D^2f is the derivative of the function $Df: B_r(x) \rightarrow \text{Lin}(U, V)$,

$$D^2f(x) \in \text{Lin}(U, \text{Lin}(U, V)).$$

Thus, for $u \in U$,

$$D^2f(x)[u] \in \text{Lin}(U, V);$$

and then for $w \in U$,

$$(D^2f(x)[u])[w] \in V.$$

$\therefore D^2f(x)$ may be viewed as the mapping from $U \times U$ to V defined by

$$D^2f(x)[u, w] := (D^2f(x)[u])[w].$$

We put the pair (u, w) in the bracket on the l.h.s. above in anticipation of

Theorem A5.10.1. Let $f: D \rightarrow V$ have a second derivative
at $x \in D \subset U$. Then $D^2f(x): U \times U \rightarrow V$ defined by

$$D^2f(x)[u, w] = (D^2f(x)[u])[w]$$

is bilinear.

Proof. The linearity in the second variable is trivial. To see

that there also is linearity in the first variable, consider

$$\begin{aligned}
 D^2 f(x) [\alpha_1 u_1 + \alpha_2 u_2, w] &= (D^2 f(x) [\alpha_1 u_1 + \alpha_2 u_2]) [w] \quad (\text{defn. of } D^2 f(x)[u, w]) \\
 &= (\alpha_1 D^2 f(x) [u_1] + \alpha_2 D^2 f(x) [u_2]) [w] \quad (D^2 f(x) \in \text{Lin}(U, \text{Lin}(U, V))) \\
 &= \alpha_1 (D^2 f(x) [u_1]) [w] + \alpha_2 (D^2 f(x) [u_2]) [w] \quad (\text{defns. of addition and scalar mult. in } \text{Lin}(U, \text{Lin}(U, V))) \\
 &= \alpha_1 D^2 f(x) [u_1, w] + \alpha_2 D^2 f(x) [u_2, w] \quad (\text{defn. of } D^2 f(x)[u, w]). \quad \square
 \end{aligned}$$

As an important consequence of the mean-value theorem for derivatives, we have

Theorem A5.10.2. Let $f: D \rightarrow V$ have a second derivative at each $x \in D \subset U$, and let the function defined on D by $x \mapsto D^2 f(x)$ be continuous on D . Then $D^2 f(x)$ viewed as a bilinear map on $U \times U$ to V is symmetric in the sense that at each $x \in D$

$$D^2 f(x) [u, w] = D^2 f(x) [w, u] \quad \forall u, w \in U.$$

Proof. See Lang's Analysis II. $\square$

We can go on to higher derivatives in the natural way. Suppose that $D_{\underline{f}}^1 : B_{r_1}(\underline{x}) \rightarrow V$ is differentiable on some open ball $B_{\frac{r_2}{2}}(\underline{x}) \subset D \subset U$. This sets up the mapping

$$D_{\underline{f}}^2 : B_{\frac{r_2}{2}}(\underline{x}) \rightarrow \text{Lin}(U, \text{Lin}(U, V)),$$

If $D_{\underline{f}}^2$ is differentiable at $\underline{x}$, then its derivative there is the third derivative

$$D_{\underline{f}}^3(\underline{x}) := D D_{\underline{f}}^2(\underline{x}) \in \text{Lin}(U, \text{Lin}(U, \text{Lin}(U, V))).$$

We can keep this going until some derivative is not differentiable. A useful and standard definition is

Definition A5.10.2. The function $\underline{f} : D \rightarrow V$ is said to be of class C^p on $D \subset U$, and we write

$$\underline{f} \in C^p(D, V) \quad)^2$$

if for each $k = 0, 1, 2, \dots, p$, $D^k \underline{f}(\underline{x})$ exists at each $\underline{x} \in D$
and $D^k \underline{f}$ is continuous on D .

In view of Theorem A5.2.4, we could characterize class C^p functions by simply saying that $D^p \underline{f}$ exists and is

¹ Class C^1 functions are often said to be continuously differentiable or smooth.

² Sometimes, simply $\underline{f} \in C^p(D)$ or $\underline{f} \in C^p$. Occasionally, it is

continuation of footnote:

useful to note that $\underline{f}$ is into some subset of $\mathcal{V}$, say $\mathcal{Q}$;
then we write $\underline{f} \in C^p(\mathcal{Q}, \mathcal{Q})$.

continuous on D ; the continuity of the lower derivatives is automatic.

Theorem A5.10.3. With the natural definitions for addition, scalar multiplication, zero function, and equality, the function class $CP(D, V)$ is a linear space,

Proof. Exercise A5.10.4. $\square$

Let us consider a particular example which we shall need later. In Theorem A5.6.1, we saw that the tensor square function $\underline{F} : \text{Lin } V_n \rightarrow \text{Lin } V_n$ defined by

$$\underline{F}(\underline{X}) = \underline{X}^2$$

is differentiable everywhere on $\text{Lin } V_n$, and

$$D\underline{F}(\underline{X})[\underline{U}] = \underline{X}\underline{U} + \underline{U}\underline{X}.$$

Thus,

$$\underline{G} := D\underline{F} : \text{Lin } V_n \rightarrow \text{Lin}(\text{Lin } V_n, \text{Lin } V_n);$$

and

$$\underline{G}(\underline{X} + \underline{V})[\underline{U}] = (\underline{X} + \underline{V})\underline{U} + \underline{U}(\underline{X} + \underline{V})$$

$$= \underline{X}\underline{U} + \underline{V}\underline{U} + \underline{U}\underline{X} + \underline{U}\underline{V} \quad (\text{tensor mult.})$$

$$= (\underline{X}\underline{U} + \underline{U}\underline{X}) + (\underline{V}\underline{U} + \underline{U}\underline{V}) \quad (\text{commutativity and associativity})$$

$$= \underline{G}(\underline{X})[\underline{U}] + \underline{G}(\underline{V})[\underline{U}] \quad (\text{definition of } \underline{G}).$$

$\therefore$ by the definition of addition of linear transformations,

$$\underline{G}(\underline{X} + \underline{V}) = \underline{G}(\underline{X}) + \underline{G}(\underline{V}); \quad (*)$$

which we could have written down to begin with by the linearity of $\underline{G}$ (in $\underline{X}$), (*) and the linearity of $\underline{G} \Rightarrow$ that $\underline{G}$ is differentiable on $\mathcal{V}_n$ and

$$D\underline{G}(\underline{X})[\underline{V}] = \underline{G}(\underline{V})$$

or

$$D^2 \underline{F}(\underline{X})[\underline{V}] = \underline{G}(\underline{V})$$

or

$$(D^2 \underline{F}(\underline{X})[\underline{V}])[\underline{W}] = \underline{V}\underline{W} + \underline{W}\underline{V}$$

or

$$D^2 \underline{F}(\underline{X})[\underline{V}, \underline{W}] = \underline{V}\underline{W} + \underline{W}\underline{V},$$

which demonstrates the bilinearity and the symmetry of $D^2 \underline{F}(\underline{X})[\cdot, \cdot]$.

Since $D^2 \underline{F}(\underline{X})$ is independent of $\underline{X}$,

$$D^2 \underline{F}(\underline{X}) = \underline{Q}, \quad)'$$

In summary, we have

¹cf. Theorem 5.5.2.

Theorem AS.10.4. Let $\underline{F} : \text{Lin } V_n \rightarrow \text{Lin } V_n$ be the
tensor square function defined by

$$\underline{F}(\underline{x}) = \underline{x}^2,$$

Then

$$\underline{F} \in C^\infty(\text{Lin } V_n, \text{Lin } V_n), \quad)^1$$

and

$$D\underline{F}(\underline{x})[\underline{u}] = \underline{x}\underline{u} + \underline{u}\underline{x},$$

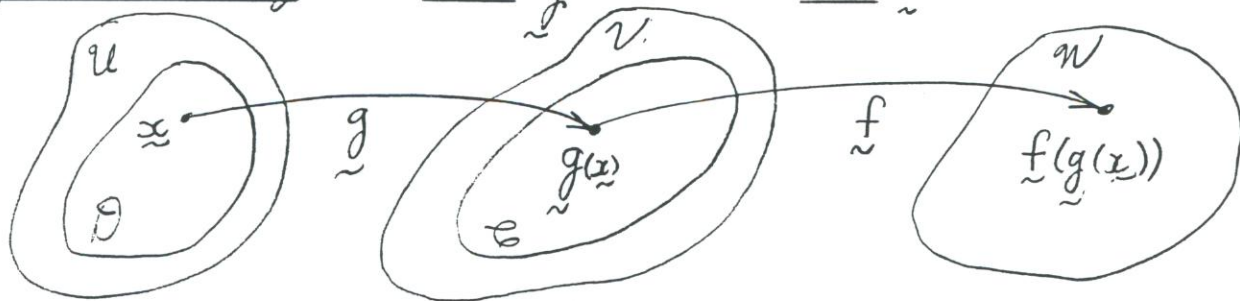
$$D^2\underline{F}(\underline{x})[\underline{v}] = D\underline{F}(\underline{v}),$$

$$D^3\underline{F}(\underline{x})[\underline{w}] = \underline{0},$$

Insert material on pp. AS.10.8a-d

We also shall have occasion to need higher derivatives of composite functions.

Theorem AS.10.7. Let U, V , and W be finite-dimensional
linear spaces. Let D be an open subset of U and C an
open subset of V . Let $g: D \rightarrow C$ and $f: C \rightarrow W$



¹Meaning that all derivatives of $\underline{F}$ exist and are continuous.
on $\text{Lin } V_n$.

Another example which we shall need later is the tensor inverse function $F: \text{Inv } V_n \rightarrow \text{Inv } V_n$ defined by

$$F(\underline{x}) = \underline{x}^{-1}.$$

By Theorem A5.8.3, F is differentiable everywhere on $\text{Inv } V_n$, and

$$DF(\underline{x})[\underline{v}] = -\underline{x}^{-1}\underline{v}\underline{x}^{-1}.$$

Thus,

$$\underline{G} := DF : \text{Lin } V_n \rightarrow \text{Lin}(\text{Lin } V_n, \text{Lin } V_n);$$

and

$$\underline{G}(\underline{x} + \underline{v})[\underline{u}] = -(\underline{x} + \underline{v})^{-1}\underline{u}(\underline{x} + \underline{v})^{-1}$$

$$= -F(\underline{x} + \underline{v})\underline{u}F(\underline{x} + \underline{v}) \quad (\text{substitution})$$

$$= -\{F(\underline{x}) + DF(\underline{x})[\underline{v}] + o(\underline{v})\}\underline{u}\{F(\underline{x}) + DF(\underline{x})[\underline{v}] + o(\underline{v})\} \quad (\text{defn. of derivative})$$

$$= -\{F(\underline{x})\underline{u}F(\underline{x}) + F(\underline{x})\underline{u}DF(\underline{x})[\underline{v}] + F(\underline{x})o(\underline{v})$$

$$+ DF(\underline{x})[\underline{v}]\underline{u}F(\underline{x}) + DF(\underline{x})[\underline{v}]\underline{u}DF(\underline{x})[\underline{v}] + DF(\underline{x})[\underline{v}]o(\underline{v})$$

(tensor mult.)

$$+ o(\underline{v})\underline{u}F(\underline{x}) + o(\underline{v})\underline{u}DF(\underline{x})[\underline{v}] + o(\underline{v})\underline{u}o(\underline{v})\}$$

$$= \underline{G}(\underline{x})[\underline{u}] - F(\underline{x})\underline{u}DF(\underline{x})[\underline{v}] - DF(\underline{x})[\underline{v}]\underline{u}F(\underline{x}) + o(\underline{v}) \text{ as } \underline{v} \rightarrow \underline{0}. \quad (\text{Exercise A5.10.5})$$

Now, we have a bit of a technical problem. In the previous

example of the tensor square function, at this stage we could separate out the $\underline{U}$. That cannot be done here, and what we need is

Theorem A5.10.5. Let $f: \mathcal{O} \rightarrow \mathcal{V}$ be differentiable on some open ball $B_r(\underline{x}) \subset \mathcal{O} \subset \mathcal{U}$. If $\forall \underline{u}, \underline{v} \in \mathcal{U}$

$$Df(\underline{x} + \underline{v})[\underline{u}] - Df(\underline{x})[\underline{u}] = \underline{L}(\underline{x})[\underline{u}, \underline{v}] + o(\underline{v}) \text{ as } \underline{v} \rightarrow \underline{0}.$$

where $\underline{L}(\underline{x}): \mathcal{U} \times \mathcal{U} \rightarrow \mathcal{V}$ is bilinear, then Df is differentiable at $\underline{x}$ and

$$D^2 f(\underline{x})[\underline{u}, \underline{v}] = \underline{L}(\underline{x})[\underline{u}, \underline{v}].$$

Proof. Consider the hypothesis

$$Df(\underline{x} + \underline{v})[\underline{u}] - Df(\underline{x})[\underline{u}] = \underline{L}(\underline{x})[\underline{u}, \underline{v}] + o(\underline{v}).$$

$o(\underline{v})$ stands for the value at $\underline{v}$ of some function which in the present context can be expected to depend on $\underline{x}$ and $\underline{u}$. Moreover, on examination of the rest of the terms in the above equation, we see that the dependence on $\underline{v}$ is linear and that $o(\underline{v}) \in \mathcal{V}$. Thus,

$$o(\underline{v}) = \hat{\underline{L}}(\underline{x}, \underline{v})[\underline{u}],$$

where

$$\hat{\underline{L}}(\underline{x}, \underline{v}) \in \text{Lin}(\mathcal{U}, \mathcal{V}).$$

And since

$$\hat{L}(\underline{x}, \underline{v})[\underline{u}] = \alpha \underline{v},$$

$$\hat{L}(\underline{x}, \underline{v}) = o(\underline{v}) \text{ as } \underline{v} \rightarrow 0 \quad (\text{Exercise A5.10.6}).$$

Now the hypothesis can be written as

$$Df(\underline{x} + \underline{v})[\underline{u}] - Df(\underline{x})[\underline{u}] = L(\underline{x})[\underline{u}, \underline{v}] + \hat{L}(\underline{x}, \underline{v})[\underline{u}],$$

and $\therefore$

$$Df(\underline{x} + \underline{v}) - Df(\underline{x}) = L(\underline{x})[\cdot, \underline{v}] + \hat{L}(\underline{x}, \underline{v})$$

by the linearity (in $\underline{u}$). Since $L(\underline{x})[\cdot, \underline{v}]$ is linear in $\underline{v}$ and since $\hat{L}(\underline{x}, \underline{v}) = o(\underline{v})$ as $\underline{v} \rightarrow 0$, Df is differentiable at $\underline{x}$, and

$$D^2f(\underline{x})[\underline{v}] = L(\underline{x})[\cdot, \underline{v}].$$

Then

$$(D^2f(\underline{x})[\underline{v}])[\underline{u}] = L(\underline{x})[\underline{u}, \underline{v}],$$

but

$$(D^2f(\underline{x})[\underline{v}])[\underline{u}] = D^2f(\underline{x})[\underline{v}, \underline{u}] \quad (\text{definition})$$

$$= D^2f(\underline{x})[\underline{u}, \underline{v}] \quad (\text{symmetry of } D^2). \quad \square$$

Returning now to the analysis of $\mathcal{G} = Df$ on p. A5.10.8 a, we can read off that Df is differentiable on $\text{Inv } V_n$ and that

$$D^2F(\underline{x})[\underline{u}, \underline{v}] = -\underline{F}(\underline{x}) \underline{u} D\underline{F}(\underline{x})[\underline{v}] - D\underline{F}(\underline{x})[\underline{v}] \underline{u} \underline{F}(\underline{x}).$$

Because of the simple way in which $\underline{F}$ and $D\underline{F}$ are involved, it is clear that the differentiation can be repeated indefinitely; and thus we have

Theorem A5.10.6. Let $\underline{F}: \text{Inv } V_n \rightarrow \text{Inv } V_n$ be the tensor inverse function defined by

$$\underline{F}(\underline{x}) = \underline{x}^{-1}$$

Then

$$\underline{F} \in C^\infty(\text{Inv } V_n, \text{Inv } V_n),$$

$$D\underline{F}(\underline{x})[\underline{u}] = -\underline{x}^{-1} \underline{u} \underline{x}^{-1},$$

and

$$D^2F(\underline{x})[\underline{u}, \underline{v}] = -\underline{F}(\underline{x}) \underline{u} D\underline{F}(\underline{x})[\underline{v}] - D\underline{F}(\underline{x})[\underline{v}] \underline{u} \underline{F}(\underline{x}).$$

return to p. A5.10.8

have the differentiability properties

$$\underline{f} \in C^m(\mathcal{O}, V) \quad \text{and} \quad \underline{g} \in C^n(\mathcal{C}, W),$$

where $m, n \geq 2$, Then the composite function $\underline{h} = \underline{f} \circ \underline{g} : \mathcal{O} \rightarrow W$
defined by

$$\underline{h}(\underline{x}) = \underline{f}(\underline{g}(\underline{x})), \quad \underline{x} \in \mathcal{O}$$

has the differentiability property

$$\underline{h} \in C^p(\mathcal{O}, W), \quad p = \min\{m, n\}.$$

Moreover,

$$D\underline{h}(\underline{x}) = D\underline{f}(\underline{g}(\underline{x})) \circ D\underline{g}(\underline{x})$$

and

$$D^2\underline{h}(\underline{x}) = D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})] \circ D\underline{g}(\underline{x}) + D\underline{f}(\underline{g}(\underline{x})) \circ D^2\underline{g}(\underline{x}).$$

Proof. By the chain rule (Theorem A5.7.1), $\underline{h}$ is differentiable at each $\underline{x} \in \mathcal{O}$ and

$$D\underline{h}(\underline{x}) = D\underline{f}(\underline{g}(\underline{x})) \circ D\underline{g}(\underline{x}).$$

Consequently,

$$D\underline{h}(\underline{x} + \underline{u}) = D\underline{f}(\underline{g}(\underline{x} + \underline{u})) \circ D\underline{g}(\underline{x} + \underline{u}) \quad (\text{substitution})$$

$$= D\underline{f}(\underline{g}(\underline{x}) + D\underline{g}(\underline{x})[\underline{u}] + o(\underline{u})) \circ \{D\underline{g}(\underline{x}) + D^2\underline{g}(\underline{x})[\underline{u}] + o(\underline{u})\} \quad (\text{differentiability})$$

$$= \{ D\underline{f}(\underline{g}(\underline{x})) + D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}] + o(\underline{u})] + o(D\underline{g}(\underline{x})[\underline{u}] + o(\underline{u})) \} \circ \{ - \}$$

(differentiability)

$$= \{ D\underline{f}(\underline{g}(\underline{x})) + D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] + D^2\underline{f}(\underline{g}(\underline{x})) [o(\underline{u})] + o(D\underline{g}(\underline{x})[\underline{u}] + o(\underline{u})) \} \circ \{ - \}$$

(linearity of derivative)

$$= \{ D\underline{f}(\underline{g}(\underline{x})) + D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] + o(\underline{u}) \} \circ$$

$$\{ D\underline{g}(\underline{x}) + D^2\underline{g}(\underline{x})[\underline{u}] + o(\underline{u}) \} \quad (\text{Exercise A5.10.7})$$

$$= D\underline{f}(\underline{g}(\underline{x})) \circ D\underline{g}(\underline{x}) + D\underline{f}(\underline{g}(\underline{x})) \circ D^2\underline{g}(\underline{x})[\underline{u}] + D\underline{f}(\underline{g}(\underline{x})) \circ o(\underline{u})$$

$$+ D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] \circ D\underline{g}(\underline{x}) + D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] \circ D^2\underline{g}(\underline{x})[\underline{u}]$$

$$+ D^3\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] \circ o(\underline{u}) + o(\underline{u}) \circ D\underline{g}(\underline{x}) + o(\underline{u}) \circ D^2\underline{g}(\underline{x})[\underline{u}] + o(\underline{u}) \circ o(\underline{u})$$

(composition of linear transformations, cf. Theorem A3.5.1)

$$= D\underline{h}(\underline{x}) + D\underline{f}(\underline{g}(\underline{x})) \circ D^2\underline{g}(\underline{x})[\underline{u}]$$

(Exercise A5.10.8)

$$+ D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] \circ D\underline{g}(\underline{x}) + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

$\therefore D\underline{h}$ is differentiable at $\underline{x}$ and

$$D^2\underline{h}(\underline{x})[\underline{u}] = D\underline{f}(\underline{g}(\underline{x})) \circ D^2\underline{g}(\underline{x})[\underline{u}] + D^2\underline{f}(\underline{g}(\underline{x})) [D\underline{g}(\underline{x})[\underline{u}]] \circ D\underline{g}(\underline{x}).$$

Now since $\underline{f} \in C^2(\mathcal{D}, V)$, its second derivative is symmetric (Theorem A5.10.1), and

$$D^2 \underline{f}(\underline{g}(\underline{x})) [D \underline{g}(\underline{x}) [\underline{u}]] \circ D \underline{g}(\underline{x}) = D^2 \underline{f}(\underline{g}(\underline{x})) [D \underline{g}(\underline{x})] \circ D \underline{g}(\underline{x}) [\underline{u}].$$

$$\therefore D^2 \underline{h}(\underline{x}) [\underline{u}] = D \underline{f}(\underline{g}(\underline{x})) \circ D^2 \underline{g}(\underline{x}) [\underline{u}] + D^2 \underline{f}(\underline{g}(\underline{x})) [D \underline{g}(\underline{x})] \circ D \underline{g}(\underline{x}) [\underline{u}]$$

or

$$D^2 \underline{h}(\underline{x}) = D \underline{f}(\underline{g}(\underline{x})) \circ D^2 \underline{g}(\underline{x}) + D^2 \underline{f}(\underline{g}(\underline{x})) [D \underline{g}(\underline{x})] \circ D \underline{g}(\underline{x})$$

Clearly, we can keep repeating the procedure until either $\underline{f}$ or $\underline{g}$ runs out of derivatives. $\square$

Supplementary Reading

Same as for §5.2

A5.11.1. The Inverse Function Theorem. The Tensor Square Root

Throughout this section $f: D \rightarrow V$, where V is a finite-dimensional space and $D \subset V$ is open. It is important to realize that, in contrast with most of this chapter, here the domain space and the range space are the same.

We shall have several occasions to use the following classical result.

Theorem A5.11.1, (The Inverse Function Theorem) Let $f \in C^p(D, V)$, $p \geq 1$. Let $x_0 \in D$, and suppose that $Df(x_0)$ is invertible¹. Then $\exists$ an open ball centered at x_0 and contained in D , say $B_r(x_0)$, $\ni$

(i) f is one-to-one on $B_r(x_0)$;

(ii) the image $f(B_r(x_0))$ is open;

(iii) $g \in C^p(f(B_r(x_0)), V)$, where g is the inverse² of the restriction of f to $B_r(x_0)$;

(iv) for $x \in B_r(x_0)$,

$$Dg(f(x)) = (Df(x))^{-1}.$$

¹ Definition A3.9.1

² g . Theorem A1.2.2.

While we have assembled adequate machinery to carry out a proof, the proof is fairly long; and $\therefore$ the reader is referred to the references. Our statement of the inverse function theorem is essentially identical to the one in FLEMING's text; however, the proof in LANG's text, even though it is given a general BANACH space setting, is closest to our approach.

A useful corollary of the inverse function theorem is given by

Theorem A5.11.2. Let $f \in C^p(D, V)$, $p \geq 1$ be one-to-one. If $Df(x)$ is invertible at each $x \in D$, then
 $f^{-1} \in C^p(f(D), V)$.

Proof. Since f is one-to-one, $\exists$ an inverse function f^{-1} . This global inverse necessarily agrees with the local inverses guaranteed by the inverse function theorem. $\square$

For applications of the inverse function theorem or the above corollary, it often is helpful to recognize that since V is finite-dimensional, the invertibility of the linear transformation

$$Df(x_0) : V \rightarrow V$$

is equivalent to

$$\det Df(x_0) \neq 0$$

by Theorem A3.9.9, and by Theorem A3.8.7 the determinantal condition is equivalent to

$$Df(\underline{x}_0)[\underline{u}] = \underline{0} \Rightarrow \underline{u} = \underline{0}.$$

As an example of the use of the inverse function theorem, we consider the differentiability properties of the positive definite tensor square root $F(\underline{X}) = \sqrt{\underline{X}}$, which is uniquely defined¹ on $\text{Psym } V_n$ into $\text{Psym } V_n$ by

$$[F(\underline{X})]^2 = \underline{X}.$$

Clearly, F is one-to-one and its inverse is the tensor square function

$$F^{-1}(\underline{Y}) = \underline{Y}^2 =: \underline{G}(\underline{Y}), \quad \underline{Y} \in \text{Psym } V_n.$$

Now while the tensor square root is relatively complicated, its inverse, the tensor square, is not. In fact, by Theorem A5.10.4,

$$\underline{G} \in C^\infty(\text{Psym } V_n, \text{Sym } V_n); \quad)^2$$

actually, $\underline{G}(\text{Psym } V_n) \subset \text{Psym } V_n$. $\therefore$ by applying Theorem A5.11.2 to $\underline{G}(\underline{Y}) = \underline{Y}^2$, we have that its inverse $F(\underline{X}) = \sqrt{\underline{X}}$ also will be class C^∞ provided $D\underline{G}(\underline{Y})$ is invertible for each $\underline{Y} \in \text{Psym } V_n$. In view of the remarks above, it suffices to show that at each $\underline{Y} \in \text{Psym } V_n$

¹ See Theorem A3.15.3.

² We temporarily take it for granted that $\text{Psym } V_n$ is open in $\text{Sym } V_n$.

$$D\mathcal{G}(\underline{Y})[\underline{U}] = \underline{0} \Rightarrow \underline{U} = \underline{0}.$$

Accordingly, with the aid of Theorem A5.10.4, we suppose

$$D\mathcal{G}(\underline{Y})[\underline{U}] = \underline{Y}\underline{U} + \underline{U}\underline{Y} = \underline{0}. \quad (*)$$

Let $(\lambda, \underline{e})$ be an eigenpair of $\underline{Y}$. Then $(*) \Rightarrow$

$$\underline{Y}\underline{U}\underline{e} + \underline{U}(\lambda\underline{e}) = \underline{0}$$

or

$$\underline{Y}(\underline{U}\underline{e}) = -\lambda(\underline{U}\underline{e}).$$

Thus, if $\underline{U}\underline{e} \neq \underline{0}$, then $(-\lambda, \underline{U}\underline{e})$ is also an eigenpair of $\underline{Y}$. But by Theorem A3.15.1, all of the eigenvalues of $\underline{Y} \in \text{Psym } V_n$ must be strictly positive; $\therefore -\lambda$ cannot be an eigenvalue. Consequently, for any eigenvalue $\underline{\lambda}$ of $\underline{Y}$,

$$\underline{U}\underline{e} = \underline{0}. \quad (**)$$

Let $\{\underline{\tilde{e}}_{\langle 1 \rangle}, \underline{\tilde{e}}_{\langle 2 \rangle}, \dots, \underline{\tilde{e}}_{\langle n \rangle}\}$ be an eigenvector basis¹ of $\underline{Y}$. Then for any $\underline{x} \in V_n$,

$$\underline{U}\underline{x} = \underline{U}(x^{\langle i \rangle} \underline{\tilde{e}}_{\langle i \rangle}) = \underline{0};$$

i.e., $\underline{U} = \underline{0}$,

¹cf. Theorem A3.13.1.

Hence, we have established

Theorem A5.11.3.¹ The positive definite tensor square root function $\underline{F}: \text{Psym } V_n \rightarrow \text{Psym } V_n$ defined by

$$\underline{F}(\underline{X}) = \sqrt{\underline{X}}$$

is differentiable everywhere on $\text{Psym } V_n$; in fact,

$$\underline{F} \in C^\infty(\text{Psym } V_n, \text{Sym } V_n).$$

The above proof is nonconstructive in that it establishes that the positive definite tensor square root $\underline{F}$ has a derivative, but it does not tell us what the derivative is. However, knowing that $\underline{F}$ is differentiable, we can differentiate

$$[\underline{F}(\underline{X})]^2 = \underline{X}$$

by the product rule; and with some effort this leads to an explicit formula for $D\underline{F}(\underline{X})[\underline{U}]$.²

¹ This result appeared in print for the first time in GURTIN's An Introduction to Continuum Mechanics, he attributes it to NOLL.

² See HOGER and CARLSON, "On the derivative of the square root of a tensor and Gao's rate theorems.", J. Elasticity 14 (1984) 329-336.

It remains to establish

Theorem A5.11.4. $\text{Psym } \mathcal{V}_n$ is open in $\text{Sym } \mathcal{V}_n$.

Proof. In view of Definition A5.1.3, we must show that each element of $\text{Psym } \mathcal{V}_n$ is the center of an open ball which is contained in $\text{Psym } \mathcal{V}_n$. I.e., we must show that for each $\underline{S}_0 \in \text{Psym } \mathcal{V}_n$
 $\exists$ an $r_0 > 0 \exists$

$$\{ \underline{S} \in \text{Sym } \mathcal{V}_n : |\underline{S} - \underline{S}_0| < r_0 \} \subset \text{Psym } \mathcal{V}_n,$$

Accordingly, let $\underline{S}_0 \in \text{Psym } \mathcal{V}_n$. Then by Definition A3.15.1,

$$\underline{u} \cdot \underline{S}_0 \underline{u} > 0 \quad \forall \underline{u} \in \mathcal{V}_n - \{0\}.$$

Consider the function $f_u: \text{Sym } \mathcal{V}_n \rightarrow \mathbb{R}$ defined for each $\underline{u} \in \mathcal{V}_n - \{0\}$ by the evaluation rule

$$f_u(\underline{S}) = \underline{u} \cdot \underline{S} \underline{u}.$$

Since $\underline{S}_0 \in \text{Psym } \mathcal{V}_n$,

$$f_u(\underline{S}_0) > 0 \quad \forall \underline{u} \in \mathcal{V}_n - \{0\}.$$

It is not difficult to show that f_u is continuous and then to conclude that $f_u(\underline{S}) > 0$ throughout some sufficiently small open ball centered at $\underline{S}_0$, which means that

every element of this ball is positive definite or that the ball is contained in $\text{Psym } V_n$. The details of the demonstration are left for the reader as Exercise A5.11.1. For a model, see the proof of Theorem A5.8.2. $\square$

Supplementary Reading

Same as for §5.2 except delete TRUESDELL and NOLL and add

FLEMING, Functions of Several Variables

A5.12. Functions of Several Variables. Partial Derivatives

Many of the functions that appear in continuum mechanics are most conveniently viewed as functions of several variables. We can embed this within the framework that we have set up for differentiation in the previous sections as follows.

Theorem A5.12.1. Let $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ be n (generally different) linear spaces. Define $\mathcal{U}$ to be the n -fold set-product

$$\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n = \{(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) : \underline{x}_1 \in \mathcal{U}_1, \underline{x}_2 \in \mathcal{U}_2, \dots, \underline{x}_n \in \mathcal{U}_n\}. \quad)^1$$

Equip $\mathcal{U}$ with the operations of addition and scalar multiplication defined by

$$\underline{x} + \underline{y} = (\underline{x}_1 + \underline{y}_1, \underline{x}_2 + \underline{y}_2, \dots, \underline{x}_n + \underline{y}_n)$$

and

$$\alpha \underline{x} = (\alpha \underline{x}_1, \alpha \underline{x}_2, \dots, \alpha \underline{x}_n),$$

where $\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) \in \mathcal{U}$, $\underline{y} = (\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n) \in \mathcal{U}$, and $\alpha \in \mathbb{R}$.

Then $\mathcal{U}$ is a linear space, it is called the product space of $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$

Proof. Exercise A5.12.1. $\square$

We note that the operations of addition and scalar

¹Gf. Definition A1.1.10.

multiplication in $\mathcal{U}$ are based on their counterparts in the linear spaces $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$. Of course, equality in $\mathcal{U}$ is covered by equality for lists (Definition A1.1.8); i.e.,

$$(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) = (\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n) \text{ if } \underline{x}_i = \underline{y}_i \ \forall i \in \mathbb{I}_n.$$

The next result is not unexpected.

Theorem A5.12.2. Refer to Theorem A5.12.1. If each of the linear spaces $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ is finite-dimensional, then the product space $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$ also is finite-dimensional; and

$$\dim \mathcal{U} = \sum_{i=1}^n (\dim \mathcal{U}_i).$$

Proof. Exercise A5.12.2. $\square$

If each of the $\mathcal{U}_i$ is an inner product space, we can use the individual inner products to define an inner product on $\mathcal{U}$.

Theorem A5.12.3. Refer to Theorem A5.12.1. If each of the linear spaces $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ is an inner product space, then

$$\underline{x} \cdot \underline{y} = \underline{x}_1 \cdot \underline{y}_1 + \underline{x}_2 \cdot \underline{y}_2 + \dots + \underline{x}_n \cdot \underline{y}_n$$

defines an inner product on the product space $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$.

Proof. Exercise A5.12.3. $\square$

As usual, when the product space $\mathcal{U}$ is an inner product space, we will work with the inner product norm,¹

$$|\underline{x}| = \sqrt{\underline{x} \cdot \underline{x}},$$

and the norm induced metric,²

$$d(\underline{x}, \underline{y}) = |\underline{x} - \underline{y}|.$$

With a notion of distance available, we are prepared for

Theorem A5.12.4, Refer to Theorem A5.12.1. Let each of the linear spaces $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ be an inner product space, and for each $i \in \mathbb{I}_n$ let $\mathcal{D}_i$ be an open subset of $\mathcal{U}_i$. Then

$$\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n$$

is an open subset of the product space $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$.

Proof. Exercise A5.12.9. $\square$

For the remainder of this section, it is understood that $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ are finite-dimensional inner product spaces, and that the subset $\mathcal{D}_i \subset \mathcal{U}_i$ is open in $\mathcal{U}_i$ ($i \in \mathbb{I}_n$).

¹Gf. Theorem A2.3.8.

²Gf. Theorem A5.1.1.

Furthermore, the corresponding structure induced on the product space $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$ as described above is to be utilized. $\mathcal{V}$ always denotes a finite-dimensional inner product space. Finally, the agreements of §A5.2 concerning the special spaces $\mathbb{R}$ and $\text{Lin } \mathcal{W}$ are to be honored.

Definition A5.12.1. Let $f: \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n \rightarrow \mathcal{V}$. For same $i \in \mathbb{I}_n$, fix $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{i-1}, \underline{x}_{i+1}, \dots, \underline{x}_n$ and consider the partial map

$$\underline{x}_i \mapsto f(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_i, \dots, \underline{x}_n) \quad)^1$$

¹Of course, this strange notation emphasizes that we are looking at $f(\underline{x})$ and $\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$. Conventionally, one simply writes $f(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$. We shall be pedantic about this for awhile.

of $\mathcal{U}_i$ into V . If this map is differentiable at $\underline{x}_i$, its derivative is called the first partial derivative of f at $\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_i, \dots, \underline{x}_n)$ w.r.t. the i th variable and is denoted by $D_i f(\underline{x})$.

Of course, $D_i f(\underline{x}) \in \text{Lin}(\mathcal{U}_i, V)$,

Theorem A5.12.5. Let $f: \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n \rightarrow V$ be differentiable at $\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$. Then each of its first partial derivatives exists at $\underline{x}$; in fact,

$$D_i f(\underline{x})[\underline{u}] = Df(\underline{x})[(\underline{0}, \underline{0}, \dots, \underset{\substack{\uparrow \\ i\text{th place}}}{\underline{u}}, \underline{0}, \underline{0}, \dots, \underline{0})], \quad \underline{u} \in \mathcal{U}_i.$$

Proof. Since f is differentiable at $\underline{x}$, for any sufficiently small $\underline{h} = (\underline{h}_1, \underline{h}_2, \dots, \underline{h}_n) \in \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$,

$$f(\underline{x} + \underline{h}) - f(\underline{x}) = Df(\underline{x})[\underline{h}] + o(\underline{h}) \text{ as } \underline{h} \rightarrow \underline{0},$$

and

$$Df(\underline{x}) \in \text{Lin}(\mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n, V).$$

In particular, for $\underline{h}_i = (\underline{0}, \underline{0}, \dots, \underline{0}, \underline{u}, \underline{0}, \underline{0}, \dots, \underline{0})$

$$f(\underline{x}_1, \dots, \underline{x}_{i-1}, \underline{x}_i + \underline{u}, \underline{x}_{i+1}, \dots, \underline{x}_n) - f(\underline{x}_1, \dots, \underline{x}_n)$$

$$= Df(\underline{x})[(\underline{0}, \dots, \underline{0}, \underline{u}, \underline{0}, \dots, \underline{0})] + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

Consider the function $D_i \underline{f}(\underline{x}) : \mathcal{U}_i \rightarrow V$ defined by

$$D_i \underline{f}(\underline{x})[\underline{u}] = D \underline{f}(\underline{x})[\underline{0}, \dots, \underline{0}, \underline{u}, \underline{0}, \dots, \underline{0}].$$

The linearity of $D \underline{f}(\underline{x})$ obviously carries over to $D_i \underline{f}(\underline{x})$.
 $\therefore$ the partial map

$$\underline{x}_i \mapsto \underline{f}(\underline{x}_1, \dots, \underline{x}_{i-1}, \underline{x}_i, \underline{x}_{i+1}, \dots, \underline{x}_n)$$

is differentiable at $\underline{x}_i$, and its derivative at $\underline{x}_i$ is $D_i \underline{f}(\underline{x})$ as defined above. $\square$

Theorem A5.12.6. Let $\underline{f} : \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n \rightarrow V$. Then
 $\underline{f} \in C^p(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, V)$ ($p \geq 1$) iff for each $i \in \mathbb{I}_n$ the first partial derivative

$$D_i \underline{f} \in C^{p-1}(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, \text{Lin}(\mathcal{U}_i, V)).$$

In this case, for each

$$\underline{x} \in \mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n \text{ and } \underline{u} = (\underline{u}_1, \underline{u}_2, \dots, \underline{u}_n) \in \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$$

$$D \underline{f}(\underline{x})[\underline{u}] = \sum_{i=1}^n D_i \underline{f}(\underline{x})[\underline{u}_i].$$

Proof. First suppose that $\underline{f} \in C^p(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, V)$. Then
 by Theorem A5.12.5, $D_i \underline{f}(\underline{x})$ exists and is given by

$$D_i \underline{f}(\underline{x})[\underline{u}] = D \underline{f}(\underline{x})[\underline{0}, \dots, \underline{0}, \underline{u}, \underline{0}, \dots, \underline{0}],$$

which makes it clear that $D_i \underline{f} \in C^{p-1}(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, \text{Lin}(\mathcal{U}_i, V))$.

Conversely, assume that each partial map is of class CP. For ease of notation, let us consider the case $n=2$. Let $\underline{x} = (\underline{x}_1, \underline{x}_2) \in \mathcal{D}_1 \times \mathcal{D}_2$. Since $\mathcal{D}_1 \times \mathcal{D}_2$ is open, $(\underline{x}_1, \underline{x}_2)$ is the center of a ball contained in $\mathcal{D}_1 \times \mathcal{D}_2$. Then for sufficiently small $\underline{h} = (\underline{h}_1, \underline{h}_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, $(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \underline{h}_2) \in \mathcal{D}_1 \times \mathcal{D}_2$. In fact, the whole line segment from $(\underline{x}_1, \underline{x}_2)$ to $(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \underline{h}_2)$ will be contained in $\mathcal{D}_1 \times \mathcal{D}_2$. Then by the mean-value theorem¹, we can write

$$\begin{aligned} f(\underline{x} + \underline{h}) - f(\underline{x}) &= f(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \underline{h}_2) - f(\underline{x}_1, \underline{x}_2) \\ &= f(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \underline{h}_2) - f(\underline{x}_1 + \underline{h}_1, \underline{x}_2) + f(\underline{x}_1 + \underline{h}_1, \underline{x}_2) - f(\underline{x}_1, \underline{x}_2) \\ &= \int_0^1 D_2 f(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \tau \underline{h}_2) [\underline{h}_2] d\tau \\ &\quad + \int_0^1 D_1 f(\underline{x}_1 + \tau \underline{h}_1, \underline{x}_2) [\underline{h}_1] d\tau. \end{aligned} \quad)^2$$

Since $D_2 f : \mathcal{D}_1 \times \mathcal{D}_2 \rightarrow \text{Lin}(\mathcal{U}_2, \mathcal{V})$ is continuous, the function

$g_\tau : \mathcal{H} \rightarrow \text{Lin}(\mathcal{U}_2, \mathcal{V})$ defined for each $\tau \in [0, 1]$ by

¹ Theorem A5.9.1.

² Actually, we need that the line segment from $\underline{x}_i$ to $\underline{x}_i + \underline{h}_i$ is contained in $\mathcal{D}_i$. Prove this as Exercise A5.12.5.

$$g_{\tau}(\underline{h}_1, \underline{h}_2) = D_2 \underline{f}(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \tau \underline{h}_2) - D_2 \underline{f}(\underline{x}_1, \underline{x}_2)$$

is continuous and $g_{\tau}(\underline{h}_1, \underline{h}_2) \rightarrow 0$ as $(\underline{h}_1, \underline{h}_2) \rightarrow 0$.

Here, $\mathcal{H}$ is a sufficiently small ball in $\mathcal{U}_1 \times \mathcal{U}_2$ centered at $(0, 0)$. Thus, the first integral above can be written as

$$\begin{aligned} \int_0^1 D_2 \underline{f}(\underline{x}_1 + \underline{h}_1, \underline{x}_2 + \tau \underline{h}_2) [\underline{h}_2] d\tau &= \int_0^1 D_2 \underline{f}(\underline{x}_1, \underline{x}_2) [\underline{h}_2] d\tau \\ &\quad + \int_0^1 g_{\tau}(\underline{h}_1, \underline{h}_2) [\underline{h}_2] d\tau \\ &= D_2 \underline{f}(\underline{x}_1, \underline{x}_2) [\underline{h}_2] + \int_0^1 g_{\tau}(\underline{h}_1, \underline{h}_2) [\underline{h}_2] d\tau. \end{aligned}$$

From the defining formula for $g_{\tau}(\underline{h}_1, \underline{h}_2)$, we see that $g_{(\cdot)}(\underline{h}_1, \underline{h}_2)$ is a continuous function on $[0, 1]$. $\therefore$

$$\begin{aligned} \left| \int_0^1 g_{\tau}(\underline{h}_1, \underline{h}_2) [\underline{h}_2] d\tau \right| &\leq \int_0^1 |g_{\tau}(\underline{h}_1, \underline{h}_2) [\underline{h}_2]| d\tau \\ &\leq \max_{\tau \in [0, 1]} |g_{\tau}(\underline{h}_1, \underline{h}_2) [\underline{h}_2]| \\ &= |g_{\tau^*}(\underline{h}_1, \underline{h}_2) [\underline{h}_2]| \\ &\leq K(\tau^*, \underline{h}_1, \underline{h}_2) |\underline{h}_2| \leq K(\tau^*, \underline{h}_1, \underline{h}_2) |\underline{h}|. \end{aligned}$$

¹By our choice of inner product (Theorem A5.12.3), $|\underline{h}|^2 = |\underline{h}_1|^2 + |\underline{h}_2|^2$.

Here, $K(\bar{c}, \underline{h}_1, \underline{h}_2)$ is the bound for the linear transformation

$$g_{\bar{c}}(\underline{h}_1, \underline{h}_2) \in \text{Lin}(\mathcal{U}_2, \mathcal{V})$$

that is delivered by Theorem A3.4.2. Examination of the proof of that theorem reveals that

$$K(\bar{c}, \underline{h}_1, \underline{h}_2) = |m g_{\bar{c}}(\underline{h}_1, \underline{h}_2)[\underline{e}]|,$$

where $\underline{e}$ is a particular element of $\mathcal{U}_2$ and m is the dimension of $\mathcal{U}_2$. Since $g_{\bar{c}}(\underline{h}_1, \underline{h}_2) \rightarrow 0$ as $(\underline{h}_1, \underline{h}_2) \rightarrow 0$,

$$K(\bar{c}, \underline{h}_1, \underline{h}_2) \rightarrow 0 \text{ as } (\underline{h}_1, \underline{h}_2) \rightarrow 0.$$

$\therefore$

$$\int_0^1 g_{\bar{c}}(\underline{h}_1, \underline{h}_2)[\underline{h}_2] d\tau = o(\underline{h}) \text{ as } \underline{h} \rightarrow 0.$$

Similarly,

$$\int_0^1 D_1 f(\underline{x}_1 + \tau \underline{h}_1, \underline{x}_2)[\underline{h}_1] d\tau = D_1 f(\underline{x}_1, \underline{x}_2)[\underline{h}_1] + o(\underline{h}) \text{ as } \underline{h} \rightarrow 0.$$

Hence,

$$\begin{aligned} f(\underline{x} + \underline{h}) - f(\underline{x}) &= D_1 f(\underline{x}_1, \underline{x}_2)[\underline{h}_1] + D_2 f(\underline{x}_1, \underline{x}_2)[\underline{h}_2] + o(\underline{h}) \\ &= D f(\underline{x})[\underline{h}] + o(\underline{h}) \text{ as } \underline{h} \rightarrow 0, \end{aligned}$$

where $Df(\underline{x}): \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{V}$ is defined by

$$Df(\underline{x})[\underline{u}] = D_1 f(\underline{x})[\underline{u}_1] + D_2 f(\underline{x})[\underline{u}_2], \quad \underline{u} = (\underline{u}_1, \underline{u}_2).$$

Now,

$$Df(\underline{x})[\alpha \underline{u} + \beta \underline{v}] = D_1 f(\underline{x})[\alpha \underline{u}_1 + \beta \underline{v}_1] + D_2 f(\underline{x})[\alpha \underline{u}_2 + \beta \underline{v}_2]$$

$$= \alpha D_1 f(\underline{x})[\underline{u}_1] + \beta D_1 f(\underline{x})[\underline{v}_1]$$

$$+ \alpha D_2 f(\underline{x})[\underline{u}_2] + \beta D_2 f(\underline{x})[\underline{v}_2]$$

$$= \alpha Df(\underline{x})[\underline{u}] + \beta Df(\underline{x})[\underline{v}].$$

So $Df(\underline{x})$ is linear, and we conclude that f is differentiable at $\underline{x}$. Also the mapping $Df: \mathcal{D}_1 \times \mathcal{D}_2 \rightarrow \text{Lin}(\mathcal{U}_1 \times \mathcal{U}_2, \mathcal{V})$ is continuous by the continuity of the first partial derivatives.

$\therefore f \in C^1(\mathcal{D}_1 \times \mathcal{D}_2, \mathcal{V})$. We can get that $f \in C^p(\mathcal{D}_1 \times \mathcal{D}_2, \mathcal{V})$

by mathematical induction. Work this out as Exercise AS.12.6. $\square$

To see the implications of the symmetry of the second derivative in the present context, consider $f \in C^2(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, \mathcal{V})$. By Theorem AS.12.6,

$$D_i f \in C^1(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, \text{Lin}(\mathcal{U}_i, \mathcal{V}))$$

and

$$D_j D_i f \in C^0(\mathcal{D}_1 \times \mathcal{D}_2 \times \dots \times \mathcal{D}_n, \text{Lin}(\mathcal{U}_j, \text{Lin}(\mathcal{U}_i, \mathcal{V}))),$$

Thus, for $\underline{u}_j \in \mathcal{U}_j$

$$D_j D_i f(\underline{x})[\underline{u}_j] \in \text{Lin}(\mathcal{U}_i, \mathcal{V});$$

and for $\underline{u}_i \in \mathcal{U}_i$

$$(D_j D_i f(\underline{x})[\underline{u}_j])[\underline{u}_i] \in \mathcal{V}.$$

Let us define

$$D_j D_i f(\underline{x}) [\underline{u}_j, \underline{u}_i] = (D_j D_i f(\underline{x}) [\underline{u}_j]) [\underline{u}_i] .$$

$D_j D_i f(\underline{x}) [\cdot, \cdot]$ is obviously linear in its second variable, and linearity in the first variable follows from a straight forward calculation (cf. the proof of Theorem A5.10.1). As a consequence of Theorems A5.10.2 and A5.12.5, we get the symmetry

$$D_j D_i f(\underline{x}) [\underline{u}_j, \underline{u}_i] = D_i D_j f(\underline{x}) [\underline{u}_i, \underline{u}_j] .$$

Establish this as Exercise A5.12.7.

In summary, we have

Theorem A5.12.7. Let $f \in C^2(D_1 \times D_2 \times \dots \times D_n, V)$. For
each $\underline{x} \in D_1 \times D_2 \times \dots \times D_n$ and for each $i, j \in \mathbb{N}_n$, define the function

$$D_i D_j f(\underline{x}) [\cdot, \cdot] : \mathcal{U}_i \times \mathcal{U}_j \rightarrow V$$

by

$$D_i D_j f(\underline{x}) [\underline{u}_i, \underline{u}_j] = (D_i D_j f(\underline{x}) [\underline{u}_i]) [\underline{u}_j] .$$

Then $D_i D_j f(\underline{x}) [\cdot, \cdot]$ is bilinear, and it is symmetric in
the sense that

$$D_j D_i f(\underline{x}) [\underline{u}_j, \underline{u}_i] = D_i D_j f(\underline{x}) [\underline{u}_i, \underline{u}_j] \quad \forall \underline{u}_i \in \mathcal{U}_i, \underline{u}_j \in \mathcal{U}_j .$$

The case where $\mathcal{U} = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$ is of special interest.

Definition A5.12.2. Let $f: D_1 \times D_2 \times \dots \times D_n \rightarrow V$, where each $D_i \subset \mathbb{R}$ is open. For $\underline{x} = (x_1, x_2, \dots, x_n) \in D_1 \times D_2 \times \dots \times D_n$, fix $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ and consider the partial map

$$x_i \mapsto \tilde{f}((x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n))$$

on D_i to V . The ordinary derivative¹ of this map at x_i ,

$$\tilde{f}_{,i}(\underline{x}) := \lim_{h \rightarrow 0} \frac{1}{h} [\tilde{f}((x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n)) - \tilde{f}((x_1, \dots, x_n))],$$

is called the ordinary first partial derivative of f at $\underline{x} = (x_1, x_2, \dots, x_n)$ w.r.t. the i th variable.

On combining Theorem A5.3.1 with Definitions A5.12.1 and A5.12.2, we get

Theorem A5.12.3. Let $f: D_1 \times D_2 \times \dots \times D_n \rightarrow V$, where each D_i is open in $\mathbb{R}$. If $\tilde{f}$ has an ordinary first partial derivative at $\underline{x} = (x_1, x_2, \dots, x_n)$ w.r.t. the i th variable, then the partial map

$$x_i \mapsto \tilde{f}((x_1, x_2, \dots, x_n))$$

is differentiable at x_i and

¹ Cf. the introductory remarks of §5.3.

$$D_i f(\underline{x})[\alpha] = \alpha f_{,i}(\underline{x}).$$

Conversely, if the partial map

$$x_i \mapsto f(x_1, x_2, \dots, x_n)$$

is differentiable at x_i , then f has an ordinary first partial derivative at $\underline{x} = (x_1, x_2, \dots, x_n)$ w.r.t. the i^{th} variable
and

$$f_{,i}(\underline{x}) = D_i f(\underline{x})[1].$$

Supplementary Reading

Same as for §A5.2

Sometimes it is the range space V which has a product structure, say

$$V = W = W_1 \times W_2 \times \dots \times W_m,$$

where each W_i is a finite-dimensional linear space. Then $\underline{f}: D \rightarrow W$ means

$$\underline{f}(\underline{x}) = (\underline{f}_1(\underline{x}), \underline{f}_2(\underline{x}), \dots, \underline{f}_m(\underline{x})),$$

where $\underline{f}_i: D \rightarrow W_i$. Of course, if we do addition, scalar multiplication, and inner products in the natural way, then W is indeed a finite-dimensional inner product space.¹ This arrangement is understood in the sequel.

Theorem A5.13.1. Let $\underline{f}: D \rightarrow W_1 \times W_2 \times \dots \times W_m$ and write

$$\underline{f}(\underline{x}) = (\underline{f}_1(\underline{x}), \underline{f}_2(\underline{x}), \dots, \underline{f}_m(\underline{x})), \quad \underline{x} \in D.$$

Then $\underline{f}$ is differentiable at $\underline{x}$ iff each $\underline{f}_i$ is differentiable at $\underline{x}$. In this case,

$$D\underline{f}(\underline{x})[\underline{u}] = (D\underline{f}_1(\underline{x})[\underline{u}], \dots, D\underline{f}_m(\underline{x})[\underline{u}]).$$

Proof. For notational simplicity, we consider the case $m=2$. Suppose that $\underline{f} = (\underline{f}_1, \underline{f}_2)$ is differentiable at $\underline{x}$. Then

¹Gf. the discussion of $U = U_1 \times U_2 \times \dots \times U_n$ at the beginning of §A5.12.

$$\begin{aligned} \underline{f}(\underline{x} + \underline{u}) - \underline{f}(\underline{x}) &= (\underline{f}_1(\underline{x} + \underline{u}), \underline{f}_2(\underline{x} + \underline{u})) - (\underline{f}_1(\underline{x}), \underline{f}_2(\underline{x})) \\ &= (\underline{f}_1(\underline{x} + \underline{u}) - \underline{f}_1(\underline{x}), \underline{f}_2(\underline{x} + \underline{u}) - \underline{f}_2(\underline{x})) = D\underline{f}(\underline{x})[\underline{u}] + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}. \quad (\#) \end{aligned}$$

Now $D\underline{f}(\underline{x})[\underline{u}] \in \mathcal{W}_1 \times \mathcal{W}_2$; so for given $\underline{f}$ and $\underline{x}$,

$$D\underline{f}(\underline{x})[\underline{u}] = (\underline{w}_1(\underline{u}), \underline{w}_2(\underline{u})). \quad (\#\#)$$

Then by the linearity of $D\underline{f}(\underline{x})$,

$$\begin{aligned} (\underline{w}_1(\alpha \underline{u} + \beta \underline{v}), \underline{w}_2(\alpha \underline{u} + \beta \underline{v})) &= D\underline{f}(\underline{x})[\alpha \underline{u} + \beta \underline{v}] = \alpha D\underline{f}(\underline{x})[\underline{u}] + \beta D\underline{f}(\underline{x})[\underline{v}] \\ &= \alpha (\underline{w}_1(\underline{u}), \underline{w}_2(\underline{u})) + \beta (\underline{w}_1(\underline{v}), \underline{w}_2(\underline{v})) = (\alpha \underline{w}_1(\underline{u}) + \beta \underline{w}_1(\underline{v}), \alpha \underline{w}_2(\underline{u}) + \beta \underline{w}_2(\underline{v})); \end{aligned}$$

and by the meaning of equality in $\mathcal{W}_1 \times \mathcal{W}_2$,

$$\underline{w}_i(\alpha \underline{u} + \beta \underline{v}) = \alpha \underline{w}_i(\underline{u}) + \beta \underline{w}_i(\underline{v}).$$

Thus, each of the maps $\underline{u} \mapsto \underline{w}_i(\underline{u})$ is linear. Now return to $(\#)$ and $(\#\#)$ and apply the meaning of equality in $\mathcal{W}_1 \times \mathcal{W}_2$ to get

$$\underline{f}_i(\underline{x} + \underline{u}) - \underline{f}_i(\underline{x}) = \underline{w}_i(\underline{u}) + o(\underline{u}) \text{ as } \underline{u} \rightarrow \underline{0}.$$

Since $\underline{w}_i: \mathcal{U} \rightarrow \mathcal{W}_i$ is linear, $\underline{f}_i$ is differentiable at $\underline{x}$ and

$$D\underline{f}_i(\underline{x})[\underline{u}] = \underline{w}_i(\underline{u});$$

and on comparison with $(\#\#)$, we have

$$Df(\underline{x})[\underline{u}] = (Df_1(\underline{x})[\underline{u}], \dots, Df_m(\underline{x})[\underline{u}]).$$

The converse assertion is left as Exercise A5.13.1. $\square$

Higher derivatives work the same way. Suppose, e.g., that

$$f \in C^2(\mathcal{O}, \mathcal{W}_1 \times \mathcal{W}_2 \times \dots \times \mathcal{W}_m)$$

Then $Df: \mathcal{O} \rightarrow \text{Lin}(\mathcal{U}, \mathcal{W}_1 \times \mathcal{W}_2 \times \dots \times \mathcal{W}_m)$ is also differentiable, and by the above theorem

$$\begin{aligned} Df(\underline{x}+\underline{v})[\underline{u}] - Df(\underline{x})[\underline{u}] &= (Df_1(\underline{x}+\underline{v})[\underline{u}], \dots, Df_m(\underline{x}+\underline{v})[\underline{u}]) \\ &\quad - (Df_1(\underline{x})[\underline{u}], \dots, Df_m(\underline{x})[\underline{u}]) \end{aligned}$$

$$\begin{aligned} &= ((Df_1(\underline{x}+\underline{v}) - Df_1(\underline{x}))[\underline{u}], \dots, (Df_m(\underline{x}+\underline{v}) - Df_m(\underline{x}))[\underline{u}]) \\ &= ((D^2f_1(\underline{x})[\underline{v}] + o(\underline{v}))[\underline{u}], \dots, (D^2f_m(\underline{x})[\underline{v}] + o(\underline{v}))[\underline{u}]) \quad \leftarrow \text{use same argument as in the last proof} \\ &= ((D^2f_1(\underline{x})[\underline{v}])[\underline{u}], \dots, (D^2f_m(\underline{x})[\underline{v}])[\underline{u}]) + o(\underline{v}) \text{ as } \underline{v} \rightarrow \underline{0} \\ &= (D^2f(\underline{x})[\underline{v}])[\underline{u}] + o(\underline{v}) \text{ as } \underline{v} \rightarrow \underline{0}. \end{aligned}$$

Thus,

$$(D^2f(\underline{x})[\underline{v}])[\underline{u}] = ((D^2f_1(\underline{x})[\underline{v}])[\underline{u}], \dots, (D^2f_m(\underline{x})[\underline{v}])[\underline{u}])$$

or

$$D^2f(\underline{x})[\underline{v}, \underline{u}] = (D^2f_1(\underline{x})[\underline{v}, \underline{u}], \dots, D^2f_m(\underline{x})[\underline{v}, \underline{u}])$$

in the language of Theorem A5.10.1. These remarks lead us to

Theorem A5.13.2. Let $\underline{f}: D \rightarrow \mathcal{W}_1 \times \mathcal{W}_2 \times \dots \times \mathcal{W}_m$ and write

$$\underline{f}(\underline{x}) = (\underline{f}_1(\underline{x}), \underline{f}_2(\underline{x}), \dots, \underline{f}_m(\underline{x})), \underline{x} \in D.$$

Then $\underline{f} \in C^p(D, \mathcal{W}_1 \times \dots \times \mathcal{W}_m)$ ($p \geq 1$) iff for each $i \in \mathbb{I}_m$
 $\underline{f}_i \in C^p(D, \mathcal{W}_i)$. In the case of $p = 2$,

$$D^2 \underline{f}(\underline{x})[\underline{v}, \underline{u}] = (D^2 \underline{f}_1(\underline{x})[\underline{v}, \underline{u}], \dots, D^2 \underline{f}_m(\underline{x})[\underline{v}, \underline{u}]).$$

Proof. Exercise A5.13.2. $\square$

In continuum mechanical applications, we will often have compositions $\tilde{h} = \tilde{f} \circ \tilde{g}$ where the range space of $\tilde{g}$ and the domain space of $\tilde{f}$ are the same product space. Let

$$\tilde{g}: \theta \rightarrow \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_n \quad \text{and} \quad \tilde{f}: \mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_n \rightarrow \mathcal{W}$$

where $\mathcal{C}_i$ is open in $\mathcal{V}_i$. Suppose that $\tilde{g}$ is differentiable at $\underline{x}$ and that $\tilde{f}$ is differentiable at $\tilde{g}(\underline{x})$. Then by the chain rule (Theorem A5.7.1), $\tilde{h}$ is differentiable at $\underline{x}$ and

$$D\tilde{h}(\underline{x})[\underline{u}] = D\tilde{f}(\tilde{g}(\underline{x})) [D\tilde{g}(\underline{x})[\underline{u}]]. \quad (b)$$

We want to know what this means in terms of the functions in the list $\tilde{g} = (\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_n)$ and the partial derivatives $D_i \tilde{f}$ of $\tilde{f}$.

For notational simplicity let us take $n=2$. By Theorem A5.13.1, the differentiability of $\tilde{g} \Rightarrow$ that $\tilde{g}_i$ is differentiable and

$$D\tilde{g}(\underline{x})[\underline{u}] = (D\tilde{g}_1(\underline{x})[\underline{u}], D\tilde{g}_2(\underline{x})[\underline{u}]), \quad (bb)$$

By Theorem A5.12.6, the differentiability of $\tilde{f} \Rightarrow$ the existence of the partial derivatives $D_i \tilde{f}$ and

$$D\tilde{f}(\tilde{g}(\underline{x}))[(\underline{v}_1, \underline{v}_2)] = D_1 \tilde{f}(\tilde{g}(\underline{x}))[\underline{v}_1] + D_2 \tilde{f}(\tilde{g}(\underline{x}))[\underline{v}_2], \quad (bbb)$$

Assembling (bb) and (bbb) into (b), we get

$$Dh(\underline{x})[\underline{u}] = D_1 \underline{f}(g(\underline{x})) [Dg_1(\underline{x})[\underline{u}]] + D_2 \underline{f}(g(\underline{x})) [Dg_2(\underline{x})[\underline{u}]].$$

Hence, we have proven

Theorem A5.13.3. Let $\mathcal{U}, \mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_n$, and $\mathcal{W}$ be
finite-dimensional linear spaces, Let $\mathcal{D}$ be an open subset of $\mathcal{U}$
and $\mathcal{C}_i$ be an open subset of $\mathcal{V}_i$ ($i=1, 2, \dots, n$). Let $g: \mathcal{D} \rightarrow \mathcal{C}_1 \times \mathcal{C}_2 \times \dots \times \mathcal{C}_n$
be differentiable at $\underline{x} \in \mathcal{D}$ and $\underline{f}: \mathcal{C}_1 \times \mathcal{C}_2 \times \dots \times \mathcal{C}_n \rightarrow \mathcal{W}$ be
differentiable at $g(\underline{x}) = (g_1(\underline{x}), g_2(\underline{x}), \dots, g_n(\underline{x}))$. Then the
composite function

$$\underline{h} = \underline{f} \circ g : \mathcal{D} \rightarrow \mathcal{W}$$

defined by

$$\underline{h}(\underline{x}) = \underline{f}(g(\underline{x}))$$

is differentiable at $\underline{x}$ and

$$Dh(\underline{x})[\underline{u}] = \sum_{i=1}^n D_i \underline{f}(g(\underline{x})) [Dg_i(\underline{x})[\underline{u}]].$$

Supplementary Reading

Same as for §A5.2

A5.14. Identification with the Component Approach

Let us return to our general setting of

$$\underline{f}: \mathcal{D} \rightarrow \mathcal{V},$$

where $\mathcal{D}$ is open in $\mathcal{U}$ and all of the agreements of §5.2 are in force. Often, the finite-dimensionality of $\mathcal{U}$ and $\mathcal{V}$ is used up front; and $\underline{f}$ is studied through its components, which can be thought of as real-valued functions of several real variables. To get to this from our more modern approach will be a good test of our understanding of the previous sections of this chapter. It also will provide us with a useful computational tool.

Let us take the domain space $\mathcal{U}$ to be m -dimensional with basis $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_m\}$. Then for $\underline{x} \in \mathcal{U} \exists$ unique $x^1, x^2, \dots, x^m \in \mathbb{R} \ni$

$$\underline{x} = x^1 \underline{e}_1 + x^2 \underline{e}_2 + \dots + x^m \underline{e}_m;$$

in fact,

$$x^i = \underline{e}^i \cdot \underline{x}, \quad)^1$$

where $\{\underline{e}^1, \underline{e}^2, \dots, \underline{e}^m\}$ is the basis of $\mathcal{U}$ dual to $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_m\}$.

Now, we can think of the components $x^1, x^2, \dots, x^m$ as a

¹Cf. Theorem A2.5.8.

list. Formally, we define

$$\hat{\underline{x}} = (x^1, x^2, \dots, x^m) \in \mathbb{R}^m = \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{m\text{-times}} \quad)^1$$

Even more formally, we have a function

$$\hat{\underline{e}} : \mathcal{U} \rightarrow \mathbb{R}^m$$

defined by

$$\hat{\underline{e}}(\underline{x}) = \hat{\underline{x}} = (\underline{e}^1 \cdot \underline{x}, \underline{e}^2 \cdot \underline{x}, \dots, \underline{e}^m \cdot \underline{x}) . \quad (1)$$

By the uniqueness of components, this function is one-to-one; and the inverse function is

$$\underline{e} : \mathbb{R}^m \rightarrow \mathcal{U}$$

defined by

$$\underline{e}(\hat{\underline{x}}) = \underline{e}((x^1, x^2, \dots, x^m)) = \underline{x} = x^1 \underline{e}_1 + x^2 \underline{e}_2 + \dots + x^m \underline{e}_m . \quad (2)$$

Of course, all we have done here is to exhibit an isomorphism between the m -dimensional linear spaces $\mathcal{U}$ and $\mathbb{R}^m$.)²

Next, we evaluate the function $\underline{f}$ at $\underline{x} \in \mathcal{D}$:

$$\underline{f}(\underline{x}) = \underline{f}(\underline{e}(\hat{\underline{x}})) =: \underline{\hat{f}}(\hat{\underline{x}}) = \underline{\hat{f}}(\hat{\underline{e}}(\underline{x})) , \quad (3)$$

¹ Cf. Theorem A2.1.1.

² Cf. Theorem A2.2.13.

Hence, $\underline{f}$ is the composite function $\underline{\hat{f}} \circ \underline{\hat{e}}$, and by the chain rule differentiability of $\underline{f}$ rests on the differentiability of $\underline{\hat{f}}$ and $\underline{\hat{e}}$. Actually, we have here the special situation where the range space of $\underline{\hat{e}}$ and the domain space of $\underline{\hat{f}}$ are each the product space $\mathbb{R}^m$, and Theorem A5.13.3 applies. For ease in using this result, write

$$\underline{\hat{e}}(\underline{x}) = (e^1(\underline{x}), e^2(\underline{x}), \dots, e^m(\underline{x})), \quad (4)$$

where for $i \in \mathbb{I}_m$ e^i is the function on $\mathcal{U}$ ¹ to $\mathbb{R}$ defined by

$$e^i(\underline{x}) = \underline{e}^i \cdot \underline{x}.$$

By the linearity, e^i is differentiable everywhere on $\mathcal{U}$ and

$$D e^i(\underline{x})[\underline{u}] = \underline{e}^i \cdot \underline{u} = u^i, \quad)^2 \quad (5)$$

Then by Theorem A5.13.1, $\underline{\hat{e}}$ also is differentiable in $\mathcal{U}$ and

$$D \underline{\hat{e}}(\underline{x})[\underline{u}] = (u^1, u^2, \dots, u^m) =: \underline{\hat{u}}.$$

The higher derivatives of e^i and $\underline{\hat{e}}$ all vanish.

Next, we investigate the function

¹Of course, here we need only the subset $\mathcal{D}$.

²Cf. Theorem A5.5.3.

$$\hat{f} : \hat{D} \rightarrow V$$

defined by

$$\hat{f}(\hat{x}) = f(e(\hat{x})),$$

where $\hat{D}$ is the subset of $\mathbb{R}^m$ for which the corresponding elements of $\mathcal{U}$ lie in the domain D of f ; i.e.,

$$\hat{D} = \{ \hat{x} \in \mathbb{R}^m : e(\hat{x}) \in D \}.$$

This set is called the inverse image of D under e . Since e is defined on $\mathbb{R}^m$ with range in $\mathcal{U}$, and since e is continuous on $\mathbb{R}^m$ (an immediate consequence of its obvious linearity¹), it follows from a well known theorem of advanced calculus² that $\hat{D}$ is open in $\mathbb{R}^m$ because D is open in $\mathcal{U}$.

INSERT material on p. A5.14.4a

Exercise A5.14.1. Can you construct a direct proof of the fact that $\hat{D}$ is open?

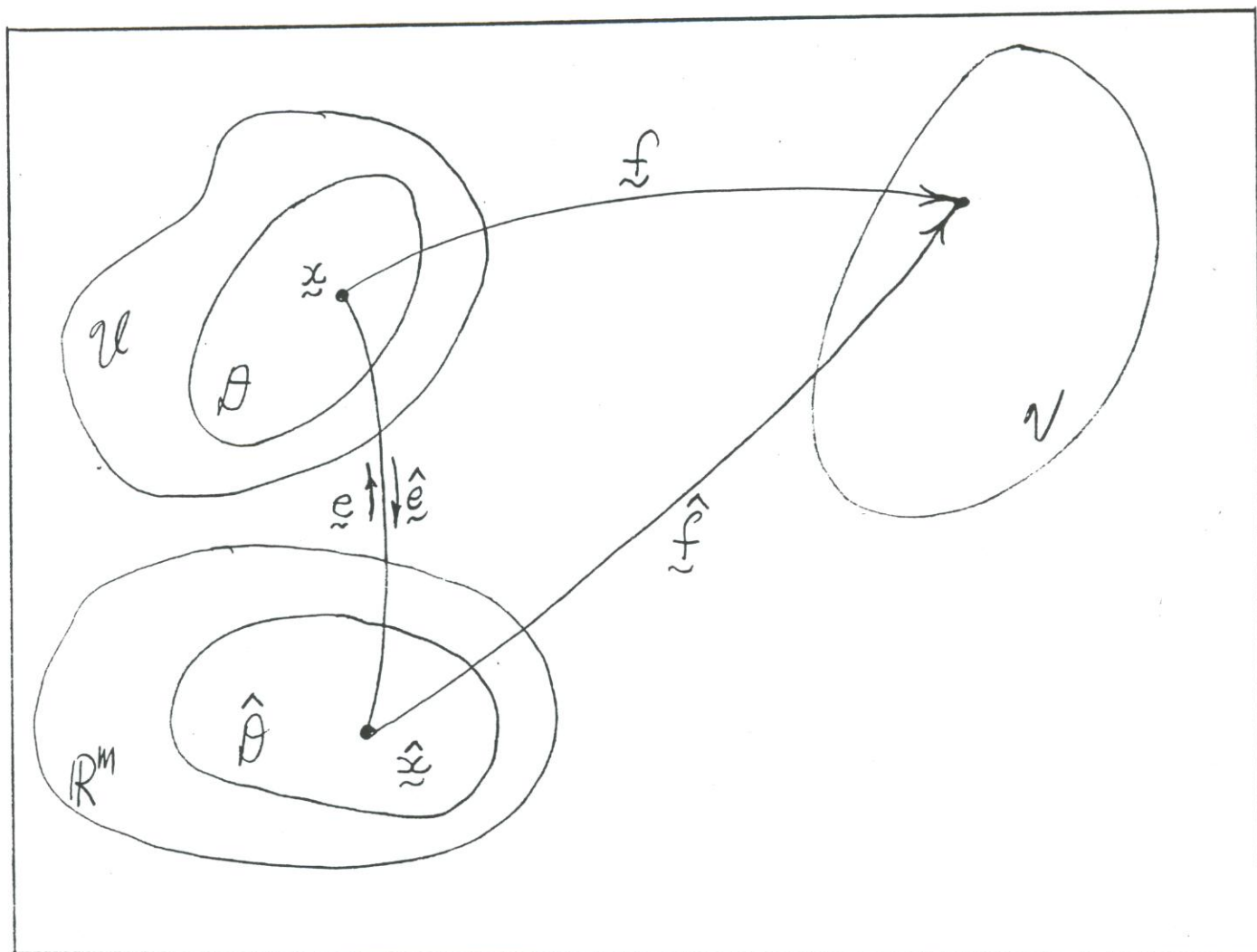
Now, $\hat{D} \subset \mathbb{R}^m = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$, but this does not mean that $\exists D_i \subset \mathbb{R} \ni$

$$\hat{D} = D_1 \times D_2 \times \dots \times D_m.$$

¹ Cf. Theorems A3.4.1 and A3.4.2

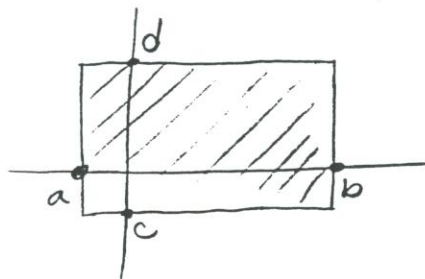
² E.g., see p. 161 of BARTLE's The Elements of Real Analysis.

INSERT for p. A5.14.4

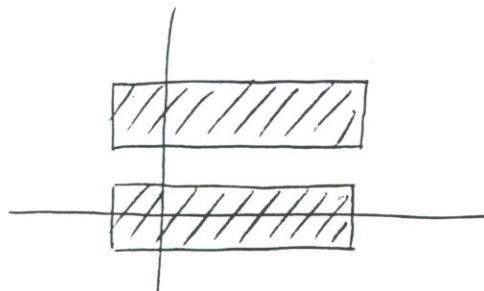


return to p. A5.14.4

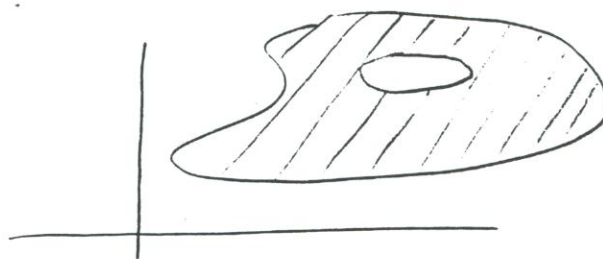
This is easy to see by drawing pictures that correspond to $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, which we think of as the Cartesian plane. Subsets of $\mathbb{R}^2$ of the form $(a,b) \times (c,d)$ are simply rectangles:



If the intervals are not connected, then we pick up multiply connected rectangles; e.g.:



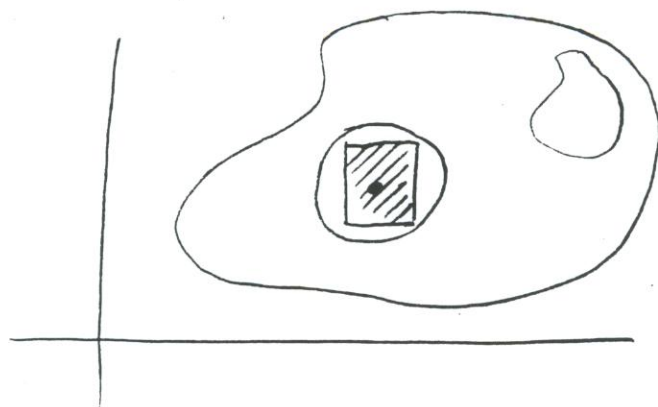
It is clear that we cannot describe a general subset of $\mathbb{R}^2$



as a simple set product.

On the other hand, any interior point is contained in a ball (a disc in $\mathbb{R}^2$) that lies in the

set, and we can put a rectangle inside the ball.



To proceed then, we consider the restriction of $\hat{f}$ to an open "rectangle"

$$\hat{Q} = \mathcal{I}_1 \times \mathcal{I}_2 \times \dots \times \mathcal{I}_m,$$

where each $\mathcal{I}_i$ is an open interval of $\mathbb{R} \ni \hat{Q} \subset \hat{D}$ and $\hat{x} \in \hat{Q}$.¹ We do not introduce any special terminology for this restriction.

By Theorem A5.12.6, $\hat{f} \in C^p(\hat{Q}, \mathcal{V})$ iff

$$D_i \hat{f} \in C^{p-1}(\hat{Q}, \text{Lin}(\mathcal{I}_i, \mathcal{V})), \quad i \in \mathbb{I}_m.$$

Since $\hat{Q} \subset \mathbb{R}^m$, we can use Theorem A5.12.8 to conclude that $\hat{f} \in C^p(\hat{Q}, \mathcal{V})$ iff $\hat{f}$ is continuous on $\hat{Q}$ and has continuous partial derivatives on $\hat{Q}$ of all orders up to and including p . ordinary

¹We have motivated this possibility, but we have not proven it.

Let us take the range space $\mathcal{V}$ to be n -dimensional with basis $\{\underline{\varepsilon}_1, \underline{\varepsilon}_2, \dots, \underline{\varepsilon}_n\}$, then for $\underline{v} \in \mathcal{V} \exists$ unique $v^1, v^2, \dots, v^n \in \mathbb{R} \ni$

$$\underline{v} = v^1 \underline{\varepsilon}_1 + v^2 \underline{\varepsilon}_2 + \dots + v^n \underline{\varepsilon}_n;$$

in fact,

$$v^i = \underline{\varepsilon}_i^i \cdot \underline{v},$$

where $\{\underline{\varepsilon}_1^i, \underline{\varepsilon}_2^i, \dots, \underline{\varepsilon}_n^i\}$ is the basis of $\mathcal{V}$ dual to $\{\underline{\varepsilon}_1, \underline{\varepsilon}_2, \dots, \underline{\varepsilon}_n\}$.

In particular, for $\underline{x} \in \mathcal{D}$

$$\underline{f}(\underline{x}) = \sum_{j=1}^n f^j(\underline{x}) \underline{\varepsilon}_j$$

where

$$f^j(\underline{x}) = \underline{\varepsilon}_j^j \cdot \underline{f}(\underline{x})$$

But $\underline{x} = \underline{e}(\hat{x})$ and $\underline{f}(\underline{e}(\hat{x})) = \hat{\underline{f}}(\hat{x})$, so that

$$\hat{\underline{f}}(\hat{x}) = \sum_{j=1}^n \hat{f}^j(\hat{x}) \underline{\varepsilon}_j, \quad (7)$$

where

$$\hat{f}^j(\hat{x}) = f^j(\underline{e}(\hat{x})) = \underline{\varepsilon}_j^j \cdot \hat{\underline{f}}(\hat{x}).$$

This makes it clear that $\hat{\underline{f}}$ is continuous and has continuous ordinary partial derivatives on $\hat{\mathcal{D}}$ of all orders up to and including p iff exactly the same is true of each component function $\hat{f}^j$.

Now $\hat{x}$ is an arbitrary element of the open set $\hat{\mathcal{D}}$, and

a product set $\hat{\mathcal{Q}}$ can be set up at any $\hat{x}$. Thus, it makes sense to say that $\hat{f}$ is continuous and has continuous ordinary partial derivatives in $\hat{\mathcal{D}}$ of all orders up to and including p iff the same can be said of the component functions $\hat{f}^j$, and this is equivalent to $\hat{f} \in C^p(\hat{\mathcal{D}}, \mathcal{V})$.

Let us suppose that $\hat{f} \in C^p(\hat{\mathcal{D}}, \mathcal{V})$. Then (7) and Theorem A5.12.8 $\Rightarrow$

$$D_i \hat{f}(\hat{x})[\alpha] = \alpha \hat{f}_{,i}(\hat{x}) = \alpha \sum_{j=1}^n \hat{f}^{j}_{,i}(\hat{x}) \underline{e}_j, \quad (8)$$

Now we are finally in position to apply the chain rule in the form of Theorem A5.13.3 to $\underline{f} = \hat{f} \circ \hat{e}$. By (4), the result is

$$\begin{aligned} D \underline{f}(\underline{x})[\underline{u}] &= \sum_{i=1}^m D_i \hat{f}(\hat{e}(\underline{x})) [D e^i(\underline{x})[\underline{u}]] \\ &= \sum_{i=1}^m \omega^i \sum_{j=1}^n \hat{f}^{j}_{,i}(\hat{x}) \underline{e}_j \quad ((5) \text{ and } (8)) \\ &= \sum_{i=1}^m \sum_{j=1}^n \hat{f}^{j}_{,i}(\hat{x}) (\underline{e}^i \cdot \underline{u}) \underline{e}_j \quad (\text{Theorem A2.5.8}) \\ &= \sum_{i=1}^m \sum_{j=1}^n \hat{f}^{j}_{,i}(\hat{x}) (\underline{e}_j \otimes \underline{e}^i) \underline{u} \quad (\text{Exercise A5.10.1}), \end{aligned}$$

or

$$D \underline{f}(\underline{x}) = \sum_{i=1}^m \sum_{j=1}^n \hat{f}^{j}_{,i}(\hat{x}) \underline{e}_j \otimes \underline{e}^i \quad (9)$$

where

$$\hat{x} = \hat{e}(x) = (x^1, x^2, \dots, x^m) = (e^1 \cdot x, e^2 \cdot x, \dots, e^m \cdot x) \quad (1)$$

and

$$x = e(\hat{x}) = x^1 e_1 + x^2 e_2 + \dots + x^m e_m \quad (2)$$

Thus, $\hat{f} \in C^1(\hat{D}, V)$ (equivalently, each $\hat{f}^j$ continuous with continuous ordinary first partial derivatives on $\hat{D}$) $\Rightarrow$ $f \in C^1(D, V)$. Clearly, we can start with (9) and repeat the procedure; i.e.,

$$\hat{f} \in C^p(\hat{D}, V) \Rightarrow f \in C^p(D, V).$$

We can also go in the opposite direction. Suppose $f \in C^p(D, V)$, and consider

$$\hat{f}(\hat{x}) = f(e(\hat{x})).$$

Since

$$e(\hat{x}) = x^1 e_1 + x^2 e_2 + \dots + x^m e_m$$

$$e_{j,i}(\hat{x}) = e_i.$$

So by Theorem A5.12.8,

$$D_i e(\hat{x})[\alpha] = \alpha e_i;$$

and then by Theorem A5.12.6, e is differentiable on $\hat{D}$ and

$$D e(\hat{x})[\hat{u}] = \sum_{i=1}^m u^i e_i =: u, \quad (10)$$

where

$$\hat{\underline{u}} = (u^1, u^2, \dots, u^m).$$

The higher derivatives of $\underline{e}$ are all zero.

Next, consider

$$\hat{f}^j(\hat{\underline{x}}) = \underline{\varepsilon}^j \cdot \hat{\underline{f}}(\hat{\underline{x}}) = \underline{\varepsilon}^j \cdot \underline{f}(\underline{e}(\hat{\underline{x}})).$$

Since $\underline{f}$ is differentiable, it follows from the chain rule and the appropriate product rule (Theorem AS.6.5 (i)) that $\hat{f}^j$ also is differentiable, and

$$\begin{aligned} D\hat{f}^j(\hat{\underline{x}})[\hat{\underline{u}}] &= \underline{\varepsilon}^j \cdot D\underline{f}(\underline{e}(\hat{\underline{x}}))[D\underline{e}(\hat{\underline{x}})[\hat{\underline{u}}]] \\ &= \underline{\varepsilon}^j \cdot D\underline{f}(\underline{e}(\hat{\underline{x}}))[\underline{u}] \quad ((10)) \end{aligned}$$

$$= \underline{\varepsilon}^j \cdot D\underline{f}(\underline{e}(\hat{\underline{x}}))\left[\sum_{i=1}^m u^i \underline{e}_i\right] \quad ((10))$$

$$= \sum_{i=1}^m \underline{\varepsilon}^j \cdot D\underline{f}(\underline{e}(\hat{\underline{x}}))[\underline{e}_i] u^i \quad (\text{linearity of the derivative and the inner product}).$$

$\hat{f}^j$, Because of the product structure of the domain space $\mathbb{R}^m$ of $\hat{f}^j$,

$$D\hat{f}^j(\hat{\underline{x}})[\hat{\underline{u}}] = \sum_{i=1}^m D_i \hat{f}^j(\hat{\underline{x}})[u^i] \quad (\text{Theorem AS.12.6}).$$

and since $\mathbb{R}^m = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$,

$$D_i \hat{f}^j(\hat{\underline{x}})[u^i] = \hat{f}^j_{,i}(\hat{\underline{x}}) u^i \quad (\text{no sum}) \quad (\text{Theorem AS.12.8}).$$

$$D\hat{f}^j(\hat{x})[\hat{u}] = \sum_{i=1}^m \hat{f}_{,i}^j(\hat{x}) u^i,$$

On comparison with the result above, we have

$$\sum_{i=1}^m \underline{e}_i^j \cdot Df(\underline{e}(\hat{x}))[\underline{e}_i] u^i = \sum_{i=1}^m \hat{f}_{,i}^j(\hat{x}) u^i.$$

Then by the arbitrariness of $(u^1, u^2, \dots, u^m)$,

$$\hat{f}_{,i}^j(\hat{x}) = \underline{e}_i^j \cdot Df(\underline{e}(\hat{x}))[\underline{e}_i], \quad (11)$$

which is equivalent to (9).

Since $\underline{f} \in C^p(\mathcal{O}, \mathcal{V})$, $D\underline{f}$ is continuous on $\mathcal{O}$; and since $\underline{e}$ is linear, it is continuous on $\hat{\mathcal{O}}$. Then by (11), $\hat{f}_{,i}^j$ is continuous on $\hat{\mathcal{O}}$. Clearly, we can start with (11) and go on to higher derivatives, and the ordinary partial derivatives of $\hat{f}^j$ inherit the smoothness properties of $\underline{f}$. Thus, $\underline{f} \in C^p(\mathcal{O}, \mathcal{V}) \Rightarrow$ each $\hat{f}^j$ is continuous and has continuous partial derivatives on $\hat{\mathcal{O}}$ of all orders up to and including p .

Assembling this material, we get the following two results.

Theorem AS.14.1. Consider $\underline{f}: \mathcal{O} \rightarrow \mathcal{V}$, where $\mathcal{O}$ is open in $\mathcal{U}$. Let $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_m\}$ and $\{\underline{e}^1, \underline{e}^2, \dots, \underline{e}^m\}$ be a basis and its dual for the domain space $\mathcal{U}$, and let $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$

¹ Cf. Exercise AS.10.2.

and $\{\underline{\varepsilon}^1, \underline{\varepsilon}^2, \dots, \underline{\varepsilon}^n\}$ be a basis and its dual for the range space V . Associate $\mathcal{U}$ and $\mathbb{R}^m$ through the isomorphism
 $\underline{\hat{e}}: \mathcal{U} \rightarrow \mathbb{R}^m$ defined by

$$\underline{\hat{e}}(\underline{x}) = \underline{\hat{e}}(x^1 \underline{e}_1 + x^2 \underline{e}_2 + \dots + x^m \underline{e}_m) = (x^1, x^2, \dots, x^m) = \underline{\hat{x}},$$

whose inverse $\underline{e}: \mathbb{R}^m \rightarrow \mathcal{U}$ is defined by

$$\underline{e}(\underline{\hat{x}}) = \underline{e}((x^1, x^2, \dots, x^m)) = x^1 \underline{e}_1 + x^2 \underline{e}_2 + \dots + x^m \underline{e}_m = \underline{x}.$$

For $\underline{x} \in D$, write

$$\underline{f}(\underline{x}) = \sum_{j=1}^n f^j(\underline{x}) \underline{\varepsilon}_j,$$

and for $j \in \mathbb{I}_n$ define $\hat{f}^j: \hat{D} \rightarrow \mathbb{R}$ by

$$\hat{f}^j(\underline{\hat{x}}) = f^j(\underline{e}(\underline{\hat{x}})) = \underline{\varepsilon}^j \cdot \underline{f}(\underline{e}(\underline{\hat{x}}))$$

so that

$$\underline{f}(\underline{x}) = \sum_{j=1}^n \hat{f}^j(\underline{\hat{x}}) \underline{\varepsilon}_j,$$

where $\hat{D}$ is the open subset of $\mathbb{R}^m$ defined by

$$\hat{D} = \{\underline{\hat{x}} \in \mathbb{R}^m : \underline{e}(\underline{\hat{x}}) \in D\}.$$

If $\underline{f}$ is differentiable at $\underline{x} \in D$, then each component function $\hat{f}^j$ has ordinary first partial derivatives at $\underline{\hat{x}} = \underline{\hat{e}}(\underline{x}) \in \hat{D}$.
In fact,

$$\hat{f}_{,i}^j(\hat{x}) = \underline{e}^j \cdot D \underline{f}(\underline{e}(\hat{x}))[\underline{e}_i], \quad i \in \mathbb{I}_m, j \in \mathbb{I}_n.$$

and

"Conversely", if each component function $\hat{f}^j$ has continuous ordinary first partial derivatives in an open ball centered at $\hat{x}$ and contained in $\hat{D}$, then $\underline{f}$ is differentiable at $\underline{x} = \underline{e}(\hat{x}) \in D$.

$$D \underline{f}(\underline{x}) = \sum_{i=1}^m \sum_{j=1}^n \hat{f}_{,i}^j(\hat{e}(\underline{x})) \underline{e}_j \otimes \underline{e}^i.$$

Of course, we used the quotation marks above because the second assertion was not strictly the converse of the first one. We get a more "symmetric" theorem by stating our results in the $\mathbb{C}^p$ spaces.

Theorem A5.14.2. Refer to the notation of theorem A5.14.1. Then $\underline{f} \in C^p(D, V)$ if each component function $\hat{f}^j$ is continuous and has continuous ordinary partial derivatives on $\hat{D}$ of all orders up to and including p .

In practice, we usually drop the formal correspondence between $\underline{x} = x^1 \underline{e}_1 + x^2 \underline{e}_2 + \dots + x^m \underline{e}_m \in D \subset \mathcal{U}$ and $\hat{x} = (x^1, x^2, \dots, x^m) \in \hat{D} \subset \mathbb{R}^m$ and simply write

$$\underline{f}(\underline{x}) = \sum_{j=1}^n f^j(x^1, x^2, \dots, x^m) \underline{e}_j.$$

? For $\underline{u} \in \mathcal{U}$ and $\underline{v} \in \mathcal{V}$ with $\mathcal{U}$ an inner product space, $\underline{u} \otimes \underline{v} \in \text{Lin}(\mathcal{U}, \mathcal{V})$ is defined by $(\underline{u} \otimes \underline{v}) \underline{w} = (\underline{w} \cdot \underline{u}) \underline{v}$, See Exercise A5.10.1.

Note that the domain and range spaces U and V in Theorems A5.14.1 and 2 are arbitrary finite-dimensional linear spaces. In particular, either can be identified with special spaces such as $\mathbb{R}$ and $\text{Lin } V_n$. The same can be said for

Theorem A5.14.3. Refer to the notation of Theorem A5.14.1.
If f is twice differentiable at $\underline{x} \in D$, then each component function f^j has ordinary second partial derivatives at $\hat{\underline{x}} = \hat{e}(\underline{x}) \in \hat{D}$.

In fact, $\hat{f}^j_{,ik}(\hat{\underline{x}}) = \underline{\varepsilon}^j \cdot \{ (D^2 f(\underline{e}\hat{\underline{x}})) [\underline{e}_k] [\underline{e}_i] \}, j \in \mathbb{I}_n; i, k \in \mathbb{I}_m$

and $D^2 f(\underline{x}) = \sum_{k=1}^m \sum_{i=1}^m \sum_{j=1}^n \hat{f}^j_{,ik}(\hat{e}(\underline{x})) (\underline{\varepsilon}_j \otimes \underline{e}^i) \otimes \underline{e}^k.)'$

If $f \in C^2(D, V)$, then the ordinary second partial derivatives of each component function f^j have the symmetry

$$\hat{f}^j_{,ki} = \hat{f}^j_{,ik} \quad \text{on } \hat{D}.$$

¹ For $\underline{L} \in \text{Lin}(U, V)$ where U is an inner product space, $\underline{L} \otimes \underline{u} \in \text{Lin}(U, \text{Lin}(U, V))$ is defined by
 $(\underline{L} \otimes \underline{u}) \underline{z} = (\underline{u} \cdot \underline{z}) \underline{L}.$

Chapter A6

Point Spaces

A6.1. Definition of a Point Space

We have repeatedly drawn on our experience with the Euclidean plane of high school geometry to motivate the formulation of various concepts. This was especially so for the development of linear spaces in §A2.1. There we viewed a vector as the arrow connecting two points. Here we turn things around. We take the point of view that we know all about vectors, and we base our axiomatic formulation of point spaces on the theory of linear spaces. All we need besides the notion that any two points determine a vector is the idea that we can start at a given point and follow a given vector to get to a new point.

Definition A6.1.1. Let V be a real linear space. Then a point space¹ with translation space² V is a nonempty set, say $\mathcal{Q} = \{p, q, \dots\}$, whose elements are called points³,

¹ Often, affine space.

² Sometimes, difference space.

³ Our use of "hollow" or "double stemmed" letters to denote points is unusual. Most continuum mechanics use lower case, bold faced letters; while mathematicians generally use upper case, light faced letters.

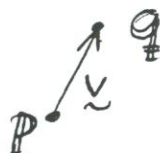
equipped with a function on $\mathcal{Q} \times \mathcal{Q}$ to $\mathcal{V}$, denoted by

$$(p, q) \mapsto \overrightarrow{pq}$$

with the function value $\overrightarrow{pq}$ called the point difference of p and q , which satisfies the following axioms:

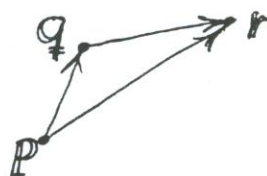
(P1) Given $p \in \mathcal{Q}$ and $\underline{v} \in \mathcal{V}$, $\exists$ a unique point $q \in \mathcal{Q} \ni$

$$\overrightarrow{pq} = \underline{v};$$



(P2) $\forall p, q, r \in \mathcal{Q}$,

$$\overrightarrow{pq} + \overrightarrow{qr} = \overrightarrow{pr}.$$



It often is useful to write the operation

$$(p, q) \mapsto \overrightarrow{pq} = \underline{v} \in \mathcal{V}$$

in the form:

$$q - p = \overrightarrow{pq} = \underline{v}.$$

In this sense, we speak of taking the difference of two points to get a vector¹.

¹In geometrical contexts, we shall usually refer to the elements of a linear space as vectors.

Similarly, the content of axiom (P1); i.e., the accepted fact that given $p \in \mathcal{Q}$ and $v \in V \exists$ unique $q \in \mathcal{Q} \Rightarrow$

$$\overrightarrow{pq} = v,$$

is suggestively expressed as

$$p + v = p + \overrightarrow{pq} = q,$$

In this sense, we speak of adding a vector to a point to get a new point.

It is natural to refer to p and q as the initial point and final point, respectively, of $\overrightarrow{pq}$.

The following theorem contains some of the obviously important properties of point spaces.

Theorem A6.1.1. Let $p, q, r, s \in \mathcal{Q}$, where $\mathcal{Q}$ is a point space with translation space V . Then (and $u, v \in V$)

$$(i) \overrightarrow{pp} = \mathbf{0}, \quad)^1$$

$$(ii) p - p = \mathbf{0},$$

¹Here, of course, $\mathbf{0}$ denotes the zero vector of the associated translation space V .

$$(iii) \quad p + 0 = p$$

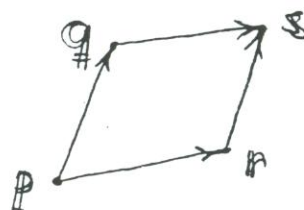
$$(iv) \quad \overrightarrow{pq} = 0 \Rightarrow p = q,$$

$$(v) \quad q - p = 0 \Rightarrow p = q,$$

$$(vi) \quad \overrightarrow{qp} = -\overrightarrow{pq},$$

$$(vii) \quad p - q = -(q - p),$$

$$(viii) \quad \overrightarrow{pq} = \overrightarrow{rs} \Rightarrow \overrightarrow{pr} = \overrightarrow{qs},$$



$$(ix) \quad (p - q) - (r - s) = (p - r) - (q - s),$$

$$(x) \quad (p - s) - (r - s) = p - r,$$

$$(xi) \quad (p + u) - q = (p - q) + u,$$

$$(xii) \quad (p + u) - (p + v) = u - v,$$

$$(xiii) \quad (p + u) - p = u.$$

INSERT material on p. A6.1.4a

Proof. (i): Setting $p = q$ in (P2) yields

$$\overrightarrow{pp} + \overrightarrow{pr} = \overrightarrow{pr}$$

$\Rightarrow \overrightarrow{pp} = 0$ by the uniqueness of the zero vector for v .

INSERT for p. A6.1.4

$$(xiv) \quad p + (\underline{u} + \underline{v}) = (p + \underline{u}) + \underline{v} ,$$

$$(xv) \quad p - q + (r - s) \Rightarrow p - r = q - s ,$$

$$(xvi) \quad p + \underline{u} = q \Rightarrow \underline{u} = q - p ,$$

$$(xvii) \quad p + \underline{u} = p + \underline{v} \Rightarrow \underline{u} = \underline{v} ,$$

$$(xviii) \quad p - q = r - q \Rightarrow p = r ,$$

$$(xix) \quad p - q = p - r \Rightarrow q = r .$$

return to p. A6.1.4

(ii): Same as (i) since $p - p = \overrightarrow{pp}$.

$$\begin{aligned}
 \text{(iii):} \quad q &= p + \overrightarrow{pq} && (\text{Axiom (P1)}) \\
 &= q + \overrightarrow{qq} && (\text{choose } p = q) \\
 &= q + 0 && ((i)).
 \end{aligned}$$

$$\begin{aligned}
 \text{(iv):} \quad q &= p + \overrightarrow{pq} && (\text{Axiom (P1)}) \\
 &= p + 0 && (\text{hypothesis}) \\
 &= p && ((iii)).
 \end{aligned}$$

(v): Same as (iv) since $\overrightarrow{pq} = q - p$

(vi): Set $r = p$ in (P2) to get

$$\begin{aligned}
 \overrightarrow{pq} + \overrightarrow{qp} &= \overrightarrow{pp} \\
 &= 0 && ((i))
 \end{aligned}$$

$$\Rightarrow \overrightarrow{qp} = -\overrightarrow{pq} \text{ by the uniqueness of negatives in } V.$$

(vii): Same as (vi) since $\overrightarrow{pq} = q - p$.

(viii): By (P2),

$$\left. \begin{array}{l} \overrightarrow{pq} + \overrightarrow{qs} = \overrightarrow{ps} \\ \overrightarrow{pr} + \overrightarrow{rs} = \overrightarrow{ps} \end{array} \right\} \Rightarrow \overrightarrow{pq} + \overrightarrow{qs} = \overrightarrow{pr} + \overrightarrow{rs}$$

$$\therefore \overrightarrow{pq} = \overrightarrow{rs} \Rightarrow \overrightarrow{pr} = \overrightarrow{qs},$$

$$(ix): \overrightarrow{qp} + \overrightarrow{pr} = \overrightarrow{qs} + \overrightarrow{sr} \quad ((P2))$$

$$\begin{aligned} \Rightarrow \overrightarrow{qp} - \overrightarrow{sr} &= \overrightarrow{qs} - \overrightarrow{pr} \\ &= -\overrightarrow{sq} + \overrightarrow{rp} \quad ((vi)) \end{aligned}$$

But $\overrightarrow{qp} = p - q$, etc.; thus,

$$(p - q) - (r - s) = (p - r) - (q - s).$$

(x): Choose $q = s$ in (ix) to get

$$\begin{aligned} (p - s) - (r - s) &= (p - r) - (s - s) \\ &= (p - r) - \mathbf{0} \quad ((ii)) \\ &= (p - r) \quad (\text{property of } \mathbf{0}). \end{aligned}$$

(xi): Set $p + \underline{u} = r$ or $\underline{u} = \overrightarrow{pr}$, Then

$$\overrightarrow{qr} = \overrightarrow{qp} + \overrightarrow{pr} \quad (P2))$$

$$\begin{aligned}
 &= \vec{q} \vec{p} + \underline{u} \quad (\text{substitution}) \\
 \Rightarrow \quad \vec{r} - \vec{q} &= (\vec{p} - \vec{q}) + \underline{u} \quad (\text{notation}).
 \end{aligned}$$

Then the substitution $\vec{r} = \vec{p} + \underline{u}$ gives the result.

(xii): In (xi) take $\vec{q} = \vec{p} + \underline{v}$ to get

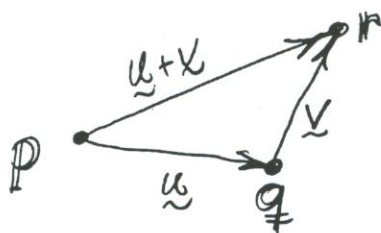
$$\begin{aligned}
 (\vec{p} + \underline{u}) - (\vec{p} + \underline{v}) &= (\vec{p} - \vec{q}) + \underline{u} \\
 &= -(\vec{q} - \vec{p}) + \underline{u} \quad (\text{vii}) \\
 &= -\underline{v} + \underline{u} \quad (\vec{q} = \vec{p} + \underline{v} \text{ means } \underline{v} = \vec{q} - \vec{p})
 \end{aligned}$$

(xiii): In (xii) take $\underline{v} = \underline{0}$ to get

$$\begin{aligned}
 (\vec{p} + \underline{u}) - (\vec{p} + \underline{0}) &= \underline{u} - \underline{0} \\
 &= \underline{u} \quad (\text{property of } \underline{0}).
 \end{aligned}$$

The final result is obtained on noting that $\vec{p} + \underline{0} = \vec{p}$ by (iii).

(xiv): The following figure may be helpful.



Given $p \in \mathcal{A}$ and $u, v \in \mathcal{V}$, define

$$q = p + u \quad \text{and then} \quad r = q + v,$$

which means

$$\overrightarrow{pq} = u \quad \text{and} \quad \overrightarrow{qr} = v.$$

Add these last two equations¹ to get

$$\begin{aligned} u + v &= \overrightarrow{pq} + \overrightarrow{qr} \\ &= \overrightarrow{pr} \quad (\text{Axiom P2}), \end{aligned}$$

but this means

$$\begin{aligned} p + (u + v) &= r \\ &= q + v \quad (\text{substitution}) \\ &= (p + u) + v \quad (\text{substitution}). \end{aligned}$$

(xv): Call $r - s = v$. Then the hypothesis

$$p = q + (r - s)$$

¹ See Theorem A2.1.8,

implies

$$p = q + \underline{v},$$

which means

$$p - q = \underline{v}$$

$$= r - s \quad (\text{substitution}).$$

In view of (ix), this last result implies

$$p - r = q - s.$$

(xvi) - (xix): Exercises AG.1.1 - AG.1.4, respectively. $\square$

A6.1.10

Supplementary Reading

BOWEN and WANG, Introduction to Vectors and Tensors,
Vol. 2, Vector and Tensor Analysis

COLEMAN, MARKOVITZ, and NOLL, Viscometric Flows of Non-Newtonian
Fluids (Appendix on Mathematical Concepts)

GREUB, Linear Algebra

GURTIN, An Introduction to Continuum Mechanics

LICHNEROWICZ, Tensor Calculus

MARTIN and MIZEL, Introduction to Linear Algebra

NOMIZU, Fundamentals of Linear Algebra

MacLANE and BIRKHOFF, Algebra

NOLL, Finite-Dimensional Spaces

A6.2. Finite-Dimensional Point Spaces, Cartesian Coordinate Systems

Points model locations, and, at least for such purposes as modeling the space in which we live, the point spaces of the previous section are more general than we need. The first simplification is to make them finite-dimensional.

Definition A6.2.1. Let $\mathcal{Q}$ be a point space with translation space V . If the linear space V is n -dimensional, then the point space $\mathcal{Q}$ is n -dimensional. Otherwise, $\mathcal{Q}$ is infinite-dimensional.

It is often useful to identify points with the vectors in the associated translation space that connect them to some distinguished point.

Definition A6.2.2. Let $\odot$ be a fixed point in a point space $\mathcal{Q}$ with translation space V . Then the vector $\overrightarrow{\odot p} \in V$ is the position vector of the point $p \in \mathcal{Q}$ w.r.t. $\odot$.

Working with position vectors (rather than points) allows one to draw on all of the machinery associated with linear spaces. This procedure is so natural that physicists and engineers usually do not bother to distinguish between points and their position vectors. On the other hand, mathematicians often use the terminology point as a generic synonym, especially as a synonym for vector. Confusion is possible.

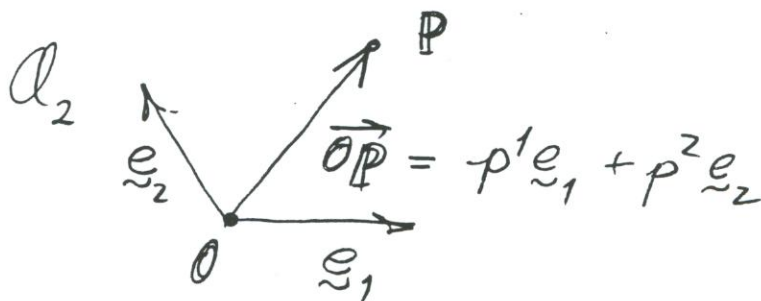
It is worth noting that the idea of position vector is independent of dimension, and it could just as well have been introduced in the previous section, except for the fact that it fits better here. When finite-dimensionality is incorporated, we are lead to

Definition A6.2.3. Let $\mathcal{Q}_n$ be an n -dimensional point space with translation space $\mathcal{V}_n$. Any point $\mathcal{O} \in \mathcal{Q}_n$ together with any basis $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$ for $\mathcal{V}_n$ is called a reference frame for $\mathcal{Q}_n$ and is denoted by

$$\{\mathcal{O}; \underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}.$$

The point $\mathcal{O}$ is called the origin of the reference frame. Given $P \in \mathcal{Q}_n$, the components of the position vector $\vec{\mathcal{O}P} \in \mathcal{V}_n$ w.r.t. the basis $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$ are the coordinates of point P w.r.t. the reference frame $\{\mathcal{O}; \underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$.

The situation in $\mathcal{Q}_2$ might be pictured as below.



Here, the coordinates of P w.r.t. the reference frame $\{\mathcal{O}; \underline{e}_1, \underline{e}_2\}$ are the numbers p^1 and p^2 .

Since the components of a vector relative to a given basis are unique, we have

Theorem A6.2.1, The coordinates of a point in a finite-dimensional point space w.r.t. a given reference frame are unique.

Because of this uniqueness of coordinates, the following definition is meaningful.

Definition A6.2.4, Let $\{O; \underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$ be a reference frame for an n -dimensional point space $\mathcal{A}_n$ so that the position vector w.r.t. the origin O of any point $P \in \mathcal{A}_n$ admits the representation

$$\overrightarrow{OP} = p^i \underline{e}_i \quad)'$$

Then the one-to-one correspondence between points in $\mathcal{A}_n$ and lists in $\mathbb{R}^n$ defined by

$$P \mapsto (p^1, p^2, \dots, p^n)$$

with inverse

$$(p^1, p^2, \dots, p^n) \mapsto O + p^i \underline{e}_i = P$$

is called a Cartesian coordinate system for $\mathcal{A}_n$.

¹Here, as usual, the summation convention is in force (see p. A2.5.12).

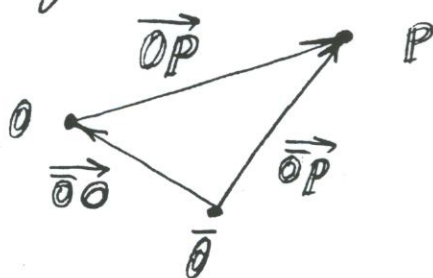
A change of reference frame will generally involve a change of origin as well as a change of basis for the translation space. Let $\{\mathcal{O}; \underline{e}_i\}$ and $\{\bar{\mathcal{O}}; \bar{\underline{e}}_i\}$ be two reference frames for the n -dimensional point space $\mathcal{Q}_n$ with translation space $\mathcal{V}_n$. Write

$$\vec{\mathcal{O}\bar{\mathcal{O}}} = \alpha^i \bar{\underline{e}}_i \quad \text{and} \quad \underline{e}_i = \alpha^j_i \bar{\underline{e}}_j,$$

which is always possible since $\vec{\mathcal{O}\bar{\mathcal{O}}}, \underline{e}_i \in \mathcal{V}_n$ and $\{\bar{\underline{e}}_i\}$ is a basis for $\mathcal{V}_n$.¹ Given $P \in \mathcal{Q}_n$, write

$$\vec{\mathcal{O}P} = p^i \underline{e}_i \quad \text{and} \quad \vec{\bar{\mathcal{O}}P} = \bar{p}^i \bar{\underline{e}}_i,$$

so that $(p^1, p^2, \dots, p^n)$ and $(\bar{p}^1, \bar{p}^2, \dots, \bar{p}^n)$ are the coordinates of P w.r.t. the reference frames $\{\mathcal{O}; \underline{e}_i\}$ and $\{\bar{\mathcal{O}}; \bar{\underline{e}}_i\}$, respectively.



The various pieces of information above get related through the observation that by Axiom (P2)

¹ Cf. the treatment of change of basis for $\mathcal{V}_n$ in §A2.6. There we could do more because we had introduced an inner product for $\mathcal{V}_n$. Because of the finite-dimensionality that could be done here too;

$$\vec{OP} = \vec{OO} + \vec{OP}.$$

By substitution,

$$\bar{p}^i \bar{e}_i = \alpha^i \bar{e}_i + p^i e_i = \alpha^i \bar{e}_i + p^i \alpha_j^i \bar{e}_j.$$

Consequently,

$$\bar{p}^i = \alpha^i + \alpha_j^i p^j,$$

Thus, we have

Theorem A6.2.2. (Transformation Rule for Cartesian Coordinates) Let $\{O; e_1, e_2, \dots, e_n\}$ and $\{\bar{O}; \bar{e}_1, \bar{e}_2, \dots, \bar{e}_n\}$, related by

$$\vec{OO} = \alpha^i \bar{e}_i \text{ and } e_i = \alpha_j^i \bar{e}_j,$$

be two reference frames for an n -dimensional point space $\mathcal{A}$
Then the coordinates $(p^1, p^2, \dots, p^n)$ and $(\bar{p}^1, \bar{p}^2, \dots, \bar{p}^n)$
of a point $P \in \mathcal{A}_n$ w.r.t. the frames $\{O; e_1, e_2, \dots, e_n\}$ and
 $\{\bar{O}; \bar{e}_1, \bar{e}_2, \dots, \bar{e}_n\}$, respectively, are related by

$$\bar{p}^i = \alpha^i + \alpha_j^i p^j, \quad i \in \mathbb{I}_n. \quad)^1$$

Supplementary Reading

Same as for §A6.1

¹Such transformations (i.e., constant function + linear function) are said to be affine.

A6.3. Euclidean Point Spaces. Point Set Topology

In the previous section, we specialized the general concept of a point space by requiring its translation space to be finite-dimensional. Another simplification results if the translation space is taken to be an inner product space.¹

Definition A6.3.1. Let $\mathcal{Q}$ be a point space with translation space V . If the linear space V is an inner product space, then the point space $\mathcal{Q}$ is Euclidean. If V is an n -dimensional inner product space, then $\mathcal{Q}$ is an n -dimensional Euclidean point space.

Of course, the existence of an inner product allows us to speak of lengths and angles (in the translation space). The distance between two points is simply the length of the vector that connects them. Formally, we have

Definition A6.3.2. The distance between two points p and q in a Euclidean space is

$$d(p, q) = |\vec{pq}| = \sqrt{\vec{pq} \cdot \vec{pq}}.$$

Given the notational agreements of §A6.1, we can also write

¹Since the existence of an inner product supplies more structure, perhaps we should speak of additional complexity instead.

$$d(p, q) = |q - p|.$$

In view of motivation of the notion of metric spaces in §A5.1, the following result comes as no surprise.

Theorem A6.3.1. Let E be a Euclidean point space. Then the distance function $d: E \times E \rightarrow \mathbb{R}$ defined by

$$d(p, q) = |\vec{pq}|$$

has the following properties:

(i) Symmetry, $\forall p, q \in E$

$$d(q, p) = d(p, q);$$

(ii) Positive Definiteness, $\forall p, q \in E$

$$d(p, q) \geq 0 \text{ with } d(p, q) = 0 \text{ iff } p = q;$$

(iii) Triangle Inequality, $\forall p, q, r \in E$

$$d(p, r) \leq d(p, q) + d(q, r).$$

Thus, E is a metric space w.r.t. the distance function defined by $d(p, q) = |\vec{pq}|$. This distance function is called the Euclidean metric.

Proof. Exercise A6.3.1. $\square$

Since a Euclidean point space is a metric space, we can do the usual point set topology on it. Some topological notions have been introduced already in §A5.1; namely, open sets and the interior of a set. The following concepts are introduced here in the context of a Euclidean point space, but we emphasize that they remain valid without change (unless explicitly mentioned otherwise) in the setting of an abstract metric space.

For the remainder of this section, E denotes a Euclidean point space. In order to proceed, we recall¹ that if $q \in E$ and $r \in \mathbb{R}$ with $r > 0$, then the subset of E defined by

$$B_r(q) = \{p \in E : d(p, q) < r\}$$

is the open ball of radius r centered at q .

Definition A6.3.3. A subset $A \subset E$ is bounded if it is contained in an open ball; i.e., if $\exists q \in E$ and $r \in \mathbb{R}$ with $0 < r$ ²
 $\Rightarrow A \subset B_r(q)$.

¹ Definition A5.1.2.

² Of course, since $r \in \mathbb{R}$, $r < \infty$.

Definition A6.3.4. A point $p \in E$ is an accumulation
point¹ of a subset $A \subset E$ if every open ball centered at p
 contains at least one point of A other than p . Note that
 p need not belong to A .

An interesting result which helps to explain both the
 significance of accumulation points and the terminology is

Theorem A6.3.2. Let p be an accumulation point of a subset
 $A \subset E$. Then each open ball centered at p contains infinitely
many points of A .

Proof. Exercise A6.3.2. (Hint: Argue by contradiction.) $\square$

Definition A6.3.5. A subset $A \subset E$ is closed if it contains
all of its accumulation points.

Definition A6.3.6. Let E_n be an n -dimensional Euclidean
point space. Then a subset $A \subset E_n$ is compact if it is
closed and bounded.

In a general metric space, the concept of a compact
 subset has a more involved definition, which by the
 famous HEINE-BOREL theorem is equivalent to closed and
 bounded when the metric space is a finite-dimensional
 Euclidean point space.

¹Often, limit point.

Definition A6.3.7. The closure $\bar{A}$ of a subset $A \subset E$ is the union of A and its set of accumulation points,

If the following theorem were not true, we would need some changes in our terminology.

Theorem A6.3.3. The closure $\bar{A}$ of a subset $A \subset E$ is closed.

Proof. Exercise A6.3.3. $\square$

The boundary of a subset $A \subset E$, roughly speaking, should be the "interface" between A and the rest of E . Of course, the rest of E is simply the complement of A w.r.t. E ; i.e., the set $E - A$, whose interior is often suggestively called the exterior of A .

Definition A6.3.8. The boundary of a subset $A \subset E$ is the set

$$\partial A = \{p \in E : \forall r > 0, B_r(p) \cap A \neq \emptyset \text{ and } B_r(p) \cap (E - A) \neq \emptyset\}.$$

In words, p is a boundary point of A if every open ball centered at p intersects both A and $E - A$.

Recall that the interior¹ of a subset $A \subset E$ is denoted by $\bar{A}$. Then as an alternative characterization of the

¹See Definition A5.1.4.

boundary of A , we have

Theorem A6.3.4. Let $A \subset E$. Then

$$\partial A = \bar{A} - \overset{\circ}{A} \quad \text{and} \quad \bar{A} = \overset{\circ}{A} \cup \partial A,$$

Proof. Exercise A6.3.4. $\square$

Some additional useful results are given by

Theorem A6.3.5. Let $A \subset E$. Then

(i) A is closed iff $\bar{A} = A$;

(ii) A is closed iff $E - A$ is open;

(iii) A is closed iff $\partial A \subset A$;

(iv) $\partial(E - A) = \partial A$;

(v) ∂A is closed.

Proof. Exercises A6.3.5-9. $\square$

Let us develop too simple a picture of those matters, we point out that if a set has any "isolated" points, then they are included in its boundary. To be more precise, we make

Definition A6.3.9. A point $p \in A \subset E$ is an isolated point of the subset A if $\exists$ an open ball centered at p which contains no points of A other than p ; i.e., if $\exists r \in \mathbb{R}$ with $r > 0 \ni B_r(p) \cap A = \{p\}$.

Quite easily then, we have

Theorem A6.3.6. Let p be an isolated point of a subset $A \subset E$. Then $p \in \partial A$.

Proof. Exercise A6.3.10. $\square$

Roughly speaking, isolated points of a subset are not "connected" to the rest of the set. In §5.9, we had occasion to develop the notion of a polygonally connected subset. For later purposes, we will need

Definition A6.3.10. A subset $A \subset E$ is (smoothly) pathwise connected if for each pair of points $p, q \in A$ $\exists$ a (smooth) continuous function $g: [0, 1] \rightarrow A$ $\ni$ $g(0) = p$ and $g(1) = q$.

Certainly, at least for open subsets, the properties polygonally connected, pathwise connected, and smoothly pathwise connected are mutually equivalent. To just say that a subset is connected is slightly different. For this and a fuller development of point set topology see the references.

Supplementary Reading

Same as for § AG.1 plus

BARTLE, The Elements of Real Analysis

DIEUDONNÉ, Foundations of Modern Analysis

FLEMING, Functions of Several Variables

MICHEL and HERGET, Mathematical Foundations in Engineering
and Science

NAYLOR and SELL, Linear Operator Theory in Engineering and Science

SIMMONS, Introduction to Topology and Modern Analysis