0. THE BACKGROUND: VECTOR CALCULUS

The reader of this book is assumed to have a working knowledge of vector calculus. The book is intended to explain wide generalisations of that subject. In this chapter we identify some aspects of the subject which are not always treated adequately in elementary accounts. These will be the starting points for several later developments.

1. Vectors

The word "vector" is used in slightly different ways in pure mathematics, on the one hand, and in applied mathematics and physics, on the other. The usefulness of the vector concept in applications is enhanced by the convention that vectors may be located at different points of space: thus a force may be represented by a vector located at its point of application. Sometimes a distinction is drawn between "free" vectors and "bound" vectors. By a free vector is meant one which may be moved about in space provided that its length and direction are kept unchanged. A bound vector is one which acts at a definite point.

In the mathematical theory of vector spaces these distinctions are unknown. In that context all vectors, insofar as they are represented by directed line segments, must originate from the same point, namely the zero vector, or origin. Only the parallelogram rule of vector addition makes sense, not the triangle rule.

Closely connected with these distinctions is a difficulty about the representation of the ordinary space of classical physics and the space-time of special relativity theory. On the one hand, one finds it more or less explicitly stated that space is homogeneous: the laws of physics do not prefer any one point of space, or of space-time, over any other. On the other hand, almost any quantitative treatment of a physical problem begins with a particular choice of coordinate axes—a choice which singles out some supposedly unremarkable point for the privileged role of origin. The underlying difficulty here is that the vector space $\mathbf{R}^3$ is not quite appropriate as a model for ordinary physical space. The kind of space which is homogeneous, in which the whereabouts of the origin of coordinates is an act of choice not a dictate of the structure, which retains sufficient vectorial properties to model physical space, and in which a sensible distinction between free and bound vectors can be made, is called an affine space; it will be discussed in detail in Chapter 1. (The concept of a vector space, and the notation $\mathbf{R}^3$, are explained in Note 2 at the end of Chapter 1.)

It is unfortunate that distinctions which appear merely pedantic in the straightforward context of $\mathbf{R}^3$ are sometimes important for generalisations. The scalar product, also called the inner or dot product, is so familiar that it is difficult to keep in mind that $\mathbf{R}^3$ may be given many different scalar products, with similar properties, or no scalar product at all. The scalar product is a secondary structure: if one fails

to recognise this one cannot exploit fully the relationship between a vector space and its dual space (the dual space is also defined in Note 2 to Chapter 1).

In other terms, the matrix product of a row and a column vector, resulting in a number, may be constructed without the introduction of any secondary structure, but the scalar product of two column vectors, also resulting in a number, cannot. The first makes use only of vector space notions, combining an element of a vector space, represented as a column vector, and an element of its dual space, represented as a row vector. The product, called a pairing, is represented by matrix multiplication. In tensor calculus this would be expressed by the contraction of a contravariant and a covariant vector. The second requires the additional concept of a scalar product. It is surprising how rich a geometry may be developed without the introduction of a scalar product: after this chapter, we do not introduce scalar products again until Chapter 7. It is also instructive to see which notions of vector algebra and calculus really depend on the scalar product, or on the metrical structure, of Euclidean space.

From the outset we shall distinguish notationally between the underlying n-dimensional vector space of n-tuples, and the same space with the scalar product added, by writing $\mathbf{R}^n$ for the former but $\mathcal{E}^n$ for the latter.

2. Derivatives

Let f be a function on $\mathcal{E}^3$; $\operatorname{grad} f$ is the column vector of its partial derivatives, evaluated at any chosen point. Let $\mathbf{v}$ be a unit vector at that point, with a chosen direction. Then the directional derivative of f in the chosen direction is the scalar product $\mathbf{v} \cdot \operatorname{grad} f$.

A more general directional derivative may be defined by dropping the requirement that $\mathbf{v}$ be a unit vector. This directional derivative may be interpreted as the derivative along any curve which has $\mathbf{v}$ as tangent vector at the point in question, the curve not being necessarily parametrised by its length. If $\mathbf{v}$ is regarded as a velocity vector of a particle then $\mathbf{v} \cdot \operatorname{grad} f$ is the time rate of change of f along the path of the particle. However, the directional derivative may perfectly well be constructed without appealing to the scalar product, by taking the partial derivatives as components of a row vector. This vector is called the differential of f; in these terms the directional derivative is simply the pairing of the tangent vector and the differential. Having no scalar product, one cannot sustain the usual interpretation of the gradient as the normal to a surface $f = \text{constant}$, but the differential may still be used to specify the tangent plane to this surface at the point in question. The main advantage of this point of view is that it is the starting point for a general theory which encompasses non-metrical versions of grad, curl and div, and much more besides.

In vector calculus one sees pretty clear hints of close connections between grad, curl and div, but in the usual treatments they are often not much more than hints. We have in mind for example the relations $\operatorname{curl} \circ \operatorname{grad} = \mathbf{0}$ and $\operatorname{div} \circ \operatorname{curl} = 0$, and the assertions that a vector field is a gradient if and only if its curl vanishes and a curl if and only if its divergence vanishes. These relations all fall into place in the development of the exterior calculus, which is undertaken in Chapters 4 and 5.

We return to consideration of the directional derivative, but from a different point of view. The directional derivative operator associated with a vector field $\mathbf{X}$ will for the time being be denoted by $\mathbf{X}\cdot\mathrm{grad}$, as before, but now we insist on regarding the components of $\mathrm{grad}f$ as the components of the differential, so that there is no need to introduce the scalar product to construct directional derivatives. We list some properties of operators of this type, as applied to functions:

(1) $\mathbf{X}\cdot\mathrm{grad}$ maps functions to functions
(2) $\mathbf{X}\cdot\mathrm{grad}$ is a linear operator, and is linear in $\mathbf{X}$
(3) $(f\mathbf{X})\cdot\mathrm{grad} = f(\mathbf{X}\cdot\mathrm{grad})$
(4) $(\mathbf{X}\cdot\mathrm{grad})(f_1f_2) = (\mathbf{X}\cdot\mathrm{grad}\,f_1)f_2 + f_1(\mathbf{X}\cdot\mathrm{grad}\,f_2)$ (Leibniz's rule).

The composition of directional derivative operators, and their commutation properties, are not often discussed in standard treatments of vector calculus. The composite $(\mathbf{X}\cdot\mathrm{grad})\circ(\mathbf{Y}\cdot\mathrm{grad})$ of two operators is not a directional derivative operator, because it takes second derivatives of any function on which it acts, while directional derivative operators take only first derivatives. However, the commutator

$$(\mathbf{X}\cdot\mathrm{grad})\circ(\mathbf{Y}\cdot\mathrm{grad}) - (\mathbf{Y}\cdot\mathrm{grad})\circ(\mathbf{X}\cdot\mathrm{grad})$$

is a directional derivative operator, which is to say that it is of the form $\mathbf{Z}\cdot\mathrm{grad}$ for some vector field $\mathbf{Z}$. The vector field $\mathbf{Z}$ depends on $\mathbf{X}$ and $\mathbf{Y}$ (and on their derivatives). It is usual to denote the commutator by the use of square brackets, and to extend this notation to the vector fields, writing

$$[\mathbf{X}\cdot\mathrm{grad}, \mathbf{Y}\cdot\mathrm{grad}] = [\mathbf{X},\mathbf{Y}]\cdot\mathrm{grad}.$$

It is not difficult to compute the components of $[\mathbf{X},\mathbf{Y}]$ in terms of the components of $\mathbf{X}$ and $\mathbf{Y}$; this, and the significance and properties of the brackets of vector fields, is discussed at length in Chapter 3.

The directional derivative operator may be applied to a vector field as well as to a function. The Cartesian components of a vector field are functions, and the operator is applied to them one by one: if $\mathbf{E}_1$, $\mathbf{E}_2$, and $\mathbf{E}_3$ are the usual coordinate vector fields and

$$\mathbf{Y} = Y_1\mathbf{E}_1 + Y_2\mathbf{E}_2 + Y_3\mathbf{E}_3$$

then

$$(\mathbf{X}\cdot\mathrm{grad})\mathbf{Y} = (\mathbf{X}\cdot\mathrm{grad}\,Y_1)\,\mathbf{E}_1 + (\mathbf{X}\cdot\mathrm{grad}\,Y_2)\,\mathbf{E}_2 + (\mathbf{X}\cdot\mathrm{grad}\,Y_3)\,\mathbf{E}_3.$$

This operation has properties similar to those of the directional derivative as applied to functions:

(1) $\mathbf{X}\cdot\mathrm{grad}$ maps vector fields to vector fields
(2) $\mathbf{X}\cdot\mathrm{grad}$ is a linear operator and is linear in $\mathbf{X}$
(3) $(f\mathbf{X})\cdot\mathrm{grad} = f(\mathbf{X}\cdot\mathrm{grad})$
(4) $\mathbf{X}\cdot\mathrm{grad}(f\mathbf{Y}) = (\mathbf{X}\cdot\mathrm{grad}\,f)\mathbf{Y} + f(\mathbf{X}\cdot\mathrm{grad}\,\mathbf{Y})$.

However, the conventional use of the same symbol $\mathbf{X}\cdot\mathrm{grad}$ for what are really two different operators—the directional derivatives of functions and of vector fields—makes the last of these appear more like Leibniz's rule than it really is: on the right hand side each of the two usages of $\mathbf{X}\cdot\mathrm{grad}$ occurs.

The properties of the directional derivative of vector fields listed above are typical of the properties of a covariant derivative; this subject is developed in Chapters 2, 5 and 7, and generalised in Chapters 9, 11, 13 and 15. The application of $\mathbf{X}\cdot\mathrm{grad}$ to vector fields in $\mathbf{R}^3$ is the simplest example of a covariant derivative. The interaction of the covariant derivative with scalar products is exemplified by the formula (in $\mathcal{E}^3$)

$$(\mathbf{X}\cdot\mathrm{grad})(\mathbf{Y}\cdot\mathbf{Z}) = (\mathbf{X}\cdot\mathrm{grad}\,\mathbf{Y})\cdot\mathbf{Z} + \mathbf{Y}\cdot(\mathbf{X}\cdot\mathrm{grad}\,\mathbf{Z}).$$

Note that in this formula the two different meanings of $\mathbf{X}\cdot\mathrm{grad}$ again occur: on the left it acts on a function, on the right, on vector fields. The commutator of two such operators, acting on vector fields, is given by the same formula as for the action on functions:

$$[\mathbf{X}\cdot\mathrm{grad}, \mathbf{Y}\cdot\mathrm{grad}] = [\mathbf{X},\mathbf{Y}]\cdot\mathrm{grad}.$$

This formula, which is not typical of covariant differentiation formulae, expresses the fact that ordinary Euclidean space is flat, not curved.

We have adopted the usual convention of vector calculus that vectors and vector fields are printed in boldface type. We shall continue to follow this convention, but only for the vectors and vector fields in $\mathcal{E}^3$ with which vector calculus deals: in more general situations vectors and vector fields will be printed in ordinary italic type.

3. Coordinates

One of the byproducts of the approach to be developed here is that the expression in curvilinear coordinates of such constructions as grad, curl and div, which can appear puzzling, becomes relatively straightforward. Coordinate transformations, and the way in which quantities transform in consequence, have an important part to play in the developing argument. However, we do not generally define objects in terms of their transformation properties under change of coordinates, as would be the practice in tensor calculus. We develop the idea that since no one coordinate system is preferable to another, objects of interest should be defined geometrically, without reference to a coordinate system, and their transformation properties deduced from the definitions. In tensor calculus, on the other hand, the transformation law is the primary focus, and generally the basis for the definition of objects.

The arena for most of the geometry described below is (finally) the differentiable manifold, in which coordinates exist locally, but no assumption is made that a single coordinate system may be extended to cover the whole space. The homogeneity of affine space is thus taken many steps further in the definition of a differentiable manifold.

We shall also attempt to give some indications of global matters, which tensor calculus rarely does, it being ill adapted for that purpose. On the other hand, the results we obtain often have tensor calculus equivalents, which will frequently be revealed in the exercises; but our approach is, in a word, geometrical. Our exposition is intended to illustrate Felix Klein's remark that "The geometric properties of any figure must be expressible in formulas which are not changed when one changes the

coordinate system ... conversely, any formula which in this sense is invariant under a group of coordinate transformations must represent a geometric property".

4. The Range and Summation Conventions

Throughout this work we shall use certain conventions regarding indices which simplify the representation of sums, and result in considerable savings of space and effort. These are the range and summation conventions, often associated with the name Einstein. The reader who is already familiar with tensor calculus will need no instruction in their use. For other readers, not so prepared, we describe their operation here.

It is simplest to begin with an example. Consider the matrix equation

$$v = \lambda(u).$$

Here u is supposed to be a column vector, of size n say (or $n \times 1$ matrix); λ is an $m \times n$ matrix; and v, therefore, is a column vector of size m ($m \times 1$ matrix). This equation may be interpreted as expressing how each individual component of v is determined from the components of u via λ. To write down that expression explicitly one introduces notation for the components of u and v and the elements of λ: say u^a to stand for the ath component of u ($a = 1, 2, \ldots, n$); v^α to stand for the αth component of v ($\alpha = 1, 2, \ldots, m$); and λ^α_a to stand for the (α, a) element of λ, that is, the element in the αth row and ath column. The matrix equation above is then equivalent to the m equations

$$v^\alpha = \sum_{a=1}^{n} \lambda^\alpha_a u^a.$$

The range convention arises from the realisation that it is not necessary to state, at each occurrence of a set of equations like this, that there are m equations involved and that the truth of each is being asserted. This much could be guessed from the appearance of the index α on each side of the equation: for α is a free index, unlike a which is subject to the summation sign. On the other hand, the summation convention follows from the observation that whenever a summation occurs in an expression of this kind it is a summation over an index (here a) which occurs precisely twice in the expression to be summed. Thus summation occurs only where there is a repeated index; and when an index is repeated summation is almost always required. Under these circumstances the summation symbol $\sum_{a=1}^{n}$ serves no useful function, since summation may be recognised by the repetition of an index; it may therefore be omitted.

Thus the component equation above is written, when range and summation conventions are in force, in the simple form

$$v^\alpha = \lambda^\alpha_a u^a.$$

The presence of the repeated index a on the right hand side implies summation over its permitted range of values $1, 2, \ldots, n$ by virtue of the summation convention; while the presence of the free index α on both sides of the equation implies equality for each value $1, 2, \ldots, m$ that it can take, by virtue of the range convention.

In general, the range and summation conventions work as follows. If, in an equation involving indexed quantities, there are free (unrepeated) indices, then the equation holds for all values in the ranges of all free indices, these ranges having been declared previously: this is the *range convention*. Where, in an expression involving indexed quantities, any index is repeated, summation over all possible values in the range of that index is implied, the range again having been declared previously: this is the *summation convention*.

The ranges of indices governed by the range and summation conventions will always be finite: thus only finite sums are involved, and there is no problem of convergence.

Operation of the range and summation conventions in practice is relatively straightforward. One or two rules—often best employed as running checks on the correctness of a calculation—should be mentioned. The number of free indices on the two sides of an equation must be the same; and of course each different free index in an expression must be represented by a different letter. Repeated indices in an expression may occur only in pairs. Replacement of a letter representing an index by another letter is allowed, provided that all occurrences of the letter are changed at the same time and in the same way, and provided that it is understood that the new letter has the same range of values as the one it replaces. The most convenient practice to adopt, where indices with different ranges are involved in a single calculation, is to reserve a small section of a particular alphabet to represent indices with a given range. Thus in the case discussed above one could take a, b, c to range and sum from 1 to n, and α, β, γ to range and sum from 1 to m; then $v^\beta = \lambda^\beta_c u^c$ would mean exactly the same as $v^\alpha = \lambda^\alpha_a u^a$.

From a given expression containing two free indices with the same ranges, a new expression may be formed by making them the same, that is, by taking a sum: this process is known as *contraction*. For example, from the components μ^a_b of a square matrix one may form the number μ^c_c, its trace.

Three points should be made about the way the summation convention is employed in this book. In the first place, we have so arranged matters that the pair of repeated indices implying a summation will (almost always) occur with one index in the upper position and one in the lower. This will already be apparent from the way we have chosen to write the matrix equation above, when some such thing as $v_\alpha = \lambda_{\alpha a} u_a$ might have been expected. The point is related to the importance of distinguishing between a vector space and its dual (column versus row vectors) mentioned several times earlier in this chapter. This distinction is introduced into the notation for components by using an index in the upper position (u^a, v^α) for components of a column vector. For the components of a row vector we shall place the index in the lower position, thus: c_α. Then the multiplication of the matrix λ by a row vector c (of length m), on the left, gives a row vector (of length n) whose components are $c_\alpha \lambda^\alpha_a$. Notice that the type of the resulting vector (row rather than column) is correctly indicated by the position of the free index a.

The pairing of a row and a column vector (in other words, a $1 \times m$ and an $m \times 1$ matrix) by matrix multiplication, mentioned in Section 1, is represented by an expression $c_\alpha v^\alpha$, which conforms to our rule. On the other hand, the scalar

product of two column vectors, $\sum_{\alpha=1}^{m} v^\alpha w^\alpha$, cannot be correctly so represented without the introduction of further quantities. What is required is a two-index object, say $\delta_{\alpha\beta}$, with

$$\delta_{11} = \delta_{22} = \ldots = \delta_{mm} = 1, \text{ but } \delta_{\alpha\beta} = 0 \text{ if } \alpha \neq \beta;$$

with the aid of this the expression $\delta_{\alpha\beta} v^\alpha w^\beta$ can be correctly formed. This has the same value as $\sum_{\alpha=1}^{m} v^\alpha w^\alpha$; but the point of the remark is to show again, this time through the application of the summation convention, how the pairing of vectors and duals differs from a scalar product. The extra piece of machinery required in the case of the scalar product, represented by $\delta_{\alpha\beta}$ above, is the Euclidean metric.

The second point we should mention about our use of the range and summation conventions is that, whereas in tensor calculus they are used almost exclusively with indexed quantities which are collections of numbers (or functions), we shall use them with other types of object. For example, basis vectors for an n-dimensional vector space may be written $\{e_a\}$, where a ranges and sums from 1 to n; then any vector u in the space may be written $u = u^a e_a$, where the u^a, its components with respect to the basis, are numbers, but the e_a are vectors.

The third point to watch out for is that an expression such as (x^c) is frequently used to stand for $(x^1, x^2, \ldots, x^n)$. Furthermore, the value of a function of n variables, say f, at (x^c) will be denoted $f(x^c)$. In this situation the index c is subject neither to the summation nor to the range convention. In such a context (x^c) is usually to be thought of as the set of coordinates of a point in some space. Where elements of $\mathbf{R}^n$ are being used as coordinates rather than as the components of velocity vectors or differentials, for example, the distinctions made earlier between vector spaces and their duals, or between column and row vectors, no longer have the same importance.

Note to Chapter 0

Klein's remark is in his splendid *Elementary mathematics from an advanced standpoint, part II, Geometry* (Klein [1939]) p 25.

1. AFFINE SPACES

When one first begins to learn mechanics one is confronted with a space—the "ordinary" space in which mechanical processes take place—which in many ways resembles a vector space, but which lacks a point playing the distinctive role of zero vector. The resemblance lies in the vector addition properties of displacements and of quantities derived from them such as velocities, accelerations and forces. The difference lies in the fact that the mechanical properties of a system are quite independent of its position and orientation in space, so that its behaviour is unaffected by choice of origin. Of course the Sun, or the mass centre of the Solar System, plays a role in the formulation of the Kepler problem of planetary motion, but the relative motion of the planets does not depend on whether displacements are measured from the Sun or from some other point. Nor does it depend on the choice of origin for time measurements.

The same is true in special relativity theory. Here also the behaviour of a physical system is unaffected by the choice of space-time origin.

In neither case can there be ascribed to any point the distinctive position and properties ascribed to the zero vector in a vector space; nor can any meaning be given to the addition of points as if they were vectors. Nevertheless, one learns to manipulate vectors in ordinary Euclidean space or in Minkowski space-time and to give physical significance to these manipulations, without perhaps paying too much attention to the precise nature of the underlying space or space-time. When one wants to be more systematic, however, it is necessary to establish the precise relation between the vectors and the space. A satisfactory construction must allow for vector addition of displacements but may not single out any point with special properties. The result is called an affine space.

It is true that the limitations imposed by formation in an affine mould are too severe for some applications. This became apparent during the course of the nineteenth century, when various generalisations were developed. One line of development culminated in the work of Ricci and Levi-Civita on the tensor calculus, which was exploited by Einstein in the invention of general relativity theory; another line led to the work of Lie in group theory, another to the work of É. Cartan in differential geometry, yet another to the work of Poincaré and Birkhoff in celestial mechanics. The generalisations which were developed include much of the subject matter of the later part of this book (and much else). To a great extent these generalisations may be attained by modifying one or another property of an affine space, so we start with that. Most of the techniques needed in the later work may be explained quite easily in the affine case and extended without much effort. The more general spaces introduced later are called "manifolds". They are defined in Chapter 10. In the first nine chapters we shall develop the differential geometry of affine spaces in a form suitable for applications and adaptable to generalisation.

To start with, the concepts to be explained do not require assumptions of a metrical character—no scalar product or measure of length is required—and so they will be applicable later on in both the Euclidean and the Minkowskian contexts.

1. Affine Spaces

In this section we define affine spaces and introduce coordinates natural to them called affine coordinates.

Affine space defined. We are to define a space $\mathcal{A}$ in which displacements may be represented by vectors. As a model for displacements we shall take a real vector space $\mathcal{V}$ of finite dimension n. We shall not choose any particular basis in the vector space $\mathcal{V}$, so it is not merely a fixed copy of the real number space $\mathbf{R}^n$. From experience in mechanics, one might hope that displacements in $\mathcal{A}$ would enjoy these properties:

(1) a succession of displacements is achieved by the addition of vectors (the triangle law for displacements)

(2) displacement by the zero vector leaves a point where it is

(3) if any two points are chosen, there is a unique displacement from one to the other.

A formal definition which embodies these properties may be given in terms of the effect on any point $x \in \mathcal{A}$ of displacement by the vector $v \in \mathcal{V}$. We shall write $x + v$ to denote the point to which x is displaced, and regard the operation of displacement as a map $\mathcal{A} \times \mathcal{V} \to \mathcal{A}$ by $(x, v) \mapsto x + v$. The definition runs as follows: a set $\mathcal{A}$ is called an *affine space* modelled on the real vector space $\mathcal{V}$ if there is a map, called an *affine structure*, $\mathcal{A} \times \mathcal{V} \to \mathcal{A}$, denoted additively: $(x, v) \mapsto x + v$, with the properties

(1) $(x + v) + w = x + (v + w)$ for all $x \in \mathcal{A}$ and all $v, w \in \mathcal{V}$

(2) $x + 0 = x$ for all $x \in \mathcal{A}$, where $0 \in \mathcal{V}$ is the zero vector

(3) for any pair of points $x, x' \in \mathcal{A}$ there is a unique element of $\mathcal{V}$, denoted $x' - x$, such that $x + (x' - x) = x'$.

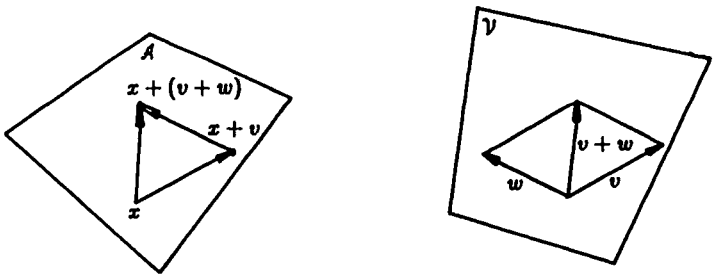

Fig. 1 Points and displacements in an affine space.

Note that the plus sign is given two distinct meanings in this definition: in the expression $x + v$ it denotes the displacement of a point of $\mathcal{A}$ by an element of $\mathcal{V}$, while in the expression $v+w$ it denotes the addition of two elements of $\mathcal{V}$. Moreover, although the displacement from x to x' is denoted $x' - x$, there is no sense in which points of $\mathcal{A}$ may be added together. It is essential to keep these distinctions clearly in mind when working with an affine space.

An affine space $\mathcal{A}$ modelled on a vector space $\mathcal{V}$ of dimension n is also said to be of dimension n. One writes $\dim \mathcal{A} = n$ (and $\dim \mathcal{V} = n$).

Exercise 1. Let $\mathcal{A}$ be an affine space modelled on a vector space $\mathcal{V}$. Show that if, for some $x \in \mathcal{A}$ and some $v \in \mathcal{V}$, $x + v = x$, then $v = 0$. Thus displacement by any vector other than the zero vector leaves no point where it is. □

Exercise 2. Let x_0 be any chosen point of $\mathcal{A}$. Show that the map $\mathcal{A} \to \mathcal{V}$ by $x \mapsto x - x_0$ is bijective (onto and 1 : 1). □

Affine coordinates. In order to deal with specific problems one is likely, sooner or later, to want to introduce coordinates into affine spaces. It is true that one can sometimes go a long way in solving mechanics problems without using coordinates, but even so an adept choice of coordinates may much simplify a problem. The same is true in special relativity theory. What is desirable, on the other hand, is to formulate the problem in the first place without using coordinates, so as to be able to recognise whatever symmetry or invariance properties it may have.

Among all the coordinates which may be introduced into an affine space there are those, called affine coordinates, which are especially well-adapted to its structure. These coordinates will be explained here.

A choice of affine coordinates requires only a choice of origin and a choice of axes. The choice of axes is merely a choice of basis for the underlying vector space. If x_0 is any chosen point of $\mathcal{A}$ and $\{e_1, e_2, \ldots, e_n\}$ is any chosen basis for $\mathcal{V}$ then any point x in $\mathcal{A}$ may be written $x = x_0 + (x - x_0)$, and since $x - x_0$ belongs to $\mathcal{V}$ it may in turn be written $x - x_0 = x^a e_a$, where the x^a are its components with respect to the chosen basis $\{e_a\}$. Here for the first time we employ the summation convention explained in Section 4 of Chapter 0: because the index a is repeated, summation from 1 to n (= $\dim \mathcal{A}$) is understood. Thus $x^a e_a$ is short for $\sum_{a=1}^{n} x^a e_a$.

The components $(x^1, x^2, \ldots, x^n)$ are called the *affine coordinates* of x. The point x_0, whose coordinates are evidently $(0, 0, \ldots, 0)$, is called the *origin* of affine coordinates. An assignment of coordinates associates a set of n real numbers $(x^1, x^2, \ldots, x^n)$ to each point of $\mathcal{A}$, and so may be described by a bijective map $\mathcal{A} \to \mathbf{R}^n$ (compare Exercise 2). Thus the dimension of an affine space, like the dimension of the vector space on which it is modelled, is equal to the number of coordinates needed to specify a point of the space.

The notion of dimension and the possibility of using coordinates carry over to manifolds; however, one distinctive property of affine coordinates—that they are valid everywhere in the space at once—will not generalise.

We shall for the time being use $(x^1, x^2, \ldots, x^n)$, often abbreviated to (x^a), to denote the affine coordinates of a general point of $\mathcal{A}$ relative to some system of affine coordinates. Each choice of origin of $\mathcal{A}$ and of basis of $\mathcal{V}$ determines a choice