

Chapter 1

Models and Ideas of Classical Mechanics

1.1 Orientation

An introductory section is usually written to convince readers that their lives will be incomplete unless they buy the book. At some risk, the present authors would like to state that a reader will profit most from this book if he or she seeks a clearer view of continuum mechanics from two concurrent viewpoints: that of the engineer, and that of the mathematician.

Continuum mechanics, which began with simple problems in hydromechanics and the strength of materials, spans multiple theories including elasticity, plasticity, and viscoelasticity. It employs models describing not only how objects deform under load, but their thermal, electric, and magnetic properties. The various sub-theories within continuum mechanics can reach high degrees of complexity. We will ultimately focus on the linearized theory of elasticity, a departure point for many extensions of basic continuum mechanics.

Any engineer will know at least some elements of continuum mechanics. (It is worth noting that James Clerk Maxwell utilized notions from hydrodynamics when formulating his famous equations of electromagnetism.) The understanding of a typical applied mathematician is, however, quite different. To a mathematician working in the theory of shells, say, the whole subject may commence with a statement of the form “The following system of partial differential equations describes a shell in equilibrium. Supplementing these with the boundary conditions, we arrive at the boundary value problem considered in the next 300 pages.” Then the mathematician can forget what was denoted by u or F : he or she can begin to play with equations in a manner completely divorced from physical considerations.

Engineers and mathematicians are therefore unlikely to understand each

other, even when discussing the same problem. The commonality between their worlds is minimal. The purpose of the present book is to address this unfortunate gap. Only a better mutual understanding can further the collaborative efforts of these two technical communities. Mathematicians should understand that the physical models they view as axioms are actually derived under rather crude assumptions. A mathematician can run into trouble by oversimplifying the behavior of a real object or unwittingly introducing highly artificial features into a model. Engineers, on the other hand, should start to understand why mathematicians spend so much time talking about things like the ill-posedness of a problem, or the weak convergence of a sequence of approximations.

Our ideal reader will possess a wealth of curiosity and a desire to apply it to the fascinating gulf that still exists between real physics and rigorous mathematics. With that thought firmly in mind, let us begin.

1.2 Some Words on the Fundamentals of Our Subject

Among the most primitive of the sciences that treat the behavior of bodies in space, the *theory of elasticity* is simple in some respects and complex in others. It describes the motion and deformation of real bodies, but does so by dealing with idealizations. Neglecting atomic structure, the subject treats the motion and deformation of geometrical figures; unlike geometry, however, it attributes the properties of mass and elasticity to the parts of such figures. We might say that the theory of elasticity deals with spatial transformations of geometrical figures having these mechanical properties.

In this book we shall consider some principal models and mathematical questions in the theory of elasticity. Just as pure mathematics had its roots in the ideas of arithmetic and elementary geometry, and developed these so far that a novice may not see connections between the former and the more advanced parts of modern mathematics, elasticity was based on the ideas of *classical mechanics*. Classical mechanics also treats real natural objects, but using highly simplified models. The set of all the models in continuum mechanics constitutes a hierarchy entailing increasing complexity but still resting on the laws of motion and equilibrium of real bodies (which continuum mechanics inherited from classical mechanics). So before proceeding to the theory of elasticity, we should touch on a few essential points from classical mechanics.

First, we should point out that mathematicians are not alone in hav-

ing to deal with abstractions. The objects of mechanics may have more elaborate properties than those of pure mathematics, but they are *still* abstractions. Such properties were assigned only after long experience with the behavior of real bodies. They were also influenced by the mathematical tools available for their study. Essential portions of mathematics, in turn, were developed to meet the needs of mechanics, and the interplay between these subjects is still strong. Although various viewpoints are possible, it can be argued that classical mechanics is now a branch of mathematics. It is common for sciences to branch out into separate areas initially, only to reunite after reaching a more mature stage. Indeed, the ultimate aim of any natural science is the study of one thing: Nature. Some mathematicians believe they study an ideal world, but this latter “world” is an attempt to describe Nature in a certain way.

So what are the objects of classical mechanics? We shall not delve into the notions of space and time here. These may seem simple and evident to students today, but classical mechanicians up to and including Newton did not regard them as such. Although many of the ideas in mechanics were elaborated long before Newton, we frequently refer to him as the founder of classical mechanics. In fact, Newton collected known results and created a general approach to modern mechanics just as Euclid did for geometry.

The most elementary object that exists in equilibrium or moves through the space and time of classical mechanics is the *material point* or *mass point*. We shall refer to it loosely as a *particle*. As with a geometric point, a mass point has no spatial dimension — although it does have finite mass. The notion of *mass* is regarded as primitive (and undefinable) in mechanics. In elementary books we encounter statements to the effect that mass is “the measure of inertia” of a body. But such “definitions” are meaningless (despite the comfort we often take in them).

Next come collections of mass points and, after that, *rigid bodies* (i.e., bodies that cannot be deformed). Newton used the term “corpuscle” instead of “mass point”. He avoided the term “rigid body” as well. But time changes everything and we are discussing the present form of classical mechanics.

In many mechanics books, a rigid body is defined as a collection of mass points whose relative positions are fixed. Even for a body that could realistically be considered as a finite collection of particles, however, the definition is not complete until we specify *how* the particles can interact. But the practical necessity of considering bodies that appear as geometric figures having continuous mass distributions basically forces us to employ

limit passages from finite sets of mass points to continuous bodies. To justify such passages we must bring in additional assumptions which, from a mathematical viewpoint, could be regarded as new axioms of classical mechanics. Overall, however, it is more convenient to take the rigid body itself as a primary notion. We then formulate as *axioms* for a *continuous* rigid body the properties *derived* for a rigid body composed of a finite number of particles. This is implicitly done in almost any book on theoretical mechanics. None of these constructs — point mass, rigid body, etc. — exist in Nature, but all can serve as good approximations to real bodies.

Classical mechanics studies the motion and equilibrium of its objects under the action of *forces*. Forces likewise may not truly exist in Nature, but the force concept gives us a way to describe the effects of bodies or fields on the motion of a given body. Force is another primitive notion in mechanics; it is left undefined, but certain properties are attributed to it. In that sense it is similar to the primitive notions of pure geometry, such as those of point, line, and plane.

With the advent of relativity, classical mechanics lost much of its status as an exact science. This hardly affected its usefulness as an engineering tool. We could maintain, furthermore, that classical mechanics *still is* an exact science in the same sense as mathematics is. Its structure, in fact, is similar to that of a branch of mathematics: it has a set of primitive notions (space, time, particle, force, etc.), as well as a set of axioms. Unlike the axioms of mathematics, however, the axioms of mechanics are sometimes left unstated.

Before embarking on the theory of elasticity, we shall provide an overview of the conceptual base on which it rests. This includes a collection of topics from classical mechanics, along with certain tools of the theory of elasticity that happened to arise in the context of classical mechanics.

1.3 Metric Spaces and Spaces of Particles

Newtonian mechanics considers the motion of mass points and rigid bodies in an absolute space. Of course, this implies that the latter exists and has properties like those of the space of ordinary Euclidean geometry. We call such a space a (*Newtonian*) *reference frame*. If we consider one absolute space in which a system of particles moves, then there exist (infinitely many) other absolute spaces in which this system can be taken as moving. The spaces themselves translate with respect to one another at a constant

velocity. We call any two such frames *inertial reference frames*. Note that one inertial frame cannot rotate with respect to another; rotation about some axis implies that the velocity of various points is proportional to their distances from the axis, and this is obviously not the same for all points.

Newton's first law implies that there is no *preferred* absolute space: no experiment can distinguish one inertial reference frame from another. In particular, it is impossible to determine which reference frame might be "stationary" in an absolute sense. Nonetheless, it is conventional to construct a reference frame that is "stationary with respect to the distant stars" (or even "stationary with respect to the Earth's surface," although any point of that surface executes a complex motion produced by the Earth's rotation about its axis, its revolution about the Sun, etc.). In Newton's time, a good number of stars appeared to be fixed in position, so they were used to mark out a reference frame. Today we know that all stars are in motion, but the idea is still convenient for ordinary calculations.

When an ideal particle has a fixed position in an absolute space, it coincides with a point of the space. The space itself is *isotropic* and *homogeneous* — its properties are the same in all directions at all points — so the only meaningful relation between any two of its points is one of separation distance. We wish to apply the notion of distance to other objects not necessarily related to geometrical space, so for the mass points we will generalize it as follows. Suppose that to any pair of points A and B we assign a nonnegative finite number denoted by $d(A, B)$. In this way we get a function in two variables that is defined for each pair of points in the space. It is called a distance function or, in mathematics, a *metric*, if it satisfies three axioms of the usual distance employed in geometry:

- M1. $d(A, B) \geq 0$, with $d(A, B) = 0$ if and only if A and B coincide;
- M2. $d(B, A) = d(A, B)$;
- M3. $d(A, B) \leq d(A, C) + d(C, B)$, where C is any other point of the space.

Exercise 1.3.1. *Demonstrate that M1 can be changed to " $d(A, B) = 0$ if and only if the points A and B coincide." So $d(A, B) \geq 0$ is a consequence of the altered system of axioms.*

The absolute space is physically empty, composed only of fictitious points. Material points are always associated with material objects; the reference frame is a mental construction for the sake of expediency. Let us take a fixed time instant. Now we can consider only the set S of mass points, which could be finite or infinite, and pair this set with a metric d

that was defined in the absolute space (but is now applied only to mass points). In mathematics, we denote such a pair by the symbol (S, d) and call it a *metric space*. When there is no chance for confusion (i.e., when only one metric d is employed in a given discussion), we loosely refer to S itself as a “metric space”.

The notion of metric space is general and can be used with sets S that do not consist of spatial or mass points. The elements of S can be of any nature if an appropriate metric d can be defined. In the term “metric space”, the word “space” is simply a synonym for “set”. Hence, by definition, even a set consisting of just one mass point is a metric space. Indeed, labeling this point A , we could define the metric by setting $d(A, A) = 0$. The reader can verify that M1–M3 hold in this simple case.

When dealing with metric spaces we often borrow mental pictures from elementary geometry. For example, we can define a *ball* having center x_0 and radius $r > 0$ as the set of points x of the metric space that fall within distance r of x_0 . The ball is *open* if the inequality $d(x, x_0) < r$ is used; it is *closed* if the inequality $d(x, x_0) \leq r$ is used. Sometimes we require the notion of a *neighborhood* of x_0 . By this we mean a subset of the space that contains some open ball with center x_0 and nonzero radius.

We have said that a metric can take any form satisfying the necessary axioms. Consider, for instance, a set of mass points whose motion is confined to the surface of a sphere. In this case it is natural to measure distance along the great circle that connects any two points (of the two possible arcs along the great circle, we must take the shorter one in order to satisfy the metric axioms). This is essentially how we measure ordinary distances between points on the Earth’s surface.

As another example we could consider how distances should be measured in a town where the streets form a uniform rectangular grid. A metric can be defined as the minimal distance between any two points *when measured along the grid lines*. If we introduce Cartesian coordinates in the plane and identify points with these coordinates, e.g.,

$$A = (a_1, a_2), \quad B = (b_1, b_2),$$

then we can represent the “taxicab metric” by the expression

$$d_1(A, B) = |b_1 - a_1| + |b_2 - a_2|. \quad (1.3.1)$$

The reader can also verify that the function

$$d_p(A, B) = (|b_1 - a_1|^p + |b_2 - a_2|^p)^{1/p} \quad (1.3.2)$$

is a valid metric for any fixed $p \geq 1$. When $p = 2$ we get the Euclidean distance. The reader should guard against a tendency to accept such statements without actually checking for satisfaction of the axioms. For the p -metric above, the only nontrivial axiom to check is the “triangle inequality” M3. Satisfaction of this follows from Minkowski’s inequality

$$\left(\sum_{i=1}^m |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^m |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^m |b_i|^p \right)^{1/p} \quad (1.3.3)$$

which holds for any $p \geq 1$ and any two sets of real numbers a_1, \dots, a_m and b_1, \dots, b_m .

Exercise 1.3.2. *Demonstrate that for $0 < p < 1$, the function $d_p(A, B)$ cannot serve as a metric for points on the plane.*

To introduce Cartesian coordinates as we have done above, we must appoint an origin. We noted previously, however, that all points in an absolute Newtonian space stand on an equal footing. So the choice of coordinate origin is arbitrary and has no ultimate physical significance. When we write $A = (a_1, a_2, a_3)$ we in reality introduce a directed line segment extending from the frame origin to the point A . We can denote this segment by $a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ where \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 are unit vectors along the orthogonal frame axes. We can even draw this vector in the geometrical space, where the mass points are, but must keep in mind that it merely symbolizes a correspondence between a certain vector and the position of a point as mentioned above; in particular it does not belong to our initial set of mass points in the space. The reader has surely made use of “position vectors” in solving mechanics problems. The concept is useful because it allows us to impose all the machinery of vector algebra on a space that really possesses only the metric property. In order to make the best possible use of this correspondence between mass points and vectors, we should introduce it in such a way that it is one-to-one and preserves the distance (metric) between pairs of respective elements. A one-to-one correspondence between two metric spaces in which distance is preserved is said to be an *isometric* correspondence.

We started with a simple metric space having no algebraic structure and arrived at another metric space with algebraic structure. We shall continue to work with the latter space, loosely regarding position vectors as points. Again, this is permissible only because of the isometric correspondence mentioned above. Just as there are no position vectors in the space of mass points, there are no mass points in the space of position vectors. It

is natural to employ a mixture of the two different kinds of objects only because we are used to drawing points and vectors on the same plane. But from a mathematical viewpoint, the vectors and points are objects of different natures and are treated using very different tools.

Let us quickly summarize. The principal objects of classical mechanics are mass points, which are described using a reference frame that has been imposed on an idealized absolute space. In this space, a free mass point (i.e., one that is not experiencing forces or collisions with other bodies) maintains both its speed and its direction of motion. From a mathematical viewpoint, however, classical mechanics deals with the images of those mass points under a one-to-one correspondence with a space of position vectors. The rules for working with these vectors are taken from the theory of vector spaces, but are supplemented by the rules of mechanics itself.

1.4 Vectors and Vector Spaces

Many of us were exposed to the vector concept in high school mathematics. Unfortunately, beginning students are prone to assign the term “vector” to any arrow drawn on the chalkboard. A nice demonstration that such an arrow need not represent a vector was given by A.P. Minakov. Sketching the perpendicular intersection of two one-way streets with a shop standing on one corner, Minakov had his students imagine traffic flows of 30 cars per minute down one street and 40 cars per minute down the other. He noted that nothing would prevent anyone from labeling these flows with appropriately sized arrows. He was quick to point out, however, that if these arrows represented vectors then a resultant flow of $(30^2 + 40^2)^{1/2} = 50$ cars per minute would be entering the doors of the shop! Clearly we cannot apply vector addition to just any quantities that happen to be represented by directed line segments. Quantities have a vectorial nature only when we can carry out vectorial operations with them. These include vector addition, subtraction, and multiplication by a scalar.

The lesson here is that one cannot perform mathematical operations on objects without first verifying that these objects share all properties required for validity of the operations. This holds for the formation of a metric and for the treatment of quantities as vectors. So what is a vector — or, more precisely, a linear space of vectors? In mathematics, an element of an n -dimensional Euclidean space is a special object denoted variously by symbols such as \mathbf{x} , \bar{x} , \underline{x} , or \vec{x} . But the mere use of notation does not

automatically make something a vector. With vectors, we must be able to carry out two principal operations: vector addition and scalar multiplication. These operations are, in turn, subject to the axioms of a vector space. Let $V \neq \emptyset$ be a set along with suitably defined operations of addition and scalar multiplication. That is,

- (a) to each pair $x, y \in V$ there corresponds a unique vector $x + y$, and
- (b) to each $x \in V$ and each scalar λ there corresponds a unique vector λx .

This structure — consisting of the elements with two operations of addition and multiplication by scalars (real or complex) — can be a vector space only if the following hold:

- (1) V is algebraically closed with respect to the two operations. That is, $x + y$ and αx both belong to V for any $x, y \in V$ and any scalar α .
- (2) Addition is both commutative and associative; that is, we have

$$x + y = y + x, \quad x + (y + z) = (x + y) + z,$$

for any $x, y, z \in V$.

- (3) There is an additive identity element in V . This unique element is called the *zero vector* and is denoted by 0 ; it has the property that $x + 0 = x$ for any $x \in V$.
- (4) Each $x \in V$ has a unique additive inverse in V . This vector is denoted by $-x$ and has the property that $x + (-x) = 0$.
- (5) If $x, y \in V$ and α, β are any scalars, then

- (1) $\alpha(x + y) = \alpha x + \alpha y$,
- (2) $(\alpha + \beta)x = \alpha x + \beta x$,
- (3) $(\alpha\beta)x = \alpha(\beta x)$.

Moreover, we have $1x = x$.

Of course, these axioms are so simple that they obviously hold for ordinary vectors in two or three dimensions. But such formalization allows us to apply them to more abstract sets. Indeed, the notion of vector space applies not only to sets of forces or position vectors, but also to finite (or infinite) sets of trigonometric polynomials of the form

$$\sum_k (a_k \sin kx + b_k \cos kx).$$

When considered on some interval $a \leq x \leq b$ (finite or infinite), a set of these polynomials can constitute a vector space (of dimension $2n$ if we sum over k from 1 to n only). Here, however, the use of arrows to represent

vectors would be fruitless. We urge the reader to verify the axioms of a vector space for this example and thereby justify labeling trigonometric polynomials as vectors (even though in many situations it would not be advisable to do so).

We have referred to vector space dimension. This notion relates, as the reader knows, to that of linear independence. A set of vectors $\{x_1, \dots, x_m\}$ is said to be *linearly independent* if from the equation

$$c_1x_1 + \dots + c_nx_m = 0$$

with scalar coefficients c_k it follows that $c_1 = \dots = c_m = 0$. The *dimension* n of a vector space is the maximal number linearly independent vectors in the space. A set of n linearly independent vectors is called a *basis* of the n -dimensional space; any vector from the space can be uniquely represented as a linear sum of the basis vectors.

If we cannot find a finite n for the dimension of the space, we call the space *infinite dimensional*. Here the problem of basis is not simple, however. Above we considered the $2n$ -dimensional space of trigonometric polynomials. For some problems this space is of great interest; the trigonometric polynomials are used to represent solutions to differential equations of the hyperbolic or parabolic type (and not only these, of course, but this is where interest in such polynomials originated). But infinite polynomials

$$b_0 + \sum_{k=1}^{\infty} (a_k \sin kx + b_k \cos kx),$$

called *Fourier series*, are also employed. In calculus, these series are considered apart from differential equations. Instead, they are used to represent a 2π -periodic continuous function, and it is shown that the Fourier coefficients a_k and b_k are defined uniquely. The set of continuous 2π -periodic functions is obviously a vector space. Furthermore, it has infinite dimension since a finite set of functions $1, \sin x, \sin 2x, \dots, \sin rx, \cos x, \cos 2x, \dots, \cos rx$ is linearly independent. Thus we have found an infinite set of linearly independent “vectors” (i.e., continuous functions) in the space.

Exercise 1.4.1. *Propose a few metrics over spaces of trigonometric polynomials.*

1.5 Normed Spaces and Inner Product Spaces

It is clear that the space of functions continuous on a segment $[a, b]$ is also an infinite-dimensional vector space. An extension of this idea is the space of vector functions — functions taking values from a vector space such as \mathbb{R}^3 or \mathbb{R}^2 — that depend continuously on some parameter. If the parameter is time t , such a space can be regarded as the set of all continuous trajectories of a point in space for t in some segment such as $[0, T]$. The space of continuous vector functions on $[a, b]$ is important in mechanics. We often must characterize the difference between two trajectories $\mathbf{f} = \mathbf{f}(t)$ and $\mathbf{g} = \mathbf{g}(t)$, not only at a given time instant (which could be done with the metrics we have considered for the mass points), but “in total” on the segment. We could accomplish this with the metric space notion and introduce, say,

$$d(\mathbf{f}, \mathbf{g}) = \max_{t \in [a, b]} |\mathbf{f}(t) - \mathbf{g}(t)|. \quad (1.5.1)$$

Here, instead of the absolute value, we could use any metric on \mathbb{R}^n to characterize the distance between points on the trajectories at the instant t . But the vectorial structure of the space leads us to use a particular kind of metric, one based on the norm that appears in linear algebra.

A *norm* on a vector space is a function that assigns to every element x in the space a finite nonnegative number $\|x\|$. This function must satisfy the following axioms:

- N1. $\|x\| \geq 0$, with $\|x\| = 0$ if and only if $x = 0$;
- N2. $\|\lambda x\| = |\lambda| \|x\|$ for any scalar λ ;
- N3. $\|x + y\| \leq \|x\| + \|y\|$ for any two vectors x, y in the space.

A vector space V , when paired with a norm $\|\cdot\|$, is called a *normed space*. From a mechanical viewpoint, the notion of norm brings in the idea of the *homogeneity* of space. First, if two pairs of elements have equal differences then the norms of these differences will be equal, regardless of the regions of space from which we take the elements:

$$\|(x + z) - (y + z)\| = \|x + z - y - z\| = \|x - y\|.$$

Second, axiom N2 guarantees homogeneity with respect to multiplication by a scalar. The triangle inequality N3 extends the usual triangle axiom of Euclidean space. Note that every normed space is automatically a metric space. Indeed, the function

$$d(x, y) = \|x - y\| \quad (1.5.2)$$

is easily seen to satisfy M1–M3 on page 5.

Exercise 1.5.1. *Prove that N1 can be changed to “ $\|x\| = 0$ if and only if $x = 0$ ”. That is, $\|x\| \geq 0$ is a consequence of the new set of axioms.*

Exercise 1.5.2. *Show that the inequality*

$$\left| \|x\| - \|y\| \right| \leq \|x - y\| \quad (1.5.3)$$

holds for any $x, y \in V$.

It is clear that the metric (1.5.1) is induced by the norm

$$\|\mathbf{f}\|_{C(a,b)} = \max_{t \in [a,b]} |\mathbf{f}(t)|, \quad (1.5.4)$$

and therefore the space of continuous vector functions on $[a, b]$ with this norm is a normed space. It is usually denoted by $C(a, b)$. We stress that this notation indicates not only that a set of continuous vector functions is under consideration, but that the norm (1.5.4) is assumed as well. Perhaps it would be more reasonable to write $C[a, b]$, but our notation is traditional and in this book we deal exclusively with compact domains such as closed and bounded intervals. The subscript $C(a, b)$ is appended to the norm symbol because we shall introduce other norms on the same set of vector functions. For example, the norm

$$\|\mathbf{f}\|_{L^2(a,b)} = \left(\int_a^b |\mathbf{f}(t)|^2 dt \right)^{1/2} \quad (1.5.5)$$

characterizes the difference between two continuous vector functions in an integral rather than a pointwise sense. Another difference between (1.5.5) and (1.5.4) will become apparent when we consider the results of performing limit passages for sequences of elements.

Exercise 1.5.3. *On the set of all functions continuous on $[0, 1]$, introduce*

$$\|f\| = \sup_{x \in [0,1]} \frac{|f(x)|}{x}.$$

Is the result a normed space?

Remark 1.5.1. When we say that some quantity like a norm is “defined on” a space, we mean that it must be defined (hence finite) at every point of the space. \square

There is another integral quantity whose relationship to the norm (1.5.5) is analogous to that between the ordinary dot product in \mathbb{R}^n and the ordinary length of a vector. It is given by

$$(\mathbf{f}, \mathbf{g})_{L^2(a,b)} = \int_a^b \mathbf{f}(t) \cdot \mathbf{g}(t) dt,$$

and we have

$$(\mathbf{f}, \mathbf{f})_{L^2(a,b)} = \|\mathbf{f}\|_{L^2(a,b)}^2.$$

This, along with the linearity of $(\mathbf{f}, \mathbf{g})_{L^2(a,b)}$ with respect to the arguments \mathbf{f} and \mathbf{g} , suggests that we could use this integral form in the same way we use a dot product in \mathbb{R}^n .

Exercise 1.5.4. *Using a uniform Riemann sum approximation to the integral, confirm that the analogy between the dot product and the norm in \mathbb{R}^n really corresponds to the relation between $(f, g)_{L^2(a,b)}$ and $\|f\|_{L^2(a,b)}$ for ordinary functions continuous on $[a, b]$.*

Let us introduce the general case covering such analogies to the dot product. An *inner product space* is a vector space V together with a function (f, g) defined for any pair of elements $f, g \in V$; this function, termed an *inner product*, satisfies the following axioms:

- I1. $(f, f) \geq 0$ for all $f \in V$, with $(f, f) = 0$ if and only if $f = 0$;
- I2. $(g, f) = (f, g)$ for all $f, g \in V$;
- I3. $(\lambda f + \mu g, h) = \lambda(f, h) + \mu(g, h)$ for all $f, g, h \in V$ and real scalars λ, μ .

In much of this book we employ real spaces. It is worth mentioning, however, that for a complex space we need only change axiom I2 to read

$$\text{I2}'. (g, f) = \overline{(f, g)} \text{ for all } f, g \in V$$

and then change the real scalars in I3 to complex scalars.

The inner product structure lets us introduce the idea of orthogonality between elements of a vector space. We say that f and g are *orthogonal* if

$$(f, g) = 0. \tag{1.5.6}$$

This extends the familiar condition $\mathbf{f} \cdot \mathbf{g} = 0$ in \mathbb{R}^3 . In \mathbb{R}^3 , of course, we can go further and introduce the full notion of angle. In general this is not possible; however, the orthogonality idea deserves special mention because it lets us carry out orthogonal projections even in an abstract space.

Exercise 1.5.5. Let e be a unit vector: $(e, e) = 1$. Prove that $f - (f, e)e$ is orthogonal to e . This shows that the operation $(f, e)e$ is analogous to orthogonal projection onto an axis in \mathbb{R}^3 .

The next thing to notice is that an inner product space is automatically a normed space under the *natural norm*

$$\|f\| = (f, f)^{1/2} \quad (1.5.7)$$

where the positive square root is taken. Hence it is also a metric space under the induced metric

$$d(f, g) = \|f - g\| = (f - g, f - g)^{1/2}. \quad (1.5.8)$$

As always, we cannot simply state that $(f, f)^{1/2}$ is a norm; we must prove that the axioms hold. Verification of N1 and N2 is trivial, but this is not the case for N3. The triangle inequality is equivalent to

$$\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2,$$

which, in terms of the inner product, can be rewritten as

$$(x + y, x + y) \leq (x, x) + 2\|x\|\|y\| + (y, y).$$

By I1–I3 we have

$$(x + y, x + y) = (x, x) + 2(x, y) + (y, y).$$

We see that N3 is satisfied if we can establish the

Cauchy–Buniakowski–Schwarz inequality. *We have*

$$|(x, y)| \leq \|x\|\|y\|. \quad (1.5.9)$$

Equality holds if x or y is zero, or if there is a constant c such that $y = cx$.

Proof. Clearly (1.5.9) holds as an equality for $x = 0$. Now assume $x \neq 0$ and consider the vector

$$\begin{aligned} z &= y - \frac{(y, x)}{\|x\|^2}x \\ &= (y, e)e \text{ where } e = \frac{x}{\|x\|}. \end{aligned}$$

By Exercise 1.5.5, we have $(z, x) = 0$. By I1,

$$0 \leq \|z\|^2 = \left(y - \frac{(y, x)}{\|x\|^2}x, y - \frac{(y, x)}{\|x\|^2}x \right) = (y, y) - \frac{(y, x)(x, y)}{\|x\|^2},$$

which is equivalent to (1.5.9) squared. Equality in (1.5.9) holds only when $\|z\| = 0$; this means that $z = 0$, hence $y = cx$ with $c = (y, x)/\|x\|^2$. \square

Exercise 1.5.6. Prove (1.5.9) for a complex space V .

Some equalities from elementary geometry extend to an inner product space. One is the *parallelogram equality*: the sum of the squares of the diagonals of a parallelogram is twice the sum of the squares of its sides. The reader should prove this in abstract form.

Exercise 1.5.7. Show that

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2). \quad (1.5.10)$$

Because in a real inner product space

$$4(x, y) = \|x + y\|^2 - \|x - y\|^2, \quad (1.5.11)$$

we can represent an inner product using only its norm. Clearly the form on the right-hand side of (1.5.11) does not satisfy the axioms of the inner product in just any normed space, hence not every normed space is an inner product space.

It is useful to present one special inner product space.

The space l^2

This space consists of all the real infinite sequences $X = (x_1, x_2, x_3, \dots)$ for which the series

$$\sum_{k=1}^{\infty} |x_k|^2 \quad (1.5.12)$$

converges. The inner product of X with the sequence $Y = (y_1, y_2, y_3, \dots)$ is given by

$$(X, Y) = \sum_{k=1}^{\infty} x_k y_k. \quad (1.5.13)$$

We emphasize that from the set of all infinite sequences we select *only* those for which (1.5.12) is convergent.

The space l^2 is quite special. Let us explain why, using the results and terminology that will appear later in the book. In a Hilbert space H with an orthonormal basis (e_1, e_2, e_3, \dots) (i.e., a separable Hilbert space), any $x \in H$ can be represented in the form

$$x = \sum_{k=1}^{\infty} x_k e_k \quad (1.5.14)$$

with uniquely defined Fourier coefficients $x_k = (x, e_k)$. Moreover

$$\|x\|^2 = \sum_{k=1}^{\infty} |x_k|^2 \quad (1.5.15)$$

and so we obtain a one-to-one correspondence between H and l^2 . This means we could present the entire theory of separable Hilbert spaces using only l^2 .

In a similar fashion we can introduce normed spaces l^p , $p \geq 1$, starting with the same set of sequences X but under the condition that the series $\sum_{k=1}^{\infty} |x_k|^p$ must converge. The norm is given by

$$\|X\|_{l^p} = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \quad (p \geq 1). \quad (1.5.16)$$

The inner product of l^2 can be used with a space l^p for $p > 2$, but such a space is incomplete under the induced (i.e., l^2) norm. We will understand what this means after introducing Banach and Hilbert spaces.

Exercise 1.5.8. *Show that for any integer k the elements 1 , $\sin kx$, and $\cos kx$ are mutually orthogonal in $L^2(-\pi, \pi)$. Find the unit basis vectors of this space and calculate the projections of the above elements along the directions defined by the basis. In this way we obtain the Fourier coefficients of a function given on $(-\pi, \pi)$. The present general viewpoint (with the inner product) allows us to consider other expansions of continuous and discontinuous functions. Such expansions (in particular, involving the use of orthogonal polynomials) are widely used in analysis.*

1.6 Forces

The term “force” lacks a rigorous definition. But we think of force as the quantity that effects the motions or deformations of bodies and characterizes their mutual interaction. That forces have a vectorial nature is also well known (in fact, force was the prototype for the general notion of a vector). By this we mean that the *resultant* of several forces acting on the same particle can be found by vector addition. The original force system can be replaced by its resultant, and the motion of the particle will be unchanged.

Since forces are applied to certain points, the addition of two forces that act on different mass points would be senseless. Therefore the sets of forces acting on different particles constitute different vector spaces.

In mathematics, vector quantities do not carry physical units. Hence their norms, which can be regarded as characterizing their sizes or intensities, are dimensionless numbers. But many mechanical quantities do carry units and this is the case with force. The SI base unit of force is the Newton (N). The need to perform extensive numerical calculations now requires engineers to convert many of their equations and relations to *dimensionless form*: a form in which everything is expressed in relative figures in such a way that variable quantities typically lie near unity. This has its advantages for calculation, but can obscure what happens with real objects. Of course, some routine calculations can be performed entirely by a computer — even to the plotting of final graphs. But the analysis of intermediate and final results is often easier using quantities whose physical meanings are clear.

We will frequently use energy norms. These inherit the dimensional units of the corresponding energy quantities. We will establish various inequalities among the energy norms, and constants will appear in these relations. It is important to understand that such constants often carry dimensional units and would therefore have different values in other unit systems. And, of course, we cannot directly compare the values of constants that have different dimensions.

We will not stop to review the familiar processes of adding or subtracting forces that act on the same particle. However, we should mention some issues concerning rigid bodies. To say that force is a vector is really an oversimplification. With rigid bodies, we run into various modifications of the vector concept: we must distinguish between sliding vectors, free vectors, etc. This is due to the peculiarities inherent in the effects produced by forces acting on rigid bodies (in particular, the possibility of inducing rotation). Before discussing this further, let us recall that when we depict a force vector acting on a body, we in fact superpose pictures for two different spaces: an absolute space, and a space of force vectors. From a logical standpoint, this particular combination of pictures is even “worse” than that of points and their position vectors. In applications, however, convenience always triumphs over formal requirements, so for mathematicians there is no recourse other than attempting to justify such “illegal” actions. Engineers often make use of objects or tools that are imperfect from a mathematical viewpoint. In the more extreme cases, entirely new branches of mathematics have been created in response to this. A good example was the δ -function, used so intensively in physics that it gave rise to the theory of *distributions* or *generalized functions*.

For a system of separate mass points, it is forbidden to shift a force

from one point to another. With a rigid body, because of the constraints connecting the points, it is possible but not completely straightforward: clearly an *arbitrary* shift in the point of application of a force on a rigid body can introduce a rotational tendency that was not present originally. This leads us to consider another characteristic of force: the *moment* it produces about a point. In elementary physics we learn that moment equals “force times lever arm.” This definition suffices for planar structures where clockwise or counterclockwise rotation are the only two possibilities. In general, however, the moment of a force \mathbf{F} about a point O is the vector quantity

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}. \quad (1.6.1)$$

Here \mathbf{r} locates the point of application of \mathbf{F} with respect to O . The reference point O is arbitrary but typically placed at the coordinate origin.

When forces are applied to a particle, the resultant force is the simple vector sum of all forces acting. We cannot distinguish whether the particle moves under the action of some number of forces, or under the action of their resultant. What is the simplest force complex to which we can reduce the action of some set of forces acting on a rigid body, in such a way that the resulting motion (i.e., the acceleration of all points of the body) is indistinguishable from that produced by the original force set? The answer, it turns out, consists of a resultant force and a resultant “couple.”

Let us mention, first, that long experimentation brought physicists to the idea that a force acting on a rigid body can be shifted along its own *line of action*. That is, the point of force application can be moved along this line without affecting the resulting motion.¹ A vector that can be “attached” at any point of its line of action without affecting other characteristics of a problem is called a *sliding vector*. A good deal more is required if we wish to move the point of application off the original line of action. Suppose a force \mathbf{F} acts at a point A on a body and we want to shift this force in a parallel fashion to some other point B (see Fig. 1.1). We begin by introducing a pair of forces $\pm\mathbf{F}$ at B . This is certainly permissible since the effects of these additional forces completely cancel. But now we can regard the pair of forces consisting of \mathbf{F} acting at A and $-\mathbf{F}$ acting at B as a *couple* C . So when we transfer the point of application of the original force \mathbf{F} from A to B , we must, in effect, compensate through the introduction of C . Continuing to refer to Fig. 1.1, we can see that force couples possess some very important properties. First, the resultant force associated with

¹This is not true for a deformable body.

\mathcal{C} is zero. Second, the moment of \mathcal{C} is given by

$$\mathbf{M} = -\overline{AB} \times \mathbf{F}.$$

We can obtain this by considering the sum of the moments of each of the forces introduced above. Indeed, the moment of \mathbf{F} acting at A is $\overline{OA} \times \mathbf{F}$, and the moment of $-\mathbf{F}$ acting at B is $\overline{OB} \times (-\mathbf{F})$; hence the sum of these moments is $\overline{OA} \times \mathbf{F} + \overline{OB} \times (-\mathbf{F}) = -\overline{AB} \times \mathbf{F}$.

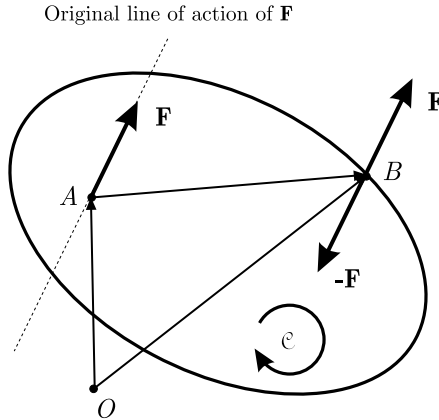


Fig. 1.1 Shifting a force off its line of action on a rigid body. When \mathbf{F} is transferred from point A to point B , the couple \mathcal{C} appears.

The motion of the rigid body under the force \mathbf{F} applied at A is exactly the same as that under \mathbf{F} applied at B in the presence of the couple \mathcal{C} . It turns out, however, that \mathcal{C} is completely characterized by its moment \mathbf{M} . This means we can replace \mathcal{C} by any other force couple (i.e., by any other pair of forces that have equal magnitudes, opposite directions, and non-coincident lines of action) that happens to produce the same moment \mathbf{M} , and the motion of the body will remain the same. Furthermore, we can attach the moment \mathbf{M} to any desired point of the body. Whereas force is an example of a sliding vector, the moment of a couple is a *free vector*. A free vector is one that we can attach to any point of a body without changing the other characteristics of the problem.

Several moments \mathbf{M}_i can be added according to the usual rules of vector addition. This means that any set of forces acting on a rigid body can be replaced by a single resultant force \mathbf{F}_R and a single couple having moment \mathbf{M}_R . Indeed, we can transfer all the given forces to any fixed point D .

Each time a force is moved off its line of action, a couple appears. The force vectors can be summed at D to obtain \mathbf{F}_R , and the moments of the couples can be summed to obtain \mathbf{M}_R . Again, the original set of forces and the pair $(\mathbf{F}_R \text{ at } D, \mathbf{M}_R)$ will be equivalent in the sense that each will have precisely the same effect on the rigid body. We should add that, whereas the magnitude and direction of \mathbf{F}_R do not depend on the point to which it is attached, the corresponding characteristics of \mathbf{M}_R do depend on the line of action of \mathbf{F}_R .

Consideration of the set (space) of forces acting on a rigid body has led us to a “vector addition” different from that used in pure mathematics. In particular, the notion of couple enters the picture as we transfer all force vectors to a common point in order to facilitate their combination. So a force \mathbf{F} applied to a rigid body at some point A possesses characteristics beyond those of an ordinary vector. It has magnitude, a line of action along which it can slide without affecting the resulting motion, and a moment $\mathbf{M} = \overline{OA} \times \mathbf{F}$ about any arbitrary reference point O . From a mathematical viewpoint then, the force vector is a distinctly different object from the vector of formal linear algebra. It does obey a set of well-defined rules, however, and it is somewhat surprising that mathematicians have not seized the opportunity to use these rules as the basis for a new abstract formal system. Possibly this happened because these rules appear to hold only for forces, hence they were left for mechanicians to consider and employ.

Exercise 1.6.1. *Two forces of different magnitude and opposite direction are applied to a rigid body. Can their action be equivalent to only the resultant force (without a couple)? If so, when?*

We have said that mechanics could be based on an explicit set of axioms. Apparently, there have been no serious attempts to select the minimal set of axioms. However, for the equivalence of sets of forces acting on a rigid body, an attempt at axiomatization was made by the Polish mathematician Stephan Banach. Banach initially received an engineering diploma and only afterwards became a mathematician. He lectured on mechanics and even wrote a book (*Mechanics*) on the subject. Although much of this book was written from a traditional mechanical viewpoint, it also contains interesting glimpses of a purely mathematical approach. One passage, concerning the equivalence of sets of forces acting on a rigid body, is worth quoting here:

“In order to deduce the conditions for the equilibrium of a rigid body, we shall assume the following hypotheses:

- I. To a system of forces acting on a rigid body which is in equilibrium we can add (or remove from the system) without disturbing equilibrium:
 - (a) two forces equal in magnitude and acting along the same line, but oppositely directed;
 - (b) several forces having a common point of application and whose sum is zero.
- II. Zero forces balance one another; in other words: if no forces act on a rigid body, then the body can remain in equilibrium.

These hypothesis can be verified experimentally. We shall deduce from them the necessary and sufficient conditions for the equilibrium of forces.”

1.7 Equilibrium and Motion of a Rigid Body

It is impossible to tell whether a rigid body moves under the action of some set of forces (and couples), or only under the action of the pair discussed above (resultant force and resultant couple). If no resultant force acts on a body, it remains in a state of “uniform motion”. But the meaning of this phrase is not as simple as it is in the case of a particle. If the resultant for a mass point is zero, it moves along a straight line at constant speed with respect to an inertial frame (or is in equilibrium when this speed is zero: we can treat both cases in a unified manner and speak only about equilibrium). This is Newton’s first law. But the situation is different for a rigid body. A vanishing resultant is not enough for a rigid body to be in equilibrium with respect to an inertial frame, since the body can rotate.

Suppose we can neglect the size of a rigid body and consider it as a particle. In this case we neglect all couples (their moments become zero), and apply all forces to the particle. Thus we can replace the forces with a single resultant. The mass of the particle should be taken as the whole mass of the rigid body. This is

$$M = \int_V \rho(\mathbf{r}) dV, \quad (1.7.1)$$

where $\rho = \rho(\mathbf{r})$ is the mass density of the body as a function of position, and V is the volume occupied by the body. In this case we obtain the motion of the body in an “integral sense” where rotation is neglected. In actuality the body may rotate, of course, but there is one point whose motion coincides precisely with that of the “equivalent” mass point under the action of the

resultant. This is the *center of mass*, given by

$$\mathbf{r}_M = \frac{1}{M} \int_V \rho(\mathbf{r}) \mathbf{r} dV. \quad (1.7.2)$$

Newton's first law is traditionally formulated for a mass point. To formulate it for a rigid body, we should say that the center of mass moves along a straight line with constant velocity if the resultant force acting on the body is zero. (Again, however, the body may rotate about its center of mass.) In standard textbooks on classical mechanics we find that the linear momentum of the body now remains constant during the motion.

The center of mass is a particularly convenient point to attach the resultant force. The resultant couple is then called the *principal couple*. In classical mechanics it is shown that if the force resultant is zero, then the magnitude and direction of the resultant couple do not depend on the point to which all the forces were referred to obtain the resultant.

Our present goal is to formulate the conditions for equilibrium of a rigid body. Suppose the body is stationary with respect to a stationary frame of the absolute space and that no forces act on it. Since there are no forces, the body remains in equilibrium. But we have said that it is impossible to distinguish whether a rigid body is under the action of some set of forces or under the action of their resultant force and couple. Thus if these latter quantities are together zero, it is equivalent to the case of the absence of any forces; hence a body in equilibrium will remain in equilibrium if the resultant force and couple vanish. The condition that the resultant force and couple both vanish is equivalent to the statement that the body remains in equilibrium. Forces acting on a rigid body that satisfy this condition are said to be *forces in equilibrium*.

We mentioned that if the resultant is zero, the resultant couple does not depend on the point of reduction of the forces. Transferring all the forces so that their lines of action pass through the frame origin, we find that the moment of the corresponding couples equals the moment of the forces with respect to the origin. So the condition for equilibrium of a rigid body can be written as

$$\sum_i \mathbf{F}_i = \mathbf{0}, \quad \sum_i (\mathbf{r}_i \times \mathbf{F}_i) = \mathbf{0}, \quad (1.7.3)$$

where \mathbf{r}_i locates the point of application of \mathbf{F}_i for each i . In component form this gives six equations. Equations (1.7.3) are used for three-dimensional objects. For two-dimensional problems we get three scalar equations: two for the components of the resultant force, and one for the couple.

Equilibrium problems for rigid bodies normally involve geometrical constraints as well as applied forces. Usually the effects of constraints are translated into reaction forces and moments, which are then found using the equilibrium equations (1.7.3). Thus, for a three-dimensional problem involving a rigid body, we can find up to six unknown components of reaction forces or moments, and for a two-dimensional problem only up to three. If a structure consists of several joined parts, we should also introduce reactions in the joints and consider each body as separate under the actions of all forces and reactions. If the number of equations is equal to the number of unknown components and the system of equations is linearly independent, it can be solved and the reactions determined. This is a typical problem of classical mechanics; such problems are common in both civil and mechanical engineering. However, problems where the number of equations is less than the number of reaction components are even more frequent in practice. For such a problem we cannot find the reaction by solving the system of equilibrium equations; instead we must bring in the laws of deformation of the structure and use a model of a deformable body. Among the simplest of such models (though still not simple) is that of a linearly elastic body.

1.8 D'Alembert's Principle

It is clear that the derivatives of the position vectors of the same mass point in two inertial frames differ by the velocity of their relative motion. However, the acceleration of the mass point is the same in both frames and appears in the mathematical formulation of Newton's second law for the motion of a particle having mass m :

$$\mathbf{F} = m \frac{d^2 \mathbf{r}}{dt^2}. \quad (1.8.1)$$

The position vector \mathbf{r} is often called the *radius vector*; it can be drawn in the absolute space as a directed segment starting at the origin of the immovable frame and ending on the mass point. See Fig. 1.2.

The derivative of a vector function $\mathbf{f} = \mathbf{f}(t)$ is defined by analogy with that for a scalar function:

$$\frac{d\mathbf{f}(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t}. \quad (1.8.2)$$

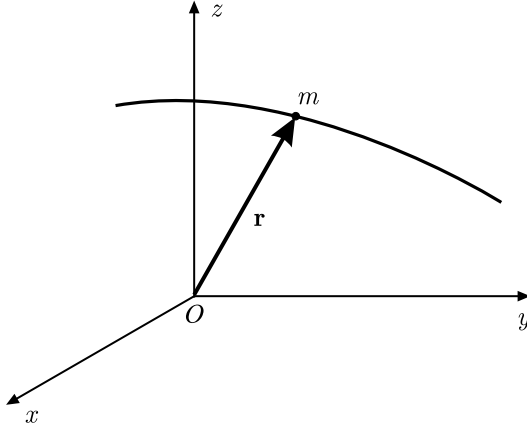


Fig. 1.2 Trajectory of a moving particle.

If we express the function in component form as

$$\mathbf{f} = \sum_{k=1}^3 f^k \mathbf{e}_k \quad (1.8.3)$$

and the basis vectors \mathbf{e}_k do not depend on t , then

$$\frac{d\mathbf{f}(t)}{dt} = \sum_{k=1}^3 \frac{df^k(t)}{dt} \mathbf{e}_k. \quad (1.8.4)$$

We shall now invoke Einstein's summation rule: when we see repeating sub- and superscripts in a term involving components of vectors, such as $a^i b_i$, we should perform a sum over i , with i taking values from 1 to the dimension of the space in which the vectors are considered. We therefore write

$$\frac{d\mathbf{f}(t)}{dt} = \frac{df^k(t)}{dt} \mathbf{e}_k. \quad (1.8.5)$$

Exercise 1.8.1. Write out $d^2\mathbf{f}(t)/dt^2$ in component form when the frame basis is (a) time independent, and (b) time dependent.

Newton's second law for a particle can be rewritten in the form

$$\mathbf{F} - m \frac{d^2\mathbf{r}}{dt^2} = \mathbf{0}. \quad (1.8.6)$$

This simple transformation, after introduction of the notation

$$\mathbf{F}_I = -m \frac{d^2\mathbf{r}}{dt^2}, \quad (1.8.7)$$

brings us to

$$\mathbf{F} + \mathbf{F}_I = \mathbf{0}, \quad (1.8.8)$$

which looks precisely like the equilibrium equation for the particle. This transformation was proposed by d'Alembert. The expression \mathbf{F}_I is called the *inertia force*. Equation (1.8.8) can be regarded as a statement of

d'Alembert's principle. *During the motion, the set of all forces (including the inertia force) forms a system of forces in equilibrium.*

Thus, the problem of finding the acceleration of a particle reduces to the problem of equilibrium of the system of all forces acting on the particle.

Of course, there is little point in using this principle for a free mass point. It does offer advantages, however, when constraints are present (e.g., when points form a rigid body). Let us consider this possibility further.

1.9 The Motion of a System of Particles

Consider the motion of a more complex object: a finite system of n particles. Of course, for each particle we could think of simply writing down Newton's second law. However, we wish to focus on the interaction between particles and the properties that result from this.

The theory for a finite system should inherit some features from the theory for a single mass point. In particular, when we consider the system as a mass unit without extent, as is done for the distant stars, the equations should reduce to those for a mass point. Indeed, we shall see that the center of mass of the system moves exactly as a material point having mass equal to the total mass of the system and acted upon by the resultant force. There is a hierarchy in the theories of classical mechanics.

Let the i th particle have mass m_i and position vector \mathbf{r}_i with respect to the origin of an inertial frame. Consider the forces acting on this particle, including those produced by the actions of the other particles in the system. If the distances between pairs of particles are all preserved during the motion (because of massless constraints), we call the system a rigid body; we shall not, however, limit ourselves to this case. To characterize the system as a whole at a time instant t , let us introduce the total mass

$$M = \sum_{i=1}^n m_i \quad (1.9.1)$$

and the position \mathbf{r}_C of the center of mass defined by

$$M\mathbf{r}_C = \sum_{i=1}^n m_i \mathbf{r}_i. \quad (1.9.2)$$

The position of the i th particle relative to the center of mass is given by the vector

$$\boldsymbol{\rho}_i = \mathbf{r}_i - \mathbf{r}_C. \quad (1.9.3)$$

Then, since (1.9.2) can be written as

$$\sum_{i=1}^n m_i \mathbf{r}_i - \sum_{i=1}^n m_i \mathbf{r}_C = \mathbf{0}, \quad (1.9.4)$$

we have

$$\sum_{i=1}^n m_i \boldsymbol{\rho}_i = \mathbf{0}. \quad (1.9.5)$$

Exercise 1.9.1. Putting $\boldsymbol{\rho}_i = (\xi_i, \eta_i, \zeta_i)$, expand (1.9.5) in a Cartesian frame. Note that in this frame the center of mass is the origin.

Henceforth we shall use an overdot notation for time derivatives. Equation (1.9.2) holds at any instant. Differentiating it with respect to time t , we get

$$M\dot{\mathbf{r}}_C = \sum_{i=1}^n m_i \dot{\mathbf{r}}_i.$$

Consequently,

$$\sum_{i=1}^n m_i \dot{\boldsymbol{\rho}}_i = \mathbf{0}. \quad (1.9.6)$$

A second differentiation gives

$$M\ddot{\mathbf{r}}_C = \sum_{i=1}^n m_i \ddot{\mathbf{r}}_i$$

and

$$\sum_{i=1}^n m_i \ddot{\boldsymbol{\rho}}_i = \mathbf{0}. \quad (1.9.7)$$

These are only kinematical characteristics of the center of mass of the system. They do not depend on the forces acting on the particles.

We divide the forces acting on the system into two classes. The first, the class of *internal forces*, includes forces that arise between the particles of the system because of constraints or other effects. We denote a force acting on the i th particle from the j th particle by \mathbf{F}_{ij} (noting, of course, that $\mathbf{F}_{ii} = \mathbf{0}$ for any i). Secondly, we have the *external forces*. We account for these by simply assuming that a resultant force \mathbf{F}_i acts on the i th particle.

By Newton's third law, the internal forces should be balanced in the sense that

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji}. \quad (1.9.8)$$

As is common in classical mechanics, we *assume* both of these forces share the same line of action connecting the particles. This is a restrictive assumption, but it makes sense in physics where in statics we meet only central forces between particles like gravitation or electrical attraction. Although the equality (1.9.8) makes sense, it is only an assumption; in mathematics we would call it an axiom with far-reaching consequences. Its validity is not so evident, say, for dynamical processes.

By this assumption, the sum of all internal forces is zero:

$$\sum_{i,j=1}^n \mathbf{F}_{ij} = \mathbf{0}. \quad (1.9.9)$$

This is often called d'Alembert's principle.² Although we obtained it from (1.9.8), it may be taken as an independent principle; in fact, it represents a weaker assumption than that in which the forces are assumed to be central and obey Newton's third law. We are about to see that (1.9.9) forms the background for the derivation of some principal conservation laws in mechanics.

Newton's second law for the i th particle is

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i + \sum_{j=1}^n \mathbf{F}_{ij}. \quad (1.9.10)$$

We have noted that the addition of forces acting on different particles is senseless, because the force acting on one particle cannot directly affect another one. However, this operation begins to make sense when we consider a system of particles, as it gives us some characteristics of the system as a whole. We have introduced the mass of the system and the position vector

²It is unrelated to the similarly named principle in § 1.8. d'Alembert's name appears throughout mechanics.

of the center of mass. Let us introduce, similarly, the following characteristics:

(1) the total linear momentum

$$\sum_{i=1}^n m_i \mathbf{v}_i \equiv \sum_{i=1}^n m_i \dot{\mathbf{r}}_i = M \dot{\mathbf{r}}_C; \quad (1.9.11)$$

(2) the total angular momentum

$$\sum_{i=1}^n \mathbf{r}_i \times m_i \mathbf{v}_i \equiv \sum_{i=1}^n \mathbf{r}_i \times m_i \dot{\mathbf{r}}_i; \quad (1.9.12)$$

(3) the resultant of external forces

$$\mathbf{F}_R = \sum_{i=1}^n \mathbf{F}_i; \quad (1.9.13)$$

(4) the total moment of external forces (or total torque)

$$\sum_{i=1}^n \mathbf{r}_i \times \mathbf{F}_i. \quad (1.9.14)$$

First we sum all the equations (1.9.10):

$$\sum_{i=1}^n m_i \ddot{\mathbf{r}}_i = \sum_{i=1}^n \left(\mathbf{F}_i + \sum_{j=1}^n \mathbf{F}_{ij} \right).$$

By (1.9.9) we have

$$\sum_{i=1}^n m_i \ddot{\mathbf{r}}_i = \sum_{i=1}^n \mathbf{F}_i.$$

Hence

$$\frac{d}{dt} \sum_{i=1}^n m_i \mathbf{v}_i = \mathbf{F}_R \quad (1.9.15)$$

or

$$\frac{d}{dt} (M \mathbf{v}_C) = \mathbf{F}_R, \quad \mathbf{v}_C = \dot{\mathbf{r}}_C. \quad (1.9.16)$$

So the motion of the center of mass of the system of particles depends only on the resultant external force acting on the system, and coincides with the motion of a particle of mass M under the same resultant force. From this we obtain a crucial result.

Conservation of linear momentum. If $\mathbf{F}_R = \mathbf{0}$, then the linear momentum remains constant and the center of mass moves along a straight line with constant velocity.

This follows from the fact that

$$\frac{d}{dt}(M\mathbf{v}_C) = \mathbf{0},$$

hence $M\mathbf{v}_C$ is a constant. Similar reasoning brings us to the conservation of total angular momentum. We begin with the elementary transformation

$$\frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \mathbf{v} \times m\mathbf{v} + \mathbf{r} \times m\dot{\mathbf{v}} = \mathbf{r} \times m\ddot{\mathbf{r}}, \quad \mathbf{v} = \dot{\mathbf{r}}. \quad (1.9.17)$$

Next, cross the vector \mathbf{r}_i into both sides of (1.9.10) and sum over i :

$$\sum_{i=1}^n (\mathbf{r}_i \times m_i \ddot{\mathbf{r}}_i) = \sum_{i=1}^n (\mathbf{r}_i \times \mathbf{F}_i) + \sum_{i=1}^n \mathbf{r}_i \times \left(\sum_{j=1}^n \mathbf{F}_{ij} \right).$$

By (1.9.17), the sum on the left is the time derivative of the total angular momentum:

$$\frac{d}{dt} \sum_{i=1}^n (\mathbf{r}_i \times m_i \mathbf{v}_i).$$

The first sum on the right is the total moment of the external forces. Consider the second sum on the right. Together with the term $\mathbf{r}_i \times \mathbf{F}_{ij}$ it contains a dual term $\mathbf{r}_j \times \mathbf{F}_{ji}$. By (1.9.8) their sum is

$$\mathbf{r}_i \times \mathbf{F}_{ij} + \mathbf{r}_j \times \mathbf{F}_{ji} = (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ij} = \mathbf{0},$$

since \mathbf{F}_{ij} is assumed to be parallel to the vector $\mathbf{r}_i - \mathbf{r}_j$ connecting the i th and j th particles. So the second sum on the right is zero and we have

$$\frac{d}{dt} \sum_{i=1}^n (\mathbf{r}_i \times m_i \mathbf{v}_i) = \sum_{i=1}^n (\mathbf{r}_i \times \mathbf{F}_i), \quad (1.9.18)$$

which involves only the external forces. In particular we have the following.

Conservation of total angular momentum. If the total moment of the external forces acting on a system of n particles is zero,

$$\sum_{i=1}^n (\mathbf{r}_i \times \mathbf{F}_i) = \mathbf{0},$$

then the total angular momentum

$$\sum_{i=1}^n (\mathbf{r}_i \times m_i \mathbf{v}_i)$$

remains constant.

Exercise 1.9.2. *Supply the proof.*

Equation (1.9.18) describes the total angular momentum with respect to the origin of the inertial frame. Let us find the total angular momentum with respect to the center of mass of the system. We substitute

$$\mathbf{r}_i = \mathbf{r}_C + \boldsymbol{\rho}_i, \quad \mathbf{v}_i = \mathbf{v}_C + \dot{\boldsymbol{\rho}}_i, \quad \mathbf{v}_C = \dot{\mathbf{r}}_C,$$

into (1.9.18):

$$\frac{d}{dt} \sum_{i=1}^n [(\mathbf{r}_C + \boldsymbol{\rho}_i) \times m_i(\mathbf{v}_C + \dot{\boldsymbol{\rho}}_i)] = \sum_{i=1}^n [(\mathbf{r}_C + \boldsymbol{\rho}_i) \times \mathbf{F}_i].$$

This, after simple transformation, brings

$$\begin{aligned} & \frac{d}{dt} \left\{ \mathbf{r}_C \times \sum_{i=1}^n m_i \mathbf{v}_C + \mathbf{r}_C \times \sum_{i=1}^n m_i \dot{\boldsymbol{\rho}}_i + \sum_{i=1}^n m_i \boldsymbol{\rho}_i \times \mathbf{v}_C + \sum_{i=1}^n (\boldsymbol{\rho}_i \times m_i \dot{\boldsymbol{\rho}}_i) \right\} \\ &= \mathbf{r}_C \times \sum_{i=1}^n \mathbf{F}_i + \sum_{i=1}^n (\boldsymbol{\rho}_i \times \mathbf{F}_i). \end{aligned}$$

Using (1.9.5) and (1.9.6) we get

$$\left[\frac{d}{dt} (\mathbf{r}_C \times M \mathbf{v}_C) - \mathbf{r}_C \times \mathbf{F}_R \right] + \frac{d}{dt} \sum_{i=1}^n (\boldsymbol{\rho}_i \times m_i \dot{\boldsymbol{\rho}}_i) = \sum_{i=1}^n (\boldsymbol{\rho}_i \times \mathbf{F}_i).$$

The bracketed difference on the left is zero, which follows from (1.9.16) if we form the cross product with \mathbf{r}_C and use (1.9.17). Thus

$$\frac{d}{dt} \sum_{i=1}^n (\boldsymbol{\rho}_i \times m_i \dot{\boldsymbol{\rho}}_i) = \sum_{i=1}^n (\boldsymbol{\rho}_i \times \mathbf{F}_i). \quad (1.9.19)$$

Although (1.9.18) and (1.9.19) are similar in form, the latter shows no dependence on the motion of the center of mass. Hence the rotation of the system about the center of mass is independent of the motion of the center of mass. This holds for a system of particles and for a rigid body in particular.

Finally, we mention that by introducing inertial forces for each particle we can rewrite the equations of motion for the system in a form that coincides formally with the equations of equilibrium for the system. This set of equations constitutes, as in the case of one particle, d'Alembert's principle.

1.10 The Rigid Body

Let us reconsider the notion of a rigid body. In many books on theoretical mechanics, a rigid body is defined as a set of particles connected in such a way that their mutual distances are fixed. While this may seem acceptable, it fails to specify how the particles can react with one another (i.e., with something akin to (1.9.8)). The “definition” of a rigid body in mechanics should include not only the idea of constant shape, but the mechanism of force transmission between parts of the body.

The same textbooks often derive results for a rigid body consisting of a finite number of particles, then pass directly to the theory of a “continuous” body. It is supposed that this transition is done through an elementary limit passage. They first think of approximating the body as a finite collection of “particles”; each particle is actually an elemental volume over which the “mass density” is essentially constant. Assuming all mutual distances between particles are fixed during any motion, they calculate the total linear and angular momenta of the system. These are expressed in terms of finite summations, which are taken to become Riemann volume integrals during the subsequent limit passage. The listener is expected to accept the entire procedure without question, especially if he or she has a good calculus background. It turns out, however, that there is reason for concern: when we pass from a system containing finitely many objects to one containing infinitely many objects, we can encounter unexpected changes in qualitative behavior. In particular, we must question whether the properties of the internal forces should carry over. The interior state of a deformable solid cannot be described using only “central forces” (i.e., forces such as the \mathbf{F}_{ij} that we used to describe a finite system of particles) or, indeed, using forces alone. We must employ a stress tensor: an entity that inherits some properties of force, but with other properties of its own.

The behavior of forces inside a rigid body cannot be derived straightforwardly from the corresponding picture for a finite system of particles. Rather, it must be formulated as a kind of axiom that reflects our primary interest in determining the integral characteristics of the motion. In particular, we can take the “simplest” case where the relations between parts of the body are thought to mimic those that would apply to the internal forces in a finite system. But again, this amounts essentially to formulating an axiom. We find that to describe the motion of a rigid body (which we consider as a system of n particles at fixed mutual separations while preserving our assumption on the internal forces \mathbf{F}_{ij}), it is necessary to invoke

only (1.9.16) and (1.9.18), or equivalently

$$\frac{d}{dt}(M\mathbf{v}_C) = \mathbf{F}_R, \quad \frac{d}{dt} \sum_{i=1}^n \boldsymbol{\rho}_i \times m_i \dot{\boldsymbol{\rho}}_i = \sum_{i=1}^n \boldsymbol{\rho}_i \times \mathbf{F}_i. \quad (1.10.1)$$

Denoting

$$\mathbf{G}_C = \sum_{i=1}^n \boldsymbol{\rho}_i \times m_i \dot{\boldsymbol{\rho}}_i, \quad \mathbf{M}_C = \sum_{i=1}^n \boldsymbol{\rho}_i \times \mathbf{F}_i,$$

where \mathbf{G}_C is the angular (kinetic) momentum of the body with respect to the center of mass C and \mathbf{M}_C the resultant angular moment of the external forces with respect to C , we can rewrite (1.10.1) as

$$\frac{d}{dt}(M\mathbf{v}_C) = \mathbf{F}_R, \quad \frac{d}{dt}\mathbf{G}_C = \mathbf{M}_C. \quad (1.10.2)$$

Using the kinematical analysis for a rigid body, which can be found in any textbook on classical mechanics, we can express \mathbf{G}_C in terms of the angular velocity vector $\boldsymbol{\omega}$ for the body, which has three components in three-dimensional space. So in scalar form, we get six simultaneous differential equations in the six unknown components of \mathbf{v}_C and $\boldsymbol{\omega}$. These components are uniquely defined by the equations and corresponding initial values, so we have a description of the motion of a rigid body. While the form of these equations seems simple, the study of rigid body motion occupies much space in a typical textbook. Our goal is to discuss only some main ideas.

We should add that here the use of a Cartesian frame for the description of relative motion of the body is not the best choice. And this leads us to the idea that to describe the motion of bodies, we must introduce other parameters such as the familiar Euler angles α, β, γ . In fact this is the first step toward Lagrangian mechanics. The latter is a consequence of Newtonian mechanics that makes use of generalized coordinates to describe the motion of objects. It simplifies the solution of many problems.

We may regard (1.10.1) as consequences of the dynamic equations for a system of particles. But normally (1.10.2) are applied to continuous rigid bodies that occupy volumes or that are idealized as surface or linear mass distributions. In such cases we may consider the terms of (1.10.1) as Riemann-sum-type approximations to corresponding volume, surface, or line integrals and obtain (1.10.2) using a limit passage. However, these equations must be regarded as new axioms for rigid body mechanics. Indeed, internal force terms such as \mathbf{F}_{ij} are absent. This results from the assumptions on the nature of the internal forces. Inside a rigid body, however, it is strange to make any such assumptions; rather, it is preferable to

begin with (1.10.2) as we will eventually use them to obtain relations for deformable bodies where such simple assumptions cannot be applied at all. So we will regard (1.10.2) as essentially axiomatic in nature.

Unique specification of the position of a rigid body requires six independent parameters. In contrast, the position of a particle in space can be specified by three coordinates or other parameters that define these coordinates uniquely. The position of a system consisting of some number of particles and rigid bodies can be uniquely defined through some minimal number of parameters. In this case we speak of the *configuration* of the system; the space in which these parameters take their values is known as the *configuration space*. The *minimal* number of *independent* parameters describing the system uniquely is called the *degree of freedom* of the system.

The degree of freedom is less than or equal to the total number of parameters that describes each item of the system. For example, if a particle can move only on some surface, then the description of its position requires just two parameters: coordinates of the point on the surface. If it can move only along some curve, a single parameter is required. The same holds for rigid bodies: if a point of a rigid body is fixed, we need just three parameters (e.g., the angles that define its position uniquely). Therefore the degree of freedom of a body with a fixed point is three.

The degree of freedom of a system of bodies in mechanics plays the same role for its configuration space as the dimension plays for a vector space: it shows the number of quantities we must know in order to determine some object in the space uniquely. The configuration space is not a vector space.

Exercise 1.10.1. *What is the degree of freedom of a moving “rigid” segment?*

1.11 Motion of a System of Particles; Comparison of Trajectories; Notion of Operator

A curve describing the motion of a particle in space is called a *trajectory*. If we mark, in configuration space, the points taken by a system of particles during its motion, we have what could be called the system trajectory. Since the ordinary differential equations of motion for particles are of second order, to define the motion uniquely we must provide two initial conditions

for each component of the position vector for each particle. We have

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i + \sum_{j=1}^n \mathbf{F}_{ij} \quad (i = 1, \dots, n), \quad (1.11.1)$$

where the forces on the right can depend on time t and the position vectors of the particles (we continue to denote by \mathbf{F}_{ij} the force exerted on the i th particle by the j th particle and take $\mathbf{F}_{ii} = \mathbf{0}$). We typically specify, for each particle, the position and velocity vectors at some initial time:

$$\mathbf{r}_i(0) = \mathbf{a}_i, \quad \dot{\mathbf{r}}_i(0) = \mathbf{b}_i \quad (i = 1, \dots, n), \quad (1.11.2)$$

where $\mathbf{a}_i, \mathbf{b}_i$ are given constants describing the initial position and velocity of the i th particle, respectively. The resulting problem is called a *Cauchy problem* or *initial value problem* in ordinary differential equation theory. The theory of Cauchy problems offers theorems covering existence and uniqueness of solution and providing for the continuous dependence of solutions on small changes in the initial conditions, masses, and external forces. There are problems to which these theorems do not apply, but they usually suffice for applications (note that the forces can depend on the \mathbf{r}_i and their derivatives, so the equations can be complex).

In other kinds of problems, conditions are posed at various time instants. Such problems are called *boundary value problems*, because the conditions are commonly posed at two points often designated as “initial” and “final” points (although more than two points may be involved). Boundary value problems can have nonunique solutions, and their theory is not as clean as that of Cauchy problems. Nonetheless, a great deal of effort has been directed towards them and much is now known about both their theoretical and practical treatment.

Suppose we are dealing with a particle system whose motion problem has a unique solution under given initial or boundary conditions and under a set of forces. If these parameters change, so does the system trajectory or velocity. How should we measure this change on the time interval $[0, T]$? If we are interested only in the difference between the particle positions, we can measure the deviation between two trajectories $\mathbf{r} = \mathbf{r}_1(t)$ and $\mathbf{r} = \mathbf{r}_2(t)$ using a *uniform norm*. For a single particle we can use

$$\|\mathbf{r}_2 - \mathbf{r}_1\| = \max_{t \in [0, T]} \|\mathbf{r}_2(t) - \mathbf{r}_1(t)\|_{\mathbb{R}^3}, \quad (1.11.3)$$

where $\|\cdot\|_{\mathbb{R}^3}$ is a norm on \mathbb{R}^3 . We require continuity of the vector functions $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ on the finite segment $[0, T]$, which is guaranteed by general theorems covering many motion problems for systems of particles having

unique solutions. The rectilinear motion of a particle with given initial and final times a and b brings us to the uniform norm on the *space of continuous functions* on the segment $[a, b]$. The resulting normed space is denoted by $C(a, b)$. The norm of a function $f = f(t)$ in this space is given by

$$\|f\|_{C(a,b)} = \max_{t \in [a,b]} |f(t)|. \tag{1.11.4}$$

Later, when we consider a deformable body occupying a volume V , it will be convenient to introduce the uniform norm for functions $f = f(\mathbf{x})$ continuous on V . We will take V to be *compact* (i.e., closed and bounded) and use

$$\|f\|_{C(V)} = \max_{\mathbf{x} \in V} |f(\mathbf{x})| \tag{1.11.5}$$

for the norm.

Exercise 1.11.1. *Verify that (1.11.5) satisfies the norm axioms.*

If we must evaluate differences in velocities as well as positions, then for continuously differentiable vector functions $\mathbf{r} = \mathbf{r}_1(t)$ and $\mathbf{r} = \mathbf{r}_2(t)$ we can introduce another norm involving first derivatives. For the one-dimensional case, a norm on the space of functions continuously differentiable on $[a, b]$, denoted by $C^{(1)}(a, b)$, can be introduced in various ways. These include

$$\|f\|_{C^{(1)}(a,b)} = \max_{t \in [a,b]} |f(t)| + \max_{t \in [a,b]} |f'(t)| \tag{1.11.6}$$

and

$$\|f\|_{C^{(1)}(a,b)} = \max \left\{ \max_{t \in [a,b]} |f(t)|, \max_{t \in [a,b]} |f'(t)| \right\}, \tag{1.11.7}$$

which are equivalent in the sense of convergence of sequences of differentiable functions. (We will formalize this notion of equivalent norms in Definition 1.12.1.) They can be extended to the space of functions having continuous derivatives up to order m on a compact set $V \subset \mathbb{R}^n$. We have

$$\|f\|_{C^{(m)}(V)} = \max_{\mathbf{x} \in V} |f(\mathbf{x})| + \sum_{|\alpha| \leq m} \max_{\mathbf{x} \in V} |D^\alpha f(\mathbf{x})|, \tag{1.11.8}$$

where the *multi-index notation* D^α is understood as follows:

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}, \quad \alpha = (\alpha_1, \dots, \alpha_n), \quad |\alpha| = \alpha_1 + \dots + \alpha_n.$$

Such norms are used when we must characterize a solution beyond simply a deviation under a change in parameters of the problem. The importance

of smoothness follows from the fact that the smoother a solution (i.e., the larger the number m in $C^{(m)}(V)$), the better the convergence of approximation schemes used to seek the solution numerically. Note that the norms are written out for dimensionless variables; otherwise, we would append coefficients to account for differences in units between the terms.

Exercise 1.11.2. *Verify that (1.11.8) satisfies the norm axioms.*

Let us return to the initial value problem (1.11.1)–(1.11.2) and suppose that the \mathbf{F}_i depend only on t . Assume these vector functions can be defined independently. Then to a set of forces $\{\mathbf{F}_i\}$ given on $[0, T]$ there corresponds a unique set of trajectories $\{\mathbf{r}_i\}$ on $[0, T]$. This reminds us of the definition of a function, where to each point of some set called the domain of the function there corresponds a unique value of the function. In this case, however, a set of vector functions $\{\mathbf{F}_i\}$ stands in correspondence with another set of vector functions $\{\mathbf{r}_i\}$. We cannot say that to a value of $\{\mathbf{F}_i\}$ at an instant t there corresponds a value $\{\mathbf{r}_i\}$ at the same instant t , since $\{\mathbf{r}_i\}$ at t depends on all the values taken by $\{\mathbf{F}_i\}$ on $[0, t]$. This dependence cannot be described in terms of an ordinary function. We shall call the correspondence an *operator* (or sometimes *mapping*, or *map*) if it is uniquely defined.

To really define a function, we must specify not only a rule of correspondence between two sets, but the sets as well. The same holds for an operator. An operator A has a domain $D(A)$ and a range $R(A)$. In this book $D(A)$ will be a subset of a normed space X , while $R(A)$ will lie in a normed space Y (which may coincide with X). The correspondence itself, taking each point $x \in D(A)$ into a uniquely defined point $y \in R(A)$, will be denoted

$$y = A(x). \quad (1.11.9)$$

We say that A *acts from* X *to* Y . If $Y = X$, we say that A *acts in* X .

A mapping A acting from a normed space X to the scalars \mathbb{R} or \mathbb{C} is called a *functional*. The branch of mathematics known as *functional analysis* had its origins in the study of such mappings.

In linear algebra we treat operators represented by matrices. An $n \times n$ matrix M applied to a vector $\mathbf{x} \in \mathbb{R}^n$ yields another vector $\mathbf{y} \in \mathbb{R}^n$. The reader must be aware that a matrix is not an operator but only a component representative of the operator corresponding to some basis in \mathbb{R}^n . The matrix elements play the same role as the components of a vector in a fixed basis: when we change the basis (and we can do this independently in the domain \mathbb{R}^n and in the range \mathbb{R}^n), the matrix elements also change. They

must do so in such a way that for any \mathbf{x} given in some basis, application of this representative matrix yields \mathbf{y} given in another basis — but this \mathbf{y} must not depend on the choice of bases. Clearly, transformations of representative matrices must obey certain rules, but not necessarily those for transformations of vectors. The operators corresponding to such matrices are also known as tensors of the second rank. In the theory of tensors, vectors constitute tensors of the first rank and scalars constitute tensors of zero rank. Although we can consider the correspondence M as a function acting from \mathbb{R}^n to \mathbb{R}^n , we prefer to call it a matrix operator. We know that a matrix operator is a linear transformation whose degree of continuity — that is, how a change in \mathbf{x} affects the change in \mathbf{y} — is measured by its norm $\|M\|$ (defined below). All such notions as continuity, linearity, and norm can be extended to the general case. They are mostly simple restatements of concepts from calculus or linear algebra.

So let us consider an operator A from a normed space X to a normed space Y . First we extend the ordinary notion of continuity at a point. Recall that a real-valued function f of a real variable x is said to be continuous at a point x_0 of its domain if to every positive number ε there corresponds a positive number δ (which may depend on ε) such that $|f(x) - f(x_0)| < \varepsilon$ whenever $|x - x_0| < \delta$.

Definition 1.11.1. An operator A acting from X to Y is *continuous* at $x_0 \in X$ if to every $\varepsilon > 0$ there corresponds $\delta = \delta(\varepsilon)$ such that

$$\|A(x) - A(x_0)\| < \varepsilon \text{ whenever } \|x - x_0\| < \delta.$$

The definition for an ordinary function was extended by using the norm in place of the absolute value operation. To emphasize that the spaces X and Y may have different norms, we could have written

$$\|A(x) - A(x_0)\|_Y < \varepsilon \text{ whenever } \|x - x_0\|_X < \delta$$

instead of the above. However, we shall follow the usual practice and attach subscripts to norm symbols only when the spaces involved may not be clear from the context.

Definition 1.11.2. Let $\{x_n\}$ be a sequence in X . We say that $\{x_n\}$ *converges* to x_0 and write

$$x_0 = \lim_{n \rightarrow \infty} x_n$$

if for any $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that

$$\|x_n - x_0\| < \varepsilon \text{ whenever } n > N.$$

We may also write $x_n \rightarrow x_0$ as $n \rightarrow \infty$.

Note that we have introduced the limit passage in a normed space using only the operations of classical analysis; indeed, only operations with numbers — the values of the norm — are involved.

Having the notion of sequence convergence, we may introduce

Definition 1.11.3. An operator A acting from X to Y is *sequentially continuous* at $x_0 \in X$ if $A(x_n) \rightarrow A(x_0)$ as $n \rightarrow \infty$ for any sequence $\{x_n\}$ such that $x_n \rightarrow x_0$ as $n \rightarrow \infty$.

Lemma 1.11.1. *The two types of continuity are equivalent.*

Proof. First let us show that continuity implies sequential continuity. Let A be continuous at x_0 and take any convergent sequence $\{x_n\}$ with $x_n \rightarrow x_0$. Let $\varepsilon > 0$ be given. By continuity there exists δ such that $\|A(x_n) - A(x_0)\| < \varepsilon$ whenever $\|x_n - x_0\| < \delta$. But, by convergence of $\{x_n\}$, there exists N such that this last inequality holds whenever $n > N$. Therefore, to any $\varepsilon > 0$ there corresponds N such that $\|A(x_n) - A(x_0)\| < \varepsilon$ whenever $n > N$. So A is sequentially continuous at x_0 .

Conversely, let us show that sequential continuity implies continuity. Suppose A is not continuous at x_0 . Then there exists $\varepsilon = \varepsilon_0$ with the property that for every positive integer n there is a point x_n inside the ball $\|x - x_0\| < 1/n$ such that $\|A(x_n) - A(x_0)\| \geq \varepsilon_0$. The sequence $\{x_n\}$ thus constructed is convergent to x_0 , but it is false that $A(x_n) \rightarrow A(x_0)$. Therefore A is not sequentially continuous at x_0 . \square

In view of the lemma, we will refer to sequential continuity as simply “continuity” and use whichever formulation is convenient.

Definition 1.11.4. We say that A is *linear* if

$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y) \quad (1.11.10)$$

for any two scalars α and β and any two elements $x, y \in X$.

For such an operator we often write Ax instead of $A(x)$. One useful observation is

Lemma 1.11.2. *If a linear operator A is continuous at $x = 0$, it is continuous on the entire space X .*

This follows immediately from the relation

$$A(x) - A(x_0) = A(x - x_0),$$

since we can think of $x - x_0$ as a vector y that becomes arbitrarily small when we make x arbitrarily close to x_0 .

Why is the notion of continuity seldom stressed for matrix operators in linear algebra? In fact the issue is trivial: any $n \times n$ matrix A having only finite elements, which represents an operator A in some basis, corresponds to a continuous operator.

Now we introduce

Definition 1.11.5. An operator A from a normed space X to a normed space Y is *bounded* if there is a positive constant c such that

$$\|Ax\| \leq c \|x\| \quad \text{for all } x \in X. \quad (1.11.11)$$

If (1.11.11) holds, then A is continuous at $x = 0$ by Definition 1.11.1 with $\delta = \varepsilon/c$ and hence is continuous on X .

For linear operators, we have

Theorem 1.11.1. *A continuous linear operator A from X to Y is bounded.*

Proof. A is continuous at $x = 0$. Take $\varepsilon = 1$; by definition there exists $\delta > 0$ such that $\|Ax\| \leq 1$ whenever $\|x\| < \delta$. For every nonzero $x \in X$, the norm of $x^* = \delta x / (2 \|x\|)$ is

$$\|x^*\| = \|\delta x / (2 \|x\|)\| = \delta/2 < \delta,$$

so $\|Ax^*\| \leq 1$. By linearity of A , this gives us

$$\|Ax\| \leq \frac{2}{\delta} \|x\|,$$

which is (1.11.11) with $c = 2/\delta$. □

Definition 1.11.6. The infimum of the set of constants c for which (1.11.11) holds is called the *norm* of A .

Alternatively the number $\|A\|$ is a norm if, for any $x \in X$, we have

$$\|Ax\| \leq \|A\| \|x\|, \quad (1.11.12)$$

and, for any $\varepsilon > 0$, we can find x_ε such that

$$\|Ax_\varepsilon\| > (\|A\| - \varepsilon) \|x_\varepsilon\|. \quad (1.11.13)$$

Clearly, when we use a term like operator “norm” we should prove that it really has all norm properties and is suitably defined on some linear space.

Here the space is the set of all continuous linear operators, with operations of addition and scalar multiplication patterned after those for matrices:

$$(A + B)x = Ax + Bx, \quad (\alpha A)x = \alpha(Ax).$$

The reader should prove that $\|A\|$ satisfies N1–N3. The set of all continuous linear operators from a normed space X to a normed space Y is denoted $L(X, Y)$. If $Y = X$, the notation is $L(X)$.

Exercise 1.11.3. *Prove that the norm of an operator A can be defined as*

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \quad \text{or} \quad \|A\| = \sup_{\|x\|\leq 1} \|Ax\|. \quad (1.11.14)$$

Although other norms can be placed on $L(X, Y)$, we will typically use the above norm which is related to the norms on X and Y .

1.12 Matrix Operators and Matrix Equations

Many continuum mechanics problems cannot be solved analytically. Even when analytic solution is possible, engineers often prefer approximate numerical simulations that yield instructive pictures of system behavior. For example, the motion of a particle system can be studied by applying approximate methods to the Cauchy problem for the corresponding ordinary differential equations; the problem is thereby reduced to a discrete one, to be integrated in finite time steps using finite difference approximations of the derivatives. For a Cauchy problem this is done successively beginning with the initial point, whereas for a boundary value problem we must satisfy the conditions at the final point of the system trajectory. The latter leads to a system of equations which, in the case of a general particle system with forces of attraction, etc., are transcendental as a rule. Treatment is seldom straightforward.

In continuum mechanics many engineering problems, like those of the theory of elasticity, are described by linear equations. Available solution methods include the finite element method, the boundary element method, the finite difference method, etc. All lead to systems of simultaneous linear equations that can be written as

$$a_{ij}x_j = f_i \quad (1.12.1)$$

or in matrix form as

$$A\mathbf{x} = \mathbf{f}. \quad (1.12.2)$$

In this way, a matrix A approximates the original operator of the boundary value problem. Formally, (1.12.2) looks like an equation for vector quantities \mathbf{f} and \mathbf{x} in a space \mathbb{R}^n of higher dimension. But \mathbf{x} and \mathbf{f} are not real vectors, since their various components refer to different points of the body and we cannot work with them (for example, transforming between coordinate systems) as we do with vectors if we wish to properly preserve their physical meanings. Still, as soon as we write down the matrix equation it begins to live a life of its own and we can treat it using customary methods. For a time, we can virtually forget how the matrix equation originated — until, of course, we reach the point where we must interpret the final results in light of the original model.

Let us calculate a few matrix norms. Suppose, in analogy with (1.5.4), we define a norm on \mathbb{R}^n using

$$\|\mathbf{x}\| = \max_{1 \leq i \leq n} |x_i| \quad (1.12.3)$$

for $\mathbf{x} = (x_1, \dots, x_n)$. Note the norm is written for a fixed basis of \mathbb{R}^n that may be non-orthogonal. Writing $A = (a_{ij})$, which represents the operator \mathbf{A} so that the i th component of the image $\mathbf{y} = \mathbf{A}\mathbf{x}$ in the same basis is $\sum_{j=1}^n a_{ij}x_j$, we have

$$\begin{aligned} \|\mathbf{A}\mathbf{x}\| &= \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij}x_j \right| \\ &\leq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}x_j| \\ &\leq \left(\max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \right) \max_{1 \leq j \leq n} |x_j|, \end{aligned}$$

hence

$$\|\mathbf{A}\mathbf{x}\| \leq \mu \|\mathbf{x}\| \quad (1.12.4)$$

where

$$\mu = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

In linear algebra, μ is called the “max row sum” of the matrix A . Note that $\mu \geq \|A\|$ by definition of $\|A\|$. But we can go further and show that $\mu = \|A\|$. To see that equality holds in (1.12.4) for some nonzero \mathbf{x} , let k

be the value of i for which the max row sum occurs:

$$\mu = \sum_{j=1}^n |a_{kj}|.$$

Then choose the components of \mathbf{x} according to the rule

$$x_i = \begin{cases} +1, & a_{ki} \geq 0, \\ -1, & a_{ki} < 0. \end{cases}$$

For this \mathbf{x} , we have $\|\mathbf{x}\| = 1$ and $\|\mathbf{Ax}\| = \mu$, so equality holds.

Exercise 1.12.1. Show that the quantity $\|\mathbf{x}\|$ specified in (1.12.3) satisfies norm axioms N1–N3. Repeat for

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \tag{1.12.5}$$

where $1 \leq p < \infty$. The latter norm, of course, induces a metric given by $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p$. A special case of this metric was encountered in (1.3.2).

For the norm (1.12.5), we get

$$\|\mathbf{Ax}\|_p = \left(\sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}x_j \right|^p \right)^{1/p} \leq \left[\sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}x_j| \right)^p \right]^{1/p}.$$

Next we apply Hölder’s inequality for sums. If $p > 1$, $q > 1$, and

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then for any two sets of real numbers a_1, \dots, a_m and b_1, \dots, b_m , we have

$$\sum_{i=1}^m |a_i b_i| \leq \left(\sum_{i=1}^m |a_i|^p \right)^{1/p} \left(\sum_{i=1}^m |b_i|^q \right)^{1/q}. \tag{1.12.6}$$

Hence

$$\left(\sum_{j=1}^n |a_{ij}x_j| \right)^p \leq \left(\sum_{j=1}^n |a_{ij}|^q \right)^{p/q} \left(\sum_{j=1}^n |x_j|^p \right)$$

where $q = p/(p - 1)$, so

$$\|\mathbf{Ax}\|_p \leq \left[\sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^q \right)^{p/q} \right]^{1/p} \|\mathbf{x}\|_p.$$

By this and the well-known conditions for equality in Hölder's inequality, we conclude that

$$\|A\| = \left[\sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^q \right)^{p/q} \right]^{1/p}.$$

If $p = 2$, then $q = 2$ and the formulas above reduce to

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}, \quad \|A\| = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}.$$

The norm $\|\mathbf{x}\|_2$ is induced by an inner product: viz.,

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i.$$

In contrast, each of the norms $\|\mathbf{x}\|_p$ for $p \neq 2$ cannot be induced by an inner product.

It is worth emphasizing that the operator norm depends on the underlying norm imposed on \mathbb{R}^n .

Exercise 1.12.2. Consider a matrix operator A acting between different normed spaces. Both spaces consist of n -tuples as above, but have respective norms given by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad \|\mathbf{y}\|_r = \left(\sum_{i=1}^n |y_i|^r \right)^{1/r}.$$

Assuming $\mathbf{y} = A\mathbf{x}$, estimate $\|A\|$.

We mentioned equivalent norms for the spaces of differentiable functions and \mathbb{R}^n . Let us introduce a strict definition.

Definition 1.12.1. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$, imposed on the same set of vectors X , are *equivalent* if there exist positive constants c and C such that the inequality

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1 \tag{1.12.7}$$

holds for all $x \in X$.

Clearly, when a sequence $\{x_n\}$ converges to x_0 in one norm then it also converges to x_0 in any equivalent norm.

Exercise 1.12.3. Show that any two norms in \mathbb{R}^n are equivalent.

Using some basis of a finite dimensional space, we can introduce a one-to-one correspondence between it and \mathbb{R}^n or \mathbb{C}^n that preserves algebraic operations and the norms. From this (and the equivalence of all norms on \mathbb{R}^n or \mathbb{C}^n) it follows that on any finite-dimensional space all norms are equivalent. This is false for infinite-dimensional spaces. Moreover, *the equivalence of all norms on a linear space means the space is finite-dimensional.*

As we have said, numerical approaches to the linear problems of continuum mechanics effectively reduce these problems to matrix equations. Nonlinear problems also reduce, as a rule, to the solution of linear matrix equations that arise as intermediate problems. So an ability to solve matrix equations is essential. The relevant methods fall under the heading of numerical linear algebra and lie outside the scope of the present book. The reader can see any textbook on numerical analysis for a full discussion.

1.13 Complete Spaces

We expect a numerical approach to yield an approximation to a true solution. This approximation could be good or bad, however, and the best one can hope for is a reliable estimate of the error. Often we are merely assured that a method can, in principle, yield a sequence of approximations convergent to the true solution. Even so, it is evident that the numerical implementation of such a method may not converge to the true solution: roundoff error alone can prevent this. It can destroy a solution to a simultaneous system of equations when the dimension reaches a certain size.

Practitioners commonly judge the convergence of an approximation sequence by comparing successive terms of the sequence. When the difference seems “small enough” for the purpose at hand, computation is halted. So the analyst simply watches the successive differences between approximations and waits until some stopping criterion has been satisfied. For problems involving matrix equations, these differences are typically gauged using one of the norms on \mathbb{R}^n ; either absolute errors or relative errors (obtained by dividing the difference by the norm of one of the solutions) can be used. If the calculations were perfect, without roundoff or truncation error, the analyst could continue the process indefinitely. The best he or she could hope for, however, is to observe the pattern typical of what we call a “Cauchy sequence” in calculus. This leads us to reframe the concept in the more general metric space setting.

Definition 1.13.1. Let $\{x_n\}$ be a sequence of points in a metric space (S, d) . We say that $\{x_n\}$ is a *Cauchy sequence* if to each $\varepsilon > 0$, there corresponds a number $N = N(\varepsilon)$ such that $d(x_n, x_m) < \varepsilon$ whenever $m > N$ and $n > N$.

This definition practically coincides with the usual one given in calculus. We know that in \mathbb{R} or \mathbb{R}^n the concepts of “Cauchy sequence” and “convergent sequence” are essentially equivalent. Is this true in a general metric space? Let us explicitly generalize the idea of convergence.

Definition 1.13.2. Let $\{x_n\}$ be a sequence of points in a metric space (S, d) . We say that $\{x_n\}$ is *convergent* if there is a point $x \in S$ having the property that, to each $\varepsilon > 0$, there corresponds a number $N = N(\varepsilon)$ such that $d(x_n, x) < \varepsilon$ whenever $n > N$. In this case the point x is called the *limit* of $\{x_n\}$, and we write

$$\lim_{n \rightarrow \infty} x_n = x$$

or $x_n \rightarrow x$ as $n \rightarrow \infty$.

Observe that $x_n \rightarrow x$ if and only if $d(x_n, x) \rightarrow 0$, by definition of the ordinary limit in \mathbb{R} . The reader should also be aware that we sometimes write

$$\lim_{m, n \rightarrow \infty} d(x_n, x_m) = 0, \quad \text{or} \quad d(x_n, x_m) \rightarrow 0 \text{ as } m, n \rightarrow \infty,$$

if $\{x_n\}$ is a Cauchy sequence.

Again, in the Euclidean space \mathbb{R}^n every Cauchy sequence is convergent and vice versa. It is clear that even in a general metric space, every convergent sequence is a Cauchy sequence; we formulate this as

Exercise 1.13.1. Suppose $x_n \rightarrow x$ as $n \rightarrow \infty$ in a general metric space (S, d) . Show that $\{x_n\}$ is a Cauchy sequence in (S, d) .

What about the converse: are Cauchy sequences always convergent? We consider a couple of examples. Recall that $C(a, b)$ stands for the space of continuous functions defined on the closed interval $[a, b]$ with the “max norm”

$$\|f\| = \max_{t \in [a, b]} |f(t)|. \quad (1.13.1)$$

Here a sequence of functions $f_n = f_n(t)$ is a Cauchy sequence if

$$\max_{t \in [a, b]} |f_n(t) - f_m(t)| \rightarrow 0 \quad \text{as } m, n \rightarrow \infty.$$

For any $t \in [a, b]$, the sequence $\{f_n(t)\}$ is a numerical Cauchy sequence. Hence it has a unique limit that we denote as $f(t)$. This $f(t)$ is a function on $[a, b]$. Clearly $\{f_n(t)\}$ converges to $f(t)$ in the sense that

$$\sup_{t \in [a, b]} |f_n(t) - f(t)| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

But we do not know whether $f(t)$ is continuous on $[a, b]$. We now refer to a theorem from classical analysis:

Theorem 1.13.1 (Weierstrass). *Suppose a sequence of continuous functions $\{f_n(t)\}$, defined on a closed and bounded interval $[a, b]$, is uniformly convergent to a limit function $f(t)$. Then $f(t)$ is also continuous on $[a, b]$.*

Because uniform convergence is precisely equivalent to convergence in the max metric of $C(a, b)$, every Cauchy sequence taken from $C(a, b)$ is convergent to an element of the space.

But now consider the linear space of functions continuous on $[-1, 1]$ with the “ L^1 -norm”

$$\|f\|_1 = \int_{-1}^1 |f(t)| dt. \tag{1.13.2}$$

Exercise 1.13.2. *Show that (1.13.2) satisfies N1–N3. The reason for the “ L_1 ” designation and the subscript “1” on the norm symbol will become clear in § 1.15.*

In this new normed space we consider a sequence $\{f_n(t)\}$ given by

$$f_n(t) = \begin{cases} 0, & -1 \leq t < 0, \\ nt, & 0 \leq t \leq 1/n, \\ 1, & 1/n < t \leq 1, \end{cases} \quad (n = 1, 2, 3, \dots).$$

If $m > n$, then

$$\begin{aligned} d(f_m, f_n) &= \int_0^{1/m} |mt - nt| dt + \int_{1/m}^{1/n} |1 - nt| dt \\ &= \frac{1}{2} \left(\frac{1}{n} - \frac{1}{m} \right) \rightarrow 0 \quad \text{as } m, n \rightarrow \infty, \end{aligned}$$

so $\{f_n(t)\}$ is a Cauchy sequence. But $f_n(t) \rightarrow U(t)$, where $U(t)$ is the Heaviside unit step function defined by

$$U(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Indeed,

$$d(f_n, U) = \int_0^{1/n} |nt - 1| dt = \frac{1}{2n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

However, the limit function $U(t)$ is *not* continuous on $(-1, 1)$. We have established

Lemma 1.13.1. *Let S be the space of functions continuous on $[-1, 1]$ with the L^1 -norm. There is a Cauchy sequence in S whose limit lies outside S .*

To establish a theorem we must provide a full proof. On the other hand, a proposition can be invalidated through just one counterexample. We have shown that, in general, not all Cauchy sequences converge. What shall we do then? It makes sense to select the class of spaces where the desired equivalence does hold.

Definition 1.13.3. A metric space S is *complete* if every Cauchy sequence taken from S converges to a limit in S . If S is not complete, it is *incomplete*. A normed space that is complete in its natural metric is a *Banach space*. An inner product space complete in its natural norm (i.e., complete in the metric induced by that norm) is a *Hilbert space*.

We pause to note that the idea of metric space completeness was known well before Stefan Banach (1892–1945). However, Banach was the first to perceive the true usefulness of complete normed spaces. Banach was educated as an engineer — he even published a book on classical mechanics. This background helped him understand the importance of the spaces that now bear his name. Unlike Banach, Hilbert was a pure mathematician. For many years he was considered the best in the world. His ideas set the stage for much of 20th century mathematics.

Whenever we encounter a new space, we should verify whether it is complete. Our mere introduction of the “completeness” notion does not mean such verification will be easy. We cannot, for example, immediately generalize our conclusion regarding $C(a, b)$ to the space $C(\Omega)$ where Ω is a compact subset of \mathbb{R}^n — at least not without being aware that Weierstrass’s theorem generalizes appropriately to the multivariable case. The same holds for a generalization to the space of continuous functions on Ω with metric

$$\|f - g\|_1 = \int_{\Omega} |f(\mathbf{x}) - g(\mathbf{x})| dx_1 \cdots dx_n. \quad (1.13.3)$$

In this case, however, incompleteness can indeed be shown.

We should make another point. We examined two metric spaces above, both constructed using the same base set S (i.e., the set of functions continuous on $[a, b]$). The metrics were different, however, and this allowed us to find a sequence $\{f_n(t)\}$ that happens to be a Cauchy sequence in one space but not in the other. In this sense there is a lack of equivalence between the norms on these two spaces.

It is evident that equivalent norms provide the same convergence properties. Recalling that all norms on \mathbb{R}^n are equivalent, we could ask what makes our present examples non-equivalent. The answer is that, in contrast to \mathbb{R}^n , these spaces are infinite-dimensional.

Exercise 1.13.3. *Prove that the norms (1.13.1) and (1.13.2) on the set of continuous functions are not equivalent. (Hint: construct a sequence of elements that all have unit norm under (1.13.1), but whose norms under (1.13.2) tend to zero.)*

Of course, one might suggest that we simply avoid the L^1 norm. But norms of integral type are important. Is there a better way to circumvent the difficulties associated with the incompleteness of such spaces? It turns out that there is a powerful theorem which will permit us to “extend” an incomplete space to a resulting complete space, the latter containing (at least essentially — we shall clarify this below) the elements of the original incomplete space. This construction is not unfamiliar, since we tacitly make use of it when dealing with the real number system. We take up the full details in the next section.

1.14 Completion Theorem

Although irrational numbers such as π and $\sqrt{2}$ are truly numbers, we do not specify their values merely by giving them symbols. We can, however, approximate them to any desired accuracy. Indeed we can find a Cauchy sequence whose limit is irrational, but the best we can do to state the actual limit is to assign it a name (such as “ π ” or “ $\sqrt{2}$ ”). So, from this viewpoint, an irrational number is *defined* through an approximation sequence. But the choice of sequence is obviously non-unique; many different sequences can define the same irrational number. One way around this difficulty is to introduce *equivalent Cauchy sequences*. In this approach, any two Cauchy sequences approaching the same limit are said to be equivalent; we can collect all these sequences into equivalence classes and identify each

irrational number with one of the classes. Each rational number can be identified with an equivalence class as well. This idea lies at the base of the metric space completion theorem.

Before stating the theorem we introduce some terminology. Some of these concepts have been mentioned above, but we pause to formalize them.

Definition 1.14.1. Two sequences $\{x_n\}, \{y_n\}$ in a metric space (S, d) are said to be *equivalent* if $d(x_n, y_n) \rightarrow 0$ as $n \rightarrow \infty$. Given any Cauchy sequence $\{x_n\}$ in S , we can gather into an equivalence class X all Cauchy sequences in S that are equivalent to $\{x_n\}$. We then refer to any Cauchy sequence from X as a *representative* of X . Note that to any $x \in S$ there corresponds a *stationary* equivalence class containing the “stationary” Cauchy sequence x, x, x, \dots

We think in terms of metric spaces primarily in those situations (e.g., the study of convergence) where the distance between elements is crucial. If a one-to-one *distance preserving* correspondence exists between metric spaces, we can work with the elements of either space. This is true even if the spaces have elements of distinctly different natures — we can work with the elements of one space and understand that, since distances are of main concern, all statements we make relative to that space also hold for the corresponding elements of the other space. With this in mind we state

Definition 1.14.2. A mapping from one metric space to another is an *isometry* if it preserves distances; that is, f is an isometry from a metric space (S_1, d_1) to a metric space (S_2, d_2) if

$$d_2(f(x), f(y)) = d_1(x, y) \quad (1.14.1)$$

for every pair of points x, y taken from S_1 . If an isometry is also a one-to-one correspondence, it is a *one-to-one isometry* and the two metric spaces involved are *isometric*.

We know that a given real number can be approximated to any desired accuracy by a rational number. In short, “the rationals are dense in the reals.” The notion of denseness can be extended to more general sets.

Definition 1.14.3. Let X and Y be two subsets of a metric space (S, d) . We say that X is *dense in* Y if for any point $y \in Y$ and any $\varepsilon > 0$, we can find a point $x \in X$ such that $d(x, y) < \varepsilon$.

Now we can talk about metric space completion. The main result is

Theorem 1.14.1. *Let S be a metric space. There is a one-to-one isometry between S and a set \tilde{S} which is dense in a complete metric space S^* .*

This is the *completion theorem*. It guarantees that any metric space can be completed. Of course, if S is already complete then there is nothing to prove, so the result is interesting only when S is incomplete. The proof is rather long and we subdivide it into digestible portions.

We begin by introducing the elements of S^* , using the idea of approximating irrational numbers with classes of equivalent Cauchy sequences of rational numbers. Given any particular Cauchy sequence $\{x_n\}$ in the original space S , we form the equivalence class X mentioned in Definition 1.14.1. We now view X as a single element of the space S^* and construct the remaining elements similarly. So S^* consists of equivalence classes of Cauchy sequences taken from the original space S .

Definition 1.14.4. We call S^* the *completion* of S .

Of course, we will have to define a suitable metric on S^* and then *show* that S^* is complete in this metric.

Now we define \tilde{S} . Corresponding to any $x \in S$, there is a stationary Cauchy sequence x, x, x, \dots . This sequence would, by our procedure above, generate an equivalence class to be included in the completion space S^* . The subset of S^* consisting of only the stationary equivalence classes is denoted by \tilde{S} . There is clearly a one-to-one correspondence between S and \tilde{S} , and we will ultimately show that \tilde{S} is dense in S^* . We begin by introducing a suitable metric on S^* (and \tilde{S}).

Lemma 1.14.1. *Let $X, Y \in S^*$. The function $D(X, Y)$ given by*

$$D(X, Y) = \lim_{n \rightarrow \infty} d(x_n, y_n), \quad (1.14.2)$$

where $\{x_n\}$ and $\{y_n\}$ are arbitrary representatives of X and Y , respectively, is a metric on S^ .*

Proof. We first show that the proposed metric is well-defined; i.e., that the indicated limit exists and is independent of the choice of representative sequences. The triangle inequality for the metric d (on S) allows us to write

$$d(x_n, y_n) \leq d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n)$$

so that

$$d(x_n, y_n) - d(x_m, y_m) \leq d(x_n, x_m) + d(y_m, y_n).$$

Interchanging m and n and using the symmetry of the metric, we obtain

$$-[d(x_n, y_n) - d(x_m, y_m)] \leq d(x_n, x_m) + d(y_m, y_n).$$

Therefore

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m). \quad (1.14.3)$$

Because $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences, we know that $d(x_n, x_m) \rightarrow 0$ and $d(y_n, y_m) \rightarrow 0$ as $m, n \rightarrow \infty$. It follows from (1.14.3) that $\{d(x_n, y_n)\}$ is a Cauchy sequence of real numbers. So the limit in (1.14.2) exists by completeness of \mathbb{R} . To see that it does not depend on the choice of representatives, we take any two other representatives $\{x'_n\}$ and $\{y'_n\}$ from X and Y , respectively, and use the inequality

$$|d(x_n, y_n) - d(x'_n, y'_n)| \leq d(x_n, x'_n) + d(y_n, y'_n)$$

to get

$$|d(x_n, y_n) - d(x'_n, y'_n)| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This shows that

$$\lim_{n \rightarrow \infty} d(x'_n, y'_n) = \lim_{n \rightarrow \infty} d(x_n, y_n).$$

So $D(X, Y)$ is well-defined. Does it really satisfy the axioms of a metric? First, the inequality $D(X, Y) \geq 0$ follows from passage to the limit as $n \rightarrow \infty$ in the corresponding inequality $d(x_n, y_n) \geq 0$ that is satisfied by d for each n . If $X = Y$ then we certainly have $D(X, Y) = 0$ (since we can choose the same representative sequence from both X and Y). Conversely, the statement $D(X, Y) = 0$ implies that any two representatives $\{x_n\}$ and $\{y_n\}$ give $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$. By Definition 1.14.1 these representatives are equivalent and we conclude that $X = Y$. So metric axiom M1 is satisfied. Satisfaction of M2 follows from the definition of D and the symmetry of d :

$$D(X, Y) = \lim_{n \rightarrow \infty} d(x_n, y_n) = \lim_{n \rightarrow \infty} d(y_n, x_n) = D(Y, X).$$

Finally, the triangle inequality

$$D(X, Y) \leq D(X, Z) + D(Z, Y)$$

follows from passage to the limit as $n \rightarrow \infty$ in the inequality

$$d(x_n, y_n) \leq d(x_n, z_n) + d(z_n, y_n).$$

So D is a suitable metric on S^* . Since \tilde{S} is a subset of S^* , we can also employ D as the metric on \tilde{S} . □

Lemma 1.14.2. *The space (S^*, D) is complete.*

Proof. We will show that an arbitrary Cauchy sequence $\{X^i\}$ in S^* is convergent. From each X^i we take a representative $\{x_j^{(i)}\}$ and, from this, an element x_{k_i} such that $k_i > k_{i-1}$ and $d(x_{k_i}, x_j^{(i)}) < 1/i$ for all $j > k_i$. To see that $\{x_{k_i}\}$ is a Cauchy sequence in S , we denote by X_{k_i} the equivalence class containing the stationary sequence $(x_{k_i}, x_{k_i}, \dots)$ and write

$$\begin{aligned} d(x_{k_i}, x_{k_j}) &= D(X_{k_i}, X_{k_j}) \\ &\leq D(X_{k_i}, X^{k_i}) + D(X^{k_i}, X^{k_j}) + D(X^{k_j}, X_{k_j}) \\ &\leq \frac{1}{i} + D(X^{k_i}, X^{k_j}) + \frac{1}{j} \\ &\rightarrow 0 \quad \text{as } k_i, k_j \rightarrow \infty. \end{aligned}$$

Denoting by X the class determined by $\{x_{k_i}\}$, we have

$$\begin{aligned} D(X^{k_i}, X) &\leq D(X^{k_i}, X_{k_i}) + D(X_{k_i}, X) \\ &\leq \frac{1}{i} + D(X_{k_i}, X) \\ &= \frac{1}{i} + \lim_{j \rightarrow \infty} d(x_{k_i}, x_{k_j}) \\ &\rightarrow 0 \quad \text{as } i \rightarrow \infty. \end{aligned}$$

So $X^{k_i} \rightarrow X$ in the metric of S^* and as $\{X^i\}$ is a Cauchy sequence then $X^i \rightarrow X$ as $i \rightarrow \infty$. □

Lemma 1.14.3. *The set \tilde{S} is dense in the set S^* , relative to the metric D .*

Proof. Let $X \in S^*$ be given. We select a representative $\{x_n\}$ from X , and for each n denote by X_n the stationary equivalence class containing (x_n, x_n, \dots) . Because

$$D(X_n, X) = \lim_{m \rightarrow \infty} d(x_n, x_m) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

we can approximate X as closely as desired by the elements X_n taken from \tilde{S} . □

Lemma 1.14.4. *The spaces (S, d) and (\tilde{S}, D) are isometric.*

Proof. The one-to-one correspondence between S and \tilde{S} is defined by pairing with any $x \in S$ the element $X \in \tilde{S}$ that contains (x, x, \dots) . Given

any $x, y \in S$, we can take their images X and Y under the correspondence and write

$$D(X, Y) = \lim_{n \rightarrow \infty} d(x, y) = d(x, y)$$

to see that it preserves distances. \square

Theorem 1.14.1 follows from Lemmas 1.14.1–1.14.4. Since it is formulated for a general metric space, it holds for all particular cases. If a space has additional properties, these typically transfer to the completion space as well.

The most important property of this kind is linearity. Suppose S is a linear metric space with the additional operations of summation $x + y$ and multiplication λx by a (real or complex) number. It is clear that the same operations can be introduced in S^* for the elements X, Y and, moreover, that the above correspondence between S and \tilde{S} preserves these algebraic operations. Hence the completion is also a vector space.

Particularly important are the normed spaces. Each is a vector space and, in addition, has a natural metric $d(x, y) = \|x - y\|$. Here we can also apply Theorem 1.14.1. Let us consider this case. Everything stated for metric spaces continues to hold, of course, but there are additional observations as well. The metric (1.14.2) now takes the form

$$D(X, Y) = \lim_{n \rightarrow \infty} d(x_n, y_n) = \lim_{n \rightarrow \infty} \|x_n - y_n\|.$$

This raises the question whether $D(X, Y)$ can be considered as the norm of the element $X - Y$: we denote this by $\|X - Y\|_*$. The reader should verify that $\|X\|_*$ satisfies N1–N3. Hence S^* , resulting from application of Theorem 1.14.1, is a Banach space.

Similar reasoning holds for an inner product space: Theorem 1.14.1 yields a Hilbert space S^* whose inner product between X and Y is defined by a limit passage in the inner product over representative Cauchy sequences for these elements. The result is independent of the choice of representatives and satisfies I1–I3.

In what follows, we will complete various normed and inner product spaces. Let us formulate appropriate versions of the theorem.

Theorem 1.14.2. *Let S be a normed space. There is a one-to-one isometry between S and a set \tilde{S} which is dense in a Banach space S^* . Algebraic operations between elements are preserved under this correspondence. The norm in S^* is given by*

$$\|X\|_* = \lim_{n \rightarrow \infty} \|x_n\|, \quad (1.14.4)$$

where $\{x_n\}$ is any representative Cauchy sequence taken from the class X .

Theorem 1.14.3. *Let S be an inner product space. There is a one-to-one isometry between S and a set \tilde{S} which is dense in a Hilbert space S^* . Algebraic operations between elements are preserved under this correspondence. The inner product in S^* is given by*

$$(X, Y)_* = \lim_{n \rightarrow \infty} (x_n, y_n), \quad (1.14.5)$$

where $\{x_n\}$ and $\{y_n\}$ are any representative Cauchy sequences taken from the classes X and Y , respectively.

Exercise 1.14.1. *Denote by P the set of all polynomials with real coefficients on the closed segment $[0, 1]$. Observe that P is a linear space. Supply it with the norm of $C(0, 1)$, that is, $\|p\| = \max_{x \in [0, 1]} |p(x)|$. Clearly the resulting normed space is not complete. Describe the space that results from the completion theorem in this case.*

Let us sketch a solution. Clearly, by the completion theorem, the reader will get a complete space consisting of all the classes of equivalent Cauchy sequences of polynomials. However, by the Weierstrass theorem, any continuous function in $C(0, 1)$ can be uniformly approximated by a polynomial to within any desired accuracy. This means that P is dense in $C(0, 1)$, which is a complete space. The latter means that any Cauchy sequence in $C(0, 1)$, including sequences of elements of P , has a continuous function as a limit. It is easy to show that this limit does not depend on the choice of a representative sequence from a class of the completion space. So here any class of the completion space can be identified with a continuous function from $C(0, 1)$. We can regard the completion space obtained by the theorem as another form of representation of the space $C(0, 1)$.

1.15 Lebesgue Integration and the L^p Spaces

Suppose Ω is a closed and bounded (i.e., compact), Jordan measurable subset of \mathbb{R}^n . Let S be the collection of all functions $f(\mathbf{x})$ continuous on Ω and thus absolutely integrable over Ω in the Riemann sense:

$$\int_{\Omega} |f(\mathbf{x})| \, d\Omega < \infty. \quad (1.15.1)$$

Since the integral expression above is a valid norm on S , we can consider the normed space $(S, \|\cdot\|_1)$ where

$$\|f\|_1 = \int_{\Omega} |f(\mathbf{x})| d\Omega. \quad (1.15.2)$$

Lemma 1.13.1 states that $(S, \|\cdot\|_1)$ is incomplete when Ω is a one-dimensional interval $[a, b]$. The same holds for $\Omega \subset \mathbb{R}^n$ with any finite n . We can apply Theorem 1.14.2 and extend the operation of integration to the elements of the resulting Banach space. The integral we obtain is called the *Lebesgue integral*, after H. Lebesgue who introduced it from another standpoint. Let us denote the elements of the completion by $F(\mathbf{x})$ and introduce

Definition 1.15.1. $L^1(\Omega)$ is the Banach space formed by completing the normed space $(S, \|\cdot\|_1)$. The norm on $L^1(\Omega)$ is given by

$$\|F\|_1 = \int_{\Omega} |F(\mathbf{x})| d\Omega. \quad (1.15.3)$$

For brevity, we often write $L(\Omega)$ instead of $L^1(\Omega)$.

The sense of integration on the right side of (1.15.3) is given by Theorem 1.14.2; it is therefore

$$\int_{\Omega} |F(\mathbf{x})| d\Omega = \lim_{n \rightarrow \infty} \int_{\Omega} |f_n(\mathbf{x})| d\Omega, \quad (1.15.4)$$

where $\{f_n(\mathbf{x})\}$ is a representative sequence (i.e., of continuous functions from S) taken from the class of equivalent Cauchy sequences $F(\mathbf{x})$. We will call the value

$$\int_{\Omega} |F(\mathbf{x})| d\Omega$$

the Lebesgue integral of $|F(\mathbf{x})|$.

We pause to observe that an element of $L(\Omega)$ is not an ordinary function. It is an *equivalence class* of Cauchy sequences of continuous functions. But we do employ function notation as indicated above. This is convenient if one maintains the correct interpretation of the symbolism. For example, the sum $F(\mathbf{x}) + G(\mathbf{x})$ of two elements is understood to be the equivalence class determined by a representative Cauchy sequence $\{f_n(\mathbf{x}) + g_n(\mathbf{x})\}$, where $\{f_n(\mathbf{x})\}$ and $\{g_n(\mathbf{x})\}$ determine $F(\mathbf{x})$ and $G(\mathbf{x})$, respectively.

We have introduced the integral for $|F(\mathbf{x})|$. Let us introduce the value of the integral

$$\int_{\Omega} F(\mathbf{x}) \, d\Omega$$

for elements $F(\mathbf{x}) \in L(\Omega)$ themselves. The result will also be equivalent to the Lebesgue integral of a function. Taking a representative $\{f_n(\mathbf{x})\}$ from $F(\mathbf{x})$, we use the modulus inequality

$$\left| \int_{\Omega} f(\mathbf{x}) \, d\Omega \right| \leq \int_{\Omega} |f(\mathbf{x})| \, d\Omega \quad (1.15.5)$$

to show that the numerical sequence $\{\int_{\Omega} f_n(\mathbf{x}) \, d\Omega\}$ is a Cauchy sequence:

$$\begin{aligned} \left| \int_{\Omega} f_n(\mathbf{x}) \, d\Omega - \int_{\Omega} f_m(\mathbf{x}) \, d\Omega \right| &= \left| \int_{\Omega} [f_n(\mathbf{x}) - f_m(\mathbf{x})] \, d\Omega \right| \\ &\leq \int_{\Omega} |f_n(\mathbf{x}) - f_m(\mathbf{x})| \, d\Omega \\ &= \|f_n - f_m\|_1 \\ &\rightarrow 0 \quad \text{as } m, n \rightarrow \infty. \end{aligned}$$

Definition 1.15.2. The quantity

$$\int_{\Omega} F(\mathbf{x}) \, d\Omega \equiv \lim_{n \rightarrow \infty} \int_{\Omega} f_n(\mathbf{x}) \, d\Omega \quad (1.15.6)$$

is uniquely determined by $F(\mathbf{x})$ and is called the *Lebesgue integral* of $F(\mathbf{x})$ over Ω .

Again, this is equivalent to the Lebesgue integral as presented in real function theory.

Remark 1.15.1. The classical theory of Lebesgue integration begins with the notion of Lebesgue measurability of a set in \mathbb{R}^n . This differs from Jordan measurability in the way that the elementary domains, used for defining whether a domain is measurable, contain a countable set of elementary parallelepipeds (inscribed and circumscribed) unlike the Jordan case where only finite sets of these are used. The classical theory introduces the notion of Lebesgue integral, but the “functions” involved are not simple functions (just as they are not in our approach); rather, a function here is a collection of all those that are equal *almost everywhere* (i.e., except on a set of Lebesgue measure zero). Sets of Lebesgue measure zero can be complicated, and hence the corresponding “functions” for which

Lebesgue integration is introduced can differ substantially from ordinary functions. For example, the set of all rational numbers on the segment $[0, 1]$ has Lebesgue measure zero. Thus the functions participating in the classical theory of Lebesgue integration are also some classes of equivalent functions and offer no advantages over those used in our approach. When we say that the integrals are equivalent we mean that the resulting spaces can be placed in one-to-one correspondence in such a way that the integrals for the corresponding elements are equal. The correspondence is also isometric and preserves linear operations over the elements. Moreover, in the case of a continuous function for which the Riemann integral exists, both approaches to the Lebesgue integral yield a value equal to this Riemann integral. \square

Note the following.

- (1) If we take an element of $L(\Omega)$ that contains a stationary sequence of elements of S , this integral is equal to the ordinary Riemann integral of the function of the stationary sequence. Thus it extends the notion of Riemann integral.
- (2) In $L(\Omega)$ all the functions (elements) are absolutely integrable.

$L(\Omega)$ belongs to a class of Banach spaces that are denoted by $L^p(\Omega)$, $p \geq 1$. In particular, $L(\Omega) \equiv L^1(\Omega)$. For a fixed $p \geq 1$, the space $L^p(\Omega)$ is the completion of S with respect to the metric induced by the norm

$$\|f\|_p = \left(\int_{\Omega} |f(\mathbf{x})|^p d\Omega \right)^{1/p}. \quad (1.15.7)$$

The norm of an equivalence class $F(\mathbf{x}) \in L^p(\Omega)$ is given by

$$\|F\|_p = \left(\int_{\Omega} |F(\mathbf{x})|^p d\Omega \right)^{1/p}. \quad (1.15.8)$$

Here integration is now understood in the Lebesgue sense:

$$\int_{\Omega} |F(\mathbf{x})|^p d\Omega = \lim_{n \rightarrow \infty} \int_{\Omega} |f_n(\mathbf{x})|^p d\Omega, \quad (1.15.9)$$

where $\{f_n(\mathbf{x})\}$ is any representative of $F(\mathbf{x})$.

Exercise 1.15.1. Show that this integral is well-defined; i.e., that the limit on the right exists and is unique for any given $F(\mathbf{x}) \in L^p(\Omega)$. Hence the norm (1.15.8) is well-defined.

We now state some important facts regarding the $L^p(\Omega)$ spaces. First, for compact Ω these spaces are nested in the sense that

$$L^p(\Omega) \subseteq L^r(\Omega) \quad \text{for } 1 \leq r \leq p. \tag{1.15.10}$$

Second, a sufficient condition for existence of the integral

$$\int_{\Omega} F(\mathbf{x})G(\mathbf{x}) \, d\Omega$$

is that

$$F(\mathbf{x}) \in L^p(\Omega) \text{ and } G(\mathbf{x}) \in L^q(\Omega)$$

for p and q such that

$$\frac{1}{p} + \frac{1}{q} = 1 \text{ and } p > 1.$$

In this case Hölder’s inequality for integrals

$$\left| \int_{\Omega} F(\mathbf{x})G(\mathbf{x}) \, d\Omega \right| \leq \left(\int_{\Omega} |F(\mathbf{x})|^p \, d\Omega \right)^{1/p} \left(\int_{\Omega} |G(\mathbf{x})|^q \, d\Omega \right)^{1/q} \tag{1.15.11}$$

holds, with equality if and only if $F(\mathbf{x}) = \lambda G(\mathbf{x})$ for some number λ .

Exercise 1.15.2. Use (1.15.11) to prove (1.15.10); i.e., there exists a constant $C_{p,r}$ not dependent on $F(\mathbf{x})$ such that

$$\left(\int_{\Omega} |F(\mathbf{x})|^r \, d\Omega \right)^{1/r} \leq C_{p,r} \left(\int_{\Omega} |F(\mathbf{x})|^p \, d\Omega \right)^{1/p}$$

when $1 \leq r \leq p$.

The space $L^2(\Omega)$ deserves mention since it is a Hilbert space. (The spaces $L^p(\Omega)$ for $p \neq 2$ are Banach spaces but their norms cannot be induced by suitable inner products.) If we begin with a base set S of complex functions, Theorem 1.14.3 yields a complex Hilbert space with inner product

$$(F, G) = \lim_{n \rightarrow \infty} \int_{\Omega} f_n(\mathbf{x}) \overline{g_n(\mathbf{x})} \, d\Omega = \int_{\Omega} F(\mathbf{x}) \overline{G(\mathbf{x})} \, d\Omega. \tag{1.15.12}$$

Complex conjugation can be omitted to obtain the correct expression for the inner product on the real version of $L^2(\Omega)$.

Fredholm's operator in $L^p(\Omega)$

Let us apply Hölder's inequality to find the norm of Fredholm's operator. Equations of the general form

$$U(\mathbf{x}) + \int_{\Omega} K(\mathbf{x}, \mathbf{y})U(\mathbf{y}) d\Omega = G(\mathbf{x}), \quad (1.15.13)$$

where $K(\mathbf{x}, \mathbf{y})$ and $G(\mathbf{x})$ are given and $U(\mathbf{y})$ is the unknown sought, occur commonly in mathematical physics (and continuum mechanics in particular). They are known as *Fredholm integral equations of the second kind*. We can write (1.15.13) in the slightly more abstract form

$$U(\mathbf{x}) + AU(\mathbf{x}) = G(\mathbf{x}) \quad (1.15.14)$$

where A is the linear integral operator given by

$$AU(\mathbf{x}) = \int_{\Omega} K(\mathbf{x}, \mathbf{y})U(\mathbf{y}) d\Omega. \quad (1.15.15)$$

This is Fredholm's integral operator. Depending on the application, A can be considered as acting in various spaces of functions. It may or may not be continuous, depending on the properties of its *kernel* $K(\mathbf{x}, \mathbf{y})$. Let us obtain a condition on $K(\mathbf{x}, \mathbf{y})$ sufficient to ensure that A is bounded when it acts in the space $L^p(\Omega)$ (i.e., maps elements $U \in L^p(\Omega)$ into images $AU \in L^p(\Omega)$). We have

$$\|AU\|_p = \left(\int_{\Omega} \left| \int_{\Omega} K(\mathbf{x}, \mathbf{y})U(\mathbf{y}) d\Omega \right|^p d\Omega \right)^{1/p}.$$

But by Hölder's inequality we can write

$$\left| \int_{\Omega} K(\mathbf{x}, \mathbf{y})U(\mathbf{y}) d\Omega \right|^p \leq \left(\int_{\Omega} |K(\mathbf{x}, \mathbf{y})|^q d\Omega \right)^{p/q} \left(\int_{\Omega} |U(\mathbf{y})|^p d\Omega \right)$$

where $q = p/(p-1)$, so

$$\|AU\|_p \leq \left[\int_{\Omega} \left(\int_{\Omega} |K(\mathbf{x}, \mathbf{y})|^q d\Omega \right)^{p/q} d\Omega \right]^{1/p} \|U\|_p.$$

The needed condition for continuity of A is

$$\left[\int_{\Omega} \left(\int_{\Omega} |K(\mathbf{x}, \mathbf{y})|^q d\Omega \right)^{p/q} d\Omega \right]^{1/p} < \infty.$$

In fact, it can be shown that the quantity on the left equals $\|A\|$.

The idea of Lebesgue integration can be extended to the case of non-compact domains Ω . The above-stated fact about $\|A\|$ holds even if Ω is not compact, since Hölder's inequality continues to hold in that case. The reader should note the similarity in form between the corresponding norms of the Fredholm operator and the matrix operator in § 1.12.

Vectorial versions of (1.15.13) are also important in mechanics.

1.16 Orthogonal Decomposition of Hilbert Space

The Hilbert space $L^2(\Omega)$ plays an important role in modern theoretical investigations of partial differential equations. In this book we will encounter other Hilbert spaces that relate to the energy functionals of the various objects of continuum mechanics.

A Hilbert space possesses an important functional, the inner product, whose existence we have not used so far. However, many properties that hold in \mathbb{R}^3 — for example, those relating to projections, component representations of vectors, etc. — can be extended to general Hilbert spaces. Here we consider the decomposition of a Hilbert space into a sum of mutually orthogonal subspaces. In \mathbb{R}^3 a proper subspace might be a set of vectors acting in a direction parallel to a line or a plane. Suppose U is a subspace of vectors whose line of action is parallel to a plane through the origin, and that V consists of vectors whose line of action is parallel to a line perpendicular to that plane. Then any $\mathbf{v} \in V$ has the property that $(\mathbf{v}, \mathbf{u}) = 0$ for all $\mathbf{u} \in U$. Two subspaces related in this way are said to be *mutually orthogonal*. In this case any $\mathbf{x} \in \mathbb{R}^3$ can be written uniquely as a sum

$$\mathbf{x} = \mathbf{u} + \mathbf{v}, \quad \mathbf{u} \in U, \mathbf{v} \in V. \quad (1.16.1)$$

We say that \mathbb{R}^3 has been decomposed as a *direct sum* of the subspaces U and V , and write

$$\mathbb{R}^3 = U \dot{+} V. \quad (1.16.2)$$

These ideas extend to a general Hilbert space as follows.

Definition 1.16.1. Let V be a subspace of a Hilbert space H . A vector $x \in H$ is *orthogonal to V* if $(x, v) = 0$ for every $v \in V$. Let U be another subspace of H . We say that U and V are *mutually orthogonal subspaces*, and write $U \perp V$, if every $u \in U$ is orthogonal to V . Finally, if U and V have the property that any $x \in H$ can be expressed uniquely in the form

$x = u + v$ for some $u \in U$ and $v \in V$, we say that H has an *orthogonal decomposition* as the direct sum $H = U \dot{+} V$.

The orthogonal projection of a vector $\mathbf{x} \in \mathbb{R}^3$ onto a subspace M of \mathbb{R}^3 has a few properties by which we can define the projection uniquely. One is that the difference between \mathbf{x} and its projection \mathbf{m}_0 , i.e., $\mathbf{x} - \mathbf{m}_0$, has the least length of all vectors $\mathbf{x} - \mathbf{m}$, where $\mathbf{m} \in M$. It turns out that the orthogonal projection of a vector on a subspace of a Hilbert space can be defined in the same way. Thus we begin with the question of minimization of the functional $\|x - m\|$ over $m \in M$, a closed subspace of H .

Exercise 1.16.1. *Verify that the inequality $ax^2 + bx + c \geq c$ holds for all real x if and only if $b = 0$ and $a \geq 0$. (This result will be needed in the proof of Theorem 1.16.1.)*

Theorem 1.16.1. *Suppose M is a closed subspace of a Hilbert space H and $x \in H$. There is a unique element $m_0 \in M$ such that $\|x - m\|$ is minimized when $m = m_0$. Furthermore, m_0 is the unique “minimizing vector” if and only if $x - m_0$ is orthogonal to M .*

Proof. We prove this when H is a real Hilbert space. The case where $x \in M$ is trivial (simply take $m_0 = x$), so we assume $x \notin M$. Define

$$\delta = \inf_{m \in M} \|x - m\|.$$

Now the distance between any two elements $m_i, m_j \in M$ can be expressed using

$$\|m_j - m_i\|^2 = \|(m_j - x) + (x - m_i)\|^2$$

where, by the parallelogram law, the right side satisfies

$$\begin{aligned} & \| (m_j - x) + (x - m_i) \|^2 + \| (m_j - x) - (x - m_i) \|^2 \\ &= 2 \|x - m_j\|^2 + 2 \|x - m_i\|^2. \end{aligned}$$

This means that

$$\begin{aligned} \|m_j - m_i\|^2 &= 2 \|x - m_j\|^2 + 2 \|x - m_i\|^2 - 4 \left\| x - \frac{m_i + m_j}{2} \right\|^2 \\ &\leq 2 \|x - m_j\|^2 + 2 \|x - m_i\|^2 - 4\delta^2. \end{aligned} \quad (1.16.3)$$

(Here we have used the fact that $(m_i + m_j)/2$ lies in M , since M is a subspace.) By definition of δ we can take a sequence $\{m_i\}$ in M such that $\|x - m_i\| \rightarrow \delta$. Such a sequence is a Cauchy sequence by (1.16.3).

Furthermore, because H is complete $\{m_i\}$ must converge and its limit m_0 must belong to M since M is closed. By continuity of the norm we have $\|x - m_0\| = \delta$.

Equation (1.16.3) can also be used to prove uniqueness of the minimizing vector. If $m_0, \bar{m}_0 \in M$ are any two minimizing vectors, we can set $m_i = m_0$ and $m_j = \bar{m}_0$ and obtain

$$\|\bar{m}_0 - m_0\|^2 \leq 2 \|x - \bar{m}_0\|^2 + 2 \|x - m_0\|^2 - 4\delta^2 \leq 2\delta^2 + 2\delta^2 - 4\delta^2 = 0,$$

hence $\bar{m}_0 = m_0$.

We finish the proof by showing that m_0 is the unique minimizing vector if and only if $x - m_0$ is orthogonal to M . Indeed, let $m \in M$. If m_0 is a minimizing vector then for any α we have

$$\|x - m_0\|^2 \leq \|x - m_0 - \alpha m\|^2 = \|x - m_0\|^2 - 2\alpha(x - m_0, m) + \alpha^2 \|m\|^2.$$

This is a quadratic inequality; by Exercise 1.16.1 it can hold for all α if and only if the coefficient of the first power of α , namely $(x - m_0, m)$, is zero. This means that any $m \in M$ is orthogonal to $x - m_0$, and thus $x - m_0$ is orthogonal to M .

Conversely, let $(x - m_0, m) = 0$ for any $m \in M$. Denote $m_1 = m - m_0$. Then

$$\|x - m\|^2 = \|x - m_0 - m_1\|^2 = \|x - m_0\|^2 - 2(x - m_0, m_1) + \|m_1\|^2.$$

As $(x - m_0, m_1) = 0$ we get

$$\|x - m\|^2 = \|x - m_0\|^2 + \|m_1\|^2$$

and so m_0 is the needed minimizer. □

Definition 1.16.2. Given a subspace M of a Hilbert space H , the set of all $x \in H$ that are orthogonal to M is called the *orthogonal complement* of M and is denoted by M^\perp .

Exercise 1.16.2. Let M be a closed subspace of H . Show that M^\perp is a closed subspace of H . Hence M and M^\perp are orthogonal subspaces of H .

Now we can state the *orthogonal decomposition theorem*.

Theorem 1.16.2. If M is a closed subspace of a Hilbert space H , then

$$H = M \dot{+} M^\perp. \tag{1.16.4}$$

Hence any $x \in H$ has a unique representation $x = m + n$, where $m \in M$ and $n \in M^\perp$.

Proof. Let $x \in H$. According to Theorem 1.16.1 there is a unique $m_0 \in M$ such that $x - m_0$ is orthogonal to M . Writing $x = m_0 + (x - m_0)$ we see that x is decomposed uniquely into a component in M and a component in M^\perp . \square

Exercise 1.16.3. Our proof was given for a real Hilbert space. Supply the proof for a complex Hilbert space.

1.17 Work and Energy

It was not a simple task to devise a way of measuring the action performed by a person or machine. Eventually, the measure we now call *work* was introduced. It involves the notion that a force (or set of forces) acts on an object and thereby moves it. In the simplest case the force is constant and acts on a particle in the direction of its motion. Then the product of the force F and the distance s through which the force has shifted the particle is called its work:

$$\mathcal{W} = Fs. \quad (1.17.1)$$

When the force is variable and so depends on the length parameter s as some function $F = F(s)$, it is reasonable to introduce work as an integral:

$$\mathcal{W} = \int_A^B F(s) ds, \quad (1.17.2)$$

where A, B are the initial and final points of the particle trajectory.

When the direction of the force does not coincide with the direction of motion of the particle, a natural generalization is to introduce the force vector \mathbf{F} and represent a small piece of the trajectory as an elemental displacement $d\mathbf{r}$. Then the work can be written as a dot product:

$$\mathcal{W} = \int_A^B \mathbf{F}(s) \cdot d\mathbf{r}. \quad (1.17.3)$$

In fact, nothing prevents us from introducing other measures of the action of a force, but this particular one is intimately related to the quantity we call *energy*.

The notion of energy occurs in all the physical sciences. It is applied in many situations, but has no strict definition. We say that energy is conserved after observing a wide variety of processes in Nature. In mechanics, however, the law of conservation of mechanical energy follows directly from

mathematical transformations that yield energy integrals in various situations; these integrals, in turn, are found to be related to work as introduced in (1.17.3).

The notion of mechanical energy is extremely important. We use it not only as a measure of something conserved during the motions and deformations of bodies, but to characterize the differences between states of a body. Moreover, we will introduce normed spaces that employ energy-related expressions for norms and inner products.

Let us illustrate how the above mentioned work–energy relation arises in simple problems. Consider the motion of a particle, having mass m , under the action of a force \mathbf{F} . We shall see how the simplest form of the energy conservation law arises in mechanics. The equation of motion is

$$m\ddot{\mathbf{r}} = \mathbf{F}. \quad (1.17.4)$$

We represent $d\mathbf{r}$ along the trajectory as

$$d\mathbf{r} = \dot{\mathbf{r}} dt.$$

Dot multiplying (1.17.4) by this and integrating with respect to time over $[t_0, t_1]$, we get

$$\int_{t_0}^{t_1} m\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} dt = \int_{t_0}^{t_1} \mathbf{F} \cdot d\mathbf{r}.$$

On the right we have the work of the force \mathbf{F} acting during the time interval $[t_0, t_1]$. On the left, the simple transformation

$$m\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} = \frac{d}{dt} \frac{m\dot{\mathbf{r}}^2}{2} \equiv \frac{d}{dt} \frac{m\mathbf{v}^2}{2}$$

yields, after integration,

$$\left. \frac{m\mathbf{v}^2}{2} \right|_{t=t_1} - \left. \frac{m\mathbf{v}^2}{2} \right|_{t=t_0} = \int_{t_0}^{t_1} \mathbf{F} \cdot d\mathbf{r}. \quad (1.17.5)$$

It follows that if \mathbf{F} is zero or orthogonal to the trajectory at all times, the quantity $m\mathbf{v}^2/2$ stays constant. This is the *kinetic energy* of the particle. The last equation states that the change in kinetic energy during some time interval is equal to the work of the force during that same interval. This is one formulation of the law of energy conservation.

We know that total energy (i.e., the sum of potential and kinetic energy terms) is conserved. Let us consider a simple problem: the oscillations,

along a straight line, of a particle attached to a spring. *Hooke's law* relates the extension x of the spring to the applied force F :

$$F = kx. \quad (1.17.6)$$

By Newton's third law the spring exerts a force $-kx$ on the particle, and the equation of motion of a particle of mass m is

$$m\ddot{x}(t) = F_0(t) - kx(t). \quad (1.17.7)$$

The active force $F_0(t)$ is assumed given. Before repeating the transformations done above, let us introduce the potential

$$\mathcal{V} = \frac{1}{2}kx^2 \quad (1.17.8)$$

corresponding to the elastic force $-kx$. The name "potential" indicates that its derivative with respect to x gives us the force expression:

$$\frac{d\mathcal{V}}{dx} = -(-kx). \quad (1.17.9)$$

The equation of motion of the particle attached to the spring can be rewritten in the form

$$m\ddot{x}(t) + \frac{d\mathcal{V}(x)}{dx} = F_0(t). \quad (1.17.10)$$

Now we multiply through by

$$dx(t) = \dot{x}(t) dt$$

and integrate along the trajectory over the time $[t_0, t_1]$. We get

$$\int_{t_0}^{t_1} m\ddot{x}(t)\dot{x}(t) dt + \int_{t_0}^{t_1} \frac{d\mathcal{V}(x)}{dx} dx = \int_{t_0}^{t_1} F_0(t) dx(t)$$

which brings us, after integration, to

$$\left[\frac{mv^2}{2} + \mathcal{V}(x) \right] \Big|_{t=t_1} - \left[\frac{mv^2}{2} + \mathcal{V}(x) \right] \Big|_{t=t_0} = \int_{t_0}^{t_1} F_0(t) dx(t) \quad (1.17.11)$$

where

$$v(t) = \dot{x}(t).$$

The expression on the right-hand side of (1.17.11) is the work of the active force $F_0(t)$ during $[t_0, t_1]$. On the left we see the particle's kinetic energy $\mathcal{K} = mv^2/2$. We see the sum $\mathcal{K} + \mathcal{V}$ evaluated at the final and initial points of the time period under consideration. Calling \mathcal{V} the potential energy and $\mathcal{K} + \mathcal{V}$ the total energy, we come to a well-known statement of energy

conservation: the change in total energy of the “particle-spring” system during $[t_0, t_1]$ is equal to the work performed by the external force during that same period. Of course, the conservation of total energy holds when the work is zero, e.g., when $F_0 = 0$; in this case $\mathcal{K} + \mathcal{V}$ stays constant over time.

We called \mathcal{V} the potential energy. It is clearly associated with the spring and not the particle. Of course, we are free to assign names in any desired way, but should have some justification. We have said that the energy relates to the work done by forces. So why is $\mathcal{V} = kx^2/2$ called potential energy? Consider the work done by the external force while stretching the spring by an amount x . At the final position, the extension of the spring is x , and the force to maintain this extension would be $F = kx$. A naive application of the “work equals force times distance” idea would yield a value of $kx \cdot x = kx^2$ for the work done. But \mathcal{V} contains an additional factor of $1/2$. Why? The answer is that we cannot apply the force kx at once: at first we need only a small force, near zero, to produce a bit of extension. If we apply the force kx right away we will produce motion, but we suppose that there is no motion. So our external force should increase from zero to kx in such a way that at every moment we have a state of equilibrium.³ Then the total work of the external force is

$$\mathcal{W} = \int_0^x k\xi \, d\xi = \frac{kx^2}{2}. \quad (1.17.12)$$

This coincides with the value of \mathcal{V} as introduced above.

The reader is aware of elementary physics problems in which a particle moves vertically through the Earth’s gravitational field; in such cases we also consider the total energy of the particle to be the sum of its kinetic and potential energies, and the potential energy term is analogous to that found above for the mass-spring system. We also introduce a gravitational potential

$$\mathcal{V}(z) = mgy, \quad (1.17.13)$$

where y is height above the Earth’s surface. The total energy in this way

³Here we ran into a typical snag that is common in statics and thermodynamics: we essentially treat a moving system as though it were in true static equilibrium at any instant of the motion. Formally, this can be done in two ways: (1) we can consider certain masses that are involved to be zero (as we have done with the mass of the spring), or (2) we can assume extremely slow motion and consider all inertial forces to be zero (although it is not altogether clear that we can do this when we observe finite changes at the conclusion of the motion).

is

$$\mathcal{E} = \frac{mv^2}{2} + \mathcal{V}(y). \quad (1.17.14)$$

So the two problems exhibit the structure for total energy, which is conserved during motion. In this sense they are analogous. Note that in both cases the total energy is related not only to the particle but to “the sources” of external force (i.e., the spring and gravitational field, respectively). In many problems we can regard the forces as “external” in nature and thereby introduce a potential-type function. Then \mathcal{V} plays the role of the potential energy of the system under consideration, but in fact it relates to the energy of some external objects that “emanate” those forces somehow. This notion of the potential of external forces is extremely useful in Lagrangian mechanics.

1.18 Virtual Work Principle

For a system of n particles in equilibrium, the resultant force acting on the i th particle is zero:

$$\mathbf{F}_i = \mathbf{0} \quad (i = 1, \dots, n). \quad (1.18.1)$$

If the motions of the particles are unconstrained, we can denote by $\delta \mathbf{r}_i$ the (arbitrary) permissible motion of the i th particle and write all the equilibrium equations as the single equation

$$\sum_{i=1}^n \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0. \quad (1.18.2)$$

This holds for all possible $\delta \mathbf{r}_i$, and from it we can recover (1.18.1) since we can appoint each $\delta \mathbf{r}_i$ independently. Equation (1.18.2) expresses the *virtual work principle* (VWP) for the equilibrium of a system of independent particles. We see that its terms express the work of the forces \mathbf{F}_i over the displacements given by the vectors $\delta \mathbf{r}_i$. This is called *virtual work*, a name we shall soon explain.

The transformation from (1.18.1) to (1.18.2) offers no real advantages in this case. But the situation is different when constraints on the motion are present.

We have mentioned the constraints under which a system of particles becomes a rigid body. In mechanics there are also constraints of other

types: supports, conditions of impenetrability, etc. We shall touch upon some of the problems in which friction is negligible.

Let us consider the equilibrium of a particle under the influence of a constraint that does not involve friction. The constraint itself is defined by an equation. For example, a particle may be constrained to move without friction on some surface expressed in Cartesian coordinates by

$$F(x, y, z) = 0. \quad (1.18.3)$$

In vectorial form this looks like $F(\mathbf{r}) = 0$. The absence of friction means the reaction force \mathbf{R} of the surface on the moving particle is always directed along the surface normal \mathbf{n} . When the constrained particle is in equilibrium, the resultant force acting on it is zero. Let us call the remaining forces *active*, and denote their resultant by \mathbf{F} . We have

$$\mathbf{F} + \mathbf{R} = \mathbf{0}. \quad (1.18.4)$$

The reaction \mathbf{R} lies along \mathbf{n} and participates in the force balance along this direction, but has no component tangent to the surface; hence the projection of \mathbf{F} on the local tangent plane must be zero. The equation of the tangent plane at a point $\mathbf{r}_0 = (x_0, y_0, z_0)$ on the surface is

$$\frac{\partial F(\mathbf{r}_0)}{\partial x}(x - x_0) + \frac{\partial F(\mathbf{r}_0)}{\partial y}(y - y_0) + \frac{\partial F(\mathbf{r}_0)}{\partial z}(z - z_0) = 0.$$

In vector form this is $\nabla F(\mathbf{r}_0) \cdot (\mathbf{r} - \mathbf{r}_0) = 0$. Let us denote $\mathbf{r} - \mathbf{r}_0$ by $\delta\mathbf{r}$ so that

$$\nabla F(\mathbf{r}_0) \cdot \delta\mathbf{r} = 0. \quad (1.18.5)$$

We call a vector $\delta\mathbf{r}$ satisfying (1.18.5) a *virtual displacement* of a particle at the point \mathbf{r}_0 . In general the vector $\mathbf{r}_0 + \delta\mathbf{r}$ does not define a point on the surface, hence the displacement $\delta\mathbf{r}$ of the particle does not belong to the set of actual displacements. Usually $\delta\mathbf{r}$ is considered as an infinitesimal displacement of the particle, in this case it belongs to the surface tangent but (with the same success) it could be finite and so in general does not belong to the surface. This explains the curious term “virtual displacement”: it does not belong, in general, to the set of real displacements of the particle but is “proportional” to one of the real infinitesimal displacements. The set of all virtual displacements $\delta\mathbf{r}$ covers all directions tangent to the surface, so the condition that the projection of the active force \mathbf{F} onto the tangent plane is zero can be written in the form

$$\mathbf{F} \cdot \delta\mathbf{r} = 0 \quad (1.18.6)$$

as \mathbf{R} is orthogonal to $\delta\mathbf{r}$. In equilibrium of the particle on a surface, (1.18.6) must hold for all virtual displacements $\delta\mathbf{r}$. It is the VWP equation in this case. In form it coincides with the VWP equation for a free particle, but the set of virtual displacement vectors $\delta\mathbf{r}$ is now restricted; because of this restriction, we excluded from the equation the reaction force \mathbf{R} , and this offers some practical advantages.

If the particle is constrained to move only along a curve without friction, then the reaction cannot have components parallel to the tangent at each point. Here, the set of virtual displacements $\delta\mathbf{r}$ is restricted to the set of vectors parallel to the tangent line at any point. Reasoning similar to the above brings us to the equation of equilibrium, which coincides with (1.18.6) for the surface constraint.

In classical mechanics one also considers unilateral constraints. In the case of a surface, for example, a particle may be able to move on the surface or away from one side but cannot penetrate through to the other side. Here the set of virtual displacements is obviously not restricted to vectors lying in the tangent plane. When friction is absent, there are various arguments (more of the nature of axioms, really) supporting the notion that the work of the reaction force must be nonnegative:

$$\mathbf{R} \cdot \delta\mathbf{r} \geq 0. \quad (1.18.7)$$

Then (1.18.4) gives

$$\mathbf{F} \cdot \delta\mathbf{r} \leq 0, \quad (1.18.8)$$

which is regarded as the most general form of the virtual work principle for a particle whose position is restricted by a unilateral constraint. When treating a system of independent particles, we can write out the VWP equation (or inequality) for each particle and then add. The resulting equation (or inequality), by the independence of the virtual displacements for each particle, is equivalent to the complete set of equilibrium equations for the system. It does not contain the constraint reactions, however, so solution of the equations is simplified. Thus, for the case of a system of particles that can move only along certain curves or surfaces without friction (such systems are called *holonomic*) so that the constraints are expressed with equalities of the type $f(\mathbf{r}_1, \dots, \mathbf{r}_n, t) = 0$ (such constraints are called *geometric* or *holonomic*, as opposed to the *kinematic constraints* whose equations include velocities of points) the virtual work principle is

expressed by

$$\sum_{i=1}^n \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0. \quad (1.18.9)$$

Of course, in problem solving there is no need to try all possible virtual displacements in this equation. For each i , it is enough to take only the vectors that constitute a basis in the corresponding space of virtual displacements.

If there are unilateral constraints without friction, the virtual work principle is given by

$$\sum_{i=1}^n \mathbf{F}_i \cdot \delta \mathbf{r}_i \leq 0. \quad (1.18.10)$$

It turns out that the virtual work principle can be used not only in cases involving independent particles, but with rigid bodies or systems of such bodies under the actions of forces. It is only necessary to observe that the virtual motions of different points of a rigid body are not independent. When we do this, we can obtain the equilibrium conditions for a rigid body. The reader can consult any textbook on classical mechanics for further details. In the same way, we can consider the virtual work principle for dynamic problems when “inertial” forces are present. So the virtual work principle applies in dynamics as well (cf., equation (1.19.2)).

For a mechanical system without friction, it seemingly does not matter whether we use the virtual work principle or Newton’s laws to study particle motion. However, the virtual work principle has a broader range of application in classical mechanics. In general, the virtual work principle is not a direct consequence of Newton’s laws, although in many cases it is possible to demonstrate their equivalence as was done above. Experience shows that the virtual work principle can be taken as the base formulation for the laws of equilibrium (and, with use of d’Alembert’s principle, for the laws of motion) of particles and rigid bodies with constraints.

1.19 Lagrange’s Equations of the Second Kind

Recall that the degree of freedom of a system of particles is the minimal number of independent parameters needed to uniquely specify the position of the system. This is often less than the formal number of Cartesian components associated with the position vectors of the particles. If a particle moves along a surface, it is sufficient to know two coordinates of the parti-

cle's position on the surface. If it moves along a line, knowledge of a single coordinate suffices. Finally, if particles make up a rigid body, then their mutual separation distances are constant, so fewer (six) position parameters are needed to describe the motion. A reduction in the number of needed parameters in comparison with the number of Cartesian components with which we may describe the system is normally due to constraints imposed on the system.

Let a system of r particles have n degrees of freedom so that its position is uniquely defined by the coordinate parameters q_1, \dots, q_n . Because of possible ties between particles, we should suppose that the position vector for each particle depends on all the q_i , which are independent: that is,

$$\mathbf{r}_i = \mathbf{r}_i(q_1, \dots, q_n, t). \quad (1.19.1)$$

Of course, we could use Newton's laws to describe the motion of a system of particles having a degree of freedom less than the total number of position vector components. But it is more reasonable to derive the minimal number of necessary equations while avoiding the equations of constraint, etc. This system of equations in the variables q_i is composed of *Lagrange's equations of the second kind*. Our derivation will proceed under relatively simple assumptions on the constraints. More complex cases are treated in fundamental textbooks on classical mechanics.

Note that for these equations the parameters q_1, \dots, q_n become functions of time t , and thus we can introduce corresponding velocities $\dot{q}_1, \dots, \dot{q}_n$. In what follows we can consider \dot{q}_i to be independent of \dot{q}_j at any time t .

We begin by combining, for a system of r particles, the virtual work principle with d'Alembert's principle. Let us include inertial forces $-m_i\ddot{\mathbf{r}}_i$ in equation (1.18.2):

$$\sum_{i=1}^r (\mathbf{F}_i - m_i\ddot{\mathbf{r}}_i) \cdot \delta\mathbf{r}_i = 0. \quad (1.19.2)$$

Here $\delta\mathbf{r}_i$ is a virtual displacement for the i th particle. We assume the \mathbf{F}_i depend on the positions and velocities of the particles. Thus, in terms of the q_i , they are

$$\mathbf{F}_i = \mathbf{F}_i(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n, t). \quad (1.19.3)$$

We transform (1.19.2), with the virtual displacements represented as

$$\delta\mathbf{r}_k = \sum_{i=1}^n \frac{\partial\mathbf{r}_k}{\partial q_i} \delta q_i. \quad (1.19.4)$$

Note that the virtual displacements are taken at a fixed instant t . They are vectors proportional to infinitesimal vectors of admissible displacements at a fixed t , so we can find them by writing out the formal expression for the first differential while considering t as a fixed parameter. After writing everything in terms of the δq_i , we will select multipliers δq_i in (1.19.2). Using the mutual independence of the δq_i , we will then equate the coefficients of the δq_i to zero and obtain the needed equations.

The work of the active forces from (1.19.2) can be represented as

$$\sum_{i=1}^r \mathbf{F}_i \cdot \delta \mathbf{r}_i = \sum_{j=1}^n Q_j \delta q_j, \quad (1.19.5)$$

where Q_j is called the component of the generalized forces relating to the virtual “displacement” δq_j . By the above assumptions, Q_j depends on the q_1, \dots, q_n , the $\dot{q}_1, \dots, \dot{q}_n$, and t .

Now we transform the terms of (1.19.2) for the inertial forces. It turns out that these can be expressed in terms of the derivatives of the kinetic energy. We first use the relation (1.19.4):

$$\sum_{i=1}^r m_i \ddot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i = \sum_{i=1}^r m_i \ddot{\mathbf{r}}_i \cdot \sum_{j=1}^n \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j = \sum_{j=1}^n \delta q_j \left(\sum_{i=1}^r m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \right). \quad (1.19.6)$$

Recall that the overdot denotes a total time derivative d/dt , which differs from $\partial/\partial t$. Note that

$$\frac{d}{dt} \left(\dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \right) = \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} + \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right),$$

and therefore

$$\ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \frac{d}{dt} \left(\dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} \right) - \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right). \quad (1.19.7)$$

Next we use two formulas, proved at the end of this section:

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} = \frac{\partial \mathbf{r}_i}{\partial q_j}, \quad (1.19.8)$$

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial q_j} = \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right), \quad (1.19.9)$$

where q_i and \dot{q}_j are considered as independent variables. Applying these to the right side of (1.19.7) we get

$$\ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \frac{d}{dt} \left(\dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} \right) - \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial q_j} = \frac{d}{dt} \left(\frac{1}{2} \frac{\partial \dot{\mathbf{r}}_i^2}{\partial \dot{q}_j} \right) - \frac{\partial}{\partial q_j} \left(\frac{\dot{\mathbf{r}}_i^2}{2} \right).$$

Substituting this into (1.19.6), we get

$$\begin{aligned} \sum_{i=1}^r m_i \ddot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i &= \sum_{j=1}^n \delta q_j \sum_{i=1}^r \left\{ \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_i} \left(\frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) \right] - \frac{\partial}{\partial q_j} \left(\frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) \right\} \\ &= \sum_{j=1}^n \delta q_j \left[\frac{d}{dt} \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{E}}{\partial q_j} \right] \end{aligned}$$

where

$$\mathcal{E} = \sum_{i=1}^r \frac{1}{2} m_i \dot{\mathbf{r}}_i^2. \quad (1.19.10)$$

Combining this and (1.19.5) with (1.19.2), we derive

$$\sum_{j=1}^n \delta q_j \left[Q_j - \frac{d}{dt} \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_j} \right) + \frac{\partial \mathcal{E}}{\partial q_j} \right] = 0.$$

Finally, using the independence of the δq_i , we obtain

$$\frac{d}{dt} \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{E}}{\partial q_j} = Q_j \quad (j = 1, \dots, n). \quad (1.19.11)$$

These are Lagrange's equations of the second kind. They constitute a system of n ordinary differential equations with respect to the unknown functions $q_j(t)$. In general they contain terms involving $q_j(t)$, $\dot{q}_j(t)$, and $\ddot{q}_j(t)$, so the system is of order $2n$.

Before finishing this section, we demonstrate how the assumption of potentiality for the generalized forces leads us to something resembling the Euler–Lagrange equations for the problem of minimum of a functional (considered in § 1.20). Potentiality of the set of Q_j means there is a potential function $\mathcal{V} = \mathcal{V}(q_1, \dots, q_n)$ such that

$$Q_j = -\frac{\partial \mathcal{V}}{\partial q_j}. \quad (1.19.12)$$

In many cases \mathcal{V} is called the potential energy; it is related to the energy of the “source” of the external forces Q_j . Substituting into (1.19.11), we get

$$\frac{d}{dt} \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{E}}{\partial q_j} = -\frac{\partial \mathcal{V}}{\partial q_j} \quad (j = 1, \dots, n).$$

By assumption, \mathcal{V} does not depend on \dot{q}_j so $\partial \mathcal{V} / \partial \dot{q}_j = 0$. Introducing a new function

$$\mathcal{L} = \mathcal{E} - \mathcal{V} \quad (1.19.13)$$

called the *kinetic potential* or the *Lagrangian*, we transform (1.19.11) to

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{L}}{\partial q_j} = 0 \quad (j = 1, \dots, n). \quad (1.19.14)$$

These are Lagrange's equations of the second kind for the case of active forces having potential.

Although we derived the Lagrange equations under simplified conditions for the mechanical system, they can be extended to much less restrictive conditions. Moreover, they are used not only in classical mechanics: physicists use these and similar equations in a variety of areas.

Finally, let us derive (1.19.8) and (1.19.9). Writing

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j} = \frac{\partial}{\partial \dot{q}_j} \left(\sum_{k=1}^n \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t} \right)$$

and noting that \mathbf{r}_i does not depend on \dot{q}_j (and hence neither do $\partial \mathbf{r}_i / \partial q_j$ or $\partial \mathbf{r}_i / \partial t$), we immediately obtain (1.19.8). Relation (1.19.9) holds because we can interchange the order in mixed partial differentiation:

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) &= \sum_{k=1}^n \frac{\partial}{\partial q_k} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) \dot{q}_k + \frac{\partial}{\partial t} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) \\ &= \frac{\partial}{\partial q_j} \left(\sum_{k=1}^n \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t} \right) \\ &= \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_j}. \end{aligned}$$

To present Lagrange's equations from a variational standpoint, as a consequence of Hamilton's principle, we will need some basic notions from the calculus of variations.

1.20 Problem of Minimum of a Functional

Lagrange's equations are important in mechanics. We wish to show that they can be derived from a variational principle that can serve as the starting point for mechanics instead of Newton's laws. We begin with some elementary facts from the calculus of variations.

As with a function, we can consider the problem of extremum of a functional. For definiteness let us take a functional $F = F(x)$ given on a Banach space X . The definitions of such concepts as a point of local maximum or minimum, global maximum or minimum, etc., can be extended

from functions to functionals. For example, a point x is called a point of local minimum of $F(x)$ if there is a δ -neighborhood of x such that $F(x) \leq F(z)$ for all z in the δ -neighborhood, i.e., whenever $\|z - x\| < \delta$.

We will examine particular minimization problems for functionals in detail when considering the equilibrium problems for elastic models. Now we can find an equation that a local extreme point must satisfy (i.e., a *necessary condition* for its existence) if $F(x)$ is sufficiently smooth (differentiable). We shall not pause to define differentiability of a functional, but will proceed rather formally instead.

Suppose x_0 is a point of local minimum of $F(x)$. This means there is an $\varepsilon > 0$ having the property that if we take an element ty such that $\|ty\| \leq \varepsilon$, where t is a real number, then

$$F(x_0) \leq F(x_0 + ty). \quad (1.20.1)$$

Now let us fix y in addition to x_0 . The functional $F = F(x_0 + ty)$ becomes a simple function of the real variable t which, according to (1.20.1), is minimized when $t = 0$. If it is differentiable, the necessary condition of minimum is simply

$$\left. \frac{dF(x_0 + ty)}{dt} \right|_{t=0} = 0. \quad (1.20.2)$$

This must hold for any $y \in X$. It is, in fact, a necessary condition for x to be a minimum or maximum point of the functional.

We should note that the expression on the left-hand side of (1.20.2) is called the *Gâteaux derivative* of the functional $F(x)$. If $F(x)$ is a function in n variables, i.e., if $x = (x_1, x_2, \dots, x_n)$, this expression gives us the directional derivative in the direction $y = (y_1, y_2, \dots, y_n)$. In particular, when $y = (1, 0, \dots, 0)$ it yields the partial derivative $\partial F / \partial x_1$.

Now let us apply (1.20.2) to a simple functional that appears in any textbook on the calculus of variations:

$$F(y) = \int_a^b f(x, y, y') dx. \quad (1.20.3)$$

We shall consider $F(y)$ over a set of functions $y = y(x)$ satisfying the *Dirichlet boundary conditions*

$$y(a) = c_0, \quad y(b) = c_1. \quad (1.20.4)$$

Let us employ the space $C^{(2)}(a, b)$. The problem of minimum is formulated as

Problem 1.20.1. *Minimize the functional*

$$\int_a^b f(x, y, y') dx$$

over the set of functions $y(x) \in C^{(2)}(a, b)$ that satisfy the boundary conditions $y(a) = c_0$ and $y(b) = c_1$.

Suppose $y(x)$ is a solution. Satisfaction of (1.20.4) by the sum $y + t\varphi$ for any value of the parameter t requires that the function $\varphi = \varphi(x)$ vanish at the endpoints $x = a, b$. According to the scheme discussed above, we set

$$\left. \frac{d}{dt} \int_a^b f(x, y + t\varphi, y' + t\varphi') dx \right|_{t=0} = 0. \tag{1.20.5}$$

We can pass the derivative operator d/dt through the integral sign if $f(x, y, y')$ is continuously differentiable in y and y' (these being regarded as independent variables). The result, after taking the total derivative, is that the equation

$$\int_a^b [f_y(x, y, y')\varphi + f_{y'}(x, y, y')\varphi'] dx = 0 \tag{1.20.6}$$

must hold for any smooth function φ vanishing at $x = a, b$. Here partial derivatives with respect to y and y' are denoted by the subscripts. The left side of (1.20.6) is called the *first variation* of the functional and is denoted by $\delta F(y, \varphi)$, while φ is called the *variation* of y and in mechanics books is denoted δy . From (1.20.6), which must hold for arbitrary admissible $\varphi(x)$, we can derive a differential equation for $y(x)$. We first integrate by parts to obtain

$$\int_a^b \left[f_y(x, y, y') - \frac{d}{dx} f_{y'}(x, y, y') \right] \varphi dx = 0, \tag{1.20.7}$$

where the terms $f_{y'}\varphi|_a^b$ vanish because $\varphi(a) = \varphi(b) = 0$. Now we need the *Main Lemma* of the calculus of variations. We introduce

Definition 1.20.1. By $\mathcal{D}(0, l)$ we mean the set of functions infinitely differentiable on $(0, l)$ and vanishing in some neighborhood of the endpoints 0 and l (this neighborhood can differ for different functions in the set).

Lemma 1.20.1. *If $G = G(x)$ is continuous on $[0, l]$ and satisfies*

$$\int_0^l G(x)\varphi(x) dx = 0$$

for any $\varphi \in \mathcal{D}(0, l)$, then $G(x) = 0$ on $[0, l]$.

Proof. We suppose $G(x^*) \neq 0$ at some $x^* \in [0, l]$ and obtain a contradiction. By continuity there is a neighborhood $[x^* - \varepsilon, x^* + \varepsilon]$ of x^* throughout which the sign of $G(x)$ persists (either strictly positive or strictly negative). If we choose $\varphi(x)$ so that it, too, has a constant sign in this neighborhood and vanishes elsewhere, then the integral

$$\int_0^l G(x)\varphi(x) dx = \int_{x^*-\varepsilon}^{x^*+\varepsilon} G(x)\varphi(x) dx$$

must be nonzero since its integrand $G(x)\varphi(x)$ never changes sign. This is the desired contradiction; the proof will be complete if we can display a function $\varphi \in \mathcal{D}(0, l)$ satisfying the condition stated above. The “bell-shaped” function

$$\varphi(x) = \begin{cases} \exp\left(\frac{\varepsilon^2}{(x-x^*)^2 - \varepsilon^2}\right), & |x-x^*| < \varepsilon, \\ 0, & |x-x^*| \geq \varepsilon, \end{cases}$$

is one example. □

Because the set of admissible functions φ in (1.20.7) includes $\mathcal{D}(0, l)$, we can apply Lemma 1.20.1 to (1.20.7) and obtain

$$f_y - \frac{d}{dx}f_{y'} = 0. \tag{1.20.8}$$

This *Euler equation* is analogous to Fermat’s condition $f' = 0$ for an ordinary function. In general it is an ordinary differential equation of second order. Since the functional F will arise from a mechanical problem where we usually seek a unique solution, we understand why we specified two boundary conditions earlier. The derivative with respect to x in (1.20.8) is a total derivative: i.e., we have

$$f_y - f_{y'x} - f_{y'y}y' - f_{y'y'}y'' = 0. \tag{1.20.9}$$

In (1.20.8) we recognize Lagrange’s equation (1.19.14) when $f = \mathcal{L}$ and $y = q_i$.

Subsequently, we will need to pose a minimum problem for (1.20.3) without specifying a boundary condition at an endpoint. In fact, in addition to (1.20.8), from (1.20.5) there follow two boundary conditions called *natural boundary conditions*. To see this we return to (1.20.6) and integrate by parts *without* imposing (1.20.4). We get

$$\int_a^b \left[f_y(x, y, y') - \frac{d}{dx}f_{y'}(x, y, y') \right] \varphi dx + f_{y'}(x, y(x), y'(x))\varphi(x) \Big|_{x=a}^{x=b} = 0. \tag{1.20.10}$$

If we temporarily limit our consideration to those smooth functions $\varphi(x)$ that *do* satisfy $\varphi(a) = \varphi(b) = 0$, then we have

$$\int_a^b \left[f_y(x, y, y') - \frac{d}{dx} f_{y'}(x, y, y') \right] \varphi dx = 0. \quad (1.20.11)$$

By Lemma 1.20.1 we once again find that (1.20.8) must hold in (a, b) . Clearly, now relation (1.20.11) holds for *any* continuous φ . Thus, returning to (1.20.10), we find that

$$f_{y'}(x, y(x), y'(x))\varphi(x) \Big|_{x=a}^{x=b} = 0$$

for *any* smooth function $\varphi(x)$. The particular choices $\varphi(x) = x - b$ and $\varphi(x) = x - a$ yield, respectively,

$$f_{y'}|_{x=a} = 0, \quad f_{y'}|_{x=b} = 0. \quad (1.20.12)$$

These are the natural boundary conditions for (1.20.3).

So one can minimize F in the absence of a boundary condition on y at one of the endpoints a or b . At that endpoint the corresponding natural condition applies automatically. Note that, although the Euler equation is of second order, we *cannot* in general introduce two *initial* conditions for y ; we cannot prescribe $y(a)$ and $y'(a)$, for example, because the natural condition at b would still apply. The result would be too many conditions to define a solution of a second order ordinary differential equation. Hence we always have two boundary conditions, one at each endpoint.

The Euler equation (1.20.8) appears in all the minimum problems for the functional (1.20.3), regardless of the boundary conditions chosen, if the problem is correctly posed. For more general functionals it changes form. Listed below are additional functionals of interest along with their corresponding Euler equations.

1. If $y(x)$ is replaced by a vector function $\mathbf{y}(x) = (y_1(x), y_2(x), \dots, y_n(x))$, then a functional of the type

$$F(\mathbf{y}) = \int_a^b f(x, \mathbf{y}, \mathbf{y}') dx \quad (1.20.13)$$

results. The Euler equation can be written in vector form as

$$\nabla_{\mathbf{y}} f - \frac{d}{dx} \nabla_{\mathbf{y}'} f = 0 \quad (1.20.14)$$

or, alternatively, in scalar form as the system of equations

$$f_{y_i} - \frac{d}{dx} f_{y'_i} = 0 \quad (i = 1, \dots, n). \quad (1.20.15)$$

These, in fact, follow easily from (1.20.8): if we fix all but the i th component of the minimizing function, we obtain a functional of exactly the same form as before (but with respect to the function y_i). Natural boundary conditions for this functional can be stated as

$$f_{y'_i} \Big|_{x=a} = 0, \quad f_{y'_i} \Big|_{x=b} = 0 \quad (1.20.16)$$

for $i = 1, \dots, n$.

2. An extreme point of the functional

$$F_n(y) = \int_a^b f(x, y, y', \dots, y^{(n)}) dx \quad (1.20.17)$$

will satisfy the *Euler–Lagrange equation*

$$f_y - \frac{d}{dx} f_{y'} + \frac{d^2}{dx^2} f_{y''} - \dots + (-1)^n \frac{d^n}{dx^n} f_{y^{(n)}} = 0 \quad (1.20.18)$$

with $2n$ natural boundary conditions. The method of obtaining this is similar to that for the simplest functional. Namely, supposing y to be a minimizer of the functional, we find that $F_n(y + t\varphi)$, being a function of the real parameter t when φ is sufficiently smooth, takes its minimum at $t = 0$. This leads to a result analogous to (1.20.6). Subsequent integration by parts yields a result analogous to (1.20.10) but having $2n$ boundary terms outside the integrals. This equation holds, in particular, for all $\varphi \in \mathcal{D}(0, l)$. Now all the boundary terms vanish, and as a consequence of Lemma 1.20.1 we get the above Euler–Lagrange equation. Next, considering all the smooth functions φ , we will derive the natural boundary conditions for the functional.

In a similar manner we can find the necessary conditions of minimum for a functional defined on functions in many variables.

3. In the two-dimensional case where

$$F(u) = \iint_S f(x, y, u(x, y), u_x(x, y), u_y(x, y)) dx dy, \quad (1.20.19)$$

the equation

$$f_u - \left(\frac{\partial f_{u_x}}{\partial x} + \frac{\partial f_{u_y}}{\partial y} \right) = 0 \quad (1.20.20)$$

plays the role of the Euler equation. Here the subscripts u_x and u_y denote partial differentiation with respect to these quantities as independent variables. The operations $\partial/\partial x$ and $\partial/\partial y$, on the other hand, are *complete* partial derivatives where all the arguments of f (i.e., u, u_x, u_y) are regarded

as functions of x and y and the chain rule is applied. Natural boundary conditions can be stated as

$$(f_{u_x} n_x + f_{u_y} n_y) \Big|_{\Gamma} = 0, \quad (1.20.21)$$

where Γ is the boundary of S .

Note that at each point of the boundary there is exactly one natural boundary condition; this was the case for the earlier simple problem for $f(y)$. When we use minimum energy principles to set up mechanics problems, the natural conditions represent, as a rule, force conditions on the boundary.

Many other functionals can be found in textbooks along with sufficient conditions for a function to be a minimum point. But we know enough about the calculus of variations for our immediate purposes.

Exercise 1.20.1. *Derive the natural boundary conditions for $F(\mathbf{y})$.*

Exercise 1.20.2. *Derive the component-form Euler equations for a functional containing derivative terms up to $\mathbf{y}^{(n)}$. How many boundary conditions should be given?*

Example

Consider a simple equilibrium problem for a bar (a structure that will be considered in more detail later) of length l stretched by both a distributed load $t(x)$ and forces F_0 and F_1 applied to its ends (Fig. 1.3). We first consider the case of a “free” bar under these forces, which means the bar is not clamped at any point. We will encounter similar equations, and even boundary conditions, in the equilibrium problem for a string.

The linear model assumes that during deformation the external forces do not change and are applied at the same points in space. This will be the case for all linear models in this book.

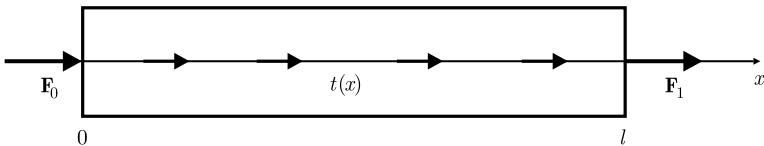


Fig. 1.3 A bar under axial loading.

When the bar is rigid and there are no geometrical constraints on its motion, the condition for equilibrium is that the resultant force must vanish:

$$\int_0^l t(x) dx + F_0 + F_1 = 0. \quad (1.20.22)$$

Let us see what happens when the bar is linearly elastic. We will frequently use the *minimum total energy principle*. For the present problem, this states that the equilibrium of the bar is reached at the point (i.e., the set of values of the displacement function $u(x)$) where the functional

$$\mathcal{E}_b(u) = \frac{1}{2} \int_0^l ES[u'(x)]^2 dx - \int_0^l t(x)u(x) dx - F_0u(0) - F_1u(l) \quad (1.20.23)$$

takes its minimum on the set of all sufficiently smooth functions $u(x)$. In deriving the functional we use Hooke's law, which relates the tension σ in a cross section of the bar with the strain $\varepsilon = u'(x)$. So the force F in the cross section due to deformation is $F = ESu'(x)$, where E is *Young's modulus* and S is the cross-sectional area.

Note that $\mathcal{E}_b(u)$ contains the non-integrated terms $F_0u(0) + F_1u(l)$. This means we cannot simply use the above formulas, but must repeat the steps that led to the necessary conditions for a minimum. We should arrive at an equilibrium equation and a set of natural boundary conditions.

So we assume a state of equilibrium described by the displacement function $u(x)$. First we consider how the *self-balance condition* (1.20.22) appears in the equilibrium problem for the bar. We fix an arbitrary $u(x)$ and consider $\mathcal{E}_b(u)$ over the set of functions $u(x) + c$ where c is an arbitrary constant representing the strain energy of the bar. The first term of $\mathcal{E}_b(u)$, the strain energy for the displacement $u(x) + c$, does not depend on c :

$$\frac{1}{2} \int_0^l ES[(u(x) + c)']^2 dx = \frac{1}{2} \int_0^l ESu'^2(x) dx.$$

Because c is arbitrary, we can get any large negative value for the quantity

$$\int_0^l t(x)[u(x) + c] dx + F_0[u(0) + c] + F_1[u(l) + c],$$

which is the work of external forces over $u(x) + c$. So the minimization problem makes sense only if the coefficient of c vanishes:

$$\int_0^l t(x) dx + F_0 + F_1 = 0.$$

Hence (1.20.22) is a necessary condition for the existence of a minimum of $\mathcal{E}_b(u)$. This equation for the elastic bar matches that for the rigid bar, confirming that the elastic model inherits the properties of the more elementary rigid model. We will see that not all elastic models (that of the membrane, for example) preserve all of the equilibrium conditions for the corresponding free rigid objects.

The requirement for external force balance can be explained in another way, if we apply the forces exactly as in Fig. 1.3; the resultant is along the x -axis. Clearly the body should move in the same direction. But we neglected the mass of the bar ($m = 0$). Consequently the bar should experience infinite acceleration, since its mass is zero but the resultant force is not. We could prevent this only by assuming the resultant force is zero. This happens for the equilibrium problems for all the free structures we will consider.

Thus we have found the force balance condition to be necessary for equilibrium. Assuming this, let us continue. According to the above theory, we implement the equation

$$\left. \frac{d\mathcal{E}_b(u + \lambda\varphi)}{d\lambda} \right|_{\lambda=0} = 0$$

for all sufficiently smooth $\varphi(x)$. Appropriate calculations yield

$$\int_0^l ESu'(x)\varphi'(x) dx - \int_0^l t(x)\varphi(x) dx - F_0\varphi(0) - F_1\varphi(l) = 0.$$

Integrating by parts in the first integral, we obtain

$$\begin{aligned} & - \int_0^l [(ESu'(x))' + t(x)] \varphi(x) dx \\ & + [ESu'(l) - F_1] \varphi(l) + [-ESu'(0) - F_0] \varphi(0) = 0. \end{aligned} \quad (1.20.24)$$

Since this holds for all smooth $\varphi(x)$, it holds in particular for those that satisfy $\varphi(0) = 0 = \varphi(l)$. For these we get

$$- \int_0^l [(ESu'(x))' + t(x)] \varphi(x) dx = 0.$$

By Lemma 1.20.1 then, we have the equilibrium equation

$$(ESu'(x))' + t(x) = 0. \quad (1.20.25)$$

This is the Euler equation for the functional. Substituting it into (1.20.24) we obtain

$$[ESu'(l) - F_1] \varphi(l) + [-ESu'(0) - F_0] \varphi(0) = 0,$$

where φ is arbitrary. By this we get two natural boundary conditions:

$$ESu'(l) = F_1, \quad ESu'(0) = -F_0. \quad (1.20.26)$$

These have a clear mechanical sense: the external forces at the endpoints equal the tension forces at those same points. The negative sign in the second condition stands in accord with the algebraic sign rule for the strength of materials. We derived it without reference to that rule (cf., § 3.2).

Again, we have derived exactly two natural boundary conditions. What if we fix the end at $x = 0$? We should repeat all the above, requiring $\varphi(x) = 0$ at $x = 0$. The derivation would yield the equilibrium (Euler) equation for $0 < x < l$ and the right-end natural condition $ESu'(l) = F_1$. There would still be precisely one condition for each endpoint, as expected, since the Euler equation is of second order.

Mechanical considerations are used to explain why an equilibrium solution for a free bar exists only for special choices of the external loads. In general, they can throw light on the physical origins of many conditions that seem to arise artificially in pure mathematics. Newton said that it is useful to solve problems, meaning the problems of real life.

Exercise 1.20.3. *Prove that if a solution of the Euler equation with natural boundary conditions exists, it minimizes the functional $\mathcal{E}_b(u)$.*

1.21 Hamilton's Principle

The pioneers of modern mechanics were sure that the universe was created in the most economical fashion and that all processes occur in an optimal manner. The existence of optimum principles was part of the ideology of Metaphysics and their manifestations were everywhere sought. While it is not our goal to discuss all the extremal principles of mechanics, we should touch on one closely related to the calculus of variations. This is *Hamilton's principle of stationary action*, which can be regarded as a simple consequence of the equations derived in § 1.20. On the other hand, it can also be regarded as the basis from which some portion of classical mechanics can be developed. It has many restrictive assumptions and the interested reader can pursue these in various textbooks.

This principle allows us to obtain the Lagrangian equations by seeking the stationary point of a functional called the *action*. The action \mathcal{W} is given

by

$$\mathcal{W} = \int_{t_1}^{t_2} \mathcal{L} dt, \quad (1.21.1)$$

where \mathcal{L} is the Lagrangian of a system of particles; here we regard the integration variable t as time and its limits as arbitrary but temporarily fixed instants. It should be emphasized that \mathcal{W} is required to have a *stationary value* (not necessarily a minimum), which means that its first variation must vanish and the Euler–Lagrange equations must hold. To exclude any additional conditions at the endpoints of the time interval, we consider only motions in which all particles in the system start and finish at the positions taken by the real particles in the actual motion. So the trajectories of admissible motions in the space of parameters q_j , given by the functions $q_1(t), \dots, q_n(t)$ on the segment $[t_0, t_1]$, must start at $(q_1(a), \dots, q_n(a))$ and finish at $(q_1(b), \dots, q_n(b))$. We also take them to be sufficiently smooth as usual.

Let us compose the Euler–Lagrange equations for the problem of minimum of the functional

$$\int_{t_0}^{t_1} \mathcal{L}(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n, t) dt$$

under the stated assumptions. By (1.20.15) we obtain

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{L}}{\partial q_j} = 0 \quad (j = 1, \dots, n),$$

which coincides with (1.19.14). We may now state

Hamilton’s stationary action principle. *Among all trajectories that start and finish along with the real trajectory, the actual trajectory yields a stationary value for the action functional \mathcal{W} .*

For an ordinary function, the fact that the derivative vanishes at some point does not mean the function takes a minimum value there. A similar statement can be made for a variational problem. A real trajectory is not necessarily a point of maximum or minimum of the functional. Hamilton’s principle shows that a real trajectory of the system is one of its extremals (i.e., satisfies (1.19.14)).

We conclude this section by mentioning the “variational principles of mechanics.” The calculus of variations deals with the minimization and maximization of functionals. The derivation of necessary conditions of

minimum leads to equations similar to (1.20.10), which contain admissible perturbations of the unknown functions. These functionals are linear with respect to the variations. The use of various versions of the Main Lemma yields differential equations, as a consequence.

In mechanics, certain equations can be obtained as Euler–Lagrange equations of functionals; however, the variational problems are only to find stationary points of a functional, not necessarily minima or maxima. Moreover, in mechanics there are integro-differential equations that resemble the equality of the first variation of a functional to zero but such that the expressions are not the first variation of any functional. Nonetheless, from such equations we can still use the Main Lemma to derive mechanically meaningful differential equations. They fall under the heading of the *variational principles of mechanics*. An example is the virtual work principle as applied to non-elastic bodies. Such “variational principles” are widely used in the generalized setup of boundary value problems, and in the construction of numerical solution methods.

1.22 Energy Conservation Revisited

The notion of energy plays a central role in science. In “physics” books whose contents are mainly mathematical, we find discussions of energy and its transformations — largely offered as illustrations of how mathematical tools can generate physically meaningful relations. In this book we take energy as a central quantity. We omit many important portions of classical mechanics and consider only those that relate to the contents of the book. Not surprisingly, the idea of energy turns out to be extremely fruitful in the mathematical analysis of mechanical problems. We therefore return to the main principle of physics: that of energy conservation. Let us examine, in general form, the equations that give rise to the notion of energy and to its all-important conservation law. The reader is asked to work the following preparatory exercise.

Exercise 1.22.1. *We say that a function $F = F(x_1, \dots, x_n)$ is homogeneous of degree r if there is a constant r such that $F(cx_1, \dots, cx_n) = c^r F(x_1, \dots, x_n)$ whenever $c > 0$. Euler’s theorem states that if F is differentiable and homogeneous of degree r , then*

$$\sum_{k=1}^n x_k \frac{\partial F}{\partial x_k} = rF.$$

Show that $g(x, y) = x^2 + 2xy + 3y^2$ is homogeneous of degree 2 and verify Euler's theorem for this function.

We would like to derive the law of energy conservation for a system of material particles having n degrees of freedom. Let us consider the relatively simple case of a system under stationary *holonomic constraints*; this means the constraint equations do not depend explicitly on time t . So we can express the position vector of a particle as a function of the variables q_1, \dots, q_n :

$$\mathbf{r}_i = \mathbf{r}_i(q_1, \dots, q_n) \quad (i = 1, \dots, n).$$

We recall that the kinetic energy \mathcal{E} of a system of r particles having masses m_i and position vectors \mathbf{r}_i is

$$\mathcal{E} = \sum_{i=1}^r \frac{1}{2} m_i \dot{\mathbf{r}}_i^2.$$

Substituting

$$\dot{\mathbf{r}}_i = \sum_{j=1}^n \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j$$

into \mathcal{E} , we find that \mathcal{E} is a quadratic form with respect to the variables \dot{q}_j and having coefficients that depend only on the variables q_k . That is,

$$\mathcal{E} = \frac{1}{2} \sum_{i,j=1}^n a_{ij} \dot{q}_i \dot{q}_j$$

where $a_{ij} = a_{ij}(q_1, \dots, q_n)$. The time derivative of \mathcal{E} is

$$\begin{aligned} \frac{d\mathcal{E}}{dt} &= \sum_{i=1}^n \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_i} \ddot{q}_i + \frac{\partial \mathcal{E}}{\partial q_i} \dot{q}_i \right) \\ &= \frac{d}{dt} \left(\sum_{i=1}^n \frac{\partial \mathcal{E}}{\partial \dot{q}_i} \dot{q}_i \right) - \sum_{i=1}^n \left(\frac{d}{dt} \frac{\partial \mathcal{E}}{\partial \dot{q}_i} - \frac{\partial \mathcal{E}}{\partial q_i} \right). \end{aligned}$$

Since \mathcal{E} is homogeneous of degree two with respect to the \dot{q}_i , Euler's theorem yields

$$\frac{\partial \mathcal{E}}{\partial \dot{q}_i} \dot{q}_i = 2\mathcal{E}.$$

Substituting this, and using the Lagrange equations (1.19.11) for the second term

$$\frac{d}{dt} \left(\frac{\partial \mathcal{E}}{\partial \dot{q}_j} \right) - \frac{\partial \mathcal{E}}{\partial q_j} = -\frac{\partial \mathcal{V}}{\partial q_j},$$

we get

$$\frac{d\mathcal{E}}{dt} = \frac{d(2\mathcal{E})}{dt} + \sum_{i=1}^n \frac{\partial \mathcal{V}}{\partial q_i} \dot{q}_i.$$

If \mathcal{V} does not depend on t explicitly, then

$$\frac{d\mathcal{V}}{dt} = \sum_{i=1}^n \frac{\partial \mathcal{V}}{\partial q_i} \dot{q}_i$$

and hence

$$\frac{d}{dt}(\mathcal{E} + \mathcal{V}) = 0. \quad (1.22.1)$$

So $\mathcal{E} + \mathcal{V}$ is time-independent, which is a statement of energy conservation. Again, however, we proceeded under certain assumptions: (1) \mathcal{E} is a homogeneous quadratic form with respect to the \dot{q}_i , having coefficients that do not depend explicitly on time t ; (2) the forces are potential, and the potential does not depend explicitly on t either. A system satisfying these assumptions is said to be *conservative*. Thus we have established that

For a conservative system of particles, $\mathcal{E} + \mathcal{V}$ the sum of kinetic energy and the potential function is conserved over time.

The function \mathcal{V} is often called the potential energy. This is reasonable in view of its role in the sum above. But there is a deeper reason for this name, which we will understand through consideration of the following problem.

In § 1.17 we mentioned energy conservation for a point mass m projected vertically upward through the gravitational field. The result is

$$\frac{mv^2}{2} + mgh = \frac{mv_0^2}{2} + mgh_0,$$

where v and h are the vertical velocity and height of the body, respectively, and v_0, h_0 are their initial values. Here $\mathcal{V} = mgy$; taking the $+y$ -direction upwards, we see that $\partial \mathcal{V} / \partial y = -(-mg)$ where $-mg$ is the weight force acting on the body. We recall that mgh is sometimes called the potential energy of the particle. It is, however, related to the work of the gravitational force and the expression $mg(h - h_0)$ is the corresponding change in the kinetic energy of the Earth. Despite the fact that this results in a negligible effect on the Earth's motion, the problem really does involve two bodies in principle. Here the Earth is something like an infinite source of produced work whose state does not change because of the work done. So the term "potential energy" is a convenient way to talk about the change in kinetic energy of the other bodies acting on the system under consideration, without explicitly mentioning those bodies or their states.