1 Mathematics and engineers

Is it justified to have a course in mathematics for engineers? Would it not be better to provide a course in electrical calculations for electrical engineers, chemical calculations for chemical engineers and so on through all the departments? The reason for teaching mathematics is that mathematics is concerned not with particular situations but with patterns that occur again and again. This is most obvious in elementary mathematics. In arithmetic, the number 40 may occur as $40, 40 horsepower, 40 tons, 40 feet, 40 atoms, 40 ohms and so on indefinitely. It would be most wasteful if we decided not to teach a child the general idea of 40, but left this idea to be explained in every activity involving counting or measurement.

Advanced mathematics cannot claim the universal relevance that arithmetic has. There are mathematical topics vital for aerodynamics that leave the production engineer cold. An engineer cannot simply decide to learn mathematics. He must judge wisely what mathematics will serve him best. His aim is not only to find mathematics that will help him frequently now, but also to guess what mathematics is most likely to help in industries and processes still to be invented.

Linear algebra qualifies on both counts; it is already used in most branches of engineering, and has every prospect of continuing to be.

Very many situations in engineering involve the mathematical concept of mapping. We shall restrict ourselves to the simplest class of these, namely *linear mappings*. We shall explain what these are and how to recognize an engineering situation in which they occur. And again in the interests of simplicity, we will not consider linear mappings in general, but restrict ourselves to linear mappings involving only a finite number of dimensions.

2 Mappings

Obviously we can find many applications of the scheme

$$\text{input} \xrightarrow{\text{Process}} \text{output}.$$

The process may involve actual material, as when certain components or ingredients (input) are used to produce some manufactured article (output). Or the process may

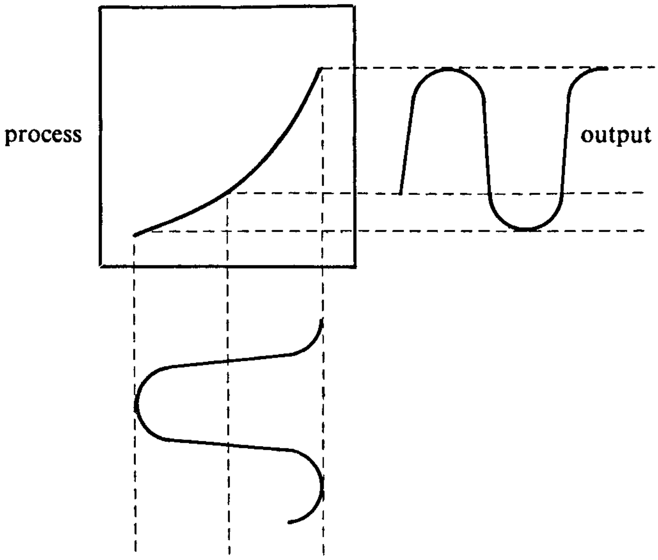

Fig. 1

be one of calculation, the input certain data, the output an answer. Another type of example would be a public address system; the input being the sounds made by a speaker or performer, the process being one of amplification and perhaps distortion, the output the sounds or noise emerging.

Diagrammatically, this last example might be shown as in Fig. 1.

Here distortion is occurring. The input is supposed to be a pure sine wave; the output certainly is not.

This is an example of a *mapping*, but it would not seem sensible to describe it as a linear mapping. We notice, for instance, that the graph representing the process is *not* a straight line.

There are two ways in which we might go about getting a straight-line graph. One way would probably be expensive – to use extremely good apparatus, which would yield a straight-line graph. The other way to avoid distortion is simpler and cheaper – to turn down the volume control. This means that we operate with a very small part of the curve. It usually happens that a small part of a curve is nearly (though of course not exactly) straight.

In fact, most of the examples we think of, that seem to arise in the first way (a graph that is really straight) are probably instances of the second way (using a small piece of a curve). Here we are thinking not only of our particular example – sound reproduction – but of all the cases in science where a straight-line law seems to hold. Someone might suggest, say, Hooke's Law: that for a stretched spring, force is proportional to deformation. But if we allow sufficiently large forces to act, the graph connecting force and deformation looks as is shown in Fig. 2. It seems reasonable to suppose that, if we

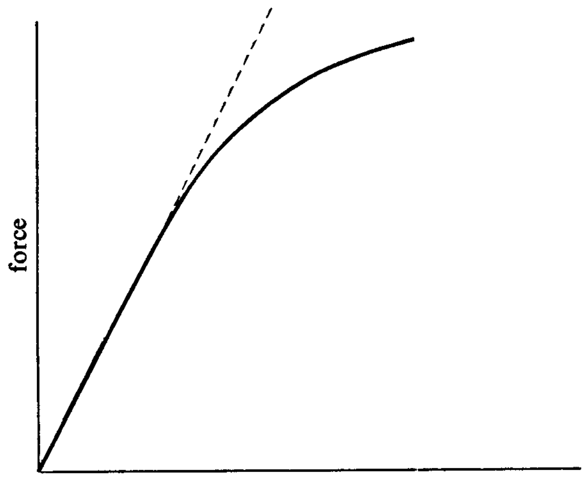

Fig. 2

could make very precise measurements, we would find a slight curvature in every part of the graph. But Hooke's formula is very convenient: it is easy to use and, for forces of moderate size, it gives all the accuracy we need.

In the same way, the formula $V = gt$ for a body falling from rest is linear, and is useful in many small-scale situations, but would be entirely misleading if used by astronauts. Ohm's Law, giving current and voltage proportional, is linear, but heavy currents are liable to produce changes in resistance and ultimately of course to melt the conductor.

Linear algebra then, like calculus, relies on the fact that, very often, a small part of a curve can be efficiently approximated by a straight line, a small part of a surface by a plane, and so on. (Question for discussion: what does 'and so on' mean here?) The qualification 'very often' is necessary; mathematicians have studied curves that are infinitely wriggly; for these no part, however small, can be approximated by a line. The functions corresponding to such graphs could appear only in quite sophisticated engineering problems.

The utility of linear algebra thus depends on a very general consideration – the tendency of functions to have good linear approximations. This argument is not tied to any branch of science, or to any particular department of engineering, or to the present state of the engineer's art; even if the whole of our present technology became obsolete, the argument would, in all probability, retain its force.

Sometimes of course an engineer has to deal with disturbances of such a scale that he cannot regard all the operations involved as effectively linear. He is then confronted with *non-linear problems*, which are much harder. As mentioned earlier, we do not plan to discuss these.

The illustrations so far given are not sufficient to make precise just what we mean by *linearity*. This idea we shall have to discuss for quite a while yet. Before we go into this it may be wise to remove a possible source of confusion. A linear mapping is a generalization of a type of function we meet in elementary algebra, namely $x \to y$, where $y = mx$ and m is constant. In elementary work, the function defined by $y = mx + b$ is usually called linear. It is important to realize that, in the language used by university mathematicians, $x \to mx + b$ would *not* be described as a linear function when $b \neq 0$. Our scheme 'input, process, output' may suggest why this is reasonable. We always suppose that if nothing goes in, nothing comes out. Zero input produces zero output. With $y = mx$, taking $x = 0$ makes $y = 0$, which is satisfactory. However, with $y = mx + b$, putting $x = 0$ gives $y = b$, which is not zero when $b \neq 0$. It has been agreed not to attach the label 'linear mapping' in this case.

3 The nature of the generalization

When $x \to y$ occurs in elementary work, it is understood that both the input x and the output y are real numbers. But the input–output idea is capable of vast generalization. Input and output could be almost anything: all we require is that the input in some way determines the output. For example, in designing an apartment building, we might be concerned about the effects of winds; the input would then be a specification of the wind acting on the building, the output perhaps the extra forces acting on the foundations as the result of such wind pressure. For an electronic computer, the programme fed in could be the input, its response the output. In an automated factory, the materials supplied could constitute the input, the goods produced the output.

It would be easy to produce many more examples; all that is necessary is that x, *what goes in* must determine y, *what comes out*; x and y can stand for extremely complicated objects or collections of data.

Any realistic person will realize that our scheme for a process or mapping, $x \to y$, is so general that very little information can be derived from it as it stands. Being told only that we have a mapping $x \to y$ is like being confronted with a machine having many controls and being told only that the machine is consistent – it always reacts in the same way to any particular setting of the controls. It may be comforting to know this, but it is not much to go on.

It is here that the restriction to *linear* mapping is helpful. But what do we mean by *linear* when we are dealing not with 'number → number' but with 'anything → anything'? Indeed, how has it come about that people have thought of using the term *linear* in such a vague and general situation?

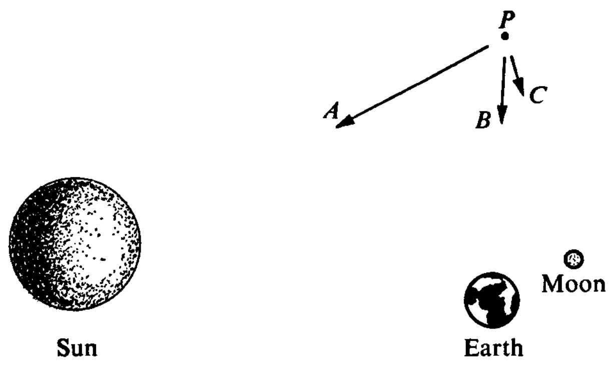

Fig. 3

To answer these questions we have to realize that this terminology arose only after many years' experience in a variety of subjects. In many sciences we meet something called a Principle of Superposition. In the theory of gravitation, for instance, we may want to find the force on an object at the point P, due to the attractions of the Sun, Earth and Moon. Let A be the force the object would experience if the Sun alone acted on it, B the force that would be produced by the Earth alone, and C that due to the Moon alone (Fig. 3). The principle of superposition states that the force due to the Sun, Earth and Moon acting simultaneously is the combined effect of the forces A, B, C. Thus one complex problem is broken into three simpler problems.

Again, in the theory of gravitation, a principle of proportion can be used. If a mass M exerts a force F on an object, then – provided the positions are unchanged – a mass of $3.7M$ will exert a force $3.7F$ on the object (Fig. 4). A mass kM would exert a force kF.

We should not think of the principles of superposition and proportion as applying to everything. For example, they do not apply to traffic.

In the situation shown in Fig. 5, it is possible that 20 cars a minute arriving from the north might pass without delay, if none came from the west. Also 30 cars a minute from the west might pass without delay if none came from the north. It would be rash to assume that, if both streams of traffic were arriving, 50 cars a minute would emerge

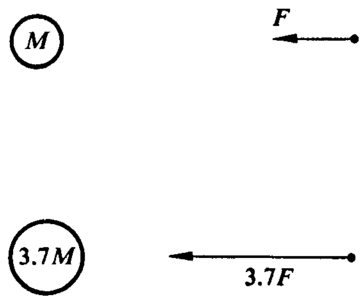

Fig. 4

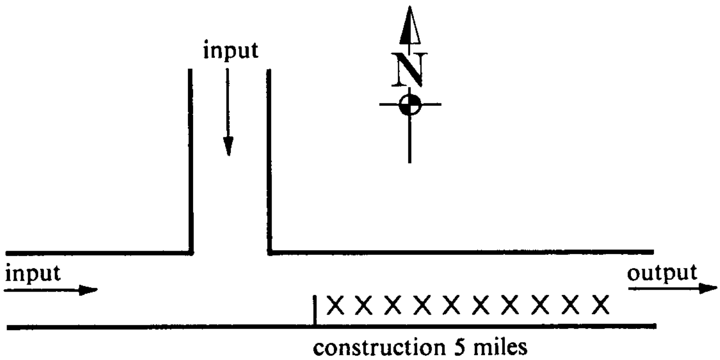

Fig. 5

beyond the construction area. Superposition does not apply. Nor does proportionality. If 10 times as many cars arrive per minute we do not expect to find 10 times as many emerging; rather we expect the rush-hour crawl.

Informally we may define a linear mapping as one arising in a situation to which the principles of superposition and proportionality apply.

Notice that, to make this explanation clear, we have to state carefully what we mean by superposition and proportionality both for input and output. In the gravitation question above, the input consists of a specification of the bodies exerting attraction. To superpose the three situations in which the Sun alone, the Earth alone and the Moon alone act on an object, we consider the Sun, Earth and Moon acting simultaneously. The output is concerned with the force experienced by the object; to apply superposition we must know how to find the combined effect of the three forces A, B and C. Someone unfamiliar with mechanics might be at a loss how to find this combined effect.

A number of situations are given below for consideration. In each case, it is necessary to spell out reasonable interpretations of superposition and proportionality both for the input and for the output. Then one has to consider whether principles of superposition and proportionality would apply (as in the gravitation problem) or would fail (as in the traffic problem).

Students may wish to select the situations that are most familiar to them.

Situations

(1) Purchasing articles at fixed prices (no reduction for quantity).
Input. Specification of articles purchased.
Output. Cost.

(2) Sand, or other heavy material, rests on a light beam (i.e. one of negligible weight). Spring balances measure the reactions at the ends (Fig. 6).

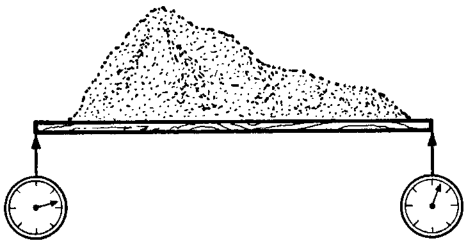

Fig. 6

Input. The distribution of weight along the beam.
Output. The readings of the spring balances.

(3) An electrical system of the type shown in Fig. 7.

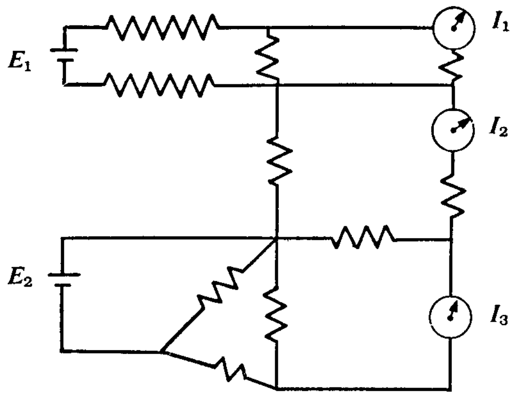

Fig. 7

Input. Voltages E_1, E_2.
Output. Currents I_1, I_2, I_3.

(4) Manufacturing.
Input. Materials required.
Output. Articles manufactured.

(5) Would it make sense to reverse input and output in Situation (4) i.e. to consider?
Input. Articles manufactured.
Output. Materials required.

(6) In order to obtain crude estimates of how quickly the temperature of a furnace is rising, the temperature (T) is measured at unit intervals of time, and the changes (ΔT) are calculated; for example

T	100		400		650		770		840
ΔT		300		250		120		70	

Input. The numbers in the T row.
Output. The numbers in the ΔT row.

(7) A student is given problems in differentiation. For example, given

$$f(x) = x^2 + 5x + 2,$$

he writes $f'(x) = 2x + 5$. (He does not make any mistakes.)
Input. The function $f(x)$.
Output. The function $f'(x)$.

(8) In an experiment, the mileage gone and the speedometer readings of a car are recorded at various times.
Input. Figures for mileage at these times.
Output. Speedometer readings at these times.

(9) In a radio programme listeners are able to phone in their opinions. In order to exclude obscenity, blasphemy, slander, sedition, etc. it is arranged that their remarks are heard on the radio ten seconds after they are spoken.
Input. The sounds made into the telephone.
Output. The sounds heard on the radio, on an occasion when nothing is censored.

(10) A number of pianos are available. A tape recorder deals faithfully with a certain range of notes on the piano. Notes above that it reproduces at half strength, and notes below that not at all.
Input. The sounds made by the pianos.
Output. The sounds as reproduced by the tape recorder.

(11) In pulse code telephony, the voice produces certain vibrations, shown in Fig. 8 as a graph. 8000 times a second a computer measures the ordinates (i.e. the y-values) shown here as upright lines.
Input. The voice graph.
Output. The values measured by the computer.

(12) Some function f, reals to reals is specified. Its values $f(0), f(1), f(2)$ are calculated.
Input. The function f.
Output. The numbers $f(0), f(1), f(2)$.

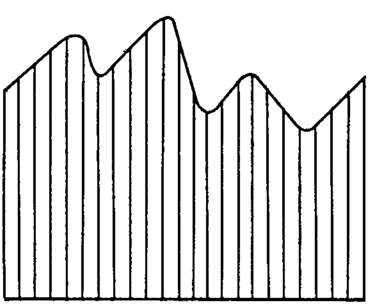

Fig. 8

(13) A function f, reals to reals, is specified. From it we calculate another function, g.
Input. The function f.
Output. The function g.

Discuss the following cases.
(*a*) $g(x) = f(2x)$.
(*b*) $g(x) = f(x - 1)$.
(*c*) $g(x) = f(x^2)$.
(*d*) $g(x) = [f(x)]^2$.

4 Symbolic conditions for linearity

Superposition and proportionality lend themselves to a very convenient representation by mathematical symbols.

In the gravitation problem, we wanted to find the attraction due to the Sun *and* the Earth *and* the Moon. In beginning arithmetic we associate 'and' with +. So it is natural to represent the superposition of Sun, Earth and Moon by $S + E + M$. The forces produced by Sun, Earth and Moon acting separately were indicated by A, B and C; it is natural to represent their combined effect by $A + B + C$, and indeed this notation will already be familiar to most students as vector addition of forces.

Thus from

$$S \to A$$
$$E \to B$$
$$M \to C$$

we conclude

$$S + E + M \to A + B + C.$$

Note that it is a subtle question of language whether we should say that the plus signs on the two sides of this statement have the same or different meanings. Someone could maintain that the meanings were the same because $S + E + M$ stands for 'the combined effect of S and E and M', while $A + B + C$ stands for 'the combined effect of A and B and C'. An opponent of this view would point out that the detailed procedure by which we find the combined effect of three forces is very different from that of considering three massive bodies existing together.

It will be seen from the list of situations given above that superposition applies to a great variety of problems. It would be very confusing to have a different sign for superposition in each of these cases. As we discussed right at the outset, mathematics is concerned with the similarities between different things, so that we learn one idea which can be used in many different circumstances. In order to bring out similarities, we use the same sign, +, whenever superposition is involved. But we must bear in mind the different systems that exist: we must never fall into the error of writing an expression that indicates the addition of a number to a force, or a mass to a voltage.

Proportionality can also be represented very simply. S standing for the Sun, $10S$ would indicate a body having the same position as the Sun but 10 times its mass. Again $10A$ would represent a force having the same direction as A but 10 times its magnitude.

The principle of proportionality shows that from the given fact

$$S \to A,$$

we may conclude

$$10S \to 10A.$$

We may now sum up our explanation of linearity.

We suppose we have an input, the elements of which may be denoted by $u, v, w, \ldots$ and an output with elements $u^*, v^*, w^*, \ldots$ In both the input and the output, we have sensible definitions of addition and of multiplication by a number, so we know what is meant by $u + v$ and ku, where k is any real number, and also what is meant by $u^* + v^*$ and ku^*. We have a mapping M which makes $u \to u^*$, $v \to v^*$, $w \to w^*$. This mapping will be called linear if for any two elements, u and v of the input, we find $u + v \to u^* + v^*$, and also that, for any number k and for any u, $ku \to ku^*$.

This is illustrated in Fig. 9.

The statement $u + v \to u^* + v^*$ means that we have two ways of finding where the mapping M sends $u + v$. First there is the obvious way, that works for any mapping. Starting with u and v in the input, we find $u + v$ which is also an element of the input. We then follow the arrow and find which element of the output we land on. The second method, which is not allowable for an arbitrary mapping but works with a linear

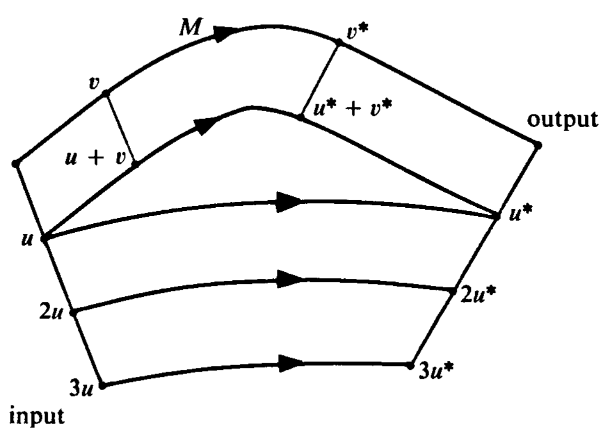

Fig. 9