1

Vectors, matrices, and derivatives

It is sometimes said that the great discovery of the nineteenth century was that the equations of nature were linear, and the great discovery of the twentieth century is that they are not.—Tom Körner, *Fourier Analysis*

1.0 Introduction

In this chapter, we introduce the principal actors of linear algebra and multivariable calculus.

By and large, first year calculus deals with functions f that associate *one* number $f(x)$ to *one* number x. In most realistic situations, this is inadequate: the description of most systems depends on many functions of many variables.

In physics, a gas might be described by pressure and temperature as a function of position and time, two functions of four variables. In biology, one might be interested in numbers of sharks and sardines as functions of position and time; a famous study of sharks and sardines in the Adriatic, described in *The Mathematics of the Struggle for Life* by Vito Volterra, founded the subject of mathematical ecology.

In microeconomics, a company might be interested in production as a function of input, where that function has as many coordinates as the number of products the company makes, each depending on as many inputs as the company uses. Even thinking of the variables needed to describe a macroeconomic model is daunting (although economists and the government base many decisions on such models). Countless examples are found in every branch of science and social science.

Mathematically, all such things are represented by functions $\mathbf{f}$ that take n numbers and return m numbers; such functions are denoted $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. In that generality, there isn't much to say; we must impose restrictions on the functions we will consider before any theory can be elaborated.

The strongest requirement one can make is that $\mathbf{f}$ should be *linear*; roughly speaking, a function is linear if when you double the input, you double the output. Such linear functions are fairly easy to describe completely, and a thorough understanding of their behavior is the foundation for everything else.

The first four sections of this chapter are devoted to laying the foundations of linear algebra. In the first three sections we introduce the main actors, vectors and matrices, and relate them to the notion of a linear function.

One problem of great interest at present is protein folding. The human genome is now known; in particular, we know the sequence of amino acids of all the proteins in the body. But proteins are only active when they are curled up in just the right way; the world's biggest computers are busy trying to derive from the sequence of amino acids just how they will fold.

Specifying the position and orientation of each of N amino acids requires $6N$ numbers; each such configuration has a potential energy, and the preferred folding corresponds to a minimum of this potential energy. Thus the problem of protein folding is essentially finding the minima of a function of $6N$ variables, where N might be 1000.

Although many approaches of this problem are actively being pursued, there is no very satisfactory solution so far. Understanding this function of $6N$ variables is one of the main challenges of the age.

Next we develop the geometrical language that we will need in multivariable calculus. The realization by René Decartes (1596–1650) that an equation can denote a curve or surface was a crucial moment in the history of mathematics, integrating two fields, algebra and geometry, that had previously seemed unrelated. We do not want to abandon this double perspective when we move to higher dimensions. Just as the equation $x^2 + y^2 = 1$ and the circle of radius 1 centered at the origin are one and the same object, we will want to speak both of the 9-dimensional unit sphere in $\mathbb{R}^{10}$ and of the equation denoting that sphere. In section 1.4 we define such notions as the length of a vector in $\mathbb{R}^n$, the length of a matrix, and the angle between two vectors. This will enable us to think and speak in geometric terms about higher-dimensional objects.

One object of linear algebra is to extend to higher dimensions the geometric language and intuition we have concerning the plane and space, familiar to us all from everyday experience.

In chapter 2, when we discuss solving systems of equations, we will also want to be able to interpret algebraic statements geometrically. You learned in school that saying that two equations in two unknowns have a unique solution is the same as saying that the two lines given by those equations intersect in a single point. In higher dimensions we will want, for example, to be able to speak of the "space of solutions" of a particular system of equations as being a four-dimensional subspace of $\mathbb{R}^7$.

In section 1.5 we will discuss sequences, subsequences, limits, and convergence. In section 1.6 we will expand on that discussion, developing the topology needed for a rigorous treatment of calculus.

Most functions are not linear, but very often they are well approximated by linear functions, at least for some values of the variables. For instance, as long as there are few hares, their number may well quadruple every three or four months, but as soon as they become numerous, they will compete with each other, and their rate of increase (or decrease) will become more complex. In the last three sections of this chapter we will begin exploring how to approximate a nonlinear function by a linear function – specifically, by its higher-dimensional derivative.

1.1 Introducing the actors: points and vectors

Much of linear algebra and multivariate calculus takes place within $\mathbb{R}^n$. This is the space of ordered lists of n real numbers.

You are probably used to thinking of a point in the plane in terms of its two coordinates: the familiar Cartesian plane with its x, y axes is $\mathbb{R}^2$. A point in space (after choosing axes) is specified by its three coordinates: Cartesian space is $\mathbb{R}^3$. Similarly, a point in $\mathbb{R}^n$ is specified by its n coordinates; it is a list of n real numbers. Such ordered lists occur everywhere, from grades on a transcript to prices on the stock exchange. Seen this way, higher dimensions are no more complicated than $\mathbb{R}^2$ and $\mathbb{R}^3$; the lists of coordinates just get longer.

The notion that one can think about and manipulate higher-dimensional spaces by considering a point in n-dimensional space as a list of its n "coordinates" did not always appear as obvious to mathematicians as it does today. In 1846, the English mathematician Arthur Cayley pointed out that a point with four coordinates can be interpreted geometrically without recourse to "any metaphysical notion concerning the possibility of four-dimensional space."

Example 1.1.1 (The stock market). The following data is from the *Ithaca Journal*, Dec. 14, 1996.

Local Nyse Stocks

	Vol	**High**	**Low**	**Close**	**Chg**
Airgas	193	24 1/2	23 1/8	23 5/8	-3/8
AT&T	36606	39 1/4	38 3/8	39	3/8
Borg Warner	74	38 3/8	38	38	-3/8
Corning	4575	44 3/4	43	44 1/4	1/2
Dow Jones	1606	33 1/4	32 1/2	33 1/4	1/8
Eastman Kodak	7774	80 5/8	79 1/4	79 3/8	-3/4
Emerson Elec.	3335	97 3/8	95 5/8	95 5/8	-1 1/8
Federal Express	5828	42 1/2	41	41 5/8	1 1/2

We can think of this table as five columns, each an element of $\mathbb{R}^8$:

"Vol" denotes the number of shares traded, "High" and "Low" the highest and lowest price paid per share, "Close" the price when trading stopped at the end of the day, and "Chg" the difference between the closing price and the closing price on the previous day.

$$\underbrace{\begin{bmatrix} 193 \\ 36606 \\ 74 \\ 4575 \\ 1606 \\ 7774 \\ 3335 \\ 5828 \end{bmatrix}}_{\text{Vol}} \quad \underbrace{\begin{pmatrix} 24\,1/2 \\ 39\,1/4 \\ 38\,3/8 \\ 44\,3/4 \\ 33\,1/4 \\ 80\,5/8 \\ 97\,3/8 \\ 42\,1/2 \end{pmatrix}}_{\text{High}} \quad \underbrace{\begin{pmatrix} 23\,1/8 \\ 38\,3/8 \\ 38 \\ 43 \\ 32\,1/2 \\ 79\,1/4 \\ 95\,5/8 \\ 41 \end{pmatrix}}_{\text{Low}} \underbrace{\begin{pmatrix} 23\,5/8 \\ 39 \\ 38 \\ 44\,1/4 \\ 33\,1/4 \\ 79\,3/8 \\ 95\,5/8 \\ 41\,5/8 \end{pmatrix}}_{\text{Close}} \quad \underbrace{\begin{bmatrix} -3/8 \\ 3/8 \\ -3/8 \\ 1/2 \\ 1/8 \\ -3/4 \\ -1\,1/8 \\ 1\,1/2 \end{bmatrix}}_{\text{Chg}}$$

Note that we use parentheses for "positional" data (for example, highest price paid per share), and brackets for "incremental" data (for example, change in price). Note also that we write elements of $\mathbb{R}^n$ as *columns*, not rows. The reason for preferring columns will become clear later: we want the order of terms in matrix multiplication to be consistent with the notation $f(x)$, where the function is placed before the variable.

Remark. Time is sometimes referred to as "the fourth dimension." This is misleading. A point in $\mathbb{R}^4$ is simply four numbers. If the first three numbers give the x, y, z coordinates, then the fourth number might give time. But the fourth number could also give temperature, or density, or some other kind of information. In addition, as shown in the above example, there is no need for any of the numbers to denote a position in physical space; in higher dimensions, it can be more helpful to think of a point as a "state" of a system. If 3356 stocks are listed on the New York Stock Exchange, the list of closing prices for those stocks is an element of $\mathbb{R}^{3356}$, and every element of $\mathbb{R}^{3356}$ is one theoretically possible state of closing prices on the stock market. (Of course, some such states will never occur; for instance, stock prices are positive.) $\triangle$

Points and vectors: positional data versus incremental data

An element of $\mathbb{R}^n$ is simply an ordered list of n numbers, but such a list can be interpreted in two ways: as a *point* representing a position or as a *vector* representing a displacement or increment.

Definition 1.1.2 (Point, vector, and coordinates). The element of $\mathbb{R}^n$ with coordinates $x_1, x_2, \cdots, x_n$ can be interpreted as the *point* $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, or as the *vector* $\vec{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, which represents an increment.

Example 1.1.3 (An element of $\mathbb{R}^2$ as a point and as a vector). The element of $\mathbb{R}^2$ with coordinates $x = 2, y = 3$ can be interpreted as the point

$\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ in the plane, as shown in figure 1.1.1. But it can also be interpreted as the instructions "start anywhere and go two units right and three units up," rather like instructions for a treasure hunt: "take two giant steps to the east, and three to the north"; this is shown in figure 1.1.2. Here we are interested in the displacement: if we start at *any* point and travel $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$, how far will we have gone, in what direction? When we interpret an element of $\mathbb{R}^n$ as a position, we call it a *point*; when we interpret it as a displacement, or increment, we call it a *vector*. $\triangle$

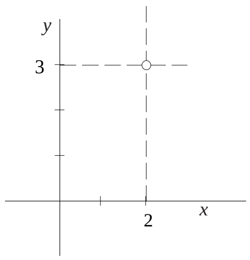

FIGURE 1.1.1.
The point $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Example 1.1.4 (Points and vectors in $\mathbb{R}^{3356}$). If 3356 stocks are listed on the New York Stock Exchange, the list of closing prices for those stocks is a point in $\mathbb{R}^{3356}$. The list telling how much each stock gained or lost compared with the previous day is also an element of $\mathbb{R}^{3356}$, but this corresponds to thinking of the element as a vector, with direction and magnitude: did the price of each stock go up or down? How much? $\triangle$

In the plane and in three-dimensional space a vector can be depicted as an arrow pointing in the direction of the displacement. The amount of displacement is the length of the arrow. This does not extend well to higher dimensions. How are we to picture the "arrow" in $\mathbb{R}^{3356}$ representing the change in prices on the stock market? How long is it, and in what "direction" does it point? We will show how to compute these magnitudes and directions for vectors in $\mathbb{R}^n$ in section 1.4.

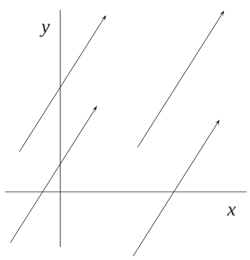

FIGURE 1.1.2.
All the arrows represent the same vector, $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

Remark. In physics textbooks and some first year calculus books, vectors are often said to represent quantities (velocity, forces) that have both "magnitude" and "direction," while other quantities (length, mass, volume, temperature) have only "magnitude" and are represented by numbers (scalars). We think this focuses on the wrong distinction, suggesting that some quantities are always represented by vectors while others never are, and that it takes more information to specify a quantity with direction than one without.

The volume of a balloon is a single number, but so is the vector expressing the difference in volume between an inflated balloon and one that has popped. The first is a number in $\mathbb{R}$, while the second is a vector in $\mathbb{R}$. The height of a child is a single number, but so is the vector expressing how much he has grown since his last birthday. A temperature can be a "magnitude," as in "It got down to -20 last night," but it can also have "magnitude and direction," as in "It is 10 degrees colder today than yesterday." Nor can "static" information always be expressed by a single number: the state of the stock market at a given instant requires one number for each stock listed – as does the vector describing the change in the stock market from one day to the next. $\triangle$