Chapter 1

Basic Matrix Operations

A *matrix* is just a rectangular array of numbers, nothing mysterious in itself. However, the algebraic interactions of such arrays, and their geometric manifestations, open up a world of fascinating mathematical phenomena. This is matrix theory, also known as linear algebra.

To keep things simple, the first part of this course will limit itself by and large to just two kinds of matrices:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{and} \quad X = \begin{bmatrix} x \\ y \end{bmatrix}.$$

When we say “matrix”, we shall usually mean the first of these types, preferring the term “vector” for the second. When dealing with vectors by themselves, we may often save space by simply writing $[x, y]$ instead of the column shown above. For typical matrix operations, however, the distinction between rows and columns is essential.

Vector quantities abound. For instance, instead of considering the value of an investment portfolio as a single sum, you might break it down into separate amounts, x for stocks and y for bonds; or, instead of looking at the total number of weeds on a lawn, you might count x dandelions and y plantains as separate populations. To study how these numbers change from one season to the next, you generally have to consider four coefficients: the growth rates of x and y by themselves, as well as the factors by which they hinder or enhance each other. This is where matrices come in; most often they occur as *multipliers of vector quantities.*

The multiplication of the vector X by the matrix A displayed above produces a new vector AX defined as follows:

$$AX = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

At first sight, this may seem somewhat arbitrary and a far cry from the usual meaning of "multiplication". However, let us note that a and d are indeed ordinary numerical multipliers of the component quantities x and y. If b and c are both $= 0$, in which case A is called *diagonal*, the components of AX are just ax and dy, and we have two ordinary multiplications stacked one upon the other — nothing new. What makes the definition of AX interesting is the presence of the terms involving b and c. For instance, the new first component $ax + by$ (say, dandelions) depends not only on the previous value of x (dandelions) but also on that of y (say, plantains). One of the major themes of this course is the search for ways of undoing, or at least restricting, this kind of linkage.

Below the ground-level of the theory to be developed, there are a couple of algebraic notions and operations which are almost self-evident. One of them is *matrix addition:* any two matrices of the same shape are "added" together by adding corresponding components or "entries". In the special case of column vectors X and X', this means:

$$\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + x' \\ y + y' \end{bmatrix}.$$

For good measure, the case of 2×2 matrices will be shown on the next page. The second of these basic operations is *scalar multiplication*, that is the *scaling* of all the entries of a matrix by the same number s. Again we show this happening to a column vector:

$$s \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} sx \\ sy \end{bmatrix},$$

with the understanding that it is also applicable to any kind of matrix M. The result of such an operation will be denoted by sM. Because of this role in the world of matrices, individual numbers are often referred to a *scalars.*

In this notation, a simple but fundamental feature of matrix action is the *linearity principle*, i.e., the validity of the formula

$$A(sX + tX') = sAX + tAX'$$

for any scalars s and t. Here is your first exercise in this course: verify this formula by writing out the details according to the definitions above.

§1. Addition and Multiplication

In this paragraph and the next one, we focus on the algebraic interactions of 2×2 matrices among themselves. Although our primary aim is to extend the basic operations of good old arithmetic to the world of matrices, the word "algebraic" is customary in this context, since it also involves some purely symbolic calculations.

Matrices can be added together in the most obvious manner. Given the matrices A and A', their *sum* $A + A'$ is defined by:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} a+a' & b+b' \\ c+c' & d+d' \end{bmatrix}. \tag{1.1}$$

As far as addition is concerned, a matrix is just like any string of numbers, its rectangular organization into rows and columns does not come into play.

Multiplication of matrices, however, is another story, the *product* of two matrices being defined as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & u \\ y & v \end{bmatrix} = \begin{bmatrix} ax+by & au+bv \\ cx+dy & cu+dv \end{bmatrix}. \tag{1.2}$$

Let A denote the first of the two matrices on the left of this equation, and let M stand for the second. Observe:

(1) the entry in row i and column j of the product AM is obtained by amalgamating the i-th *row* of the first factor A with the j-th *column* of the second factor M;

(2) the k-th column of AM is obtained by letting A act on the k-th column of M, i.e., if X and U denote the first and second columns of M, respectively, the columns of AM are AX and AU.

This "multiplication" obeys most (but not all!) of the laws familiar from working with numbers. First the bad news:

$$\boxed{AM \neq MA \quad \text{most of the time.}}$$

If $AM = MA$ does hold, we say that A and M *commute*. Alas, this does not happen very often. For illustration, let us compare left versus right multiplication of A by the *diagonal* matrix

$$D(s,t) = \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix}.$$

Using the definition given in Eq.(1.2), you can easily check the following identities

$$\begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} sa & sb \\ tc & td \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix} = \begin{bmatrix} sa & tb \\ sc & td \end{bmatrix}.$$

They show:

left (resp. right) multiplication by $D(s,t)$ scales the rows (resp. columns) by s and t.

Moreover, they reveal that matrices of the form

$$D(s,s) = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix}$$

are particularly civilized: whether used as left or right multipliers, they always have the effect of scaling all the entries of the other matrix by s. Hence they commute with any matrix, and are, in fact, the only matrices to do so (can you show this?). They are called "scalar matrices", and instead of writing $D(s,s)A$ or $AD(s,s)$ we usually just write sA.

Two extreme cases of scalar matrices are particularly important: the *identity* matrix $D(1,1)$, which has no effect as a multiplier — it preserves the identity of the other matrix — and further the *zero* matrix $D(0,0)$, whose multiplicative effect is annihilation. They are usually denoted by I and 0, respectively. Thus, $D(s,s) = sI$, and sI is the preferred notation.

Continuing the search for algebraic rules of conduct, we next come to a situation where matrices do behave like numbers, namely *distributivity* (multiplication interacting with addition):

$$A(M+L) = AM + AL \qquad \text{and} \qquad (A+B)M = AM + BM\,.$$

Both versions must be mentioned since — in the absence of commutativity — one does not follow from the other. Again, it would not be wise to accept them without a bit of reflection. If you go back to Eq.(1.2) and write out some of the details, you will come out with a better understanding of how this works. It is easy: your main challenge will probably be to find enough different lower case letters for the entries (try q, r, s, t for one set).

With distributivity in place, we can perform various low-level algebraic computations like $(A+B)(A-B) = A^2 + BA - AB - B^2$, but are severely hampered by not being able to string together more than two factors without worrying about brackets.

What we need to get over this hump is the so-called *associativity* of matrix multiplication, which says

$$L(AM) = (LA)M\,.$$

This rule is so familiar in the number context, that you might be tempted to think it is obvious. But think again: if L has the entries q, r, s, t, it asserts the equality

$$\begin{bmatrix} q & r \\ s & t \end{bmatrix} \begin{bmatrix} ax+by & au+bv \\ cx+dy & cu+dv \end{bmatrix} = \begin{bmatrix} qa+rc & qb+rd \\ sa+tc & sb+td \end{bmatrix} \begin{bmatrix} x & u \\ y & v \end{bmatrix}$$

which now looks like a minor miracle. Sharpen your pencil and check that it actually works. If you do that, you have earned the right to skip the rest of this paragraph, since it is completely dedicated to proving associativity in a more conceptual way.

PROOF.

We first reformulate the second observation following Eq.(1.2). If X and U are the columns of M, let us write $M = [X, U]$. In terms of the definition of AX given in the introduction to this chapter, we then get

$$AM = A[X, U] = [AX, AU]\,.$$

By the same token, we have $(LA)M = [(LA)X, (LA)U]$ and $L(AM) = L[AX, AU] = [L(AX), L(AU)]$. The upshot of all this is that we need only prove the identity

$$L(AX) = (LA)X\,,$$

where X is an arbitrary column.

However, even this simpler version of associativity is not yet obvious. To prove it, we analyse the action of A on X in terms of the columns of A:

$$AX = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix} = x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix} = xV + yW\,,$$

if V and W denote those columns, i.e., $A = [V, W]$. Since LA then has the columns LV and LW, it follows as above that $(LA)X = xLV + yLW$. Now we invoke the linearity principle stated in the introduction (but here applied to L) and get:

$$L(AX) = L(xV + yW) = xLV + yLW = (LA)X\,,$$

as was to be shown. We shall see later that this proof can be adjusted to work for 3×3 matrices and beyond.

We now have permission to leave off the brackets and simply write LAM for the product of those three matrices (in that order). But we want more: licence to omit brackets in a multiple matrix product $A_1 \cdots A_n$, for $n = 4$, 5, or higher. Suppose we have obtained it for all products with *less* than n factors. Then both sides of the equation

$$(A_1 \cdots A_{k-1})(A_k \cdots A_n) = (A_1 \cdots A_k)(A_{k+1} \cdots A_n)$$

make sense for any $1 < k < n$. To see that this actually *is* an equality, we try the associativity rule with $L = A_1 \cdots A_{k-1}$, $A = A_k$, and $M = A_{k+1} \cdots A_n$. It works! Hence all the different bracketings give the same result, which we can therefore write as $A_1 \cdots A_n$ as desired. To get a feel for this, you should try it on $n = 4$ and 5.

§2. Inversion and Singularity

Since we can multiply by the matrix A, it would be comforting to know that we can also "divide" by it. Is there a matrix A^{-1} which undoes multiplication by A, just as 1/3 cancels the multiplication by 3? The answer is *yes* for the first and *no* for the second of the following matrices:

$$A_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The problem with A_2 is that it is *singular* — a polite way of saying it is a killer matrix: it annihilates certain non-zero vectors X (e.g., column 2 of A_1). Such devastation cannot be undone. If you tried to resurrect X by some left multiplication of the equation $A_2X = 0$, you would only find $X = 0$. Moral: a singular matrix cannot have an inverse.

It is not hard to recognize singularity in a 2×2 matrix: any equation of the form $AX = 0$ translates into $xV + yW = 0$, i.e., $xV = -yW$ where V and W are the columns of A, while x and y are the components of X. If x, y are not both 0, we get $V = (-y/x)W$ or $(-x/y)V = W$. This shows:

> *A 2×2 matrix is singular if and only if one of its columns is a scalar multiple of the other.*

As far as inverses are concerned, singular matrices are certainly out. But does every non-singular matrix necessarily *have* an inverse? The answer (yes!) comes from the straightforward identity

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} \tag{1.3}$$

which will turn out to be a mine of information in more ways than one. The second factor appearing there is visibly concocted from the first factor A, and is known as its *adjoint*, denoted by $A^\star$. The scalar $(ad - bc)$ on the right hand side is called the *determinant*, $\det A$.

Equation (1.3) can thus be abbreviated as $AA^\star = (\det A)I$, and you should check that $A^\star A = (\det A)I$, as well. Dividing these equations by $\det A$, whenever possible, we draw the following conclusion.

$$\text{If } \det A \neq 0, \text{ put} \quad A^{-1} = \frac{1}{\det A} A^\star . \qquad \text{Then } AA^{-1} = A^{-1}A = I .$$

If $\det A = 0$, on the other hand, Eq.(1.3) reads $AA^\star = 0$, which shows that A is *singular*, hence cannot have an inverse.

Looking back to the product AM as given in Eq.(1.2), and grinding out its determinant, we obtain $(ax + by)(cu + dv) - (cx + dy)(au + bv) = adxv + bcyu - bcxv - adyu = (ad - bc)(xv - yu)$, i.e. the miraculous formula

$$\det(AM) = (\det A)(\det M) ,$$

which will be referred to as the "multiplicativity of determinants".

Our story gets yet another wrinkle from the simple observation that

$$\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} a+d & 0 \\ 0 & a+d \end{bmatrix} - \begin{bmatrix} a & b \\ c & d \end{bmatrix} ,$$

in other words, $A^\star = (\operatorname{tr} A)I - A$, where $\operatorname{tr} A$ stands for the scalar $(a + d)$, the *trace* of A.

Substituting this into $AA^\star = (\det A)I$ yields $(\operatorname{tr} A)\, A - A^2 = (\det A)I$, and hence the important *Cayley-Hamilton relation*:

$$A^2 - (\operatorname{tr} A)\, A + (\det A)\, I = 0 , \tag{1.4}$$

which will be put to work in the the examples and the next chapter.

Finally, it should be noted that inversion, insofar as it is possible, reverses the order of factors, i.e.

$$(AM)^{-1} = M^{-1}A^{-1},$$

because $M^{-1}A^{-1}AM = MIM^{-1} = I$.

Actually, this argument is incomplete: it shows only that $M^{-1}A^{-1}$ transforms AM back to I, i.e., acts as *an* inverse of AM. Who says that there can be only *one* possible inverse? Would the public be startled to read in tomorrow's paper that scientists — using a super-computer — have finally found a matrix A with two (slightly) different inverses?

Remember how this discussion began: we were looking for a matrix L which would undo the multiplication by A, i.e., change AX back to X for any X. Such a thing is called a *left inverse*, and is characterized by the condition that $LA = I$. One could also ask for a *right inverse* R, satisfying $AR = I$. On the face of it, the conditions on L and R seem quite different. Moreover, it seems conceivable that there could be several L and R.

However, if you take *any* right inverse R (for example, our A^{-1}) and multiply the equation $I = AR$ by *any* left inverse L, you get $L = L(AR) = (LA)R = R$. Hence there is only *one* inverse (if $\det A \neq 0$), and it is the A^{-1} we found.

§3. Matrices as Transformations

Geometrically a single real number x does not have much character: that is why there are not many graphic representations of life inside the number line. However, a *pair* $[x, y]$ of real numbers shows up as a *point in the plane* by the usual Cartesian coordinate trick, and planar images are exactly what our eyes (and much of our brains) are built for.

Since a picture is worth a thousand words, we can save a lot of time and effort by the use of imagery. But since pictures may also be deceptive, we shall keep our logic firmly rooted in the world of numbers, and carefully avoid mistaking geometric insights for ironclad truths.

Sometimes it is useful to interpret $[x, y]$ not as a static point but as a *vector*, i.e., a potential shift "x units east and y units north" applicable to any point. This view gives a geometric meaning to addition, namely performing one shift after the other. A vector is often depicted by an arrow showing the shift applied to a suitable point of departure (default = the origin). There is no general rule as to whether $[x, y]$ should be pictured as a point or a vector. That choice depends on the needs of the moment: the numbers do not care either way. If we wish to stress that we mean the *point*, we shall sometimes write (x, y) with round brackets.

Given a 2×2 matrix A, we now want to describe how multiplication by A *transforms* the points of the plane. We think of A as changing every point X to the "image point" AX. How is AX related to X? Before we generate a lot of words, let us look at a picture. As an example, we take the matrix

$$A = \begin{bmatrix} 1.3 & 0.4 \\ 0.1 & 1.0 \end{bmatrix}.$$

Figure 1.1 shows a coordinate system with some vectors before and after being hit by A. Actually it shows only a tiny part of the plane: on the left, the *unit square* — whose corners have the coordinates $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$ — on the right, its image. Of course, the two diagrams should really be slid one on top of the other so that the origins O coincide (after all $A0 = 0$). We have artificially separated them for the sake of clarity.

The three vectors drawn in on the left of the diagram are

$$U = \overrightarrow{PR} = \begin{bmatrix} 1.2 \\ 0.5 \end{bmatrix}, \quad V = \overrightarrow{RQ} = \begin{bmatrix} 0.8 \\ -1.0 \end{bmatrix}, \quad W = \overrightarrow{PQ} = \begin{bmatrix} 2.0 \\ -0.5 \end{bmatrix}.$$

Obverve how the fact $W = U + V$ shows up graphically: if you begin at P and follow U and then V, you arrive at the same place (namely Q) you would have reached by following W. Had you taken V first and U second — the "southern route"— you would have run from P to Q along a different triangle. If you draw all three vectors as starting at P, the sum W would show up as the diagonal in the parallelogram made up by these two triangles.

On the right of the diagram, we find AU, AV, and AW in the same relationship, although they are no longer the same vectors. The point P has changed to AP, whose coordinates in the original system can be found by applying A to P (written as a column). Ditto for Q and R.

FIGURE. 1.1. Unit square and vectors before and after being hit by A

For a systematic description of what is going on, note that any planar vector can be written as $X = xI^{(1)} + yI^{(2)}$, where $I^{(k)}$ stands for the k-th column of the identity matrix, while its image is given by $AX = xA^{(1)} + yA^{(2)}$, where $A^{(k)}$ denotes the k-th column of A. In detail:

$$X = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \Longrightarrow \quad AX = x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix}.$$

Thus AX is made up from the columns of A in the same way that X is made up from the columns of I.

Geometrically, the vectors $I^{(1)}$ and $I^{(2)}$ show up as the horizontal and vertical edges, respectively, of the unit square, while the columns $A^{(1)}$ and $A^{(2)}$ of A are represented by the corresponding edges of the parallelogram on the right of Fig.1.1. Hence, if you picture X as a diagonal arrow in a vector rectangle with sides $xI^{(1)}$ and $yI^{(2)}$, the vector AX appears as the diagonal in the parallelogram with sides $xA^{(1)}$ and $yA^{(2)}$.

In fact, if you imagine a rectangular coordinate grid overlaid on the plane as on graph paper, you will see it transformed into a (usually) skew grid made from copies of the parallelogram given by $A^{(1)}$ and $A^{(2)}$. Every X is simply transformed into the AX having the same "coordinates" x and y with respect to the new grid. This rule applies to points as well as vectors, since any point P has the same coordinates as the vector stretching from O to P.

Once this is understood, the actual graphics can be simplified to show only the unit square and its image. For illustration, consider the action of the following three matrices:

$$\begin{bmatrix} 4/5 & 0 \\ 0 & 3/2 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix},$$

which transform the unit square (dotted lines) in the manner indicated by the following sketches.

FIGURE. 1.2. Typical diagonal, shear, and rotation matrix actions

They are examples of three important special types, namely

$$D(s,t) = \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix} \qquad G_{12}(u) = \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix} \qquad R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

The diagonal matrix $D(s,t)$ affects the coordinates of a point separately, scaling each by the appropriate factor s or t. The axes stay put, the grid remains rectangular. If $s \neq t$, any vector not parallel to one of the axes is tilted toward the axis with the stronger scalar.

$G_{12}(u)$ produces a "shear", which simply moves every point $[x, y]$ horizontally by the amount uy. Every coordinate square is deformed into a parallelogram with the same base and height. Of course $G_{12}(u)$ has a sister $G_{21}(u)$ with the u in the lower left corner, which moves points vertically in a similar mannner.

$R(\theta)$ acts as a rotation of the plane through the angle θ. Since its columns are the images of $I^{(1)}$ and $I^{(2)}$, the definitions of sine and cosine show that these vectors are both rotated through θ. But then, so are $xI^{(1)}$ and $yI^{(2)}$, and therefore the whole rectangle defined by these.

These three archetypes (diagonal, shear, and rotation) will be met again and again. We shall eventually see that *any* matrix can be expressed as a product of shears and diagonals, or rotations and diagonals. In the more immediate future, every 2×2-matrix will be seen as "similar" to a scalar multiple of just one of these types.

ANGLES.

Let us take time out to examine what can be said about angles without breaking the solemn vow to keep "our logic firmly rooted in the world of numbers"— using geometry only as a source of inspiration. If you have no qualms about trusting trigonometry, you might wish to skip this discussion.

An *angle* is simply a point on the *unit circle*, i.e., a pair of numbers (x, y) such that $x^2+y^2 = 1$. As such it already has the two possible meanings of *point* or *vector*, and now gets a third: that of *angle*. While the vector $V = [x, y]$ indicates a "shift", the angle $\theta = \langle x, y\rangle$ indicates a "turn", namely the rotation given by

$$R(\theta) := \begin{bmatrix} x & -y \\ y & x \end{bmatrix} = x \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + y \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$= xI + yJ$, where J is the last of the matrices displayed above.

In this perspective, θ is just an abbreviation for $\langle x, y\rangle$, and $\cos\theta$, $\sin\theta$ are just fancy names for the components x, y. The reason for this notational extravaganza is that vectors and angles must be read very differently. For instance, the negative of the vector $V = [x, y]$ is of course $-V = [-x, -y]$, but the negative of the angle $\theta = \langle x, y\rangle$ is given by $-\theta = \langle x, -y\rangle$, which makes the reverse turn $R(\theta)^{-1}$.

Adding two angles $\theta_1 = \langle x_1, y_1\rangle$ and $\theta_2 = \langle x_2, y_2\rangle$ means performing the appropriate turns one after the other, that is, *multiplying* the matrices $R(\theta_1) = x_1 I + y_1 J$ and $R(\theta_2) = x_2 I + y_2 J$. This works out to $R(\theta_1)R(\theta_2) = (x_1x_2 - y_1y_2)I + (x_1y_2 + x_2y_1)J$, the minus sign in the first term coming from the fact that $J^2 = -I$. Since the product will still have determinant $= 1$, it is indeed a rotation — and *that* is how we define the "sum" of the two angles. In other words, we decree $R(\theta_1 + \theta_2) = R(\theta_1)R(\theta_2)$ or

$$\langle x_1, y_1\rangle + \langle x_2, y_2\rangle = \langle x_1x_2 - y_1y_2\,,\ x_1y_2 + x_2y_1\rangle\,.$$

For instance, if $\theta_1 = \langle 3/5, 4/5\rangle$ and $\theta_2 = \langle 12/13, 5/13\rangle$ we have $\theta_1 + \theta_2 = \langle 16/65, 63/65\rangle$, quite different from what we would get by vector addition. Incidentally, these angles would *measure* approximately 53.13°, 22.62°, and 75.75° — thus, the metaphorical plus-sign in $\theta_1 + \theta_2$ would correspond to an actual addition of real numbers. But angle measures, though "obvious" to the eye, are too tricky for our theory as it stands. They will be explored in Section 5 below.

§4 Examples and Exercises

Worked Examples

1. Find s and t such that

$$\begin{bmatrix} s & -1 \\ t & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}$$

is diagonal.

The off-diagonal entries in the product are $2s-6$ *and* $t+2$, *respectively. Both must be zero; hence the only possible choices are* $s = 3$ *and* $t = -2$.

2. Derive an expression for $\det(AB^{-1})$ in terms of $\det A$ and $\det B$.

Taking determinants on both sides of the equation $A = (AB^{-1}) \cdot B$, *we get* $\det A = \det(AB^{-1}) \cdot \det B$ *by the multiplicativity of determinants. Since* B^{-1} *exists, we have* $\det B \neq 0$ *and hence*

$$\det(AB^{-1}) = \frac{\det A}{\det B}\,.$$

Note: for $A = I$, *this yields the simple relation* $\det(B^{-1}) = (\det B)^{-1}$.

3. A flea market vendor buys two sizes of cotton towels for 3 and 4 dollars a piece, and sells them with a mark-up of 1 and 2 dollars, respectively. If he originally bought \$98 worth, and makes a profit of \$42, how many of each size did he have?

Let x and y be the numbers of towels in the two different sizes. His original outlay comes to $3x+4y$, his profit to $x+2y$. Equating these quantities to 98 and 42, respectively, we get the matrix equation

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 98 \\ 42 \end{bmatrix} \qquad \textit{or} \qquad \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & -4 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 98 \\ 42 \end{bmatrix} = \begin{bmatrix} 14 \\ 14 \end{bmatrix}.$$

Hence $x = y = 14$. Note: the second equation was obtained by multiplying the first equation (on the left) by the inverse of the coefficient matrix. This is not the only way of solving this kind of problem, but it is a particularly tidy one.

4. Find the matrix A of the linear transformation which interchanges the points $(3,4)$ and $(5,7)$.

Let V and W be columns with these same coordinates. We want an A such that $AV = W$ and $AW = V$. This can be summarized in the single equation $A\big[V, W\big] = \big[W, V\big]$, which can then be multiplied on the right (!) by the inverse of $\big[V, W\big]$, to yield A. In detail:

$$A \begin{bmatrix} 3 & 5 \\ 4 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 7 & 4 \end{bmatrix} \quad \Longrightarrow \quad A = \begin{bmatrix} 5 & 3 \\ 7 & 4 \end{bmatrix} \begin{bmatrix} 7 & -5 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} 23 & -16 \\ 33 & -23 \end{bmatrix}.$$

Note: in this scheme, the images AV and AW can be given arbitrarily. The main thing is that V, W be independent, so as to make the matrix $\big[V, W\big]$ invertible.

5. Find α and β so as to make $D(1,\alpha)\, G_{12}(1)\, D(1,\beta) = G_{12}(5)$.

In order to have

$$\begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix} = \begin{bmatrix} 1 & \beta \\ 0 & \alpha\beta \end{bmatrix}$$

equal to $G_{12}(5)$, we must make $\beta = 5$ and $\alpha = 1/5$. This works for any $k \neq 0$ instead of 5.

6. For the general matrix A with entries a, b, c, d find scalars r and q such that $(A - rI)^2 = qI$.

Cayley-Hamilton says: $(bc-ad)I = A^2 - (a+d)A = \big(A - \tfrac{1}{2}(a+d)I\big)^2 - \tfrac{1}{4}(a+d)^2 I$. Therefore, if we set $r = \tfrac{1}{2}(a+d)$, we obtain $(A - rI)^2 = qI$ with $q = bc - ad + \tfrac{1}{4}(a+d)^2 = bc + \tfrac{1}{4}(a-d)^2$.

7. Given that $A^2 = -I$, find the trace and the determinant of A.

Putting $\operatorname{tr} A = \tau$ *and* $\det A = \delta$, *we have* $A^2 = \tau A - \delta I$ *by Cayley-Hamilton. Now,*

$$A^2 = -I \qquad \Longrightarrow \qquad \tau A = (\delta - 1)I.$$

If τ *were non-zero, we would get* $A = \theta I$ *(with* $\theta = (\delta - 1)/\tau$*) and hence* $A^2 = \theta^2 I$, *which can never be* $= -I$. *Hence* $\tau = 0$, *and therefore also* $(\delta - 1) = 0$. *Upshot:* $\operatorname{tr} A = 0$ *and* $\det A = 1$.

8. If $M = \big[V, W\big]$ with independent columns V and W, compute $M^{-1}V$ and $M^{-1}W$.

Since $I = M^{-1}M = \big[M^{-1}V, M^{-1}W\big]$, *we have*

$$M^{-1}V = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad M^{-1}W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

9. Given a 2×2 matrix A and a column V, suppose that V and $W = AV$ are independent. Use the preceding example to compute $M^{-1}AM$, where $M = \big[V, W\big]$.

By Cayley-Hamilton, $AW = A^2V = \tau AV - \delta V$. *Therefore* $AM = \big[AV, AW\big] = \big[W, \tau W - \delta V\big]$ *with* $\tau = \operatorname{tr} A$ *and* $\delta = \det A$. *Multiplying this on the left by* M^{-1}, *we get*

$$M^{-1}W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \tau M^{-1}W - \delta M^{-1}V = \tau \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \delta \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

for the first and second column, respectively, of $M^{-1}AM$. *Altogether:*

$$M^{-1}AM = \begin{bmatrix} 0 & -\delta \\ 1 & \tau \end{bmatrix} .$$

10. Given vectors $V_1 = [x_1, y_1]$ and $V_2 = [x_2, y_2]$ with $x_1^2 + y_1^2 = x_2^2 + y_2^2 = 1$, show that there is exactly one rotation $R = xI + yJ$ such that $RV_1 = V_2$, i.e.,

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

with $x^2 + y^2 = 1$. Find x and y in terms of x_1, y_1, x_2, y_2.

Multiplying both sides of $RV_1 = V_2$ *by* J, *we obtain* $JRV_1 = RJV_1 = JV_2$ *and hence the expanded equation* $R \cdot [V_1, JV_1] = [V_2, JV_2]$. *Explicitly, this reads*

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{bmatrix} = \begin{bmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{bmatrix}$$

or $R \cdot R_1 = R_2$, *where* $R_k = x_k I + y_k J$ *for* $k = 1, 2$. *Consequently,* $R = R_2 \cdot R_1^{-1} = (x_2 I + y_2 J)(x_1 I - y_1 J)$ *is the only potential solution. Since* $J^2 = -I$, *this works out to* $xI + yJ$ *with*

$$x = x_1 x_2 + y_1 y_2 \quad \text{and} \quad y = x_1 y_2 - x_2 y_1 .$$

Now $\det R_k = x_k^2 + y_k^2 = 1$ *implies that* $x^2 + y^2 = \det R = 1$ *as well, and therefore* R *is indeed a rotation.*

Geometrically, R *is the rotation through the angle* θ *"between"* V_1 *and* V_2. *Defining the angles* θ_k *(for* $k = 1, 2$*) by* $\cos\theta_k = x_k$ *and* $\sin\theta_k = y_k$, *we must therefore take* $\theta = \theta_2 - \theta_1$.

Exercises

1. For $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ find B such that $AB = 0$ but $BA \neq 0$.
2. Find invertible matrices A and B such that AB is diagonal but BA is not.
3. Find all numbers k such that $\begin{bmatrix} 2 & k \\ k & 3 \end{bmatrix}$ is singular.
4. Find all numbers k such that $\begin{bmatrix} 2+k & 1 \\ 2 & 3+k \end{bmatrix}$ is singular.
5. Given that $u = 3x - 2y$ and $v = 2x - y$, express x and y in terms of u and v.
6. John has a sum of s cents in nickels and dimes, t coins in total. Express the number x of nickels and the number y of dimes in terms of s and t.
7. If $A = \begin{bmatrix} 3 & 2 \\ x & y \end{bmatrix}$ and $A^{-1} = \begin{bmatrix} 0 & -1 \\ u & v \end{bmatrix}$, find A.
8. Let $J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Find J^2, J^3, J^4, and derive a prescription for finding any power J^n.
9. With J as above, find $A = \begin{bmatrix} 3 & 5 \\ x & y \end{bmatrix}$ such that $AJ = JA$.
10. With J as above, compute $J^{-1} G_{12}(s) J$.
11. Compute $G_{12}(s) G_{12}(t)$ and $G_{12}(s)^n$. Find A such that $A^{-7} = G_{12}(3)$.
12. Find M such that $G_{21}(1/2) M G_{12}(1/2) = G_{12}(1) G_{21}(1)$.
13. Find a matrix A which transforms $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

14. If $V = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $W = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, find A with $AV = 2V$ and $AW = -W$.
15. With $M = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$ compute $A = MD(2, -3)M^{-1}$.
16. With $A = \begin{bmatrix} 1.7 & 0.6 \\ 0.6 & 0.8 \end{bmatrix}$ and M as above, compute AM and $M^{-1}AM$.
17. With $A = \begin{bmatrix} 1 + 2k^2 & k \\ k & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & k \\ -k & 1 \end{bmatrix}$, compute $(1 + k^2)^{-1}AB$.
18. With A as in (16), use (17) to find a rotation matrix P such that $AP = G_{12}(3/2)$.

In the following exercises, you will probably find it convenient to use the Cayley-Hamilton relation. Some of them refer to the partly given matrix

$$M = \begin{bmatrix} 2 & -1 \\ x & y \end{bmatrix}.$$

As usual, A is any 2×2 matrix with nothing given in advance.

19. Given that $\operatorname{tr} A = -1$ and $\det A = 1$ compute A^3.
20. Find x, y such that $\operatorname{tr} M = -1$ and $\det M = 1$; then determine M^3.
21. Given that $\operatorname{tr} A = 0$ and $\det A = 1$ compute A^4.
22. Using (21), find x, y such that $M^2 = -I$.
23. Find conditions on $\operatorname{tr} A$ and $\det A$ to ensure that $A^2 = A$ but $A \neq 0, I$.
24. Using (23), find x, y such that $M^n = M$, for $n = 1, 2, 3, \ldots$.
25. Prove that it is impossible to find a 2×2 matrix $A \neq I$ with *integer* entries and $A^5 = I$.

Hint: First show that $A^5 = (2\tau - \tau^3)I + (\tau^4 - 3\tau^2 + 1)A$, whenever $\det A = 1$. What kind of integer τ would allow this to equal I ?

Appendix: Measuring Arcs of Angles.

This section is dedicated to filling a gap in our account of rotations and angles: how to describe an angle by a single real number — its "arc" — instead of a pair of coordinates. Without such a numerical connection, formulas like $R(\theta)^n = R(n\theta)$ lose most of their usefulness. If you already know how this works, please feel free to skip ahead.

Here is the problem. The Cartesian plane, consists of nothing but pairs (x, y) of numbers. If $x^2 + y^2 = 1$, we may think of such a pair as an "angle" associated with the rotation matrix $R = xI + yJ$. There is no mystery about the numbers x and y: though pompously named $\cos(R)$ and $\sin(R)$, they are just data. But — within the Cartesian context — how do we measure the "arc" through which R rotates the plane?

On a physical plane (say, a piece of paper) we would use a protractor — a kind of yardstick bent into a semi-circle. Its rim has a number of equally spaced *markings*, traditionally 180 of them, and we measure an angle by *comparing* it with the angles given by these. Mathematically, the markings on a standard protractor are the powers $G, G^2, G^3, \ldots$ of a "gauge" rotation (representing 1 degree) which is one ninetieth of a right angle, i.e., $G^{90} = J$. An angle θ is then measured (approximately) by comparing $R(\theta)$ with the powers G^k to see where it fits into the sequence.

For the protractor markings, we shall abandon the impractical "degree" measure, which would force us to solve the difficult equation $G^{90} = J$. Instead, we shall construct a string of rotations $B_1, B_2, \ldots$ whose arcs are known fractions of the quadrant. Starting with $B_0 = J$, every successive B_m will be obtained from the preceding one by *bisection*, and therefore has exactly half the arc.

Bisecting a rotation matrix $A = uI + vJ$, with $v > 0$, is easy. It means finding a matrix $B = xI + yJ$ such that $y > 0$ and $B^2 = A$. Since $(xI + yJ)^2 = (x^2 - y^2)I + (2xy)J$, we must have $x^2 - y^2 = u$ and $2xy = v$. Since $\det A = 1$ is to be the square of $\det B = x^2 + y^2$, we further get $x^2 + y^2 = 1$, whence $2x^2 = 1 + u$. If both v and y are to be positive, $2xy = v$ implies that x must also be positive. The resulting equations and inequalities are uniquely solvable by

$$(\dagger) \qquad x = \sqrt{\frac{1+u}{2}}\,, \qquad y = \frac{v}{2x}\,.$$

Since this process is unique, it makes perfect sense to write $B = A^{1/2}$.

The matrices $B_1 = J^{1/2}, B_2 = J^{1/4}, \ldots, B_m = J^{1/2^m}$ so obtained have arcs measuring $1/2, 1/4, \ldots, 1/2^m$ quadrants. As k runs from 1 to 2^m, the powers B_m^k produce that many equal sectors in the first quarter circle. Therefore we shall say that the k-th power of B_m has an *arc of* $k/2^m$ *quadrants*, and write $B_m^k = R(k/2^m)$. Figure 1.3 shows the resulting "protractor" for $m = 5$. Note that its precision could be refined indefinitely by increasing m.

For such measures of arc $\mu = k/2^m$ and $\mu' = k'/2^m$, we clearly have

$$(**) \qquad R(\mu)R(\mu') = R(\mu + \mu')\,,$$

where the + now stands for the actual addition of numbers, rather than the symbolic addition of angles. Based on the obvious rule $B^k B^{k'} = B^{k+k'}$, this works for any integers k and k' without restriction — as long as we keep in mind that the circle eventually starts over again, i.e., that $R(4) = J^4 = I$. Equation $(**)$ looks even more convincing if written as $J^{\mu+\mu'} = J^{\mu}J^{\mu'}$.

Our unit of measurement is the *quadrant*, i.e., the right angle, which in conventional terms equals 90 degrees or $\pi/2$ radians. To use calculators, we must of course revert to one of these traditional scales.

At this point, you will have a fair idea of the connection between arcs and coordinates, and may wish to move on. Some of us must stay behind to argue the case that every R can indeed be approximated (to any desired

FIGURE 1.3. Semi-circle (two quadrants) divided into 64 equal sectors by the powers of B_5, each sector constituting $(1/2)^5 = 1/32$ of a right angle

precision) by the "binary" matrices B_m^k. Visually this may seem obvious, but ever since the ancient Greeks found invisible cracks in their number line, purely optical evidence has been inadmissible in mathematical courts.

First we need to deal with the *comparison* of angles in mathematical terms. We may, however, stay within the first quadrant — where sines and cosine are non-negative — since the multiplier J (that is, adding a right angle) easily moves us from one quadrant to the next. Rotations R, S in the first quadrant can be *compared* as follows: R is "greater than or equal to" S (written $R \geq S$) whenever RS^{-1} is still in this quadrant; otherwise R is "less than" S (written $R < S$). This relation is conveniently reflected by the size of the sines: $\sin(R) \geq \sin(S)$ implies $\cos(S) \geq \cos(R)$, and together these give $\sin(RS^{-1}) = \sin(R)\cos(S) - \cos(R)\sin(S) \geq 0$, with analogous implications in the case of "less than". The cosines remain non-negative both times, because $\cos(RS^{-1}) = \cos(R^{-1}S) = \cos(R)\cos(S) + \sin(R)\sin(S)$.

In approximating a rotation by special "binary" ones, the idea is to sneak up on the given R by a product of the form

$$P_m = B_1^{e_1} B_2^{e_2} B_3^{e_3} \cdots B_m^{e_m} \qquad \text{with} \qquad P_m \leq R < P_m B_m \,,$$

where each of the exponents $e_1, e_2, e_3, \ldots$ is either 0 or 1. Every step of this process goes as follows: having successfully concocted P_{m-1} (where m could be as low as 1, with $P_0 = I$), we put

$$P_m = \begin{cases} P_{m-1}B_m; & \text{if } P_{m-1}B_m \leq R, \\ P_{m-1}; & \text{if } R < P_{m-1}B_m. \end{cases} \tag{‡}$$

In the first case ($e_m = 1$) we have $P_m B_m = P_{m-1} B_m^2 = P_{m-1} B_{m-1}$, which is greater than R by construction of P_{m-1}. In the second case ($e_m = 0$)

we get $R < P_m B_m$ by default. Note: $P_m = B_m^{k_m} = R(k_m/2^m)$, where $k_m = 2^{m-1}e_1 + \cdots + 2e_{m-1} + e_m$.

It is easy and fun to write a computer program based on this process. For $R = .6I + .8J$ and $m = 21$, it yields the exponent string $e_1e_2e_3 \ldots e_{21} =$ 100101110010000000101. This is the binary expansion of $k_m/2^m = .5903344$ — which multiplied by 90 (to go from quadrants to degrees) amounts to 53.1301° — a very nice approximation to the arcsine of .8.

This was not just luck either: the proviso $P_m \leq R < P_m B_m$ will ensure that P_m always approaches R if m is chosen large enough — since B_m converges to I as m grows indefinitely, as we shall presently demonstrate. Again this is visually obvious, but since such insight (though very helpful) is not conclusive, let us take time out to settle the matter.

For $B_m = u_m I + v_m J$ and all m, we shall prove $u_m^2 \geq 1 - 1/2^m$ (and hence $v_m^2 \leq 1/2^m$). As this is clear for $m = 0$ and $m = 1$, suppose it established up to some m. The trick is to remember that $w \geq w^2$ for any w between 0 and 1 — which now gives $u_m \geq 1 - 1/2^m$. Adding 1 and dividing by 2, Eq.(†) takes us to the next rung of the ladder, namely $u_{m+1}^2 = (1 + u_m)/2 \geq 1 - 1/2^{m+1}$, and so on and on. Choosing m large enough, we can obviously make $1 - u_m \leq 2^{-m}$ and $v_m \leq 2^{-m/2}$ as small as we wish.

If continued indefinitely, the recursive process shown in Eq.(‡) yields an infinite binary expansion — i.e., a *real number* — for the exact arc measure of R. Conversely, *any* infinite string $\{e_m\}$ of zeroes and ones will define a sequence $\{P_m\}$ of rotations as above, which converges to *something* because $\{\sin(P_m)\}$ forms a non-decreasing sequence of real numbers < 1. Thus any real number μ between 0 and 1 leads — via its binary expansion — to a rotation $R(\mu)$ which may be thought of as an infinite product

$$R(\mu) = B_1^{e_1} B_2^{e_2} \cdots B_m^{e_m} \cdots = J^\mu$$

The addition law $(**)$ remains intact, because it holds for every stage of the approximation, and matrix multiplication is compatible with limits. It also shows that $R(\mu)$ varies *continuously* with μ, because $R(\mu')$ gets arbitrarily close to I for small enough μ'.

If the last two sentences seem somewhat mysterious, please disregard the clauses starting with "because". They are meant for those who know the yoga of limits and convergence, and wish to practise it on this occasion.