Chapter 2
Similarity and Standard Forms

Imagine yourself sitting in a room working away at problems about vectors and matrices. An assistant sits in another room and acts as your numerical coprocessor: when you send him vectors, he performs the required arithmetic and sends back the results. For technical reasons, the transmission of your data is encoded by a certain invertible matrix L. That is, every vector X sent by you is received as $Y = LX$ by him; every Y sent by him is received as $X = MY$ by you, where $M = L^{-1}$.

Now suppose you are exploring the nature of a matrix A by watching it multiply various vectors X. What does the Assistant see? Since he receives your X as LX and your AX as $L(AX)$, he sees you turning $Y = LX$ into $L(AX) = (LAL^{-1})Y$. In other words, he observes the multiplication of Y by the matrix

$$B = LAL^{-1}\,.$$

Conversely, whenever he uses B in his operations, the screen in your room plainly shows the action of A. From the assistant's point of view, your

$$A = MBM^{-1}$$

is an illusion produced by the encoding of his data with M. Of course, both perspectives are valid. Since the work can be done in either room, you may do it wherever it is easiest.

Matrices A and B related in this way are said to be *similar*. Given A, our aim in this chapter is to choose M so as to make $B = M^{-1}AM$ as simple as posssible for the overworked assistant. We shall find three cases, each leading to a different "standard form" for B.

§1. Eigenvectors and Diagonalization

Like *sauerkraut* and *wanderlust*, the German word *eigen* has kindly been adopted by the English language because it has no immediate relatives here. Its meaning dangles somewhere between "own" , "characteristic", and "proper".

An *eigenvector* of A is any vector $V \neq 0$ which cannot distinguish the action of A from multiplication by a certain ordinary scalar λ, i.e.,

$$AV = \lambda V \qquad \text{or} \qquad (A - \lambda I)V = 0\,. \tag{2.1}$$

Note that eigenvectors come in bunches. If $AV = \lambda V$ then also $AV' = \lambda V'$ for every scalar multiple $V' = tV$. For fixed V and variable t, all these vectors tV — interpreted as *points* in the plane — form a line through the origin, called an *eigenline* of A. They all experience the action of A as a simple scaling by the same *eigenvalue* λ.

Let us check this for the three archetypes introduced at the end of the last chapter. The diagonal matrix $D(s,t)$ has both the x-axis and the y-axis as eigenlines with $\lambda_1 = s$ and $\lambda_2 = t$, respectively. If $s \neq t$, there are no others; if $s = t$, every line through the origin is an eigenline. In either case, we can say that the number of eigenlines is at least 2.

The shear $G_{12}(u)$ leaves every point on the x-axis where it was, hence has the x-axis as an eigenline with $\lambda = 1$; if $u \neq 0$, there are no others. Finally, $R(\theta)$ turns all vectors through the angle θ, hence has no eigenlines if $\theta \neq 0$. Taking a cue from these, we shall classify 2×2 matrices according to the number of their eigenlines, saying that A represents the Case (a), (b), or (c) if the number of its eigenlines is

$$\text{(a)} \quad \geq 2 \qquad \text{or} \quad \text{(b)} \quad = 1 \qquad \text{or} \quad \text{(c)} \quad = 0\,,$$

respectively. In this section, we shall deal with Case (a).

If A has two different eigenlines, let V and W be vectors representing these, and suppose that s and t are the corresponding eigenvalues, i.e.,

$$AV = sV \qquad \text{and} \qquad AW = tW\,.$$

Forming the matrix $[V,W]$ with V and W as columns, we can rewrite this as the single equation $A[V,W] = [sV,tW] = [V,W]D(s,t)$. Lying on different lines, V and W cannot be scalar multiples of each other. Hence the matrix $[V,W]$ is invertible, and we get

$$\boxed{M^{-1}AM = D(s,t)\,, \quad \text{where } M = [V,W]\,.}$$

In the language introduced earlier: *A is similar to the diagonal matrix $D(s,t)$*. Such an A will be called *diagonalizable.*

For a numerical and pictorial example, we call once more on the matrix whose action was already shown back in Fig.1.1, namely

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

We also consider the vectors $\overrightarrow{OQ}$ and $\overrightarrow{OP}$ implicit in that diagram, but here rename them V and W (no longer having any use for the old V and W from Fig.1.1). Explicitly, under the new labels:

$$V = \begin{bmatrix} 1.2 \\ 0.3 \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} -0.8 \\ 0.8 \end{bmatrix}.$$

Now, if you transform V and W by A — do try this! — you will find $AV = 1.4\,V$ and $AW = 0.9\,W$. Hence V and W are independent eigenvectors with eigenvalues 1.4 and 0.9, respectively, and we get $M^{-1}AM =$

$$\begin{bmatrix} 1.2 & -0.8 \\ 0.3 & 0.8 \end{bmatrix}^{-1} \begin{bmatrix} 1.3 & 0.4 \\ 0.1 & 1.0 \end{bmatrix} \begin{bmatrix} 1.2 & -0.8 \\ 0.3 & 0.8 \end{bmatrix} = \begin{bmatrix} 1.4 & 0 \\ 0 & 0.9 \end{bmatrix}.$$

FIGURE 2.1. Similarity: D and A describe “the same” process viewed under different perspectives; the change of perspective is given by M (and its inverse).

Figure 2.1 is a graphic rendition of this similarity relation. The action of A is depicted by the right vertical arrow. The old unit square and its image — the major features of Fig.1.1 — are now shown by dotted lines for orientation only. The solid parallelograms are defined by V, W (top) and AV, AW (bottom). Since they are delimited by eigenlines, A does not rotate their contours, but only expands or shrinks them.

The two horizontal arrows both show the actiom of M, which produces these parallelograms from the unit square (top) and a certain rectangle (bottom),respectively. The left hand side of the diagram represents what the Assistant sees: the unit square turned into that rectangle; i. e., the action of a diagonal matrix.

By following one of the Assistant's vectors Y clockwise around the diagram, we can see how this ties in with the algebraic expression $M^{-1}AM$. First, Y turns into MY, then into $A(MY)$, and finally into $M^{-1}\big(A(MY)\big) = (M^{-1}AM)Y = D(1.4, 0.9)Y$.

The major question remaining is: how does one *find* those marvellous eigenvectors? Equation (2.1) itself does not hold out much hope, since both V and λ are unknown. This difficulty is overcome by the following sneak attack on λ alone.

Since V must be $\neq 0$ to qualify as an eigenvector, Eq.(2.1) says that λ is an eigenvalue of A if and only if the matrix $(A - \lambda I)$ is singular, i.e. $\det(A - \lambda I) = 0$. Expanding this determinant, we get $(a - \lambda)(d - \lambda) - bc = \lambda^2 - (a+d)\lambda + (ad - bc)$. Therefore eigenvalues in limited supply: they must satisfy the *characteristic equation*

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

where $\operatorname{tr} A = a + d$ and $\det A = ad - bc$ are the same coefficients we met in Eq.(1.4), the Cayley-Hamilton relation. By the usual strategems of solving quadratic equations, this simplifies to $(\lambda - r)^2 = q$, where

$$r = \frac{a+d}{2} \qquad \text{and} \qquad q = \frac{(a-d)^2 + 4bc}{4}\,. \tag{2.2}$$

There are at most two solutions $\lambda = r \pm \sqrt{q}$, and once these are known, it is easy to solve $(A - \lambda I)V = 0$ for the actual eigenvectors V. Since any such V can be scaled arbitrarily, there is a lot of choice at this stage. In our example for Fig.2.1, the vectors V and W were chosen for good looks in the graphics. For purely numerical work, we would have preferred integer components such as $[4, 1]$ and $[-1, 1]$ instead of $[1.2, 0.3]$ and $[-0.8, 0.8]$.

If $q > 0$, we get two different eigenvalues $\lambda_1 = r + \sqrt{q}$ and $\lambda_2 = r - \sqrt{q}$, and hence at least two eigenlines. Can there be more?

Suppose U, V, and W represent 3 different eigenlines. Since there cannot be 3 different eigenvalues, two of these vectors — say V and W — will have to share the same eigenvalue s. Forming $M = [V, W]$ as above, we would then have $AM = sM$, whence $A = sI$, because M is invertible.

Moral: the only 2×2 matrices with more than 2 eigenlines are the scalar matrices sI.

§2. Triangular Matrices

Before getting into the main topic of this section, it is necessary to make two general observations about similar matrices. Recall that A and B are similar if there is an invertible M such that

$$B = M^{-1}AM \quad \text{or} \quad A = MBM^{-1}\,.$$

Although we are eventually aiming at *constructing* a suitable M to relate a given A to a more convenient B, the mere definition of similarity makes no assumption about the nature of M. Any invertible matrix will do.

Nevertheless, any two similar matrices A and B have a lot in common. Since $(B - \lambda I) = M^{-1}(A - \lambda I)M$, we find that

$$\det(B - \lambda I) = \det M^{-1} \cdot \det(A - \lambda I) \cdot \det M = \det(A - \lambda I)\,.$$

Similar matrices have the same characteristic equation, hence the same determinant, the same trace, and the same eigenvalues.

They may not sport the same eigenvectors, but they will fall into the same one of the patterns (a), (b), or (c) outlined in the last section. In fact, since $\lambda Y = BY = M^{-1}AMY$ means precisely $\lambda MY = AMY$, it follows that Y is an eigenvector for B if and only if MY is one for A.

The act of changing a given A to a similar $B = M^{-1}AM$ by means of an invertible M is called *conjugating* A *with* M. Our second general observation is that you can conjugate repeatedly and always remain in the same similarity class. For instance, if $B = M^{-1}AM$ and $C = P^{-1}BP$, then $C = P^{-1}(M^{-1}AM)P = (MP)^{-1}A(MP)$, in other words, C is obtained from A by conjugating with MP.

Now let us return to our generic A and the search for convenient similar matrices $B = M^{-1}AM$. This time, however, assume only that the first column V of $M = [V, W]$ is an eigenvector, allowing the other column W to be anything except a scalar multiple of V. Since $M^{-1}M = I$, we have $M^{-1}V = I^{(1)}$, and the first column of $M^{-1}AM$ turns out as

$$M^{-1}AV = M^{-1}sV = sI^{(1)}\,,$$

where s is the eigenvalue associated with V. Had we taken W to be an eigenvector, our result would have involved *its* eigenvalue and $I^{(2)}$:

If the k-th column of M is an eigenvector of A with eigenvalue λ, the k-th column of $M^{-1}AM$ equals λ times the k-th column of I.

In other words, $B = M^{-1}AM$ will have the form

$$B = \begin{bmatrix} s & u \\ 0 & t \end{bmatrix} \qquad \text{or} \qquad B = \begin{bmatrix} s & 0 \\ u & t \end{bmatrix}, \tag{2.3}$$

depending on whether the first or second column of M is an eigenvector. In the first case B is called *upper* — in the second, *lower* — *triangular*. Such matrices wear their eigenvalues openly on the diagonal: indeed, the characteristic equation is just $\det(B - \lambda I) = (\lambda - s)(\lambda - t) = 0$.

All this is not terribly exciting if $s \neq t$, i.e., if $q > 0$ as in the last section. In that case we need not put up with an off-diagonal entry $u \neq 0$ — unless we have special reasons to conjugate with an $M = [V, W]$ in which W is not an eigenvector of A. But in Case (b), where A has only one eigenline (whence $t = s = r$), it seems the best we can do.

As a way to construct "the standard form", however, it has a serious flaw: even if we opt for one of the forms (say, the first) of B in Eq.(2.3), and even though we know its diagonal entries (both $= r$), the off-diagonal u can still be all over the place — cf. Example 7 below.

Help comes via Example 6 of Chapter 1, where we had rewritten the Cayley-Hamilton relation in the form

$$(A - rI)^2 = qI\,, \tag{2.4}$$

with r and q as in Eq.(2.2). Since $q = 0$ in Case (b), this yields a custom-made eigenvector $V = (A - rI)W$ for any W picked outside the eigenline: $V = (A - rI)W \neq 0$, but $(A - rI)V = (A - rI)^2 W = 0$.

With these choices, it is easy to check — cf. Example 8 below — that

$$M^{-1}AM = \begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix}, \quad \text{if } M = \big[(A - rI)W, W\big]\,.$$

This is our *standard form* for a 2×2 matrix in Case (b). For ease of calculation, we shall usually take W to be one of the columns of I. However, in some circumstances — e.g., if r is huge or tiny — the 1 in the upper right corner may be a poor choice, and we might have to return to the more flexible approach of Eq.(2.3).

For illustration, let A be the matrix with the entries $a = 1.3$, $b = -0.1$, $c = 0.1$, $d = 1.1$, whose characteristic equation is $\lambda^2 - 2.4\lambda + 1.44 = (\lambda - 1.2)^2 = 0$. Pick any non-eigenvector, say $W = I^{(1)}$. Then $V = (A - 1.2I)W$ is the first column of $A - 1.2I$, both of whose coordinates are $= 0.1$. Thus,

$$A = \begin{bmatrix} 1.3 & -0.1 \\ 0.1 & 1.1 \end{bmatrix}, \; M = \begin{bmatrix} 0.1 & 1 \\ 0.1 & 0 \end{bmatrix}, \; M^{-1}AM = \begin{bmatrix} 1.2 & 1 \\ 0 & 1.2 \end{bmatrix}$$

as promised. This is the pattern we shall follow in almost all computations — without, however. forgetting its potential weaknesses. In a diagram à la Fig.(2.1), for instance, the two very unequal columns of the matrix M used above would look a bit bizarre. Conjugation by the matrix

$$M_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{with} \quad M_1^{-1}AM_1 = \begin{bmatrix} 1.2 & -.2 \\ 0 & 1.2 \end{bmatrix}$$

would make a better picture, although it does not yield the standard 1 in the resulting upper triangular matrix.

§3. The Irreducible Case

If A has *no* eigenlines, where do we find a "natural" pair of vectors for our similarity matrix $M = [V, W]$? One possible idea is to start with some $W \neq 0$ and to put $V = AW$. Since there are no eigenvectors, V is never a scalar multiple of W, and $M = [V, W]$ is at least always invertible.

Let us follow this up in a particularly "easy" case: imagine a matrix E with $E^2 = -I$. By Cayley-Hamilton, this happens whenever $\operatorname{tr} E = 0$ and $\det E = 1$, i.e., the characteristic equation is the clearly unsolvable $\lambda^2 + 1 = 0$. (You should have no trouble inventing examples of such E.)

Putting $M = [V, W]$, where $V = EW$ for some $W \neq 0$, we then get $EV = E^2W = -W$, hence $E[V, W] = [-W, V]$ or

$$EM = M \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.5}$$

The matrix displayed on the right will be denoted by $-J$, so the equation says $EM = -MJ$. The minus-sign has a purely cosmetic function: it ensures that J itself represents the rotation through $+90°$ instead of $-90°$.

Turning back to our more general A, we note that $q < 0$ in Case (c), since the characteristic equation $(\lambda - r)^2 = q$ must be unsolvable. Hence $q = -s^2$ for some real number $s \neq 0$, and the modified Cayley- Hamilton relation Eq.(2.4) can be changed to

$$(A - rI)^2 = -s^2 I\,, \quad \text{or} \quad \big(s^{-1}(A - rI)\big)^2 = -I\,.$$

The road ahead is now clear: the last equation will furnish an E such that $E^2 = -I$, which will then be conjugated to $M^{-1}EM = -J$ as shown by

Eq.(2.5). The relation between A and E will turn this into a conjugation of A into a standard form. Instead of getting bogged down in petty detail — for which cf. Example 9 below — let us record the result. For any non-zero vector Y, we obtain

$$M^{-1}AM = rI + sJ = \begin{bmatrix} r & -s \\ s & r \end{bmatrix}, \quad \text{if } M = \big[(A - rI)Y, -sY\big].$$

For ease of calculation, we usually take Y to be one of the columns of I, so that $(A - rI)Y$ is just the corresponding column of $A - rI$.

Matrices of the form $rI + sJ$ are related to a rotations $R(\theta)$ as follows. For $\rho = \sqrt{r^2 + s^2}$, the numbers r/ρ and s/ρ have the property that their squares add up to 1. Hence they can be interpreted as the coordinates of a point on the unit circle, i.e., sine and cosine of a certain angle θ. Therefore

$$\begin{bmatrix} r & -s \\ s & r \end{bmatrix} = \rho \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$= \rho R(\theta)$ where θ is defined by $r/\rho = \cos\theta$ and $s/\rho = \sin\theta$. This not only gives a geometric interpretation to these matrices, it also has computational consequences — as will be seen in the next chapter.

It is time for a numerical example. We choose an A which is numerically close to the example of the last section, keeping a, c, d as before but setting $b = -0.2$. Again we have $r = 1.2$, but now q is negative: in fact, $4q = (0.2)^2 + 4(-0.2)(0.1) = -0.04$, whence $q = -0.01$. Taking $s = \sqrt{0.01} = 0.1$ and $Y = I^{(1)}$, we wind up with

$$A = \begin{bmatrix} 1.3 & -0.2 \\ 0.1 & 1.1 \end{bmatrix}, \; M = \begin{bmatrix} 0.1 & -0.1 \\ 0.1 & 0 \end{bmatrix}, \; M^{-1}AM = \begin{bmatrix} 1.2 & -0.1 \\ 0.1 & 1.2 \end{bmatrix},$$

which is $\rho R(\theta)$, where $\rho = \sqrt{1.45} \approx 1.204$ and $\theta = \tan^{-1} 0.1/1.2 \approx 4.76°$.

Remark. The big difference between the scenario of the present section ($q < 0$) and that of Section 1 ($q > 0$) is the fact that, in the real number system, negatives have no square roots. This not so in the so-called *complex* number system, where everything has a square root.

If all this algebra were done with complex numbers, any $q \neq 0$ would lead to distinct eigenvalues $\lambda_1 \neq \lambda_2$ and hence to a diagonalization $A = MDM^{-1}$ with *complex* matrices M and D. No change would be needed in the formal computations, as they depend only on the feasibility of certain algebraic operations (namely, $+, -, \times, \div, \sqrt{\ }$) for the scalars used. Complex numbers will be introduced and explored in the appendix below. However, this book will avoid their systematic use as scalars.

§4. Examples and Exercises

Worked Examples

The first three examples will deal with the following matrices:

$$A_1 = \begin{bmatrix} 6 & -2 \\ 2 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 5 & -2 \\ 2 & 1 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 5 & -3 \\ 3 & 1 \end{bmatrix} .$$

For each of these, the exercise consists in writing down the characteristic equation, deciding which of the cases (a) - (c) the matrix A_k belongs to, determining eigenvalues and eigenvectors (if any), and finding an invertible matrix M such that $B_k = M^{-1}A_kM$ is of standard form. Note: the given matrices are numerically fairly close. You will find, however, that their standard forms belong to very different types.

1. *For A_1, the characteristic equation is*

$$\lambda^2 - 7\lambda + 10 = 0 \qquad \textit{or} \qquad (\lambda - 7/2)^2 = 9/4.$$

Since the right hand quantity is positive, we are in Case (a): A_1 has two eigenvalues, namely 3.5 ± 1.5, i. e. 5 and 2. To find corresponding eigenvectors, we solve the equations $(A_1 - \lambda I)X = 0$. These are

$$\begin{bmatrix} 1 & -2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \textit{and} \qquad \begin{bmatrix} 4 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for $\lambda = 5$ and $\lambda = 2$, respectively. The first of these amounts to saying that $x - 2y = 0$ (for instance $x = 2$, $y = 1$), the second one says that $2x - y = 0$ (for instance $x = 1$, $y = 2$). Thus we get two independent eigenvectors with coordinates $(2, 1)$ and $(1, 2)$, for $\lambda = 5$ and $\lambda = 2$, respectively. Using these as columns of the matrix

$$M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad \textit{we get} \qquad B_1 = M^{-1}A_1M = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} .$$

You should check this last calculation. What would happen to B_1 if you switched the columns of M?

2. *For A_2, the characteristic equation is*

$$\lambda^2 - 6\lambda + 9 = 0 \qquad or \qquad (\lambda - 3)^2 = 0.$$

Since $q = 0$, we are in Case (b): A_2 has the single eigenvalue $\lambda = 3$. To find a corresponding eigenvector, we solve the equation $(A_2 - 3I)V = 0$, i.e.,

$$\begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \qquad getting \qquad V = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for instance. To construct M, we follow the recipe given in Section 2 above: since $I^{(1)}$ is not an eigenvector, we use it as the second column of M, taking $(A_2 - 3I)Y$ for the first column. Thus,

$$M = \begin{bmatrix} 2 & 1 \\ 2 & 0 \end{bmatrix} \qquad and \qquad B_2 = M^{-1}A_2M = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} .$$

Again you should check that this works out. Note that the first column of M is an eigenvector. Can you explain this? Try to replace it by another eigenvector (say, V) and see what you get. Can you account for that?

3. *For A_3, the characteristic equation is*

$$\lambda^2 - 6\lambda + 14 = 0 \qquad or \qquad (\lambda - 3)^2 = -5.$$

Obviously we are in Case (c): A_3 has no real eigenvalues. To construct M, we follow the recipe of Section 3: take $(A_3 - 3I)Y$ for the first column, but this time make the second one equal to $-sY$. Here s can be $\pm\sqrt{5}$, and Y can be any non-zero vector. If we choose $s = \sqrt{5}$ and $Y = I^{(1)}$, we get

$$M = \begin{bmatrix} 2 & -\sqrt{5} \\ 3 & 0 \end{bmatrix} \qquad and \qquad B_3 = M^{-1}A_3M = \begin{bmatrix} 3 & -\sqrt{5} \\ \sqrt{5} & 3 \end{bmatrix}$$

$= \rho R(\theta)$, where $\rho = \sqrt{14} \approx 3.74$ and $\theta = \tan^{-1} \sqrt{5}/3 \approx 36.7°$.

What would you get if you chose $s = -\sqrt{5}$? What would happen if you took another Y (say, $I^{(2)}$)?

It is usually more interesting and instructive to do the same example in several ways than to do several examples all in the same way!

4. Without using determinants, show that

$$A = \begin{bmatrix} s+4 & 2s+2 \\ t-1 & 2t+1 \end{bmatrix}$$

always has the eigenvalue 3. Find the corresponding eigenvector.

Just check that $A-3I$ *is singular: in the present case, the second column equals twice the first. Hence, if we multiply the first column of* $A-3I$ *by* 2 *and then subtract the second, we get zero. Therefore* $(A-3I)V=0$, *if* V *has the coordianates* $(2,-1)$. *Note that this eigenvector does not depend on the values of* s *and* t.

5. What conditions on s and t ensure that A has a single eigenline in the preceding example?

Instead of computing the characteristic equation $\det(A-\lambda I)=0$, *a messy affair, we rule out a second eigenvalue* ($\neq 3$) *by demanding* $\operatorname{tr} A = 3+3$, *i.e.,* $s+2t=1$. *To avoid* $A=3I$, *we also make sure that the off-diagonal entries are not zero, i.e., that* $t\neq 1$ *and* $s\neq -1$. *Altogether:* $s=1-2t$ *with* $t\neq 1$.

6. The similarity relation $B=M^{-1}AM$ clearly implies $(sI+tB)=M^{-1}(sI+tA)M$ for any pair of scalars s and t. Let us compare other features of A and $sI+tA$.

(a) Given an eigenvector V of A, with eigenvalue λ, find the corresponding eigenvector and eigenvalue of $sI+tA$.
(b) For the matrices A_1, A_2, A_3 given above, determine the standard forms of $A'_k = I + 0.1A_k$.

(a) *The corresponding eigenvector is* V *itself:* $AV=\lambda V$ *implies* $(sI+tA)V = sV+t\lambda V = (s+t\lambda)V$, *so the corresponding eigenvalue is* $s+t\lambda$.

(b) *In the first two cases, the standard form depends only on the eigenvalues — which are* 1.5, 1.2 *for* A_1 *and* 1.3 *for* A_2. *In the third case,* $I+0.1B_3$ *clearly does have the standard form* $s'I+t'J$ *and is similar to* $I+0.1A_3$ *(this works analogously for the first case). Thus we get*

$$B'_1=\begin{bmatrix}1.5 & 0\\ 0 & 1.2\end{bmatrix},\quad B'_2=\begin{bmatrix}1.3 & 1\\ 0 & 1.3\end{bmatrix},\quad B'_3=\begin{bmatrix}1.3 & -\sqrt{.05}\\ \sqrt{.05} & 1.3\end{bmatrix}.$$

The matrices M which conjugate A_k to B_k, will also turn A'_k into $I+0.1B_k$ — which equals B'_k for $k=1$, 3. In contrast, $I+0.1B_2$ has the off-diagonal entry $u=0.1$, hence is *not* the standard form B'_2.

7. For any $u\neq 0$, find a diagonal matrix D such that

$$D^{-1}\begin{bmatrix}s & u\\ 0 & t\end{bmatrix}D=\begin{bmatrix}s & 1\\ 0 & t\end{bmatrix}.$$

In order for $D=D(\alpha,\beta)$ *to be invertible, the scalars* α *and* β *must both be* $\neq 0$. *Since left multiplication by* D *(or* D^{-1}*) scales the rows, while right multiplication scales the columns, we always have*

$$\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}^{-1} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} = \begin{bmatrix} a & \frac{\beta}{\alpha} b \\ \frac{\alpha}{\beta} c & d \end{bmatrix}.$$

Note: the diagonal entries remain unchanged, while the others are multiplied or divided by $z = \alpha/\beta$. In an upper triangular matrix only the upper right corner changes: it is divided by z. Hence we can solve our problem with any $D(\alpha, \beta)$ such that $\alpha/\beta = u$, for instance $D(u, 1)$ or $D(1, u)^{-1}$.

8. If $M = [V, W]$ is invertible with $AV = sV$, we have seen in Section 2, that $M^{-1}AM$ has the form

$$B = \begin{bmatrix} s & u \\ 0 & t \end{bmatrix}.$$

Without explicitly computing $M^{-1}AM$, find the off-diagonal entry u.

Since the second column of $B - tI$ obviously equals $uI^{(1)}$, the second column of the equation $M^{-1}(A - tI)M = B - tI$ reads $M^{-1}(A - tI)W = uI^{(1)}$, or $(A - tI)W = uV$. In particular, $u = 1$ if it happens that $V = (A - tI)W$, as in our recipe for the standard form in Case (b) with $s = t = r$.

9. Supposing that $(A - rI)^2 = -s^2 I$, let $M = [V, W]$ with $V = (A - rI)Y$ and $W = -sY$, where $Y \neq 0$ is chosen arbitrarily. Compute $M^{-1}AM$

Putting $E = -s^{-1}(A - rI)$, we have $V = (A - rI)Y = EW$. Since this choice makes $E^2 = -I$, we can use Eq.(2.5) to conclude that $EM = -MJ$, which means $(A - rI)M = sMJ$, or $M^{-1}(A - rI)M = sJ$, whence $M^{-1}AM = rI + sJ$.

Exercises

1. Find eigenvalues and eigenvectors for $A = \begin{bmatrix} 1.09 & -.04 \\ -.03 & 1.05 \end{bmatrix}$. .
2. Ditto for $A = \begin{bmatrix} 1.09 & .02 \\ -.02 & 1.05 \end{bmatrix}$.
3. Find A whose eigenlines are given by $(2, -3)$ and $(-3, 4)$ with eigenvalues 1 and -1, respectively.
4. Find A whose eigenlines are given by $(2, -3)$ and $(-3, 4)$ with eigenvalues -1 and 1, respectively.
5. For what values of k is the matrix $\begin{bmatrix} 1 & 2 \\ k & 3 \end{bmatrix}$ diagonalizable? Explain.
6. For what values of k does the matrix $\begin{bmatrix} 1 & 2 \\ k & 3 \end{bmatrix}$ have a negative eigenvalue? Explain.

7. Without using determinants, show that the matrix $A = \begin{bmatrix} p & q \\ 1-p & 1-q \end{bmatrix}$ always has the eigenvalue 1. Find a corresponding eigenvector.
8. Without computing determinants, show that the matrix A of (7) always has the eigenvector $(-1, 1)$. Find the corresponding eigenvalue.
9. For $A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$ find a vector $V = \begin{bmatrix} u \\ v \end{bmatrix}$, with $u + v = 1$, which remains unchanged when multiplied by A.
10. In Hicktown, everyone regularly drinks either Coke or Pepsi. 20% of last year's Coke drinkers switched to Pepsi this year, while 30% of former Pepsi fans changed to Coke. If the total percentage of Coke guzzlers has remained unchanged, how high is it? Explain.

For each of the following matrices A find an invertible M such that $B = M^{-1}AM$ is in one of the standard forms. Record AM, M^{-1}, and B.

11. $A = \begin{bmatrix} -2 & -2 \\ -5 & 1 \end{bmatrix}$ 12. $A = \begin{bmatrix} 3 & -5 \\ 1 & -1 \end{bmatrix}$ 13. $A = \begin{bmatrix} -12 & 7 \\ -7 & 2 \end{bmatrix}$

14. $A = \begin{bmatrix} 9 & 2 \\ -2 & 5 \end{bmatrix}$ 15. $A = \begin{bmatrix} 3 & 2 \\ -5 & 1 \end{bmatrix}$ 16. $A = \begin{bmatrix} -7 & -5 \\ 4 & 5 \end{bmatrix}$

17. $A = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix}$ 18. $A = \begin{bmatrix} -4 & 13 \\ -5 & 10 \end{bmatrix}$

For each of the statements (19) – (24) give a proof or a counterexample. The notation $A \sim B$ means that A and B are similar.

19. If A or B is invertible, then $AB \sim BA$.
20. If A and B are singular, then $AB \sim BA$.
21. If $A \sim B$ and $A' \sim B'$, then $A + A' \sim B + B'$.
22. If $A \sim B$ and $A' \sim B'$, then $AA' \sim BB'$.
23. If $bc > 0$, the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is diagonalizable.
24. If $bc < 0$, the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is not diagonalizable.
25. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a + c = b + d = k$. Taking a cue from problems (7) and (8), find eigenvalues and eigenvectors.

Appendix: Complex Numbers

1. Genesis

Mathematical evolution has repeatedly expanded the notion of *number* in order to overcome arithmetic restrictions. For instance, since the system **N** of *natural* numbers $1, 2, 3, 4 \ldots$ does not allow subtraction $a - b$ unless $b < a$, it is augmented by 0 and the negatives to form the *integers* **Z**. To permit unrestricted division (except by 0), the latter in turn is enlarged to the *rationals* **Q**, which encompass all fractions or "ratios" of integers.

Q is fine for elementary arithmetic but cannot solve simple equations like $x^2 = 2$, let alone $10^x = 3$. Hence it is extended to the *reals* **R**, which include such "irrationals" as $\sqrt{2}$ or $\log 3$. Real numbers are designed to measure real *magnitudes*. They can be depicted as points on an infinite "number line", and are often presented as "decimal expansions that go on and on indefinitely." While this is not exactly a crisp definition, it underlines the fact that every real number can be approximated by rationals to any desired degree of precision, so that any problem solvable in **R** is approximately solvable in **Q**.

R is the work-horse of undergraduate science, including this text. However, large parts of mathematics (and physics) would be seriously hobbled without recourse to the system **C** of *complex* numbers, which parametrizes a plane instead of a line, and has the virtue of providing square roots (and more) for all its inhabitants. The passage from **R** to **C** is technically far simpler than the one from **Q** to **R** — but may at first seem mysterious, because something like $\sqrt{-5}$ does not even make *approximate* sense in **R**. Although the systematic use of **C** in this text would streamline the algebra, we shall refrain from it because it would unduly complicate the geometry.

The journey from **N** to **R** took thousands of years and included long periods of confusion, which left us with such giddy terms as "irrational", "transcendental", "imaginary", etc. Every time a concept is expanded, its original meaning must be replaced by a more subtle one — which invariably raises the question of whether it all makes sense. If you are accustomed to counting sheep or stars, the idea of a "negative number" may be puzzling indeed. To make room for such a notion, the original counting numbers from **N** have to be reinterpreted as *increments* rather than absolute *quantities*. Then they can interact with *decrements* in a new happy family **Z**.

To prepare them for insertion into **C**, the familiar reals will likewise be reinterpreted, this time as matrices. Instead of seeing a real number u as a magnitude, think of it as the scalar matrix uI. This makes no difference at all to the customary arithmetical or analytical operations with these objects, but it does allow them to interact with the so-called "imaginary numbers"— which are matrices of the form vJ. Since $(vJ)^2$ equals $-v^2I$, imaginary numbers are square roots of negative real numbers (taken as scalar matrices). A *complex* number is just the sum of a real and an imaginary number, in other words, a matrix of the form $uI + vJ$.

Unfortunately, traditional notation is not as straightforward. Favouring brevity, mathematicians tend to omit any bells and whistles a number may have acquired during its evolution. For instance, we write 5 instead of +5, 5/1, or $5.000\cdots$, whether that number appears in the context of **N**, **Z**, **Q**, or **R** — and we wish to write 5 intead of $5I$ in the context of **C** as well. As to the imaginary part: if it cannot be a matrix, it will just have to sit there as an extraneous appendage. In fact, the complex number $5I + 3J$ is traditionally written as

$$5 + 3\sqrt{-1}, \qquad 5 + 3i, \qquad \text{or} \quad 5 + 3j$$

depending on the mathematical subculture in which it appears. We would prefer the last of these options (used by electrical engineers) but shall regretfully choose the second one, because it is the most popular.

In this terse notation, complex numbers appear as expressions $u + vi$, where u, v are actual real numbers but +, i are just punctuation marks with i behaving like $\sqrt{-1}$. The best way to handle them is to treat the symbol i exactly like the symbol x in computations with polynomials, and to simplify all results by substituting -1 for i^2 wherever possible. If anyone should challenge you to explain what it all means, just take him aside and tell him about the matrices $uI + vJ$, of which the mysterious $u + vi$ is but an abbreviation.

2. Properties

Addition of such "numbers" $w = u + vi$ and $w' = u' + v'i$ goes as expected:

$$(u + vi) + (u' + v'i) = [u + u'] + [v + v']\, i\,,$$

but complex multiplication follows the more intricate pattern

$$(*) \qquad (u + vi) \cdot (u' + v'i) = [uu' - vv'] + [uv' + u'v]\, i$$

because of the rule $i^2 = -1$. It is evident that $w \cdot w' = w' \cdot w$, i.e., complex multiplication is *commutative*. Moreover, since $\det(uI + vJ) = u^2 + v^2$ is positive unless $u = v = 0$, every complex number $w \neq 0$ is invertible, and commutativity allows us to write w'/w instead of $w^{-1}w' = w'w^{-1}$. In **C**, we therefore have unrestricted addition, subtraction, multiplication, and division (by non-zero elements) with all the familiar rules of arithmetic. What more could we ask from a number system? Oh yes, there is no *order*, no dichotomy between positive and negative.

But there *is* a notion of size: the *absolute value* $|w|$ of a complex number $w = uI + vJ$ is defined to be $\sqrt{u^2 + v^2}$. Since this is just the square root of $\det(uI + vJ)$, it is compatible with multiplication and division, i.e.,

$$|w \cdot w'| = |w| \cdot |w'| \quad \text{and} \quad |w^{-1}| = |w|^{-1} .$$

With respect to addition, however, it satisfies the more delicate relation

$$|w + w'| \leq |w| + |w'| , \tag{†}$$

known as the "Triangle Inequality". To prove it, we multiply both sides by the positive real number $|w|^{-1}$ (the case $w = 0$ being trivial) and obtain $|1+c| \leq 1+|c|$, with $c = w'/w$, as the thing to check. An inequality between non-negative reals is neither created nor destroyed by squaring both sides, so the question is whether $|1 + c|^2 \leq 1 + 2|c| + |c|^2$ or not. For $c = a + bi$, the left hand side of this is $(1 + a)^2 + b^2 = 1 + 2a + a^2 + b^2$, while the right comes to $1 + 2|c| + a^2 + b^2$. Therefore it all boils down to $a \leq |c|$, which is certainly true — and is a strict inequality unless $c = a$ is a non-negative real number.

3. Geometry

In some sense, a complex number $w = u + vi$ is just a pair (u, v) of real numbers, and as such can be depicted as a point or a vector in the plane. Addition is just the usual composition of vectors. To see the geometric manifestation of multiplication, however, it is good to remember that w actually stands for the matrix $uI + vJ =$

$$(**) \qquad \begin{bmatrix} u & -v \\ v & u \end{bmatrix} = \rho \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$= \rho R(\theta)$, where ρ, θ are the polar coordinates of the point (u, v). In these terms, it is easy to see how multiplication by w affects a complex number $z = x + yi$. After all, the vector (x, y) is just the first column of $xI + yJ$, and analogously the coordinates of $w \cdot z$ form the first column of the matrix $\rho R(\theta) \cdot (xI + yJ)$ — i.e., (x, y) transformed by $\rho R(\theta)$.

Conclusion: *complex multiplication by w means matrix multiplication by $uI + vJ$, i.e. rotating by θ and scaling by ρ.*

The scaling factor $\rho = \sqrt{u^2 + v^2}$ is what we had previously called the absolute value $|w|$. By the Theorem of Pythagoras, it corresponds to the *length* of the vector (u, v), which explains the term "Triangle Inequality" used above: if you picture the sum of w and w' as one arrow following another, $|w + w'|$ is the length of the third side in the triangle so obtained. Its length is at most equal to the sum of the other two lengths, with equality if and only if one of them is zero or w and w' point in the same direction (i.e., w'/w is a positive real number).

In $\mathbf{C}$ it is easy to find any n-th root, i.e. solve the equation $z^n = w$ for any given $w \in \mathbf{C}$ and $n \in \mathbf{N}$. Because of Eq.(**), you just have to choose $z = \sigma(\cos\phi + i \sin\phi)$ where $\sigma^n = \rho$ and $n \cdot \phi = \theta$, so that the underlying matrix $\sigma R(\phi)$ has it n-th power equal to $\rho R(\theta)$. Since angles obey the special rule $360° = 0°$, we actually get n different solutions by setting $\phi_k = (\theta + k \cdot 360°)/n$ for $k = 0, 1, \ldots, (n-1)$.

Working in **C**, we would never be stopped from solving a quadratic equation because of a missing square root. Forgetting — or temporarily ignoring — the matrix aspect of complex numbers, we could turn around and admit them as entries in our matrices and vectors, as a new kind of "scalar". To illustrate how this facilitates certain computations, let us look back at Case (c) in Section 3. The complex eigenvalue $\lambda_1 = r - si$ would lead to a singular complex matrix $A - \lambda_1 I$ and hence to a non-zero complex solution of $(A - \lambda_1 I)X = 0$, i.e. a complex eigenvector X. The latter can be written as $X = U + Wi$, where the vectors U and V have *real* coordinates. Now $AX = \lambda_1 X$ amounts to $AU + AWi = (rU + sW) + (-sU + rW)i$, which separates into $AU = rU + sW$ and $AW = -sU + rW$, and finally boils down to $AM = BM$, where M has columns U, V and B is as in Section 3.

If the complex plane is so great for doing algebra, why hesitate to move into it for good? The reasons are psychological and geometric: before we play with matrices whose entries can themselves be considered matrices, before we plunge into the Cartesian "plane" over **C** — whose points (z, w) have 2 complex, and hence 4 real, coordinates — we ought to develop a firm grasp of matrices and vectors in the familiar world of **R**, where scalars only scale, and not rotate to boot.

4. Polynomials

In **C** every polynomial equation $f(z) = 0$ is solvable — except, of course, when $f(z)$ is a non-zero constant. Though sometimes taken on faith, this *"Fundamental Theorem of Algebra"* is not exactly plausible. Why should $\sqrt{-1}$ open the gate for *all* polynomials? Our answer will focus on the continuous real-valued function $|f(z)|$ and use the fact that, restricted to any closed planar disc, such a function must reach a minimum. Intuitively this seems obvious. Just imagine the graph of that function stretched over the disc like a circus tent over an arena: if there are no fuzzy edges, there must be a point of lowest height.

For the rest of this dicussion, we consider a polynomial

$$f(z) = c_0 + c_l z^l + \cdots + c_h z^h, \qquad (\ddagger)$$

with non-zero complex coefficients c_0, c_l, c_h and the powers of z arranged in ascending order. Our argument will be based on the following two properties of its absolute value $|f(z)|$:

(i) $|f(z)| > |f(0)|$ for all sufficiently large values of z, and
(ii) $|f(z)| < |f(0)|$ for some very small values of z.

These have nothing to do with fancy concepts like continuity. They result from some diligent digging in **C**, which will be performed at the end of this appendix. For the moment, let us take them for granted and see how they imply the theorem. Note that (ii) depends vitally on the assumption that $f(0) = c_0 \neq 0$, without which there would be nothing to prove, anyway. Here goes.

The absolute value $|f(z)|$ attains a minimum. Being a continuous function, $|f(z)|$ takes on a minimum value $m(r)$ as z ranges over a closed disc of radius r, so $m(r) \leq |f(z)|$ for $|z| \leq r$. If r is chosen large enough, (i) guarantees that $m(r)$, which is automatically $\leq |f(0)|$, will be $\leq |f(z)|$ for $|z| > r$ as well. Therefore $m = m(r)$ is a global minimum, and there is a place $a \in \mathbf{C}$ where $|f(a)| = m$.

This minimum cannot be positive. $f(z)$ clearly has the same range of values as the polynomial $g(z) = f(a - z)$. In particular, the minimum of $|g(z)|$ is also equal to $m = |f(a)| = |g(0)|$. If m were not zero, (ii) could be applied to $g(z)$ and would deny the minimality of $|g(0)|$. Hence we conclude that $g(0) = f(a) = 0$, as was to be shown.

It remains to establish the properties (i) and (ii) used in this argument. Consider $f(z)$ written as

$$f(z) = \varphi(z) + c_h z^h \qquad \text{or} \qquad f(z) = c_0 + c_l z^l + \psi(z)\,,$$

and note that $\varphi(z)$ and $\psi(z)$ are negligible for very large z and very small z, respectively. In fact, if K is a real number greater than the sum of the absolute values $|c_0| + |c_l| + \cdots + |c_h|$, a routine (multiple) application of the Triangle Inequality shows that $|\varphi(z)| < K|z|^{h-1}$ for $|z| \geq 1$, and $|\psi(z)| < K|z|^{l+1}$ for $|z| \leq 1$. This exercise is left to the reader.

The reason behind (i) is simply that, for very large values of z, the order of magnitude of $f(z)$ is dictated entirely by the term $c_h z^h$, which overpowers $K|z|^{h-1} > |\varphi(z)|$ and outgrows any given bound.

On the other hand, (ii) is due to the fact that, for very small values of z, the behaviour of $f(z)$ resembles that of $c_0 + c_l z^l$. For a closer look, let us first consider the special case $c_0 = 1$, $c_l = -1$. Here we shall be able to find a *real* number $0 < \varepsilon < 1$ such that $|f(\varepsilon)| < |f(0)| = 1$. In fact, choosing ε so small that $K\varepsilon < 1/2$, we get $|\psi(\varepsilon)| < K\varepsilon^{l+1} < \varepsilon^l/2$ and

$$|f(\varepsilon)| \leq |1 - \varepsilon^l| + |\psi(\varepsilon)| < \left(1 - \varepsilon^l\right) + \frac{\varepsilon^l}{2} < 1.$$

If c_0, c_1 do not have the special values used in this argument, we simply apply it to the polynomial $c_0^{-1} f(wz) = 1 + c_0^{-1} w^l z^l + \cdots$, where $w^l = -c_0 c_l$. There it produces an ε such that $|c_0^{-1} f(w\varepsilon)| < 1$, whence $|f(w\varepsilon)| < |c_0|$ as required.

This article was processed by the author using the TEX macro package from Springer-Verlag.