1 GSW… Eigenvalues and Eigenvectors

Matrices with only one row or only one column (called *row vectors* or *column vectors* respectively) can be used to represent points in space by containing the co-ordinates (relative to some given axes) of the point. For example, the column vector[1]:

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

can represent a point a distance $\sqrt{8}$ from the origin, with an angle of 45 degrees to the two positive axes, x_1 and x_2.

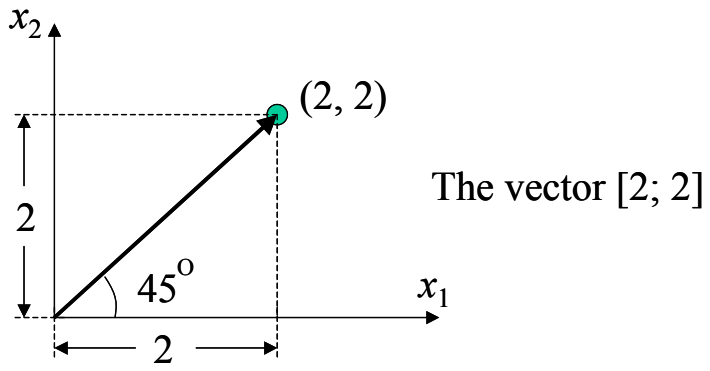

Figure 1-1 Geometric Representation of a Vector

The same idea applies to real and complex vectors of any number of dimensions (although any more than two real components, and the vector gets harder to draw).

A square matrix can be said to *act on* a vector, since multiplying a column vector by such a matrix will result in another column vector with the same number of elements, however one that will, in general, have a different direction and amplitude.

For example, consider the vector $\mathbf{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ as shown above, and the matrix $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$,

$$\mathbf{Ax} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix} \tag{0.1}$$

and this can represent a point with co-ordinates (8,2). Notice that in this case the original vector [2; 2] has changed both in amplitude and direction when pre-multiplied by the matrix **A**.

[1] I'll sometimes write column vectors in the form [2; 2] (note the semicolon between the elements) and row vectors as [2 2] (note the absence of a semicolon) since this is the easiest way to input them into MATLAB.

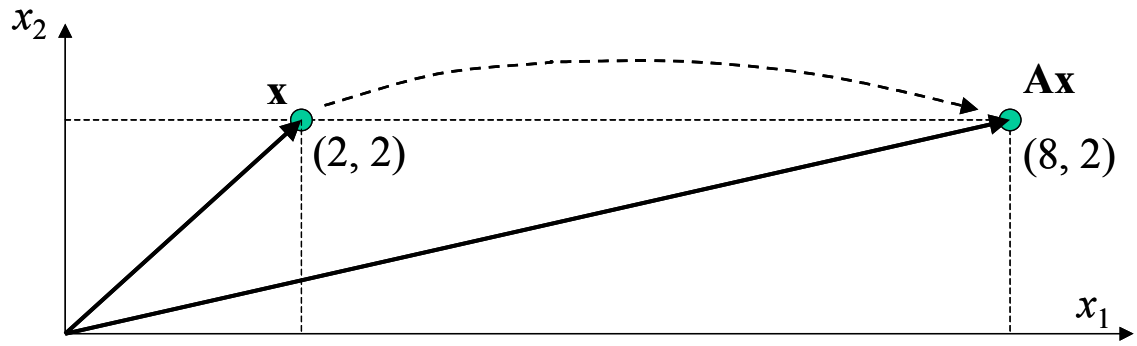

Figure 1-2 Action of a Matrix on a Vector

For any square matrices, there are some vectors that do not change their direction (except possibly to point in the opposite direction[2]) when pre-multiplied by the matrix. These vectors are known as the *eigenvectors* of the matrix. While the direction of an eigenvector does not change when pre-multiplied by a matrix (except to possibly reverse direction), the amplitude of the vector can change. The factor by which the amplitude of an eigenvector changes when pre-multiplied by the matrix is the *eigenvalue*. (Eigenvalues can be negative: it's these negative eigenvalues that result in eigenvectors reversing direction when pre-multiplied by the matrix.)

We can write:

$$\mathbf{Ax} = \lambda \mathbf{x} \tag{0.2}$$

when $\mathbf{x}$ is an eigenvector, and λ its eigenvalue.

Another useful fact: any multiple of an eigenvector is also an eigenvector:

$$\mathbf{A}\left(k\mathbf{x}\right) = k\mathbf{Ax} = k\lambda\mathbf{x} = \lambda\left(k\mathbf{x}\right) \tag{0.3}$$

It's the direction that's important for an eigenvector, not the amplitude: the amplitude can be anything. As a result, eigenvectors are sometimes quoted in a normalised form. A *normalised eigenvector* has a length of one, so that $\mathbf{x}^H\mathbf{x} = 1$.

For example, consider the matrix above, $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$. For this matrix, the two eigenvalues are 3 and –2, and they have eigenvectors are $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ respectively, since:

$$\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4+2 \\ 4-1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{0.4}$$

[2] Pointing in the opposite direction and pointing in the same direction with a negative length are the same thing in linear algebra.

$$\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2+4 \\ -2-2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix} = -2\begin{bmatrix} -1 \\ 2 \end{bmatrix} \tag{0.5}$$

The action of the matrix on these vectors is to change their lengths, but not change their direction. (With the usual caveat: the direction of the second eigenvector is reversed, but really it's in the same direction, it's just had its length multiplied by a negative number. It lies along the same straight line, it just points the other way).

To find the normalised eigenvectors, we just divide these eigenvectors by their lengths:

$$\frac{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{\sqrt{2^2+1^2}} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{0.6}$$

$$\frac{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}{\sqrt{1^2+2^2}} = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 2 \end{bmatrix} \tag{0.7}$$

1.1 Some Useful Facts about Eigenvalues

There are several useful theorems about eigenvalues and eigenvectors; what follows is a short list of some of the most useful ones.

1.1.1 How Many Eigenvalues are There?

In general, an *n*-by-*n* square matrix will have *n* eigenvalues, however, they might not all be different[3]. (However, if there are two eigenvectors with the same value, they will each have a different eigenvector.)

While it is possible to have two (or more) different eigenvectors with the same eigenvalue; it's impossible to have two different eigenvalues with the same eigenvector. If the eigenvalues are different, the eigenvectors must be different too (see the problems for a proof).

1.1.2 The Trace of a Matrix and the Sum of Eigenvalues

It's a remarkable fact that the trace of any matrix (the sum of all the terms on the main diagonal) is equal to the sum of the eigenvalues. For example, the by-now familiar matrix:

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

[3] This is one of those unhelpful statements mathematicians sometimes make. The point is that the solution to the equation that determines the eigenvalues of an *n*-by-*n* matrix is a n^{th} order polynomial, so it will have *n* roots. However, the roots might not be different: e.g. we usually say that $x^2 + 2x + 1 = 0$ has two roots, they're just both $x = -1$. The same sort of thing happens with eigenvalues.

has two eigenvalues, 3 and –2, with sum = 3 – 2 = 1. The sum of the terms on the main diagonal is 2 – 1 = 1. This is a simple way to find the last eigenvalue once all the others are known.

It's also a very useful way to put an upper limit on the value of the largest eigenvalue if you happen to know that all the eigenvalues are positive (e.g. by knowing that **A** is a positive definite matrix, see the next section): the trace of the matrix must be greater than the largest eigenvalue. This is a particularly useful result in some iterative techniques to solve the equations $\mathbf{y} = \mathbf{A}\mathbf{x}$, where it's useful to know a number bigger than the largest eigenvector[4].

1.1.3 Positive Definite Matrices and Real Positive Eigenvalues

A positive definite matrix[5] only has real positive eigenvalues. This one is easy to prove: consider the defining equation of an eigenvalue:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{0.8}$$

and pre-multiply by $\mathbf{x}^H$. That gives:

$$\mathbf{x}^H\mathbf{A}\mathbf{x} = \mathbf{x}^H\lambda\mathbf{x} = \lambda\left(\mathbf{x}^H\mathbf{x}\right) \tag{0.9}$$

Now $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is real and positive if **A** is positive definite (that's the definition of positive definite), and $\mathbf{x}^H\mathbf{x}$ is also real and positive (it's the square of the length of the vector), so the eigenvalues are the ratios of two real, positive numbers, which means they must be real and positive too.

Positive semi-definite matrices can have $\mathbf{x}^H\mathbf{A}\mathbf{x}$ equal to zero for non-zero vectors **x**, and that would imply that the eigenvalue can also be zero. However, the eigenvalues of these matrices can still never be negative.

1.1.4 Hermitian Matrices and Orthogonal Eigenvectors

Any two eigenvectors of a Hermitian[6] matrix with different eigenvalues are orthogonal.

This proof is only slightly harder that the last one. Consider two different eigenvalues λ and μ of a matrix **A** and their corresponding eigenvectors **x** and **y**, so that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad \mathbf{A}\mathbf{y} = \mu\mathbf{y} \tag{0.10}$$

Start by taking the complex transpose of the first equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, and post-multiply by **y**:

$$\mathbf{x}^H\mathbf{A}^H\mathbf{y} = \lambda\mathbf{x}^H\mathbf{y} \tag{0.11}$$

---

[4] See the chapter on Steepest Gradient Techniques for an example.

[5] Positive definite matrix: one for which for all vectors **x**, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is positive, unless **x** is the zero vector. Positive semi-definite matrix: one in which for all non-zero vectors **x**, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is not negative.

[6] Hermitian matrix: one equal to the transpose of its complex conjugate, so that $\mathbf{A}^* = \mathbf{A}^T$, or $\mathbf{A}^H = \mathbf{A}$.

then, take the second equation $\mathbf{Ay} = \mu\mathbf{y}$, and pre-multiply by $\mathbf{x}^H$:

$$\mathbf{x}^H\mathbf{Ay} = \mu\mathbf{x}^H\mathbf{y} \tag{0.12}$$

Now, if $\mathbf{A}$ is Hermitian, then $\mathbf{A}^H = \mathbf{A}$, so the left-hand sides of equation (0.12) and equation (0.11) are equal. Subtract these two equations, and we get:

$$0 = (\lambda - \mu)\mathbf{x}^H\mathbf{y} \tag{0.13}$$

If these eigenvalues have different values, then $\lambda \neq \mu$, which means $\mathbf{x}^H\mathbf{y}$ must be zero: and that's the definition of these eigenvectors being orthogonal.

1.1.5 Hermitian Matrices and Sets of Orthogonal Eigenvectors

A Hermitian matrix with n rows and n columns has a set of n mutually orthogonal eigenvectors that map n-dimensional space (i.e. any vector in n-dimensional space can be expressed in terms of a weighted sum of these eigenvectors).

If all of the eigenvalues of the matrix are different, then this property follows immediately from the last result: all the eigenvectors are orthogonal, so any n-by-n Hermitian matrix must have a set of n orthogonal eigenvectors. If you've got n orthogonal eigenvectors, they must map n-dimensional space: it's a set of axes.

However, if the matrix has two (or more) eigenvalues that are equal, then equation (0.13) suggests we could have two eigenvectors that are different but not orthogonal: and this is true, we can. Trivial example: consider the unit matrix, which doesn't change any vector at all. Every vector is therefore an eigenvector of the unit matrix, and all the eigenvalues of the unit matrix are one:

$$\mathbf{Ix} = \mathbf{x} = \lambda\mathbf{x} \qquad \Rightarrow \qquad \lambda = 1 \tag{0.14}$$

More generally, a matrix with just two equal eigenvalues has a plane of possible vectors all of which remain unchanged in direction by the operation of the matrix, and any vector in that plane is therefore an eigenvector. If a matrix has three equal eigenvalues, it has a 3-dimensional space of possible vectors all of which remain unchanged in direction. The space defined by the set of vectors that remain unchanged in direction when pre-multiplied by a matrix, but share the same eigenvalue, is known as an *eigenspace*.

In these cases we can always find a set of n orthogonal eigenvectors for an n-by-n Hermitian matrix by just choosing any set of orthogonal vectors from this eigenspace. (The set of orthogonal eigenvectors we choose is not unique: there are, for example, an infinite number of ways to pick two orthogonal vectors from a 2-dimensional plane.)

You might think this is obvious, and I'd tend to agree; but if not, then consider taking any two different eigenvectors from this eigenspace, $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are not orthogonal. Then, express the vector $\mathbf{y}$ in terms of the sum of a vector parallel to $\mathbf{x}$, and a vector perpendicular to $\mathbf{x}$, which I'll call $\mathbf{x}'$:

$$y = ax + bx' \tag{0.15}$$

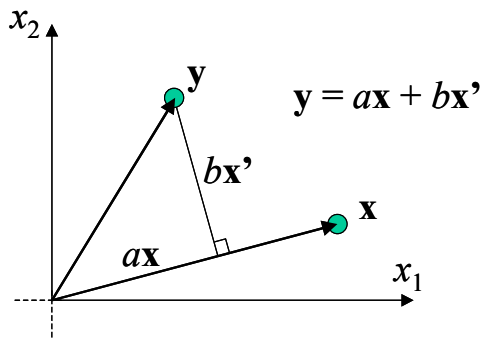

Figure 1-3 Expressing a Vector as the Sum of Two Orthogonal Eigenvectors

Then, the vector **x'** is an eigenvector of **A**, since:

$$\mathbf{A}\mathbf{x}' = \mathbf{A}\left(\frac{1}{b}\mathbf{y} - \frac{a}{b}\mathbf{x}\right) = \frac{1}{b}\mathbf{A}\mathbf{y} - \frac{a}{b}\mathbf{A}\mathbf{x} = \frac{1}{b}\lambda\mathbf{y} - \frac{a}{b}\lambda\mathbf{x} = \lambda\left(\frac{1}{b}\mathbf{y} - \frac{a}{b}\mathbf{x}\right) = \lambda\mathbf{x}' \qquad (0.16)$$

so **x** and **x'** are two orthogonal eigenvectors with the same eigenvalue. The same method can be extended to any number of dimensions. You can always find a set of orthogonal eigenvectors.

1.1.6 Non-Positive Definite Matrices

Some matrices don't have any real eigenvalues, even if all the terms in the matrix are real. For example, take the case of the matrix:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This represents a rotation through an anti-clockwise angle of θ. Any column vector multiplied by this matrix will be transformed into a column vector of the same amplitude, but pointing in a different direction.

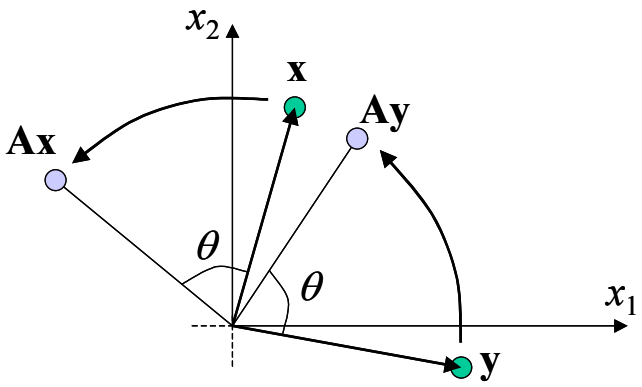

Figure 1-4 Action of a Rotation Matrix

Try and work out the eigenvalues for this case, and we get $\cos(\theta) + j\sin(\theta)$ and $\cos(\theta) - j\sin(\theta)$. Neither are real (although note that their sum is real, and equal to $2\cos(\theta)$, the trace of the matrix, as expected). The eigenvectors are $[1; j\,]$ and $[1; -j\,]$.

This isn't a surprise: there is no real direction in a two-dimensional plane that remains unchanged by this sort of rotation. However, it is interesting to note that if we allow the elements of the vectors to be complex, we can always find eigenvalues and eigenvectors for any matrix.

1.2 Expressing Vectors as the Sum of Eigenvectors

For the rest of this chapter, I'll restrict the discussion here to the most common and useful case: the Hermitian positive semi-definite matrix[7]. This means the eigenvalues are all real and positive, the eigenvectors with different eigenvalues are orthogonal, and there are a set of *n* orthogonal eigenvectors that cover the entire *n*-dimensional space.

This leads to one of the most common uses of eigenvalues: it is possible to express any vector in terms of the weighted sum of eigenvectors. If we can find a simple way to do this, then calculating the action of a matrix on a vector is straightforward: we just divide the vector into its components along the eigenvectors, multiply each component by the relevant eigenvalue, and then add all the components back up again. Simple. For example, consider the matrix **A** yet again:

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

which we know has eigenvectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Now consider the vector $\mathbf{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$. To work out **Ax**, we could express this vector as a weighed sum of the two eigenvectors:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{6}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{2}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} \tag{0.17}$$

and we can then work out the result of this matrix acting on **x** by multiplying the eigenvector components by their corresponding eigenvalues. Here we have six-fifths of the first eigenvector (which has an eigenvalue of 3); and two-fifths of the second eigenvector (with an eigenvalue of –2), so the output is:

$$\mathbf{Ax} = (3)\frac{6}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + (-2)\frac{2}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{18}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{4}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 40 \\ 10 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix} \tag{0.18}$$

---

[7] The reason this is a very common case is that these matrices result from trying to find the solution to the minimum square problem: finding the value of **x** that minimises $z = \|\mathbf{y} - \mathbf{Ax}\|^2$. Multiply this out, and differentiate with respect to the elements of **x**, and we find that the required value of **x** that minimises **z** satisfies the normal equation:

$$\mathbf{x} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{y}$$

Now, $\mathbf{A}^H\mathbf{A}$ is Hermitian, since $(\mathbf{A}^H\mathbf{A})^H = \mathbf{A}^H(\mathbf{A}^H)^H = \mathbf{A}^H\mathbf{A}$, and it's positive semi-definite, since $\mathbf{x}^H(\mathbf{A}^H\mathbf{A})\mathbf{x} = \mathbf{x}^H\mathbf{A}^H\mathbf{Ax} = (\mathbf{Ax})^H(\mathbf{Ax})$, and that's the square of the length of the vector **Ax**, which being a square can never be negative.

Of course, if you've read the chapter on "Counting the Cost" you might be wondering why anyone would ever want to do this. It only takes n^2 MAC operations to compute the product of an n-element vector and an n-by-n matrix, and that's far less than is required to work out the eigenvectors, their eigenvalues, and then work out how to express any arbitrary vector in terms of the sum of the eigenvectors.

Fair point: however, what if you wanted to calculate the result of using the same matrix to act on the same vector over and over again?

$$\mathbf{y} = \mathbf{AAAAAAAAAAAAAAAAAAAx} = \mathbf{A}^{20}\mathbf{x} \tag{0.19}$$

That takes a lot of multiplications to work out directly. However, if you happen to know the eigenvalues and eigenvectors of **A**, you can just work out the components of **x** along the eigenvectors, multiply each by the relevant eigenvalue raised to the power of 20, and add up the components again. Much easier. Would you ever want to do this? Well, in the case of an iterative algorithm (an algorithm which approaches the answer in a series of small steps by performing the same operation over and over again), possibly yes.

The only problem is that's it quite a difficult operation to work out the eigenvalues and eigenvectors for reasonable-sized matrices. Before I can talk about how to do this, I need to introduce the subject of matrix determinants.

1.3 Determinants

As we've seen, matrices operating on vectors can be thought of as changing the direction and length of the vectors. Since vectors can represent points in space, this corresponds to the matrix moving points in space: this is known as a *mapping*. If the original points lay on the surface of a volume (or the line enclosing an area in two dimensions), the mapping all these points will result in a new volume (or area). The ratio of the new volume (or area) to the original volume (or area) is a constant for all shapes, and is known as the determinant of the matrix. It's equal to the product of all the eigenvalues.

To prove this, consider a tiny, elemental volume (or area) before the mapping has taken place, with sides parallel to the eigenvectors[8], and can be represented as $a\mathbf{x}$ and $b\mathbf{y}$ where **x** and **y** are the eigenvectors of **A**.

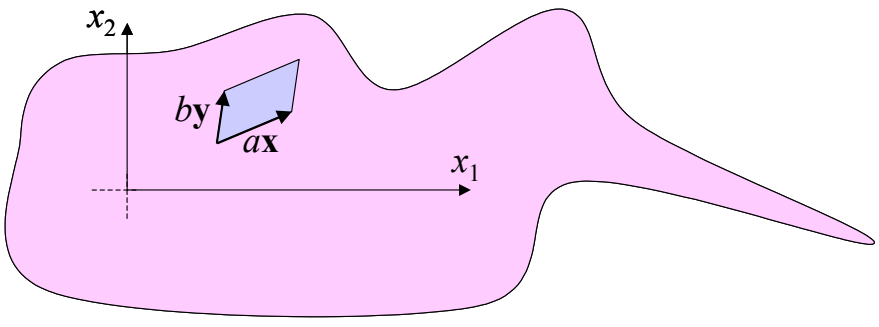

Figure 1-5 Arbitrary Area with Small Area Element

[8] For Hermitian matrices the eigenvectors will be orthogonal, but we don't need to assume that here.

(For convenience in drawing, I'll assume we've just got 2-element vectors representing points in a 2-dimensional plane, however the method can be easily extended to any number of dimensions.)

Provided a and b are small enough (and we can let them approach zero) the entire area can be expressed in terms of the sum of small areas of this shape. Consider mapping each small area element into the new space, and we'd transform this small shape into:

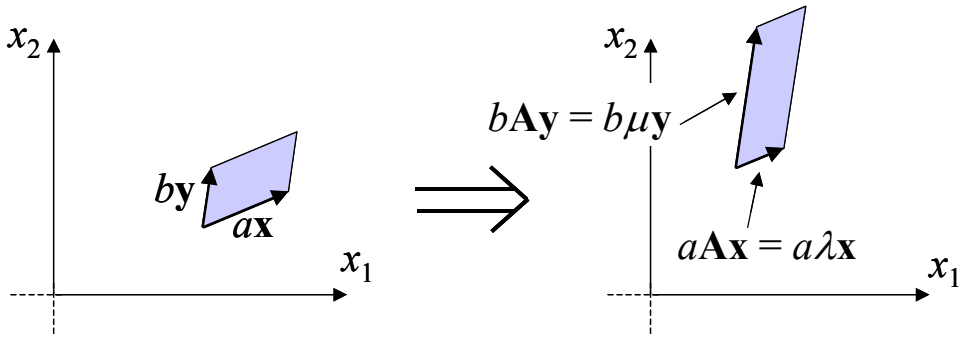

Figure 1-6 Mapping of Small Area Element

The area of the original parallelogram with sides $a\mathbf{x}$ and $b\mathbf{y}$ was:

$$\left(b|\mathbf{y}|\right)\left(a|\mathbf{x}|\right)\sin\left(\theta\right) \tag{0.20}$$

where θ is the angle between the two vectors $\mathbf{x}$ and $\mathbf{y}$. Since the two sides of the small area are eigenvectors, the angle between them doesn't change when they are both mapped by the matrix $\mathbf{A}$. However, the lengths do change: the length of the side $a\mathbf{x}$ becomes $a\mathbf{A}\mathbf{x} = a\lambda\mathbf{x}$ where λ is the eigenvalue of eigenvector $\mathbf{x}$, and the length of the side by becomes $b\mathbf{A}\mathbf{y} = b\mu\,\mathbf{y}$ where μ is the eigenvalue of eigenvector $\mathbf{y}$. Hence the area of this corresponding small element after the mapping is:

$$\left(b|\mathbf{A}\mathbf{y}|\right)\left(a|\mathbf{A}\mathbf{x}|\right)\sin\left(\theta\right) = \left(b\lambda|\mathbf{y}|\right)\left(a\mu|\mathbf{A}\mathbf{x}|\right)\sin\left(\theta\right) = \lambda\mu\left(b|\mathbf{y}|\right)\left(a|\mathbf{x}|\right)\sin\left(\theta\right) \tag{0.21}$$

which is $\lambda\mu$ times the area of the original small element. This is true for all the small areas that made up the original large shape: hence the area of the large shape has increased during the mapping by a factor of $\lambda\mu$, which is the product of the eigenvalues of the matrix $\mathbf{A}$.

1.3.1 Calculating Determinants

To calculate a determinant: multiply each element along the first row of the matrix by the determinant of the matrix composed of all the terms except the terms in the column or row of the… oh, this is getting far too complicated. I'll have to introduce the minor matrix.

1.3.1.1 The Minor Matrix

Every element of a matrix has an associated *minor* matrix. The minor matrix for any element is the entire contents of the original matrix, except the row and the column on which the element lies. For example, consider the following matrix, and the minor of the element $B_{2,3}$, which in this case, happens to be equal to two.

The minor contains all the elements of the matrix except those in the second row, and the third column. The minor of an n-by-m matrix always has $(n - 1)$ rows and $(m - 1)$ columns.

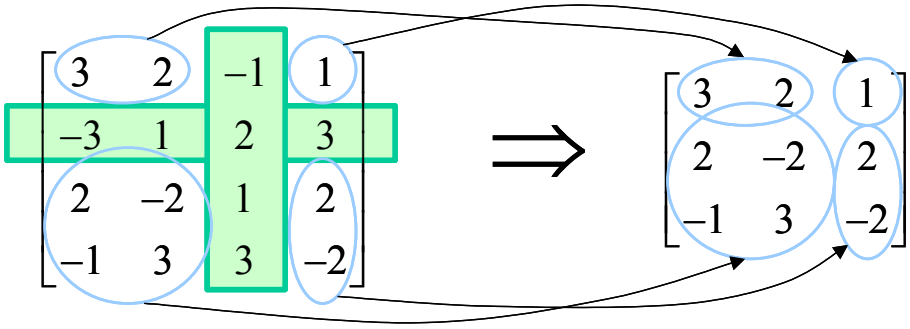

Figure 1-7 Example of a Minor Matrix of $\mathbf{A}_{2,3}$

Got the idea? Right – back to calculating determinants.

1.3.1.2 Back to Calculating Determinants

To calculate a determinant: multiply each element along the first row of the matrix by the determinant of its minor. Take the first of these products, then alternately subtract and add the next products along the row. The end result is the determinant of the original matrix.

Of course, all this does is allow us to express the determinant of a matrix in terms of the determinant of a lot of matrices (the minors) with one less row and column. How do we work out those? In terms of the determinants of matrices with one less row and column than they have. And so on, until you're trying to work out the determinant of a matrix with only one row and column: in other words, a scalar; and that's easy: the determinant of a scalar is equal to itself.

For example, consider calculating the determinant of a general 2-by-2 matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

The minor matrix for the first element a is the matrix of all terms not on the first row, and not in the first column. That's the matrix $[d]$, and it has a determinant of d. Similarly, the minor matrix for the second element b is just $[c]$. Alternately adding and subtracting the product of the terms and their minors along the first row gives:

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a(d) - b(c) = ad - bc \tag{0.22}$$

Extending the idea to a three-by-three matrix:

$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = a\left(\det\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b\left(\det\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c\left(\det\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right) \tag{0.23}$$

$$= a(ei - fh) - b(di - fg) + c(dh - eg)$$

and the determinants of larger matrices can be determined in a similar way, it just takes rather a long time for large matrices.

1.3.1.3 Some Important Things about Determinants

Any matrix with a non-zero determinant can be inverted, so having a non-zero determinant means the matrix is full rank (this also means that only square matrices have determinants,

since only square matrices can be inverted). Any matrix with a determinant of zero is *singular*, and cannot be inverted.

If a matrix **A** has a determinant of det(**A**), then the volume enclosed by any set of points **Ax** is det(**A**) times greater than the volume enclosed by the set of points **x**. Similarly, the volume enclosed by any set of points **BAx** is det(**BA**) times greater than the volume enclosed by the points **x**. But this is also det(**B**) times the volume enclosed by the points **Ax**, which is det(**A**) times greater than the original volume. This gives the important result that:

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \tag{0.24}$$

1.4 Determining Eigenvalues and Eigenvectors

Working out the eigenvalues and eigenvectors of a matrix is not a simple task, and takes a large number of individual calculations to do. For this reason, many practical algorithms use a different approach based on iteration. First, however, the direct method:

1.4.1 The Direct Method

Given a matrix **A**, the eigenvectors can be determined by noting that:

$$\mathbf{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{0.25}$$

which with a simple manipulation gives:

$$(\mathbf{A} - \lambda\mathbf{I})\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{0.26}$$

Let me write that equation in a slightly different way:

$$(\mathbf{A} - \lambda\mathbf{I})\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{0.27}$$

which makes it a bit more obvious that any eigenvector of **A** is also an eigenvector of the matrix (**A** – λ**I**), but now with an eigenvalue of zero. Think about what that means in terms of the determinant of (**A** – λ**I**): if a matrix has an eigenvalue of zero, then any n-dimensional space is going to map to an (n – 1)-dimensional space using this vector, since any component parallel to the eigenvector with a zero eigenvalue is going to end up as zero. A 2-dimensional plane, when acted on by such a matrix, will end up as a straight line; a 3-dimensional space will end up as a plane, and so on.

The determinant is the ratio of the volume of a space after the action of the matrix **A** to the volume of the original space. That means the determinant of this matrix (**A** – λ**I**) must be zero. 2-dimensional planes have a 3-dimensional volume of zero. Straight lines have an area of zero.

So, to find the eigenvalues of a matrix **A**, we need to find the values λ that make the determinant of the matrix (**A** – λ**I**) equal to zero. This leads to the *characteristic equation* of the matrix:

$$\det\left(\mathbf{A}-\lambda\mathbf{I}\right)=0 \tag{0.28}$$

solving which produces the eigenvalues. For example, consider the matrix **A** above yet again:

$$\begin{aligned}\det\left(\mathbf{A}-\lambda\mathbf{I}\right)&=\det\left(\begin{bmatrix}2-\lambda & 2\\ 2 & -1-\lambda\end{bmatrix}\right)=0\\ &=\left(2-\lambda\right)\left(-1-\lambda\right)-4\\ &=\lambda^2-\lambda-6\\ &=\left(\lambda-3\right)\left(\lambda+2\right)\end{aligned} \tag{0.29}$$

giving the possible eigenvalues 3 and –2.

Knowing the eigenvalues, the eigenvectors can be calculated by direct substitution, for example for the eigenvalue of three:

$$\begin{bmatrix}2-3 & 2\\ 2 & -1-3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}-1 & 2\\ 2 & -4\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=0 \tag{0.30}$$

$$\begin{aligned}-x_1+2x_2&=0\\ 2x_1-4x_2&=0\end{aligned} \tag{0.31}$$

You might notice that these two equations are not linearly independent. We can ignore one of them, set the value of one element of **x** to anything we like, and then solve the others. For example, let $x_1 = 1$, then $x_2 = 0.5$, giving the eigenvector:

$$\mathbf{x}=\begin{bmatrix}1\\ 0.5\end{bmatrix} \tag{0.32}$$

Finding all the eigenvalues and eigenvectors of a large matrix by this process is very time-consuming, and is another thing we usually try to avoid doing if at all possible.

1.4.2 Finding the Eigenvector with the Largest Eigenvalue by Iteration

In the case where there is one dominant eigenvector, a simple iterative procedure can find the largest eigenvector[9]. All you have to do is pick a suitable first guess for the eigenvector[10], and repeatedly pre-multiply this vector by the matrix.

The idea is that the component of the original guess vector parallel to the eigenvector with the largest eigenvalue will be multiplied by more than the components of the original guess vector

[9] Strictly speaking, this method finds the eigenvector with the greatest absolute eigenvalue, for example in the case of a matrix with one eigenvalue of –3 and another of +2, it would find the eigenvector corresponding to the eigenvalue of –3. However, since we're considering positive semi-definite matrices here, and they don't have any negative eigenvalues, I can just talk about finding the largest one.

[10] Any vector which cannot be expressed as a linear sum of all the other eigenvectors will do fine.

parallel to all the other eigenvectors, so it will end up bigger. Repeat this process a large number of times, and the end result is a vector that is almost parallel to the eigenvector with the largest eigenvalue. It's called the *power method*.

Perhaps an example might help, and we might as well stick to the same one:

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

Consider as a first guess the vector **x** = [1; 0]. We can express this vector as the sum of the two eigenvectors, as:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{2}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} \tag{0.33}$$

multiply this by the matrix **A**, and we get:

$$\mathbf{A}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{2}{5}\mathbf{A}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{5}\mathbf{A}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{6}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{2}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \tag{0.34}$$

multiply it by A again:

$$\mathbf{A}^2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{6}{5}\mathbf{A}\begin{bmatrix} 2 \\ 2 \end{bmatrix} + \frac{2}{5}\mathbf{A}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{18}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{4}{5}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix} \tag{0.35}$$

and so on. Plot these on a diagram, and we'll see that the angle between our guess and the eigenvector with the largest eigenvalue (in this case [2; 1]) is getting smaller each time.

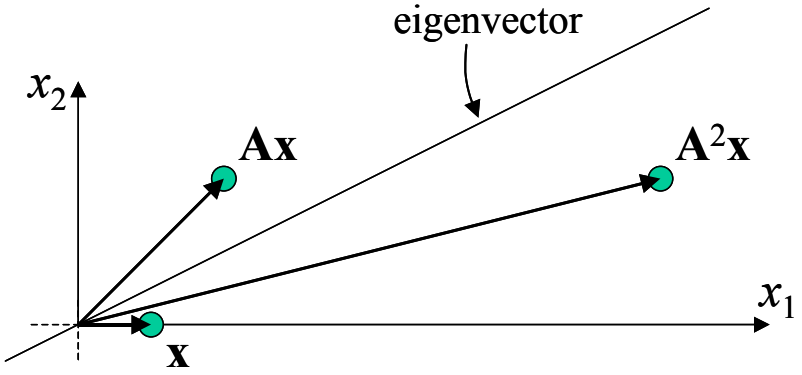

Figure 1-8 Iterating to the Largest Eigenvalue

The other thing we notice is that the vector is getting longer and longer all the time (as expected: the largest eigenvalue is greater than one). If we're trying to do this calculation in hardware, or with a limited number of bits to store each number, this isn't such a good idea, it would be better to limit the length of the vector. One way to do this is to normalise the length of the eigenvector at each step, so in effect the iteration is:

$$\mathbf{x} \rightarrow \frac{\mathbf{Ax}}{\|\mathbf{Ax}\|} \tag{0.36}$$

That, however, requires the calculation of a square root at every stage. Expensive. A less expensive way is to divide by the square of the length of the vector after every other stage; although in practice it's usually easier just to test if the vector has got too big, and if it has, divide all the elements by a power of two[11] (2, 4, 8, 16, etc).

Starting from $\mathbf{x} = [1\ 0]^T$, and repeating this iteration lots of times gives:

Iteration	$\mathbf{x}$	$\mathbf{Ax}$	Angle to eigenvector $\begin{bmatrix}2 & 1\end{bmatrix}^T$
1	$\begin{bmatrix}1 & 0\end{bmatrix}^T$	$\begin{bmatrix}2 & 2\end{bmatrix}^T$	18.4 degrees
2	$\begin{bmatrix}2 & 2\end{bmatrix}^T$	$\begin{bmatrix}8 & 2\end{bmatrix}^T$	-12.5 degrees
3	$\begin{bmatrix}8 & 2\end{bmatrix}^T$	$\begin{bmatrix}20 & 14\end{bmatrix}^T$	8.4 degrees
4	$\begin{bmatrix}20 & 14\end{bmatrix}^T$	$\begin{bmatrix}68 & 26\end{bmatrix}^T$	-5.6 degrees
5	$\begin{bmatrix}68 & 26\end{bmatrix}^T$	$\begin{bmatrix}188 & 110\end{bmatrix}^T$	3.8 degrees
6	$\begin{bmatrix}188 & 110\end{bmatrix}^T$	$\begin{bmatrix}596 & 266\end{bmatrix}^T$	-2.5 degrees
10	$\begin{bmatrix}15644 & 8078\end{bmatrix}^T$	$\begin{bmatrix}47444 & 23210\end{bmatrix}^T$	-0.5 degrees

Writing out the maths behind this method: consider a matrix $\mathbf{A}$, which has a set of eigenvectors $\mathbf{e_i}$. (Note $\mathbf{e_i}$ is not the i^{th} element of a vector $\mathbf{e}$, that would be written as e_i. This is $\mathbf{e_i}$, the i^{th} eigenvector of $\mathbf{A}$.) If the eigenvectors span space (which for the case of a Hermitian positive semi-definite matrix, they do), we can write any vector $\mathbf{x}$ as the weighted sum of the eigenvalues:

$$\mathbf{x} = \sum_i a_i \mathbf{e_i} \tag{0.37}$$

If the eigenvector $\mathbf{e_i}$ has eigenvalue λ_i, then the action of the matrix $\mathbf{A}$ on vector $\mathbf{x}$ is:

$$\mathbf{Ax} = \sum_i a_i \mathbf{A e_i} = \sum_i a_i \lambda_i \mathbf{e_i} \tag{0.38}$$

Repeating the process gives the vectors:

$$\mathbf{A}^n \mathbf{x} = \sum_i a_i \mathbf{A}^n \mathbf{e_i} = \sum_i a_i \lambda_i^n \mathbf{e_i} \tag{0.39}$$

To determine the direction of this vector, consider dividing all the elements by the constant factor λ_b^n, where λ_b is the largest eigenvalue:

[11] Dividing by a power of two is easy for computers and hardware that works in binary – you just rotate all the bits right. For example, divide 204 (11001100 in binary) by four, and the answer is 51 (110011 in binary).

$$\frac{\mathbf{A}^n\mathbf{x}}{\lambda_b{}^n} = \sum_i a_i \left(\frac{\lambda_i}{\lambda_b}\right)^n \mathbf{e_i} = a_b\mathbf{e_b} + \sum_{i\neq b} a_i \left(\frac{\lambda_i}{\lambda_b}\right)^n \mathbf{e_i} \tag{0.40}$$

Since λ_b is the largest eigenvalue, (λ_i / λ_b) is always less than one, so provided n is large enough, all these terms will tend towards zero, and can be neglected compared to $a_b\mathbf{e_b}$. The result is the eigenvector with the largest eigenvalue.

If we want the normalised eigenvector, we can just divide it by its length, and we can find the corresponding eigenvalue from the formula:

$$\mathbf{A}\mathbf{e_b} = \lambda_b\mathbf{e_b} \tag{0.41}$$

This method usually works much faster than the direct method. Only two problems really: the iteration can converge very slowly when there are two large eigenvectors that are almost equal (in this case one of the other eigenvalues has a value of (λ_i / λ_b) very close to one); and it only finds the eigenvector with the largest eigenvalue.

1.4.3 Determining the Other Eigenvectors

If we're working with a positive definite matrix, so we know the eigenvectors are all orthogonal, the simplest way to find another eigenvector is to subtract the component of the starting guess parallel to the known eigenvector at each stage of the iteration process. That leaves only the components of all the other eigenvectors, and repeating this process should provide the eigenvector with the second largest eigenvalue.

For example, consider the matrix:

$$\mathbf{A} = \begin{bmatrix} 6 & 0 & -2 \\ 0 & 2 & 2 \\ -2 & 2 & 3 \end{bmatrix} \tag{0.42}$$

This is Hermitian positive definite, so all eigenvalues will be real and positive, and all eigenvectors orthogonal (or can be chosen to be orthogonal). In this case, there is a particularly simple way to find out the components of a vector **x** that lie along the direction of an eigenvector $\mathbf{e}_1$: just calculate $(\mathbf{x}.\mathbf{e}_1)\mathbf{e}_1$ where $(\mathbf{x}.\mathbf{e}_1)$ is the inner (dot) product of the vectors **x** and $\mathbf{e}_1$.

In the case, we could start with a guess vector of $\mathbf{x} = [1\ 0\ 0]^T$, and run the iteration $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x}$ ten times, and we'd get an estimate of the first eigenvector (the one with the largest eigenvalue) of:

$$\mathbf{e}_1 \approx \mathbf{A}^{10}\mathbf{x} = \begin{bmatrix} 295123044 & -58687316 & -152123034 \end{bmatrix} \tag{0.43}$$

normalise this, and we get an estimate of the first eigenvalue:

$$\mathbf{e}_1 \approx \begin{bmatrix} 0.8464 & -0.1917 & -0.4969 \end{bmatrix} \tag{0.44}$$

To find the second largest eigenvalue, we take the first guess of $[1\ 0\ 0]^T$, and subtract the component of this guess along the eigenvalue we already know:

$$\mathbf{x} \rightarrow \mathbf{A}\left(\mathbf{x} - \left(\mathbf{x}.\mathbf{e}_1\right)\mathbf{e}_1\right) \tag{0.45}$$

repeat this ten times, and normalise the result, and we end up with an estimate of the second eigenvector of:

$$\mathbf{e}_2 \approx \begin{bmatrix} 0.4816 & 0.6720 & 0.5625 \end{bmatrix} \tag{0.46}$$

With an estimate of the first two eigenvectors, we can subtract the component of our guess along both of these eigenvectors, and that should leave us with a vector that tends towards the last eigenvector:

$$\mathbf{x} \to \mathbf{A}\left(\mathbf{x} - \left(\mathbf{x}.\mathbf{e}_1\right)\mathbf{e}_1 - \left(\mathbf{x}.\mathbf{e}_2\right)\mathbf{e}_2\right) \tag{0.47}$$

Set $\mathbf{x}$ back to $[1\ 0\ 0]^T$, and repeat this ten times and normalise the result, and we end up with:

$$\mathbf{e}_3 \approx \begin{bmatrix} 0.2261 & -0.7154 & -0.6611 \end{bmatrix} \tag{0.48}$$

The real eigenvectors of this matrix are:

$$\begin{aligned} \mathbf{e}_1 &\approx \begin{bmatrix} 0.8460 & -0.1922 & -0.4973 \end{bmatrix} \\ \mathbf{e}_2 &\approx \begin{bmatrix} 0.4828 & 0.6718 & 0.5618 \end{bmatrix} \\ \mathbf{e}_3 &\approx \begin{bmatrix} 0.2261 & -0.7154 & -0.6611 \end{bmatrix} \end{aligned} \tag{0.49}$$

so the approximate answer wasn't bad. The main problem with this technique is the accumulation of errors: a small error in the first eigenvector can produce much larger error in subsequent eigenvectors, especially if the eigenvalues of the last few eigenvectors are very small.

1.4.4 Adapting the Power Method

There is another way we can adapt the power method to find the other eigenvalues (or at least some of them). This relies on the useful facts that if $\mathbf{x}$ is an eigenvector of the matrix $\mathbf{A}$ with an eigenvalue of λ, then:

1. $\mathbf{x}$ is also an eigenvector of the matrix $\mathbf{A}^{-1}$, with an eigenvalue of λ^{-1}
2. $\mathbf{x}$ is also an eigenvector of the matrix $(\mathbf{A} - \alpha\mathbf{I})^{-1}$, with an eigenvalue of $1 / (\lambda - \alpha)$

(for the proofs, see the problems. Note the first fact is just a special case of the second fact, with $\alpha = 0$.)

Using the first of these results, we can immediately find the eigenvector with the smallest absolute value eigenvalue. Just use the same power method as before, only use $\mathbf{A}^{-1}$ instead of $\mathbf{A}$. (Of course this means we have to invert the matrix $\mathbf{A}$, which is expensive.)

Using the iterative method on the matrix $(\mathbf{A} - \alpha\mathbf{I})^{-1}$ will find the eigenvalue with the largest value of $1 / (\lambda - \alpha)$: in other words, the eigenvalue closest to α. For positive definite matrices, we will already know the largest and smallest eigenvalues, so we could start by looking half-way between them, and see what comes up. This process can then be repeated, looking directly in-between all values of α tried so far, until all the eigenvalues are found. It doesn't always work, but for small matrices, it can usually find the eigenvectors comparatively quickly.

1.4.5 The QR-Method

I should just mention that one of the most common ways to determine eigenvalues and eigenvectors is not described in this chapter, since it's beyond the scope of this book. It's an extension of the power method called the QR algorithm, and it works using the QR-decomposition, which decomposes a matrix into the product of an orthogonal matrix (**Q**) and an upper triangular matrix (**R**). This iterative method finds all the eigenvectors and eigenvalues.

1.5 Problems

1) A matrix can have two different eigenvectors with the same eigenvalue. However, prove that no matrix can have two identical eigenvectors with different eigenvalues.

2) Prove that all the eigenvalues of any Hermitian matrix are real.

3) If $\mathbf{A} = \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}$, evaluate the determinant of **A**. What are the eigenvalues and eigenvectors of **A**?

4) Try working out the largest eigenvalue of the matrices $\begin{bmatrix} 11 & -2 & 9 \\ -4 & 4 & -5 \\ 3 & -9 & 11 \end{bmatrix}$ and $\begin{bmatrix} 16 & -6 & -7 \\ 4 & 5 & 13 \\ -1 & 2 & 7 \end{bmatrix}$ using the power method. How many iterations does it take to get within 1% of the right answer? Why is the answer so different in the two cases?

5) Prove that if **x** is an eigenvector of matrix **A** with eigenvalue λ, then **x** is also an eigenvector of the matrix $(\mathbf{A} - a\mathbf{I})^{-1}$, with an eigenvalue of $1 / (\lambda - a)$.

6) Try using the extensions to the power method described to work out all the eigenvalues of the matrix:

$$\begin{bmatrix} 12 & -4 & -3 \\ -2 & 10 & -5 \\ 4 & -12 & 9 \end{bmatrix}$$

(Hint: find the largest and smallest first, then look for one in the middle.)

7) Now try using the extensions to the power method described to work out all the eigenvalues of the matrix:

$$\begin{bmatrix} 6 & -9 & -4 \\ 0 & 2 & -12 \\ 0 & -7 & -15 \end{bmatrix}$$

What goes wrong, and why?