Chapter 11

Dimensionality Reduction

There are many sources of data that can be viewed as a large matrix. We saw in Chapter 5 how the Web can be represented as a transition matrix. In Chapter 9, the utility matrix was a point of focus. And in Chapter 10 we examined matrices that represent social networks. In many of these matrix applications, the matrix can be summarized by finding "narrower" matrices that in some sense are close to the original. These narrow matrices have only a small number of rows or a small number of columns, and therefore can be used much more efficiently than can the original large matrix. The process of finding these narrow matrices is called *dimensionality reduction*.

We saw a preliminary example of dimensionality reduction in Section 9.4. There, we discussed UV-decomposition of a matrix and gave a simple algorithm for finding this decomposition. Recall that a large matrix M was decomposed into two matrices U and V whose product UV was approximately M. The matrix U had a small number of columns whereas V had a small number of rows, so each was significantly smaller than M, and yet together they represented most of the information in M that was useful in predicting ratings of items by individuals.

In this chapter we shall explore the idea of dimensionality reduction in more detail. We begin with a discussion of eigenvalues and their use in "principal component analysis" (PCA). We cover singular-value decomposition, a more powerful version of UV-decomposition. Finally, because we are always interested in the largest data sizes we can handle, we look at another form of decomposition, called CUR-decomposition, which is a variant of singular-value decomposition that keeps the matrices of the decomposition sparse if the original matrix is sparse.

11.1 Eigenvalues and Eigenvectors of Symmetric Matrices

We shall assume that you are familiar with the basics of matrix algebra: multiplication, transpose, determinants, and solving linear equations for example. In this section, we shall define eigenvalues and eigenvectors of a symmetric matrix and show how to find them. Recall a matrix is symmetric if the element in row i and column j equals the element in row j and column i.

11.1.1 Definitions

Let M be a square matrix. Let λ be a constant and $\mathbf{e}$ a nonzero column vector with the same number of rows as M. Then λ is an *eigenvalue* of M and $\mathbf{e}$ is the corresponding *eigenvector* of M if $M\mathbf{e} = \lambda\mathbf{e}$.

If $\mathbf{e}$ is an eigenvector of M and c is any constant, then it is also true that $c\,\mathbf{e}$ is an eigenvector of M with the same eigenvalue. Multiplying a vector by a constant changes the length of a vector, but not its direction. Thus, to avoid ambiguity regarding the length, we shall require that every eigenvector be a *unit vector*, meaning that the sum of the squares of the components of the vector is 1. Even that is not quite enough to make the eigenvector unique, since we may still multiply by -1 without changing the sum of squares of the components. Thus, we shall normally require that the first nonzero component of an eigenvector be positive.

Example 11.1: Let M be the matrix

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

One of the eigenvectors of M is

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

and its corresponding eigenvalue is 7. The equation

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

demonstrates the truth of this claim. Note that both sides are equal to

$$\begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

Also observe that the eigenvector is a unit vector, because $(1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$. $\square$

11.1.2 Computing Eigenvalues and Eigenvectors

We have already seen one approach to finding an *eigenpair* (an eigenvalue and its corresponding eigenvector) for a suitable matrix M in Section 5.1: start with any unit vector $\mathbf{v}$ of the appropriate length and compute $M^i\mathbf{v}$ iteratively until it converges.[1] When M is a stochastic matrix, the limiting vector is the *principal* eigenvector (the eigenvector with the largest eigenvalue), and its corresponding eigenvalue is 1.[2] This method for finding the principal eigenvector, called *power iteration*, works quite generally, although if the principal eigenvalue (eigenvalue associated with the principal eigenvector) is not 1, then as i grows, the ratio of $M^{i+1}\mathbf{v}$ to $M^i\mathbf{v}$ approaches the principal eigenvalue while $M^i\mathbf{v}$ approaches a vector (probably not a unit vector) with the same direction as the principal eigenvector.

We shall take up the generalization of the power-iteration method to find all eigenpairs in Section 11.1.3. However, there is an $O(n^3)$-running-time method for computing all the eigenpairs of a symmetric $n \times n$ matrix exactly, and this method will be presented first. There will always be n eigenpairs, although in some cases, some of the eigenvalues will be identical. The method starts by restating the equation that defines eigenpairs, $M\mathbf{e} = \lambda\mathbf{e}$ as $(M - \lambda I)\mathbf{e} = \mathbf{0}$, where

1. I is the $n \times n$ *identity matrix* with 1's along the main diagonal and 0's elsewhere.

2. $\mathbf{0}$ is a vector of all 0's.

A fact of linear algebra is that in order for $(M - \lambda I)\mathbf{e} = \mathbf{0}$ to hold for a vector $\mathbf{e} \neq \mathbf{0}$, the determinant of $M - \lambda I$ must be 0. Notice that $(M - \lambda I)$ looks almost like the matrix M, but if M has c in one of its diagonal elements, then $(M - \lambda I)$ has $c - \lambda$ there. While the determinant of an $n \times n$ matrix has $n!$ terms, it can be computed in various ways in $O(n^3)$ time; an example is the method of "pivotal condensation."

The determinant of $(M - \lambda I)$ is an nth-degree polynomial in λ, from which we can get the n values of λ that are the eigenvalues of M. For any such value, say c, we can then solve the equation $M\mathbf{e} = c\,\mathbf{e}$. There are n equations in n unknowns (the n components of $\mathbf{e}$), but since there is no constant term in any equation, we can only solve for $\mathbf{e}$ to within a constant factor. However, using any solution, we can normalize it so the sum of the squares of the components is 1, thus obtaining the eigenvector that corresponds to eigenvalue c.

Example 11.2: Let us find the eigenpairs for the 2×2 matrix M from Example 11.1. Recall $M =$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

[1] Recall M^i denotes multiplying by the matrix M i times, as discussed in Section 5.1.2.

[2] Note that a stochastic matrix is not generally symmetric. Symmetric matrices and stochastic matrices are two classes of matrices for which eigenpairs exist and can be exploited. In this chapter, we focus on techniques for symmetric matrices.

Then $M - \lambda I$ is

$$\begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix}$$

The determinant of this matrix is $(3-\lambda)(6-\lambda) - 4$, which we must set to 0. The equation in λ to solve is thus $\lambda^2 - 9\lambda + 14 = 0$. The roots of this equation are $\lambda = 7$ and $\lambda = 2$; the first is the principal eigenvalue, since it is the larger. Let $\mathbf{e}$ be the vector of unknowns

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

We must solve

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix}$$

When we multiply the matrix and vector we get two equations

$$\begin{aligned} \text{3x+2y} &= \text{7x} \\ \text{2x+6y} &= \text{7y} \end{aligned}$$

Notice that both of these equations really say the same thing: $y = 2x$. Thus, a possible eigenvector is

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

But that vector is not a unit vector, since the sum of the squares of its components is 5, not 1. Thus to get the unit vector in the same direction, we divide each component by $\sqrt{5}$. That is, the principal eigenvector is

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

and its eigenvalue is 7. Note that this was the eigenpair we explored in Example 11.1.

For the second eigenpair, we repeat the above with eigenvalue 2 in place of 7. The equation involving the components of $\mathbf{e}$ is $x = -2y$, and the second eigenvector is

$$\begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$$

Its corresponding eigenvalue is 2, of course. □

11.1.3 Finding Eigenpairs by Power Iteration

We now examine the generalization of the process we used in Section 5.1 to find the principal eigenvector, which in that section was the PageRank vector – all we needed from among the various eigenvectors of the stochastic matrix of the Web. We start by computing the principal eigenvector by a slight generalization of the approach used in Section 5.1. We then modify the matrix to, in effect, remove the principal eigenvector. The result is a new matrix whose principal eigenvector is the second eigenvector (eigenvector with the second-largest eigenvalue) of the original matrix. The process proceeds in that manner, removing each eigenvector as we find it, and then using power iteration to find the principal eigenvector of the matrix that remains.

Let M be the matrix whose eigenpairs we would like to find. Start with any nonzero vector $\mathbf{x}_0$ and then iterate:

$$\mathbf{x}_{k+1} := \frac{M\mathbf{x}_k}{\|M\mathbf{x}_k\|}$$

where $\|N\|$ for a matrix or vector N denotes the *Frobenius norm*; that is, the square root of the sum of the squares of the elements of N. We multiply the current vector $\mathbf{x}_k$ by the matrix M until convergence (i.e., $\|x_k - x_{k+1}\|$ is less than some small, chosen constant). Let $\mathbf{x}$ be $\mathbf{x}_k$ for that value of k at which convergence is obtained. Then $\mathbf{x}$ is (approximately) the principal eigenvector of M. To obtain the corresponding eigenvalue we simply compute $\lambda_1 = \mathbf{x}^{\mathrm{T}} M\mathbf{x}$, which is the equation $M\mathbf{x} = \lambda\mathbf{x}$ solved for λ, since $\mathbf{x}$ is a unit vector.

Example 11.3 : Take the matrix from Example 11.2:

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

and let us start with $\mathbf{x}_0$ a vector with 1 for both components. To compute $\mathbf{x}_1$, we multiply $M\mathbf{x}_0$ to get

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

The Frobenius norm of the result is $\sqrt{5^2 + 8^2} = \sqrt{89} = 9.434$. We obtain $\mathbf{x}_1$ by dividing 5 and 8 by 9.434; that is:

$$\mathbf{x}_1 = \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix}$$

For the next iteration, we compute

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix} = \begin{bmatrix} 3.286 \\ 6.148 \end{bmatrix}$$

The Frobenius norm of the result is 6.971, so we divide to obtain

$$\mathbf{x}_2 = \begin{bmatrix} 0.471 \\ 0.882 \end{bmatrix}$$

We are converging toward a normal vector whose second component is twice the first. That is, the limiting value of the vector that we obtain by power iteration is the principal eigenvector:

$$\mathbf{x} = \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix}$$

Finally, we compute the principal eigenvalue by

$$\lambda = \mathbf{x}^{\mathrm{T}} M \mathbf{x} = \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} = 6.993$$

Recall from Example 11.2 that the true principal eigenvalue is 7. Power iteration will introduce small errors due either to limited precision, as was the case here, or due to the fact that we stop the iteration before reaching the exact value of the eigenvector. When we computed PageRank, the small inaccuracies did not matter, but when we try to compute all eigenpairs, inaccuracies accumulate if we are not careful. □

To find the second eigenpair we create a new matrix $M^* = M - \lambda_1 \mathbf{x}\mathbf{x}^{\mathrm{T}}$. Then, use power iteration on M^* to compute its largest eigenvalue. The obtained $\mathbf{x}*$ and $\lambda*$ correspond to the second largest eigenvalue and the corresponding eigenvector of matrix M.

Intuitively, what we have done is eliminate the influence of a given eigenvector by setting its associated eigenvalue to zero. The formal justification is the following two observations. If $M^* = M - \lambda \mathbf{x}\mathbf{x}^{\mathrm{T}}$, where $\mathbf{x}$ and λ are the eigenpair with the largest eigenvalue, then:

1. $\mathbf{x}$ is also an eigenvector of M^*, and its corresponding eigenvalue is 0. In proof, observe that

$$M^*\mathbf{x} = (M - \lambda \mathbf{x}\mathbf{x}^{\mathrm{T}})\mathbf{x} = M\mathbf{x} - \lambda \mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{x} = M\mathbf{x} - \lambda\mathbf{x} = 0$$

   At the next-to-last step we use the fact that $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$ because $\mathbf{x}$ is a unit vector.

2. Conversely, if $\mathbf{v}$ and λ_v are an eigenpair of a symmetric matrix M other than the first eigenpair $(\mathbf{x}, \lambda)$, then they are also an eigenpair of M^*. *Proof*:

$$M^*\mathbf{v} = (M^*)^{\mathrm{T}}\mathbf{v} = (M - \lambda \mathbf{x}\mathbf{x}^{\mathrm{T}})^{\mathrm{T}}\mathbf{v} = M^{\mathrm{T}}\mathbf{v} - \lambda\mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{v}) = M^{\mathrm{T}}\mathbf{v} = \lambda_v \mathbf{v}$$

   This sequence of equalities needs the following justifications:

   (a) If M is symmetric, then $M = M^{\mathrm{T}}$.

   (b) The eigenvectors of a symmetric matrix are *orthogonal*. That is, the dot product of any two distinct eigenvectors of a matrix is 0. We do not prove this statement here.

Example 11.4: Continuing Example 11.3, we compute

$$M^* = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - 6.993 \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} =$$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} 1.397 & 2.795 \\ 2.795 & 5.589 \end{bmatrix} = \begin{bmatrix} 1.603 & -0.795 \\ -0.795 & 0.411 \end{bmatrix}$$

We may find the second eigenpair by processing the matrix above as we did the original matrix M. □

11.1.4 The Matrix of Eigenvectors

Suppose we have an $n \times n$ symmetric matrix M whose eigenvectors, viewed as column vectors, are $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Let E be the matrix whose ith column is $\mathbf{e}_i$. Then $EE^{\mathrm{T}} = E^{\mathrm{T}}E = I$. The explanation is that the eigenvectors of a symmetric matrix are *orthonormal*. That is, they are orthogonal unit vectors.

Example 11.5 : For the matrix M of Example 11.2, the matrix E is

$$\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

E^{T} is therefore

$$\begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

When we compute EE^{T} we get

$$\begin{bmatrix} 4/5 + 1/5 & -2/5 + 2/5 \\ -2/5 + 2/5 & 1/5 + 4/5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The calculation is similar when we compute $E^{\mathrm{T}}E$. Notice that the 1's along the main diagonal are the sums of the squares of the components of each of the eigenvectors, which makes sense because they are unit vectors. The 0's off the diagonal reflect the fact that the entry in the ith row and jth column is the dot product of the ith and jth eigenvectors. Since eigenvectors are orthogonal, these dot products are 0. $\square$

11.1.5 Exercises for Section 11.1

Exercise 11.1.1 : Find the unit vector in the same direction as the vector $[1, 2, 3]$.

Exercise 11.1.2 : Complete Example 11.4 by computing the principal eigenvector of the matrix that was constructed in this example. How close to the correct solution (from Example 11.2) are you?

Exercise 11.1.3 : For any symmetric 3×3 matrix

$$\begin{bmatrix} a - \lambda & b & c \\ b & d - \lambda & e \\ c & e & f - \lambda \end{bmatrix}$$

there is a cubic equation in λ that says the determinant of this matrix is 0. In terms of a through f, find this equation.

Exercise 11.1.4: Find the eigenpairs for the following matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{bmatrix}$$

using the method of Section 11.1.2.

! **Exercise 11.1.5:** Find the eigenpairs for the following matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 6 \end{bmatrix}$$

using the method of Section 11.1.2.

Exercise 11.1.6: For the matrix of Exercise 11.1.4:

(a) Starting with a vector of three 1's, use power iteration to find an approximate value of the principal eigenvector.

(b) Compute an estimate the principal eigenvalue for the matrix.

(c) Construct a new matrix by subtracting out the effect of the principal eigenpair, as in Section 11.1.3.

(d) From your matrix of (c), find the second eigenpair for the original matrix of Exercise 11.1.4.

(e) Repeat (c) and (d) to find the third eigenpair for the original matrix.

Exercise 11.1.7: Repeat Exercise 11.1.6 for the matrix of Exercise 11.1.5.

11.2 Principal-Component Analysis

Principal-component analysis, or PCA, is a technique for taking a dataset consisting of a set of tuples representing points in a high-dimensional space and finding the directions along which the tuples line up best. The idea is to treat the set of tuples as a matrix M and find the eigenvectors for MM^{T} or $M^{\mathrm{T}}M$. The matrix of these eigenvectors can be thought of as a rigid rotation in a high-dimensional space. When you apply this transformation to the original data, the axis corresponding to the principal eigenvector is the one along which the points are most "spread out," More precisely, this axis is the one along which the variance of the data is maximized. Put another way, the points can best be viewed as lying along this axis, with small deviations from this axis. Likewise, the axis corresponding to the second eigenvector (the eigenvector corresponding to the second-largest eigenvalue) is the axis along which the variance of distances from the first axis is greatest, and so on.

We can view PCA as a data-mining technique. The high-dimensional data can be replaced by its projection onto the most important axes. These axes are the ones corresponding to the largest eigenvalues. Thus, the original data is approximated by data that has many fewer dimensions and that summarizes well the original data.

11.2.1 An Illustrative Example

We shall start the exposition with a contrived and simple example. In this example, the data is two-dimensional, a number of dimensions that is too small to make PCA really useful. Moreover, the data, shown in Fig. 11.1 has only four points, and they are arranged in a simple pattern along the 45-degree line to make our calculations easy to follow. That is, to anticipate the result, the points can best be viewed as lying along the axis that is at a 45-degree angle, with small deviations in the perpendicular direction.

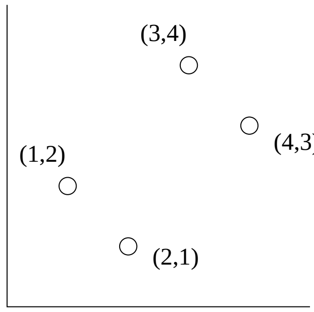

Figure 11.1: Four points in a two-dimensional space

To begin, let us represent the points by a matrix M with four rows – one for each point – and two columns, corresponding to the x-axis and y-axis. This matrix is

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

Compute $M^{\mathrm{T}}M$, which is

$$M^{\mathrm{T}}M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$$

We may find the eigenvalues of the matrix above by solving the equation

$$(30 - \lambda)(30 - \lambda) - 28 \times 28 = 0$$

as we did in Example 11.2. The solution is $\lambda = 58$ and $\lambda = 2$.

Following the same procedure as in Example 11.2, we must solve

$$\begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 58 \begin{bmatrix} x \\ y \end{bmatrix}$$

When we multiply out the matrix and vector we get two equations

$$\begin{aligned} 30x{+}28y &= 58x \\ 28x{+}30y &= 58y \end{aligned}$$

Both equations tell us the same thing: $x = y$. Thus, the unit eigenvector corresponding to the principal eigenvalue 58 is

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

For the second eigenvalue, 2, we perform the same process. Multiply out

$$\begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix}$$

to get the two equations

$$\begin{aligned} 30x{+}28y &= 2x \\ 28x{+}30y &= 2y \end{aligned}$$

Both equations tell us the same thing: $x = -y$. Thus, the unit eigenvector corresponding to the principal eigenvalue 2 is

$$\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

While we promised to write eigenvectors with their first component positive, we choose the opposite here because it makes the transformation of coordinates easier to follow in this case.

Now, let us construct E, the matrix of eigenvectors for the matrix $M^{\mathrm{T}}M$. Placing the principal eigenvector first, the matrix of eigenvectors is

$$E = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Any matrix of *orthonormal vectors* (unit vectors that are orthogonal to one another) represents a rotation and/or reflection of the axes of a Euclidean space. The matrix above can be viewed as a rotation 45 degrees counterclockwise. For example, let us multiply the matrix M that represents each of the points of Fig. 11.1 by E. The product is

$$ME = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

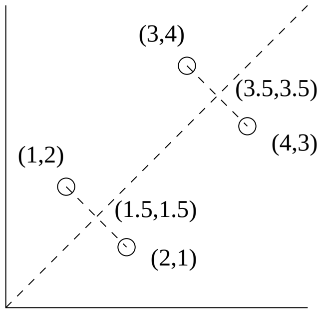

Figure 11.2: Figure 11.1 with the axes rotated 45 degrees counterclockwise

We see the first point, $[1, 2]$, has been transformed into the point

$$[3/\sqrt{2}, 1/\sqrt{2}]$$

If we examine Fig. 11.2, with the dashed line representing the new x-axis, we see that the projection of the first point onto that axis places it at distance $3/\sqrt{2}$ from the origin. To check this fact, notice that the point of projection for both the first and second points is $[1.5, 1.5]$ in the original coordinate system, and the distance from the origin to this point is

$$\sqrt{(1.5)^2 + (1.5)^2} = \sqrt{9/2} = 3/\sqrt{2}$$

Moreover, the new y-axis is, of course, perpendicular to the dashed line. The first point is at distance $1/\sqrt{2}$ above the new x-axis in the direction of the y-axis. That is, the distance between the points $[1, 2]$ and $[1.5, 1.5]$ is

$$\sqrt{(1 - 1.5)^2 + (2 - 1.5)^2} = \sqrt{(-1/2)^2 + (1/2)^2} = \sqrt{1/2} = 1/\sqrt{2}$$

Figure 11.3 shows the four points in the rotated coordinate system.

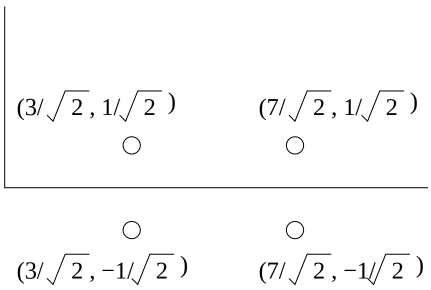

Figure 11.3: The points of Fig. 11.1 in the new coordinate system

The second point, $[2, 1]$ happens by coincidence to project onto the same point of the new x-axis. It is $1/\sqrt{2}$ below that axis along the new y-axis, as is confirmed by the fact that the second row in the matrix of transformed points is $[3/\sqrt{2}, -1/\sqrt{2}]$. The third point, $[3, 4]$ is transformed into $[7/\sqrt{2}, 1/\sqrt{2}]$ and the fourth point, $[4, 3]$, is transformed to $[7/\sqrt{2}, -1/\sqrt{2}]$. That is, they both project onto the same point of the new x-axis, and that point is at distance $7/\sqrt{2}$ from the origin, while they are $1/\sqrt{2}$ above and below the new x-axis in the direction of the new y-axis.

11.2.2 Using Eigenvectors for Dimensionality Reduction

From the example we have just worked out, we can see a general principle. If M is a matrix whose rows each represent a point in a Euclidean space with any number of dimensions, we can compute $M^{\mathrm{T}}M$ and compute its eigenpairs. Let E be the matrix whose columns are the eigenvectors, ordered as largest eigenvalue first. Define the matrix L to have the eigenvalues of $M^{\mathrm{T}}M$ along the diagonal, largest first, and 0's in all other entries. Then, since $M^{\mathrm{T}}M\mathbf{e} = \lambda\mathbf{e} = \mathbf{e}\lambda$ for each eigenvector $\mathbf{e}$ and its corresponding eigenvalue λ, it follows that $M^{\mathrm{T}}ME = EL$.

We observed that ME is the points of M transformed into a new coordinate space. In this space, the first axis (the one corresponding to the largest eigenvalue) is the most significant; formally, the variance of points along that axis is the greatest. The second axis, corresponding to the second eigenpair, is next most significant in the same sense, and the pattern continues for each of the eigenpairs. If we want to transform M to a space with fewer dimensions, then the choice that preserves the most significance is the one that uses the eigenvectors associated with the largest eigenvalues and ignores the other eigenvalues.

That is, let E_k be the first k columns of E. Then ME_k is a k-dimensional representation of M.

Example 11.6 : Let M be the matrix from Section 11.2.1. This data has only two dimensions, so the only dimensionality reduction we can do is to use $k = 1$; i.e., project the data onto a one dimensional space. That is, we compute ME_1 by

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} \\ 3/\sqrt{2} \\ 7/\sqrt{2} \\ 7/\sqrt{2} \end{bmatrix}$$

The effect of this transformation is to replace the points of M by their projections onto the x-axis of Fig. 11.3. While the first two points project to the same point, as do the third and fourth points, this representation makes the best possible one-dimensional distinctions among the points. □

11.2.3 The Matrix of Distances

Let us return to the example of Section 11.2.1, but instead of starting with $M^{\mathrm{T}}M$, let us examine the eigenvalues of MM^{T}. Since our example M has more rows than columns, the latter is a bigger matrix than the former, but if M had more columns than rows, we would actually get a smaller matrix. In the running example, we have

$$MM^{\mathrm{T}} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 4 & 11 & 10 \\ 4 & 5 & 10 & 11 \\ 11 & 10 & 25 & 24 \\ 10 & 11 & 24 & 25 \end{bmatrix}$$

Like $M^{\mathrm{T}}M$, we see that MM^{T} is symmetric. The entry in the ith row and jth column has a simple interpretation; it is the dot product of the vectors represented by the ith and jth points (rows of M).

There is a strong relationship between the eigenvalues of $M^{\mathrm{T}}M$ and MM^{T}. Suppose $\mathbf{e}$ is an eigenvector of $M^{\mathrm{T}}M$; that is,

$$M^{\mathrm{T}}M\mathbf{e} = \lambda\mathbf{e}$$

Multiply both sides of this equation by M on the left. Then

$$MM^{\mathrm{T}}(M\mathbf{e}) = M\lambda\mathbf{e} = \lambda(M\mathbf{e})$$

Thus, as long as $M\mathbf{e}$ is not the zero vector $\mathbf{0}$, it will be an eigenvector of MM^{T} and λ will be an eigenvalue of MM^{T} as well as of $M^{\mathrm{T}}M$.

The converse holds as well. That is, if $\mathbf{e}$ is an eigenvector of MM^{T} with corresponding eigenvalue λ, then start with $MM^{\mathrm{T}}\mathbf{e} = \lambda\mathbf{e}$ and multiply on the left by M^{T} to conclude that $M^{\mathrm{T}}M(M^{\mathrm{T}}\mathbf{e}) = \lambda(M^{\mathrm{T}}\mathbf{e})$. Thus, if $M^{\mathrm{T}}\mathbf{e}$ is not $\mathbf{0}$, then λ is also an eigenvalue of $M^{\mathrm{T}}M$.

We might wonder what happens when $M^{\mathrm{T}}\mathbf{e} = \mathbf{0}$. In that case, $MM^{\mathrm{T}}\mathbf{e}$ is also $\mathbf{0}$, but $\mathbf{e}$ is not $\mathbf{0}$ because $\mathbf{0}$ cannot be an eigenvector. However, since $\mathbf{0} = \lambda\mathbf{e}$, we conclude that $\lambda = 0$.

We conclude that the eigenvalues of MM^{T} are the eigenvalues of $M^{\mathrm{T}}M$ plus additional 0's. If the dimension of MM^{T} were less than the dimension of $M^{\mathrm{T}}M$, then the opposite would be true; the eigenvalues of $M^{\mathrm{T}}M$ would be those of MM^{T} plus additional 0's.

$$\begin{bmatrix} 3/\sqrt{116} & 1/2 & 7/\sqrt{116} & 1/2 \\ 3/\sqrt{116} & -1/2 & 7/\sqrt{116} & -1/2 \\ 7/\sqrt{116} & 1/2 & -3/\sqrt{116} & -1/2 \\ 7/\sqrt{116} & -1/2 & -3/\sqrt{116} & 1/2 \end{bmatrix}$$

Figure 11.4: Eigenvector matrix for MM^{T}

Example 11.7: The eigenvalues of MM^{T} for our running example must include 58 and 2, because those are the eigenvalues of $M^{\mathrm{T}}M$ as we observed in Section 11.2.1. Since MM^{T} is a 4×4 matrix, it has two other eigenvalues, which must both be 0. The matrix of eigenvectors corresponding to 58, 2, 0, and 0 is shown in Fig. 11.4. $\Box$

11.2.4 Exercises for Section 11.2

Exercise 11.2.1: Let M be the matrix of data points

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$

(a) What are $M^{\mathrm{T}}M$ and MM^{T}?

(b) Compute the eigenpairs for $M^{\mathrm{T}}M$.

! (c) What do you expect to be the eigenvalues of MM^{T}?

! (d) Find the eigenvectors of MM^{T}, using your eigenvalues from part (c).

! Exercise 11.2.2: Prove that if M is any matrix, then $M^{\mathrm{T}}M$ and MM^{T} are symmetric.

11.3 Singular-Value Decomposition

We now take up a second form of matrix analysis that leads to a low-dimensional representation of a high-dimensional matrix. This approach, called *singular-value decomposition* (SVD), allows an exact representation of any matrix, and also makes it easy to eliminate the less important parts of that representation to produce an approximate representation with any desired number of dimensions. Of course the fewer the dimensions we choose, the less accurate will be the approximation.

We begin with the necessary definitions. Then, we explore the idea that the SVD defines a small number of "concepts" that connect the rows and columns of the matrix. We show how eliminating the least important concepts gives us a smaller representation that closely approximates the original matrix. Next, we see how these concepts can be used to query the original matrix more efficiently, and finally we offer an algorithm for performing the SVD itself.

11.3.1 Definition of SVD

Let M be an $m \times n$ matrix, and let the rank of M be r. Recall that the *rank* of a matrix is the largest number of rows (or equivalently columns) we can choose