Diagonalizing symmetric matrices

In this writeup, for dot products we'll use inner-product notation: $\mathbf{u} \cdot \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$.

1. Rotations

Definition. A list of vectors in $\mathbf{R}^n$ is **orthonormal** if the vectors are mutually perpendicular and all of length 1.

It can be shown that orthonormal vectors are linearly independent.

Recall that you can test perpendicularity (orthogonality) of two vectors by checking that their dot product is 0; also, the dot product of a vector with itself is its length squared. Therefore:

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are orthonormal $\Leftrightarrow \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \begin{cases} 1 & \text{if } \ i = j \\ 0 & \text{if } \ i \neq j \end{cases}$

Definition. An $n \times n$ matrix R is a **rotation matrix** if

(i) the columns of R are orthonormal, and

(ii) $\det R = 1$.

For 2×2 matrices, rotations are just what you are used to, R_θ. However, using (i) and (ii) you can tell if a matrix is a rotation matrix without knowing the angle. For example, $R = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}$ is a rotation, as you can check.

2. Symmetric matrices

As you know, even for a 2×2 matrix A, the properties of A might not be ideal. There might not be a basis of linearly independent eigenvectors, or the eigenvalues and eigenvectors might be complex, or perhaps the eigenspaces of A are not perpendicular. In contrast, symmetric matrices are guaranteed to have the best possible properties:

Theorem ("spectral theorem"). If A is a real symmetric matrix, then

(1) the eigenvalues of A are real numbers,

(2) the eigenspaces of A are perpendicular to one another,

(3) A can be diagonalized using a rotation matrix; in other words, $R^{-1}AR = D$ for some rotation matrix R.

The name "spectral theorem" comes from advanced applications in physics where the eigenvalues determine frequencies, like a color spectrum of light.

Problem BB-1. Explain how (3) implies (1) and (2). (Method: In (3), does D have real entries? Are the eigenvectors perpendicular?)

The nice thing in applications is that there is is not much to do to find R; as the theorem says, the eigenspaces are automatically perpendicular to one another.

Example: $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. Here in the past we found that we could use $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$. But P is not a rotation matrix, since it violates both (i) and (ii). The columns are perpendicular as expected, though. To fix (i), find the length of each column and divide by it to make all column lengths be 1:

$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

Now check the determinant. It's -1, so change the sign of a column, say the second. We get

$R = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$.

Now R does fit the definition of a rotation. Notice that $R = R_{45^\circ}$.

Problem BB-2. Diagonalize $A = \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix}$ with a rotation matrix.

3. More on orthonormal sets of vectors and rotations

Problem BB-3. Show that an orthonormal set of vectors is linearly independent.

(Method: Given a linear combination = 0, take the inner product of both sides with the first vector and expand, using the fact that the inner product is linear in each entry. Now try the second vector, etc.)

In the definition of a rotation matrix, notice that by (i), the matrix transformation T_R takes the standard basis to an orthonormal basis, so T_R acts "rigidly". By (ii), orientation is preserved. More generally,

Definition. A matrix whose columns are orthonormal is called an *orthogonal* matrix (not a good terminology, since it really should be an "orthonormal" matrix). Thus a rotation matrix is an orthogonal matrix with determinant 1. The following fact can be proved:

Theorem. For an $n \times n$ real matrix P, the following are equivalent:

(1) The columns of P are orthonormal (i.e., P is an orthogonal matrix);

(2) $P^t P = I$;

(3) P is invertible and $P^{-1} = P^t$;

(4) the rows of P are orthonormal;

(5) T_P is a rigid transformation, meaning that all distances are preserved.

Problem BB-4. Prove (1) $\Leftrightarrow$ (2). (Think about what happens in multiplying when P is 2×2 or 3×3.)

Problem BB-5. Find the inverse of the rotation $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

Proposition. If P is an orthogonal matrix then $\det P = \pm 1$.

Problem BB-6. Prove this proposition. (Start from (2) and use the facts that $\det AB = \det A \, \det B$ and that $\det A^t = \det A$.)

So, rotations are the case $\det P = +1$.

4. An application

A quadratic expression such as $6x^2 + 4xy + 9y^2$ is said to be a "quadratic form".

A quadratic form can be rewritten as $\mathbf{x}^t A \mathbf{x}$, where A is symmetric and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ (a column vector) so $\mathbf{x}^t$ is a row vector. For example, $6x^2 + 4xy + 9y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. (Try multiplying out to check.) The coefficient of xy is split in half and each half becomes an off-diagonal entry.

To analyze the graph of an equation such as $6x^2 + 4xy + 9y^2 = 1$, write it in matrix form, diagonalize the matrix, and make the substitution $\mathbf{x} = R\mathbf{r}$, where R is a rotation matrix.

Example: Describe the graph of $3x^2 + 2xy + 3y^2 = 1$.

Solution: Rewrite as $\mathbf{x}^t A\mathbf{x} = 1$ for $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. We saw that $R^{-1}AR = D$ for $R = R_{45^\circ}$ and $D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$. Substitute in $\mathbf{x} = R\mathbf{r}$. We get $(R\mathbf{r})^t A(R\mathbf{r}) = 1$, or $\mathbf{r}^t(R^t AR)\mathbf{r} = 1$. Since $R^t = R^{-1}$, we get $\mathbf{r}^t D\mathbf{r} = 1$, or $4r^2 + 2s^2 = 1$. This can be rewritten as

$\frac{r^2}{a^2} + \frac{s^2}{b^2} = 1$ for $a = \frac{1}{2}$, $b = \frac{1}{2}\sqrt{2}$,

so the shape is an ellipse with semimajor axis $\frac{1}{2}\sqrt{2}$ and semiminor axis $\frac{1}{2}$. But the ellipse is slanted. How? Since $a < b$, the end of the semimajor axis is where $r = 0$, $s = b = \frac{1}{2}\sqrt{2}$; here $\mathbf{x} = R\begin{bmatrix} 0 \\ \frac{1}{2}\sqrt{2} \end{bmatrix} = [\mathbf{v}_1|\mathbf{v}_2]\begin{bmatrix} 0 \\ \frac{1}{2}\sqrt{2} \end{bmatrix} = \frac{1}{2}\sqrt{2}\mathbf{v}_2$, where $\mathbf{v}_2$ is the second eigenvector, which is the same as the second column of R. Therefore the end of the semimajor axis of R is at $\frac{1}{2}\sqrt{2}\begin{bmatrix} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$. You can check that this point is on the graph.

Problem BB-7. In this example, find the end of the semiminor axis.

Problem BB-8. (a) In general, the transformed equation is $\lambda_1 r^2 + \lambda_2 s^2 = 1$. Explain why, by discussing the substitution.

(b) How can you tell from the eigenvalues whether the equation is going to be an ellipse or a hyperbola or maybe some degenerate case? (In fact, what kinds of graphs can arise from (a)?)

Problem BB-9. (a) Describe the shape of the graph of $6x^2 + 4xy + 9y^2 = 1$. If it is an ellipse, give semimajor and semiminor axis lengths. (For this you can use eigenvalues alone.)

(b) How is this graph slanted? (You'll need the eigenvectors.)

5. Proof of the Spectral Theorem

We'll concentrate on the 2×2 case.

Problem BB-10. For a general symmetric matrix $\begin{bmatrix} a & b \\ b & d \end{bmatrix}$,

(a) find the eigenvalues in terms of a, b, d. (You will need to use the quadratic formula to get the roots of the characteristic polynomial.) Simplify algebraically to show that the discriminant (the part inside the square root) is the sum of two squares.

(b) How do you know from (a) that the eigenvalues are real?

(c) Show that if the two eigenvalues are equal, then the matrix is a scalar matrix. (Method: Notice that the eigenvalues are equal when the discriminant is zero.)

Now for perpendicularity of eigenvectors:

Proposition. If A is a symmetric matrix, then two eigenvectors belonging to distinct eigenvalues are perpendicular.

Problem BB-11. Show that if A is symmetric, then for any two vectors $\mathbf{u}, \mathbf{v}$ we have $\langle \mathbf{u}, A\mathbf{v} \rangle = \langle A\mathbf{u}, \mathbf{v} \rangle$.

(Method: If we regard $\mathbf{u}, \mathbf{v}$ as column vectors, $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^t\mathbf{v}$. What happens when you put A in, in either location?)

Now to prove the Proposition:

If $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$ and $\lambda_1 \neq \lambda_2$, then start from

$\langle A\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, A\mathbf{v}_2 \rangle$. This becomes

$\langle \lambda_1\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \lambda_2\mathbf{v}_2 \rangle$. Since the dot product is linear in each entry, we get

$\lambda_1\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \lambda_2\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$. Then

$(\lambda_1 - \lambda_2)\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$.

Since $\lambda_1 - \lambda_2 \neq 0$, we must have $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. Therefore $\mathbf{v}_1 \perp \mathbf{v}_2$.

How does all this prove the Spectral Theorem for a 2×2 matrix A? You know from Problem BB-10 that the eigenvalues are real. Also, from the same problem you know that if the eigenvalues are equal then A is *already* diagonal, so you can diagonalize it using I, which is a rotation. On the other hand, if the eigenvalues are different, then the Proposition shows that the eigenvectors are perpendicular. Now you can make the matrix $P = [\mathbf{v}_1|\mathbf{v}_2]$, scale the lengths of the columns to make them length 1, and then negate $\mathbf{v}_2$ if necessary to make the determinant be 1.