4 Orthogonal diagonalization

Our excursion into orthogonality was motivated by the following result.

Proposition 7 *Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be vectors from distinct eigenspaces of the symmetric matrix A. Then*

$$\boldsymbol{v} \cdot \boldsymbol{w} = 0.$$

Proof. We have $A\boldsymbol{v} = c_1\boldsymbol{v}$, $A\boldsymbol{w} = c_2\boldsymbol{w}$, and $c_1 \neq c_2$. Now

$$\boldsymbol{w} \cdot (A\boldsymbol{v}) = \boldsymbol{v} \cdot (A\boldsymbol{w})$$

from Proposition 1 (ii). The left side is $\boldsymbol{w} \cdot c_1\boldsymbol{v} = c_1\boldsymbol{w} \cdot \boldsymbol{v} = c_1\boldsymbol{v} \cdot \boldsymbol{w}$ and the right side is $\boldsymbol{v} \cdot c_2\boldsymbol{w} = c_2\boldsymbol{v} \cdot \boldsymbol{w}$. So

$$(c_1 - c_2)\boldsymbol{v} \cdot \boldsymbol{w} = 0.$$

Since $c_1 - c_2 \neq 0$, we must have $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.

Example 11 Verify the result of Proposition 7 for

$$A = \begin{bmatrix} f & g \\ g & h \end{bmatrix}$$

where $g \neq 0$.

Solution. The characteristic equation is

$$c^2 - (f+h)c + fh - g^2 = 0.$$

The solutions of the quadratic equation are

$$c_1 = \frac{f+h+\sqrt{(f+h)^2 - 4fh + 4g^2}}{2}, c_2 = \frac{f+h-\sqrt{(f+h)^2 - 4fh + 4g^2}}{2}$$

which we rewrite as

$$c_1 = \frac{f+h+\sqrt{(f-h)^2 + 4g^2}}{2}, c_2 = \frac{f+h-\sqrt{(f-h)^2 + 4g^2}}{2}.$$

Since the number D under the square root sign is positive, $c_1 \neq c_2$.

Eigenspace for c_1. The first equation of the system $(c_1I - A)\boldsymbol{x} = \boldsymbol{0}$ is

$$\left(\frac{h-f}{2} + \frac{\sqrt{D}}{2}\right) x_1 - gx_2 = 0.$$

A basis of the eigenspace is $\boldsymbol{v}_1 = \left(g, \frac{h-f}{2} + \frac{\sqrt{D}}{2}\right)$.

Eigenspace for c_2. Repeating the above calculation with $\sqrt{D}$ replaced by $-\sqrt{D}$, a basis of the eigenspace is $\boldsymbol{v}_2 = \left(g, \frac{h-f}{2} - \frac{\sqrt{D}}{2}\right)$. Now

$$\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = g^2 + \frac{(h-f)^2}{4} - \frac{D^2}{4} = 0.$$

Definition 4 Let A be $n \times n$ and let V be a subspace of $\mathbb{R}^n$. We write $A : V \to V$ if $A\boldsymbol{v}$ is in V whenever $\boldsymbol{v}$ is in V.

Example 12 Let A be the matrix of a rotation of $\mathbb{R}^3$ about the axis L. Then $A : L \to L$ and $A : L^\perp \to L^\perp$.

We now come to an interesting result with a geometrical flavor.

Proposition 8 *Let V be a subspace of $\mathbb{R}^n$. Write S for the set of $\boldsymbol{v}$ in V with $|\boldsymbol{v}| = 1$. Let A be a symmetric matrix, $A : V \to V$. The vector $\boldsymbol{v}_1$ in S where $\boldsymbol{v}^tA\boldsymbol{v}$ is largest is an eigenvector of A.*

Proof. The result is obvious if $\dim V = 1$; $A\boldsymbol{v}_1$ is in V and must be a multiple of $\boldsymbol{v}_1$. So we may suppose that $\dim V \geq 2$.

Firstly, there *is* a point $\boldsymbol{v}_1$ in S where $\boldsymbol{v}^tA\boldsymbol{v}$ is largest. This follows from a standard result in calculus about a continuous function on a closed bounded set. Pick any $\boldsymbol{w}$ in S with $\boldsymbol{w} \cdot \boldsymbol{v}_1 = 0$.

Let

$$\boldsymbol{z} = \boldsymbol{v}_1 \cos y + \boldsymbol{w} \sin y,$$

where y is an arbitrary angle. We have $|\boldsymbol{z}|^2 = |\boldsymbol{v}_1|^2 \cos^2 y + |\boldsymbol{w}|^2 \sin^2 y = \cos^2 y + \sin^2 y = 1$, so $\boldsymbol{z}$ is also in S. Consequently $\boldsymbol{z}^tA\boldsymbol{z}$ has a maximum value when $y = 0$ and $\boldsymbol{z} = \boldsymbol{v}_1$. Now

$$\begin{aligned}\boldsymbol{z}^tA\boldsymbol{z} &= (\boldsymbol{v}_1 \cos y + \boldsymbol{w} \sin y)^t A(\boldsymbol{v}_1 \cos y + \boldsymbol{w} \sin y)\\ &= \boldsymbol{v}_1^t A\boldsymbol{v}_1 \cos^2 y + 2\boldsymbol{v}_1^t A\boldsymbol{w} \sin y \cos y + \boldsymbol{w}^t A\boldsymbol{w} \sin^2 y\end{aligned}$$

using Proposition 1 (ii). The derivative of the right side is

$$-2\boldsymbol{v}_1^t A\boldsymbol{v}_1 \sin y \cos y + 2\boldsymbol{v}_1^t A\boldsymbol{w}(\cos^2 y - \sin^2 y) + 2\boldsymbol{w}^t A\boldsymbol{w} \sin y \cos y.$$

The derivative at $y = 0$ is $2\boldsymbol{v}_1^t A\boldsymbol{w}$. Since this is where the maximum occurs,

$$\boldsymbol{v}_1^t A\boldsymbol{w} = 0 = \boldsymbol{w} \cdot A\boldsymbol{v}_1.$$

Summarizing, *any* $\boldsymbol{w}$ *in* S *that is orthogonal to* $\boldsymbol{v}_1$ *is orthogonal to* $A\boldsymbol{v}_1$. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_j$ be an orthonormal basis of V. Then $A\boldsymbol{v}_1$ is a vector in V that is orthogonal to $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$. So

$$A\boldsymbol{v}_1 = (A\boldsymbol{v}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (A\boldsymbol{v}_1 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (A\boldsymbol{v}_1 \cdot \boldsymbol{v}_r)\boldsymbol{v}_r$$

(Proposition 5)

$$= (A\boldsymbol{v}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1.$$

This proves that $\boldsymbol{v}_1$ is an eigenvector of A.

Example 13 Let

$$A = \begin{bmatrix} 8 & & & \\ & 5 & & \\ & & 7 & \\ & & & -3 \end{bmatrix}$$

and $V = \text{Span}\{\boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4\}$. Then $A : V \to V$. Now

$$\boldsymbol{v}^t A\boldsymbol{v} = 5v_2^2 + 7v_3^2 - 3v_4^2$$

for $\boldsymbol{v} = (0, v_2, v_3, v_4)$ in V. If S is the set of $(0, v_2, v_3, v_4)$ with $v_2^2 + v_3^2 + v_4^2 = 1$, then

$$\boldsymbol{v}^t A\boldsymbol{v} = 5v_2^2 + 7v_3^2 - 3v_4^2 \le 7(v_2^2 + v_3^2 + v_4^2) = 7$$

with equality at $(0, 0, \pm 1, 0)$. The proposition tells us that the vectors $(0, 0, \pm 1, 0)$ are eigenvalues of A, which is easy to verify.

Proposition 8 is a stepping stone to the following important result.

Proposition 9 *Let A be $n \times n$ symmetric. There is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

Proof. We apply Proposition 8 repeatedly. In the first step, $V = \mathbb{R}^n$. Obviously, $A : V \to V$, so Proposition 8 supplies an eigenvector $\boldsymbol{v}_1$. In the next step, $V = (\mathrm{Span}\{\boldsymbol{v}_1\})^\perp$; Proposition 8 supplies an eigenvector $\boldsymbol{v}_2$ in V, as we see in the next paragraph. In the jth step, $V = (\mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}\})^\perp$; where the unit vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$ have been chosen already and satisfy $A\boldsymbol{v}_i = c_i\boldsymbol{v}_i$.

We need to be sure that Proposition 8 is applicable to $V = (\mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}\})^\perp$, for each $j = 2, \ldots, n$. If $\boldsymbol{v}$ is in V, then Proposition 1 (ii) gives

$$\boldsymbol{v}_i \cdot A\boldsymbol{v} = \boldsymbol{v} \cdot A\boldsymbol{v}_i = \boldsymbol{v} \cdot c_i\boldsymbol{v}_i = 0 \quad (i = 1, \ldots, j-1).$$

So $A\boldsymbol{v}$ is orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$, and is a vector in V. This gives the required hypothesis $A : V \to V$. Now Proposition 8 supplies a unit eigenvector $\boldsymbol{v}_j$ in V. By definition of V, $\boldsymbol{v}_j \cdot \boldsymbol{v}_i = 0$ for $i < j$. The process concludes when $j = n$.

Example 14 Let $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$. Find an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of A.

Solution. The characteristic polynomial of A is

$$\det \begin{bmatrix} c-4 & -2 & -2 \\ -2 & c-4 & -2 \\ -2 & -2 & c-4 \end{bmatrix} = \det \begin{bmatrix} c-2 & 2-c & 0 \\ -2 & c-4 & -2 \\ 0 & 2-c & c-2 \end{bmatrix} \begin{matrix} \scriptstyle \mathrm{I-II} \\ \scriptstyle \mathrm{III-II} \\ \end{matrix}$$

$$= (c-2)^2 \det \begin{bmatrix} 1 & -1 & 0 \\ -2 & c-4 & -2 \\ 0 & -1 & 1 \end{bmatrix} \text{(factors from rows 1, 3)}$$

$$= (c-2)^2 \det \begin{bmatrix} 1 & -1 & 0 \\ 0 & c-6 & -2 \\ 0 & -1 & 1 \end{bmatrix} \begin{matrix} \scriptstyle \mathrm{II+2I} \\ \\ \end{matrix} = (c-2)^2(c-8).$$

Eigenspace for $c = 2$. The coefficient matrix of $(2I - A)\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} -2 & -2 & -2 \\ -2 & -2 & -2 \\ -2 & -2 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \scriptstyle \mathrm{II-I} \\ \scriptstyle \mathrm{III-I} \\ \scriptstyle \mathrm{I}\times -\frac{1}{2} \end{matrix}.$$

A basis of the eigenspace is $\boldsymbol{w}_1 = (-1, 1, 0), \boldsymbol{w}_2 = (-1, 0, 1)$. Let

$$\begin{aligned} \boldsymbol{v}_1 = \boldsymbol{w}_1, \boldsymbol{v}_2 &= \boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\, \boldsymbol{v}_1 \\ &= (-1, 0, 1) - \frac{1}{2}\,(-1, 1, 0) = \left(-\frac{1}{2}, -\frac{1}{2}, 1\right). \end{aligned}$$

This gives an orthogonal basis $(-1, 1, 0), (-1, -1, 2)$ of the eigenspace. Normalize to obtain an orthonormal basis $\boldsymbol{u}_1 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \boldsymbol{u}_2 = \left(-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)$.

Eigenspace for $c = 8$. The coefficient matrix of $(8I - A)\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & -2 \\ 0 & 6 & -6 \\ 0 & -6 & 6 \end{bmatrix} \begin{array}{l} \scriptstyle \mathrm{I} \leftrightarrow \mathrm{III} \\ \scriptstyle \mathrm{II} - \mathrm{I} \\ \scriptstyle \mathrm{III} + 2\mathrm{I} \\ \scriptstyle \mathrm{I} \times -\frac{1}{2} \end{array} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{array}{c} \scriptstyle \mathrm{III} + \mathrm{II} \\ \scriptstyle \mathrm{II} \times \frac{1}{6} \\ \scriptstyle \mathrm{I} - \mathrm{II} \end{array}.$$

A basis of the eigenspace is $\boldsymbol{u}_3 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$, which is orthogonal to $\boldsymbol{u}_1, \boldsymbol{u}_2$, as predicted in Proposition 7.

Proposition 9 tells us that when A is symmetric, there is a diagonalization

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

where the columns of P are an orthonormal basis of eigenvectors and D is diagonal. Of course P is an orthogonal matrix, so $P^{-1} = P^t$, and

$$A = PDP^t.$$

This equation (with D diagonal, P orthogonal) is an **orthogonal diagonalization** of A. In Example 14, we obtain

$$A = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 2 & & \\ & 2 & \\ & & 8 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Are there any nonsymmetric matrices that have an orthogonal diagonalization? The answer is a resounding no. If $A = PDP^t$ (P orthogonal, D diagonal), then

$$A^t = (PDP^t)^t = (P^t)^t D^t P^t = PDP^t = A,$$

and A is symmetric!

It is interesting that Proposition 9 contains the statement that the characteristic polynomial of a symmetric matrix has n real zeros, counted with multiplicity. If there were less than n real zeros, we would not be able to pick n linearly independent eigenvectors, because of Proposition 5 of Chapter 6. In the next chapter we examine this again, from the point of view of complex numbers.

5 Diagonalizing a quadratic form

Let $Q(\boldsymbol{y})$ be a quadratic form in $\boldsymbol{y} = (y_1, \ldots, y_n)$ and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be any basis of $\mathbb{R}^n$. Let

$$\boldsymbol{y} = S\boldsymbol{x}$$

be the change of coordinates corresponding to

$$y_1\boldsymbol{e}_1 + \cdots + y_n\boldsymbol{e}_n = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n. \tag{14}$$

If we write Q as a function of $\boldsymbol{x}$,

$$Q(\boldsymbol{y}) = (S\boldsymbol{x})^t AS\boldsymbol{x} = \boldsymbol{x}^t(S^tAS)\boldsymbol{x} = Q_1(\boldsymbol{x}), \text{ say.}$$

Of course S^tAS is symmetric, so the change of coordinates gives a quadratic form Q_1 with matrix S^tAS. Note that $Q_1(\boldsymbol{x}) = Q(S\boldsymbol{x})$.

Example 15 Let $\boldsymbol{v}_1 = (1, 3), \boldsymbol{v}_2 = (2, 5)$,

$$Q(y_1, y_2) = y_1^2 - 3y_1y_2 - y_2^2.$$

Then $S = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}, A = \begin{bmatrix} 1 & -3/2 \\ -3/2 & -1 \end{bmatrix}$. Thus $Q(\boldsymbol{y}) = Q_1(\boldsymbol{x})$ where Q_1 has matrix

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & -3/2 \\ -3/2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} -17 & -59/2 \\ -59/2 & -51 \end{bmatrix}.$$

In other words,

$$Q_1(x_1, x_2) = -17x_1^2 - 59x_1x_2 - 51x_2^2.$$

You should check that this expression matches

$$Q(S\boldsymbol{x}) = (x_1 + 2x_2)^2 - 3(x_1 + 2x_2)(3x_1 + 5x_2) - (3x_1 + 5x_2)^2.$$