THE EIGENVALUE PROBLEM

Let A be an $n \times n$ square matrix. If there is a number λ and a column vector $v \neq 0$ for which

$$Av = \lambda v$$

then we say λ is an *eigenvalue* of A and v is an associated *eigenvector*. Note that if v is an eigenvector, then any nonzero multiple of v is also an eigenvector for the same eigenvalue λ.

Example: Let

$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix} \tag{1}$$

The eigenvalue-eigenvector pairs for A are

$$\begin{aligned} \lambda_1 = 2, &\qquad v^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \lambda_2 = 0.5, &\qquad v^{(2)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned} \tag{2}$$

Eigenvalues and eigenvectors are often used to give additional intuition to the function

$$F(x) = Ax, \qquad x \in \mathbb{R}^n \text{ or } \mathbb{C}^n$$

Example. The eigenvectors in the preceding example (2) form a *basis* for $\mathbb{R}^2$. For $x = [x_1, x_2]^T$,

$$\begin{aligned} x &= c_1 v^{(1)} + c_2 v^{(2)} \\ c_1 &= \frac{x_1 + x_2}{2}, \quad c_2 = \frac{x_2 - x_1}{2} \end{aligned}$$

Using (2), the function

$$F(x) = Ax, \qquad x \in \mathbb{R}^2$$

can be written as

$$\begin{aligned} F(x) &= c_1 A v^{(1)} + c_2 A v^{(2)} \\ &= 2 c_1 v^{(1)} + \frac{1}{2} c_2 v^{(2)}, \quad x \in \mathbb{R}^2 \end{aligned}$$

The eigenvectors provide a better way to understand the meaning of $F(x) = Ax$.

See the following Figure 1 and Figure 2.

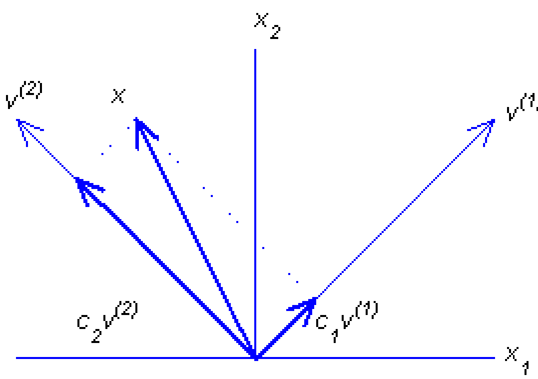

Figure 1: Decomposition of x

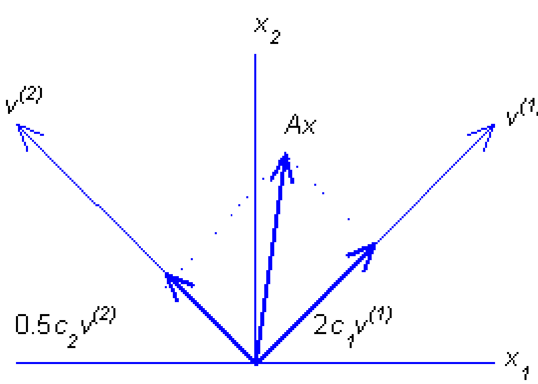

Figure 2: Decomposition of Ax

How to calculate λ and v? Rewrite $Av = \lambda v$ as

$$(\lambda I - A)v = 0, \qquad v \neq 0 \tag{3}$$

a homogeneous system of linear equations with the coefficient matrix $\lambda I - A$ and the nonzero solution v. This can be true if and only if

$$f(\lambda) \equiv \det(\lambda I - A) = 0$$

The function $f(\lambda)$ is called the *characteristic polynomial* of A, and its roots are the eigenvalues of A.
Assuming A has order n,

$$f(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0 \tag{4}$$

For the case $n = 2$,

$$\begin{aligned} f(\lambda) &= \det \begin{bmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{bmatrix} \\ &= (\lambda - a_{11})(\lambda - a_{22}) - a_{21}a_{12} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{21}a_{12} \end{aligned}$$

The formula (4) shows that a matrix A of order n can have at most n distinct eigenvalues.

Example. Let

$$A = \begin{bmatrix} -7 & 13 & -16 \\ 13 & -10 & 13 \\ -16 & 13 & -7 \end{bmatrix} \tag{5}$$

Then

$$\begin{aligned} f(\lambda) &= \det \begin{bmatrix} \lambda + 7 & -13 & 16 \\ -13 & \lambda + 10 & -13 \\ 16 & -13 & \lambda + 7 \end{bmatrix} \\ &= \lambda^3 + 24\lambda^2 - 405\lambda + 972 \end{aligned}$$

is the characteristic polynomial of A.

The roots are

$$\lambda_1 = -36, \qquad \lambda_2 = 9, \qquad \lambda_3 = 3 \tag{6}$$

Finding an eigenvector. For $\lambda = -36$, find an associated eigenvector v by solving $(\lambda I - A)v = 0$, which becomes

$$(-36I - A)v = 0$$

$$\begin{bmatrix} -29 & -13 & 16 \\ -13 & -26 & -13 \\ 16 & -13 & -29 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

If $v_1 = 0$, then the only solution is $v = 0$. Thus $v_1 \neq 0$, and we arbitrarily choose $v_1 = 1$. This leads to the system

$$\begin{aligned} -13v_2 + 16v_3 &= 29 \\ -26v_2 - 13v_3 &= 13 \\ -13v_2 - 29v_3 &= -16 \end{aligned}$$

The solution is $v_2 = -1$, $v_3 = 1$. Thus the eigenvector v for $\lambda = -36$ is

$$v^{(1)} = [1, -1, 1]^T \tag{7}$$

or any nonzero multiple of this vector.

SYMMETRIC MATRICES

Recall that A *symmetric* means $A^{\mathsf{T}} = A$, assuming that A contains only real number entries. Such matrices are very common in applications.

Example. A general 3×3 symmetric matrix looks like

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

A general $n \times n$ matrix $A = \left[a_{i,j}\right]$ is symmetric if and only if

$$a_{i,j} = a_{j,i}, \qquad 1 \le i, j \le n$$

Following is a very important set of results for symmetric matrices, explaining much of their special character.

THEOREM. Let A be a real, symmetric, $n \times n$ matrix. Then there is a set of n eigenvalue-eigenvector pairs $\{\lambda_i, v^{(i)}\}$, $1 \leq i \leq n$ that satisfy the following properties.

(i) The numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all of the roots of the characteristic polynomial $f(\lambda)$ of A, repeated according to their multiplicity. Moreover, all the λ_i are real numbers.

(ii) If the column matrices $v^{(i)}$, $1 \leq i \leq n$ are regarded as vectors in n-dimensional space, then they are mutually perpendicular and of length 1:

$$\begin{aligned} v^{(i)\mathsf{T}} v^{(j)} &= 0, \qquad 1 \leq i, j \leq n, \quad i \neq j \\ v^{(i)\mathsf{T}} v^{(i)} &= 1, \qquad 1 \leq i \leq n \end{aligned}$$

(iii) For each column vector $x = [x_1, x_2, \ldots, x_n]^{\mathsf{T}}$, there is a unique choice of constants $c_1, \ldots, c_n$ for which

$$x = c_1 v^{(1)} + \cdots + c_n v^{(n)}$$

The constants are given by

$$c_i = \sum_{j=1}^{n} x_j v_j^{(i)} = x^{\mathsf{T}} v^{(i)}, \qquad 1 \le i \le n$$

where $v^{(i)} = [v_1^{(i)}, \ldots, v_n^{(i)}]^{\mathsf{T}}$.

(iv) Define the matrix U of order n by

$$U = [v^{(1)}, v^{(2)}, \ldots, v^{(n)}] \tag{8}$$

Then

$$U^{\mathsf{T}} A U = D \equiv \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

and

$$U U^{\mathsf{T}} = U^{\mathsf{T}} U = I \tag{9}$$

$A = U D U^{\mathsf{T}}$ is a useful decomposition of A.

Example. Recall

$$A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$$

and its eigenvalue-eigenvector pairs

$$\lambda_1 = 2, \qquad v^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 0.5, \qquad v^{(2)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Figure 1 illustrates that the eigenvectors are perpendicular. To have them be of length 1, replace the above by

$$\lambda_1 = 2, \qquad v^{(1)} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\lambda_2 = 0.5, \qquad v^{(2)} = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

that are multiples of the original $v^{(i)}$.

The matrix U of (8) is given by

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

Easily, $U^{\mathsf{T}}U = I$.

Also,

$$\begin{aligned} & U^{\mathsf{T}}AU \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \end{aligned}$$

as specified in (9).

NONSYMMETRIC MATRICES

Nonsymmetric matrices have a wide variety of possible behaviour. We illustrate with two simple examples some of the possible behaviour.

Example. We illustrate the existence of complex eigenvalues. Let

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

The characteristic polynomial is

$$f(\lambda) = \det \begin{bmatrix} \lambda & -1 \\ 1 & \lambda \end{bmatrix} = \lambda^2 + 1$$

The roots of $f(\lambda)$ are complex,

$$\lambda_1 = i \equiv \sqrt{-1}, \qquad \lambda_2 = -i$$

and corresponding eigenvectors are

$$v^{(1)} = \begin{bmatrix} i \\ -1 \end{bmatrix}, \qquad v^{(2)} = \begin{bmatrix} i \\ 1 \end{bmatrix}$$

Example. For A an $n \times n$ nonsymmetric matrix, there may not be n independent eigenvectors. Let

$$A = \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$$

where a is a constant.

Then $\lambda = a$ is the eigenvalue of A with multiplicity 3, and any associated eigenvector must be of the form

$$v = c\,[1,\ 0,\ 0]^T$$

for some $c \neq 0$.

Thus, up to a nonzero multiplicative constant, we have only one eigenvector,

$$v = [1,\ 0,\ 0]^T$$

for the three equal eigenvalues $\lambda_1 = \lambda_2 = \lambda_3 = a$.

THE POWER METHOD

This numerical method is used primarily to find the eigenvalue of largest magnitude, if such exists.

We assume that the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of an $n \times n$ matrix A satisfy

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n| \tag{10}$$

Denote the eigenvector for λ_1 by $v^{(1)}$. We define an iteration method for computing improved estimates of λ_1 and $v^{(1)}$.

Choose $z^{(0)} \approx v^{(1)}$, usually chosen randomly. Define

$$w^{(1)} = Az^{(0)}$$

Let α_1 be the maximum component of $w^{(1)}$, in size. If there is more than one such component, choose the first such component as α_1. Then define

$$z^{(1)} = \frac{1}{\alpha_1} w^{(1)}$$

Repeat the process iteratively. Define

$$w^{(m)} = Az^{(m-1)} \tag{11}$$

Let α_m be the maximum component of $w^{(m)}$, in size. Define

$$z^{(m)} = \frac{1}{\alpha_m} w^{(m)} \tag{12}$$

for $m = 1, 2, \ldots$ Then, roughly speaking, the vectors $z^{(m)}$ will converge to some multiple of $v^{(1)}$.

To find λ_1 by this process, also pick some nonzero component of the vectors $z^{(m)}$ and $w^{(m)}$, say component k; and fix k. Often this is picked as the maximal component of $z^{(l)}$, for some large l. Define

$$\lambda_1^{(m)} = \frac{w_k^{(m)}}{z_k^{(m-1)}}, \qquad m = 1, 2, \ldots \tag{13}$$

where $z_k^{(m-1)}$ denotes component k of $z^{(m-1)}$.

It can be shown that $\lambda_1^{(m)}$ converges to λ_1 as $m \to \infty$.

Example. Recall the earlier example $A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$.
Double precision was used in the computation, with rounded values shown in the table given here.

Table 1: Power method example

m	$z_1^{(m)}$	$z_2^{(m)}$	$\lambda_1^{(m)}$	$\lambda_1^{(m)} - \lambda_1^{(m-1)}$	*Ratio*
0	1.0	.5			
1	1.0	.84615	1.62500		
2	1.0	.95918	1.88462	$2.60E-1$	
3	1.0	.98964	1.96939	$8.48E-2$	0.33
4	1.0	.99740	1.99223	$2.28E-2$	0.27
5	1.0	.99935	1.99805	$5.82E-3$	0.26
6	1.0	.99984	1.99951	$1.46E-3$	0.25
7	1.0	.99996	1.99988	$3.66E-4$	0.25

Note that in this example, $\lambda_1 = 2$ and $v^{(1)} = [1,1]^T$, and the numerical results are converging to these values.
The column of successive differences $\lambda_1^{(m)} - \lambda_1^{(m-1)}$ and their successive ratios are included to show that there is a regular pattern to the convergence.

CONVERGENCE

Assume A is a real $n \times n$ matrix.

It can be shown, by induction, that

$$z^{(m)} = \sigma_m \cdot \frac{A^m z^{(0)}}{\|A^m z^{(0)}\|}, \qquad m \geq 1 \tag{14}$$

where $\sigma_m = \pm 1$.

Further assume that the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ satisfy (10), and also that there are n corresponding eigenvectors $\left\{v^{(1)}, \ldots, v^{(n)}\right\}$ that form a basis for $\mathbb{R}^n$.

Thus

$$z^{(0)} = c_1 v^{(1)} + \cdots + c_n v^{(n)} \tag{15}$$

for some choice of constants $\{c_1, \ldots, c_n\}$.

Assume $c_1 \neq 0$, something that a truly random choice of $z^{(0)}$ will usually guarantee.

Apply A to $z^{(0)}$ in (15), to get

$$\begin{aligned} Az^{(0)} &= c_1 Av^{(1)} + \cdots + c_n Av^{(n)} \\ &= \lambda_1 c_1 v^{(1)} + \cdots + \lambda_n c_n v^{(n)} \end{aligned}$$

Apply A repeatedly to get

$$\begin{aligned} &A^m z^{(0)} = \lambda_1^m c_1 v^{(1)} + \cdots + \lambda_n^m c_n v^{(n)} \\ &= \lambda_1^m \left[c_1 v^{(1)} + \left(\frac{\lambda_2}{\lambda_1}\right)^m c_2 v^{(2)} + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^m c_n v^{(n)} \right] \end{aligned}$$

From (14),

$$\begin{aligned} &z^{(m)} = \sigma_m \left(\frac{\lambda_1}{|\lambda_1|}\right)^m \times \\ &\qquad \frac{c_1 v^{(1)} + \left(\frac{\lambda_2}{\lambda_1}\right)^m c_2 v^{(2)} + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^m c_n v^{(n)}}{\left\| c_1 v^{(1)} + \left(\frac{\lambda_2}{\lambda_1}\right)^m c_2 v^{(2)} + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^m c_n v^{(n)} \right\|} \end{aligned} \tag{16}$$

As $m \to \infty$, the terms $(\lambda_j/\lambda_1)^m \to 0$, for $2 \le j \le n$, with $(\lambda_2/\lambda_1)^m$ the largest. Also,

$$\sigma_m \left(\frac{\lambda_1}{|\lambda_1|}\right)^m = \pm 1$$

Thus as $m \to \infty$, most terms in (16) are converging to zero. Cancel c_1 from numerator and denominator, obtaining

$$z^{(m)} \approx \hat{\sigma}_m \frac{v^{(1)}}{\|v^{(1)}\|}$$

where $\hat{\sigma}_m = \pm 1$.

If the normalization of $z^{(m)}$ is modified, to always have some particular component be positive, then

$$z^{(m)} \to \pm \frac{v^{(1)}}{\left\|v^{(1)}\right\|} \equiv \hat{v}^{(1)} \tag{17}$$

with a fixed sign independent of m. Our earlier normalization of sign, dividing by α_m, will usually accomplish this, but not always.

The error in $z^{(m)}$ will satisfy

$$\|z^{(m)} - \hat{v}^{(1)}\| \le c \left|\frac{\lambda_2}{\lambda_1}\right|^m, \qquad m \ge 0 \tag{18}$$

for some constant $c > 0$.

CONVERGENCE OF $\lambda_1^{(m)}$

A similar convergence analysis can be given for $\left\{\lambda_1^{(m)}\right\}$, with the same kind of error bound.

Moreover, if we also assume

$$|\lambda_1| > |\lambda_2| > |\lambda_3| \geq |\lambda_4| \geq \cdots \geq |\lambda_n| \geq 0 \qquad (19)$$

then

$$\lambda_1 - \lambda_1^{(m)} \approx c\left(\frac{\lambda_2}{\lambda_1}\right)^m \qquad (20)$$

for some constant c, as $m \to \infty$.

The error decreases geometrically with a ratio of λ_2/λ_1. In the earlier example with $A = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{bmatrix}$,

$$\lambda_2/\lambda_1 = 0.5/2 = .25$$

which is the ratio observed in Table 1.

Example. Consider the symmetric matrix

$$A = \begin{bmatrix} -7 & 13 & -16 \\ 13 & -10 & 13 \\ -16 & 13 & -7 \end{bmatrix}$$

From earlier, the eigenvalues are

$$\lambda_1 = -36, \qquad \lambda_2 = 9, \qquad \lambda_3 = 3$$

The eigenvector $v^{(1)}$ associated with λ_1 is

$$v^{(1)} = [1, -1, 1]^T$$

The results of using the power method are shown in the following Table 2.

Note that the ratios of the successive differences of $\lambda_1^{(m)}$ are approaching

$$\frac{\lambda_2}{\lambda_1} = -0.25$$

Also note that location of the maximal component of $z^{(m)}$ changes from one iteration to the next.

The initial guess $z^{(0)}$ was chosen closer to $v^{(1)}$ than would be usual in actual practice. It was so chosen for purposes of illustration.

Table 2: Power method example

m	$z_1^{(m)}$	$z_2^{(m)}$	$z_3^{(m)}$
0	1.000000	-0.800000	0.900000
1	-0.972477	1.000000	-1.000000
2	1.000000	-0.995388	0.993082
3	0.998265	-0.999229	1.000000
4	1.000000	-0.999775	0.999566
5	0.999891	-0.999946	1.000000
6	1.000000	-0.999986	0.999973
7	0.999993	-0.999997	1.000000

m	$\lambda_1^{(m)}$	$\lambda_1^{(m)} - \lambda_1^{(m-1)}$	*Ratio*
1	-31.80000		
2	-36.82075	$-5.03E+0$	
3	-35.82936	$9.91E-1$	-0.197
4	-36.04035	$-2.11E-1$	-0.213
5	-35.99013	$5.02E-2$	-0.238
6	-36.00245	$-1.23E-2$	-0.245
7	-35.99939	$3.06E-3$	-0.249

AITKEN EXTRAPOLATION

From (20),

$$\lambda_1 - \lambda_1^{(m+1)} \approx r(\lambda_1 - \lambda_1^{(m)}), \qquad r = \lambda_2/\lambda_1 \quad (21)$$

for large m. Choose r using

$$r \approx \frac{\lambda_1^{(m+1)} - \lambda_1^{(m)}}{\lambda_1^{(m)} - \lambda_1^{(m-1)}} \quad (22)$$

as with Aitken extrapolation in § 3.4 on linear iteration methods.

Using this r, solve for λ_1 in (21),

$$\begin{aligned} \lambda_1 &\approx \frac{1}{1-r}\left[\lambda_1^{(m+1)} - r\lambda_1^{(m)}\right] \\ &= \lambda_1^{(m+1)} + \frac{r}{1-r}\left[\lambda_1^{(m+1)} - \lambda_1^{(m)}\right] \end{aligned} \quad (23)$$

This is *Aitken's extrapolation formula*. It also gives us the *Aitken error estimate*

$$\lambda_1 - \lambda_1^{(m+1)} \approx \frac{r}{1-r}\left[\lambda_1^{(m+1)} - \lambda_1^{(m)}\right] \quad (24)$$

Example. In Table 2, take $m+1=7$. Then (24) yields

$$\lambda_1 - \lambda_1^{(7)} \approx \frac{r}{1-r}\left[\lambda_1^{(7)} - \lambda_1^{(6)}\right]$$
$$= \frac{-0.249}{1+0.249}[0.00306] \doteq -0.00061$$

which is the actual error.

Also the Aitken formula (23) will give the exact answer for λ_1, to seven significant digits.