

Quick Review of Eigenvalues and Eigenvectors

The relationship between errors in measurement and corresponding errors in parameters estimated using those measurements often depends on properties of a matrix — in particular the eigenvalues and eigenvectors of that matrix. Here is a quick review of some of the relevant facts.

Consider an $n \times n$ symmetric real matrix M . The vector \mathbf{e} is an eigenvector with eigenvalue λ if

$$M\mathbf{e} = \lambda\mathbf{e}$$

Note that if \mathbf{e} is an eigenvector then so is $k\mathbf{e}$ for any non-zero k . Hence the magnitude of these vectors is of no significance, and we may as well use unit vectors when it is useful to do so.

To solve the above equation for the eigenvectors and eigenvalues of the matrix M , we can rewrite it in the form

$$(M - \lambda I)\mathbf{e} = \mathbf{0}$$

where I is the $n \times n$ identity matrix and $\mathbf{0}$ is a vector of zeroes.

This set of linear equations is homogeneous since the constant term in each equation is zero. If we can invert $(M - \lambda I)$, then we can multiply this inverse by $\mathbf{0}$ to obtain the “trivial” solution $\mathbf{0}$ for \mathbf{e} . The only way non-zero solutions are possible is if the matrix is singular, that is, when

$$\det(M - \lambda I) = 0$$

The determinant can be written as a sum of terms, each of which is a product of matrix elements, one from each row. Since λ occurs in the elements on the diagonal, terms of up to order n in λ may occur in these products. So the determinant is a polynomial of order n in λ . This polynomial will have n roots. We will assume that these roots are distinct.

As a simple illustration, consider the 2×2 real symmetric matrix

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Its eigenvalues can be found by finding the roots of

$$\det \begin{pmatrix} a - \lambda & b \\ b & c - \lambda \end{pmatrix} = 0$$

that is,

$$\lambda^2 - (a + c)\lambda + (ac - b^2) = 0$$

The two roots are,

$$\lambda_{+,-} = \frac{(a + c) \pm \sqrt{(a - c)^2 + 4b^2}}{2}$$

or, if we let $d = \sqrt{(a - c)^2 + 4b^2}$, then

$$\lambda_{+,-} = \frac{(a + c) \pm d}{2}$$

If λ_+ and λ_- are the values with the + and – sign respectively in front of the square-root, then $\lambda_+ > \lambda_-$.

Corresponding to each eigenvalue λ_i will be an eigenvector \mathbf{e}_i obtained by solving the set of n homogeneous linear equations

$$(M - \lambda_i I)\mathbf{e}_i = 0$$

One way to find an eigenvector is to note that it must be orthogonal to each of the rows of the matrix $(M - \lambda_i I)$.

Again, as a simple illustration, consider the 2×2 real symmetric matrix above. We have to solve

$$\begin{pmatrix} a - \lambda & b \\ b & c - \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

A vector orthogonal to the first row is given by

$$\begin{pmatrix} -b \\ a - \lambda \end{pmatrix}$$

The two eigenvectors can now be determined by substituting the values λ_+ and λ_- determined above. Similarly, a vector orthogonal to the second row is

$$\begin{pmatrix} c - \lambda \\ -b \end{pmatrix}$$

The two apparently different directions for the eigenvector are actually the same if λ is an eigenvalue, since then $b/(a - \lambda) = (c - \lambda)/b$.

The eigenvectors can thus be written in either of the forms

$$\begin{pmatrix} 2b \\ (c - a) \pm d \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} (a - c) \pm d \\ 2b \end{pmatrix}$$

The magnitude squared of the first form is $2d(d \pm (c - a))$, while the magnitude squared of the second is $2d(d \pm (a - c))$. The unit eigenvectors can thus be written in either of the forms:

$$\frac{1}{\sqrt{2d}\sqrt{d \pm (c - a)}} \begin{pmatrix} \pm 2b \\ d \pm (c - a) \end{pmatrix}$$

or

$$\frac{1}{\sqrt{2d}\sqrt{d \pm (a - c)}} \begin{pmatrix} d \pm (a - c) \\ \pm 2b \end{pmatrix}$$

In each occurrence of \pm the plus sign is chosen when the plus sign is chosen in the expression for the eigenvalue, while the minus sign is chosen when the minus sign is chosen in the expression for the eigenvalue. Any non-zero multiple of these vectors is, of course, also an eigenvector (Note that the two forms happen to give opposite directions when $b < 0$).

Next, consider

$$(M\mathbf{e}_1) \cdot \mathbf{e}_2 = (\mathbf{e}_1^T M^T \mathbf{e}_2)^T = (\mathbf{e}_2^T M \mathbf{e}_1) = (\mathbf{e}_2^T M^T \mathbf{e}_1) = (M\mathbf{e}_2) \cdot \mathbf{e}_1$$

Now

$$(M\mathbf{e}_1) \cdot \mathbf{e}_2 = \lambda_1 \mathbf{e}_1 \cdot \mathbf{e}_2$$

while

$$(M\mathbf{e}_2) \cdot \mathbf{e}_1 = \lambda_2 \mathbf{e}_2 \cdot \mathbf{e}_1$$

Consequently,

$$\lambda_1 \mathbf{e}_1 \cdot \mathbf{e}_2 = \lambda_2 \mathbf{e}_1 \cdot \mathbf{e}_2$$

or

$$(\lambda_1 - \lambda_2) \mathbf{e}_1 \cdot \mathbf{e}_2 = 0$$

which implies that $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$, if $\lambda_1 \neq \lambda_2$. That is, eigenvectors corresponding to distinct eigenvalues are orthogonal.

The above argument does not guarantee orthogonality of eigenvectors if a root of the polynomial has multiplicity two. In this case, any vector in a plane is an eigenvector, and we can arbitrarily pick two orthogonal vectors in that plane as eigenvectors. The same method can be used with roots of higher multiplicity to obtain a set of mutually orthogonal eigenvectors.

Since the n eigenvectors of an $n \times n$ real symmetric matrix are orthogonal to one another, they form a basis, and can be used to express an arbitrary vector \mathbf{v} in the form

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{e}_i$$

for some set of α_i s. Now

$$\mathbf{v} \cdot \mathbf{e}_i = \sum_{j=1}^n \alpha_j (\mathbf{e}_j \cdot \mathbf{e}_i) = \alpha_i$$

since $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ unless $i = j$. We conclude that $\alpha_i = \mathbf{v} \cdot \mathbf{e}_i$. Next we see that

$$M\mathbf{v} = \sum_{i=1}^n \alpha_i M\mathbf{e}_i = \sum_{i=1}^n \alpha_i \lambda_i \mathbf{e}_i$$

We see that the eigenvalue λ_i specifies how much a component of the vector \mathbf{v} in the direction of the corresponding eigenvector \mathbf{e}_i is “amplified” when multiplied by the matrix M .

Next, note that

$$\mathbf{v} = \sum_{i=1}^n (\mathbf{v} \cdot \mathbf{e}_i) \mathbf{e}_i = \sum_{i=1}^n \mathbf{e}_i (\mathbf{e}_i \cdot \mathbf{v}) = \left(\sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T \right) \mathbf{v}$$

We conclude that we can write the identity matrix in terms of a sum of dyadic products of eigenvectors:

$$I = \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T$$

We can write the matrix M also in terms of a sum of dyadic products:

$$M = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T$$

This can be verified by multiplying M by each of the eigenvectors \mathbf{e}_j :

$$M \mathbf{e}_j = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T \mathbf{e}_j = M \mathbf{e}_j = \sum_{i=1}^n \lambda_i \mathbf{e}_i (\mathbf{e}_i \cdot \mathbf{e}_j) = \lambda_j \mathbf{e}_j$$

since $(\mathbf{e}_i \cdot \mathbf{e}_j) = 0$ unless $i = j$.

Finally, we can write the inverse of the matrix M also in terms of a sum of dyadic products:

$$M^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i^T$$

This expression can be verified by premultiplying it by M .

$$MM^{-1} = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T \sum_{j=1}^n \frac{1}{\lambda_j} \mathbf{e}_j \mathbf{e}_j^T = \sum_{i=1}^n \sum_{j=1}^n \frac{\lambda_i}{\lambda_j} \mathbf{e}_i \mathbf{e}_i^T \mathbf{e}_j \mathbf{e}_j^T = \sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T$$

which is the expression for the identity matrix I given above. The same result is obtained when postmultiplying by M .

If computation of estimates of unknown parameters depends on solving linear equations, then clearly small eigenvalues of the corresponding coefficient matrix present a problem since they lead to large amplification of measurement error.

Importantly, error amplification may differ in different directions of parameter space, with results along some directions sometimes being quite stable even when results along other directions are not.

In the case of optical flow, for example, the component of motion in the direction of the brightness gradient is locally well defined, while the component of motion along the isophotes is not. This can be shown by analysis of the eigenvectors and eigenvalues of the 2×2 matrix

$$\begin{pmatrix} \iint E_x^2 dx dy & \iint E_x E_y dx dy \\ \iint E_y E_x dx dy & \iint E_y^2 dx dy \end{pmatrix}$$

when $E_y : E_x$ is constant in the region of integration.