COMS10003 : Linear Algebra

Eigenvalues and Eigenvectors

Andrew Calway

April 27, 2015

Introduction

In this final section on linear algebra we are going to look at eigenvalues and eigenvectors. The latter are special: they are those vectors, associated with a given matrix, which after being multiplied by the said matrix give a vector in the same direction, i.e. a scalar multiple of itself. In other words, if we think of a matrix as defining a linear transformation, then eigenvectors are those vectors which don't change as a result of the transformation, bar a scale factor. The scale factor is in fact the eigenvalue associated with the eigenvector. It turns out that these special vectors turn up in many places in computer science and so again it is important that you have a good understanding of the fundamentals and the language we use to describe them. As always, I have made use of several textbooks from my bookshelves and these are listed below.

Theory and problems of linear algebra by Seymour Lipschutz, McGraw-Hill,1981.

Linear Algebra and Probability for CS Applications by Ernest Davis, CRC Press, 2012.

Coding the Matrix by Philip N Klein, Newtonian Press, 2013.

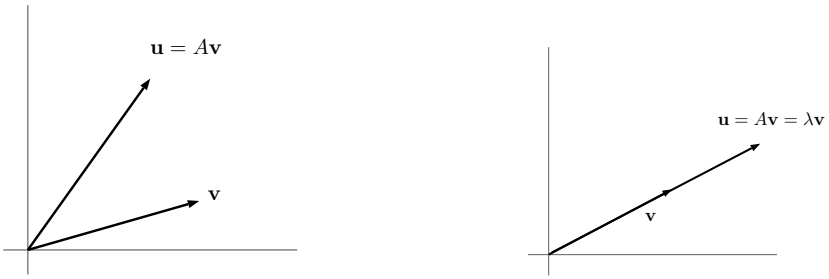
Eigenvalues and Eigenvectors

In general the application of a linear transformation, respresented by a matrix A say, results in a new vector in a different direction, e.g. $\mathbf{u} = A\mathbf{v}$, as illustrated in Fig. 1a for the 2-D case. However, for a given transformation, there can exist vectors which transform to a vector in the same direction, i.e. for the scalar λ, $A\mathbf{v} = \lambda\mathbf{v}$, as shown in Fig. 1b. The scalars λ and vectors $\mathbf{v}$ are known as *eigenvalues* and *eigenvectors*, respectively, of the matrix A.

Note that if $\mathbf{v}$ is an eigenvector of A then all scalar multiplies of $\mathbf{v}$ are also eigenvectors, i.e. for some scalar a

$$A\mathbf{v} = \lambda\mathbf{v} \qquad\qquad A(a\mathbf{v}) = \lambda(a\mathbf{v})$$

t



(a) (b)

Figure 1: In general linear transformations result in a vector ina different direction as in (a) but for some special vectors, the direction is not changed by the transformation as in (b).

Hence for a given eigenvalue, there are an infinite number of possible eigenvectors all in the same direction.

Finding Eigenvalues and Eigenvectors

We need to find the scalars λ and the vectors $\mathbf{v}$ which satisfy $A\mathbf{v} = \lambda\mathbf{v}$ for some matrix A. Rewriting this as

$$A\mathbf{v} = \lambda I\mathbf{v}$$

where I is the identity matrix, then we get

$$(A - \lambda I)\mathbf{v} = 0$$

The trivial solution to this is $\mathbf{v} = 0$, i.e. the zero vector. However this isn't interesting. Hence we seek values of λ for which the above equation is satisfied for non-zero vectors $\mathbf{v}$. For this, we can see that the matrix $(A - \lambda I)$ can't have an inverse - if it did, then $\mathbf{v} = 0$, i.e.

$$(A - \lambda I)^{-1}(A - \lambda I)\mathbf{v} = \mathbf{v} = 0$$

We say that the matrix $(A - \lambda I)$ needs to be *singular*, i.e. not have an inverse.

Now let us consider the case of 2×2 matrices. Recall from the notes on "Matrices", that the inverse of a 2×2 matrix is given by

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

and that the matrix is singular if the determinant $|A| = ad - bc = 0$.

To find our eigenvalues we therefore require that

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0$$

and hence from the definition of the determinant we get

$$\begin{aligned}(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} &= \lambda^2 - \lambda(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}a_{21} \\ &= \lambda^2 - \lambda(a_{11} + a_{22}) + |A| = 0\end{aligned}$$

The eigenvalues are therefore the values of λ that are solutions to the above quadratic equation, i.e. the roots of the equation. The equation is known as the *characteristic polynomial*. It follows that there are two such eigenvalues which may be real or complex.

It turns out that we can also define determinants for matrices larger than 2×2 and for a $n\times n$ matrix this defines a characteristic polynomial of degree n and the n roots are the required eigenvalues. We shall look at the 3×3 case once we have done an example. Determinants of $n\times n$ matrices get somewhat involved - see `http://en.wikipedia.org/wiki/Determinant`.

Note that once we have the eigenvalues, then finding the eigenvectors is straightforward - we can go back to the equation $(A - \lambda I)\mathbf{v} = 0$ and solve for $\mathbf{v}$, i.e. in general, we will have n equations in n unknowns. Of course, we will only be able to find the relationship between the components of $\mathbf{v}$, which defines its direction, since there will be infinitely many such vectors all in the same direction.

Example

Find the eigenvalues and eigenvectors of the matrix

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

We have

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{vmatrix} = 0$$

giving the characteristic polynomial

$$(1-\lambda)(2-\lambda) - 6 = 0$$
$$\lambda^2 - 3\lambda - 4 = (\lambda - 4)(\lambda + 1) = 0$$

and hence the eigenvalues of A are $\lambda_1 = 4$ and $\lambda_2 = -1$.

To find the corresponding eigenvectors we substitute each eigenvalue into $(A - \lambda I)\mathbf{v} = 0$ and then solve for $\mathbf{v}$. For example, for $\lambda_1 = 4$ we get

$$\begin{bmatrix} 1-4 & 2 \\ 3 & 2-4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

giving

$$\begin{aligned} -3v_1 + 2v_2 &= 0 \\ 3v_1 - 2v_2 &= 0 \end{aligned}$$

Hence we can immediately see that vectors with $2v_2 = 3v_1$ will satisfy both equations and hence for this eigenvalue the eigenvectors are of the form $(2a, 3a)$ for scalars a.

Similarly, for $\lambda_2 = -1$, we get $v_1 = -v_2$ and hence eigenvectors of the form $(a, -a)$.

To summarise, to find eigenvalues and eigenvectors for the matrix A we perform the following steps

1. Compute the determinant $|A - \lambda I|$ to the characteristic polynomial.
2. Find the roots of the polynomial. These are the eigenvalues of A.
3. For each eigenvalue, solve $(A - \lambda I)\mathbf{v} = 0$ to give the corresponding eigenvectors.

Eigenvalues of 3×3 matrices

For a 3×3 matrix A, its determinant can be defined as follows

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + a_{i3}A_{i3} \tag{1}$$

where A_{ij} is known as a *cofactor*

$$A_{ij} = (-1)^{i+j}|M_{ij}|$$

and M_{ij} is the *minor* matrix which results from removing row i and column j from matrix A, e.g. M_{12} is

$$M_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}$$

We can select any row to compute the determinant using eqn (1) and in fact a similar rule applies for selecting a column. As before, if the determinant is zero, then the matrix A is singular and doesn't have an inverse - see the extra notes on *Determinants and Cramer's Rule*. Hence we can use this to find the eigenvalues and eigenvectors for a 3×3 matrix.

Example

For the matrix A

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$$

we need to find the roots of its characteristic polynomial

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & -3 & 3 \\ 3 & -5-\lambda & 3 \\ 6 & -6 & 4-\lambda \end{vmatrix} = (1-\lambda)A_{11} - 3A_{12} + 3A_{13} = 0$$

From above we have

$$A_{11} = \begin{vmatrix} -5-\lambda & 3 \\ -6 & 4-\lambda \end{vmatrix} \qquad A_{12} = -\begin{vmatrix} 3 & 3 \\ 6 & 4-\lambda \end{vmatrix} \qquad A_{13} = \begin{vmatrix} 3 & -5-\lambda \\ 6 & -6 \end{vmatrix}$$

giving

$$A_{11} = (-5-\lambda)(4-\lambda)+18 \qquad A_{12} = -3(4-\lambda)+18 \qquad A_{13} = -18-6(-5-\lambda)$$

The determinant of A is therefore

$$\begin{aligned} |A| &= (1-\lambda)((-5-\lambda)(4-\lambda)+18) - 3(-3(4-\lambda)+18) + 3(-18-6(-5-\lambda)) \\ &= (\lambda+2)^2(\lambda-4) \end{aligned}$$

and hence the eigenvalues of A are $\lambda_1 = -2$ and $\lambda = 4$. Substituting into $(A-\lambda I)\mathbf{v} = 0$ for $\lambda = -2$ gives

$$\begin{bmatrix} 3 & -3 & 3 \\ 3 & -3 & 3 \\ 6 & -6 & 6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Note that the 3 rows of the matrix in the above equation are all dependent, giving eigenvectors such that $v_1 - v_2 + v_3 = 0$. There are 2 degrees of freedom here and so all the eigenvectors lie in a plane - we say that the eigenvalue $\lambda_1 = -2$ is *of order 2*, which we can spot from the $(\lambda+2)^2$ term in the characteristic polynomial. We can choose an arbitrary basis for the plane by selecting two independent vectors which satisfy $v_1 - v_2 + v_3 = 0$ and hence get a general expression for the eigenvectors, e.g.

$$\mathbf{v} = a\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

For the other eigenvalue $\lambda = 4$ we get

$$\begin{bmatrix} -3 & -3 & 3 \\ 3 & -9 & 3 \\ 6 & -6 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which has the solutions $\mathbf{v} = a(1,1,2)$.

Independence of Eigenvectors

An important property of eigenvectors is the following:

> Non-zero eigenvectors corresponding to different eigenvalues of a matrix A are linearly independent, i.e. if $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ are the n eigenvectors for eigenvalues λ_1, λ_2, ..., λ_n, then $\mathbf{v}_i.\mathbf{v}_j = 0$ for all $i = j$.

Proof

To show the above, consider taking the weighted sum of two of the eigenvectors. If they are dependent then we will be able to find non-zero weights such that the sum is equal to zero, i.e.

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = 0$$

Let's now apply matrix A to the left-hand side to give

$$c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 = 0$$

Multiplying the first of the above two equations by λ_2 and substracting from the second equation gives

$$c_1(\lambda_1 - \lambda_2)\mathbf{v}_1 = 0$$

Since λ_1 and λ_2 are different and $\mathbf{v}_1$ is non-zero, then $c_1 = 0$ and hence from the first equation, $c_2 = 0$, proving that the two eigenvectors must be linearly independent. We can apply the same argument to any pair of eigenvectors, showing that they are all linearly independent.

Diagonalisation, Similarity and Powers of Matrices

We now look at important application of eigenvalue analysis - diagonalisation. Suppose an $n \times n$ matrix A has n independent eigenvectors, then if these vectors are used as the columns of a matrix P, the matrix $P^{-1}AP$ is a diagonal matrix whose diagonal components are the eigenvalues of A, i.e.

$$B = P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

where λ_i is the eigenvalue corresponding to the eigenvector in the ith column of P. Note that P must have an inverse since its n columns are independent. We say that the matrix P diagonalises A and the process is sometimes known as *eigendecomposition*.

Proof

Let $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ be the eigenvectors of A, so P is given by

$$P = [\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \dots \;\; \mathbf{v}_n]$$

Multiplying P by A then gives

$$\begin{aligned} AP &= [\lambda_1\mathbf{v}_1 \;\; \lambda_2\mathbf{v}_2 \;\; \dots \;\; \lambda_n\mathbf{v}_n] \\ &= [\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \dots \;\; \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix} = PB \end{aligned}$$

and hence that $P^{-1}AP = P^{-1}PB = B$.

Given the above relationship, the matrices A and B are said to be *similar*. This derives from the fact that they represent the *same linear transformation* but with respect to *different basis*, where B represents the transformation w.r.t the basis formed by the eigenvectors of A. Refer to the textbooks for more details.

One particular use of diagonalisation is that it helps us to easily form powers of $n \times n$ matrices. For example, if the matrix A can be diagonalised such that $A = PBP^{-1}$, then A to the power of 2 is

$$A^2 = AA = (PBP^{-1})(PBP^{-1}) = PBBP^{-1}$$

since $P^{-1}P = I$ and hence in general $A^k = PB^kP^{-1}$. Why is this useful? Because the kth power of the diagonal matrix B is easy to compute, i.e.

$$B^k = \begin{bmatrix} \lambda_1^k & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^k & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^k & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$$

thus reducing the kn^3 multiplications needed to compute A^k the 'long way' to $kn + n^2 + n^3$ multiplications, i.e. for large n a reduction by a factor of k. You should convince yourself of this.

We now take a quick look at two practical applications of eigenvalue analysis from computer science - *principal component analysis* (PCA) and the Pagerank algorithm which underlies the way Google ranks web pages when returning search results.

Principal Component Analysis

We shall introduce the ideas behind PCA by considering a simple 2-D example. Imagine we are given a set of 2-D points with average $(0, 0)$ such as that illustrated in Fig. 2. We would like to achieve *dimensionality reduction* by representing the points by a single number but in such a way that we could reconstruct the points with minimum error, i.e. an 'optimal' 1-D representation.

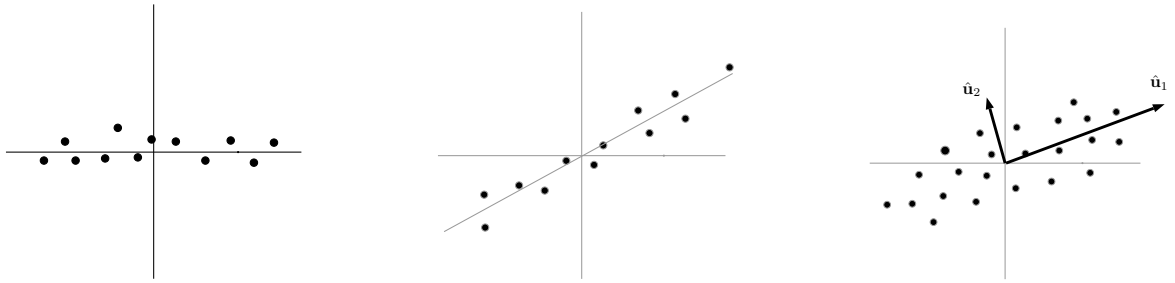
For example, in the case of Fig. 2a, we could use the x coordinates of the points and we would do quite well. But this is not the case for Fig. 2b - in that case we would be better to represent the points by their projection onto the line which passes through the origin at an angle θ.

As the above suggests, for any set of points, the 'best' 1-D representation is the projection of the points onto a line through the origin such that the sum of the squared errors between the projections and the original points is minimised. If we denote the line by a unit vector $\mathbf{u}$, then this can be expressed as

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \sum_{i=1}^{n} |\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{u})\mathbf{u}|^2 = \arg\min_{\mathbf{u}} \sum_{i=1}^{n} (\mathbf{x}_i^T\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{u})^2)$$

As we are interested in finding the best $\mathbf{u}$ we only need to worry about the 2nd term on the right-hand side and so the above is the same as

t



(a) (b) (c)

Figure 2: Simple 2-D PCA. The points in (a) would be well represented by their x coordiantes, whilst those in (b) would be better represented by their projection onto the line at some angle through the origin. In general, projections onto the eigenvector corresponding to the largest eigenvalue of the covariance matrix of the points, $\mathbf{u}_1$ in (c), give the optimal representation. These and the projections onto the other eigenvector $\mathbf{u}_2$ are the *principal components* of the data.

$$\hat{\mathbf{u}} = \arg\max_{\mathbf{u}} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{u})^2 = \arg\max_{\mathbf{u}} \sum_{i=1}^{n} \mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u} = \arg\max_{\mathbf{u}} \mathbf{u}^T (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T) \mathbf{u}$$

The term $C = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$ is a 2×2 matrix and is known as the *covariance matrix* of the points - we will return to this in a later lecture. For now, note that the right-hand side of the above equation corresponds to transforming $\mathbf{u}$ by the matrix C and then projecting it onto $\mathbf{u}$. It turns out that the vector $\mathbf{u}$ which maximises this projection is the eigenvector of C corresponding to the largest eigenvalue - $C\mathbf{u} = \lambda_{max}\mathbf{u}$ is then parallel to $\mathbf{u}$ and hence maximises the projection.

The projections of the data points onto the vector $\hat{\mathbf{u}}$ are known as *principal components*. It turns out that the other eigenvector of the covariance matrix C is orthogonal to the first and projections onto that are also known as principal components - in contrast, they maximise the squared error from the original data. An example is shown in Fig. 2c - note how the two eigenvectors characterise the shape of the set of points.

PCA is used in many areas of computer science. One notable example is in *Face Recognition* in which PCA dimensionality reduction is used to 'compress' high dimension 'image vectors' into a representation in terms of principal components - known as *eigenfaces* - to enable fast searching in face databases - see for example `http://en.wikipedia.org/wiki/Eigenface`.

t

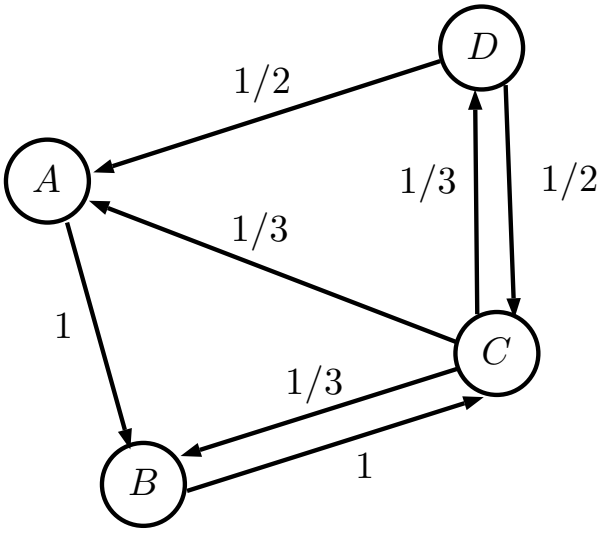


Figure 3: Linked web pages represented as a graph, with directed edges representing links from contained within one page to another page.

The Pagerank Algorithm

When searching the web, we are typically looking for web pages that contain certain words. Thus text matching can be used to find possible web pages of interest. However as there are likely to be huge numbers of these, the question arises as to how to rank the pages in some sort of order?

One way to do this is to rank them in order of *importance*, where the latter for a give page is determined by how many other pages contain a link to it - if the link to a page appears on many other pages, it must be 'important'.

However, links from some pages may be more significant than others - if a link is made from a very important page, then that must count more than a link from a page which has no links. This model underlies the ranking system used by Google and it is defined by the Pagerank algorithm.

To illustrate, consider the simple web illustrated as a graph in Fig. 3, where the nodes are pages and the directed edges indicate when pages contain links to another pages. For example, page C contains links to pages A, B and D.

In Pagerank, pages have an importance and that importance is transfered evenly amongst the pages to which it links. The importance for a page is then the sum of its transfered importance.

For example, the importance of page C is made up of the importance of page B plus half the importance of page D (since page D contains links to two pages). Denoting page importance by x_A, x_B, etc, this can be written as

$$\begin{bmatrix} x_A \\ x_B \\ x_C \\ x_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1/3 & 1/2 \\ 1 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} x_A \\ x_B \\ x_C \\ x_D \end{bmatrix}$$

Hence the importance for each page is given by the eigenvector of the 4×4 matrix in the above equation corresponding to the eigenvalue with value 1. The matrix is known as a *transition matrix* and it can be shown that its larget eigenvalue is equal to 1.

In the above case the corresponding eigenvector is approximately $(0.35, 0.58, 0.70, 0.23)$, indicating that page C has the highest importance. This can be seen by noting that although it is linked to by the same number of other pages as page B, which is ranked 2nd, the total incoming distribution is higher, 1.5 vs 1.3, and that it also gains all of page A's importance via page B.

It turns out that in practice, a number of modifications need to be made to the algorithm to make it effective, but the above captures the underlying principle. It also turns out that since the transition matrix tends to be huge (it's a big web), the rank eigenvector is not computed directly - it is approximated by repeated application of the transition matrix in what is known as the *power method*. See for example `http://en.wikipedia.org/wiki/PageRank` for more details of this and the modifications.
