smfl12.tex

Lecture 12. 11.2.2011

Orthogonal Matrices.

A square matrix A is *orthogonal* if

$$A^T = A^{-1},$$

or equivalently, if

$$A^T A = AA^T = I.$$

Then $|A^T A| = |A^T||A| = |A|.|A| = |I| = 1$, $|A|^2 = 1$, $|A| = \pm 1$ (we take the $+$ sign).

If $A = (a_1, \ldots, a_n)$ (row of column vectors, so A^T is the column of row-vectors a_i^T) is orthogonal, $A^T A = I$, i.e.

$$\begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} (a_1, \ldots, a_n) = I,$$

$a_i^T a_j = \delta_{ij}$: the columns of A are orthogonal to each other, and similarly the rows are orthogonal to each other.

Note. If A, B are orthogonal, so is AB, since $(AB)^T AB = B^T A^T AB = B^T B = I$.

Generalised inverses.

The theory above partially extends to non-square matrices, and matrices not of full rank. For A $m \times n$, call A^- a *generalised inverse* if

$$AA^- A = A.$$

We quote:
1. Generalised inverses always exist (but need not be unique),
2. If the linear equation

$$Ax = b$$

is consistent (has at least one solution), then a particular solution is given by

$$x = A^- b.$$

Eigenvalues and eigenvectors.

If A is square, and

$$Ax = \lambda x \qquad (x \neq 0),$$

λ is called an *eigenvalue* (latent value, characteristic value, e-value) of A, x an *eigenvector* (latent vector, characteristic vector, e-vector) (determined only to within a non-zero scalar factor c, as $A(cx) = \lambda(cx)$). Then

$$(A - \lambda I)x = 0$$

has non-zero solutions x, so infinitely many solutions cx, so $A-\lambda I$ is singular:

$$|A - \lambda I| = 0.$$

If A is $n \times n$, this is a polynomial equation of degree n in λ. By the Fundamental Theorem of Algebra (see e.g. M2PM3 L19-L20), there are n roots $\lambda_1, \ldots, \lambda_n$ (possibly complex, counted according to multiplicity).

A matrix A is singular iff the linear equation $Ax = 0$ has some non-zero solution x. This is the condition for 0 to be an eigenvalue:

a matrix is singular iff 0 is an eigenvalue.

Since the coefficient of λ^n in the polynomial $p(\lambda) := |A - \lambda I|$ is $(-)^n$, $p(\lambda)$ factorises as

$$p(\lambda) := |A - \lambda I| = \prod_1^n (\lambda - \lambda_i).$$

Put $\lambda = 0$:

$$|A| = \prod_1^n \lambda_i :$$

the determinant is the product of the eigenvalues.

Match the coefficients of $(-\lambda)^{n-1}$: in the RHS, we get a λ_i term for each i, so the coefficient is $\sum_i \lambda_i$, the sum of the eigenvalues. In the LHS, we get an a_{ii} term for each i, so the coefficient is $\sum a_{ii}$, the sum of the diagonal elements of A, which is called the *trace* of A. Comparing:

$$\text{tr } A = \sum_i \lambda_i :$$

the trace is the sum of the eigenvalues.

Properties.

1. If A is *symmetric*, eigenvectors x_i, x_j corresponding to distinct eigenvalues λ_i, λ_j are *orthogonal*.

Proof. $Ax_i = \lambda_i x_i$, so $x_i^T A^T = \lambda_i x_i^T$, or $x_i^T A = \lambda_i x_i^T$ as A is symmetric. So $x_i^T A x_j = \lambda_i x_i^T x_j$. Interchanging i and j and transposing (or arguing as above), $x_i^T A x_j = \lambda_j x_i^T x_j$. Subtract: $(\lambda_i - \lambda_j) x_i^T x_j = 0$, giving $x_i^T x_j = 0$ as $\lambda_i \neq \lambda_j$. //

2. If A is *real and symmetric*, its eigenvalues are *real*. For $Ax = \lambda x$; taking complex conjugates gives $A\overline{x} = \overline{\lambda}\overline{x}$ as A is real. Transposing, as A is symmetric, this gives $\overline{x}^T A = \overline{\lambda}\overline{x}^T$. So $\overline{x}^T A x = \overline{\lambda}\overline{x}^T x$. Also $Ax = \lambda x$, so $\overline{x}^T A x = \lambda \overline{x}^T x$. Subtract: $0 = (\overline{\lambda} - \lambda)\overline{x}^T x$. But if x has jth element $x_j + iy_j$, $\overline{x}^T x = \sum_j (x_j^2 + y_j^2)$, which is non-zero as x is non-zero. So $\overline{\lambda}^T = \lambda$, and λ is real. //

Note. The same proof shows that if A is *anti-symmetric* – $A^T = -A$ – the eigenvalues are purely imaginary.

3. If A is *real and orthogonal*, its eigenvalues are of *unit modulus*: $|\lambda| = 1$.

Proof. If $Ax = \lambda x$, $A\overline{x} = \overline{\lambda}\overline{x}$ as A is real, so $\overline{x}^T A^T = \overline{x}^T \overline{\lambda}$. So $\overline{x}^T A^T A x = \overline{x}^T \overline{\lambda}.\lambda x$, which as A is orthogonal is $\overline{x}^T x = \overline{\lambda}\lambda.\overline{x}^T x$. Divide by $\overline{x}^T x = \sum_i x_i^2 > 0$ (as $x \neq 0$): $\overline{\lambda}.\lambda = |\lambda|^2 = 1$. //

4. If C, A are similar ($C = B^{-1}AB$), A has eigenvalues λ and eigenvectors x – then C has eigenvalues λ and eigenvectors $B^{-1}x$.

Proof. $|A - \lambda I| = 0$, so $|C - \lambda I| = |B^{-1}AB - \lambda B^{-1} I B| = |B^{-1}||A - \lambda I||B| = 0$. So C has eigenvalues λ. $C(B^{-1}x) = (B^{-1}AB)(B^{-1}x) = B^{-1}Ax = B^{-1}\lambda x = \lambda(B^{-1}x)$, so C has eigenvectors $B^{-1}x$. //

Corollary. Similar matrices have the same determinant and trace.

Proof. These are the product and sum of the eigenvalues. //

5. If A is non-singular, the eigenvalues of A^{-1} are the reciprocals λ^{-1} of the eigenvalues λ of A, and the eigenvectors are the same.

Proof. $Ax = \lambda x$, so $x = A^{-1}\lambda x$, so $A^{-1}x = \lambda^{-1}x$. //

Theorem (Spectral Decomposition, or Jordan Decomposition). A symmetric matrix A can be decomposed as

$$A = \Gamma \Lambda \Gamma^T = \sum \lambda_i \gamma_i \gamma_i^T,$$

where $\Lambda = diag(\lambda_i)$ is the diagonal matrix of eigenvalues λ_i and $\Gamma = (\gamma_1, \ldots, \gamma_n)$ is an orthogonal matrix whose columns γ_i are standardised eigenvectors ($\gamma_i^T \gamma_i = 1$).

We refer for proof to any standard text on Linear Algebra, or on Multivariate Analysis in Statistics. As a corollary, one can show that for A

symmetric, its rank $r(A)$ is the number of non-zero eigenvalues.
Square root of a matrix.

If A is symmetric, with decomposition as above, and we define $\Lambda^{1/2} := diag(\lambda_i^{1/2})$, then putting

$$A^{1/2} := \Gamma\Lambda^{1/2}\Gamma^T,$$

$$\begin{aligned} A^{1/2}A^{1/2} &= \Gamma\Lambda^{1/2}\Gamma^T\Gamma\Lambda^{1/2}\Gamma^T \\ &= \Gamma\Lambda^{1/2}\Lambda^{1/2}\Gamma^T \qquad (\Lambda \text{ is orthogonal}) \\ &= \Gamma\Lambda\Gamma^T \qquad (\Lambda = diag(\lambda_i)) \\ &= A. \end{aligned}$$

We call $A^{1/2}$ the *square root* of A. If also A is non-singular (so no eigenvalue is 0, so each λ_i^{-1} is defined), write

$$A^{-1/2} := \Gamma\Lambda^{-1/2}\Gamma^T.$$

A similar argument shows that

$$A^{-1/2}A^{-1/2} = A^{-1},$$

so we call $A^{-1/2}$ the square root of A^{-1}, and the inverse square root of A.
Positive definite matrices.

If A $(n \times n)$ is real and symmetric, A is *positive definite* (respectively *non-negative definite*) if

$$x^TAx > 0 \qquad (\text{respectively } \geq 0) \qquad \text{for all non-zero } x.$$

Here $x^TAx = \sum_{i,j=1}^n x_i a_{ij} x_j = \sum_{i=1}^n a_{ii}x_i^2 + \sum_{i \neq j} a_{ij}x_ix_j$ is a *quadratic form* in the n variables $x_1, \ldots, x_n$ (one can replace $\sum_{i \neq j}$ by $2\sum_{i<j}$).

By the Spectral Decomposition Theorem,

$$\begin{aligned} x^TAx &= x^T\Gamma\Lambda\Gamma^Tx = y^T\Lambda y \qquad (y := \Gamma^Tx) \\ &= \sum \lambda_i y_i^2. \end{aligned}$$

So A is non-negative definite (positive definite) iff $\sum_i \lambda_i y_i^2 \geq 0$ for all y (> 0 for all non-zero y) iff all $\lambda_i \geq 0$ (> 0):
Proposition. A real symmetric matrix A is non-negative definite (positive definite) iff all its eigenvalues are non-negative (positive).