MAS212 Linear Algebra I Lecture Notes

11. Orthogonality

Contents

Orthogonality is a very important generalization of the geometrical concept of the *right-angle*: two vectors at right-angles are said to be *orthogonal*. Two orthogonal vectors are "as independent as possible", which is a stronger notion that linear independence: orthogonality implies linear independence, but not vice versa. However, there is no concept of angle in a vector space so we need some additional structure; this is the concept of an *inner product*, which can be imposed on a vector space to give an *inner product space*. An inner product is a very restricted product that always involves precisely two vectors and cannot be generalized to arbitrary numbers of vectors. An inner product of two vectors is a scalar, and so is sometimes called a *scalar product*. When working with Euclidean vectors it is usually denoted by a dot and is then often called a *dot product*. The term *scalar* or *dot product* is usually used in more concrete geometrical contexts and the term *inner product* is usually used in more abstract vector or function space contexts.

Inner Products and Norms

There are two common notations for an inner product of two vectors u and v: either (u, v) or $\langle u, v \rangle$. Both are ambiguous, since the same notation is used for other concepts, but of the two the second is less ambiguous so we will use it; do not confuse it with the notation for the vector space spanned by the two vectors! In this course we will consider only real inner product spaces, although the definition of inner product extends easily to complex vector spaces (except that complex conjugation becomes important).

Definition of Real Inner Product

Let V be a vector space over the real field $\boldsymbol{R}$. The *real inner product* is a (bilinear) function from $V \times V$ to $\boldsymbol{R}$ denoted by $\langle u, v \rangle$ such that the following axioms hold:

- $\langle k_1 u_1 + k_2 u_2, v \rangle = k_1 \langle u_1, v \rangle + k_2 \langle u_2, v \rangle$ for all $u_1, u_2, v \in V$, $k_1, k_2 \in R$ [linearity];
- $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$ [symmetry];
- $0 \le \langle u, u \rangle$ for all $u \in V$ and $\langle u, u \rangle = 0$ if and only if $u = 0$ [positive definiteness].

Linearity and symmetry together imply *bilinearity*, which means linearity with respect to each of the two vectors. There are many possible realizations of inner product, but we will consider only the following, which is the most common and generally most useful.

Definition of Standard Inner Product in R^n

If $(x = (x_1, x_2, \ldots, x_n)) \in R^n$, $(y = (y_1, y_2, \ldots, y_n)) \in R^n$ then the standard inner product is defined to be $\langle x, y\rangle = \sum_{i=1}^{n} x_i\, y_i$, namely the scalar or dot product used in Euclidean geometry. If x and y are represented as *column* vectors then $\langle x, y\rangle = (x^T) \, . \, y$.

Reminder about Transposition

We will use the concept of the *transpose* of a matrix a lot in this chapter, which is not surprising because its fundamental significance is very much tied up with inner product spaces. For further details of this connection see the last section in these lectures. The important thing to remember is that transposition exchanges rows and columns, which effectively flips a matrix about its principal diagonal, and in particular it turns a column vector into a row vector and vice versa. The only way to multiply two vectors using matrix multiplication is to transpose one of the vectors so that the first vector in the product is a row vector and the second is a column vector; the result is the scalar or dot product of the two vectors. Transposition also plays a crucial role in the symmetry of a square matrix. Remember that transposing a product reverses the factors, which we will prove and use later.

Another geometrical concept that we need to generalize is that of *length*, for which the general mathematical term is *norm*. Again, this concept is not defined for a general vector space, but a vector space on which a norm is defined is called a *normed vector space*. An inner product implies a norm.

Definition of Norm

Let V be a vector space over the real field $\boldsymbol{R}$. A *norm* is a function from V to $\boldsymbol{R}$ denoted by $\|v\|$ such that the following axioms hold:

- $0 \le \|v\|$ for all $v \in V$ and $\|v\| = 0$ if and only if $v = 0$ [positive definiteness];
- $\|k\,v\| = |k|\,\|v\|$ for all $v \in V$, $k \in R$ [scalar multiple];
- $\|u + v\| \le \|u\| + \|v\|$ for all $u, v \in V$ [triangle inequality].

Again, there are many possible realizations of a norm, but we will consider only the following, which is identical to the geometrical length.

Definition of Euclidean Norm

Let V be a vector space over the real field $\boldsymbol{R}$. The *Euclidean norm* is the function from V to $\boldsymbol{R}$ defined by $\|v\| = \sqrt{\langle v, v\rangle}$.

That this is a well defined norm follows fairly directly from the definition of inner product; it is necessary to take the square root in order to satisfy the scalar multiple requirement. In this course, *norm* will always mean the *Euclidean norm*. It is also sometimes called the 2-norm because it is the special case with $p = 2$ of the p-norm defined for $v \in R^n$ by $||v||_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}$.

We are now in a position to make the following definitions.

Definition of Orthogonal, Normalized, Unit and Orthonormal Vectors

Let V be a vector space.

- Two vectors $u, v \in V$ are *orthogonal* if their inner product vanishes, i.e. $\langle u, v \rangle = 0$.
- A set of vectors is orthogonal if every pair of vectors in the set is orthogonal.
- A vector $v \in V$ is *normalized* or is a *unit vector* if it has unit norm, i.e. $||v|| = 1$.
- Two vectors $u, v \in V$ are *orthonormal* if each vector is normalized and their inner product vanishes, i.e. $||u|| = 1$, $||v|| = 1$ and $\langle u, v \rangle = 0$.
- A set of vectors is orthonormal if every vector in the set is normalized and every pair of vectors in the set is orthogonal.

Any vector can be normalized by dividing it by its norm, i.e. $\frac{v}{||v||}$ necessarily has norm 1.

Examples

$\{(1, 0), (0, 1)\}$ and $\{(1, 1), (-1, 1)\}$ are two orthogonal sets of vectors in R^2. The set $\{(1, 0), (0, 1)\}$ is also orthonormal but the set $\{(1, 1), (-1, 1)\}$ is not, since each vector in the latter set has norm $\sqrt{2}$. However, each vector can be normalized by dividing it by its norm and $\left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right\}$ is an orthonormal set.

An orthonormal basis for a vector space has some advantages and the standard basis for R^n is orthonormal, e.g. the standard basis for R^3 is $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

Real Symmetric Matrices

Matrices that have pure real elements and are symmetric are important in applications and have a number of useful properties that make them particularly simple to work with.

Definition of Symmetric and AntiSymmetric Matrices

A square $n \times n$ matrix A is *symmetric* if $A^T = A$, i.e. $a_{i,j} = a_{j,i}$, $i, j = 1 .. n$ or *antisymmetric* (which is sometimes called *skew-symmetric*) if $A^T = -A$, i.e. $a_{i,j} = -a_{j,i}$, $i, j = 1 .. n$.

A symmetric or antisymmetric matrix must be square for its transpose to have the same shape and a symmetric / antisymmetric matrix has mirror symmetry / antisymmetry about its principal diagonal. The principal diagonal of an antisymmetric matrix must vanish, since antisymmetry means that $a_{ii} = -a_{ii}$ => $a_{ii} = 0$. Any square matrix A can be written as the sum of a symmetric and an antisymmetric matrix in the form $A = \frac{(A + A^T)}{2} + \frac{(A - A^T)}{2}$, since the matrix $(A + A^T)$ is clearly symmetric and the matrix $(A - A^T)$ is clearly antisymmetric.

Examples

Symmetric and antisymmetric 2×2 matrices have the following general forms:

$$\text{symmetric: } \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \text{antisymmetric: } \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$$

A general 2×2 matrix A can be decomposed into its symmetric part, AS, and its antisymmetric part, AA, as follows:

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

$$AS := \begin{bmatrix} a & \frac{b}{2} + \frac{c}{2} \\ \frac{b}{2} + \frac{c}{2} & d \end{bmatrix}$$

$$AA := \begin{bmatrix} 0 & \frac{b}{2} - \frac{c}{2} \\ \frac{c}{2} - \frac{b}{2} & 0 \end{bmatrix}$$

$$AS + AA = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

One application of real symmetric matrices is as representations of quadratic forms, which are homogeneous polynomials of degree 2, e.g.

$$[x, y] . \begin{bmatrix} a & b \\ b & c \end{bmatrix} . \begin{bmatrix} x \\ y \end{bmatrix} = x^2\, a + 2\, y\, x\, b + y^2\, c$$

A polynomial of this form equated to zero is the general equation of a conic section or quadric curve (i.e. an ellipse, parabola or hyperbola) centred at the origin. This generalizes to higher dimensions, e.g. quadric surfaces in three dimensions. The properties of the underlying real symmetric matrix relate closely to the properties of the corresponding quadric curve or surface; for example, the eigenvectors specify natural axes.

Eigenvalues and Eigenvectors of Real Symmetric Matrices

The eigenvalues and eigenvectors of real symmetric matrices are particularly simple, one consequence of which is that all real symmetric matrices can be diagonalized. Most proofs of properties of real symmetric matrices involve transposition.

Lemma: Transposing a Product Reverses the Factors

If A, B are matrices such that their product exists then $(A \,.\, B)^T = (B^T) \,.\, (A^T)$.

Proof

A matrix product corresponds to a sum over neighbouring indices and transposing corresponds to interchanging the indices, hence

$$((A \,.\, B)^T)_{i,j} = (A \,.\, B)_{j,i} = \sum_k (A)_{j,k} \, (B)_{k,i} = \sum_k (A^T)_{k,j} \, (B^T)_{i,k} =$$
$$\sum_k (B^T)_{i,k} \, (A^T)_{k,j} = ((B^T) \,.\, (A^T))_{i,j}.$$

Remark

This result generalizes to a product of any number of matrices, which need not be square provided they are conformable for multiplication in the order specified. In particular, the first matrix can be a row vector and the last can be a column vector.

Proposition: A Real Symmetric Matrix has Real Eigenvalues

If A is a real symmetric matrix and there exists a vector $x \neq 0$ such that $A \,.\, x = \lambda\, x$ then λ is real.

Proof

Let * denote complex conjugate. Premultiplying $A \,.\, x = \lambda\, x$ by the row vector x^{*T} gives

$$(x^{*T}) \,.\, A \,.\, x = \lambda\, ((x^{*T}) \,.\, x). \qquad (1)$$

Transposing and conjugating this equation (using the previous lemma) gives

$$(x^{*T}) \,.\, (A^{*T}) \,.\, x = \lambda^* \, ((x^{*T}) \,.\, x) \qquad (2)$$

But A is real and symmetric, so $A^{*T} = A$ and (2) becomes

$$(x^{*T}) \,.\, A \,.\, x = \lambda^* \, ((x^{*T}) \,.\, x).$$

Subtracting this equation from (1) gives $0 = (\lambda - \lambda^*)\, ((x^{*T}) \,.\, x)$. But $(x^{*T}) \,.\, x = \sum_i |x_i|^2 \neq 0$ since $x \neq 0$, so $\lambda^* = \lambda$ and hence λ is real.

Remark

The eigenvectors of a real symmetric matrix can be chosen to be real since the equation they satisfy is real.

Proposition: A Real Symmetric Matrix has Orthogonal Eigenvectors

If A is a real symmetric matrix with eigenvectors x, y corresponding to *distinct* eigenvalues λ, μ then

$\langle x, y \rangle = 0$, i.e. the eigenvectors are orthogonal.

Proof

$A . x = \lambda\, x$ and $A . y = \mu\, y$. Premultiply these equations by the row vectors y^T and x^T respectively to give $(y^T) . A . x = \lambda\, ((y^T) . x)$ and $(x^T) . A . y = \mu\, ((x^T) . y)$. Transpose the second equation and subtract to give $0 = (\lambda - \mu)\, ((y^T) . x)$. If the eigenvalues are distinct then $\lambda - \mu \neq 0$ so $(y^T) . x = 0$, i.e. $\langle y, x \rangle = 0$ and the eigenvectors are orthogonal.

Orthogonal Matrices

An orthogonal matrix is an important special case of an invertible matrix.

Definition of Orthogonal Matrix

A real square matrix A is *orthogonal* if $A \,.\, (A^T) = I = (A^T) \,.\, A$, i.e. $A^{(-1)} = A^T$.

Clearly, it is much easier to invert an orthogonal matrix than a general matrix!

Proposition: the Rows and Columns of an Orthogonal Matrix are Orthonormal

If R_i represents row i and C_j represents column j of an $n{\times}n$ orthogonal matrix A then $\{ R_i, i = 1 \,..\, n \}$ and $\{ C_j, j = 1 \,..\, n \}$ are orthonormal sets of vectors.

Proof

By the rules of matrix multiplication, $(A \,.\, (A^T))_{i,j} = \langle R_i, R_j \rangle$. But we know that $A \,.\, (A^T) = I$, hence $\langle R_i, R_j \rangle = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases}$, i.e. the rows of A are orthonormal. If A is orthogonal then so is A^T, hence the columns are also orthonormal.

Proposition: an Orthogonal Matrix represents a Rotation

A linear mapping represented by an orthogonal matrix preserves the lengths of vectors and therefore corresponds to a pure rotation about the origin, i.e. if A is an orthogonal $n{\times}n$ matrix then $||A \,.\, x|| = ||x||$ for all $x \in R^n$.

Proof

Recall that $||v|| = \sqrt{\langle v, v \rangle}$ and $\langle v, v \rangle = (v^T) \,.\, v$, so that $||v||^2 = (v^T) \,.\, v$. Then $||A \,.\, x||^2 = (x^T) \,.\, (A^T) \,.\, A \,.\, x = (x^T) \,.\, x = ||x||^2$.

Orthogonal Diagonalization of Real Symmetric Matrices

A real symmetric matrix has the important property that it can be diagonalized by a real orthogonal similarity transformation. We will consider this in two stages: first the simple case that the matrix has distinct eigenvalues and then the general case. Note that a general square matrix cannot necessarily be diagonalized by any similarity transformation, so the diagonalizability property of all real symmetric matrices makes them a very important special case!

Proposition: Orthogonal Diagonalizability of Some Real Symmetric Matrices

If a real symmetric $n{\times}n$ matrix A has n distinct eigenvalues then it is diagonalizable by a real orthogonal similarity transformation.

Proof

We have already proved that if an $n{\times}n$ matrix A has n distinct eigenvalues then it is diagonalizable by a similarity transformation using the matrix of eigenvectors. If the matrix A is real symmetric then its eigenvectors are orthogonal and can be chosen to be real. If they are normalized before constructing the similarity transformation matrix then it will be a real orthogonal matrix.

Example 1

Let

$$A := \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Its eigenvalues and corresponding eigenvectors are

$$\mathit{eigenvalues} := 5, 0$$

$$\mathit{eigenvectors} := \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

The eigenvectors are orthogonal, because their inner (dot) product is

$$\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} . \begin{bmatrix} -2 \\ 1 \end{bmatrix} = 0$$

Normalizing (using the Euclidean or 2-norm) and using the eigenvectors as the columns of a matrix X gives

$$X := \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}$$

The inverse P of X is given simply by its transpose (since X is orthogonal), which we can check:

$$P := \begin{bmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}$$

$$P . X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then P defines an orthogonal transformation by which A is similar to the diagonal matrix of its eigenvalues:

$$P . A . (P^T) = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$$

Note that a zero eigenvalue does not cause any difficulties (although it means that the matrix A is singular).

Example 2

This example is more complicated, and leads us to the final theorem of the course. Let

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

This matrix has only 2 distinct eigenvalues, −1 and 2. The eigenvalue −1 has a 1-dimensional eigenspace spanned by the first eigenvector (1, 1, 1) and the eigenvalue 2 has a 2-dimensional eigenspace spanned by the second and third eigenvectors (−1, 0, 1) and (−1, 1, 0). Because the eigenvalues are not all distinct the eigenvectors are not necessarily all orthogonal: the first is orthogonal to the second and third, but the second and third are not orthogonal (although they are linearly independent). Normalizing the first and second eigenvectors (as row vectors) gives

$$p_1 := \left[\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right]$$

$$p_2 := \left[-\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}\right]$$

We can replace the third eigenvector by a linear combination of the second and third eigenvectors that is orthogonal to the second eigenvector. If we want the vector $v + k\,u$ to be orthogonal to u then we require $(u\,.\,v) + k\,(u\,.\,u) = 0$ => $k = -\frac{u\,.\,v}{u\,.\,u}$, i.e. if we replace v by $v - \frac{(u\,.\,v)\,u}{u\,.\,u}$ then the new v is orthogonal to u. (This is a simple example of a general procedure called the Gram-Schmidt Orthogonalization Process.) Therefore, if we replace the third eigenvector

$$p_3 := [-1, 1, 0]$$

by $p_3 - \frac{((p_2)\,.\,(p_3))\,p_2}{(p_2)\,.\,(p_2)}$ to give

$$p_3 := \left[\frac{-1}{2}, 1, \frac{-1}{2}\right]$$

then we have constructed an orthogonal set of eigenvectors. Normalizing the new third eigenvector gives the matrix P, whose *rows* are the orthonormal eigenvectors p_1, p_2, p_3, as

$$P := \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{3} & -\frac{\sqrt{6}}{6} \end{bmatrix}$$

We can check that this is an orthogonal matrix:

$$P\,.\,(P^T) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then P defines an orthogonal transformation by which A is similar to the diagonal matrix of its eigenvalues:

$$P \,.\, A \,.\, (P^T) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The previous example indicates a general procedure for diagonalizing a real symmetric matrix even when the eigenvalues are not all distinct and the following theorem proves that this is always possible.

Theorem: Orthogonal Diagonalizability of All Real Symmetric Matrices

A real square matrix is orthogonally similar to a real diagonal matrix if and only if it is symmetric.

Proof*

(a) Orthogonally diagonalizable => symmetric:

If Λ is diagonal and $A = P . \Lambda . (P^T)$ then $A^T = P . \Lambda . (P^T) = A$.

(b) Symmetric => orthogonally diagonalizable:

Let A be a real symmetric $n \times n$ matrix. We prove the claim by induction on n. At least one (real) eigenvalue λ_1 and one (real) eigenvector x_1 exist such that $A . (x_1) = \lambda_1 x_1$. Extend x_1 to an orthonormal basis for R^n, which can be done as follows. If the ith entry in x_1 is nonzero then extend x_1 to a basis by including the standard basis vectors $e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n$. From this basis, use the Gram-Schmidt Orthogonalization Process to construct an orthogonal basis and normalize each vector in the basis so that $(x_i^T) . (x_j) = 1$ if $i = j$ and 0 otherwise. Construct an orthogonal matrix X by taking the vectors x_i as its columns. Then

$(X^T) . A . X = (X^T) . (A . (x_1), A . (x_2), \ldots, A . (x_n)) =$

$$\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} . [\lambda_1 x_1, A . (x_2), \ldots, A . (x_n)] = \begin{bmatrix} \lambda_1 & k_2 & \ldots & k_n \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix}$$

for some $k_2, \ldots, k_n \in R$. But $(X^T) . A . X$ is a real symmetric matrix (since $((X^T) . A . X)^T = (X^T) . A . X$), hence $k_2 = 0, \ldots, k_n = 0$ and

$$(X^T) . A . X = \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix}$$

where A_1 is also a real symmetric matrix.

The induction hypothesis, which is trivially true when $n = 1$, implies that A_1 is orthogonally similar to a real diagonal matrix Λ_1, i.e. there exists an orthogonal matrix X_1 such that $(X_1^T) . (A_1) . (X_1) = \Lambda_1$. Define

$$X'_1 = \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix}$$

Then

$$(X'^T_1) . (((X^T) . A . X)) . (X'_1) = \begin{bmatrix} 1 & 0 \\ 0 & X_1^T \end{bmatrix} . \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix}, =$$

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & (X_1^T) . (A_1) . (X_1) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_1 \end{bmatrix}, = \Lambda$$

If we define $P = (X'^{T}_{1}) \,.\, (X^T)$ then $P \,.\, A \,.\, (P^T) = \Lambda$ is a diagonal matrix and P is an orthogonal matrix.

The Transpose of a Linear Map*

Transposing a matrix clearly interchanges the roles of the domain and codomain in some sense, but to make this precise requires the concept of *linear functionals* and *dual spaces*. If V is a vector space over a field K then a linear map $\phi : V \to K$ (regarding K as a one-dimensional vector space) is called a *linear functional* (or linear form). The set of all linear functionals on a vector space V over a field K forms another vector space over K (as we saw earlier), which is called the *dual space* of V and is denoted V^*. The transpose of a mapping α from U to V is the mapping α^T from V^* to U^*. A $1 \times n$ row matrix represents a mapping from K^n to K, i.e. a linear functional, and corresponds to the transpose of a column vector. Hence, the inner product $\langle x, y \rangle = (x^T) \,.\, y$ is the value of the linear functional that is represented by the transpose of the vector x applied to the vector y. This, in outline, is the fundamental significance of transposition and why it arises in the context of inner products.