Lecture 2. Orthogonal Vectors and Matrices

Since the 1960s, many of the best algorithms of numerical linear algebra have been based in one way or another on orthogonality. In this lecture we present the ingredients: orthogonal vectors and orthogonal (unitary) matrices.

Adjoint

The *complex conjugate* of a scalar z, written $\overline{z}$ or z^*, is obtained by negating its imaginary part. For real z, $\overline{z} = z$.

The *hermitian conjugate* or *adjoint* of an $m \times n$ matrix A, written A^*, is the $n \times m$ matrix whose i, j entry is the complex conjugate of the j, i entry of A. For example,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad \Longrightarrow \qquad A^* = \begin{bmatrix} \overline{a}_{11} & \overline{a}_{21} & \overline{a}_{31} \\ \overline{a}_{12} & \overline{a}_{22} & \overline{a}_{32} \end{bmatrix}.$$

If $A = A^*$, A is *hermitian*. By definition, a hermitian matrix must be square. For real A, the adjoint simply interchanges the rows and columns of A. In this case, the adjoint is also known as the *transpose*, and is written A^T. If a real matrix is hermitian, that is, $A = A^T$, then it is also said to be *symmetric*.

Most textbooks of numerical linear algebra assume that the matrices under discussion are real and thus use principally ^{T} instead of *. Since most of the ideas to be dealt with are not intrinsically restricted to the reals, however, we have followed the other course. Thus, for example, in this book a row vector will usually be denoted by, say, a^* rather than a^T. The reader who prefers to imagine that all quantities are real and that * is a synonym for ^{T} will rarely get into trouble.

Inner Product

The *inner product* of two column vectors $x, y \in \mathbb{C}^m$ is the product of the adjoint of x by y:

$$x^*y = \sum_{i=1}^{m} \overline{x}_i y_i. \tag{2.1}$$

The Euclidean length of x may be written $\|x\|$ (vector norms such as this are discussed systematically in the next lecture), and can be defined as the square root of the inner product of x with itself:

$$\|x\| \;=\; \sqrt{x^*x} \;=\; \left(\sum_{i=1}^{m} |x_i|^2\right)^{1/2}. \tag{2.2}$$

The cosine of the angle α between x and y can also be expressed in terms of the inner product:

$$\cos\alpha = \frac{x^*y}{\|x\|\,\|y\|}. \tag{2.3}$$

At various points of this book, as here, we mention geometric interpretations of algebraic formulas. For these geometric interpretations, the reader should think of the vectors as real rather than complex, although usually the interpretations can be carried over in one way or another to the complex case too.

The inner product is *bilinear*, which means that it is linear in each vector separately:

$$\begin{aligned} (x_1 + x_2)^*y &= x_1^*y + x_2^*y, \\ x^*(y_1 + y_2) &= x^*y_1 + x^*y_2, \\ (\alpha x)^*(\beta y) &= \overline{\alpha}\beta x^*y. \end{aligned}$$

We shall also frequently use the easily proved property that for any matrices or vectors A and B of compatible dimensions,

$$(AB)^* = B^*A^*. \tag{2.4}$$

This is analogous to the equally important formula for products of invertible square matrices,

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{2.5}$$

The notation A^{-*} is a shorthand for $(A^*)^{-1}$ or $(A^{-1})^*$; these two are equal, as can be verified by applying (2.4) with $B = A^{-1}$.

Orthogonal Vectors

A pair of vectors x and y are said to be *orthogonal* if $x^*y = 0$. If x and y are real, this means they lie at right angles to each other in $\mathbb{R}^m$. Two sets of vectors X and Y are orthogonal (also stated "X is orthogonal to Y") if every $x \in X$ is orthogonal to every $y \in Y$.

A set of nonzero vectors S is *orthogonal* if its elements are pairwise orthogonal, i.e., if for $x, y \in S$, $x \neq y \Rightarrow x^*y = 0$. A set of vectors is *orthonormal* if it is orthogonal and in addition every $x \in S$ has $\|x\| = 1$.

Theorem 2.1. *The vectors in an orthogonal set S are linearly independent.*

Proof. If the vectors in S are not independent, then some $v_k \in S$ can be expressed as a linear combination of other members $v_1, \ldots, v_n \in S$,

$$v_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} c_i v_i.$$

Since $v_k \neq 0$, $v_k^* v_k = \|v_k\|^2 > 0$. Using the bilinearity of inner products and the orthogonality of S, we calculate

$$v_k^* v_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} c_i v_k^* v_i = 0,$$

which contradicts the assumption that the vectors in S are nonzero. $\square$

As a corollary of Theorem 2.1 it follows that if an orthogonal set $S \subseteq \mathbb{C}^m$ contains m vectors, then it is a basis for $\mathbb{C}^m$.

Components of a Vector

The most important idea to draw from the concepts of inner products and orthogonality is this: inner products can be used to decompose arbitrary vectors into orthogonal components.

For example, suppose that $\{q_1, q_2, \ldots, q_n\}$ is an orthonormal set, and let v be an arbitrary vector. The quantity $q_j^* v$ is a scalar. Utilizing these scalars as coordinates in an expansion, we find that the vector

$$r = v - (q_1^* v) q_1 - (q_2^* v) q_2 - \cdots - (q_n^* v) q_n \qquad (2.6)$$

is orthogonal to $\{q_1, q_2, \cdots, q_n\}$. This can be verified by computing $q_i^* r$:

$$q_i^* r \;=\; q_i^* v - (q_1^* v)(q_i^* q_1) - \cdots - (q_n^* v)(q_i^* q_n).$$

This sum collapses, since $q_i^* q_j = 0$ for $i \neq j$:

$$q_i^* r \;=\; q_i^* v - (q_i^* v)(q_i^* q_i) \;=\; 0.$$

Thus we see that v can be decomposed into $n+1$ orthogonal components:

$$v \;=\; r + \sum_{i=1}^{n} (q_i^* v) q_i \;=\; r + \sum_{i=1}^{n} (q_i q_i^*) v. \tag{2.7}$$

In this decomposition, r is the part of v orthogonal to the set of vectors $\{q_1, q_2, \ldots, q_n\}$, or equivalently to the subspace spanned by this set of vectors, and $(q_i^* v) q_i$ is the part of v in the direction of q_i.

If $\{q_i\}$ is a basis for $\mathbb{C}^m$, then n must be equal to m and r must be the zero vector, so v is completely decomposed into m orthogonal components in the directions of the q_i:

$$v \;=\; \sum_{i=1}^{m} (q_i^* v) q_i \;=\; \sum_{i=1}^{m} (q_i q_i^*) v. \tag{2.8}$$

In both (2.7) and (2.8) we have written the formula in two different ways, once with $(q_i^* v) q_i$ and again with $(q_i q_i^*) v$. These expressions are equal, but they have different interpretations. In the first case, we view v as a sum of coefficients $q_i^* v$ times vectors q_i. In the second, we view v as a sum of orthogonal projections of v onto the various directions q_i. The ith projection operation is achieved by the very special rank-one matrix $q_i q_i^*$. We shall discuss this and other projection processes in Lecture 6.

Unitary Matrices

A square matrix $Q \in \mathbb{C}^{m\times m}$ is *unitary* (in the real case we also say *orthogonal*) if $Q^* = Q^{-1}$, i.e, if $Q^*Q = I$. In terms of the columns of Q, this product can be written

$$\left[\begin{array}{c} q_1^* \\ \hline q_2^* \\ \hline \vdots \\ \hline q_m^* \end{array}\right] \left[\begin{array}{c|c|c|c} & & & \\ q_1 & q_2 & \cdots & q_m \\ & & & \end{array}\right] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

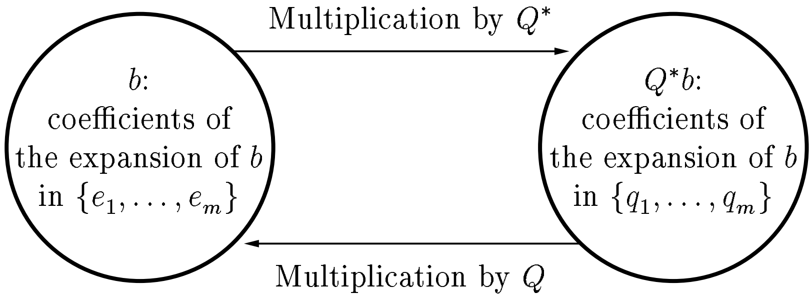
In other words, $q_i^* q_j = \delta_{ij}$, and the columns of a unitary matrix Q form an orthonormal basis of $\mathbb{C}^m$. The symbol δ_{ij} is the *Kronecker delta*, equal to 1 if $i = j$ and 0 if $i \neq j$.

Multiplication by a Unitary Matrix

In the last lecture we discussed the interpretation of matrix-vector products Ax and $A^{-1}b$. If A is a unitary matrix Q, these products become Qx and Q^*b, and the same interpretations are of course still valid. As before, Qx is the linear combination of the columns of Q with coefficients x. Conversely,

Q^*b is the vector of coefficients of the expansion of b in the basis of columns of Q.

Schematically, the situation looks like this:



These processes of multiplication by a unitary matrix or its adjoint preserve geometric structure in the Euclidean sense, because inner products are preserved. That is, for unitary Q,

$$(Qx)^*(Qy) = x^*y, \tag{2.9}$$

as is readily verified by (2.4). The invariance of inner products means that angles between vectors are preserved, and so are their lengths:

$$\|Qx\| = \|x\|. \tag{2.10}$$

In the real case, multiplication by an orthogonal matrix Q corresponds to a rigid rotation (if $\det Q = 1$) or reflection (if $\det Q = -1$) of the vector space.

Exercises

1. Show that if a matrix A is both triangular and unitary, then it is diagonal.

2. The Pythagorean theorem asserts that for a set of n orthogonal vectors $\{x_i\}$,

$$\|\sum_{i=1}^{n} x_i\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$

(a) Prove this in the case $n = 2$ by an explicit computation of $\|x_1 + x_2\|^2$.

(b) Show that this computation also establishes the general case, by induction.

3. Let $A \in \mathbb{C}^{m\times m}$ be hermitian. An eigenvector of A is a nonzero vector $x \in \mathbb{C}^m$ such that $Ax = \lambda x$ for some $\lambda \in \mathbb{C}$, the corresponding eigenvalue.

(a) Prove that all eigenvalues of A are real.

(b) Prove that if x and y are eigenvectors corresponding to distinct eigenvalues, then x and y are orthogonal.

4. What can be said about the eigenvalues of a unitary matrix?

5. Let $S \in \mathbb{C}^{m\times m}$ be *skew-hermitian*, i.e., $S^* = -S$.

   (a) Show by using Exercise 1 that the eigenvalues of S are pure imaginary.

   (b) Show that $I - S$ is nonsingular.

   (c) Show that the matrix $Q = (I-S)^{-1}(I+S)$, known as the *Cayley transform* of S, is unitary. (This is a matrix analogue of a linear fractional transformation $(1+s)/(1-s)$, which maps the left half of the complex s-plane conformally onto the unit disk.)

6. If u and v are m-vectors, the matrix $A = I + uv^*$ is known as a *rank-one perturbation of the identity*. Show that if A is nonsingular, then its inverse has the form $A^{-1} = I + \alpha uv^*$ for some scalar α, and give an expression for α. For what u and v is A singular? If it is singular, what is $\operatorname{null}(A)$?

7. A *Hadamard matrix* is a matrix whose entries are all ± 1 and whose transpose is equal to its inverse times a constant factor. It is known that if A is a Hadamard matrix of dimension $m > 2$, then m is a multiple of 4, but it is an unsolved problem whether there is a Hadamard matrix for every such m, though examples are known for all cases $m \le 424$.

   Show that the following recursive description provides a Hadamard matrix of each dimension $m = 2^k$, $k = 0, 1, 2, \ldots$.

$$H_0 = \begin{bmatrix} 1 \end{bmatrix}, \qquad H_{k+1} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix}.$$