Lecture 1: Linear Algebra

Standard texts on Linear Algebra and Algebra are [1, 2].

1 Preliminaries

1.1 Vectors and matrices

We shall use $\mathbb{R}$ to denote the set of real numbers and $\mathbb{C}$ to denote the set of complex numbers. For any $c = a + bi \in \mathbb{C}$, the *complex conjugate* of c, denoted by $\bar{c}$ is defined to be $\bar{c} = a - bi$. The *modulus* of c, denoted by $|c|$, is $\sqrt{a^2 + b^2}$. It is easy to see that $|c|^2 = c\bar{c}$.

If we mention the word "vector" alone, it is understood to be a column vector. An n-dimensional vector x has n entries in some field of numbers, such as $\mathbb{R}$ or $\mathbb{C}$:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The set of all n-dimensional vectors over $\mathbb{R}$ (respectively $\mathbb{C}$) is denoted by $\mathbb{R}^n$ (respectively $\mathbb{C}^n$). They are also called *real vectors* and *complex vectors*, respectively.

Similar to vectors, matrices need an underlying field. We thus have complex matrices and real matrices just as in the case of vectors. In fact, an n-dimensional vector is a $n \times 1$ matrix. In the discussion of the following paragraph, the concepts of complex conjugates, transposes, and conjugate transposes also apply to vectors in this sense.

Given an $m \times n$ matrix $A = (a_{ij})$, the *complex conjugate* $\bar{A}$ of A is a matrix obtained from A by replacing each entry a_{ij} of A by the corresponding complex conjugate $\bar{a}_{ij}$. The *transpose* A^T of A is the matrix obtained from A by turning its rows into columns and vice versa. For example,

$$A = \begin{bmatrix} 0 & 3 & 1 \\ -2 & 0 & 1 \end{bmatrix}, A^T = \begin{bmatrix} 0 & -2 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}.$$

The *conjugate transpose* A^* of A is defined to be $\bar{A}^T$. A square matrix A is *symmetric* if $A = A^T$, and it is *Hermitian* if $A = A^*$.

Given a real vector $x \in \mathbb{R}^n$, the *length* $\|x\|$ of x is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1}$$

Notice that $\|x\|^2 = xx^T$. When x is a complex vector, we use x^* instead of x^T. Hence, in general we define $\|x\| = \sqrt{xx^*} = \sqrt{x^*x}$. (You should check that $xx^* = x^*x$, and that it is a real number so that the square root makes sense.)

The length $\|x\|$ is also referred to as the L_2-*norm* of vector x, denoted by $\|x\|_2$. In general, the L_p-*norm* $\|x\|_p$ of an n-dimensional vector x, $p = 1, 2, \dots$, is defined to be

$$\|x\|_p := \left(|x_1|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}, \tag{2}$$

and

$$\|x\|_\infty := \max_{i=1..n} |x_i|. \tag{3}$$

The following identities are easy to show, yet of great importance. Given an $p \times q$ matrix A and an $q \times r$ matrix B, we have

$$\begin{aligned} (AB)^T &= B^T A^T & (4)\\ (AB)^* &= B^* A^* & (5) \end{aligned}$$

(Question: what's the dimension of $(AB)^T$ and $(AB)^*$?)

A square matrix A is said to be *singular* if there is no unique solution to the equation $Ax = b$. It does not matter what b is. The uniqueness of solutions to $Ax = b$ is an intrinsic property of A alone. If there is one and only one x such that $Ax = b$, then A is said to be *non-singular*.

1.2 Determinant and trace

Given a square matrix $A = (a_{ij})$ of order n, the equation $Ax = 0$ has a unique solution x if and only if $\det A \neq 0$, where $\det A$ denotes the *determinant* of A, which is defined by

$$\det A = \sum_{\pi \in S_n} (-1)^{\text{sign}(\pi)} \prod_{i=1}^{n} a_{i\pi(i)}. \tag{6}$$

Here, S_n denotes the set of all permutations on the set $[n] = \{1, \dots, n\}$. (S_n is more often referred to as the *symmetric group* of order n.) Given a permutation $\pi \in S_n$, the number of *inversions* of π is the number of pairs $(\pi(i), \pi(j))$ for which $i < j$ and $\pi(i) > \pi(j)$. The *sign* of a permutation π, denoted by $\text{sign}(\pi)$ is -1 if there are an odd number of inversions, and 1 if there are an even number of inversions.

Let us take an example for $n = 3$. In this case S_n consists of 6 permutations:

$$S_n = \{123, 132, 213, 231, 312, 321\}.$$

Notationally, if $\pi = 132$, then $\pi(1) = 1$, $\pi(2) = 3$, and $\pi(3) = 2$; and, $\text{sign}(\pi) = -1$ since there is only one "out-of-order" pair $(3, 2)$ in the permutation. To be more precise, $\text{sign}(123) = 1$, $\text{sign}(132) = -1$, $\text{sign}(312) = 1$, $\text{sign}(213) = -1$, $\text{sign}(231) = 1$, $\text{sign}(321) = -1$.

With

$$A = \begin{bmatrix} 0 & 3 & 1 \\ -2 & 0 & 1 \\ -1 & 2 & 2 \end{bmatrix}$$

we have

$$\begin{aligned} \det A &= a_{11}a_{22}a_{33} + (-1)a_{11}a_{23}a_{32} + (-1)a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + \\ &\quad a_{13}a_{21}a_{32} + (-1)a_{13}a_{22}a_{31} \\ &= 0 \cdot 0 \cdot 2 + (-1) \cdot 0 \cdot 1 \cdot 2 + (-1) \cdot 3 \cdot (-2) \cdot 2 + 3 \cdot 1 \cdot (-1) + \\ &\quad 1 \cdot (-2) \cdot 2 + (-1) \cdot 1 \cdot 0 \cdot (-1) \\ &= 5 \end{aligned}$$

The *trace* of a square matrix A, denoted by $\text{tr}\, A$ is the sum of its diagonal entries. The matrix A above has

$$\text{tr}\, A = 0 + 0 + 2 = 2.$$

1.3 Combinations of vectors and vector spaces

A vector w is a *linear combination* of m vectors $v_1, \dots, v_m$ if w can be written as

$$w = a_1 v_1 + a_2 v_2 + \dots a_m v_m. \tag{7}$$

The number a_i is called the *coefficient* of the vector v_i in this linear combination. Note that, as usual, we have to fix the underlying field such as $\mathbb{R}$ or $\mathbb{C}$. If, additionally, we also have $a_1 + a_2 + \dots + a_m = 1$, then w is called an *affine combination* of the v_i.

A *canonical combination* is a linear combination in which $\lambda_j \geq 0, \forall j$; and a *convex combination* is an affine combination which is also canonical. The *linear (affine, canonical, convex) hull* of $\{v_1, \dots, v_m\}$ is the set of all linear (affine, canonical, convex) combinations of the v_j. Note that in the above definitions, m could be infinite. The convex hull of a finite set of vectors is called a *cone*, or more specifically a *convex polyhedral cone*.

A *real vector space* is a set V of real vectors so that a linear combination of any subset of vectors in V is also in V. In other words, vector spaces have to be *closed* under taking linear combinations. Technically speaking, this is an incomplete definition, but it is sufficient for our purposes. One can also replace the word "real" by "complex". A *subspace* of a vector space V is a subset of V which is closed under taking linear combinations.

Given a set $V = \{v_1, \dots, v_m\}$ of vectors, the set of all linear combinations of the v_j forms a vector space, denoted by span (V), or span $(v_1, \dots, v_m)$. The *column space* of a matrix A is the span of its column vectors. The *row space* of A is defined similarly. Note that equation $Ax = b$ (with A not necessarily a square matrix) has a solution if and only if b lies in the column space of A. The coordinates of x form the coefficients of the column vectors of A in a linear combination to form b.

A set $V = \{v_1, \dots, v_m\}$ of (real, complex) vectors is said to be *linearly independent* if

$$a_1 v_1 + a_2 v_2 + \dots a_m v_m = 0 \text{ only happens when } a_1 = a_2 = \dots a_m = 0.$$

Otherwise, the vectors in V are said to be (linearly) *dependent*.

The *dimension* of a vector space is the maximum number of linearly independent vectors in the space. The *basis* of a vector space V is a subset $\{v_1, \dots, v_m\}$ of V which is linearly independent and span $(v_1, \dots, v_m) = V$. It is easy to show that m is actually the dimension of V. A vector space typically has infinitely many bases. All bases of a vector space V have the same size, which is also the dimension of V.

The dimensions of a matrix A's column space and row space are equal, and is referred to as the *rank* of A. This fact is not very easy to show, but not difficult either. Gaussian elimination is of great use here.

The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces by themselves. A set of m vectors in $\mathbb{R}^n$ must be linearly dependent if $m > n$. Similarly, a set of m vectors in $\mathbb{C}^n$ must be linearly dependent if $m > n$.

Given a basis B of V and some vector $v \in V$, there is exactly one way to write v as a linear combination of vectors in B. (Why?)

1.4 Inverses

Given a square matrix A, the *inverse* of A, denoted by A^{-1} is a matrix B such that

$$AB = BA = I, \text{ or } AA^{-1} = A^{-1}A = I.$$

If A and B both have inverses, then the inverse of AB can be calculated easily:

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{8}$$

The same rule holds for 3 or more matrices. For example,

$$(ABCD)^{-1} = D^{-1}C^{-1}B^{-1}A^{-1}.$$

If A has an inverse, it is said to be *invertible*. Not all matrices are invertible. There are many conditions to test if a matrix has an inverse, including: non-singularity, non-zero determinant, non-zero eigenvalues, linearly independent column vectors, linearly independent row vectors.

2 Eigenvalues and eigenvectors

In this section, we shall be concerned with square matrices only, unless stated otherwise. We shall use I to denote the identity matrix.

The *eigenvalues* of a matrix A are the numbers λ such that the equation $Ax = \lambda x$, or $(\lambda I - A)x = 0$, has a non-zero solution vector, in which case the solution vector x is the corresponding *eigenvector*.

Since the all-0 vector, denoted by $\vec{0}$, is always a solution, it would be the only solution if $\det(\lambda I - A) \neq 0$. Hence, the eigenvalues are solutions to the equation $\det(\lambda I - A) = 0$, i.e. $p_A(\lambda) = 0$ where

$$p_A(\lambda) := \det(\lambda I - A)$$

is called the *characteristic polynomial* of A. For example, if

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

then,

$$p_A(\lambda) = \det \begin{bmatrix} \lambda - 2 & -1 \\ +2 & \lambda - 3 \end{bmatrix} = (\lambda - 2)(\lambda - 3) + 2 = \lambda^2 - 5\lambda + 8.$$

If we work on the complex numbers, then equation $p_A(\lambda) = 0$ always has n roots (up to multiplicities). However, we shall be concerned mostly with matrices which have real eigenvalues. We shall establish sufficient conditions for a matrix to have real eigenvalues, as shall be seen in later sections.

Theorem 2.1. *Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of an $n \times n$ complex matrix A, then*

(i) $\lambda_1 + \cdots + \lambda_n = tr\ A$.

(ii) $\lambda_1 \dots \lambda_n = \det A$.

Proof. In the complex domain, $p_A(\lambda)$ has n complex roots since it is a polynomial of degree n. The eigenvalues $\lambda_1, \dots, \lambda_n$ are the roots of $p_A(\lambda)$. Hence, we can write

$$p_A(\lambda) = \prod_i (\lambda - \lambda_i) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

It is evident that

$$\begin{aligned} c_{n-1} &= -(\lambda_1 + \cdots + \lambda_n) \\ c_0 &= (-1)^n \lambda_1 \dots \lambda_n. \end{aligned}$$

On the other hand, by definition we have

$$p_A(\lambda) = \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \dots & \dots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{bmatrix}.$$

Expanding $p_A(\lambda)$ in this way, the coefficient of λ^{n-1} (which is c_{n-1}) is precisely $-(a_{11} + a_{22} + \cdots + a_{nn})$; and the coefficient of λ^0 (which is c_0) is $(-1)^n \det A$ (think carefully about this statement!). $\square$

2.1 The diagonal form

Proposition 2.2. *Suppose the $n \times n$ matrix A has n linearly independent eigenvectors $x_1, \dots, x_n$. Let S be the matrix whose columns are the vectors x_i, then $S^{-1}AS = \Lambda$ where Λ is a diagonal matrix with $(\Lambda)_{ii} = \lambda_i$.*

Proof. The diagonal matrix with $\lambda_1, \dots, \lambda_n$ as diagonal entries is often denoted by diag $(\lambda_1, \dots, \lambda_n)$. Note that since the column vectors of S are independent, S is invertible and S^{-1} makes sense. We want to show $S^{-1}AS = \Lambda$, which is the same as showing $AS = S\Lambda$. Since $Ax_i = x_i\lambda_i$, it follows that

$$AS = A \begin{bmatrix} | & \dots & | \\ x_1 & \dots & x_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ Ax_1 & \dots & Ax_n \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} | & \dots & | \\ \lambda_1 x_1 & \dots & \lambda_n x_n \\ | & \dots & | \end{bmatrix} = S\Lambda.$$

□

In general, if a matrix S satisfies the property that $S^{-1}AS$ is a diagonal matrix, then S is said to *diagonalize* A, and A is said to be *diagonalizable*. It is easy to see from the above proof that if A is diagonalizable by S, then the columns of S are eigenvectors of A; moreover, since S is invertible by definition, the columns of S must be linearly independent. In other words, a matrix is diagonalizable if and only if it has n inpdependent eigenvectors.

Proposition 2.3. *If $x_1, \dots x_k$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots \lambda_k$, then $x_1, \dots x_k$ are linearly independent.*

Proof. When $k = 2$, suppose $c_1x_1 + c_2x_2 = 0$. Multiplying by A gives $c_1\lambda_1x_1 + c_2\lambda_2x_2 = 0$. Subtracting λ_2 times the previous equation we get

$$c_1(\lambda_1 - \lambda_2)x_1 = 0.$$

Hence, $c_1 = 0$ since $\lambda_1 \neq \lambda_2$ and $x_1 \neq 0$. The general case follows trivially by induction. □

Proposition 2.4. *If $\lambda_1, \dots \lambda_n$ are eigenvalues of A, then $\lambda_1^k, \dots \lambda_n^k$ are eigenvalues of A^k. If S diagonalizes A, i.e. $S^{-1}AS = \Lambda$, then $S^{-1}A^kS = \Lambda^k$*

Proof. Trivial. □

2.2 Symmetric and Hermitian matrices

For any two vectors $x, y \in \mathbb{C}^n$, the *inner product* of x and y is defined to be

$$x^*y = \bar{x}^Ty = \bar{x_1}y_1 + \dots + \bar{x_n}y_n$$

Two vectors are *orthogonal* to one another if their inner product is 0. The vector $\vec{0}$ is orthogonal to all vectors. Two orthogonal non-zero vectors must be linearly independent. For, if $x^*y = 0$ and $ax+by = 0$, then $0 = ax^*x + bx^*y = ax^*x$. This implies $a = 0$, which in turns implies $b = 0$ also. With the same reasoning, one easily shows that a set of pairwise orthogonal non-zero vectors must be linearly independent.

If A is any complex matrix, recall that the *Hermitian transpose* A^* of A is defined to be $\bar{A}^T$, and that A is said to be *Hermitian* if $A = A^*$. A real matrix is Hermitian if and only if it is symmetric. Also notice that the diagonal entries of a Hermitian matrix must be real, because they are equal to their respective complex conjugates. The next lemma lists several useful properties of a Hermitian matrix.

Lemma 2.5. *Let A be a Hermitian matrix, then* *(i) for all $x \in \mathbb{C}^n$, x^*Ax is real.*

(ii) every eigenvalue of A is real.

(iii) the eigenvectors of A, if come from distinct eigenvalues, are orthogonal to one another.

Proof. It is straightforward that

(i) $(x^*Ax)^* = x^*A^*x^{**} = x^*Ax$.

(ii) $Ax = \lambda x$ implies $\lambda = \frac{x^*Ax}{x^*x}$.

(iii) Suppose $Ax = \lambda_1 x$, $Ay = \lambda_2 y$, and $\lambda_1 \neq \lambda_2$, then

$$(\lambda_1 x)^* y = (Ax)^* y = x^* A y = x^*(\lambda_2 y).$$

Hence, $(\lambda_1 - \lambda_2)x^*y = 0$, implying $x^*y = 0$.

□

2.3 Orthonormal and unitary matrices

A real matrix Q is said to be *orthogonal* if $Q^TQ = I$. A complex matrix U is *unitary* if $U^*U = I$. In other words, the columns of U (and Q) are *orthonormal*. Clearly being orthogonal is a special case of being unitary. We state without proof a simple proposition.

Proposition 2.6. *Let U be a unitary matrix, then*

(i) $(Ux)^*(Uy) = x^*y$, and $\|Ux\|^2 = \|x\|^2$.

(ii) Every eigenvalue λ of U has modulus 1 (i.e. $|\lambda| = \lambda^*\lambda = 1$).

(iii) Eigenvectors corresponding to distinct eigenvalues of U are orthogonal.

(iv) If U' is another unitary matrix, then UU' is unitary.

2.4 The Spectral theorem

Two matrices A and B are said to be *similar* iff there is an invertible matrix M such that $M^{-1}AM = B$. Thus, a matrix is diagonalizable iff it is similar to a diagonal matrix. Similarity is obviously an equivalent relation. The following proposition shows what is common among matrices in the same similarity equivalent class.

Proposition 2.7. *If $B = M^{-1}AM$, then A and B have the same eigenvalues. An eigenvector x of A corresponds to an eigenvector $M^{-1}x$ of B.*

Proof. $Ax = \lambda x$ implies $(M^{-1}A)x = \lambda M^{-1}x$, or $(BM^{-1})x = \lambda(M^{-1}x)$. □

An eigenvector corresponding to an eigenvalue λ is called a λ-eigenvector. The vector space spanned by all λ-eigenvectors is called the λ-eigenspace. We shall often use V_λ to denote this space if the underlying matrix is clear from context.

Corollary 2.8. *If A and B are similar, then the corresponding eigenspaces of A and B have the same dimension.*

Proof. Suppose $B = M^{-1}AM$, then the mapping $\phi : x \to M^{-1}x$ is an invertible linear transformation from one eigenspace of A to the corresponding eigenspace of B.[1] □

If two matrices A and B are similar, then we can say a lot about A if we know B. Hence, we would like to find B similar to A where B is as "simple" as possible. The first "simple" form is the upper-triangular form, as shown by the following Lemma, which is sometime referred to as the Jacobi Theorem.

Lemma 2.9 (Schur's lemma). *For any $n \times n$ matrix A, there is a unitary matrix U such that $B = U^{-1}AU$ is upper triangular. Hence, the eigenvalues of A are on the diagonal of B.*

Proof. We show this by induction on n. The lemma holds when $n = 1$. When $n > 1$, over $\mathbb{C}$ A must have at least one eigenvalue λ_1. Let x_1' be a corresponding eigenvector. Use *Gram-Schmidt* process to extend x_1' to an orthonormal basis $\{x_1, x_2, \dots, x_n\}$ of $\mathbb{C}^n$. Let U_1 be the matrix whose columns are these vectors in order. From the fact that $U_1^{-1} = U_1^*$, it is easy to see that

$$U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ 0 & * & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots \\ 0 & * & * & \dots & * \end{bmatrix}.$$

Now, let $A' = (U_1^{-1}AU_1)_{11}$ (crossing off row 1 and column 1 of $U_1^{-1}AU_1$). Then by induction there exists an $(n-1) \times (n-1)$ unitary matrix M such that $M^{-1}A'M$ is upper triangular. Let U_2 be the $n \times n$ matrix obtained by adding a new row and new column to M with all new entries equal 0 except $(U_2)_{11} = 1$. Clearly U_2 is unitary and $U_2^{-1}(U_1^{-1}AU_1)U_2$ is upper triangular. Letting $U = U_1U_2$ completes the proof. □

The following theorem is one of the most important theorems in elementary linear algebra, beside the Jordan form.

Theorem 2.10 (Spectral theorem). *Every real symmetric matrix can be diagonalized by an orthogonal matrix, and every Hermitian matrix can be diagonalized by a unitary matrix:*

$$\textit{(real case)} \quad Q^{-1}AQ = \Lambda \quad \textit{(complex case)} \quad U^{-1}AU = \Lambda$$

Moreover, in both cases all the eigenvalues are real.

Proof. The real case follows from the complex case. Firstly, by Schur's lemma there is a unitary matrix U such that $U^{-1}AU$ is upper triangular. Moreover,

$$(U^{-1}AU)^* = U^*A^*(U^{-1})^* = U^{-1}AU,$$

i.e. $U^{-1}AU$ is also Hermitian. But an upper triangular Hermitian matrix must be diagonal. The realness of the eigenvalues follow from Lemma 2.5. □

References

[1] M. ARTIN, *Algebra*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1991.

[2] G. STRANG, *Linear algebra and its applications*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, second ed., 1980.

[1] I have not define linear transformation yet. The thing to remember is that if there is an invertible linear transformation from one vector space to another, then the two vector spaces have the same dimension. Invertible linear transformations are like isomorphisms or bijections, in some sense. A curious student should try to prove this fact directly without using the term linear transformation.