Linear Algebra Supplementary Notes

Spring 2009

David Lyons
Mathematical Sciences
Lebanon Valley College

Linear Algebra
Supplementary Notes

Spring 2009

David Lyons
Mathematical Sciences
Lebanon Valley College

References given in the margins are for the text: Strang, *Linear Algebra and Its Applications*, 4th edition

Contents

1 Vectors, Linear Maps, and Matrices

Euclidean Space

Let $\mathbf{R}$ denote the set of real numbers. For a given positive integer n, the set

$$\mathbf{R}^n = \underbrace{\mathbf{R} \times \mathbf{R} \times \cdots \times \mathbf{R}}_{n \text{ factors}} = \{(x_1, x_2, x_3, \ldots, x_n)\}$$

of ordered n-tuples of real numbers is called ***n-dimensional Euclidean space*** or *Euclidean n-space*.

The numbers $\{x_i\}$ are called the ***coordinates*** of the point $x = (x_1, \ldots, x_n)$. The space $\mathbf{R}^1 = \mathbf{R}$ is the real number line, $\mathbf{R}^2$ is the plane of high school geometry and algebra, and $\mathbf{R}^3$ is the mathematical abstraction for the familiar 3-space in which we live. The space $\mathbf{R}^0$ is defined to be the one point set $\mathbf{R}^0 = \{0\}$.

Vector Operations

Points in Euclidean space are sometimes called ***vectors***, and real numbers are sometimes called ***scalars***. In multivariate calculus and physics courses, vectors are often denoted using an arrow decoration like "$\vec{x}$", but it is also common to omit any decorations, as we choose to do in these notes. Given two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ and a scalar α, we define two operations called ***scaling*** and ***vector addition***. The vector αx is defined to be

(1.1) $$\alpha x = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

We say that αx is the vector x *scaled by the factor* α. The vector $x + y$, called the ***vector sum of x and y***, is defined to be

(1.2) $$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Scalar multiplication and vector addition obey the following distributive law, which is easy to verify.

(1.3) **Distributive law for vector operations.** For any scalar α and vectors x, y of the same dimension, we have

$$\alpha(x + y) = \alpha x + \alpha y.$$

The vector $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith coordinate and zeroes in all other coordinates is called the ***ith standard basis vector*** in $\mathbf{R}^n$. In $\mathbf{R}^2$, the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$ are also called $\mathbf{i}$ and $\mathbf{j}$, respectively. In $\mathbf{R}^3$, the standard basis vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ are also called $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, respectively. Given a vector $x = (x_1, \ldots, x_n)$, we have the following representation of x as a sum of scalar multiples of the standard basis vectors (note that the summation sign indicates ***vector*** addition).

(1.4) $$x = \sum_{i=1}^{n} x_i e_i$$

The ***inner product*** or ***dot product*** of two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is defined to be the scalar quantity

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Linear Maps and Matrices

Because vector operations are useful, it is natural to consider functions or maps that respect vector operations. We call these maps *linear*.

(1.5) **Definition of Linear Map.** A function or map $L: \mathbf{R}^n \to \mathbf{R}^m$ is called ***linear*** if

$$\begin{aligned} L(\alpha x) &= \alpha L(x) \\ L(x+y) &= L(x) + L(y) \end{aligned}$$

for all vectors x, y in $\mathbf{R}^n$ and scalars α in $\mathbf{R}$. We say that L ***preserves*** or ***respects*** vector operations of scaling and addition.

Given a vector $x = (x_1, \ldots, x_n)$ and a linear map $L: \mathbf{R}^n \to \mathbf{R}^m$, we have

(1.6)
$$\begin{aligned} L(x) &= L(x_1, x_2, \ldots, x_n) \\ &= L\left(\sum_{j=1}^{n} x_j e_j\right) \\ &= \sum_{j=1}^{n} L(x_j e_j) \\ &= \sum_{j=1}^{n} x_j L(e_j) \end{aligned}$$

A consequence of this equation is that a linear map is determined by its values on the standard basis vectors $e_1, e_2, \ldots, e_n$. We can write an explicit formula for the coordinates $(y_1, y_2, \ldots, y_m)$ of the value $y = L(x)$. Let $f_1, f_2, \ldots, f_m$ denote the standard basis vectors for $\mathbf{R}^m$. Then we have

(1.7)
$$\begin{aligned} y_i &= f_i \cdot L(x) \\ &= f_i \cdot \sum_{j=1}^{n} x_j L(e_j) \\ &= \sum_{j=1}^{n} x_j \; f_i \cdot L(e_j). \end{aligned}$$

This last expression shows that the values of a linear function are completely determined by the numbers

(1.8)
$$a_{ij} = f_i \cdot L(e_j)$$

where i is in the range $1 \le i \le m$ and j is in the range $1 \le j \le n$. We call the rectangular array of numbers

(1.9)
$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

the ***matrix*** for L, and denote it by $[L]$ or $[a_{ij}]$. The numbers a_{ij} are called the ***entries*** of the matrix. Rows of the matrix are numbered top to bottom, and columns are numbered left to right. A matrix with m rows and n columns is called an $m \times n$ matrix.

Written out fully, the equations for the value $y = L(x)$ are the following.

(1.10)
$$\begin{array}{ccccccccc} y_1 & = & a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n \\ y_2 & = & a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n \\ & \vdots & & & & & & & \\ y_i & = & a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{in}x_n \\ & \vdots & & & & & & & \\ y_m & = & a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n \end{array}$$

The expressions on the right sides are dot products.

(1.11)
$$y_i = (\text{the } i\text{th row of } [L]) \cdot x$$

For the special case when the input vector x is a standard basis vector e_j, we see that

(1.12)
$$y = L(e_j) = (a_{1j}, a_{2j}, \ldots, a_{mj})$$

or, in words,

(1.13)
$$L(e_j) \text{ is the } j\text{th column of } [L].$$

We summarize the relationship between linear maps and matrices.

(1.14) **Proposition.** There is a one-to-one correspondence between linear maps $\mathbf{R}^n \to \mathbf{R}^m$ and $m \times n$ matrices, as follows. Given a linear map L, it corresponding matrix $[a_{ij}]$ is given by equation (1.8). Conversely, given a matrix $[a_{ij}]$ the corresponding linear map L is given by $L(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$ where $y_1, \ldots, y_m$ are given by equations (1.10).

Exercises

1. Write each of the following vectors x in the form $(x_1, x_2, \ldots, x_n)$ and $\sum x_i e_i$. For $n = 2, 3$, also write x using $\mathbf{i}, \mathbf{j}, \mathbf{k}$ notation.
   Example: Given $x = 3(2, 4, -1)$, write $x = (6, 12, -3) = 6e_1 + 12e_2 - 3e_3 = 6\mathbf{i} + 12\mathbf{j} - 3\mathbf{k}$.
   (a) $x = (3, 2) - (5, -2)$
   (b) $x = 2(-1, 2, 1) + 3(2, -2, 0)$
   (c) $x = 2e_1 - 3e_2 + 4e_4 - (e_1 - e_2 + e_3)$
2. Prove the distributive law (1.3).
3. Verify equation (1.4).
4. Show that for $x = (x_1, \ldots, x_n)$, we have $x_i = e_i \cdot x$.
5. Show that
   $$e_i \cdot e_j = \delta_{ij}$$
   where δ_{ij}, called the ***Kronecker delta***, is given by
   $$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

6. Let $L: \mathbf{R}^2 \to \mathbf{R}^2$ be a linear map such that $L(e_1) = (2,3)$ and $L(e_2) = (-1,-2)$. Find $L(1,2)$.

7. Let $L: \mathbf{R}^3 \to \mathbf{R}$ be a linear map. Find $L(\mathbf{k})$ if $L(\mathbf{i}) = 2$, $L(\mathbf{j}) = -1$, and $L(1,2,3) = 0$.

8. Let $L: \mathbf{R}^3 \to \mathbf{R}^2$ be a linear map with $L(e_1) = (1,2)$, $L(e_2) = (-1,1)$, and $L(e_3) = (0,1)$.

   (a) Write the matrix for L.

   (b) Evaluate $L(2,1,3)$.

   (c) Evaluate $L(0,1,1)$.

9. Let $L: \mathbf{R}^2 \to \mathbf{R}^3$ be a linear map with the following matrix.

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

   (a) Evaluate $L(1,2)$.

   (b) Evaluate $L(-2,1)$.

[2T p.126]

10. Show that the two linearity properties in the definition (1.5) of linear map are equivalent to the single property

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \tag{1.15}$$

   for all vectors x, y and scalars α, β.

11. Justify each equality in (1.6).

12. In formulas (1.7) and (1.8), what is the difference between e_i and f_i? Aren't both of these vectors with 1 in the ith coordinate and 0 elsewhere?

13. The dot product has the following properties that look like the properties in the definition of linear map.

$$\begin{aligned} u \cdot (\alpha v) &= \alpha u \cdot v \\ u \cdot (v + w) &= u \cdot v + u \cdot w \end{aligned}$$

   for all u, v, w in $\mathbf{R}^n$ and scalars α.

   (a) Show that these properties hold.

   (b) Exactly where in this section did we use these properties?

14. Justify the steps of the derivation (1.7).

15. Verify (1.13).

16. Prove that the map $r: \mathbf{R}^2 \to \mathbf{R}^2$ which rotates the plane 1/4 turn counterclockwise about the origin is a linear map. Find the matrix for this linear map.

2 Matrix Multiplication

The purpose of this section is to motivate and explain the fundamental operation of matrix algebra called matrix multiplication. Matrix multiplication arises naturally as the answer to a question about the composition of linear maps.

Let $T\colon \mathbf{R}^u \to \mathbf{R}^v$ and $S\colon \mathbf{R}^v \to \mathbf{R}^w$ be linear maps with matrices $[T] = [t_{ij}]$ and $[S] = [s_{ij}]$. It is a simple exercise to check that the composition $S \circ T\colon \mathbf{R}^u \to \mathbf{R}^w$ is also a linear map. Therefore the composition $S \circ T\colon \mathbf{R}^u \to \mathbf{R}^w$ has a matrix $[S \circ T]$. Intuition says that the entries of $[S \circ T]$ must be formed from the coefficients t_{ij} and s_{ij}, but how?

First we need to establish a few names. Let $R = S \circ T$ and let $[R] = [r_{ij}]$. Let $e_1, \dots, e_u$ be the standard basis for $\mathbf{R}^u$, let $f_1, f_2, \dots, f_v$ be the standard basis for $\mathbf{R}^v$, and and let $g_1, \dots, g_w$ be the basis for $\mathbf{R}^w$. We will use the formula (1.8)

$$r_{ij} = g_i \cdot R(e_j)$$

to solve the problem. We start with $R(e_j) = S(T(e_j))$. Formula (1.13) says $T(e_j)$ is the jth column of the matrix $[T]$. This is

$$T(e_j) = (t_{1j}, t_{2j}, \dots, t_{vj}) = t_{1j} f_1 + t_{2j} f_2 + \cdots + t_{vj} f_v.$$

Thus we have

$$\begin{aligned} R(e_j) &= S(T(e_j)) \\ &= S(t_{1j}, t_{2j}, \dots, t_{vj}) \\ &= S(t_{1j} f_1 + t_{2j} f_2 + \cdots + t_{vj} f_v) \\ &= t_{1j} S(f_1) + t_{2j} S(f_2) + \cdots + t_{vj} S(f_v). \end{aligned}$$

The next step is to take the dot product with g_i, then recognize that $g_i \cdot S(f_k) = s_{ik}$.

$$\begin{aligned} r_{ij} &= g_i \cdot R(e_j) \\ &= g_i \cdot (t_{1j} S(f_1) + t_{2j} S(f_2) + \cdots + t_{vj} S(f_v)) \\ &= t_{1j}\ g_i \cdot S(f_1) + t_{2j}\ g_i \cdot S(f_2) + \cdots + t_{vj}\ g_i \cdot S(f_v) \\ &= t_{1j} s_{i1} + t_{2j} s_{i2} + \cdots + t_{vj} s_{iv}. \end{aligned}$$

The pattern looks nicer if we switch right and left in each summand in the last expression.

$$r_{ij} = s_{i1} t_{1j} + s_{i2} t_{2j} + \cdots + s_{iv} t_{vj}$$

We recognize the right side of the last expression as the dot product of the ith row of $[S]$ with the jth column of $[T]$.

(2.16) $$r_{ij} = (i\text{th row of } [S]) \cdot (j\text{th column of } [T])$$

[1D p.24] This result motivates the following definition.

(2.17) **Definition of Matrix Multiplication.** Given a $w \times v$ matrix A with entries a_{ij} and a $v \times u$ matrix B with entries b_{ij}, the *product matrix* AB is defined to be the $w \times u$ matrix with the following entry in the ith row and jth column.

$$a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{iv} b_{vj} = \sum_{k=1}^{v} a_{ik} b_{kj}$$

Here is a restatement of the connection between composition of linear maps and matrix multiplication. This is a sort of "chain rule" for linear algebra.

(2.18) **Composition and Matrix Multiplication.** Let S and T be linear maps such that their composition is defined. Then we have

$$[S \circ T] = [S][T].$$

Matrix multiplication can be used to compute values for linear maps. First, we make vectors into matrices. Given a vector $x = (x_1, x_2, \ldots, x_n)$, we define the ***matrix*** for x, denoted $[x]$, to the the $n \times 1$ matrix

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

A matrix with one column is called a ***column matrix***. The identification of vectors with column matrices is standard and nearly universal. It is normal usage to use the symbol x to refer to the vector and the column matrix without bothering to use square brackets or other notation to make a distinction. With this invention, we have a way to compute $L(x)$. The proof is an exercise.

(2.19) **Evaluation and Matrix Multiplication.** The matrix product $[L][x]$ is the column matrix for $L(x)$. In other words,

$$[L(x)] = [L][x].$$

Block Multiplication

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix, so that the product AB is defined. If we subdivide A and B into blocks (submatrices) of compatible sizes, we can perform matrix multiplication using the blocks. Here is a picture of the situation.

$$\underbrace{\left[\begin{array}{c|c} A_{11} & A_{12} \\ m_1 \times n_1 & m_1 \times n_2 \\ \hline A_{21} & A_{22} \\ m_2 \times n_1 & m_2 \times n_2 \end{array}\right]}_{A} \underbrace{\left[\begin{array}{c|c} B_{11} & B_{12} \\ n_1 \times p_1 & n_1 \times p_2 \\ \hline B_{21} & B_{22} \\ n_2 \times p_1 & n_2 \times p_2 \end{array}\right]}_{B}$$

The result is the following.

$$AB = \left[\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right] \tag{2.20}$$

Exercises

1. Perform the matrix multiplications below.

   (a)

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

(b)

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

2. Let $L: \mathbf{R}^2 \to \mathbf{R}$ and $K: \mathbf{R} \to \mathbf{R}^3$ be linear maps such that $L(\mathbf{i}) = 1$, $L(\mathbf{j}) = -2$ and $K(1) = (2, 1, 3)$.
   (a) Write the matrices for L and K.
   (b) Find the matrix for $K \circ L$.
   (c) Find $(K \circ L)(2, -1)$.
3. Prove that the composition of two linear maps is a linear map.
4. Prove (2.19).
5. Let r_θ and r_φ be rotations of the plane about the origin by θ and φ radians, respectively. Use matrix multiplication to derive the formulas for cosines and sines of a sum of two angles.
6. Let z_0 be a fixed complex number, and let $f: \mathbf{C} \to \mathbf{C}$ be the map given by $f(w) = z_0 w$. Let $I: \mathbf{R}^2 \to \mathbf{C}$ be the identification of $\mathbf{R}^2$ with $\mathbf{C}$ given by $I(x, y) = x + iy$. Let $L = I^{-1} \circ f \circ I$.
   (a) Verify that the map $L: \mathbf{R}^2 \to \mathbf{R}^2$ is linear.
   (b) Find the matrix for L.
   (c) Let $z_0 = e^{i\theta}$, where θ is a real number. Verify that L is the same thing as r_θ from the previous problem.
7. Prove (2.20).

3 Complex Numbers

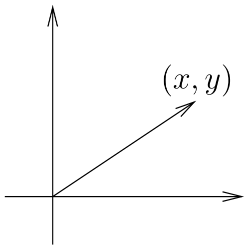


Figure 1
The vector (x, y)

Motivation

The Euclidean line is the set used to measure a great many "real world" phenomena such as time, distance, mass, temperature, and so on. It is for this reason that we call the line *real.* Analysis of mathematical models involving functions on the real line is made rich and powerful by the algebraic structure of the real numbers. Key features of this algebra are the operations of addition and multiplication, together with the distributive law which governs their interaction.

The Euclidean plane is the set used to measure any kind of data that consists of *pairs* of real numbers. Examples include graphs of functions on the real line and two-dimensional geometric figures. A natural question is whether the algebra of the line extends in any useful ways to the plane. The answer is yes. The extension of the algebra of the reals to the plane forms the *complex* numbers.

Vectors and vector operations

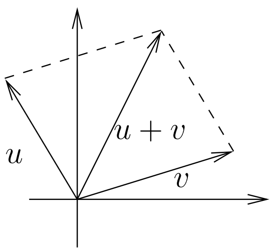


Figure 2
The parallelogram law
of vector addition

The complex numbers, denoted $\mathbf{C}$, is defined to be the set of points in the Euclidean plane, that is, the set

$$\mathbf{R}^2 = \mathbf{R} \times \mathbf{R} = \{(x, y) \mid x, y \in \mathbf{R}\}$$

where $\mathbf{R}$ denotes the set of real numbers. We can think of a point (x, y) as a *vector*, or directed line segment, which begins at the origin and ends at (x, y). In drawings, we represent vectors by arrows as in Figure 1. This comes from physics where vectors model physical quantities such as displacement, velocity or force.

The way in which physical vector quantities combine leads to the operation of *vector addition.* The sum of two vectors (a, b), (c, d) is defined to be

$$(a, b) + (c, d) = (a + c, b + d). \tag{3.1}$$

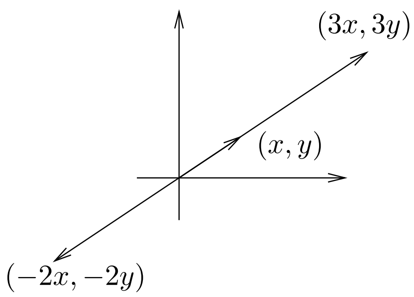


Figure 3
Scalar multiples of
a vector

Geometrically, two vectors v, w and their sum $v + w$ form two sides and a diagonal of a parallelogram; for this reason, equation (3.1) is called the *parallelogram law.* See Figure 2.

Given a vector $v = (x, y)$ and a real number k, the *scalar product* of k times v, denoted kv, is defined to be (kx, ky). The vector kv is $|k|$ times as long as v and points in the same direction as v if k is positive and in the opposite direction from v if k is negative. See Figure 3.

Vector addition and scalar multiplication obey the *distributive law.* For any two vectors v, w and any real number k, we have

$$k(v + w) = kv + kw. \tag{3.2}$$

Notice that vector operations say how to add two points in the plane, and how to multiply a point in the plane by a real number, but *not* how to multiply two vectors. We explain how to multiply complex numbers in §3 below.

Points, vectors and scalars as complex numbers

While points and vectors may seem to be different types of objects, the set which represents them both is the same, namely $\mathbf{R}^2$. We refer to an ordered pair (x, y) alternately as a point or as a vector, depending on convenience. This takes some getting used to, but it turns out to be useful to think of complex numbers both ways.

Similarly, a scalar $k \in \mathbf{R}$ and a vector $v \in \mathbf{R}^2 = \mathbf{C}$ seem to be different kinds of things. It is also natural, however, to think of the real number line as a *subset* of the plane. The x-axis, that is, the set of points of the form $\{(x, 0) \mid x \in \mathbf{R}\}$, is in one-to-one correspondence with the real numbers via $(x, 0) \leftrightarrow x$. In this way, we think of the real number x as the complex number $(x, 0)$. This is what we mean when we say that that the Euclidean plane *extends* the Euclidean line.

To summarize this subsection: Complex numbers may be thought of as either points or vectors in the plane; real numbers are also complex numbers since points on the real line (the x-axis) are also points in the plane.

Polar coordinates

The *norm* of a point $p = (x, y)$, denoted $|p|$, is defined to be

$$|p| = \sqrt{x^2 + y^2}.$$

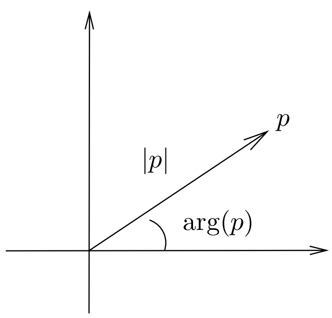


Figure 4
Norm and argument

Geometrically, the norm of a point is its length as a vector, which is the same as its distance from the origin $(0, 0)$. The *argument* of p, denoted $\arg(p)$, is the measure in radians of the directed angle with vertex at the origin, whose initial ray is the positive x-axis and whose terminal ray passes through p. See Figure 4. We agree that two arguments are the same if they differ by an integer multiple of 2π.

The norm and argument of a point are called the *polar coordinates* of that point. By contrast, the numbers x, y in the ordered pair $(x, y) \in \mathbf{R}^2$ are called the *rectangular* or *Cartesian* coordinates of the point. In these notes, we shall use the notation $P_{(r,\theta)}$ to denote the point whose polar coordinates are (r, θ).

Multiplication

The "obvious" way to attempt multiplication of vectors is to mimic vector addition by defining $(a, b) \cdot (c, d) = (ac, bd)$. However, this leads to an unacceptable conflict, as we shall describe.

Let k be a real number and let $v = (x, y)$ be a vector. As a point in the plane, the real number k is $(k, 0)$, so we would have $kv = (k, 0) \cdot (x, y) = (kx, 0)$. On the other hand, we have already defined (scalar multiplication) kv to be $k \cdot (x, y) = (kx, ky)$. But $(kx, 0)$ does not equal (kx, ky) unless one or both of k, y are zero. This example says that if we want scalar multiplication of vectors to be compatible with multiplying pairs of complex numbers, we must look for a different way to multiply points than component-wise.

To see another way, we examine some multiplications of real numbers using polar coordinates.

$$\begin{array}{rcccccl} P_{(2,0)} \cdot P_{(3,0)} & = & 2 \cdot 3 & = & 6 & = & P_{(6,0)} \\ P_{(2,0)} \cdot P_{(3,\pi)} & = & 2 \cdot -3 & = & -6 & = & P_{(6,\pi)} \\ P_{(2,\pi)} \cdot P_{(3,\pi)} & = & -2 \cdot -3 & = & 6 & = & P_{(6,2\pi) \,=\, P(6,0)} \end{array}$$

These examples suggest that in polar coordinates, norms multiply and arguments add. That is the inspiration for the following definition of multiplication.

$$P_{(r,\theta)} \cdot P_{(s,\varphi)} = P_{(rs,\theta+\phi)} \tag{3.3}$$

This multiplication does indeed extend the ordinary multiplication of real numbers, and resolves the conflict from the previous paragraph. It is not immediately obvious, but it turns out that other desirable properties of the algebra of the real numbers also hold in the plane. In particular, extended addition and multiplication satisfy the distributive law

$$z(u+v) = zu + zv \tag{3.4}$$

for all z, u, v in the plane. Note that we normally omit the dot between two points and write zw to denote the product of two points z and w, just as we do for real numbers.

Real and imaginary parts, rectangular form

The complex number $(0,1) = P_{(1,\pi/2)}$ is denoted by the symbol i. It has the curious property that its square is negative one. Note that the number $z = (x, y)$ is equal to $x(1,0) + y(0,1) = x + yi$. The real number x is called the *real part* and the real number y is called the *imaginary part* of $z = x + yi$, and we write $\mathrm{Re}(z) = x$ and $\mathrm{Im}(z) = y$. A number of the form $(0, y) = yi$ is called *pure imaginary.* The coordinates (x, y) of z are called the *rectangular* or *Cartesian* coordinates, and the expression $x + yi$ is called the *rectangular form* of z.

Next we demonstrate how to multiply two complex numbers in rectangular form without converting to polar coordinates. Let $(x, y) = x+yi$ and let $(u, v) = u+vi$ be two complex numbers. Using the distributive law, we have

$$\begin{array}{rcll} (x,y)\cdot(u,v) & = & (x+yi)(u+vi) & \\ & = & xu + yui + xvi + yvi^2 & \text{(distributing)} \\ & = & xu + yui + xvi - yv & (i^2 = -1) \\ & = & (xu - yv) + (yu + xv)i & \text{(collecting real and imaginary parts).} \end{array}$$

Exponential notation and polar form

The elementary functions of a single real variable, including polynomials, rational functions, sine, cosine, and the natural exponential function can be extended to functions of a complex variable. This is done in the subject of *complex analysis.* It turns out that the natural exponential function (that is, the function that sends the complex number z to the complex number e^z) has the following formula for pure imaginary exponents.

$$e^{it} = \cos t + i \sin t \qquad \text{for } t \in \mathbf{R} \tag{3.5}$$

Equation (3.5) is not supposed to be obvious; it takes a bit of work to even define what e^z means for a complex number z. However, it is not necessary to know the theory of the complex exponential function to use (3.5). For now, you may think of (3.5) as a definition of the symbols e^{it}.

Recall that $p = (\cos t, \sin t)$ is the point on the unit circle intersected by the terminal ray of a directed angle of t radians in standard position. See Figure 5. It follows that $e^{it} = P_{(1,t)}$ has norm 1 and argument t. If z is an arbitrary complex number with norm r and argument θ, then we have

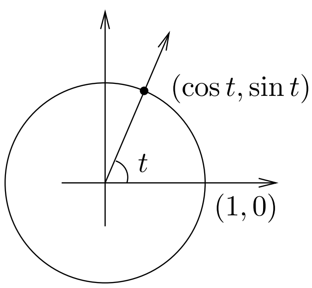


Figure 5
Point on
the unit circle

$$z = P_{(r,\theta)} = rP_{(1,\theta)} = re^{i\theta}. \tag{3.6}$$

The expression on the right-hand side of the above equation is called the *polar form* of the complex number z.

Conjugation

Let $z = (x, y) = P_{(r,\theta)} = x + iy = re^{i\theta}$ be a complex number. The *conjugate* of z, denoted $\overline{z}$ is defined to be

$$\overline{z} = (x, -y) = P_{(r,-\theta)} = x - iy = re^{-i\theta}. \tag{3.7}$$

Geometrically, the conjugate of a point is the reflection of that point across the x-axis. Here is a useful relation that involves the conjugate.

$$z\overline{z} = |z|^2 \tag{3.8}$$

Exercises

1. What is the difference between polar coordinates and polar form? What is the difference between rectangular coordinates and rectangular form? Write formulas for converting from polar to rectangular coordinates and vice-versa.

2. Express each of the following in rectangular and polar form.

   (a) $3(2 - i) + 6(1 + i)$
   (b) $\left(2e^{i\pi/6}\right)\left(3e^{-i\pi/3}\right)$
   (c) $(2 + 3i)(4 - i)$
   (d) $(1 + i)^3$

3. Prove the following property of norm, for all complex numbers z, w.

   $$|zw| = |z||w|$$

   Do the proof using rectangular and polar forms. Which is easier?

4. Prove the following property of norm, called the *triangle inequality*. For any two complex numbers z, w, we have

   $$|z + w| \le |z| + |w|.$$

5. Prove (3.8).

6. Let p and q be complex numbers. Prove that the distance (ordinary distance between points in the plane) between p and q is $|p - q|$. Hint: Use rectangular form.

7. Verify the distributive law (3.4). Suggestion: First prove case (i) where z is a real number. Next prove case (ii) where z is has norm 1 (use the fact that the diagonal of a rotated parallelogram is the rotation of the diagonal of the original parallelogram. Finally prove the general case where $z = re^{i\theta}$.

8. This exercise outlines the definitions and properties of complex division.

   (a) Notice that the real number 1, considered as the complex number (1,0), has the property that $1z = z$ for any complex number z. For this reason, 1 is called a *multiplicative identity* for **C**. Are there any other multiplicative identities? That is, does there exist any other complex number u with the property that $uz = z$ for every complex number z? If so, find one. If not, explain why none exists.

   (b) Given two complex numbers u and v with $v \neq 0$, the quotient of u divided by v, denoted u/v, is defined to be the complex number z with the property that $u = vz$. Write expressions for $z = u/v$ and $w = 1/v$ in polar form if $u = re^{i\theta}$ and $v = se^{i\varphi}$.

   Note a new definition: we call $1/v$ the *reciprocal* or *multiplicative inverse* of v, and also write it as v^{-1}. Notice that multiplicative identity, division and multiplicative inverse are defined the same as for real numbers.

   (c) Suppose that z, w are nonzero complex numbers. Prove that

   $$(1/z)(1/w) = 1/(zw).$$

   (d) Find the multiplicative inverse of $z = x + iy$ in rectangular form (assume $z \neq 0$). Hint: Multiply $1/z$ by $\overline{z}/\overline{z}$ and use the previous exercise. This is called *rationalizing the denominator*.

9. Express each of the following in rectangular and polar form.

   (a) $\dfrac{2+i}{3-i}$

   (b) $\dfrac{1+2i}{1-2i}$

   (c) $\dfrac{2e^{i\pi/4}}{3e^{-i\pi/2}}$

10. Verify the following formulas. For any complex number z, we have

    (a) $\operatorname{Re}(z) = \dfrac{z+\overline{z}}{2}$, and

    (b) $\operatorname{Im}(z) = \dfrac{z-\overline{z}}{2i}$.

11. Given a nonzero complex number z, explain why z has exactly two square roots, and explain how to find them.

12. Find all complex solutions of the following equations.

(a) $z^2 + 3z + 5 = 0$

(b) $(z - i)(z + i) = 1$

(c) $\dfrac{2z + i}{-z + 3i} = z$

13. Derive the double angle identities for $\cos 2\theta$ and $\sin 2\theta$ by computing $(e^{i\theta})^2$ two ways: in polar form and in rectangular form. Then compare real and imaginary parts.

14. Graph the solutions to the following complex equations.

(a) $|z - 2| = 3$

(b) $|4z - 2i| = 3$

(c) $\operatorname{Im}(z) = 3$

(d) $\operatorname{Im}(2e^{i\pi/4}z - 2 + 3i) = 0$

4 Change of Basis

Matrix for a linear map

[Sec. 2.6]

Let V, W be vector spaces with bases $v_1, v_2, \ldots, v_n$ for V and $w_1, w_2, \ldots, w_m$ for W, and let $L: V \to W$ be a linear map (that is, L satisfies $L(x+y) = L(x)+L(y)$ and $L(\alpha x) = \alpha L(x)$ for all vectors x, y in V and all scalars α). Let v be the name of the basis $v_1, v_2, \ldots, v_n$ (be careful! each of the v_j is a *vector*, and v is an *ordered list* of vectors) and likewise let w denote the basis $w_1, \ldots, w_m$. Since w is a basis for W, we can write $L(v_1), L(v_2), \ldots, L(v_n)$ as linear combinations of the w_j.

$$\begin{aligned}
L(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\
L(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\
&\vdots \\
L(v_j) &= a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m \\
&\vdots \\
L(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m
\end{aligned}$$

In summation notation, this is

(4.1) $$L(v_j) = \sum_{i=1}^{m} a_{ij}w_i, \qquad 1 \le j \le n.$$

[2U p.129]

We define the ***matrix for L with respect to the bases v and w***, denoted $[L]_{v \text{ to } w}$, to be the matrix whose i, j entry is a_{ij}.

(4.2) $$[L]_{v \text{ to } w} = [a_{ij}]$$

A special case is when $V = W$ and $v_j = w_j$ for all j. Then the matrix $[L]_{v \text{ to } v}$ is also called ***the matrix for L with respect to the basis v***, or even more simply, ***the matrix for L*** when the basis is understood. The significance of matrix representation is that matrix multiplication computes values of L, in the following way. Let x be a vector in V and let $y = L(x)$. Write x and y in terms of the given bases.

$$\begin{aligned}
x &= x_1v_1 + x_2v_2 + \cdots + x_nv_n \\
y &= y_1w_1 + y_2w_2 + \cdots + y_mw_m
\end{aligned}$$

Let $[x]_v$ $[y]_w$ denote the column vectors

(4.3) $$[x]_v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad [y]_w = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Then we have

(4.4) $$[L]_{v \text{ to } w}\,[x]_v = [y]_w.$$

Matrix multiplication also computes the composition of linear maps. Let $L: V \to W$ be a linear map, where v, w are bases as above. Let $M: W \to U$ be another linear map, and let $u = (u_1, u_2, \ldots, u_t)$ be a basis for U. Then we have

$$[M \circ L]_{v \text{ to } u} = [M]_{w \text{ to } u}[L]_{v \text{ to } w}. \tag{4.5}$$

[2V p.131] Here is the proof. It is an exercise for the reader to justify each step. Let $[b_{ij}] = [M]_{w \text{ to } u}$ and let $[c_{ij}] = [M \circ L]_{v \text{ to } u}$. We have

$$\begin{aligned}
(M \circ L)(v_j) &= M\left(\sum_{k=1}^{m} a_{kj} w_k\right) \\
&= \sum_{k=1}^{m} a_{kj} M(w_k) \\
&= \sum_{k=1}^{m} \left(a_{kj} \sum_{i=1}^{t} b_{ik} u_i\right) \\
&= \sum_{k=1}^{m} \left(\sum_{i=1}^{t} b_{ik} a_{kj} u_i\right) \\
&= \sum_{i=1}^{t} \left(\sum_{k=1}^{m} b_{ik} a_{kj} u_i\right) \\
&= \sum_{i=1}^{t} \left(\sum_{k=1}^{m} b_{ik} a_{kj}\right) u_i.
\end{aligned}$$

Therefore, we have $c_{ij} = \sum_{k=1}^{m} b_{ik} a_{kj}$, which is the same as (4.5).

Change of basis and matrices

[Sec. 5.6]

Consider the special case of the identity map $I: V \to V$ and two bases $v = (v_1, v_2, \ldots, v_n)$, $w = (w_1, w_2, \ldots, w_n)$ for V. The matrix $[I]_{v \text{ to } w}$ for the identity map with respect to the bases v and w is called the ***change of basis matrix from*** v ***to*** w. Its entries a_{ij} are given by writing the v_j as linear combinations of the w_i.

$$\begin{aligned}
I(v_1) = v_1 &= a_{11} w_1 + a_{21} w_2 + \cdots + a_{n1} w_n \\
v_2 &= a_{12} w_1 + a_{22} w_2 + \cdots + a_{n2} w_n \\
&\vdots \\
v_j &= a_{1j} w_1 + a_{2j} w_2 + \cdots + a_{nj} w_n \\
&\vdots \\
v_n &= a_{1n} w_1 + a_{2n} w_2 + \cdots + a_{nn} w_n
\end{aligned}$$

In summation notation, this is

$$I(v_j) = v_j = \sum_{i=1}^{n} a_{ij} w_i, \qquad 1 \le j \le n. \tag{4.6}$$

We conclude with a formula that gives the relationship between two matrices for the same linear map with respect to two bases. Let $L: V \to V$ be a linear map, let $B = [L]_{v \text{ to } v}$ be the matrix for L with respect to the basis v and let $A = [L]_{w \text{ to } w}$ be the matrix for L with respect to the basis w. Let $M = [I]_{v \text{ to } w}$ be the change of basis matrix from v to w. Then we have

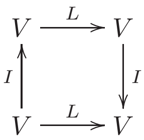
$$B = M^{-1}AM, \qquad \text{which is the same as} \tag{4.7}$$

$$[L]_{v \text{ to } v} = [I]_{w \text{ to } v}[L]_{w \text{ to } w}[I]_{v \text{ to } w}. \tag{4.8}$$

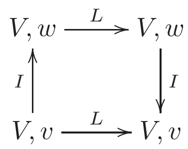
[5Q p.294] To see why, consider the equation

$$L = I \circ L \circ I. \tag{4.9}$$

It may help to visualize (4.9) as a square diagram of functions



where the right-hand side of (4.9) begins in the lower left corner of the square, goes up the left edge, across the top, then down the right edge to the lower right corner of the square. The left-hand side of (4.9) is the function along the bottom edge of the square. Choose the basis v for V in the lower left and lower right corner of the square, and choose the basis w for V in the upper two corners of the square. Here is the square again, indicating the chosen bases.



Applying (4.5) to the right-hand side of (4.9) , we have the right-hand side of (4.8). Clearly this equals the left-hand side, and (4.8) is established. To show that (4.8) is the same as (4.7), we need to show that

$$[I]_{w \text{ to } v} = \left([I]_{v \text{ to } w}\right)^{-1}. \tag{4.10}$$

The proof is left as an exercise.

This discussion establishes the fact that two square matrices are similar if and only if they represent the same linear map with respect to two bases.

Exercises

1. Verify (4.4).
2. Verify the proof of (4.5).
3. Verify the proof of (4.8).
4. Prove (4.7). Hint: apply (4.5) to $I \circ I: V \to V$ with the right choice of bases on the two copies of V.

5 Schur, Spectral, and Singular Value Decompositions

Schur Decomposition

[5R p.296]

(5.1) **Schur Decomposition.** Let A be an $n \times n$ matrix with complex entries. There is a unitary matrix U and an upper triangular matrix T such that

$$A = UTU^H.$$

The proof is by induction on n. The theorem is trivially true for $n = 1$. Now suppose the theorem is true for all $k \times k$ matrices for $k < n$, and let A be $n \times n$.

The fundamental theorem of algebra guarantees that A has an eigenvalue, say λ, because eigenvalues are roots of the characteristic polynomial $\det(A - xI)$ and the fundamental theorem of algebra says that every polynomial with complex coefficients has a root. Choose a unit length eigenvector u_1 with eigenvalue λ, and choose $u_2, \ldots, u_n$ so that $u_1, \ldots, u_n$ forms an orthonormal basis for $\mathbf{C}$. Let U_1 be the unitary matrix whose columns are $u_1, \ldots, u_n$. We have

$$U_1^H A U_1 = \left[\begin{array}{c|c} \lambda & x_2\ x_3\ \cdots\ x_n \\ \hline 0 & M \end{array}\right]$$

where $x = [x_2\ x_3\ \cdots\ x_n]$ is a row vector of length $n-1$, 0 is a column vector of length $n-1$ with all zero entries, and M is an $(n-1) \times (n-1)$ matrix.

Applying the inductive hypothesis to M, there is a unitary V and an upper triangular S such that

$$M = VSV^H.$$

Let U_2 be the unitary matrix

$$U_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V \end{array}\right].$$

Let T be the upper triangular matrix

$$T = \left[\begin{array}{c|c} \lambda & xV \\ \hline 0 & S \end{array}\right].$$

Now we have

$$\begin{aligned} U_2 T U_2^H &= \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V \end{array}\right] \left(\left[\begin{array}{c|c} \lambda & xV \\ \hline 0 & S \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V^H \end{array}\right] \right) \\ &= \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & V \end{array}\right] \left[\begin{array}{c|c} \lambda & xVV^H \\ \hline 0 & SV^H \end{array}\right] \quad \text{(use block multiplication)} \\ &= \left[\begin{array}{c|c} \lambda & x \\ \hline 0 & VSV^H = M \end{array}\right] = U_1^H A U_1. \end{aligned}$$

Setting $U = U_1 U_2$, equating first and last expressions in the chain of equalities above, and solving for A, we have

$$A = UTU^H$$

as desired.

Spectral Theorem

[5S p.297]

(5.2) **Spectral Theorem for Hermitian Matrices.** Let A be an $n \times n$ Hermitian matrix (Hermitian means $A = A^H$). Then there is a unitary matrix U and a diagonal matrix Λ such that

$$A = U\Lambda U^H.$$

If A is real symmetric, U can be taken to be a real orthogonal matrix.

The proof uses the Schur decomposition. First, we show that if A is Hermitian, then T (in the Schur decomposition) is Hermitian. Indeed, we have $T = U^H AU$, so

$$T = U^H AU = U^H A^H U = (U^H AU)^H = T^H$$

where we used the Hermicity of A in the middle of the above string of equalities.

Second, we note that a triangular T that is Hermitian must be diagonal.

This completes the proof of the Spectral Theorem for the Hermitian case. We complete the real symmetric case in two steps.

Step A. If B is Hermitian (in particular, if B is real symmetric), its eigenvalues are real.

Step B. If a real square matrix B, viewed as a complex matrix, has a real eigenvalue, then it has a corresponding eigenvector with real entries.

Steps A and B then imply that the unitary U in the statement of the Spectral Theorem can be taken to be real orthogonal when A is real symmetric.

Proof of Step A. Let λ be an eigenvalue of a Hermitian matrix B, and let x be a corresponding eigenvector. Then we have

$$\lambda|x|^2 = x^H Bx = x^H B^H x = (Bx)^H x = \overline{\lambda}|x|^2.$$

Since $x \neq 0$, we have $\lambda = \overline{\lambda}$, so λ is real.

Proof of Step B. If B has a real eigenvalue λ, then $\det(B - \lambda I) = 0$. This means the matrix $B - \lambda I$ has a nontrivial nullspace in $\mathbf{R}^n$. Any nonzero vector in the nullspace is a real eigenvector for B with eigenvalue λ.

This completes the proof of the Spectral Theorem for Hermitian (including real symmetric) matrices.

The decomposition given by the Spectral Theorem holds for the more general class of normal matrices. Here is the statement.

(5.3) **Spectral Theorem for Normal Matrices.** Let A be an $n \times n$ normal matrix with complex entries (normal means $AA^H = A^H A$). Then there is a unitary matrix U and a diagonal matrix Λ such that

$$A = U\Lambda U^H.$$

The proof is the same as the Hermitian case, with two modifications. We must show

(i) If A is normal, then T (in the Schur decomposition) is normal.

[5.6 Exer.20 p.303] (ii) If T is upper triangular and normal, then T is diagonal. .

Proof of (i). We have $T = U^H AU$, so

$$TT^H = (U^H AU)(U^H AU)^H = U^H AA^H U = U^H A^H AU = (U^H AU)^H(U^H AU) = T^H T$$

where we used the normality of A in the middle of the above string of equalities.

Proof of (ii). We do this in three steps.

Step 1. We show that if N is normal, then $|Nx| = |N^H x|$ for all x.

Step 2. It follows that for normal N, the length of the ith row is equal to the length of the ith column.

Step 3. It follows that a normal and upper triangular T must be diagonal.

Proof of Step 1. We have

$$|Nx|^2 = (Nx)^H(Nx) = x^H N^H Nx = x^H NN^H x = (N^H x)^H(N^H x) = |N^H x|^2.$$

Since $|Nx|$ and $|N^H x|$ are nonnegative numbers we conclude that $|Nx| = |N^H x|$.

Proof of Step 2. Simply note that Ne_i is the ith column of N, and $N^H e_i$ is the ith column of N^H, which is the conjugate of the ith row of N. Apply the result of Step 1.

Proof of Step 3. We show that the ith row of T has at most one nonzero entry, and that is in column i. We use induction on i. For $i = 1$, the length of the ith column is $|t_{11}|^2$ (because T is upper triangular). The length of the ith row is $|t_{11}|^2$ plus the sum of the squares of the norms of all the other entries in row 1. The result of Step 2 implies all entries in row 1 must be zero except for t_{11}. Inductively, suppose that all rows $i < k$ have zero in all off-diagonal entries, and let $i = k$. The length of the kth column is $|t_{kk}|^2$ (use the inductive hypothesis to see that the kth *column* has zero in all off-diagonal entries) and the length of the kth row is $|t_{kk}|^2$ plus the sum of the squares of the norms of the off-diagonal entries in the kth row. Step 2 implies the off-diagonal entries in the kth row must be zero, and this completes the inductive step.

This completes the proof of the Spectral Theorem for normal matrices.

Singular Value Decomposition

[Sec. 6.3]

(5.4) **Singular Value Decomposition.** Let A be a matrix with complex entries, not necessarily square. Then A can be factored

$$A = U\Sigma V^H$$

where U, V are unitary and Σ is real with nonnegative entries on the main diagonal and zero entries off the main diagonal. The nonzero entries of Σ are the square roots of the eigenvalues of AA^H and $A^H A$. The columns of U are an orthonormal basis of eigenvectors for AA^H, and the columns of V are an orthonormal basis of eigenvectors of $A^H A$. If A has real entries, then U, V can be taken to be real orthogonal matrices.

The proof begins by applying the Spectral Theorem to the Hermitian matrix $A^H A$ to obtain

$$A^H A = V\Lambda V^H$$

with V unitary (real orthogonal if A is real) and Λ the diagonal matrix of eigenvalues of $A^H A$.

Claim 1. The eigenvalues of $A^H A$ are real and nonnegative. Indeed, let λ be an eigenvalue with eigenvector x. We have

$$|Ax|^2 = (Ax)^H(Ax) = x^H(A^H A)x = \lambda|x|^2.$$

Equating the first and last expressions above, we see that λ is real and nonnegative.

Claim 2. The nonzero eigenvalues of $A^H A$ are the same as the nonzero eigenvalues of AA^H. To see this, again let λ be a nonzero eigenvalue of $A^H A$ with eigenvector x. Multiplying A on the left on both sides of $A^H Ax = \lambda x$ yields $A(A^H A)x = A\lambda x$, which we can read as

$$(AA^H)(Ax) = \lambda(Ax)$$

which means λ is an eigenvalue of AA^H since Ax is not zero. A similar argument shows that a nonzero eigenvalue for AA^H is also an eigenvalue for $A^H A$.

Let $\lambda_1, \ldots, \lambda_n$ (suppose A is $m \times n$, so $A^H A$ is $n \times n$) be the eigenvalues of $A^H A$. Reorder the eigenvalues, if necessary, so that $\lambda_1, \ldots, \lambda_r$ are the nonzero eigenvalues (this also reorders the columns of V, which are the corresponding eigenvectors), and let $\sigma_j = \sqrt{\lambda_j}$ for $1 \leq j \leq r$ (taking square roots is possible by Claim 1). Let v_j denote the jth column of V.

For each j in the range $1 \leq j \leq r$, let

$$u_j = \frac{Av_j}{\sigma_j}.$$

Claim 3. The vector u_j is a unit eigenvector for AA^H with eigenvalue λ. To see that u_j is an eigenvalue, we just check:

$$AA^H(u_j) = AA^H Av_j/\sigma_j = A(A^H Av_j)/\sigma_j = A(\lambda_j v_j)\sigma_j = \lambda_j u_j.$$

Now we check the length:

$$|u_j|^2 = |Av_j|^2/\lambda_j = v_j^H A^H Av_j/\lambda_j = |v_j|^2 = 1.$$

Choose an orthonormal basis $u_{r+1}, \ldots, u_m$ for the zero eigenspace of AA^H so that the set $u_1, \ldots, u_m$ is an orthonormal basis for $\mathbf{C}^m$ consisting of eigenvectors of AA^H, and let U be the unitary (or real orthogonal, if A is real) matrix whose columns are $u_1, \ldots, u_m$. Then we have

$$AA^H = U\Lambda' U^H$$

where Λ' is diagonal, with nonzero entries the same as Λ. Finally, let Σ be an $m \times n$ matrix with main diagonal entries $\sigma_1, \ldots, \sigma_r$ (padded with zeros, if necessary, to complete the main diagonal) and zero in all off-diagonal entries. By our construction of u_j, we have

$$Av_j = \sigma_j u_j$$

for $1 \leq j \leq r$. Reading Av_j as $AV(e_j)$ and reading $s_j u_j$ as $U\Sigma(e_j)$, we have $AV(e_j) = U\Sigma(e_j)$ for $1 \leq j \leq r$. Since $AV(e_j) = 0 = U\Sigma(e_j)$ for $r+1 \leq j \leq n$, we have equality of matrices

$$AV = U\Sigma.$$

Multiplying both sides on the right by V^H yields the Singular Value Decomposition, and the proof is complete.

Comment: The numbers $\sigma_1, \ldots, \sigma_r$ are called the ***singular values*** of A. In the construction of U, V, Σ, we have the freedom to arrange that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

This is useful in applications of the Singular Value Decomposition where the largest singular values have the most significance.

Exercises

1. Illustrate the process of Schur decomposition with two 3×3 examples where T is diagonal in one example and not diagonal in the other.
2. Find an example of a 2×2 real matrix that is normal but not symmetric.
3. Let B be a real symmetric matrix.
   (a) Show directly (not just applying Step A in the proof of the Spectral Theorem) that the eigenvalues of B are real.
   (b) Let λ be an eigenvalue for B. Let
   $$V_\lambda = \{v \colon Bv = \lambda v\}$$
   and let
   $$W = (V_\lambda)^\perp = \{w \colon w^T v = 0 \text{ for all } v \text{ in } V_\lambda\}.$$
   Show that $BW \subset W$, that is, $Bw \in W$ for all w in W.
4. Let A be an $m \times n$ matrix with complex entries.
   (a) Show directly (not just applying Claim 1 in the proof of the Singular Value Decomposition) that the eigenvalues of AA^H are real and nonnegative.
   (b) Show directly (not just applying Claim 2 in the proof of the Singular Value Decomposition) that if λ is a nonzero eigenvalue for AA^H is also an eigenvalue for $A^H A$.
5. Let A be a 6×8 matrix with real entries and only two nonzero singular values. Explain how the Singular Value Decomposition can be used to encode all the information in A using only 28 real numbers instead of the $6 \cdot 8 = 48$ original entries.

6 Spring 2009 Miscellaneous Items

6.1 Definition of Vector Space

[2.1 p.69]

Strang's Chapter 2 is titled "Vector Spaces," and section 2.1 sort of gives a definition, but the pieces are scattered. Here is the complete definition in one place.

A ***real vector space*** is a nonempty set V with two operations.

(i) ***Scalar Multiplication.*** Given a real number α and a vector x in V, the vector αx in V is defined.

(ii) ***Vector Addition.*** Given two vectors x, y in V, the vector $x + y$ in V is defined.

[Exer. 5 p.74]

These operations must satisfy the following axioms, for all x, y, z in V and α, β in $\mathbf{R}$.

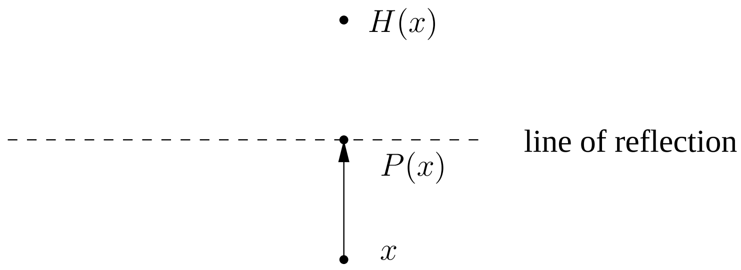
1. $x + y = y + x$ (commutativity of vector addition)
2. $x + (y + z) = (x + y) + z$ (associativity of vector addition)
3. There is a unique vector, denoted 0, and called the ***zero vector*** such that $x + 0 = x$ (identity element for vector addition).
4. For each vector x there is a unique $-x$ such that $x + (-x) = 0$ (additive inverses for vector addition).
5. $1x = x$
6. $(\alpha\beta)x = \alpha(\beta x)$
7. $\alpha(x + y) = \alpha x + \alpha y$
8. $(\alpha + \beta)x = \alpha x + \beta x$

Of course, $\mathbf{R}^n$, with its usual scaling and vector addition, satisfies this definition, so $\mathbf{R}^n$ is a vector space. Other important examples are $\mathbf{R}^\infty$ (infinite sequences of real numbers), the set of $m \times n$ matrices, and the set of real-valued functions on a chosen set X.

6.2 Reflection Across a Line

[p.126]

In Strang's Section 2.6, he discusses several linear transformations of the plane $\mathbf{R}^2$. He gives the formula $H = 2P - I$, where P is projection onto the line $y \cos\theta = x \sin\theta$ and H is the reflection across that line. To see why this is true, just look at this picture.



The arrow in the picture is the vector $P(x) - x$ (in general, given $x = (x_1, x_2)$ and $y = (y_1, y_2)$, the vector $y - x = (y_1 - x_1, y_2 - x_2)$ is the arrow with its tail at x and its head at y). You can see that $H(x) - x = 2(P(x) - x)$, or $(H - I)(x) = (2P - 2I)(x)$. Since this is true for all x, we have

$$H - I = 2P - 2I,$$

or $H = 2P - I$, as claimed.