

Linear Algebra

© 1999 by Bong Lian. All rights reserved.

Note to students: I recommend the following strategy if you are using my book for Math 22a. Prepare before each class. That means you should read (carefully!) the sections to be covered before coming to class. More importantly, try the exercises even though you may not be able to get them all. Having thought through the material by yourself makes it a lot easier for you to understand the lectures. It is also a very economical way to learn the subject. For every hour you spend preparing before class, the pay-off could easily be a saving of two to three hours after class.

Warning: If you are printing this book from a pdf file, I recommend that you first save the file to your desktop, reopen it and then send it to a printer. Failure to do so can sometimes result in unreadable text. You should also print a few test pages first, before printing the entire book.

Contents

Chapter 1. Linear Equations	4
1.1. Equations	4
1.2. Geometry of equations	5
1.3. Equations vs. solutions	11
1.4. A general procedure – row operations	12
1.5. Returning to linear systems	17
1.6. Homogeneous systems	19
1.7. A preview	22
1.8. Homework	22
Chapter 2. Vectors	26
2.1. Three basic operations	26
2.2. Lengths, angles and basic inequalities	30
2.3. Vector geometry	33
2.4. Linear combinations	36
2.5. Homework	39
Chapter 3. Matrices	45
3.1. Matrix operations and linear equations	45
3.2. Algebraic and vectorial operations and their properties	48
3.3. Inverting a matrix	52
3.4. Transpose	56
3.5. Markov process	59
3.6. Linear transformations	60
3.7. Properties preserved by linear transformations	62
3.8. Homework	64
Chapter 4. Linear Subspaces	71
4.1. Addition and scaling	71
4.2. Matrices and linear subspaces	73
4.3. Linear independence	74
4.4. Bases and dimension	76
4.5. Matrices and bases	80
4.6. The rank of a matrix	84
4.7. Orthogonal complement	88
4.8. Coordinates and change of basis	91
4.9. Sums and direct sums	93
4.10. Orthonormal bases	97
4.11. Least Square Problems	99
4.12. Solutions	100
4.13. Homework	102
Chapter 5. Determinants	108
5.1. Permutations	108
5.2. The sign function	109

5.3. Sum over permutations	112
5.4. Determinant as a multilinear alternating function	113
5.5. Computational consequences	115
5.6. Theoretical consequences	117
5.7. Minors	120
5.8. Geometry of determinants	122
5.9. Appendix	126
5.10. Homework	129
Chapter 6. Eigenvalue Problems	136
6.1. Characteristic polynomial	136
6.2. Diagonalizable matrices	140
6.3. Symmetric matrices	143
6.4. Diagonalizability of symmetric matrices	145
6.5. Homework	147
Chapter 7. Abstract Vector Spaces	152
7.1. Basic definition	152
7.2. Bases and dimension	158
7.3. Inner Products	161
7.4. Lengths, angles and basic inequalities	163
7.5. Orthogonal sets	165
7.6. Orthonormal bases	167
7.7. Orthogonal complement	169
7.8. Homework	169
Chapter 8. Linear Maps	176
8.1. Basic definitions	176
8.2. A dimension relation	180
8.3. Composition	186
8.4. Linear equations	189
8.5. The Hom space	190
8.6. Induced maps	196
8.7. Tensor product spaces	198
8.8. Homework	199
Chapter 9. Determinants For Linear Maps	204
9.1. p -forms	204
9.2. The determinant	213
9.3. Appendix	215
9.4. Homework	217
Chapter 10. Eigenvalue Problems For Linear Maps	220
10.1. Eigenvalues	220
10.2. Diagonalizable linear maps	222
10.3. Symmetric linear maps	223
10.4. Homework	225

Chapter 11. Jordan Canonical Form	227
11.1. Complex numbers	227
11.2. Linear algebra over \mathbf{C}	229
11.3. Similarity	231
11.4. Invariant subspaces and cycles	232
11.5. Appendix A	240
11.6. Appendix B	243
11.7. Homework	247
Index	251

1. Linear Equations

We first discuss equations in general. We then recognize Euclidean geometry as a way to study equations in general, and linear equations in particular.

1.1. Equations

What is an equation? We're all used to examples like

$$(*) \quad x^2 + y^2 = 1$$

and we've all been indoctrinated since childhood to believe that this somehow represents a unit circle. Not so fast. How did we go from a bunch symbols (x square plus y square equals 1) to a geometrical object (a unit circle)? The next two sections explain this.

First, we must have no disagreement about what *real* numbers are, like 1, 2, π , $-\sqrt{2}$. Let's denote the set of all real by the symbol \mathbf{R} . The question: "*what is a real number?*" is legitimate, but is a subject of another course. We must also agree to accept all the familiar rules of numbers, like addition, and subtraction, etc.

An equation in two variables x, y , such as (*), is a nothing but a *question* with an input and an output (answer). It is a black box that takes in two numbers a, b . It outputs *yes*, if the *equality* $a^2 + b^2 = 1$ holds, and outputs *no*, otherwise. When the input a, b results in *yes*, we say that a, b *satisfy* the equation, and we call a, b a *solution* to the equation. Thus 1, 0 is a solution to equation (*), while 2, 1 is not. Here is another example of an equation:

$$0 = 1.$$

This equation yields the output *no*, no matter what input it takes. Thus we say that this equation has no solution.

More generally, an equation in n variables x_1, \dots, x_n is a black box, usually specified by a symbolic expression. It takes n numbers as input, and yields either *yes*, or *no*, as output. If a list of n numbers a_1, \dots, a_n result in *yes*, we call it a solution to our equation.

What is solving an equation mean? It means to describe *all* solutions. For example,

$$x = \cos t, \quad y = \sin t$$

as t ranges over all real numbers, describes all solutions to (*). Likewise,

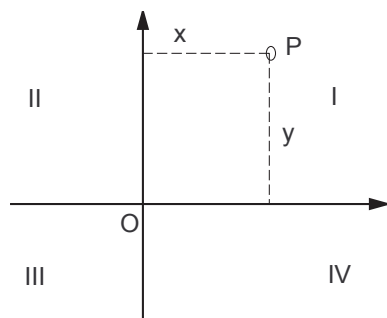
$$x = t, \quad y = -t + 1, \quad t \in \mathbf{R},$$

describes all solutions to the equation $x + y = 1$.

1.2. Geometry of equations

A powerful aid to the study of equations is geometry. Instead of treating x, y as mere symbols, we associate them to something geometrical – a plane.

In high school, we learn that the simplest way to label points in a plane is to use *Cartesian coordinates*. Let's recall what they are. First, we pick a point O in the plane, called the *origin*, which serves as our sign post. Second, we draw two axes, called *coordinate axes*, meeting perpendicularly at the origin. Each axis is given an orientation to tell us which direction of the axis is positive. The axes divide the plane into four quadrants. Third, we must agree upon a meter stick for measuring *distance*. Once the origin, the oriented axes and the unit distance are agreed upon, we have what is called a *coordinate system*. We can now use this system to label points *numerically* in the plane by this convention. Namely a point P is labelled by a pair of numbers (x, y) , where x, y are the respective distances from you have to walk, starting from O , along the direction of the *first* and the *second* axes to P . The pair (x, y) is also called the coordinates of P . The picture below depicts this. Of course, to avoid confusion, we must agree ahead of time which axis we call first axis and which second. We must also allow x or y to be negative when P lies outside the first quadrant.



Now that we have a convention for labelling points in the plane, we can use it to expand our vocabulary and to build concepts for other objects in the plane. For example, we can now describe the distance between any two points $P = (x, y)$ and $P' = (x', y')$ equationally. It is given by

$$\sqrt{(x - x')^2 + (y - y')^2}.$$

Exercise. What is the distance between P, P' when $x' = x$, i.e. when the line segment PP' is parallel to the vertical axis?

Here is another example. A circle of radius 1 (in the unit we have agreed upon) centered at the origin is the collection of points P whose distance from the origin is 1. By our convention, this means that that circle consists precisely of all solutions (a, b) to the equation

$$x^2 + y^2 = 1.$$

As we move a point in the circle, its coordinates x, y vary. Thus we also call the symbols x, y appearing in the equation above, *variables*. With this equational description, we can verify at once whether a given point $Q = (x, y)$ in the plane lies in that circle. The point $(\sqrt{2}/2, \sqrt{2}/2)$ does lie in that circle, while $(1, 1)$ does not.

This example illustrates the use of *an equation* (ie. $x^2 + y^2 = 1$) to describe *a set* (ie. the unit circle). The equation is used as *a criterion* (a point (x, y) lies on circle exactly when this criterion is fulfilled). The use of an equation as a criterion for defining a set will be a standard practice throughout this book.

Exercise. Consider the unit circle $x^2 + y^2 = 1$ relative to a given coordinate system – a choice of origin, a choice of oriented axes, and a choice of unit distance. Suppose we keep the same origin and unit distance, but rotate the axes by 90° clockwise around the origin.

What is the equation for the same circle relative to the new coordinate system? Consider the same question, but with the ellipse $\frac{x^2}{4} + y^2 = 1$.

Exercise. *A paradox?* When you look into the mirror, the image you see appears that the left and the right are interchanged, but never the top and the bottom. How would you resolve the apparent paradox that the mirror seems to have a preferred way of reflection?

Many other geometrical figures in the plane have an equational description. We know that every conic section can be described by a quadratic equation:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

where a, b, \dots, f are specified numbers. A circle is a special case of the form

$$(x - a)^2 + (y - b)^2 = r^2$$

where (a, b) is the center and r is the radius. A line is a special case of the form

$$ax + by + c = 0$$

where a, b are not both zero. This is an example of a *linear* equation. It is so-called because the variables x, y in this equation appear only to first power, rather than quadratically. This equation and its generalizations will be some of the main objects we study in this book.

It will prove convenient to allow linear equations of the form

$$ax + by + c = 0$$

without requiring that a, b be nonvanishing. Thus the equation

$$0x + 0y + 0 = 0$$

is fulfilled by any numbers x, y , and so this equation describes the entire plane. The linear equation

$$0x + 0y + 1 = 0$$

has no solution, and so it describes the empty subset of the plane.

Exercise. Where do the two lines

$$x + y = 1$$

$$x - y = 1$$

meet, if they meet?

Exercise. Where do the two lines

$$x + y = 1$$

$$2x + 2y = 1$$

meet, if they meet?

Exercise. Where do the two lines

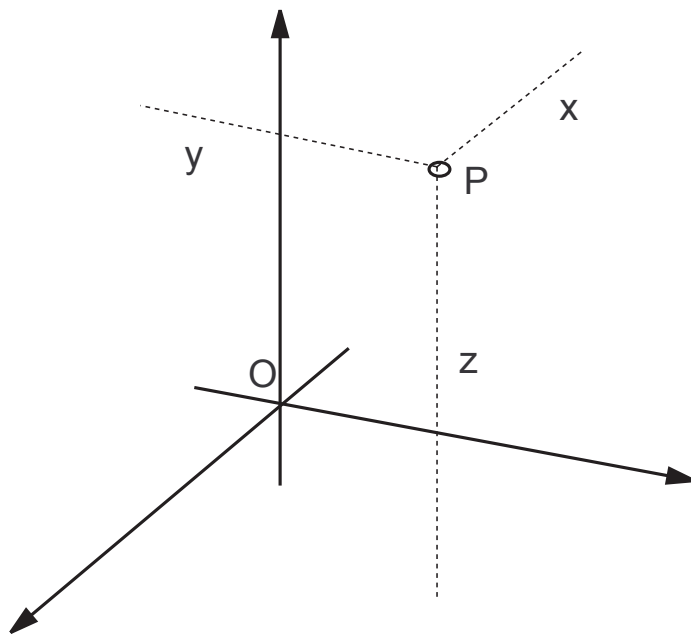
$$x + y = 1$$

$$2x + 2y = 2$$

meet, if they meet?

In the exercises above, one can answer those questions readily by inspection or by drawing pictures in the plane. The next exercise is an example of how geometry sometimes plays a crucial role in equation solving.

Exercise. Find all *rational* solutions (a, b) to the equation $x^2 + y^2 = 1$. In other words those where a, b are fractions. (Hint: Draw the unit circle and consider the line connecting $(-1, 0)$ and (a, b) . When is the slope a fraction?)



What about in a 3-space? Here we can also label points by Cartesian coordinates. As before, we pick a point O called the origin. We then choose three axes meeting perpendicularly at the origin. There is one plane passing through any two of those axes. There are three such planes, which we call coordinate planes. Together, they divide our 3-space into eight quadrants. Again, once the origin, the axes and a unit distance have been chosen, we can now label points numerically in 3-space by this convention. Namely, a point P is labelled by a triple (x, y, z) given by the distances from the three coordinate planes to P . The picture above depicts this.

We can now use these coordinates to describe objects in the 3-space, as we have done in a plane before. For example, the distance between any two points $P = (x, y, z)$ and $P' = (x', y', z')$ is

$$\sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}.$$

A plane in the 3-space can be described by an equation of the form

$$ax + by + cz + d = 0$$

where a, b, c are specified constants which are not all zero.

Example. Where do the planes

$$\begin{aligned} x + y + z &= 1 \\ (*) \quad x + y - z &= 1 \\ x - y - z &= 1 \end{aligned}$$

meet, if they meet? In this case, it may be hard to find the answer by drawing a picture. Instead, it is easier to apply the method of elimination. We want to find those (x, y, z) that satisfy all three equations. Consider the first equation. It is equivalent to $x = 1 - y - z$. We can use this to eliminate the appearance of x from the remaining equations. So our problem now becomes solving

$$\begin{aligned} x &= 1 - y - z \\ (**) \quad 1 - y - z + y - z &= 1 \\ 1 - y - z - y - z &= 1. \end{aligned}$$

Now the last two equations can be solved at once and the solution is $z = 0, y = 0$. Substitute this back into the first equation, we get $x = 1$. So the three planes meet at the point $(1, 0, 0)$.

We could have transformed the equations (*) to (**) by adding/subtracting equations. Namely, subtract the first equation from the second and the third equations, and leave the first equation alone. The result is

$$\begin{aligned}x + y + z &= 1 \\- 2z &= 0 \\- 2y - 2z &= 0\end{aligned}$$

which is just (**).

Exercise. Where do the planes

$$\begin{aligned}x + y + z &= 1 \\x + y - z &= 1 \\x + y &= 1\end{aligned}$$

meet, if they meet?

There is no reason why we should stop at 3-space.

Definition 1.1. We denote by \mathbf{R}^n the set of all n -tuples (x_1, \dots, x_n) of real numbers x_i .

The analogue of a plane in 3-space is a hyperplane in \mathbf{R}^n . A *hyperplane* is described by a single linear equation of the form

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

where a_1, \dots, a_n are given constants which are not all zero.

One of the main themes in this book is to study how hyperplanes intersect in \mathbf{R}^n . This amounts to studying the solutions to a system of such linear equations.

Definition 1.2. A linear system of m equations in n variables is a list of m equations of the form:

$$(*) \quad \begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & & \dots & & \dots & & \dots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m. \end{array}$$

where the a 's and the b 's are input data. An n -tuple $X = (x_1, \dots, x_n)$ in \mathbf{R}^n satisfying (*) is called a solution. The array of data

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right]$$

is called the augmented matrix of the linear system.

1.3. Equations vs. solutions

We have seen many samples of equations and systems of equations. *What does it mean to solve an equation?* The equation $x^2 + y^2 = 1$ in two variables describes those points which are at distance 1 from O. That is a *geometrical* description of an object. – the unit circle. Algebraically, it is also a criterion – it gives a simple test of when a point (a, b) lies in that object. It is a definition of an object, and no more than that. On the other hand, let's consider the solutions to $x^2 + y^2 = 1$. They consist of points of the form $(\cos t, \sin t)$ where t is arbitrary. Note that although this describes the very same object, the unit circle, it is a much more direct description. It gives a *parameterization* of all points simultaneously! It also tells us a way to “move” on the object: you traverse it once as the parameter t vary from 0 to 2π . This gives a useful “parametric” description of the object. This is really what solving an equation is all about – to give a useful parameterization of those points defined by the equation. Likewise solving a given system of equations means to parameterize those points which satisfies every equation in that system.

Our first task is to learn a general procedure, for solving linear systems of equations. Let's preview this in some examples.

Example. The most direct way to solve a linear system is by elimination of variables. Let's solve the system

$$x + y = 1.$$

We know that geometrically, this describes a line in \mathbf{R}^2 . But we want to write the solution algebraically. We can find x in terms of y , and get

$$x = -y + 1.$$

This yields the solution $(-y + 1, y)$ where y is *arbitrary* number. This is our general solution: it gives all possible solutions to our system.

Exercise. Solve the linear system:

$$\begin{array}{rccccrcr} x & + & y & + & z & + & t & = & 0 \\ x & + & y & - & z & & & = & 1 \\ & & y & - & z & + & t & = & 1 \end{array}$$

by elimination. After each step, write down the augmented matrix for your new system. At the end, see if you could reach the general solution $(t, -\frac{3}{2}t + \frac{1}{2}, -\frac{1}{2}t - \frac{1}{2}, t)$ where t is arbitrary.

1.4. A general procedure – row operations

We now formalize the steps performed in the preceding exercise, in order to deal with general linear systems. We will analyze an algorithm, known as Gauss elimination or row reduction. It is not only important for solving linear system, but will also be used frequently throughout this book.

The rough idea is this: *the more zeros there are in the augmented matrix the easier it is to reach the final solutions to the system.* Step by step, we would like to transform a given linear system to an equivalent (ie. without changing its solutions) linear system by manipulating the augmented matrix. In each step, our goal is to create as many zeros as possible near the lower left corner of the augmented matrix.

To do this precisely, let's first introduce some vocabulary.

Definition 1.3. *An $M \times N$ matrix A is an array of numbers arranged in M rows and N columns. The entry in the i th row and the j th column is called the (ij) entry of A . To each nonzero row of A , we assign an address between 1 and N as follows. If the leading nonzero entry of the i th row is located in the j th column, we assign the address $p_i = j$. We call that leading nonzero entry the i th pivot.*

Example. Here is a 3×3 matrix:

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

The (11) entry is 1, (12) entry is 2, (13) entry is -1, etc. The first row is $(1, 2, -1)$ and the second row is $(0, 0, 0)$. The first column is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, etc. The addresses are 1, $-$, 2.

Exercise. Let

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

List the addresses of this matrix.

A *row operation* on a matrix is one of the following operations:

R1. Interchanging two rows.

R2. Multiplying a row by a nonzero scalar.

R3. Adding a scalar multiple of one row to a different row.

Exercise. Apply R1 to A above until the first address is the minimum. For example you get:

$$B = \begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 1 \end{bmatrix}.$$

Exercise. Apply R3 to B , repeatedly if necessary, until the entries below the first pivot are all zero. You should get

$$C = \begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that for this matrix, we have $p_1 < p_i$ for $i = 2, 3, \dots$

Exercise. Now ignore the first row of C . Perform the last two steps on the remaining submatrix. When this is done, the new matrix has addresses $p_1 < p_2$. Repeat this if

necessary – ignoring the first two rows and so on. The result is

$$D = \begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The resulting matrix has the property that all zero rows are at the bottom, and that the addresses of the nonzero rows are strictly increasing.

Definition 1.4. *A matrix A is called a row echelon if all the zero rows are at the bottom, and if the addresses of the nonzero rows are strictly increasing, i.e. the pivot of a nonzero row is located further to right of the pivot in the row above it.*

Definition 1.5. *A row echelon A is called reduced if every pivot is 1, and each column containing a pivot has just one nonzero entry.*

Exercise. Apply R3 to D above until every column containing a pivot is zero everywhere below and above that pivot. You get

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Exercise. Apply R2 to E above until every pivot is 1.

Definition 1.6. *Two matrices A, A' are said to be row equivalent if A can be transformed to A' by a sequence of row operations.*

Note that if we transform A to A' by a sequence of row operations, then by reversing the operations, we can just as well transform A' to A . For example, the matrices A, B, \dots, E in the exercises above are all row equivalent.

Theorem 1.7. *Every matrix is row equivalent to a reduced row echelon.*

Proof: The proof here is a formalization of the steps illustrated in the exercises above, to deal with a general matrix.

Step 1: Given an $m \times n$ matrix, If the first k columns are all zero, then we can forget about them, and we can perform row operations on the remaining $m \times (n - k)$ submatrix without affecting those zero columns. So we may as well assume that the first column of our matrix is not all zero to begin with.

By interchanging rows (row operation R1), we can arrange that our matrix has nonzero (11) entry, ie. it reads

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

where $a_{11} \neq 0$.

Now add $-a_{21}/a_{11}$ times the 1st row to the 2nd row (row operation R2). Similarly, add $-a_{31}/a_{11}$ times the 1st row to the 3rd row, and so on, until all the entries below a_{11} in the first column are zero. Thus our original matrix is row equivalent to a matrix that reads

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{bmatrix}.$$

We then repeat the procedure with the submatrix

$$\begin{bmatrix} a'_{22} & \cdots & a'_{2n} \\ \vdots & & \vdots \\ a'_{m2} & \cdots & a'_{mn} \end{bmatrix}.$$

We continue until we reach a row echelon, say B (formally by induction).

Step 2: Suppose the addresses of the nonzero rows of $B = (b_{ij})$ are p_1, \dots, p_k . Then $1 = p_1 < p_2 < \cdots < p_k$. In particular, the p_1 th column is all zero except the entry b_{1p_1} . Now add $-b_{1p_2}/b_{2p_2}$ times the 2nd row to the 1st row (row operation R3). The new matrix has the same pivots as before. The p_2 th column is now all zero except the entry

b_{2p_2} . Similarly, Now add $-b_{1p_3}/b_{3p_3}$ times the 3rd row to the 1st row; add $-b_{2p_3}/b_{3p_3}$ times the 3rd row to the 2nd row. The new matrix has the same pivots as before.

We then repeat the procedure with the third pivot, and so on, until we reach a matrix, say C , where the p_i th column is all zero except the entry b_{ip_i} .

Step 3: Finally, scale each row of C with a nonzero scalar until each pivot is 1 (row operation R2). The resulting matrix D is a reduced row echelon. We obtain D by performing a sequence of row operations on our original matrix A , and hence A is row equivalent to D . \square

The proof above gives us an algorithm, known as *Gauss elimination* or *row reduction*, for finding a reduced row echelon which is row equivalent to a given matrix. We summarize the algorithm here:

Input: a matrix A .

- G1. Apply R1 and R3 to A , repeatedly if necessary, to yield a row echelon B .
- G2. Apply R3 to B , repeatedly if necessary, to yield a row echelon C , where each column containing a pivot is all zero except the pivot.
- G3. Apply R2 to C , repeatedly if necessary, to yield a reduced row echelon D , where each pivot is 1.

Output: D .

Exercise. Transform the following matrix to a reduced row echelon by row operations:

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 3 & 0 & -1 & 3 \end{bmatrix}.$$

Exercise. Let

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Perform a few row operations of your choice to A . Perform the same row operations to B . Perform the same row operations to the enlarged matrix $[A|B]$.

Exercise. *Row operations don't mix columns.* Suppose A, B are matrices having the same number of rows. Suppose that they transform respectively to A', B' under a single row operation. What does the enlarged matrix $[A|B]$ transform to under the same row operation?

1.5. Returning to linear systems

We now use row reduction to find the complete set of solutions to a linear system. Let S denote a linear system. Thus S is a list of m equations in n variables. Let A be the augmented matrix of S . Thus A is an $m \times (n + 1)$ matrix.

Suppose we interchange two equations in S , and call the new list S' . The linear system S' is obviously equivalent to S , since only the orders of the equations differ. This transformation $S \rightarrow S'$ of one system to the new system corresponds to transforming the augmented matrix A to another A' by a row operation R1.

Consider an equation, say

$$(*) \quad a_1x_1 + \cdots + a_nx_n = b,$$

in S . If we multiply both sides of $(*)$ by a *nonzero* scalar c , then we get an equation

$$(*') \quad ca_1x_1 + \cdots + ca_nx_n = cb.$$

This is equivalent to $(*)$ because we can recover $(*)$ by scaling both sides of $(*)'$ by $1/c$. Thus replacing $(*)$ by $(*)'$, we get a new linear system S' equivalent to S . Note that corresponding to $(*)$ is a row in A that reads

$$(a_1, \dots, a_n \mid b).$$

Replacing $(*)$ by $(*)'$ corresponding to replacing this row by

$$(ca_1, \dots, ca_n \mid cb).$$

Thus the transformation $S \rightarrow S'$ of one system to the new system corresponds to transforming the augmented matrix A to another A' by a row operation R2: scaling a row of A by c .

Consider now a pair of the equations, say

$$(*) \quad \begin{aligned} a_1x_1 + \cdots + a_nx_n &= c \\ b_1x_1 + \cdots + b_nx_n &= d, \end{aligned}$$

in S . For any scalar s , adding $sb_1x_1 + \cdots + sb_nx_n = sd$ to both sides of the first equation $(*)$, we get a new equation

$$(*') \quad (a_1 + sb_1)x_1 + \cdots + (a_n + sb_n)x_n = (c + sd).$$

We can recover $(*)$ from $(*)'$ by reversing the step. Thus replacing $(*)$ by $(*)'$, we get a new linear system S' equivalent to S . Note that corresponding to $(*)$ is a row in A that reads

$$(a_1, \dots, a_n \mid c).$$

Replacing $(*)$ by $(*)'$ corresponding to replacing this row by

$$(a_1 + sb_1, \dots, a_n + sb_n \mid c + sd).$$

The augmented matrix A' of S' is obtained from A by a row operation R3: adding s times one row of A to another row.

Thus we have seen that every row operation R1-R3 applied to the augmented matrix A corresponds to transforming S to an equivalent system S' . In other words,

Theorem 1.8. *Two linear systems having row equivalent augmented matrices have the same solutions.*

Exercise. Give two 2×3 augmented matrices which are *not* row equivalent.

A linear system is of the easiest kind when the augmented matrix is a reduced row echelon, as the following exercise illustrates.

Exercise. Let

$$E = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

Write down the linear system with augmented matrix E . Find the complete set of solutions to this system. You should get

$$(x_1, x_2, x_3, x_4) = (1, t, t, u)$$

where t, u are *free parameters*, ie. their values are arbitrary.

The following procedure give us an efficient and systematic way to solve a linear system S in n variables:

1. Write down the augmented matrix A .
2. Apply row reduction to A to get a reduced row echelon A' .
3. Write down the new linear system S' corresponding to A' .
4. Write down the complete set of solutions to S' , hence to S .

Exercise. Reconsider the linear system:

$$\begin{array}{cccccc} x & + & y & + & z & + & t & = & 0 \\ x & + & y & - & z & & & = & 1 \\ & & y & - & z & + & t & = & 1. \end{array}$$

Apply the four steps above to solve this system.

1.6. Homogeneous systems

We have seen that a linear system can always be solved by row reduction. However, the procedure doesn't tell us ahead of time whether or not our linear system actually has a solution. We find that out only at the very end of the procedure. Could we tell ahead of time if a given linear system has a solution? We will give a partial answer to this here – a more complete treatment will have to wait till later.

Definition 1.9. *An equation of the form*

$$a_1x_1 + \cdots + a_nx_n = 0$$

in n variables is called homogeneous. A linear system is called homogeneous if all its equations are homogeneous.

Obviously the values $x_1 = 0, \dots, x_n = 0$ satisfy identically every homogeneous equation in n variables. In other words,

$$(x_1, \dots, x_n) = (0, \dots, 0)$$

is always a solution to a homogeneous system. This is called the *trivial solution*.

Exercise. Solve the homogeneous linear system with the augmented matrix

$$\left[\begin{array}{cccc|c} 2 & 1 & 7 & 0 & 0 \\ 2 & 2 & 10 & 0 & 0 \\ 0 & 3 & 9 & 1 & 0 \end{array} \right].$$

What is the last column of new augmented matrix after each row operation?

Because the right hand side of each homogeneous equation is 0, the augmented matrix of a homogeneous linear system has entries 0 only in the last column. Thus if there are m homogeneous equations in n variables in the linear system, the augmented matrix reads

$$[A|O]$$

where A is an $m \times n$ matrix and O is a column with all zeros. The matrix A is called the *coefficient matrix* of the homogeneous linear system.

Recall that row operations do not mix columns. So, if A transforms to A' under a given row operation, then $[A|O]$ transforms to $[A'|O]$ under the same operation. For this reason, it is more economical to perform row operations on A only, rather than on $[A|O]$, when solving a homogeneous linear system. where A transforms to A' under the same operation.

The preceding theorem specialized to homogeneous linear systems yields

Theorem 1.10. *Two homogeneous linear systems having row equivalent coefficient matrices have the same solutions.*

Exercise. Suppose S is a *homogeneous* linear system with 2 equations in 4 variables, and suppose that its augmented matrix A is a reduced row echelon. What is the maximum number of pivots A can have? How many free parameters can the complete set of solutions have?

Theorem 1.11. *A homogeneous linear system with more variables than equations must have a nontrivial solution.*

Proof: We apply induction. Let S be a homogeneous linear system in n variables x_1, \dots, x_n .

Step 1. Suppose that S has just one equation, say

$$a_1x_1 + \cdots + a_nx_n = 0,$$

and that $n > 1$. If the a 's are all zero, then any values of x_1, \dots, x_n give a solution. If the a 's are not all zero, say $a_n \neq 0$, then the values $x_1 = 1, \dots, x_{n-1} = 1, x_n = -\frac{1}{a_n}(a_1 + \cdots + a_{n-1})$, give a solution to S .

Step 2. Now suppose that S has m equations and that $n > m$. As before, we can pick an equation, say

$$a_1x_1 + \cdots + a_nx_n = 0,$$

and solve for one of the variables, say x_n :

$$(*) \quad x_n = -\frac{1}{a_n}(a_1x_1 + \cdots + a_{n-1}x_{n-1}).$$

Then use this to eliminate x_n from the remaining $m - 1$ equations. What we get is a new linear system S' with $m - 1$ equations in $n - 1$ variables x_1, \dots, x_{n-1} . Since $n - 1 > m - 1$, this system has more variables than equations. By induction, we conclude that S' has a nontrivial solution given by some values of x_1, \dots, x_{n-1} . Now note that S is equivalent to S' together with $(*)$. So by setting x_n to the value given by $(*)$, we obtain a nontrivial solution to S . \square

Exercise. Use row reduction to find the complete set of solutions to the system:

$$\begin{array}{cccccc} x & + & y & + & z & + & t & + & u & = & 0 \\ x & + & 2y & - & z & & & & & = & 0 \\ & & - & y & - & z & + & t & + & u & = & 0. \end{array}$$

There should be two free parameters t, u in your solutions. Write down the solution corresponding to $t = 1, u = 0$. Write down the solution corresponding to $t = 0, u = 1$. We call them the “basic” solutions. Based on this exercise, can you guess a relationship between the number variables, the number of free parameters, and the number of pivots in the reduced row echelon of the augmented matrix?

Verify that $(0, 0, 0, 1, -1)$ is a solution. Can you write this solution as a “combination” of your basic solutions?

1.7. A preview

This chapter teaches us a mechanical way to solve any given linear system. *How many solutions are there? How do all the solutions hang together?* With the tools introduced in this chapter, these questions can be answered on the case-by-case basis, at best. In the next three chapters, we will learn a general approach to quantify *the shape and size of solution sets to linear systems*. We'll use the language of vectors (Chapter 2) and matrices (Chapter 3) for this quantification. This will allow us to classify and catalog all linear systems in some coherent way (Chapter 4). We then introduce the notions of determinant (Chapter 5) and eigenvalues (Chapter 6) which yield further important methods for analysing linear systems. The second half of the book (Chapters 7-10) brings everything in the first half up to an abstract level. Chapter 11 is on a more advanced topic. Here we must work with square matrices of complex numbers. By associating certain numerical data to these matrices, we have a way to decide when two matrices are “similar”.

1.8. Homework

1. Find the points in \mathbf{R}^2 at which the following line and circle meet:

$$x + y = 1, \quad x^2 + y^2 = 4.$$

2. Consider the following line and a circle of radius r in \mathbf{R}^2 :

$$2x + y = 1, \quad x^2 + y^2 = r^2.$$

Find all the values of radii r for which the two objects meet. Where do they meet?

3. Consider the following two planes and a sphere of radius r in \mathbf{R}^3 :

$$x + y + z = 0, \quad x + y - z = 2, \quad x^2 + y^2 + z^2 = r^2.$$

Find all the values of radii r for which all three objects meet. Where do they meet?

4. Consider the ellipse $\frac{x^2}{4} + y^2 = 1$ relative to a given coordinate system – a choice of origin, a choice of oriented axes, and a choice of unit distance. Suppose we shift the origin to the point $(x, y) = (1, 2)$ and then rotate the axes 90° counter-clockwise around the new origin. What is the equation for the same ellipse relative to the new coordinate system? (Hint: Draw a picture! Answer: $(X + 2)^2 + \frac{(Y-1)^2}{4} = 1$.)

5. Find the complete set of solutions to the system:

$$\begin{aligned} x + y + z + t &= 0 \\ x + y + 2z + 2t &= 0 \\ x + y + 2z - t &= 0. \end{aligned}$$

6. Find two “basic” solutions to the following system:

$$\begin{aligned} x + y + z + t + u &= 0 \\ 3x + 2y - z - 4t + 3u &= 0 \\ 2x - y - 9z + t + 9u &= 0. \end{aligned}$$

7. Write the solution $(12, -11, 1, 1, -3)$ as a “combination” of your two basic solutions from the preceding problem. To what values of your free parameters does this solution correspond?

8. Solve the linear system with the given augmented matrix:

$$(a) \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 1 & 1 & 1 & 0 \\ 5 & 7 & 9 & 0 \end{array} \right] \quad (b) \left[\begin{array}{cccc|c} 1 & 2 & 3 & 1 & 8 \\ 1 & 3 & 0 & 1 & 7 \\ 1 & 0 & 2 & 1 & 0 \end{array} \right] \quad (c) \left[\begin{array}{ccc|c} 1 & -2 & 3 & 4 \\ 2 & -1 & -3 & 5 \\ 3 & 0 & 1 & 2 \\ 3 & -3 & 0 & 7 \end{array} \right].$$

9. Find an equation for the line passing through the two points $(1, 2)$ and $(\frac{1}{2}, 0)$ in \mathbf{R}^2 .

10. Find an equation for the plane passing through the three points $(1, 0, 0)$, $(-1, 2, 0)$, and $(0, 1, -1)$.

11. Consider the function

$$f(x) = a \sin x + b \cos x$$

where a, b are constants to be determined. Suppose that $f(0) = 1$ and $f(\pi/4) = -1$. Find a, b .

12. (Calculus required) Consider the function

$$f(x) = a e^x + b e^{-x}$$

where a, b are constants to be determined. Suppose that $f(0) = 0$, $f'(0) = 1$. Find a, b .

13. Consider the polynomial function

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$$

where the a_i are constants to be determined. Suppose that $f(1), f(2), f(4)$ are all zero, and that $f(0) = -8$. Find the values of the a_i . Show that

$$f(x) = (x - 1)(x - 2)(x - 4).$$

14. Write the expression $\frac{-3}{x^2 - x - 2}$ as a partial fraction, ie. find the numbers a, b so that

$$\frac{-3}{x^2 - x - 2} = \frac{a}{x + 1} + \frac{b}{x - 2}$$

for all x .

15. (a) Verify the identity:

$$x^4 + 3x^3 + x^2 - 3x - 2 = (x + 1)^2(x - 1)(x + 2).$$

- (b) Find the numbers a, b, c, d so that the following identity holds:

$$\frac{x^3 + 6x^2 + 12x + 5}{x^4 + 3x^3 + x^2 - 3x - 2} = \frac{a}{(x + 1)^2} + \frac{b}{x + 1} + \frac{c}{x - 1} + \frac{d}{x + 2}.$$

16. Write down your favorite reduced row echelon B . Apply 10 of your favorite row operations to B to get a new matrix A . Now row reduce A to see if you get back B .

17. * Prove that row operation R1 can be replaced by a series of row operations involving R2 and R3 only. Therefore, theoretically R2 and R3 are sufficient to transform any given matrix to its reduced row echelon.

2. Vectors

In this chapter we study basic algebraic and geometric properties of vectors in \mathbf{R}^n .

2.1. Three basic operations

Let's recall a definition.

Definition 2.1. *We denote by \mathbf{R}^n the set of n -tuples (x_1, \dots, x_n) of real numbers. We call such an n -tuple a vector in \mathbf{R}^n .*

Let A be an $m \times n$ matrix. Since each row of A is an n -tuple, A can be thought of as a list of m vectors. In the last chapter, we have encountered two basic operations on these row vectors. Namely, we add two rows entrywise; and we multiply a row entrywise by a scalar. Let's give them names and study them closer here.

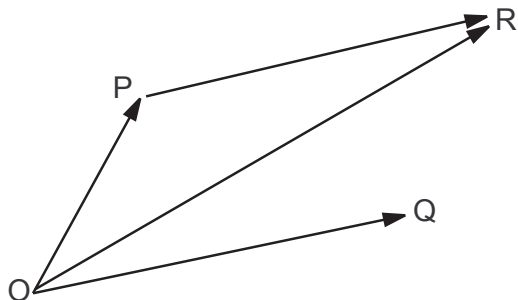
Vector addition is an operation which takes two vectors as the input and yields one vector as the output. Vector scaling is an operation which takes one vector and one number as the input and yields one vector as the output. Symbolically, these operations are defined as follows:

(i) *Vector addition:* $(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$.

(ii) *Vector scaling:* $c(x_1, \dots, x_n) = (cx_1, \dots, cx_n)$.

Geometrically vector scaling in \mathbf{R}^2 , by a number say $c > 0$, takes an arrow and yields a parallel arrow c times as long as the original one. Similarly the new arrow is

opposite in direction if $c < 0$. We illustrate vector addition in \mathbf{R}^2 in the picture below. Here, a vector P in \mathbf{R}^2 is represented by the arrow \vec{OP} running from O to the point P . The sum $R = P + Q$ of the two vectors P and Q is represented by the arrow \vec{OR} , which can be obtained by joining the two arrows \vec{OP} and \vec{OQ} , but tip to tail.



Exercise. Let $P = (1, 1)$, $Q = (2, 3)$. Find $P + Q$ by joining their arrows tip to tail. Verify that your answer is $(3, 4)$.

Exercise. Let $E_1 = (1, 0)$, $E_2 = (0, 1)$. Suppose

$$(2, 1) = aE_1 + bE_2.$$

What are the numbers a, b ?

Note that vector addition and scaling in \mathbf{R}^n are defined in terms of addition and multiplication of *numbers*. These two operations on numbers have the following familiar algebraic properties:

N1. $(x + y) + z = x + (y + z)$.

N2. $x + y = y + x$.

N3. $x + 0 = x$.

N4. $x + (-x) = 0$.

N5. $x(y + z) = xy + xz$.

N6. $xy = yx$.

N7. $(xy)z = x(yz)$.

$$\text{N8. } 1x = x.$$

These familiar properties of numbers result in similar algebraic properties for vectors in \mathbf{R}^n . To state them we introduce the following notations for vectors in \mathbf{R}^n :

$$O = (0, \dots, 0), \quad X = (x_1, \dots, x_n), \quad Y = (y_1, \dots, y_n), \quad Z = (z_1, \dots, z_n), \quad -X = (-x_1, \dots, -x_n)$$

and let a, b be numbers. Then we have

$$\text{V1. } (X + Y) + Z = X + (Y + Z).$$

$$\text{V2. } X + Y = Y + X.$$

$$\text{V3. } X + O = X.$$

$$\text{V4. } X + (-X) = O.$$

$$\text{V5. } a(X + Y) = aX + aY.$$

$$\text{V6. } (a + b)X = aX + bX.$$

$$\text{V7. } (ab)X = a(bX).$$

$$\text{V8. } 1X = X.$$

All eight algebraic properties of vector addition and scaling can be readily verified using the familiar properties of numbers N1-N8. For example V1:

$$\begin{aligned} (X + Y) + Z &= ((x_1, \dots, x_n) + (y_1, \dots, y_n)) + (z_1, \dots, z_n) \\ &= (x_1 + y_1, \dots, x_n + y_n) + (z_1, \dots, z_n) \\ &= ((x_1 + y_1) + z_1, \dots, (x_n + y_n) + z_n) \\ &= (x_1 + (y_1 + z_1), \dots, x_n + (y_n + z_n)) \\ &= (x_1, \dots, x_n) + ((y_1, \dots, y_n) + (z_1, \dots, z_n)) \\ &= X + (Y + Z). \end{aligned}$$

Property V1 allows us add three or more vectors without specifying which addition to do first. Thus we can write

$$((X + Y) + Z) + W = X + Y + Z + W.$$

Furthermore, V2 allows us to add three or more vectors without worrying about the order of the vectors. Thus we can write

$$W + Z + Y + X = X + Y + Z + W.$$

Exercise. Suppose that $X = (1, 2) - Z$ and that $Y = (1, -1) - W$. Find $X + Y + Z + W$.

Exercise. Let $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, $E_3 = (0, 0, 1)$. Suppose

$$(1, 0, -1) = aE_1 + bE_2 + cE_3.$$

what are the numbers a, b, c ?

We now define a third operation. Dot product is an operation which takes two vectors as the input and yields a number as the output. Symbolically,

$$(iii) \text{ Dot product: } X \cdot Y = x_1y_1 + \cdots + x_ny_n.$$

This operation has the following properties:

$$D1. X \cdot Y = Y \cdot X.$$

$$D2. X \cdot (Y + Z) = X \cdot Y + X \cdot Z.$$

$$D3. X \cdot (aY) = a(X \cdot Y).$$

$$D4. \text{ If } X \neq O, \text{ then } X \cdot X > 0.$$

Properties D1-D3 can be readily verified using the familiar algebraic properties of numbers above. Property D4 on the other hand is geometric, thus requires geometric properties of numbers to prove. Let's consider D4. By definition, we have

$$X \cdot X = x_1^2 + \cdots + x_n^2$$

which is a sum of nonnegative numbers. For $X \neq O$, some entry of the tuple $X = (x_1, \dots, x_n)$ is nonzero. Thus $x_i^2 > 0$ for some i , and hence $X \cdot X > 0$.

Exercise. Let $E_i = (0, \dots, 1, \dots, 0)$ be the vector in \mathbf{R}^n with every entry 0 except the i th slot which is 1. The vectors E_1, \dots, E_n are called *the standard unit vectors* in \mathbf{R}^n . What is $E_i \cdot E_j$ if $i \neq j$? if $i = j$?

Exercise. Let A_1, A_2, A_3, B be vectors in \mathbf{R}^n with $A_1 \cdot B = 1$, $A_2 \cdot B = 0$, $A_3 \cdot B = -1$. What is $(A_1 + 2A_2 + 5A_3) \cdot B$?

Exercise. Give three vectors A, B, C such that $A \cdot B = A \cdot C$ but $B \neq C$.

Example. *Linear equations and dot product.* A linear equation in n variables x_1, \dots, x_n :

$$a_1x_1 + \dots + a_nx_n = b$$

can be written as

$$A \cdot X = b$$

where $A = (a_1, \dots, a_n)$ and $X = (x_1, \dots, x_n)$. More generally, a linear system of m equations in X can now be written in the form:

$$A_1 \cdot X = b_1$$

$$A_2 \cdot X = b_2$$

$$\vdots$$

$$A_m \cdot X = b_m.$$

Exercise. Show that solutions to a given linear equation $A \cdot X = O$ are closed under vector addition and scaling. In other words if X, Y are solutions and $c \in \mathbf{R}$, then $X + Y$, cX are also solutions.

2.2. Lengths, angles and basic inequalities

As seen earlier, addition of two vectors in \mathbf{R}^2 or \mathbf{R}^3 corresponds to joining two arrows tip to tail. Many interesting laws involving distances can be discovered by drawing straight edges or arrows in \mathbf{R}^2 and \mathbf{R}^3 . For example, a triangle with one right angle has the property that the square of the hypotenuse is the sum of squares of the other two sides (Pythagoras theorem). Since arrows can be represented by vectors in \mathbf{R}^2 , these laws can be expressed in terms of vectors. These laws will be generalized to vectors in \mathbf{R}^n .

Definition 2.2. We say that $A, B \in \mathbf{R}^n$ are orthogonal if $A \cdot B = 0$. We define the length $\|A\|$ to be the real number $\sqrt{A \cdot A}$. Here we have relied on the crucial geometric

property of vectors that $A \cdot A \geq 0$. We call a vector A a unit vector if $\|A\| = 1$. We define the distance between A, B to be $\|B - A\|$.

Exercise. Use this definition to write down the distance between (x_1, x_2, x_3) and (y_1, y_2, y_3) . Do you recover the formula in chapter 1?

Exercise. Find the unit vector in the direction of $(1, 3)$.

Exercise. Use properties D1 and D3 to derive the following identity: for any scalar c and any vector A ,

$$\|cA\| = |c| \|A\|.$$

Exercise. Let A be any nonzero vector in \mathbf{R}^n . What is the length of $A/\|A\|$? Use this to find the unit vector in the direction of $(3, -1, 4)$.

Exercise. Let E_1, E_2 be the standard unit vectors in \mathbf{R}^2 . What is $\|E_1 - E_2\|$?

Exercise. Draw a triangle OAB in a plane with O at the origin. Convince yourself pictorially that

$$\|A\| + \|B\| \geq \|A - B\|.$$

Exercise. Use properties D1-D2 to derive the identity: for any vectors A, B ,

$$\|A + B\|^2 = \|A\|^2 + \|B\|^2 + 2A \cdot B.$$

Exercise. What is $A \cdot B$ if $\|A\| = \|B\| = 1$ and $\|A + B\| = \frac{3}{2}$?

Exercise. Let A, B be any vectors with $B \neq O$. Prove that if $A - cB$ is orthogonal to B , then $c = \frac{A \cdot B}{B \cdot B}$. Prove the converse: if $c = \frac{A \cdot B}{B \cdot B}$ then $A - cB$ is orthogonal to B . The number c is called the *component* of A along B , and the vector cB is called the *projection* of A along B .

Exercise. Let $A = (1, 0, -1)$. What are the components of A along E_1, E_2, E_3 ? the projections?

Theorem 2.3. (*Pythagoras theorem*) If A, B are orthogonal vectors in \mathbf{R}^n , then

$$\|A + B\|^2 = \|A\|^2 + \|B\|^2.$$

Proof: By an exercise above, we have

$$\|A + B\|^2 = \|A\|^2 + \|B\|^2 + 2A \cdot B.$$

Since A, B are orthogonal, we have $A \cdot B = 0$. \square

Theorem 2.4. (*Schwarz' inequality*) For $A, B \in \mathbf{R}^n$,

$$|A \cdot B| \leq \|A\| \|B\|.$$

Proof: If $B = O$, there is nothing to prove. Assume $B \neq 0$, and let $c = \frac{A \cdot B}{B \cdot B}$. Then $A - cB$ and cB are orthogonal. By Pythagoras,

$$\|A\|^2 = \|A - cB\|^2 + |c|^2 \|B\|^2 \geq |c|^2 \|B\|^2 = \frac{|A \cdot B|^2}{\|B\|^4} \|B\|^2.$$

Multiplying both sides by $\|B\|^2$, we get our assertion. \square

Theorem 2.5. (*Triangle inequality*) For $A, B \in \mathbf{R}^n$,

$$\|A + B\| \leq \|A\| + \|B\|.$$

Proof: By Schwarz,

$$A \cdot B \leq \|A\| \|B\|.$$

Thus we have

$$\begin{aligned} \|A + B\|^2 &= \|A\|^2 + \|B\|^2 + 2A \cdot B \\ &\leq \|A\|^2 + \|B\|^2 + 2\|A\| \|B\| = (\|A\| + \|B\|)^2. \end{aligned}$$

This implies our assertion. \square

Exercise. Show that if A, B are nonzero vectors in \mathbf{R}^n , then

$$-1 \leq \frac{A \cdot B}{\|A\| \|B\|} \leq 1.$$

Definition 2.6. If A, B are nonzero vectors in \mathbf{R}^n , we define their angle to be the number θ between 0 and π such that

$$\cos \theta = \frac{A \cdot B}{\|A\| \|B\|}.$$

Exercise. Draw a picture of a unit vector A in \mathbf{R}^2 making an angle θ with the unit vector E_1 . Verify that the preceding definition is consistent with your picture.

Exercise. What is the angle of between A, B if $A \cdot B = 0$? if $A = cB$ for some number $c > 0$? if $A = cB$ for some number $c < 0$?

Exercise. What is the cosine of the angle between A and B if $\|A\| = \|B\| = 1$, and $\|A + B\| = \frac{3}{2}$?

2.3. Vector geometry

Hyperplanes. Recall that a plane in \mathbf{R}^3 is a set of points (x, y, z) satisfying an equation of the form

$$ax + by + cz = d$$

where not all a, b, c are zero. If we put $N = (a, b, c)$, then this equation becomes

$$N \cdot X = d.$$

Let P be any point in the plane, ie. $N \cdot P = d$. Then the equation for the plane can be written as

$$N \cdot (X - P) = 0.$$

Geometrically, this equation says that if X is a point in the plane, then the edge connecting X, P is perpendicular to the vector N .

Similarly, a *hyperplane* in \mathbf{R}^n is a set of vector X satisfying an equation of the form

$$N \cdot (X - P) = 0.$$

Here N is a nonzero vector perpendicular to the hyperplane, and P is a fixed point in that hyperplane.

Distance. Find the distance between a point Q and the plane

$$N \cdot (X - P) = 0$$

in \mathbf{R}^3 . Name this plane Π . Let us first translate to the origin the edge \vec{PQ} connecting P to Q . So, P is moved to the point O and Q is moved to the point $Q - P$. Now N is a nonzero vector perpendicular to the plane. The distance between Q and the plane Π is the length $\|cN\|$, where cN is the projection of $Q - P$ along N . Recalling the value of c , we get

$$\|cN\| = |c|\|N\| = |(Q - P) \cdot N|/\|N\|.$$

Let's name the number on the right side $d(Q, \Pi)$. We have implicitly assumed that the picture that leads to this result is correct. The argument seems convincing but it's hardly obvious. It can be proven that given a plane $\Pi: N \cdot (X - P) = 0$ and a point Q in \mathbf{R}^3 , there exists one and only one point $X_0 \in \Pi$ which is nearest to Q , and whose distance to Q is precisely $d(Q, \Pi)$ as given above. In other words, there exists $X_0 \in \mathbf{R}^3$ such that

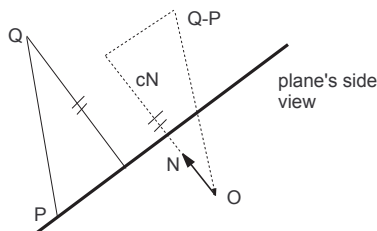
(i) $X_0 \in \Pi$;

(ii) $d(Q, \Pi) = \|Q - X_0\|$;

(iii) if $X_1 \in \Pi$, then $d(Q, \Pi) \leq \|Q - X_1\|$;

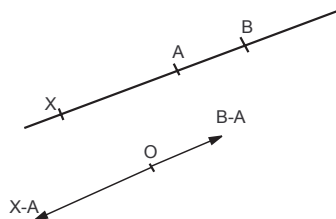
(iv) if $X_1 \in \Pi$ and if $d(Q, \Pi) = \|Q - X_1\|$, then $X_1 = X_0$.

The proof will have to wait till Chapter 4. The result also holds in \mathbf{R}^n .



Exercise. Write down a nonzero vector perpendicular to the line $y = mx + c$. Also find the distance between a point (a, b) and this line.

Lines. Given two distinct points A, B in \mathbf{R}^n , there is a unique line L in \mathbf{R}^n passing through them. The vector $B - A$ is parallel to the line.



Let X be any point on L . Then the vector $X - A$ is also parallel to L , and hence to $B - A$. Thus $X - A$ is a scalar multiple of $B - A$. Thus to each point X on L , we associate to it a scalar t such that

$$(*) \quad X - A = t(B - A), \quad \text{ie. } X = A + t(B - A).$$

Conversely, every point of the form $A + t(A - B)$ lies on L . We call $(*)$ a parametric equation for the line.

Exercise. Find a unit vector parallel to the line passing through the points $(1, 2, 3)$, $(\pi, 0, 1)$.

2.4. Linear combinations

Definition 2.7. Let $\{A_1, \dots, A_k\}$ be a set of vectors in \mathbf{R}^n , and x_1, \dots, x_k be numbers. We call

$$\sum_{i=1}^k x_i A_i = x_1 A_1 + \cdots + x_k A_k$$

a linear combination of $\{A_1, \dots, A_k\}$ with coefficients x_1, \dots, x_k .

Example. Let E_1, E_2 be the standard unit vectors in \mathbf{R}^2 . Then every vector $X = (x_1, x_2)$ in \mathbf{R}^2 is a linear combination of $\{E_1, E_2\}$ with coefficients x_1, x_2 , because

$$(x_1, x_2) = x_1(1, 0) + x_2(0, 1).$$

More generally every vector in $X = (x_1, \dots, x_n)$ in \mathbf{R}^n is a linear combination of the standard unit vectors $\{E_1, \dots, E_n\}$ with coefficients x_1, \dots, x_n .

Exercise. Write $(1, 0)$ as a linear combination of $\{(1, 1), (1, -1)\}$, ie. fill in the blanks below:

$$(1, 0) = _ (1, 1) + _ (1, -1).$$

How many ways can you do it?

Exercise. Write $(1, 0)$ as a linear combination of $\{(1, 1), (1, -1), (2, 1)\}$, ie. fill in the blanks below:

$$(1, 0) = _ (1, 1) + _ (1, -1) + _ (2, 1).$$

How many ways can you do it?

Exercise. If X and Y are linear combinations of the set $\{A_1, \dots, A_k\}$, is $X + Y$ a linear combination of that set? If c is a number, is cX a linear combination of that set?

Let $\{A_1, \dots, A_k\}$ be a set of vectors in \mathbf{R}^n . We say that this set is *orthogonal* if $A_i \cdot A_j = 0$ for all $i \neq j$. We say that this set is *orthonormal* if it is orthogonal and $A_i \cdot A_i = 1$ for all i . An example of an orthogonal set we have seen is the standard unit vectors $\{E_1, \dots, E_n\}$ in \mathbf{R}^n .

Exercise. Give a nonzero vector orthogonal to $(1, 2)$.

Exercise. Give two mutually orthogonal nonzero vectors which are themselves orthogonal to $(1, 1, 1)$.

Exercise. Let $\{A_1, \dots, A_k\}$ be an orthonormal set. Prove that

$$\left(\sum_{i=1}^k x_i A_i \right) \cdot \left(\sum_{i=1}^k y_i A_i \right) = \sum_{i=1}^k x_i y_i.$$

Theorem 2.8. (*Orthogonal sum*) Let $\{A_1, \dots, A_k\}$ be an orthonormal set, and B be a linear combination of this set. Then

$$B = \sum_{i=1}^k (B \cdot A_i) A_i$$

$$\|B\|^2 = \sum_{i=1}^k (B \cdot A_i)^2.$$

Proof: Let

$$B = x_1 A_1 + \dots + x_k A_k$$

for some numbers x_1, \dots, x_k . Then

$$B \cdot A_1 = \sum_{i=1}^k x_i A_i \cdot A_1.$$

By orthonormality, only the term with $i = 1$ survives on right hand side. So we get

$$B \cdot A_1 = x_1 A_1 \cdot A_1 = x_1.$$

Similarly dotting B with A_j , we get

$$B \cdot A_j = x_j$$

for each $j = 1, \dots, k$. The second assertion follows from the preceding exercise. \square .

Exercise. The vectors $A_1 = (1, 1)$, $A_2 = (1, -1)$ are orthogonal. Write $B = (1, 0)$ as a linear combination of A_1, A_2 .

Exercise. In the exercise just before the theorem, you have three mutually orthogonal vectors in \mathbf{R}^3 . Express $(1, 0, 0)$ as a linear combination of them.

Exercise. Let $\{A_1, \dots, A_k\}$ be an orthogonal set of nonzero vectors. Prove that if B is a linear combination of this set, then

$$B = \sum_{i=1}^k c_i A_i$$

where $c_i = \frac{B \cdot A_i}{A_i \cdot A_i}$. In other words, B is the sum of its projections $c_i A_i$ along the respective vectors A_i .

Exercise. Let B be a vector orthogonal to A_1, \dots, A_k . Show that B is orthogonal to any linear combination of $\{A_1, \dots, A_k\}$.

Theorem 2.9. (*Best approximation*) Let $\{A_1, \dots, A_k\}$ be an orthonormal set in \mathbf{R}^n , and let $B \in \mathbf{R}^n$, $x_1, \dots, x_k \in \mathbf{R}$. If $x_i \neq B \cdot A_i$ for some i , then

$$\|B - \sum_{i=1}^k (B \cdot A_i) A_i\| < \|B - \sum_{i=1}^k x_i A_i\|.$$

In other words, $\sum_i (B \cdot A_i) A_i$ is the linear combination of $\{A_1, \dots, A_k\}$ closest to B .

Proof: Consider the function

$$f(x_1, \dots, x_k) = \|B - \sum_{i=1}^k x_i A_i\|^2.$$

We will show that it attains a strict minimum exactly at the values $x_i = B \cdot A_i$. Expanding the right side, we get

$$\begin{aligned} f(x_1, \dots, x_k) &= \|B\|^2 + \sum_i x_i^2 - 2 \sum_i (B \cdot A_i) x_i \\ &= \|B\|^2 + \sum_i (x_i - B \cdot A_i)^2 - \sum_i (B \cdot A_i)^2 \\ &= f(B \cdot A_1, \dots, B \cdot A_k) + \sum_i (x_i - B \cdot A_i)^2. \end{aligned}$$

This implies that

$$f(x_1, \dots, x_k) \geq f(B \cdot A_1, \dots, B \cdot A_k)$$

and that equality holds iff $x_i = B \cdot A_i$ for all i . \square

Exercise. Consider the orthonormal set consisting of $A_1 = (1, 0, 0)$, $A_2 = \frac{1}{\sqrt{2}}(0, 1, -1)$. Find the linear combination of $\{A_1, A_2\}$ closest to the point $B = (1, 1, 0)$.

Corollary 2.10. (*Bessel's inequality*) Let $\{A_1, \dots, A_k\}$ be an orthonormal set, and B be any vector in \mathbf{R}^n . Then

$$\sum_{i=1}^k (B \cdot A_i)^2 \leq \|B\|^2.$$

Proof: In the preceding proof, we found that

$$0 \leq f(B \cdot A_1, \dots, B \cdot A_k) = \|B\|^2 - \sum_i (B \cdot A_i)^2. \quad \square$$

Exercise. Argue in one line that if both sides of Bessel's inequality are equal, then B is a linear combination of the given orthonormal set.

2.5. Homework

1. Let $A = (1, \frac{1}{2})$, $B = (-1, 2)$. Draw the points

$$A + B, \quad A - B, \quad A + 2B, \quad 2A + B, \quad A - 3B.$$

2. Which of the following pairs of vectors are perpendicular?

(a) $(1, -1, 1), (1, 1, 1)$.

(b) $(0, \pi, 1), (1, 1, -\pi)$.

(c) $(1, a, b), (0, b, -a)$.

(d) $(1, 2, 3), (\pi, 1, 0)$.

3. Which of the following pairs of planes meet? When they meet, do they meet perpendicularly?

(a) $x + y + z = 1, x - y + z = 0$.

(b) $x - 2z = 1, 2x + y + z = 3$.

(c) $x + y + z = 1$, $x + y + z = 0$.

4. For each pair of planes which meet in the previous exercise, find a unit vector parallel to the line of intersection.

5. Let L be the line which meets the plane

$$3x - 4y + z = 2$$

perpendicularly, and which passes through the point $P = (1, 2, -1)$. Find the point of intersection of L with the plane.

6. Find the distance between the following given point and plane:

(a) $(1, 0, 1)$ and $x - 2y + z = 3$.

(b) $(3, 2, 1)$ and $x + y + z = 1$.

(c) $(-1, 2, -3)$ and each of the coordinate planes.

7. Find all the y such that $\|(1, y, -3, 2)\| = 5$.

8. Let P be a given vector in \mathbf{R}^n . We define an operation, called *translation* by P as follows. Given a vector X in \mathbf{R}^n as the input, it yields $X + P$ as the output. Let $A = (-1, 2)$, $B = (1, 1)$, $C = (2, 5)$. Translate the triangle ABC by $-B$. What are the vertices of the new triangle?

9. Translate the triangle ABC above so that A gets translated to the origin O . Find the cosine of the angle at A . Find the cosine of all three angles. (You may assume that lengths and angles remain the same after translation.)

10. Find the cosine of all three angles of the triangle with vertices $(0, 1, 1)$, $(-1, 1, 1)$, $(2, 0, 1)$.

11. If A, B, C are three points in \mathbf{R}^n , find the general formula for the angle at B for the triangle ABC using translation.
12. Let $y = mx + c$ and $y = m'x + c'$ be two the equations for two lines in \mathbf{R}^2 . Show that the lines are perpendicular if and only if $mm' = -1$.
13. (a) Verify that $\{(1, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}), (0, \frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}})\}$ is an orthogonal set.
- (b) Find the linear combination of this set which best approximates the vector $(1, 0, 1)$. Do the same for $(0, 1, 1)$ and $(1, 1, 0)$.
- (c) Find a unit vector which is orthogonal to the two vectors in (a).
14. Let A, B be vectors in \mathbf{R}^n . If $\|A\|^2 = \|B\|^2 = 1$, and $\|A + B\|^2 = 3/2$, find the cosine of the angle between A, B .
15. Prove that if A, B are orthogonal iff

$$\|A - B\| = \|A + B\|.$$

That is, if $\|A - B\| = \|A + B\|$ then A, B are orthogonal; if A, B are orthogonal then $\|A - B\| = \|A + B\|$.

16. (a) Find the angle between the planes given by

$$x + y + z = 0, \quad 2x - y + z = 1.$$

- (b) Find the angle between the plane and the parametric line given by

$$x - y - z = 1, \quad X(t) = (-1, -1, 1) + t(1, 1, 1).$$

Where do they intersect?

- (c) Find the shortest distance between the point $(1, 2, 3)$ and the plane

$$x + y + z = 0.$$

17. What is the component of the vector $X = (x_1, \dots, x_n)$ along E_1 ? along E_2 ? along E_i ?
18. Write E_1, E_2 as linear combinations of $\{(1, 2), (2, 1)\}$.
19. Which of the vectors $(1, 1, -1), (1, -1, 1), (-1, 1, 1)$ are linear combinations of $\{(1, 0, 0), (0, 1, -1)\}$?
20. * Prove that if A is orthogonal to every vector in \mathbf{R}^n , then $A = O$.
21. Suppose A, B are nonzero vectors. Prove that $A = cB$ for some number c iff $|A \cdot B| = \|A\|\|B\|$.
22. Let A, B be any vectors. Prove that
- (a) $\|A + B\|^2 + \|A - B\|^2 = 2\|A\|^2 + 2\|B\|^2$.
- (b) $\|A - B\|^2 = \|A\|^2 + \|B\|^2 - 2\|A\|\|B\|\cos \theta$ where θ is the angle between A and B .
23. * Suppose A, B are nonzero vectors. Prove that $A = cB$ for some number $c > 0$ iff $\|A + B\| = \|A\| + \|B\|$.
24. * Let c be the component of A along B . Prove that
- $$\|A - cB\| \leq \|A - xB\|$$
- for any number x . That is, c is the number that minimizes $\|A - xB\|$.
25. * Let X_0 be a point in \mathbf{R}^n and r be a positive number. The *ball* of radius r centered at X_0 is the set of points X whose distance from X_0 is less than r , ie.

$$B(X_0; r) = \{X \in \mathbf{R}^n \mid \|X - X_0\| < r\}.$$

Suppose $Y \in B(X_0; r)$. Prove that $B(Y; \delta) \subset B(X_0; r)$ if and only if $\delta \leq r - \|Y - X_0\|$. (Hint: Consider the case $X_0 = 0$ first; draw a picture. In one direction, show that $B(Y; \delta) \subset B(0; r) \implies \|Y + \frac{\lambda}{\|Y\|}Y\| < r$ for all $\lambda < \delta$. In the reverse direction, use triangle inequality.)

26. Let $A, B \in \mathbf{R}^n$. The set, denoted by \overline{AB} ,

$$\{tB + (1-t)A \mid 0 \leq t \leq 1\}$$

is called the *line segment* between A, B . Draw a picture of the line segment between $(1, -2), (-1, 0)$ in \mathbf{R}^2 .

27. * A subset $S \subset \mathbf{R}^n$ is said to be *convex* if for every $A, B \in S$, the line segment between A, B is contained in S . Prove that every ball $B(X_0; r)$ in \mathbf{R}^n is convex. (Hint: $X_0 = tX_0 + (1-t)X_0$; triangle inequality.)

28. Let A, B be any vectors.

(a) Prove that

$$\|A\| - \|B\| \leq \|A - B\|.$$

(b) Prove that

$$\| \|A\| - \|B\| \| \leq \|A - B\|.$$

29. * Let A, X_0, X be vectors in \mathbf{R}^n . Let ϵ be any positive number. Prove that

$$|A \cdot X - A \cdot X_0| < \epsilon$$

whenever $\|X - X_0\| < \frac{1}{1+\|A\|}\epsilon$. (Hint: Use Schwarz' inequality.)

30. * Let $s \in \mathbf{R}$, and $A, B \in \mathbf{R}^n$. Prove that

$$(1 + \|A + B\|)^s \leq (1 + \|A\|)^s (1 + \|B\|)^{|s|}.$$

This is called the Peetre inequality. (Hint: For $s > 0$ use triangle inequality. For $s < 0$, set $A' = A + B$, $B' = -B$.)

31. A, B, C, D are points in \mathbf{R}^3 arranged so that $\|A - B\| = \|B - C\| = \|C - D\| = \|D - A\|$ and $\|A - C\| = \|B - D\|$. Describe, say by pictures, all possible such arrangements.

3. Matrices

In chapter 1, a matrix has appeared as the data that specifies a linear system. We perform row reduction on this matrix in order to find the explicit solutions to the linear system. In this chapter, we study many other ways to yield information about a matrix, hence about their associated linear system. We begin by first extending the algebraic operations on vectors in chapter 2 to algebraic operations on matrices.

3.1. Matrix operations and linear equations

Recall that an $m \times n$ *matrix* is a rectangular array of numbers

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

with m rows and n columns. We sometimes abbreviate this by writing $A = (a_{ij})$. The entry a_{ij} is located in the i th row and j th column, and it is called the (ij) entry of A .

Let A be an $m \times n$ matrix. Each row of A is a vector in \mathbf{R}^n . Thus A is a list of m row vectors in \mathbf{R}^n . If we denote the row vectors by ${}_1A, {}_2A, \dots, {}_mA$, then we write

$$A = \begin{bmatrix} {}_1A \\ {}_2A \\ \vdots \\ {}_mA \end{bmatrix}.$$

Similarly each column of A is an m tuple of numbers, hence a vector in \mathbf{R}^m . To emphasize that such an m tuple is presented in a column, we call it a column vector. Thus A is a list

of n column vectors in \mathbf{R}^m . If we denote the column vectors by A_1, \dots, A_n , then we write

$$A = [A_1, A_2, \dots, A_n].$$

A matrix is best thought of, not as a list of data, but as an operation with one input and one output. Let $A = (a_{ij})$ be an $m \times n$ matrix. The input is a *column* vector

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in \mathbf{R}^n , and the output is a vector AX in \mathbf{R}^m . We define AX to be the vector in \mathbf{R}^m with i th entry

$$\sum_{j=1}^n a_{ij}x_j = a_{i1}x_1 + \dots + a_{in}x_n.$$

There are two ways to rewrite the vector AX . First, the i th entry of AX is the dot product of the i th row ${}_iA$ of A with the vector X , so that

$$AX = \begin{bmatrix} {}_1A \cdot X \\ \vdots \\ {}_mA \cdot X \end{bmatrix}.$$

Second, AX is also the linear combination of the columns A_i of A with coefficients x_i :

$$AX = x_1A_1 + \dots + x_nA_n.$$

Dot product as matrix product. We can regard a *row* vector $R \in \mathbf{R}^n$ as a $1 \times n$ matrix. Likewise a *column* vector $C \in \mathbf{R}^n$ is an $n \times 1$ matrix. In this convention in mind, we have the identity

$$RC = [R \cdot C]$$

where the right side is the 1×1 matrix with entry given by the dot product of $R, C \in \mathbf{R}^n$. There is no confusion if we drop the bracket on the right side and just write

$$RC = R \cdot C.$$

In other words, we can drop the dot whenever we multiply a *single row* R with a *single column* C in this order. (Never do so for two rows or two columns or in the wrong order!)

Exercise. Let

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & -2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad E_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Compute AE_1 , AE_2 , AE_3 , AX . Verify that

$$A(E_1 + 2E_2 + 3E_3) = AE_1 + 2AE_2 + 3AE_3.$$

Exercise. Let A be an $m \times n$ matrix. Give the column vector X such that AX is the first column of A . Do the same for the i th column of A . What if we want AX to be 16 times the i th column of A ?

Exercise. Prove that an $n \times n$ matrix A is completely determined by the values AE_i , $1 \leq i \leq n$.

Exercise. Let I be the $n \times n$ matrix with entries 1 along the diagonal and 0 off diagonal. Thus the columns of I are the standard unit vectors E_1, \dots, E_n . The matrix I is called the *identity matrix*. If X is a vector, what is IX ?

Exercise. Let A be an $m \times n$ matrix. Suppose that $AX = O$ for *any* vector X . Show that the entries of A are all zero. This matrix is called the $m \times n$ *zero matrix*, and we denote it by \mathbf{O} .

The most important property of the matrix operation on vectors is that it is *linear*:

Theorem 3.1. *Let A be an $m \times n$ matrices, X, Y are vectors in \mathbf{R}^n , and c a scalar. Then*

$$(a) \quad A(X + Y) = AX + AY$$

$$(b) \quad A(cX) = c(AX).$$

Proof: Write

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Then

$$\begin{aligned} A(X + Y) &= (x_1 + y_1)A_1 + \cdots + (x_n + y_n)A_n \\ &= x_1A_1 + y_1A_1 + \cdots + x_nA_n + y_nA_n \\ &= x_1A_1 + \cdots + x_nA_n + y_1A_1 + \cdots + y_nA_n \\ &= AX + AY. \end{aligned}$$

Part (b) is similar, and is left as an exercise. \square

Recall that a linear system of m equations in n variables takes the form:

$$\begin{array}{cccccc}
 a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & v_1 \\
 a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & v_2 \\
 (*) & & \dots & & \dots & & \dots & & \vdots \\
 a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & v_m
 \end{array}$$

where the a 's and the v 's are input data. Put $A = (a_{ij})$, which is an $m \times n$ matrix. Put

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}.$$

Observe that if we put the right hand sides of (*) into a column vector in \mathbf{R}^m , then this column vector is AX . Thus the linear system (*) now reads

$$AX = V.$$

In particular, a homogeneous linear system now has the matrix form

$$AX = O.$$

3.2. Algebraic and vectorial operations and their properties

There are several other operations and their properties involving matrices which we will encounter later. There are two basic types of operations, which we call *algebraic* and *vectorial*. We introduce them, and then study their formal properties here. The operation of a matrix on \mathbf{R}^n introduced above falls in the first type.

Algebraic operations. Let $A = (a_{ij})$ be an $m \times n$ matrix, and $B = (b_{ij})$ be an $n \times l$ matrix. We define the *matrix product* AB to be the $m \times l$ matrix $C = (c_{ij})$ with its (ij) entry

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

There are two ways to rewrite AB . Let ${}_1A, \dots, {}_mA$ be the rows of A as before, and let B_1, \dots, B_l be the columns of B . Then note that $c_{ij} = {}_iA \cdot B_j$, so that

$$AB = \begin{bmatrix} {}_1A \cdot B_1 & {}_1A \cdot B_2 & \cdots & {}_1A \cdot B_l \\ {}_2A \cdot B_1 & {}_2A \cdot B_2 & \cdots & {}_2A \cdot B_l \\ \cdots & \cdots & \cdots & \cdots \\ {}_mA \cdot B_1 & {}_mA \cdot B_2 & \cdots & {}_mA \cdot B_l \end{bmatrix} = \begin{bmatrix} {}_1AB_1 & {}_1AB_2 & \cdots & {}_1AB_l \\ {}_2AB_1 & {}_2AB_2 & \cdots & {}_2AB_l \\ \cdots & \cdots & \cdots & \cdots \\ {}_mAB_1 & {}_mAB_2 & \cdots & {}_mAB_l \end{bmatrix}.$$

Note that the i th column of AB is also the vector AB_i , so that

$$AB = [AB_1, \dots, AB_l].$$

Warning. The matrix product of two matrices A and B is not defined unless their sizes are *compatible*! That is, the number of columns in A and the number of rows in B must be the same in order to define AB . Thus we do not multiply a 3×4 matrix with a 3×5 matrix.

Exercise. Let I be the $n \times n$ identity matrix. What is AI where A is any given $m \times n$ matrix? What is IB where B is any given $n \times l$ matrix?

Exercise. Let \mathbf{O} be the $n \times n$ zero matrix. What is $A\mathbf{O}$ where A is any given $m \times n$ matrix? What is $\mathbf{O}B$ where B is any given $n \times l$ matrix?

Vectorial operations. Let $A = (a_{ij})$ be an $m \times n$ matrix, and c be a scalar. Then we define cA to be the $m \times n$ matrix whose (ij) entry is ca_{ij} . Note that if A_1, \dots, A_n are the columns of A , then $cA = [cA_1, \dots, cA_n]$. Likewise for rows.

Let $A = (a_{ij})$, $B = (b_{ij})$ be $m \times n$ matrices. Then we define $A + B$ to be the $m \times n$ matrices whose (ij) entry is $a_{ij} + b_{ij}$. Note that if A_1, \dots, A_n are the columns of A and B_1, \dots, B_n are the columns of B , then $A_1 + B_1, \dots, A_n + B_n$ are the columns of $A + B$. Likewise for the rows. We also define $-A$ to be the matrix whose (ij) entry is the number $-a_{ij}$.

Exercise. Let

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}.$$

Find $A + B$, $2B$, $-3B$, $A + 2B$, $B - A$.

Formal properties. Let X be a column vector in \mathbf{R}^n . Then

$$\begin{aligned} (A + B)X &= x_1(A_1 + B_1) + \dots + x_n(A_n + B_n) \\ &= x_1A_1 + x_1B_1 + \dots + x_nA_n + x_nB_n \\ &= x_1A_1 + \dots + x_nA_n + x_1B_1 + \dots + x_nB_n \\ &= AX + BX. \end{aligned}$$

More generally, let $C = [C_1, \dots, C_l]$ be an $n \times l$ matrix. Then

$$\begin{aligned}(A + B)C &= [(A + B)C_1, \dots, (A + B)C_l] \\ &= [AC_1 + BC_1, \dots, AC_l + BC_l] \\ &= [AC_1, \dots, AC_l] + [BC_1, \dots, BC_l] \\ &= AC + BC.\end{aligned}$$

This yields the matrix identity:

$$(A + B)C = AC + BC.$$

Likewise, you can verify the identity

$$A(B + C) = AB + AC.$$

Exercise. Let

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Find both $(AB)C$ and $A(BC)$.

Exercise. Pick three favorite 2×2 matrices A, B, C of yours. Compute $(A + 2B)C$ and $AC + 2BC$, and compare them.

Exercise. Show that $(A + B) + C = A + (B + C)$, where A, B, C are matrices of the same size.

Exercise. Show that $A(B + C) = AB + AC$, where A, B, C are matrices of compatible sizes.

Below is a summary of basic definitions of vectorial and algebraic operations, as introduced above. Here a is any scalar, and A, B are matrices of appropriate sizes.

- $aA = [aA_1, \dots, aA_n]$
- $-A = [-A_1, \dots, -A_n]$
- $A + B = [A_1 + B_1, \dots, A_n + B_n]$

- $AB = [AB_1, \dots, AB_n]$

- $AB = \begin{bmatrix} {}_1AB \\ \vdots \\ {}_nAB \end{bmatrix}$

Below is a summary of formal properties of those operations. Each equality of matrices holds whenever both sides of the equality are defined. Here a, b are scalars and A, B, C are matrices of appropriate sizes.

- $A + (B + C) = (A + B) + C$

- $A + B = B + A$

- $A + \mathbf{O} = A$

- $A + (-A) = \mathbf{O}$

- $(a + b)A = aA + bA$

- $(ab)A = a(bA)$

- $a(A + B) = aA + aB$

- $1A = A$

- $a(AB) = (aA)B = A(aB)$

- $A(B + C) = AB + AC$

- $(A + B)C = AC + BC$

- $(AB)C = A(BC)$

Exercise. Verify that the last identity by writing down the (ij) -entry on both sides.

Exercise. Suppose A, B, C, D are $n \times n$ matrices such that $BC = I$ and $AD = I$. Find $(AB)(CD)$.

Exercise. *No cancellation law.* Find a 3 matrices A, B, C such that $B \neq C$, $A \neq \mathbf{O}$, but $AB = AC$.

Exercise. *No reciprocal law.* Give a nonzero square matrix A such that $A^2 = \mathbf{O}$.

Exercise. *No commutativity law.* Give two square matrices A, B such that $AB \neq BA$.

Exercise. *No associative law for dot product.* Give 3 vectors A, B, C such that $(A \cdot B)C \neq A(B \cdot C)$.

3.3. Inverting a matrix

Definition 3.2. A square matrix A is said to be invertible if there is a matrix B such that $AB = BA = I$. In this case, B is called an inverse of A .

Exercise. Verify that

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

are inverses of each other.

Exercise. Verify that if $ad - bc \neq 0$, then

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

are inverses of each other.

Exercise. Give a 2×2 reduced row echelon with no zero rows. How many are there? How about the case of 3×3 ?

Theorem 3.3. *Suppose A is an $n \times n$ reduced row echelon with no zero rows. Then $A = I$.*

Proof: Since A has no zero rows, its addresses p_1, \dots, p_n are all between 1 and n . Since A is a row echelon, the addresses are strictly increasing:

$$1 \leq p_1 < p_2 < \cdots < p_n \leq n.$$

This implies that $p_i = i$ for all i , ie. the pivots in A are exactly the diagonal entries. Since A is reduced, the pivots are all 1 and the i th column must be the standard vector E_i . Thus $A = I$. \square

Theorem 3.4. *Suppose A is an invertible $n \times n$ matrix. Then the following holds:*

(i) *A has a unique inverse. We denote it by A^{-1} .*

(ii) *Given any $V \in \mathbf{R}^n$, $AX = V$ has a unique solution.*

Proof: (i) Suppose B, C are both inverses of A . Then

$$AB = I = AC.$$

Multiplying both sides by B , we get

$$B(AB) = B(AC), \quad \text{or } (BA)B = (BA)C.$$

Since $BA = I$, we get $B = C$.

(ii) The linear system $AX = V$ is equivalent to the linear system

$$A^{-1}(AX) = A^{-1}V.$$

(Why?) But the second system is

$$X = A^{-1}V$$

which clearly has the vector $A^{-1}V$ as the one and only solution. Thus the system $AX = V$ must also have the same. \square

Theorem 3.5. *A square matrix A is row equivalent to I iff the linear system $AX = O$ has a unique solution.*

Proof: By Theorem 1.10 if A is row equivalent to I , then $AX = O$ and $IX = O$ have the same solution, which is obviously $X = O$.

Conversely suppose that $AX = O$ has a unique solution $X = O$. Let B be a reduced row echelon of A . Since A, B are row equivalent, the systems

$$AX = O, \quad BX = O$$

have the same solutions hence $X = O$ is the only solution to $BX = O$. So B cannot have a zero row. For otherwise $BX = O$ would be a linear system with less than n equations (a zero row will contribute no equation); and such a homogeneous linear system would have a nontrivial solution by Theorem 1.11. Thus B is a reduced row echelon square matrix with no zero rows. It follows that $B = I$. \square

Theorem 3.6. *A square matrix A is invertible iff it is row equivalent to I .*

Proof: If A is invertible then A is row equivalent to I , by the preceding two theorems. We prove the converse in two steps. For convenience, we write $P \sim Q$ if P and Q are row equivalent matrices.

Step 1. Suppose $A \sim I$. Then $[A|E_1] \sim [I|B_1]$ for some vector B_1 . Thus the two linear systems

$$AX = E_1, \quad IX = B_1$$

have augmented matrices which are row equivalent. So they have the same solution, which is $X = B_1$. Hence $AB_1 = E_1$. Similarly for each i , we have a vector B_i such that $AB_i = E_i$. This shows that

$$AB = I$$

where $B = [B_1, \dots, B_n]$.

Step 2. The equation $BX = O$ has a unique solution since multiplying both sides by A yields $ABX = IX = X = O$. By the preceding theorem, we have $B \sim I$. Applying the argument in Step 1 to the matrix B , we find that $BC = I$ for some C . Multiplying

both sides by A , we get $C = A$, hence $BA = I$. Together with $AB = I$, this means that A is invertible with $A^{-1} = B$ and $B^{-1} = A$. \square

Corollary 3.7. *If $AB = I$, then A, B are invertible and are the inverses of each other.*

Proof: This follows from Step 2 of the preceding proof. \square

Corollary 3.8. *If $[A|I] \sim [I|B]$, then B is the inverse of A .*

Proof: By assumption there is a sequence of row operations transforming $[A|I]$ to $[I|B]$. In particular the same operations transform $[A|E_i]$ to $[I|B_i]$, which means that $X = B_i$ is a solution to $AX = E_i$, i.e. $AB_i = E_i$. Thus $AB = I$. By the preceding corollary $B = A^{-1}$.

\square

Exercise. Find the inverse B of

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

Be sure to double check that $AB = I$.

Exercise. Find the inverse of

$$A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Use your answer to solve the linear systems $AX = E_i$, $i = 1, 2, 3$, where the E_i are the standard unit vectors in \mathbf{R}^3 . Solve the equation

$$AX = B$$

where B is any given vector in \mathbf{R}^3 .

Exercise. Let A, B be invertible $n \times n$ matrices. Show that $B^{-1}A^{-1}$ is the inverse of AB .

Exercise. Let A be an $n \times n$ matrix. If there is a vector $V \in \mathbf{R}^n$ such that the system $AX = V$ has a unique solution, prove that A is invertible. (cf. Theorem 3.4(ii)).

3.4. Transpose

Definition 3.9. Given an $m \times n$ matrix $A = (a_{ij})$, we define its transpose A^t to be the $n \times m$ matrix whose (ij) entry is a_{ji} . Equivalently, the i th row of A becomes the i th column of A^t .

Example. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Then

$$A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Note that $(A^t)^t = A$. More generally for any $m \times n$ matrix A , we have

$$(A^t)^t = A.$$

Theorem 3.10. For two matrices A, B of compatible sizes, we have $(AB)^t = B^t A^t$.

Proof: The (ij) entry of AB is

$$(\text{row } i \text{ of } A) \cdot (\text{column } j \text{ of } B).$$

Thus the (ij) entry of $(AB)^t$ is

$$(\text{row } j \text{ of } A) \cdot (\text{column } i \text{ of } B), \quad \text{or} \quad (\text{column } i \text{ of } B) \cdot (\text{row } j \text{ of } A).$$

But this is the same as

$$(\text{row } i \text{ of } B^t) \cdot (\text{column } j \text{ of } A^t).$$

And this is the (ij) entry of $B^t A^t$. \square

Exercise. If A, B are matrices of the same size, Show that

$$(A + B)^t = A^t + B^t.$$

If c is a scalar, show that

$$(cA)^t = c A^t.$$

Exercise. Suppose A is an invertible matrix. Prove that A^t is invertible, and that

$$(A^t)^{-1} = (A^{-1})^t.$$

Exercise. Suppose A is a 2×2 matrix such that $A^4 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. What is $(A^t)^4$?

Definition 3.11. A square matrix A is said to be orthogonal if $AA^t = I$.

Exercise. Verify that for any number θ , the matrix

$$A_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

is orthogonal.

Exercise. Let A be any matrix. Show that the (ij) entry of AA^t is

$$(\text{row } i \text{ of } A) \cdot (\text{row } j \text{ of } A).$$

Show that the (ij) entry of A^tA is

$$(\text{column } i \text{ of } A) \cdot (\text{column } j \text{ of } A).$$

Exercise. Let A be an orthogonal matrix. Show that its columns are orthonormal vectors. Show that the rows are also orthonormal vectors.

Exercise. Conversely, if the columns of a square matrix A are orthonormal vectors, show that A is an orthogonal matrix.

Exercise. How many orthogonal matrices are there, each containing the two column vectors $A_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix}^t$, $A_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \end{bmatrix}^t$?

Theorem 3.12. Suppose A is orthogonal. Then A, A^t are invertible and are the inverses of each other. Moreover, A^t is also orthogonal.

Proof: Since $AA^t = I$, it follows that A, A^t are invertible and the inverses of each other by Corollary 3.7. In particular, we also have

$$A^tA = A^t(A^t)^t = I.$$

Thus A^t is orthogonal. \square

Theorem 3.13. *Suppose A, B are both orthogonal matrices of the same size. Then AB is orthogonal.*

Proof: This follows from

$$(AB)^t(AB) = B^t A^t AB = B^t B = I. \quad \square$$

Definition 3.14. *A square matrix A is said to be symmetric if $A^t = A$. It is said to be skew-symmetric if $A^t = -A$.*

Example. Let A be a square matrix. Then

$$(AA^t)^t = (A^t)^t A^t = AA^t.$$

Thus the matrix AA^t is symmetric. We also have

$$(A + A^t)^t = A^t + A.$$

Thus the matrix $A + A^t$ is also symmetric.

Exercise. Show that the matrix $A - A^t$ is skew symmetric.

Exercise. Show that

$$A = \frac{1}{2}(A + A^t) + \frac{1}{2}(A - A^t).$$

Thus conclude that a square matrix can always be written as the sum of a symmetric and a skew-symmetric matrix.

Theorem 3.15. *Let A be a square matrix. If A is symmetric then $X \cdot AY = Y \cdot AX$ for all vectors X, Y . Conversely, if $X \cdot AY = Y \cdot AX$ for all vectors X, Y , then A is symmetric.*

Proof: Let $A = (a_{ij})$ be an $n \times n$ matrix. If A is symmetric, then for any X, Y , we have

$$X \cdot AY = X^t AY = (X^t AY)^t = Y^t A^t X = Y \cdot AX.$$

Conversely if this equation holds for any vectors X, Y , then it holds for the standard unit vectors $X = E_i, Y = E_j$ in \mathbf{R}^n . So

$$a_{ij} = E_i \cdot AE_j = E_j \cdot AE_i = a_{ji}.$$

This completes the proof. \square

Exercise. If A is symmetric and B is orthogonal matrix of the same size, then $B^{-1}AB$ is symmetric.

3.5. Markov process

Mutations. In an experiment, three strands of bacteria I, II, III, mutates among themselves in any given week as follows.

$\frac{1}{5}I$ mutates into II .

$\frac{1}{6}I$ mutates into III .

$\frac{1}{4}II$ mutates into I .

$\frac{1}{5}II$ mutates into III .

$\frac{1}{7}III$ mutates into I .

$\frac{1}{8}III$ mutates into II .

Problem. Describe the population of each strand after n weeks.

Let x_n, y_n, z_n be the populations of I,II,III in the n th week. Then in the $(n+1)$ st week, $\frac{1}{5}I$ becomes II , and $\frac{1}{6}I$ becomes III . Thus the total fraction of I mutating into other strands is

$$\frac{1}{5} + \frac{1}{6} = \frac{11}{30}.$$

Hence total fraction *remaining* as I is $1 - \frac{11}{30} = \frac{19}{30}$. But $\frac{1}{4}$ of II becomes I, and $\frac{1}{7}$ of III becomes I. Thus the population of I for the $(n+1)$ st week is

$$x_{n+1} = \frac{19}{30}x_n + \frac{1}{4}y_n + \frac{1}{7}z_n.$$

Similarly,

$$y_{n+1} = \frac{11}{20}y_n + \frac{1}{5}x_n + \frac{1}{8}z_n$$

$$z_{n+1} = \frac{41}{56}z_n + \frac{1}{6}x_n + \frac{1}{5}y_n.$$

In matrix form:

$$X_{n+1} = \begin{bmatrix} \frac{19}{30} & \frac{1}{4} & \frac{1}{7} \\ \frac{1}{5} & \frac{11}{20} & \frac{1}{8} \\ \frac{1}{6} & \frac{1}{5} & \frac{41}{56} \end{bmatrix} X_n = AX_n$$

Note that the sum of entries of each column is 1.

$$X_2 = AX_1, \quad X_3 = A^2X_1, \dots, \quad X_{n+1} = A^nX_1.$$

Thus once the initial population X_1 is known, then X_{n+1} can be computed explicitly.

3.6. Linear transformations

Exercise. Let $A_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ for any number θ . This matrix is called a rotation matrix. Draw a picture for E_1 and $A_\theta E_1$. Draw a picture for E_2 and $A_\theta E_2$. Prove that $\|X\| = \|A_\theta X\|$ for any vector X . What is the angle between X and $A_\theta X$? Verify that

$$A_\theta A_{-\theta} = A_{-\theta} A_\theta = I.$$

Throughout this section, E_1, \dots, E_n will denote the standard unit vectors in \mathbf{R}^n .

Definition 3.16. A linear transformation (or linear map) $L : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a rule which assigns a vector $L(X)$ in \mathbf{R}^m to every vector X in \mathbf{R}^n , such that L preserves vector addition and scaling. That is, for any vectors X, Y in \mathbf{R}^n and any scalar c ,

$$L(X + Y) = L(X) + L(Y), \quad L(cX) = cL(X).$$

We call the point $L(X)$ in \mathbf{R}^m the image of X .

Example. The identity transformation $I : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is the linear transformation which sends each vector to the vector itself.

Example. The zero transformation \mathbf{O} is the linear transformation which sends each vector to the zero vector.

Exercise. Suppose $L : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is a linear transformation such that

$$L(E_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad L(E_2) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Find $L \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Example. Let A be an $m \times n$ matrix. Define the rule

$$L_A : \mathbf{R}^n \rightarrow \mathbf{R}^m, \quad L_A(X) = AX.$$

By Theorem 3.1, L_A is a linear transformation.

Example. What does a linear transformation $L : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ look like? It sends the column vector E_1 to some $A_1 = L(E_1)$, and E_2 to some $A_2 = L(E_2)$. It sends $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ to $x_1A_1 + x_2A_2$. Consider the 2×2 matrix:

$$A = [A_1, A_2].$$

From the preceding example, we have a linear transformation defined by $L_A(X) = AX$.

But

$$AX = x_1A_1 + x_2A_2.$$

Thus the linear transformation L above is nothing but L_A .

Exercise. Let $L : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be a linear transformation such that $L(E_1) = (1, 1)$, $L(E_2) = (1, 2)$. Sketch the image of the line $x - y = 0$ under L . Do the same for the parametric line $X(t) = A + tB$.

Theorem 3.17. *Every linear transformation $L : \mathbf{R}^n \rightarrow \mathbf{R}^m$ can be represented by an $m \times n$ matrix A with columns $L(E_1), \dots, L(E_n)$, ie. $L = L_A$.*

Proof: Given L , let

$$A = [L(E_1), \dots, L(E_n)].$$

A vector X in \mathbf{R}^n is a linear combination of the standard unit vectors E_i :

$$X = x_1E_1 + \dots + x_nE_n.$$

By linearity of L ,

$$L(X) = x_1L(E_1) + \dots + x_nL(E_n).$$

But the right hand side is AX . So

$$L(X) = AX$$

for any X . \square

3.7. Properties preserved by linear transformations

Throughout this section, we will consider a linear transformation $\mathbf{R}^n \rightarrow \mathbf{R}^n$ represented by a given $n \times n$ matrix A .

Question. When does A preserve dot product, ie. $AX \cdot AY = X \cdot Y$ for any X, Y ?

Question. When does A preserve length?

Question. When does A preserve angle?

Question. When does A preserve volume?

We shall answer the first three questions, and must defer the last one till chapter 5.

Theorem 3.18. *A preserves dot product iff it preserves length.*

Proof: Let X, Y be arbitrary vectors in \mathbf{R}^n .

Suppose A preserves dot product. Then, in particular,

$$AX \cdot AX = X \cdot X.$$

This means that $\|AX\|^2 = \|X\|^2$, or equivalently $\|AX\| = \|X\|$. Thus A preserves length.

Conversely, suppose A preserves length. Then

$$\|A(X + Y)\|^2 = \|X + Y\|^2.$$

Expanding this, we get

$$\|AX\|^2 + \|AY\|^2 + 2AX \cdot AY = \|X\|^2 + \|Y\|^2 + 2X \cdot Y.$$

Since $\|AX\| = \|X\|$ and $\|AY\| = \|Y\|$, it follows that

$$AX \cdot AY = X \cdot Y.$$

Thus A preserves dot product. \square

Theorem 3.19. *A preserves dot product iff it is an orthogonal matrix.*

Proof: Let X, Y be arbitrary vectors in \mathbf{R}^n . Recall that $X \cdot Y = X^t Y$.

Suppose A is orthogonal. Then

$$AX \cdot AY = (AX)^t AY = X^t A^t AY = X \cdot A^t AY = X \cdot Y.$$

Thus A preserves dot product.

Conversely, suppose A preserves dot product. Then

$$AX \cdot AY = X \cdot Y.$$

As before this gives

$$X \cdot A^t AY = X \cdot Y.$$

Put $X = E_i, Y = E_j$. Then we get

$$E_i \cdot A^t A E_j = E_i \cdot E_j.$$

The left hand side is the (ij) entry of $A^t A$, and the right hand side is the (ij) entry of the identity matrix I . Thus

$$A^t A = I,$$

ie. A is orthogonal. \square

Exercise. Show that if A preserves dot product, then it preserves angle.

Theorem 3.20. *A preserves angle iff it is a nonzero multiple of an orthogonal matrix.*

Proof: Let X, Y be arbitrary nonzero vectors in \mathbf{R}^n . For clarity, let's just consider the case of $n = 2$.

Suppose $A = rB$ where B is an orthogonal matrix and r a scalar. So, B preserves dot product (preceding theorem), and thus, angle:

$$\frac{BX \cdot BY}{\|BX\| \|BY\|} = \frac{X \cdot Y}{\|X\| \|Y\|}.$$

The left hand side is equal to $\frac{AX \cdot AY}{\|AX\| \|AY\|}$, the angle between AX and AY . The right hand side is the angle between X and Y . This shows that A preserves angle.

Conversely, suppose A preserves angle. We will show that the columns of $A = [A_1, A_2]$ form an orthogonal set and have the same length l . Since $l \neq 0$ (why?), it will follow that $\frac{1}{l}A$ is an orthogonal matrix. Since A preserves angle, we have

$$\frac{AX \cdot AY}{\|AX\| \|AY\|} = \frac{X \cdot Y}{\|X\| \|Y\|}.$$

In particular, for $X = E_1, Y = E_2$, we get

$$A_1 \cdot A_2 = 0$$

ie. the columns of A are orthogonal. For $X = E_1, Y = E_1 + E_2$, we get

$$\frac{A_1 \cdot A_1}{\|A_1\| \|A_1 + A_2\|} = \frac{1}{\sqrt{2}}.$$

Squaring both sides and apply Pythagoras $\|A_1 + A_2\|^2 = \|A_1\|^2 + \|A_2\|^2$, we get

$$2\|A_1\|^2 = \|A_1\|^2 + \|A_2\|^2.$$

It follows that A_1 and A_2 have the same length. \square

Example. Consider $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. It is not orthogonal because the columns of length square 2. So A does not preserve dot product or length. But the columns form an orthogonal set and have the same length 2. So, $\frac{1}{\sqrt{2}}A$ is an orthogonal matrix. Thus A is a nonzero multiple of an orthogonal matrix.

3.8. Homework

1. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 5 & 7 \end{bmatrix}$.

(a) Find a vector $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ so that $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$.

(b) Find a vector $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ so that $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

2. (a) If $A = \begin{bmatrix} 2 & -1 \\ -4 & -2 \end{bmatrix}$, compute A^2 and A^3 .

(b) If $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix}$, compute $2A^2 - 7A$.

3. Let

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

Find A^2 , A^3 , A^4 .

4. Let

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

Find A^2 , A^4 , A^{11} .

5. Let

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

Find A^2 , A^3 .

6. Let

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

Find A^2 , A^3 , A^4 . Propose a general rule for an $n \times n$ matrix having entries below and along the diagonal all zero.

7. (a) Find a real matrix A so that

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

(b) Find a real matrix B so that

$$B^4 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

8. What is the reduced row echelon of

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 8 & 0 & 0 & 1 \end{array} \right] ?$$

What is the inverse of

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 8 \end{bmatrix} ?$$

9. Find the inverses of the following matrices, if possible.

$$(a) \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \quad (b) \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix} \quad (c) \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -2 & 0 & 0 \\ 0 & 3 & 2 & 1 \\ 1 & 2 & 1 & -2 \end{bmatrix}.$$

10. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$. Find A^{-1} , $(A^t)^{-1}$ and $(A^{-1})^t$.

11. Invert the following matrices, if possible:

$$(a) \begin{bmatrix} 1 & 11 & 7 \\ 0 & 4 & 3 \\ -1 & 2 & 0 \end{bmatrix} \quad (b) \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -2 & 0 & 0 \\ 1 & 2 & 1 & -2 \\ 0 & 3 & 2 & 1 \end{bmatrix}.$$

12. Let A, B be any symmetric matrices of the same size. Show that $AB + BA$ is symmetric.

13. If A and B are symmetric $n \times n$ matrices, must $A + B$ be symmetric? Must AB be symmetric? Explain.

14. Let A be the matrix $\begin{bmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{bmatrix}$. Is A orthogonal? Is A^5 orthogonal?

15. Let $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the map defined by

$$f(x, y) = (2x, 3y).$$

Describe the image of the circle

$$x^2 + y^2 = 1.$$

16. Let $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the map defined by

$$f(x, y) = (xy, y).$$

Describe the image of the line

$$x = 2.$$

17. Which of the following maps are linear? In (c)-(g), sketch also the image of the line $x = y$.

(a) $f : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ defined by $f(x, y, z) = (y, z)$.

(b) $f : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ defined by $f(X) = X + (1, 0, 0)$.

(c) $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined by $f(x, y) = (x - 2y, y)$.

(d) $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined by $f(x, y) = (x, x)$.

(e) $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ defined by $f(X) = X \cdot X$.

(f) $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ defined by $f(X) = (2, -1) \cdot X$.

(g) $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ defined by $f(x, y) = xy$.

18. Let $L : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be a linear map such that

$$L(1, -1) = (3, 4), \quad L(1, 2) = (0, 2).$$

Find $L(E_1)$, $L(E_2)$, $L(2, 1)$. Also find the matrix that represents L .

19. Find the matrix which transforms the vectors

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix},$$

respectively to

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

20. Give a 3×3 matrix A which preserves angle but not length.

21. Suppose A is a 4×4 matrix such that $A^t A = 2I$. Does A preserve length, or angle? Give an example of such a matrix which is not a multiple of I .

22. * Let A be a square matrix. Prove the following.

- (a) If $A^2 = \mathbf{O}$, then $I - A$ is invertible.
- (b) If $A^3 = \mathbf{O}$, then $I - A$ is invertible.
- (c) If $A^k = \mathbf{O}$ for some k , then $I - A$ is invertible.
- (d) If $A^2 + 2A + I = \mathbf{O}$, then A is invertible.
- (e) If $A^3 - A + I = \mathbf{O}$, then A is invertible.

23. * A square matrix A is said to be *similar* to B if $B = CAC^{-1}$ for some invertible matrix C . Let A be similar to B . Prove that following.

- (a) B is similar to A .
- (b) If A is invertible, then so is B .
- (c) A^t is similar to B^t .

(d) For positive integer k , A^k is similar to B^k .

(e) If $A^k = I$ for some positive integer k , then $B^k = I$.

24. * Prove that a square matrix can be written *uniquely* as a sum of a symmetric and a skew-symmetric matrix.

25. * (calculus required) Let A be an $m \times n$ matrix. Consider the function $f(X) = \|AX\|$ defined for $X \in \mathbf{R}^n$ with $\|X\| = 1$. This is a continuous function defined on the unit sphere $\|X\| = 1$. From calculus, we know that f has a maximum, i.e. there is a unit vector X_0 such that $f(X_0) \geq f(X)$ for all unit vector X . We *define*

$$\|A\| = f(X_0) = \|AX_0\|$$

and call this the norm of A .

(a) Prove that for any $X \in \mathbf{R}^n$,

$$\|AX\| \leq \|A\| \|X\|.$$

(b) Prove that if A is an orthogonal matrix, then $\|A\| = 1$. (Hint: Write $AX \cdot AX$ as a matrix product.)

(c) Prove that if A is $1 \times n$, then $\|A\|$ defined here coincides with the length of A , regarded as a vector in \mathbf{R}^n .

26. *

(a) Prove that for any matrices A, B of compatible sizes, we have

$$\|AB\| \leq \|A\| \|B\|.$$

(b) Prove that for any matrices A, B of the same size, we have

$$\|A + B\| \leq \|A\| + \|B\|.$$

27. *

- (a) Fix $i < j$. Find an $n \times n$ matrix R such that for any $n \times m$ matrix A , RA can be obtained from A by interchanging its rows i and j .
- (b) Fix i and a number $c \neq 0$. Find an $n \times n$ matrix R such that for any $n \times m$ matrix A , RA can be obtained from A by scaling its row i by c .
- (c) Fix $i < j$ and a number c . Find an $n \times n$ matrix R such that for any $n \times m$ matrix A , RA can be obtained from A by adding c times its row j to its row i .
- (d) Use (a)-(c) to give another proof that row operations do not mix columns.
28. * Let A, A', B be matrices such that $AB, A'B$ make sense. If A transforms to A' by a single row operation, prove that AB transforms to $A'B$ by the same row operation. Conclude that if A, A' are row equivalent, then $AB, A'B$ are also row equivalent.