MATH 212 – Linear Algebra I

Richard Taylor — Thompson Rivers University

last revision: November 22, 2012

1 Systems of Linear Equations

Lec. #1

In your other courses you might have already met with a problem similar to the following.

Example 1.1. Find numbers x, y that satisfy both of the following equations:

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

solution: $x = 2$, $y = 1$

There are a variety of methods for solving such a problem:

Method 1: Isolate one variable in one of the equations, and substitute into the other.

Method 2: Subtract one equation from the other.

We will want to deal with more than two equations and two unknowns. Method 1 becomes too cumbersome; instead we will build on Method 2.

Interpretation: the solution of the system of equations can be interpreted as the intersection of two lines in the xy-plane.

Example 1.2. A company owns mines A, B and C. Each mine produces gold, silver and copper, in the following amounts (in ounces of metal per ton of raw ore):

mine	gold	silver	copper
A	1	10	20
B	2	18	22
C	1	12	40

How many tons of ore should be extracted from each mine, to exactly fill an order for a total of 8 oz. gold, 78 oz. silver and 144 oz. copper?

Let x, y, z be the number of tons of raw ore from mines A, B, C, respectively. Then the total amount of gold produced will be $x + 2y + z$. So we require that $x + 2y + z = 8$ if the order for gold is to be filled exactly. We can write a similar equation for the total quantities of silver and copper produced. In summary:

$$\begin{aligned} x + 2y + z &= 8 \\ 10x + 18y + 12z &= 78 \\ 20x + 22y + 40z &= 144 \end{aligned}$$

We can solve by generalizing Method 2 above.

For any given system of linear equations, the fundamental questions are:

- does a solution exist?
- if so, how many solutions are there?
- how to we find the solution(s)?

In the first section of the course we will develop some computational tools and some theory that we can use to answer these questions.

2 Gauss-Jordan Elimination

Lec. #2

2.1 Augmented Matrix

Throughout the algebraic manipulations required to solve the previous example, many of the symbols were redundant—they act only as placeholders. All the essential information can be represented in the augmented matrix:

$$\left[\begin{array}{rrr|r} 1 & 2 & 1 & 8 \\ 10 & 18 & 12 & 78 \\ 20 & 22 & 40 & 144 \end{array}\right]$$

The variable names and equals sign are implicit. The essential operations required to solve the system can be performed by operating on rows of this matrix. This cuts down on the book-keeping.

2.2 Elementary row operations

Permissible operations on the equations become the following permissible *elementary row operations* on the augmented matrix:

1. multiply any row by a constant
2. add a multiple of any row to another
3. interchange any two rows

These operations allow us to modify the *form* of the equations *without changing their solution.*

2.3 Row reduction / Gaussian elimination

By a systematic application of these operations (the *row reduction* algorithm) we arrive at the following:

$$\left[\begin{array}{rrr|r} 1 & 2 & 1 & 8 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

This gives $z = 1$ immediately; x and y can be found by *back substitution.*

solution: $x = 3$, $y = 2$, $z = 1$

2.4 A systematic approach: the Gauss-Jordan elimination

The method of *Gauss-Jordan elimination* is a systematic approach to solving a system of linear equations as we did in the previous example.

1. Go to the first non-zero column; interchange rows so the top entry (the *pivot*) is non-zero.
2. Apply row operations to obtain all zeroes below the pivot.
3. Consider the sub-matrix below and to the right of the pivot; repeat steps 1–2 on this submatrix. Continue in this way until there are no more pivots available.
4. Scale each row so the leading entry is 1.
5. Go to the right-most pivot; use row operations to obtain all zeroes above the pivot.
6. Move one pivot to the left, and repeat. Continue in this way for all pivots.

After step 3 the augmented matrix is in *row echelon form.* It has the following properties:

- On each row, the leading entry is to the right of the leading entry in the row above, and has only zeroes below it.
- Any all-zero rows are at the bottom.

At this stage (if the solution is unique) we can read off one solution, and proceed with back-substitution. Or we can continue with steps 3-6, after which the matrix is in *reduced row echelon form* (RREF). It then has the following additional properties:

- Every leading entry is a 1, and is the only non-zero entry in its column.

From the reduced row echelon form, we can just read off the solution, if it is unique.

Note: there's more than one way to get to the RREF, so you might think there's more than one possible final answer. It turns out that for any matrix, the RREF is unique: everyone will get the same final answer, no matter what steps they take to get there (we'll prove this later).

Example 2.1. Use row operations to put the following matrix in reduced row echelon form:

$$\left[\begin{array}{rrr|r} 0 & 5 & 3 & -1 \\ 1 & 2 & 1 & 0 \\ -3 & -1 & 2 & 1 \end{array}\right]$$

solution: REF: $\left[\begin{array}{rrr|r} 1 & 2 & 1 & 0 \\ 0 & 5 & 3 & -1 \\ 0 & 0 & 2 & 2 \end{array}\right]$ RREF: $\left[\begin{array}{rrr|r} 1 & 0 & 0 & 3/5 \\ 0 & 1 & 0 & -4/5 \\ 0 & 0 & 1 & 1 \end{array}\right]$

Example 2.2. Use Gauss-Jordan elimination to solve the following linear systems:

$$\begin{cases} x_1 + x_2 + x_3 = 3 \\ 2x_1 + 3x_2 + x_3 = 5 \\ x_1 - x_2 - 2x_3 = -5 \end{cases} \qquad \begin{cases} x_1 - 2x_2 + 3x_3 = 9 \\ -x_1 + 3x_2 = -4 \\ 2x_1 - 5x_2 + 5x_3 = 17 \end{cases}$$

3 Consistency and Uniqueness

Lec. #3

3.1 Consistency

Consider the linear system we started with:

$$x - 2y = 0$$
$$x - y = 1$$

We can interpret the solution as the point of intersection of two lines:

[graph of example above]

Clearly there is exactly one solution. But consider:

$$x - 2y = 0$$
$$2x - 4y = 8.$$

These lines are parallel:

[graph of example above]

There is no point of intersection, and the system has no solution. We say this system is *inconsistent*. When we apply Gauss-Jordan to this system, we get:

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

The last row asserts $0 = 8$ (which is obviously false). We say the system is *inconsistent*. In other words, the original system is equivalent to the false statement "$0 = 8$". We can generalize this observation as follows:

- If a row echelon form of the augmented matrix has a row of the form $[0 \cdots 0 \;\; k]$ where $k \neq 0$, then the system is inconsistent.
- Conversely, there is no such row then the system is *consistent*: there exists at least one solution.

Example 3.1. For the linear system with the following augmented matrix, how many solutions are there? (interpret the result graphically.)

$$\left[\begin{array}{cc|c} 0 & 1 & 2 \\ 1 & -1 & 1 \\ 1 & 1 & 3 \end{array}\right]$$

3.2 General Solution

Consider the linear system

$$\begin{aligned} -x_1 + 3x_2 &= 1 \\ x_2 - x_3 &= 0 \\ x_1 - x_3 &= -1 \end{aligned}$$

whose RREF is

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

Notice that the final row expresses the equation "$0 = 0$", which is redundant. In essence, the Gauss-Jordan has eliminated one of the equations. Also, this is the simplest system of equations that are equivalent to the original; they can't be reduced further.

It will be useful to distinguish between the *basic variables* (x_1 and x_2) which correspond to pivot columns, and the other, *free variable* (x_3). This system has an infinite number of solutions, which can be conveniently expressed by solving for the basic variables in terms of the free variables, yielding the *general solution*

$$\begin{cases} x_1 = 3x_3 + 1 \\ x_2 = x_3 \\ x_3 \text{ is free.} \end{cases}$$

which describes every possible solution of the original system. This means x_3 can take on *any* value, while x_1 and x_2 are determined by x_3 according to the formulas above. So, for example, one solution is:

$$x_3 = 0, x_1 = 1, x_2 = 0$$

and another is

$$x_3 = 2, x_1 = 7, x_2 = 2.$$

We get a different solution for each choice of x_3; thus this system has an infinite number of solutions.

Note that the variable x_3 acts a parameter in general solution: it can take any value, but this value then determines all of the other variables. It is sometimes convenient to introduce a dummy variable to act as this parameter. Then, for example, if we let $x_3 = t$ we can write the general solution in *parametric form* as

$$\begin{cases} x_1 = 3t + 1 \\ x_2 = t \\ x_3 = t \end{cases}$$

where the parameter t can take any value.

Example 3.2. Find the general solution of the linear system

$$\begin{aligned} x_1 - x_2 - x_3 + 2x_4 &= 1 \\ 2x_1 - 2x_2 - x_3 + 3x_4 &= 3 \\ -x_1 + x_2 - x_3 &= -3, \end{aligned}$$

Lec. #4

3.3 General results

Note that for any system of linear equations, there are only three possibilities for the number of solutions: there must be either:

- exactly one solution (when the number of pivot columns equals the number of variables)
- no solutions (when the final row of a REF looks like $[0 \ \cdots \ 0 \ k]$)
- an infinite number of solutions (when the final row does not look like $[0 \ \cdots \ 0 \ k]$ and there are fewer pivot columns than the number of variables, so there is at least one free variable)

So, e.g. it is impossible for a linear system to have exactly two solutions.

Example 3.3. Consider the linear system

$$\begin{aligned} x + ky &= 2 \\ -3x + 4y &= 1. \end{aligned}$$

For what value(s) of k is there (a) a unique solution (b) no solution (c) an infinite number of solutions?

Example 3.4. Consider the linear system

$$\begin{aligned} 2x - y &= 3 \\ -4x + 2y &= k \\ 4x - 2y &= 6. \end{aligned}$$

For what value(s) of k is there (a) a unique solution (b) no solution (c) an infinite number of solutions?

3.4 Homogeneous Systems

A linear system with all all zeroes on the right-hand side is said to be *homogeneous*. For example, the system

$$\begin{aligned} 3x + y + z &= 0 \\ x - y + z &= 0 \\ x + 5y - 2z &= 0 \end{aligned}$$

is homogeneous. Homogeneous systems have some special properties that are worth noting. For example, a homogenous system always has at least one solution, the "zero solution" $x = y = z = \cdots = 0$.

Theorem 3.1. *For a homogenous linear system, if the number of equations is less than the number of variables then the system has infinitely many solutions.*

Proof. Since the augmented matrix has fewer rows than variables, the RREF will have fewer pivots (there is at most one pivot per row) than the number of variables. Thus there must be at least one free variable. The system cannot be inconsistent (the zero solution will always be a solution) so, since there is a free variable, there must be infinitely many solutions. □

4 Applications of Linear Systems

4.1 Balancing Chemical Reactions

Lec. #5

Example 4.1. To balance the chemical reaction for the combustion of propane, use variables to represent the unknown coefficients:

$$x_1\ C_3H_8 + x_2\ O_2 \longrightarrow x_3\ CO_2 + x_4\ H_2O$$

Write equations expressing the balance of numbers of each atom on both sides of the reaction:

$$\begin{aligned} 3x_1 &= x_3 \\ 8x_1 &= 2x_4 \\ 2x_2 &= 2x_3 + x_4 \end{aligned}$$

Bring all terms to the left-hand sides, write the corresponding augmented matrix, and RREF:

$$\left[\begin{array}{cccc|c} 3 & 0 & -1 & 0 & 0 \\ 8 & 0 & 0 & -2 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & -1/4 & 0 \\ 0 & 1 & 0 & -5/4 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{array}\right]$$

So x_4 is a free variable, and the general solution can be written in parametric form as

$$\begin{aligned} x_1 &= (1/4)t \\ x_2 &= (5/4)t \\ x_3 &= (3/4)t \\ x_4 &= t. \end{aligned}$$

It is convenient to choose $t = 4$, then $x_1 = 1$, $x_2 = 5$, $x_3 = 3$, $x_4 = 4$ are all integers. Any other values of x_4 could be used.

Example 4.2. Balance the following chemical reactions

$$B_2S_3 + H_2O \longrightarrow H_3BO_3 + H_2S$$

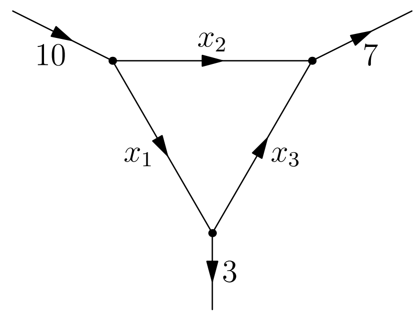
$$NaHCO_3 + H_3C_6H_5O_7 \longrightarrow Na_3C_6H_5O_7 + H_2O + CO_2$$

4.2 Network Flows

Lec. #6

e.g., traffic flow, electric circuits, computer networks. . .

Example 4.3. Solve the following network flow (i.e. find the flow in each branch of the network).



Using the principle that "flow in = flow out" at each node, we get the following equations:

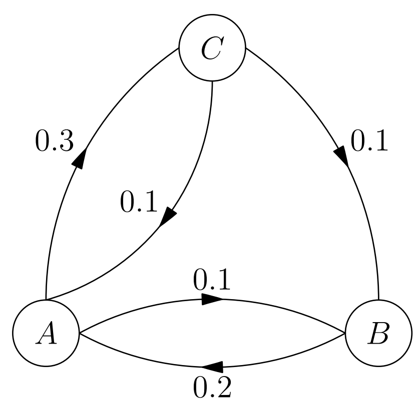
$$\begin{cases} x_1 + x_2 = 10 \\ x_2 + x_3 = 7 \\ x_1 = 3 + x_3 \end{cases} \implies \begin{bmatrix} 1 & 1 & 0 & 10 \\ 0 & 1 & 1 & 7 \\ 1 & 0 & -1 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & 1 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{cases} x_1 = 3 + t \\ x_2 = 7 - t \\ x_3 = t \end{cases}$$

If the directions of flow are required to be as shown, then the range of possible solutions becomes restricted as follows:

$$\begin{cases} 3 + t \geq 0 \\ 7 - t \geq 0 \\ t \geq 0 \end{cases} \implies 0 \leq t \leq 7 \implies \begin{cases} 3 \leq x_1 \leq 10 \\ 0 \leq x_2 \leq 7 \\ 0 \leq x_3 \leq 7 \end{cases}$$

4.3 Markov Processes

Three companies (A, B and C) compete for market share (e.g. customers). Each weak, each company loses a fixed proportion of its market share to its competitors, as shown below.



After many weeks, the distribution of market share reaches an equilibrium. What share of the market does each of the companies have?

At equilibrium, the weekly net change of each company's market share is zero. This gives the following system of equations.

$$\begin{aligned}
-0.4x_1 + 0.2x_2 + 0.1x_3 &= 0\\
0.1x_1 - 0.2x_2 + 0.1x_3 &= 0\\
0.3x_1 - 0.2x_3 &= 0
\end{aligned}$$

5 Matrix Algebra

Lec. #7

We have already seen that matrices are useful to keep track of the "book-keeping" involved in solving linear systems. Matrices have many other uses as well; to this end we need to develop our theory of matrices and the mathematical operations that can be performed with them.

5.1 Matrix addition & subtraction

It is convenient to use symbols to represent matrices, just as symbols are used to represent numbers in ordinary algebra. For example, let

$$A = \begin{bmatrix} 5 & 3 \\ 1 & -2 \\ 4 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad C = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \text{ and } \qquad c = 3.$$

Throughout this course we will follow the convention of using uppercase letters to represent matrices. Lowercase letters will be used to represent *scalars* (i.e. numbers).

Definition 5.1. Matrix addition $(A + B)$ and **subtraction** $(A - B)$ are defined by element-wise addition/subtraction of the individual components of the matrices.

So, for example,

$$A + B = \begin{bmatrix} 5 & 3 \\ 1 & -2 \\ 4 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 5 \\ 4 & 2 \\ 9 & 6 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 5 & 3 \\ 1 & -2 \\ 4 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ -2 & -6 \\ -1 & -6 \end{bmatrix}$$

Note that when two matrices are added or subtracted, the result is matrix of the same dimensions as the original two. Also, addition and subtraction are only defined for matrices that that have the same dimensions. So, with A and C as given above, both $A + C$ and $A - C$ are undefined.

Definition 5.2. Multiplication by a scalar (cA) is defined by element-wise multiplication.

So, for example,

$$cA = 3A = 3\begin{bmatrix} 5 & 3 \\ 1 & -2 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} 3\times 5 & 3\times 3 \\ 3\times 1 & 3\times -2 \\ 3\times 4 & 3\times 0 \end{bmatrix} = \begin{bmatrix} 15 & 9 \\ 3 & -6 \\ 12 & 0 \end{bmatrix}$$

Again, the result is a matrix of the same dimensions as the original.

As in regular algebra, we can combine addition and scalar multiplication to form compound expressions.

Example 5.1. Evaluate $2A - 3B$.

An expression like $2A - 3B$ can encapsulate a large number of arithmetic operations on a large amount of data, if A and B are large matrices. This is one reason matrices are useful: it provides a concise notation for specifying a lot of data manipulation. This is especially useful when we treat matrices as variables and do algebra with them (*linear algebra*).

5.2 Algebraic properties of addition and scalar multiplication

The following rules of matrix algebra are consequences of the way we have defined addition, subtraction and scalar multiplication.

Theorem 5.1. *Let A, B and C be given matrices with the same dimensions; let c and d be given scalars. Let 0 be the zero matrix. Then:*

1. $A + B = B + A$ *(commutative law of addition)*
2. $(A + B) + C = A + (B + C)$ *(associative law of addition)*
3. $A + 0 = A$
4. $A + (-A) = 0$.
5. $c(A + B) = cA + cB$ *(distributive law of scalar multiplication)*
6. $(c + d)A = cA + dA$
7. $c(dA) = (cd)A$
8. $1A = A$

We need to prove that these algebraic identiteies are indeed true in general. The proofs can be done simply (but tediously) by considering arbitrary matrices whose components are represented symbolically, and showing that both sides of each identity do indeed agree.

Example 5.2. To prove rule 1, let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

From the definition of matrix addition, the left-hand side of the identity becomes:

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

whereas

$$B + A = \begin{bmatrix} b_{11} + a_{11} & b_{12} + a_{12} & \dots & b_{1n} + a_{1n} \\ b_{21} + a_{21} & b_{22} + a_{22} & \dots & b_{2n} + a_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} + a_{m1} & b_{m2} + a_{m2} & \dots & b_{mn} + a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

where in the last step we used the commutative law for ordinary addition of scalars. Comparing final results from these two equations, we see that indeed $A + B = B + A$.

In essence, the algebraic properties listed above imply that matrices can be treated as algebraic quantities that follow the same rules as in ordinary algebra, as least as far as addition, subtraction, and scalar multiplication are concerned. That is, when doing algebra with matrices we can treat them like ordinary variables — up to a point. So, for example, we can use algebra to solve equations involving matrices:

Example 5.3. If A, B and C are matrices that satisfy the equation

$$4A + 3B = 8C$$

find an expression for A in terms of B and C.

5.3 Vectors

Matrices having just a single column play a special role, and are called *vectors*. A vector is sometimes written as a column matrix, sometimes just as a comma-separated list. For example:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix} = (1, 5, 3).$$

Throughout this course we will follow the convention of writing vectors using boldface symbols (or with an underline, when writing by hand).

We identify the individual components of a vector using subscripts. So, example, the vector $\mathbf{x}$ above has components $x_1 = 1$, $x_2 = 5$, and $x_3 = 3$. The order of the components matters: $(1, 2, 3) \neq (1, 3, 2)$.

Lec. #8

5.4 Vector form of the general solution of a linear system

Go back to a previous example:

Example 5.4. Find, in vector form, all solutions of the linear system

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0. \end{aligned}$$

RREF gives

$$\begin{bmatrix} 1 & 0 & -4/3 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

so the general solution is

$$\begin{aligned} x_1 &= \tfrac{4}{3}t \\ x_2 &= 0 \\ x_3 &= t \end{aligned}$$

where $t \in \mathbb{R}$ is an arbitrary parameter. Introducing the vector of unknowns $\mathbf{x} = (x_1, x_2, x_3)$ we can write the general solution in vector form as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}t \\ 0t \\ t \end{bmatrix} = t \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}.$$

We see that every solution is a scalar multiple of the constant vector $(\frac{4}{3}, 0, 1)$.

Example 5.5. Find, in vector form, the general solution of

$$\begin{aligned} x_1 - x_2 - x_3 + 2x_4 &= 1 \\ 2x_1 - 2x_2 - x_3 + 3x_4 &= 3 \\ -x_1 + x_2 - x_3 &= -3. \end{aligned}$$

5.5 Matrix products

Having defined addition, subtraction and scalar multiplication for matrices, we now turn to defining matrix multiplication. It turns out that the most useful way to define the product of two matrices is *not* by element-wise multiplication, as you might have expected. The matrix product has a rather more complicated definition, whose utility will become apparent as we proceed.

Definition 5.3. Given two matrices A and B, the **matrix product** (AB) is defined by

$$AB = C$$

where C_{ij}, the (i,j)'th component of C (i.e. the entry in the i'th row and j'th column) is given by the cumulative product of the i'th row of A with the j'th row of B. That is,

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

where A has n columns and B has n rows.

Important: the product AB is only defined if the number of columns of A equals the number of rows of B.

Example 5.6. Evaluate:

(a) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -3 \\ -1 & 5 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Example (c), the product of a matrix and a vector, is a common special case that's worth extra attention. Note that the product $A\mathbf{x}$ always results in a vector.

Matrix powers can be define in terms of iterated muliplication. So, for example, if A is a matrix then $A^3 = AAA$. In general, $A^n = \underbrace{AA \cdots A}_{n \text{ times}}$.

Example 5.7. Calculate A^3 where $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$.

There is another useful way to think of matrix multiplication. Suppose the columns of A are the vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$ (we write $A = [\mathbf{v_1} \ \ \mathbf{v_2} \ \cdots \ \mathbf{v_n}]$). Then

$$A\mathbf{x} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + \cdots + x_n\mathbf{v_n}.$$

So the matrix-vector product is the vector sum of the columns of A, each weighted by the corresponding component of $\mathbf{x}$ (such a weighted sum is called a *linear combination* of the columns of A.)

5.6 Algebraic properties of matrix multiplication

The following algebraic properties are consequences of our definition of matrix multiplication (although they are tedious to prove):

1. $A(BC) = (AB)C = ABC$ (associative law)
2. $A(B + C) = AB + BC$ (distributive law for left multiplication)
3. $(A + B)C = AC + BC$ (distributive law for right multiplication)
4. $k(AB) = (kA)B = A(kB)$
5. in general, $AB \neq BA$ (matrix multiplication does *not* commute)
6. $A^mA^n = A^{m+n}$
7. $(A^m)^n = A^{mn}$

Thus, under matrix multiplication, matrices can be treated as ordinary algebraic quantities *except* that matrix multiplication does not commute: left-multiplication is fundamentally different from right-multiplication; they are not interchangeable. This is very important to remember. It is the key difference between linear algebra and the 'ordinary' algebra, and accounts for much of the richness in the theory of linear algebra.

Example 5.8. Calculate AB and BA where $A = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 4 \\ 1 & 0 \end{bmatrix}$.

Example 5.9. Calculate AB and BA where $A = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}$ and $B = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$.

Lec. #9

5.7 Matrix multiplication and systems of linear equations

There is an important connection between matrix multiplication and systems of linear equations. To illustrate this, suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Note that

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix}$$

The components of this matrix look like the left-hand side of a linear system. Indeed, the the linear system

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ 3x_1 + 4x_2 &= 8 \end{aligned}$$

is equivalent to a single matrix equation:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

or simply

$$A\mathbf{x} = \mathbf{b}$$

where $\mathbf{b} = (5, 8)$.

The matrix A is called the *coefficient matrix*; its components can be read directly from the system of equations.

Any system of linear equations can be written as a single equivalent matrix equation $A\mathbf{x} = \mathbf{b}$. This turns out to be very useful.

Example 5.10. Write a system of linear equations that is equivalent to the matrix equation $A\mathbf{x} = \mathbf{b}$ with

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Example 5.11. Find matrices A and $\mathbf{b}$ so that the linear system

$$\begin{aligned} x_1 - 2x_2 &= 0 \\ 2x_1 - 4x_2 &= 8. \end{aligned}$$

is equivalent to the matrix equation .

Identifying matrix equations with linear systems has many advantages, including:

- *any* linear system can be represented by a single equation of the form $A\mathbf{x} = \mathbf{b}$, no matter how many unknowns or equations there are

- by studying algebraic properties of the equation $A\mathbf{x} = \mathbf{b}$, we can arrive at a deeper understanding of linear systems

Since any linear system can be written in matrix form as $A\mathbf{x} = \mathbf{b}$, there is much to be gained by discovering how to solve this and other matrix equations.

To solve $A\mathbf{x} = \mathbf{b}$, one is tempted to simply "divide" both sides by A. However, matrix division is an undefined operation, so this won't work. But consider the equation

$$ax = b$$

where a, x and b are simply scalars. To solve this equation we could divide both sides by a, but it is also possible to solve it using only multiplication, by multiplying both sides by the number a^{-1}:

$$ax = b \implies x = (a^{-1})b.$$

Our goal in the following sections is to do something similar for matrix equations, i.e. to solve $A\mathbf{x} = \mathbf{b}$ using only multiplication. We will do this by finding a matrix A^{-1} so that the solution can be written as $\mathbf{x} = A^{-1}\mathbf{b}$. To do this, we need to introduce the "identity matrix".

5.8 The identity matrix

Lec. #10

Example 5.12. Evaluate $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

This example can be generalized:

Definition 5.4. The $n \times n$ *identity matrix* is the (unique) matrix I such that $IA = AI = A$ or every $n \times n$ matrix A. Every identity matrix I has the form

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

For the purposes of matrix multiplication, I plays a role similar to the number 1 in ordinary multiplication. This simple idea proves to be very useful.

Note that multiplication by the identity *does* commute.

5.9 Solving matrix equations: The matrix inverse

We are now in a position to present the main tool for solving matrix equations like $A\mathbf{x} = \mathbf{b}$:

Definition 5.5. Given an $n \times n$ matrix A, the *inverse* of A (if it exists) is the (unique) $n \times n$ matrix, written A^{-1}, such that

$$AA^{-1} = A^{-1}A = I.$$

If A^{-1} exists then A is said to be *invertible* or *non-singular*.

Now consider the equation $A\mathbf{x} = \mathbf{b}$. By multiplying this equation on both sides by A^{-1} and simplifying, using the rules of matrix algebra, we can solve for $\mathbf{x}$ as follows:

$$\begin{aligned} A\mathbf{x} = \mathbf{b} &\implies A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} \\ &\implies (A^{-1}A)\mathbf{x} = A^{-1}\mathbf{b} \\ &\implies I\mathbf{x} = A^{-1}\mathbf{b} \\ &\implies \mathbf{x} = A^{-1}\mathbf{b}. \end{aligned}$$

This gives an *algebraic* method of solving a system of linear equations, without using the Gauss-Jordan algorithm, and it gives an explicit formula for the solution. All we need is a way to find A^{-1}. Often this involves a lot of work, but for special cases it's easy.

5.10 Special case: Inverse of a 2-by-2 matrix

Theorem 5.2. *If* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *and* $ab - bc \neq 0$ *then*

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

This makes it possible to solve 2-by-2 systems quickly:

Example 5.13. Let $A = \begin{bmatrix} 5 & 4 \\ 1 & 1 \end{bmatrix}$. Find A^{-1}.

Example 5.14. Use a matrix inverse to solve the following linear system:

$$\begin{aligned} 5x_1 + 4x_2 &= 6 \\ x_1 + x_2 &= 2. \end{aligned}$$

Example 5.15. Use a matrix inverse to solve the following linear system:

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ 2x_1 + 6x_2 &= 5. \end{aligned}$$

Example 5.16. Let $A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 6 & -2 & -3 \end{bmatrix}$. Check that $A^{-1} = \begin{bmatrix} -2 & -3 & 1 \\ -3 & -3 & 1 \\ -2 & -4 & 1 \end{bmatrix}$ and solve:

$$\begin{aligned} x - y &= 4 \\ x - z &= 6 \\ 6x - 2y - 3z &= 2. \end{aligned}$$

5.11 Non-invertible matrices

We have found an algebraic way of solving a system of linear equations, by transforming the system into a matrix equation:

$$A\mathbf{x} = \mathbf{b} \implies \mathbf{x} = A^{-1}\mathbf{b}.$$

Notice this this equation implies that there is a unique solution. Thus, if a linear system does *not* have a unique solution then the coefficient matrix does not have an inverse. Of course, not all matrices have inverses, since not all linear systems have unique solutions.

Definition 5.6. If a matrix A has an inverse then we say A is *invertible* or *nonsingular*. If A does not have an inverse we say A is *non-invertible* or *singular*.

Example 5.17. If we try to find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ we get a division by zero. The matrix is singular: it does not have an inverse. Consider the linear system corresponding to $A\mathbf{x} = \mathbf{b}$:

$$\begin{aligned} x_1 + 2x_2 &= b_1 \\ 2x_1 + 4x_2 &= b_2 \end{aligned}$$

$$\begin{bmatrix} 1 & 2 & b_1 \\ 2 & 4 & b_2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 2 & b_1 \\ 0 & 0 & b_2 - 2b_1 \end{bmatrix}$$

Depending on the values b_1, b_2 this system is either inconsistent or has an infinite number of solutions. Either way, there will not be a unique solution, so we cannot solve the corresponding matrix equation in the form $\mathbf{x} = A^{-1}\mathbf{b}$. Therefore A cannot have an inverse.

5.12 Algorithm for finding A^{-1}

We have seen that matrix inverses are useful for expressing the solution of linear systems, but this will only be possible if we can *find* a matrix inverse when we need one.

Here is a general method for finding A^{-1}, if it exists:

1. Form the augmented matrix $[\ A \mid I\ ]$ where the identity I has the dimensions of A.
2. Calculate the RREF of this augmented matrix.
3. If A is invertible then the resulting RREF will be $[\ I \mid A^{-1}\ ]$.

To summarize:

$$[\ A \mid I\ ] \xrightarrow{\text{RREF}} [\ I \mid A^{-1}\ ].$$

Later we'll see why this process works; for now we'll just take it for granted and use it whenever we need to compute a matrix inverse.

Example 5.18. Find the inverse of $A = \begin{bmatrix} 1 & 4 \\ 1 & 3 \end{bmatrix}$.

Example 5.19. Find the inverse of $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$ if it exists.

Example 5.20. Show that $\begin{bmatrix} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{bmatrix}$ does not have an inverse.

5.13 More general matrix equations

Lec. #11

Suppose we want to solve the matrix equation $XA = B$ for X (assuming A has in inverse)? We can't proceed as before and multiply on the left by A^{-1} to "cancel" the multiplication by A. Instead, we multiply by A^{-1} on the right

$$\begin{aligned} XA = B &\implies (XA)A^{-1} = BA^{-1} \\ &\implies X(AA^{-1}) = BA^{-1} \\ &\implies XI = BA^{-1} \\ &\implies X = BA^{-1}. \end{aligned}$$

Example 5.21. Solve the matrix equation $ABXC = D$ for X, supposing A, B and C are invertible.

5.14 Properties of the matrix inverse

Lec. #12

To reliably use inverses in matrix algebra, we will need to know their algebraic properties:

Theorem 5.3. *If A and B are invertible then:*

1. A^{-1} *is unique.*
2. $(A^{-1})^{-1} = A$.
3. $(AB)^{-1} = B^{-1}A^{-1}$.
4. $(cA)^{-1} = \frac{1}{c}A^{-1}$.
5. $(A^n)^{-1} = (A^{-1})^n$.

Proof of 1: (by contradiction) Suppose B and C are two different inverses of A. This means that $BA = I$, $AB = I$, $CA = I$, $AC = I$. Then:

$$B = IB = (CA)B = C(AB) = CI = C$$

Proof of 2: The matrix $(A^{-1})^{-1}$ is the (unique) matrix C such that $A^{-1}C = I$ and $CA^{-1} = I$. But we alreay know that A is such a matrix, so the inverse of A^{-1} is A.

Proof of 3: It suffices to demonstrate that both $(B^{-1}A^{-1})(AB) = I$ and $(AB)(B^{-1}A^{-1}) = I$:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$
$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

Property 3 can be a useful shortcut in solving matrix equations:

$$\begin{aligned} ABX = C &\implies (AB)^{-1}(AB)X = (AB)^{-1}C \\ &\implies X = (AB)^{-1}C \\ &\implies X = B^{-1}A^{-1}D \end{aligned}$$

Property 3 can also be generalized as follows: $(A_1A_2 \cdots A_n)^{-1} = A_n^{-1} \cdots A_2^{-1}A_1^{-1}$.

5.15 The Transpose

Definition 5.7. Let A be an $m \times n$ matrix. The *transpose* of A, written A^T, is the $n \times m$ matrix formed by interchanging the rows and columns of A.

Example 5.22. Find A^T and B^T given $A = \begin{bmatrix} 1 & 0 & 3 \\ 5 & 8 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$.

The transpose has the following algebraic properties:

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(kA^T) = k(A^T)$
4. $(AB)^T = B^TA^T$
5. $(A^n)^T = (A_T)^n$ if n is a non-negative integer.
6. $(A^T)^{-1} = (A^{-1})^T$.

5.16 Elementary row operations as matrix multiplication

Here we will see why it is that calculating the RREF of $[\ A \mid I\ ]$ gives $[\ I \mid A^{-1}\ ]$.

The key observation is that every elementary row operation can be represented by a certain matrix multiplication:

Example 5.23. Show that performing the row operation $R_2 + 2R_1$ on the augmented matrix

$$\left[\begin{array}{ccc|c} -1 & 3 & 2 & 5 \\ 2 & 4 & 1 & 1 \\ -4 & 9 & 0 & 2 \end{array}\right]$$

is equivalent to left multiplication by the matrix

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

In the previous example, notice that the matrix E is what we would get if we applied the same row operation to the identity matrix. In fact, this is always the case. To see why, suppose a certain row operation (or combination of row operations) corresponds to left-multiplication by some matrix E. We have, of course, that

$$E = EI,$$

but now the right-hand side is the result of applying the given row operations to I, so this equation states that we can find E by applying the corresponding row operations to I.

Example 5.24. Find the 3×3 matrix E such that left multiplication of a $3 \times n$ matrix by E corresponds to the row operations $R_1 \leftrightarrow R_3$ following by $R_2 - 5R_1$.

Now consider calculating the RREF of $[\ A \mid I\ ]$. This means finding a sequence of row operations that reduces the left side of this matrix to the identity. Equivalently, it means finding a series of matrices $E_1, E_2, \ldots, E_k$ so that left multiplication by E_1, followed by E_2, etc., reduces A to I:

$$E_k \cdots E_2 E_1 A = I.$$

Now this says that the matrix $E_k \cdots E_2 E_1$ must be A^{-1}, since multiplication of A by $E_k \cdots E_2 E_1$ gives the identity. When we calculate the RREF of $[\ A \mid I\ ]$, we apply the same row operations to both sides, so we have

$$[\ A \mid I\ ] \xrightarrow{\text{RREF}} [\ E_k \cdots E_2 E_1 A \mid E_k \cdots E_2 E_1 I] = [\ I \mid E_k \cdots E_2 E_1\ ] = [\ I \mid A^{-1}\ ].$$

Note that we have found an interesting characterization of A^{-1}: it is the matrix that performs the series of elementary row operations required tp reduce A to I. So, for example, when we solve $A\mathbf{x} = \mathbf{b}$ by multiplying both sides by $A^{-1} = E_k \cdots E_2 E_1$, we are effectively performing the row operations that result in the RREF of the augmented matrix $[\ A \mid \mathbf{b}\ ]$:

$$[\ A \mid \mathbf{b}\ ] \xrightarrow{\text{RREF}} [\ E_k \cdots E_2 E_1 A \mid E_k \cdots E_2 E_1 \mathbf{b}\ ] = [\ I \mid A^{-1}\mathbf{b}\ ].$$

In other words, A^{-1} is a record of all the row operations required to compute the RREF of the augmented matrix; calculating $A^{-1}\mathbf{b}$ performs these row operations on $\mathbf{b}$ only, resulting in the final column of the RREF, from which we can read off the solution.

6 Applications

Lec. #13

Having built up a collection of the most essential ideas and techniques of linear algebra, we're in a position to explore some applications. The techniques discussed so far are used in a huge variety of different fields; we'll only sample some of the possibilities.

6.1 Balancing chemical equations: vector form

Go back to the previous example of balancing a chemical reaction:

$$x_1 \; C_3H_8 + x_2 \; O_2 \longrightarrow x_3 \; CO_2 + x_4 \; H_2O.$$

Rather than balance all species separately, we can balance all at once using vectors. For each molecule, define a constant vector whose elements are the number of atoms of carbon, hydrogen and oxygen, respectively. This is, represent C_3H_8 with the vector $(3, 8, 0)$, O_2 with the vector $(0, 0, 2)$, etc. The balanced chemical equation then becomes the *vector equation*:

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Of course, component-by-component this represents three equations, each expressing the balance of one species. But working at the level of vectors is more concise, there's only one equation, there's a more intuitive connection to the chemical equation, and it's obvious how to generalize to more complicated reactions with more species.

Taking everything to the left-hand side we get the homogeneous matrix equation:

$$\begin{bmatrix} 3 & 0 & -1 & 0 \\ 8 & 0 & 0 & -2 \\ 0 & 2 & -2 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Computing the RREF gives, as before:

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & -1/4 & 0 \\ 0 & 1 & 0 & -5/4 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{array}\right].$$

This time, write the solution in vector form:

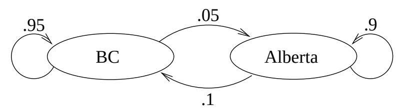
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{4}x_3 \\ \frac{5}{4}x_3 \\ \frac{3}{4}x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1/4 \\ 5/4 \\ 3/4 \\ 1 \end{bmatrix} \quad \text{where } x_3 \text{ is free.}$$

So one solution (taking $x_3 = 4$) is $\mathbf{x} = (1, 5, 3, 4)$. The structure of the entire solution set is also apparent: *every* solution is some multiple of this contant vector.

Lec. #14

6.2 Population dynamics and migration

Example 6.1. Each year 5% of BC's population moves to Alberta, while 10% of Alberta's population moves to BC. Assume, for simplicity's sake, that there is no migration to or from the rest of the world. Now there are 4,000,000 people living in BC and 3,000,000 living in Alberta.



1. How many people will be living in each province after 10 years? (Assume the foregoing assumptions continue to hold throughout this time interval.) 50 years?
2. Eventually the population distribution reaches an equilibrium. How to predict this equilibrium distribution from the transition matrix A?

To solve 1. let $x_1^{(n)}$ = number of people in BC after n years, $x_2^{(n)}$ = number of people in Alberta after n years. Let $\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)})$; so e.g. $\mathbf{x}^{(0)} = (4 \times 10^6, 3 \times 10^6)$. We can write the following equations for the populations after one year:

$$x_1^{(1)} = 0.95x_1^{(0)} + 0.1x_2^{(0)}$$
$$x_2^{(1)} = 0.05x_1^{(0)} + 0.9x_2^{(0)}$$

or in vector form:

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{bmatrix} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} = A\mathbf{x}^{(0)}.$$

The matrix A is called the *transition matrix* for this system.

By the same argument we get

$$\mathbf{x}^{(2)} = A\mathbf{x}^{(1)} = A(A\mathbf{x}^{(0)}) = A^2\mathbf{x}^{(0)}.$$

In general,

$$\mathbf{x}^{(n)} = A^n\mathbf{x}^{(0)}.$$

So to answer the question given,

$$\mathbf{x}^{(10)} = A^{10}\mathbf{x}^{(0)} = \begin{bmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{bmatrix}^{10} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \approx \begin{bmatrix} 0.732 & 0.535 \\ 0.268 & 0.465 \end{bmatrix} \begin{bmatrix} 4,000,000 \\ 3,000,000 \end{bmatrix} = \begin{bmatrix} 4,535,000 \\ 2,465,000 \end{bmatrix}$$

Note: the matrix A^{10} can be interpreted as a transition matrix, representing the net migration that has occurred after 10 years.

To answer 2. notice that if $\mathbf{x}$ is the population vector at equilibrium then

$$A\mathbf{x} = \mathbf{x} \implies (A - I)\mathbf{x} = 0 \implies \left(\begin{bmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We can't solve this by using the matrix inverse of $(A - I)$, since this would give $\mathbf{x} = (0, 0)$ (the trivial solution). In fact $(A - I)$ doesn't have an inverse, because there are an infinite number of solutions. Instead, we use Gauss-Jordan on the augmented matrix for this linear system:

$$\left[\begin{array}{rr|r} -0.05 & 0.1 & 0 \\ 0.05 & -0.1 & 0 \end{array} \right] \xrightarrow{\text{RREF}} \left[\begin{array}{rr|r} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array} \right].$$

So the general solution (in vector form) is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{where } x_2 \text{ is free.}$$

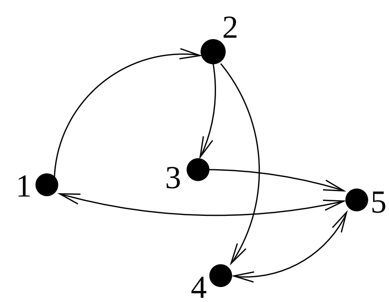
So every equilibrium solution is one where BC has twice the population of Alberta. Of course the particular value of x_2 depends on the total population of the two regions:

$$4,000,000 + 3,000,000 = 2x_2 + x_2 \implies x_2 = 7,000,000/3 \implies \mathbf{x} = \begin{bmatrix} 4,666,667 \\ 2,333,333 \end{bmatrix}$$

Lec. #15

6.3 Graphs and networks

A *graph* is a useful way to study interconnections between people and things (e.g. computers in a network). The following *directed graph* might represent possible paths of data flow in a computer network, or possible driving routes between various locations in a city, or family relationships among a group of individuals, or…



A directed graph consists of a set of *vertices* or *nodes* (numbered 1 through 5 above) and a set of *directed edges* between nodes.

Some important types of questions that might be asked about such a graph are:

- How many different paths from node 1 to node 4 take exactly 6 steps?
- How many different paths are there from node 1 to node 4 taker fewer than 6 steps?
- What is the *shortest* path from node 1 to node 4?

Of course for a simple graph we can answer these questions easily; but for a complicated graph these are hard problems. For general graphs, all of these questions can be answered using methods of matrix algebra. The key is to associate with the graph an *incidence matrix* A with elements

$$a_{ij} = \text{number of edges joining the } i\text{'th node to the } j\text{'th node.}$$

So for the graph above the incidence matrix is

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

Given A, it is possible to recover a drawing of the graph. So A is as complete a description of the graph as the picture.

Definition 6.1. A *path* or *chain* is a route (sequence of edges) from one node to another. A path that consists of n edges is called an n-chain.

Theorem 6.1. *The number of n-chains from node i to node j is equal to the element a_{ij} of A^n.*

So by direct calculation we can evaluate:

$$A^2 = \begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \end{bmatrix} \qquad A^6 = \begin{bmatrix} 10 & 0 & 4 & 14 & 0 \\ 0 & 4 & 0 & 0 & 12 \\ 6 & 0 & 2 & 8 & 0 \\ 6 & 0 & 2 & 8 & 0 \\ 0 & 6 & 0 & 0 & 16 \end{bmatrix}$$

so e.g. we can read off the number (14) of paths from node 1 to node 4 that take exactly 6 steps.

Why does this work? Consider the number of 2-chains from node 2 to node 5. Obviously there are two ($(2,3,5)$ and $(2,4,5)$); the matrix product to get element a_{25} of A^2 requires that we calculate

$$(\text{2nd row of } A) \cdot (\text{5th column of } A) = (0,0,1,1,0) \cdot (1,0,1,1,0) = 0 + 0 + 1 + 1 + 0$$

The first 1 comes from the path $(2,3,5)$; the second comes from the path $(2,4,5)$.

Theorem 6.2. *The number of paths from node i to node j that take n steps or fewer is equal to the element a_{ij} of $A + A^2 + A^3 + \cdots + A^n$.*

So by direct calculation we can find

$$A + A^2 + \cdots + A^6 = \begin{bmatrix} 15 & 6 & 6 & 21 & 15 \\ 6 & 6 & 3 & 9 & 18 \\ 9 & 3 & 3 & 12 & 9 \\ 9 & 3 & 3 & 12 & 9 \\ 9 & 9 & 3 & 12 & 24 \end{bmatrix}$$

so e.g. we can read off the number (21) of paths from node 1 to node 4 that take 6 steps or fewer.

To answer the question of shortest path from node i to node j, notice that if A has $a_{ij} \neq 0$ then there is already a path of length 1. If not, calculate A^2. If A^2 has $a_{ij} \neq 0$ then there is a path of length 2, which is the shortest path. If not, calculate A^3, etc, until you get $a_{ij} \neq 0$.

For example, to find the shortest route from node 2 to node 1, compute powers of A until the element $a_{21} \neq 0$ (we've already checked A^2):

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

So there are exactly two 3-chains from node 2 to node 1. Since there are no paths that require fewer than 3 steps, the *shortest* path from 2 to 1 requires three steps.

7 Determinants

Lec. #16

7.1 The 2×2 case

Recall the formula for the inverse of a 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The factor $ad - bc$ in the denominator is called the *determinant* of the matrix, and plays an important role. In particular, if $ab - bc = 0$ the inverse of the matrix is undefined. This provides a quick check for whether a matrix has an inverse, and hence whether the corresponding linear system has a unique solution. Determinants are important for many other reasons, so we'll spend a couple of days studying them.

Definition 7.1. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\det A = |A| = ad - bc$.

Theorem 7.1. *If A is a square matrix, then A^{-1} exists if and only if* $\det A \neq 0$.

This really becomes useful for larger matrices, because it's a much faster alternative to trying to compute A^{-1} by the Gauss-Jordan method. But how to compute $\det A$ if A is a a general $n \times n$ matrix? This requires some more definitions:

7.2 The general $n \times n$ case

Definition 7.2. Let A be an $n \times n$ matrix with elements a_{ij}. The matrix M_{ij} formed by deleting row i and column j from A is called the ij'th *minor of* A. The determinant of A, written $\det(A)$ or $|A|$, is given by

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |M_{ij}| \qquad \text{(cofactor expansion along row } i \ (i = 1, \ldots, n)) \qquad (1)$$

for any i. Equivalently, for any j,

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |M_{ij}| \qquad \text{(cofactor expansion along column } j \ (j = 1, \ldots, n)) \qquad (2)$$

So, for example if A is a 3×3 matrix then the cofactor expansion along row 1 gives

$$\det(A) = a_{11}|M_{11}| - a_{12}|M_{12}| + a_{13}|M_{13}|$$

Example 7.1. To calculate the determinant of $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$ you could expand along the first row:

$$\begin{aligned}\det A &= (1)\begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - (5)\begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + (0)\begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} \\ &= (1)\Big[(4)(0) - (-1)(-2)\Big] - (5)\Big[(2)(0) - (-1)(0)\Big] + 0 \\ &= -2\end{aligned}$$

(Note that $\det A = -2 \neq 0$ so we now know that A has an inverse.)

But It turns out that we can calculate the determinant using a cofactor expansion along any row or column:

Lec. #17

Theorem 7.2. *If A is an $n \times n$ matrix then for any k,*

$$\begin{aligned}\det A &= a_{k1}A_{k1} + a_{k2}A_{k2} + \cdots + a_{kn}A_{kn} \quad \textit{(expansion along row } k\textit{)} \\ &= a_{1k}A_{1k} + a_{2k}A_{2k} + \cdots + a_{nk}A_{nk} \quad \textit{(expansion along column } k\textit{).}\end{aligned}$$

This often makes calculating a determinant easier, because we are free to make a convenient choice of row or column along which to do the cofactor expansion — typically this is the row or column with the most zeroes.

Example 7.2. To calculate the determinant in the previous example, we can use a cofactor expansion down the third column:

$$\begin{aligned}\det A &= (0)\begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} - (-1)\begin{vmatrix} 1 & 5 \\ 0 & -2 \end{vmatrix} + (0)\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\ &= (1)\Big[(1)(-2) - (5)(0)\Big] = -2.\end{aligned}$$

Example 7.3. Calculate determinants of

$$\text{(a) } A = \begin{bmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{bmatrix} \qquad \text{(b) } B = \begin{bmatrix} 3 & 5 & -8 & 4 \\ 0 & -2 & 3 & -7 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Aside: Students who have studied recursion (e.g. computer science) will recognize the determinant calculation as a recursive algorithm: the determinant of an $n \times n$ matrix is expressed as a sum of determinants of n smaller sub-matrices of dimension $(n-1) \times (n-1)$. These in turn are expressed as determinants of yet smaller sub-matrices, of dimension $(n-2) \times (n-2)$, and so on until the algorithm terminates at the 2×2 case, for which we have a simple formula.

The number of arithmetic operations to calculate $\det A$ grows very quick as A gets bigger. Count just the multiplications (most expensive operation):

size of A	multiplications in $\det A$
2×2	2
3×3	$3 \cdot 2$
4×4	$4 \cdot (3 \cdot 2)$
$\vdots$	$\vdots$
n	$n(n-1)(n-2)\cdots 3 \cdot 2 = n!$

For the 10×10 case this is $10! = 3,628,800$. $20! \approx 2 \times 10^{18}$ (more than the age of the universe in seconds!)

7.3 Properties of determinants

The last example can be generalized:

Theorem 7.3. *If A is a triangular matrix (all zeroes either above or below the main diagonal) then $\det A$ is the product of the entries on the main diagonal.*

In particular, a triangular matrix is invertible if and only if its diagonal entries are all non-zero.

Applying elementary row operations to a matrix changes the value of its determinant in predicable ways:

Theorem 7.4. *Let A be an $n \times n$ matrix. Then*

1. *If a multiple of one row of A is added to another to produce B, then $\det A = \det B$.*
2. *If two row are interchanged to produce B, then $\det B = -\det A$.*
3. *If one row of A is multiplied by k to produce B, then $\det B = k \cdot \det A$.*

Perhaps more important are the following algebraic properties:

Theorem 7.5. *Let A be an $n \times n$ matrix. Then*

1. $\det(AB) = (\det A)(\det B)$.
2. $\det(A^{-1}) = \dfrac{1}{\det A}$.

This last is easy to prove, using the first:

$$1 = \det I = \det(AA^{-1}) = \det(A)\det(A^{-1}) \implies \det(A^{-1}) = \frac{1}{\det A}.$$

Lec. #18

7.4 Summing up

The determinant helps to tie together the disparate parts of the theory discussed so far, as summarized in the following theorem:

Theorem 7.6. *Let A be an $n \times n$ matrix. Then all of the following statements are equivalent:*

1. *A is invertible.*
2. *The linear system $A\mathbf{x} = 0$ has only the trivial solution.*
3. *The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution.*
4. *A is row equivalent to the $n \times n$ identity matrix I.*
5. *The row echelon form of A has n pivots.*
6. $\det A \neq 0$.

8 Vectors

Lec. #19

Recall: a vector is a column matrix, e.g. $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = (1, 0, 2)$.

Geometrically, 2-vectors can be identified with points in the plane; similarly, 3-vectors represent points in 3-space. It is also useful to think of a vector as displacement, e.g. relative to the origin, which specifies a magnitude and direction, but without any particular location; we represent this graphically by drawing a vector as an arrow, often but not necessarily anchored at the origin.

[pictures of 2- and 3-vectors]

8.1 Euclidean space

Definition 8.1. The set of all 2-vectors is denoted $\mathbb{R}^2$. The set of all 3-vectors is denoted $\mathbb{R}^3$. More generally, for any fixed integer $n \geq 0$, $\mathbb{R}^n$ denotes the set of all n-vectors. A set $\mathbb{R}^n$ is called a *Euclidean space.*

Geometrically, $\mathbb{R}^2$ is the set of points in the plane; $\mathbb{R}^3$ is the set of all points in 3-space. An ordinary scalar is an element of $\mathbb{R}^1$, or simply $\mathbb{R}$, the real number line.

8.2 Geometry of vector algebra

- scalar multiplication lengthens or shortens a vector while preserving its orientation (if the scalar multiple is negative, there is also a reflection)

[pictures of scalar multiplication]

- vector addition corresponds geometrically to placing vectors head to tail

[pictures of vector addition]

Vector subtraction can be viewed similarly. The difference $\mathbf{x} - \mathbf{y}$ asks"what must I add to $\mathbf{y}$ to get to $\mathbf{x}$?"

[pictures of vector subtraction]

8.3 Linear Combinations and Span

Lec. #20

Definition 8.2 (Linear Combination)**.** A *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a vector of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

where $c_1, c_2, \ldots c_n$ are scalars.

Definition 8.3 (Span)**.** The set of all possible linear combinations of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is denoted

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = \{\mathbf{v} \in V : \mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \text{ for some scalars } c_1, \ldots, c_n\}.$$

Essentially, the span of a given set of vectors is the set of points you can "get to" by forming linear combinations.

Example 8.1. Let $\mathbf{u} = (1, 3) \in \mathbb{R}^2$. Then span$\{\mathbf{u}\}$ is the set of all vectors of the form $c\mathbf{u}$ where c is a scalar. Geometrically, span$\{\mathbf{u}\}$ is the set of vectors on a line with direction $\mathbf{u}$ passing through the origin.

[*need a picture here*]

Example 8.2. Let $\mathbf{u} = (1, 2, 3),\ \mathbf{v} = (2, 0, 0) \in \mathbb{R}^3$. Then span$\{\mathbf{u}, \mathbf{v}\}$ is the set of all vectors of the form $c_1\mathbf{u} + c_2\mathbf{v}$ where c_1, c_2 are scalars. Geometrically, span$\{\mathbf{u}, \mathbf{v}\}$ is the plane through the origin containing both $\mathbf{u}$ and $\mathbf{v}$.

[*need a picture here*]

Example 8.3. Consider $\mathbf{v}_1 = (1, 2)$, $\mathbf{v}_2 = (1, -1)$. We see that span$\{\mathbf{v}_1, \mathbf{v}_2\}$ contains $\mathbf{x} = (5, 1)$, since $\mathbf{x} = 2\mathbf{v}_1 + 3\mathbf{v}_2$ (i.e. $\mathbf{x}$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$).

Example 8.4. Let $\mathbf{v}_1 = (2, 3)$, $\mathbf{v}_2 = (4, 6)$. Consider $\mathbf{x} = (4, 4)$. Is $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$?

Example 8.5. Let $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (1, -1)$. Prove that span$\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{R}^2$.

(Since you can "get to" all of $\mathbb{R}^2$ by forming linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$, we say that these vectors span $\mathbb{R}^2$.)

Example 8.6. Let $\mathbf{v}_1 = (1, 2, 0)$, $\mathbf{v}_2 = (1, 0, 1)$, $\mathbf{v}_3 = (3, 2, 2)$. Do $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$?

8.4 Linear Independence

Lec. #21

Linear independence is an important concept that's probably best introduced with a couple of examples:

Example 8.7. Let $\mathbf{u} = (1, 3),\ \mathbf{v} = (2, 6) \in \mathbb{R}^2$. Then span$\{\mathbf{u}, \mathbf{v}\}$ = span$\{\mathbf{u}\}$ = span$\{\mathbf{v}\}$ is a line through the origin with direction $\mathbf{u}$, since $\mathbf{u}$ and $\mathbf{v}$ point in the same direction. In other words, $\mathbf{u}$ and $\mathbf{v}$ don't span a plane because they don't point in independent directions.

[*need a picture here*]

Example 8.8. Let $\mathbf{u} = (1, 0, 0),\ \mathbf{v} = (0, 1, 0),\ \mathbf{w} = (1, 2, 0) \in \mathbb{R}^3$. Then span$\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is the plane with equation $z = 0$. This is the same plane obtained by considering just span$\{\mathbf{u}, \mathbf{v}\}$ because $\mathbf{w}$ does not point in a third independent direction. Loosely speaking, including $\mathbf{w}$ in the set of vectors does not increase the number of places we can get to be forming linear combinations.

[*need a picture here*]

In the previous example it's easy to see that all three vectors lie in the same plane, but consider:

Example 8.9. Suppose $\mathbf{u} = (1, 2, 3),\ \mathbf{v} = (0, 1, -1),\ \mathbf{w} = (1, 4, 1) \in \mathbb{R}^3$. Do all three vectors point in essentially independent directions, or do they all lie in a single plane? To answer this question we need to check whether any one of the vectors can be expressed as a linear combination of the other two (in which case this vector does not point in a third direction independent of the other two). In fact you can check that $\mathbf{w} = \mathbf{u} + 2\mathbf{v}$, so indeed $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$, and all three vectors lie in a single plane. In other words, the three vectors span a set in which there are only two independent directions.

The previous example motivates the following definition:

Definition 8.4 (Linear Independence)**.** Let V be a vector space. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ is said to be *linearly dependent* if

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

for some scalars $c_1, \ldots c_n$ that are not all zero. Otherwise the vectors are *linearly independent.* That is, the vectors are *linearly independent* if the equation above has only the trivial solution $c_1 = c_2 = \cdots = c_n = 0$.

The motivation behind the definition is that if it *were* possible to find scalars $c_1, \ldots, c_n$ that satisfy the equation in the definition, then it would be possible to express one of the vectors as a linear combination of the others; e.g. if $c_1 \neq 0$ then

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_n}{c_1}\mathbf{v}_n$$

so that $\mathbf{v}_1$ is a linear combination of the other vectors; that is, it doesn't point in an independent direction.

Example 8.10. Let $\mathbf{u} = (1, 2, 3),\ \mathbf{v} = (0, 1, -1),\ \mathbf{w} = (1, 4, 0) \in \mathbb{R}^3$. To test for linear independence, we could painstakingly check directly whether any of these vectors can be expressed as a linear combination of the other two (as in the previous example). It's easier to apply the definition, in which we seek numbers c_1, c_2, c_3 that satisfy the equation

$$c_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This is equivalent to a linear system with unknowns c_1, c_2, c_3, with augmented matrix

$$\left[\begin{array}{rrr|r} 1 & 0 & 1 & 0 \\ 2 & 1 & 4 & 0 \\ 3 & -1 & 0 & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{rrr|r} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

so the only solution is the trivial solution $c_1 = c_2 = c_3 = 0$. Therefore the vectors are linearly independent. In fact they span all of $\mathbb{R}^3$.

If you have two linearly independent vectors in $\mathbb{R}^2$, then it should be impossible to choose a third vector that points in a third independent direction. This idea is formalized in the following theorem.

Theorem 8.1. *A set of m vectors in $\mathbb{R}^n$ is always linearly dependent if $m > n$.*

Proof. (sketch) Following the same method as the previous example, we set up a system of equations for the unknowns $c_1, \ldots, c_m$, and the corresponding augmented matrix. We get n equations and m unknowns. If $m > n$ there will be at least one free variable, hence an infinite number of nontrivial solutions. The existence of a nontrivial solution means the vectors are linearly dependent. □

Lec. #22

An immediate consequence is the following corollary:

Corollary 8.1. *Any set of linearly independent vectors in $\mathbb{R}^n$ contains at most n vectors.*

Example 8.11. By the theorem just given, we know that the set of vectors $\mathbf{u} = (1, 2),\ \mathbf{v} = (0, 1),\ \mathbf{w} = (3, 4) \in \mathbb{R}^2$ is necessarily linearly dependent.

Linear independence gives us another way to think of existence of solutions for systems of linear equations:

Theorem 8.2. *Let A be an $n \times n$ matrix. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions if and only if the columns of the coefficient matrix A are a set of linearly dependent vectors. Conversely, if the columns of A are linearly independent then the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.*

Proof. The proof is basically the same as for Theorem 8.1 with $m = n$. □

Theorem 8.3. *Any set of n linearly independent vectors in $\mathbb{R}^n$ spans $\mathbb{R}^n$.*

8.5 Basis and Dimension

Definition 8.5 (Basis)**.** A sector of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a *basis* for a vector space V if

1. $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, and
2. $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = V$.

A basis defines a coordinate system for V, in that any $\mathbf{x} \in V$ can be expressed as a (unique) linear combination

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

The vector $\mathbf{c} = (c_1, \ldots, c_n)$ gives the *coordinates* of $\mathbf{x}$ *relative to the basis* $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Definition 8.6 (Standard Basis for $\mathbb{R}^n$)**.** In $\mathbb{R}^n$ the vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

form the *standard basis.*

Example 8.12. The vector $\mathbf{x} = (2, 5) \in \mathbb{R}^2$ can be expressed as

$$\mathbf{x} = 2\mathbf{e}_1 + 5\mathbf{e}_2.$$

The numbers 2, 5 are the familiar coordinates of $\mathbf{x}$ relative to the standard basis. An alternative basis for $\mathbb{R}^3$ is defined by the vectors $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (-1, 1)$. The same vector $\mathbf{x} = (2, 5)$ can be expressed with coordinates (c_1, c_2) relative to the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$, according to:

$$\begin{bmatrix} 2 \\ 5 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Solving this linear system gives $c_1 = 2/3$, $c_2 = -7/3$. We can think of these as the coordinates of $\mathbf{x}$ relative to a coordinate system with axes in the directions of $\mathbf{v}_1$ and $\mathbf{v}_2$. We write $(\mathbf{x})_{\mathcal{B}} = (2/3, -7/3)$ to denote that the coordinates are being specified relative to the basis $\mathcal{B}$.

Here's an immediate consequence of Theorem 8.3:

Corollary 8.2. *Every set of n linearly independent vectors in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$.*

Example 8.13. Consider the set of vectors $\mathbf{x} \in \mathbb{R}^3$ that lie on the plane

$$x + 2y + 3z = 0.$$

We've seen already that such a plane through the origin is a vector space (a subspace of $\mathbb{R}^3$); call it V. To find a basis for V, write the general solution of the equation in vector form:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2y - 3z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

Thus any point $\mathbf{x}$ in the plane can be expressed as a linear combination of $\mathbf{v}_1 = (-2, 1, 0)$ and $\mathbf{v}_2 = (-3, 0, 1)$. So $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ provides a basis for V; you can think of $\mathcal{B}$ as a coordinate system in V, with axes in the $\mathbf{v}_1$- and $\mathbf{v}_2$-directions.

Lec. #23

The number of vectors required to span a vector space V provides some measure of the "size" of V; we call this number the *dimension* of V:

Definition 8.7 (Dimension)**.** If a vector space V has a finite basis, then the *dimension* of V (written $\dim(V)$) is the number of vectors in the basis.

There's some ambiguity in this definition, since at first it seems that different bases might give different dimensions. The following theorem establishes that in fact there is just one dimension:

Theorem 8.4. *If $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ are both bases for a vector space V, then $m = n$. That is, any two bases for V have the same number of vectors.*

Example 8.14. In $\mathbb{R}^n$, a basis (e.g. the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$) consists of n vectors, so that $\dim(\mathbb{R}^n) = n$.

Example 8.15. In a previous example we found a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ for the plane V with equation $x + 2y + 3z = 0$. Therefore $\dim(V) = 2$.

A special type of subspace of $\mathbb{R}^n$ is defined by the *null space* of a matrix:

Definition 8.8 (Null Space)**.** For a given matrix A, the *solution space* or *null space* of A (written null(A)) is the set of solutions of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Example 8.16. Let $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}$. To find a basis for the null space of A and find its dimension, we first find the general solution of $A\mathbf{x} = \mathbf{0}$ as usual:

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 2 & -1 & 3 & 0 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right],$$

so the general solution is

$$\begin{cases} x = -z \\ y = z \\ z \text{ is free} \end{cases} \implies \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -z \\ z \\ z \end{bmatrix} = z \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Therefore $\mathbf{v} = (-1, 1, 1)$ is a basis for null(A), and $\dim(\text{null}(A)) = 1$.

8.6 Change of basis

Suppose we have a vector space V and a given a vector $\mathbf{x} \in V$, relative to the standard basis. Given some other basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, we would like to find the coordinates of $\mathbf{x}$ relative to $\mathcal{B}$. That is, we wish to find the coordinates $(\mathbf{x})_{\mathcal{B}} = (c_1, \dots, c_n)$ such that

$$c_1\mathbf{v}_1 + \cdots c_n\mathbf{v}_n = \mathbf{x}.$$

We can write a simpler formula for the $c_1, \dots, c_n$ by writing this equation in matrix form:

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{x}.$$

Letting $P_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $(\mathbf{x})_{\mathcal{B}} = (c_1, \dots, c_n)$, this becomes simply

$$\mathbf{x} = P_{\mathcal{B}}(\mathbf{x})_{\mathcal{B}}.$$

The matrix $P_{\mathcal{B}}$ is called the *change of basis matrix* or *transition matrix* from the standard basis to the basis $\mathcal{B}$. We can solve algebraically for $\mathbf{x}_{\mathcal{B}}$:

$$(\mathbf{x})_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{x}.$$

Example 8.17. Let $\mathbf{x} = (3, 1)$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ where $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (-1, 1)$. The transition matrix for the change of basis from the standard basis to the basis $\mathcal{B}$ is

$$P_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{so} \quad P_{\mathcal{B}}^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

Thus

$$\mathbf{x} = P_{\mathcal{B}}(\mathbf{x})_{\mathcal{B}} \implies (\mathbf{x})_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{x} = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

9 Abstract Vector Spaces

Lec. #24

9.1 Abstract vector spaces and their properties

So far we have been studying only vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ and (generally) $\mathbb{R}^n$ — the so-called *real Euclidean spaces*. Most of our considerations were motivated by either systems of linear equations or coordinate geometry. But it turns out that most of what we have done can be generalized, to the case of general *vector spaces*. While this new setting is more abstract, we gain greatly in that the ideas applicable to vectors in $\mathbb{R}^n$ (about which you have hopefully developed some intuitive ideas by now) become powerful tools applicable to much broader classes of mathematical objects. The ideas developed in this section are fundamental to many mathematical fields, including advanced calculus, differential equations, and analysis in general.

9.1.1 Definitions

Definition 9.1 (Real vector space)**.** A *real vector space* V is a set of mathematical objects, called *vectors*, together with two operations called *addition* and *scalar multiplication* that satisfy the following axioms:

1. If $\mathbf{x}$, $\mathbf{y} \in V$ then $\mathbf{x} + \mathbf{y} \in V$. *[closure under addition]*
2. If $\mathbf{x}$, $\mathbf{y} \in V$ then $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$. *[commutative law of addition]*
3. For all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$. *[associative law of addition]*
4. There is a vector $\mathbf{0} \in V$ such that for all $\mathbf{x} \in V$, $\mathbf{x} + \mathbf{0} = \mathbf{x}$. *[existence of an additive identity]*
5. If $\mathbf{x} \in V$, there is a vector $-\mathbf{x} \in V$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. *[existence of an additive inverse]*
6. If $\mathbf{x} \in V$ and α is a scalar, then $\alpha\mathbf{x} \in V$. *[closure under scalar multiplication]*
7. If $\mathbf{x}$, $\mathbf{y} \in V$ and α is a scalar then $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{x}$. *[first distributive law]*
8. If $\mathbf{x} \in V$ and α, β are scalars, then $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$. *[second distributive law]*
9. If $\mathbf{x} \in V$ and α, β are scalars, then $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$. *[associative law of scalar multiplication]*
10. For every $\mathbf{x} \in V$, $1\mathbf{x} = \mathbf{x}$. *[multiplicative identity]*

Recall that a scalar is a real number (thus the term *real* in the definition).

The axioms above are fairly obviously true for the sets $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^n$ in general, so each of these is a real vector space. However, there are much more exotic vector spaces that also satisfy these axioms. The power of the concept is that any mathematical fact that we can prove about vector spaces in general will be true for these exotic spaces, and not just for $\mathbb{R}^n$. Any theorems we prove in this section will apply to *any* vector space.

To get a sense for the generality of the concept of a vector space, it will help to consider some examples. Often the meaning of a definition becomes more clear if we consider some examples that *do not* satisfy the requirements of the definition.

Example 9.1. The unit interval $V = [0, 1] \subset \mathbb{R}$ is not a vector space. This is because V is not closed under addition: both $x = 0.5 \in V$ and $y = 0.8 \in V$, but $x + y = 1.3 \notin V$. In other words, it is possible to get outside of V using addition of elements of V.

Example 9.2. The set of non-negative integers $V = \{0, 1, 2, \ldots\}$ is a not vector space. This is because V does not contain the additive inverse: $1 \in V$ but $-1 \notin V$.

Example 9.3. The set of integers $V = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is not a real vector space. This is because V is not closed under scalar multiplication: $2 \in V$ but $0.3(2) = 0.6 \notin V$. In other words, it is possible to get outside of V using scalar multiplication.

Example 9.4. The set of points (x, y) on the line $y = mx$ is a vector space, i.e. $V = \{(x, y) \in \mathbb{R}^2 : y = mx\}$.

Most of the axioms are obvious. We'll prove some of the less obvious ones. Let $\mathbf{x} = (x_1, y_1)$ and $\mathbf{y} = (x_2, y_2) \in V$, so $y_1 = mx_1$ and $y_2 = mx_2$. To show that $\mathbf{x} + \mathbf{y} \in V$:

$$\mathbf{x} + \mathbf{y} = (x_1, mx_1) + (x_2, mx_2) = (x_1 + x_2, mx_1 + mx_2) = (x_1 + x_2, m(x_1 + x_2)) \in V.$$

To show that $-\mathbf{x} \in V$:

$$-\mathbf{x} = -(x_1, mx_1) = (-x_1, -mx_1) = (-x_1, m(-x_1)) \in V.$$

To show that $\alpha\mathbf{x} \in V$:

$$\alpha\mathbf{x} = \alpha(x_1, mx_1) = (\alpha x_1, \alpha mx_1) = (\alpha x_1, m(\alpha x_1)) \in V.$$

Clearly $\mathbf{0} = (0, 0) = (0, m0) \in V$.

The other properties follow similarly.

Notice that the line V in the previous example is equivalent to $\mathbb{R}$. One sometimes speaks of such a V as being an embedding of $\mathbb{R}$ in $\mathbb{R}^2$. More commonly we say V is a *subspace* of $\mathbb{R}^2$. Similarly a plane (a 2-dimensional vector space) in 3 dimensions is basically an embedding of $\mathbb{R}^2$ in $\mathbb{R}^3$, or a 2-dimensional subspace of $\mathbb{R}^3$:

Now some more exotic examples.

Example 9.5. Let $P_n = \{$polynomials $P(x)$: P has real coefficients and $\deg(P) \leq n\}$. Then P_n is a vector space. Clearly the sum of two polynomials of degree $\leq n$ is also a polynomial of degree $\leq n$, so P_n is closed under addition. The additive identity is given by $\mathbf{0} = 0x^n + 0x^{n-1} + \cdots + 0x + 0 \in P_n$. If $\mathbf{x} = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ then the corresponding additive inverse is $-\mathbf{x} = -a_n x^n - a_{n-1} x^{n-1} - \cdots - a_1 x - a_0$, so that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. All other properties are at least as obvious.

Example 9.6. The space $C[a, b] = \{$real-valued continuous functions on the interval $[a, b]\}$ is a real vector space. It is closed under addition because the sum of two continuous functions is also a continuous function. The addition operation is the usual addition of functions: $(f + g)(x) = f(x) + g(x)$. Similarly, scalar multiplication is defined by $(\alpha f)(x) = \alpha f(x)$. The zero element $\mathbf{0}$ is the zero function $f(x) = 0$. The additive inverse is given by $(-f)(x) = -f(x)$. With these identifications, the rest of the axioms are evident.

9.1.2 Subspaces

Lec. #25

To check that a given set is a vector space is tedious, requiring that we check many axioms. Fortunately the situation is simpler if we already have a vector space V and we wonder whether some subset $U \subset V$ is a vector space:

Theorem 9.1. *Let V be a vector space. A nonempty subset $U \subset V$ is a vector space if the following closure axioms hold:*

1. *If $\mathbf{x}$, $\mathbf{y} \in U$ then $\mathbf{x} + \mathbf{y} \in U$.*
2. *If $\mathbf{x} \in U$ and α is a scalar, then $\alpha\mathbf{x} \in U$.*

Basically, this works because U inherits all of the other axioms (algebraic properties of addition and multiplication) from V. The only axioms that might be broken by taking a subset of V are the closure axioms, so we need to check that these hold on U.

An important consequence is that every subspace of a vector space V automatically contains the $\mathbf{0}$ element. This is because for any $\mathbf{x} \in U$, closure under scalar multiplication implies that $0\mathbf{x} = \mathbf{0} \in U$ (we'll prove this in the next section).

Example 9.7. If $U \subset \mathbb{R}^3$ is the set of points lying in some plane that passes through the origin, then U is a vector space (a subspace of $\mathbb{R}^3$). To prove this, write $U = \{(x, y, z) : Ax+By+Cz = 0\}$ (points in the plane satisfy the equation of the plane in standard form). To check closure under addition, suppose $\mathbf{x} = (x_1, y_1, z_1) \in U$ and $\mathbf{y} = (x_2, y_2, z_2) \in U$. To prove that $\mathbf{x} + \mathbf{y} \in U$, write

$$\mathbf{x} + \mathbf{y} = (x_1 + x_2, y_1 + y_2, z_1 + z_2).$$

Now show that the coordinates of this vector satisfy the equation of the plane:

$$A(x_1 + x_2) + B(y_1 + y_2) + C(z_1 + z_2) = (\underbrace{Ax_1 + By_1 + Cz_1}_{=0 \text{ since } x \in U}) + (\underbrace{Ax_2 + Bx_2 + Cz_2}_{=0 \text{ since } y \in U}) = 0.$$

The check that $\alpha\mathbf{x} \in U$, write

$$\alpha\mathbf{x} = (\alpha x, \alpha y, \alpha z)$$

which is in U because

$$A(\alpha x) + B(\alpha y) + C(\alpha z) = \alpha(Ax + By + Cz) = \alpha 0 = 0.$$

Example 9.8. Let $V = C[0, 1]$ (the space of continuous functions on the interval $[0, 1]$). Let $U = \{f \in V : \int_0^1 f(x)\, dx = 0\}$. Then U is a vector space (a subspace of V). To check closure under addition, suppose f, $g \in V$. Then

$$\int_0^1 (f + g)(x)\, dx = \int_0^1 f(x)\, dx + \int_0^1 g(x)\, dx = 0 + 0 = 0,$$

so $f + g \in U$. To check closure under scalar multiplication, suppose $f \in U$. Then

$$\int_0^1 (\alpha f)(x)\, dx = \alpha \int_0^1 f(x)\, dx = \alpha 0 = 0$$

so $\alpha f \in U$.

Example 9.9. Let $V = C[0,1]$ and $U = \{f \in V : f(0) = 0\}$. Then U is a subspace of V.

Lec. #26

Example 9.10. Let $V = C[0,1]$ and $U = \{f \in V : f'(x) \text{ exists and is continuous}\}$. Then U is a subspace of V.

Given some set U that we wish to prove is a vector space, sometimes it's possible to recognize U as the intersection of two sets that have already been proved to be vector spaces. Then the following theorem guarantees that U also is a vector space.

Theorem 9.2. *Let U_1, U_2 be subspaces of a vector space V. Then $U_1 \cap U_2$ is a subspace of V.*

Example 9.11. We've already proved that $V = C[0,1] = \{\text{continuous functions on } [0,1]\}$ is a vector space. We've also proved that $U_1 = \{f \in V : \int_0^1 f(x)\,dx = 0\}$ and $U_2 = \{f \in V : f'(x) \text{ is continuous}\}$ are both subspaces of V. From the previous theorem it follows that the subset $U = \{f \in V : f \text{ has continuous first derivative and } \int_0^1 f(x)\,dx = 0\}$ is also a subspace of V, since we recognize that $U = U_1 \cap U_2$.

The span of a given set of vectors a frequently encountered subspace:

Theorem 9.3. *If V is a vector space, and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, then $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a subspace of V.*

Example 9.12. Let $\mathbf{u} = (1,2,3)$, $\mathbf{v} = (2,0,0) \in \mathbb{R}^3$. Then the plane spanned by $\mathbf{u}$ and $\mathbf{v}$ is a subspace of $\mathbb{R}^3$.

Example 9.13. Consider $V = C[0,1]$, and let $f(x) = 1$, $g(x) = x$, and $h(x) = x^2$. Then $U = \operatorname{span}\{f, g, h\}$ is the set of functions of the form $A + Bx + Cx^2$ for some coefficients A, B, C (i.e. U is the set of quadratic functions). Then U is a subspace of V. In fact U is identically the space P_2 from example 9.5.

9.1.3 Using the axioms to carry out proofs

Numerous logical consequences follow from the axioms defining a vector space. For example:

Theorem 9.4. *Let V be a vector space. Then for any $\mathbf{x} \in V$ and any scalar α,*

1. $\alpha\mathbf{0} = \mathbf{0}$
2. $0\mathbf{x} = \mathbf{0}$
3. $(-1)\mathbf{x} = -\mathbf{x}$.
4. *If $\alpha\mathbf{x} = \mathbf{0}$ then $\alpha = 0$ or $\mathbf{x} = \mathbf{0}$ (or both)*

Proof. 1. From the axioms we have $\mathbf{0} + \mathbf{0} = \mathbf{0}$, so that

$$\alpha\mathbf{0} = \alpha(\mathbf{0} + \mathbf{0}) = \alpha\mathbf{0} + \alpha\mathbf{0}.$$

Adding $-(\alpha\mathbf{0})$ to both sides we get

$$\begin{aligned} &\alpha\mathbf{0} + (-\alpha\mathbf{0}) = (\alpha\mathbf{0} + \alpha\mathbf{0}) + (-\alpha\mathbf{0}) \\ \Longrightarrow\ &\mathbf{0} = \alpha\mathbf{0} + (\alpha\mathbf{0} + (-\alpha\mathbf{0})) \\ \Longrightarrow\ &\mathbf{0} = \alpha\mathbf{0} + \mathbf{0} \\ \Longrightarrow\ &\mathbf{0} = \alpha\mathbf{0}. \end{aligned}$$

2. The proof is basically the same as above. Start with $0+0=0$, so that $0\mathbf{x} = (0+0)\mathbf{x} = 0\mathbf{x}+0\mathbf{x}$. Adding $-0\mathbf{x}$ to both sides we get $\mathbf{0} = 0\mathbf{x} + (0\mathbf{x} + (-0\mathbf{x})) = 0\mathbf{x} + \mathbf{0} = 0\mathbf{x}$.

□

It may seem that we're being inordinately fussy in carrying out the steps of the proof, but there are important reasons. Notice that every step is justified by one of the defining axioms; no algebraic "shortcuts" are taken, because it isn't apparent what shortcuts might be valid. Everything must follow directly from the defining axioms.

Every vector space is equipped with an "addition" operation we label "+", and a "zero element" we label "$\mathbf{0}$". In doing so we're re-using some familiar mathematical symbols, but they need not correspond to addition and the number 0 as we have come to know them. So no algebraic operations with them can be carried out without justification either by the defining axioms of by some theorem that follows from them. This is typical of axiomatic reasoning that is common throughout advanced mathematics.

10 Scalar Product and Orthogonality

10.1 The scalar product

The scalar- or dot-product of two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is defined by

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

(Clearly $\mathbf{x}$ and $\mathbf{y}$ must have the same dimension.)

Example 10.1. Evaluate $(1, 2, 3) \cdot (-2, 0, 5)$.

The usual laws of algebra lead to the following properties of the scalar product:

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
2. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$
3. $(a\mathbf{x}) \cdot \mathbf{y} = a(\mathbf{x} \cdot \mathbf{y})$

10.2 Length of a vector

If $\mathbf{v} = (a, b)$ is a vector in $\mathbb{R}^2$ then its *length* or *magnitude* is written

$$|\mathbf{v}| = \sqrt{a^2 + b^2}.$$

This follows from the Pythagorean Theorem. More generally, if $\mathbf{v} = (x_1, x_2, \ldots, x_n)$ is any vector in $\mathbb{R}^n$ then its magnitude is

$$|\mathbf{v}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

From the definition it follows that for any vector $\mathbf{v}$ and scalar c,

$$|c\mathbf{v}| = |c||\mathbf{v}|.$$

Notice that $|\mathbf{v}|$ can be written in terms of the dot product:

$$|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v} = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Definition 10.1. The *distance* between two vectors $\mathbf{u}$, $\mathbf{v}$ is $d(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|$.

Example 10.2. A wire is extended between the tops of two flag poles, whose tops are located at points with coordinates $(3, 4, 1)$ and $(2, 2, 2)$ in $\mathbb{R}^3$. Find the length of the wire.

Theorem 10.1 (Triangle Inequality)**.** *For any vectors* $\mathbf{u}$, $\mathbf{v}$ *we have:*

$$|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|.$$

This theorem states the intuitive fact that a straight line is the shortest distance between two points.

Proof. Use the algebraic properties of the dot product (and the Cauchy-Schwarz inequality, $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$, which we haven't discussed):

$$\begin{aligned}
|\mathbf{u} + \mathbf{v}|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\
&= |\mathbf{u}|^2 + 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2 \\
&\leq |\mathbf{u}|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + |\mathbf{v}|^2 \\
&\leq |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}| + |\mathbf{v}|^2 \\
&= (|\mathbf{u}| + |\mathbf{v}|)^2
\end{aligned}$$

□

Lec. #27

10.3 Unit vectors and direction

Definition 10.2. A vector $\mathbf{u}$ is a *unit vector* if its magnitude is 1: $|\mathbf{u}| = 1$.

Given any nonzero vector $\mathbf{v}$, we can form a unit vector having the same direction by dividing $\mathbf{v}$ by its length: $\mathbf{u} = \mathbf{v}/|\mathbf{v}|$. Then $\mathbf{u}$ points in the same direction as $\mathbf{v}$, but has unit magnitude. This is useful because it makes it easy to form other vectors, of any length we want, that point in the same direction as $\mathbf{v}$; e.g. $5\mathbf{u}$ is a vector of length 5 in the same direction as $\mathbf{v}$.

Example 10.3. You walk from the origin along a straight line through the point $\mathbf{v} = (2, 3)$. If you keep walking along this line until your distance from the origin is 10, what is your final position vector?

Definition 10.3. The *standard unit vectors* are defined as $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, $\mathbf{k} = (0, 0, 1)$. These point in the directions of the coordinate axes.

Example 10.4. Write the vector $\mathbf{v} = (4, -3, 7)$ as a combination of the standard unit vectors.

10.4 Angles between vectors

Definition 10.4. The *angle* θ between vectors $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}$$

[*need a picture here*]

Proof. (Special case of 2- and 3-D.) This follows from the cosine law, which in terms of vectors can be written:

$$|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}|\cos\theta$$

By rewriting the left-hand side using algebraic properties of the dot product:

$$\begin{aligned} |\mathbf{u} - \mathbf{v}|^2 &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} - 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\ &= |\mathbf{u}|^2 - 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2, \end{aligned}$$

the cosine law becomes:

$$\begin{aligned} &|\mathbf{u}|^2 - 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}|\cos\theta \\ &\implies -2\mathbf{u} \cdot \mathbf{v} = -2|\mathbf{u}||\mathbf{v}|\cos\theta \\ &\implies \cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} \end{aligned}$$

□

Angles between vectors really only make geometrical sense in $\mathbb{R}^2$ and $\mathbb{R}^3$; nevertheless, this definition extends the notion of "angle" to higher dimensions.

Example 10.5. Guy wires are extended from the origin to the tops of two flag poles whose tops are located at points with coordinates $(3, 4, 1)$ and $(2, 2, 2)$ in $\mathbb{R}^3$. Find the angle between the wires at the origin.

10.5 Orthogonality

Definition 10.5. Two vectors $\mathbf{u}$, $\mathbf{v}$ are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$.

In 2- and 3-D, orthogonality of vectors is equivalent to their being perpendicular, i.e. the angle θ between them is 90°. From the definition of angle between vectors:

$$\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \cos 90^\circ = 0 \Longleftrightarrow \mathbf{u} \cdot \mathbf{v} = 0.$$

Example 10.6. Let $\mathbf{u} = (1, 2, 3)$, $\mathbf{v} = (3, 2, 1)$, $\mathbf{w} = (-3, 0, 1)$. Which pairs of vectors (if any) are orthogonal?

Again, this definition only makes geometrical sense in 2- and 3-D, but it applies in higher dimensions as well. Basically, vectors are orthogonal if they point in independent directions.

Definition 10.6. Two vectors $\mathbf{u}$, $\mathbf{v}$ are *parallel* if the angle θ between them is 0 or 180°; that is, if

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \pm 1.$$

Lec. #28

Theorem 10.2 (Pythagoras)**.** *Vectors* $\mathbf{u}$*,* $\mathbf{v}$ *are orthogonal if and only if* $|\mathbf{u}+\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$.

Proof. By algebraic properties of the dot product we have:

$$\begin{aligned} |\mathbf{u}+\mathbf{v}|^2 &= (\mathbf{u}+\mathbf{v}) \cdot (\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\ &= |\mathbf{u}|^2 + 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2 \end{aligned}$$

so $|\mathbf{u}+\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$ if and only if $\mathbf{u} \cdot \mathbf{v} = 0$ (i.e. $\mathbf{u}$, $\mathbf{v}$ are orthogonal). □

10.6 Orthogonal projection

Given vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, it will be useful to have a method for resolving $\mathbf{u}$ into two components, one parallel to $\mathbf{v}$ and the other perpendicular to $\mathbf{v}$:

$$\mathbf{u} = \mathbf{u}_{\|} + \mathbf{u}_{\perp}.$$

[*need a picture here*]

Basic trigonometry leads to a simple formula for the length of $\mathbf{u}_{\|}$:

$$|\mathbf{u}_{\|}| = |\mathbf{u}|\cos\theta = |\mathbf{u}|\left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}\right) = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|}.$$

Now we can form the vector $\mathbf{u}_{\|}$ since we know both its length (by the formula above) and its direction ($\mathbf{v}/|\mathbf{v}|$). The result $\mathbf{u}_{\|}$ is called the orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$:

Definition 10.7. Let $\mathbf{u}$, $\mathbf{v}$ be nonzero vectors. The *orthogonal projection* of $\mathbf{u}$ onto $\mathbf{v}$, denoted by $\text{proj}_{\mathbf{v}}(\mathbf{u})$, is the vector

$$\text{proj}_{\mathbf{v}}(\mathbf{u}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|}\right)\left(\frac{\mathbf{v}}{|\mathbf{v}|}\right) = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v} = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}.$$

Example 10.7. Find the orthogonal projection of $\mathbf{u} = (3, 4)$ onto $\mathbf{v} = (5, 2)$.

Example 10.8. Given $\mathbf{u} = (3, -1, 4)$ and $\mathbf{v} = (4, 1, 1)$ find $\text{proj}_{\mathbf{v}}(\mathbf{u})$.

11 Lines and Planes

Lec. #29

11.1 Lines in space

Lec. #30

11.1.1 Parametric (vector) form

In two dimensions the standard form for the equation of a line is $y = mx + b$. It turns out this doesn't generalize well to describing lines in higher dimensions. For this purpose it is better to write the equation of the line in *parametric form.*

A line through the point $\mathbf{r} \in \mathbb{R}^2$ parallel to $\mathbf{v}$ can be represented by the parametric equation

$$\mathbf{x} = \mathbf{r} + t\mathbf{v}, \quad t \in \mathbb{R}$$

which describes the set of all points on the line.

This also works in three (and higher) dimensions. To describe the set of points on a line, we need only specify any point on the line and a vector giving the direction of the line.

Example 11.1. To find the parametric form of the line through the points $(3, 2, 1)$ and $(1, -1, 2)$, we first find the direction vector

$$\mathbf{v} = (3, 2, 1) - (1, -1, 2) = (2, 3, -1).$$

Then any point $\mathbf{x}$ on the line is of the form

$$\mathbf{x} = (3, 2, 1) + t(2, 3, -1), \quad t \in \mathbb{R}.$$

Component-wise, the parametric equations of the line are

$$\begin{cases} x = 3 + 2t \\ y = 2 + 3t \\ z = 1 - t. \end{cases}$$

We can readily find other points on the line, e.g. taking $t = 2$ we get $\mathbf{x} = (3 + 2\cdot 2, 2 + 3\cdot 2, 1 - 2) = (7, 8, -1)$.

11.1.2 Symmetric form

By solving for the parameter t the equation of a line can be written in *symmetric form*. Following the previous example, we get

$$t = \frac{x-3}{2} \qquad t = \frac{y-2}{3} \qquad t = \frac{1-z}{1}$$

so every point on the line must satisfy the equation

$$\frac{x-3}{2} = \frac{y-2}{3} = 1 - z.$$

From the first equality we have

$$3(x-3) = 2(y-2) \implies 3x - 2y = 5.$$

This give the equation of the line as projected on the xy-plane (i.e. top-down view). We can get projection in the xz- or yz-planes by considering other pairs of equalities.

11.1.3 Intersection of lines

Example 11.2. Consider two lines: one through the points $(0, 1, 2)$ and $(1, 2, 3)$, the other through the points $(-2, 0, 1)$ and $(1, 1, 5)$. Do these lines intersect?

It helps to first express both lines in parametric form. The direction vectors are $(1, 1, 1)$ and $(3, 1, 4)$, respectively, so in vector form the lines have parametric equations:

$$\begin{aligned} \mathbf{x} &= (0, 1, 2) + t_1(1, 1, 1) \\ \mathbf{x} &= (1, 1, 5) + t_2(3, 1, 4). \end{aligned}$$

Note that we need two parameters t_1 and t_2 (one to parametrize each line). The lines intersect if there they have a point in common; that is, if there are numbers t_1, t_2 such that both equations above describe the same point. Component-wise, this condition can be expressed as a system of linear equations:

$$\begin{cases} 0 + t_1 = 1 + 3t_2 & (x\text{-coordinate}) \\ 1 + t_1 = 1 + t_2 & (y\text{-coordinate}) \\ 2 + t_1 = 5 + 4t_2 & (z\text{-coordinate}). \end{cases}$$

The augmented matrix for this system is:

$$\begin{bmatrix} 1 & -3 & 1 \\ 1 & -1 & 0 \\ 1 & -4 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so the system is inconsistent—the lines do not intersect.

Lec. #31

11.2 Planes

11.2.1 Standard form

We have already seen how a linear equation in three variable describes a plane, e.g. the set of points $(x, y, z) \in \mathbb{R}^3$ that satisfy the equation

$$3x + 2y - 5z = 8.$$

all lie in a plane. In general, any plane in $\mathbb{R}^3$ can be described by an equation in *standard form*:

$$ax + by + cz = d,$$

for some numbers a, b, c and d.

One way to graph a plane is to plot its x-, y- and z-intercepts. E.g. for the example above the intercepts are

$$\begin{aligned} y = z = 0 &\implies x = 8/3 \ (x\text{-intercept}) \\ x = z = 0 &\implies y = 4 \ (y\text{-intercept}) \\ x = y = 0 &\implies z = -8/5 \ (z\text{-intercept}) \end{aligned}$$

[*need a picture here*]

11.3 Parametric form

Another way to describe a plane is to parametrize it much as we did for lines in the previous section. To obtain a parametric form for the equation of a plane, consider its equation in standard form as a system of linear equations (a somewhat trivial system consisting of just one equation), with one basic variable (say, x) and two free variables (y and z).

For our example above ($3x + 2y - 5z = 8$ in standard) the general solution is

$$\begin{cases} x = \frac{8}{3} - \frac{2}{3}y + \frac{5}{3}z \\ y, z \text{ are free.} \end{cases}$$

Writing the general solution in vector form we have

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{8}{3} - \frac{2}{3}y + \frac{5}{3}z \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8/3 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix}$$

where y, z are free; the free variables are parameters that parametrize the set of all points in the plane.

In general the parametric form of the equation of any plane is

$$\mathbf{x} = \mathbf{r} + s\mathbf{u} + t\mathbf{v}, \quad s, t \in \mathbb{R}.$$

In our example $\mathbf{r} = (8/3, 0, 0)$. $\mathbf{u} = (-2/3, 1, 0)$ and $\mathbf{v} = (5/3, 0, 1)$. The plane consists exactly of the set of points $\mathbf{x}$ of the form $\mathbf{x} = \mathbf{r} + s\mathbf{u} + t\mathbf{v}$ where s, t can be any real numbers. The vector $\mathbf{r}$ is some point in the plane; vectors $\mathbf{u}$, $\mathbf{v}$ specify two independent directions in the plane, along which we can travel from $\mathbf{r}$ to arrive at any other point in the plane. It makes sense that there should be two independent directions (and two corresponding parameters), since a plane is a 2-dimensional surface.

If we fix $s = 0$ then the equation reduces to $\mathbf{x} = \mathbf{r} + t\mathbf{v}$, which is the equation of a line through $\mathbf{r}$ with direction $\mathbf{v}$. This line lies in the plane. Fixing any other value of s we get a different (parallel) line, through a different point. We get another collection of lines in the plane if we hold t fixed and allow s to vary. Thus we can see that the plane consists of collections of parallel lines.

11.3.1 Normal form

In standard form the equation of a plane doesn't give us much insight into the plane's orientation. We might wonder, for example, what direction does the plane face?

A natural way to describe the orientation of a plane is by it *normal vector*, i.e. the direction of any vector perpendicular to the plane. This also leads to a natural form for the equation of the plane. Given any point $\mathbf{r}$ in the plane and the normal vector $\mathbf{n}$, we have that for any point $\mathbf{x}$ in the plane, the vector $\mathbf{x} - \mathbf{r}$ is perpendicular to $\mathbf{n}$.

[*need a picture here*]

Using the dot product to express orthogonality, we have:

$$(\mathbf{x} - \mathbf{r}) \cdot \mathbf{n} = 0.$$

This is the equation of the plane in *normal form*. The plane consists of all points $\mathbf{x}$ that satisfy this equation.

Example 11.3. The normal form of the equation of the plane that passes through the point $(2, 3, 1)$ and is oriented with normal vector $(-1, -1, 1)$ is

$$[\mathbf{x} - (2, 3, 1)] \cdot (-1, -1, 1) = 0.$$

Expanding this in terms of the components of $\mathbf{x} = (x, y, z)$ we have

$$\begin{aligned}[(x, y, z) - (2, 3, 1)] \cdot (-1, -1, 1) = 0 &\implies (x - 2, y - 3, z - 1) \cdot (-1, -1, 1) = 0 \\ &\implies -(x - 2) - (y - 3) + (z - 1) = 0 \\ &\implies -x - y + z = -4,\end{aligned}$$

which gives the equation of the line in standard form.

Example 11.4. To find the normal vector for a plane expressed in standard form, e.g.

$$3x + 2y - 5z = 8,$$

we need only rearrange the equation:

$$\begin{aligned}3x + 2y - 5z - 8 = 0 &\implies 3(x - 0) + 2(y - 4) - 5(z - 0) = 0 \\ &\implies [(x, y, z) - (0, 4, 0)] \cdot (3, 2, -5) = 0 \\ &\implies (\mathbf{x} - \mathbf{r}) \cdot \mathbf{n} = 0\end{aligned}$$

with $\mathbf{r} = (0, 4, 0)$ and $\mathbf{n} = (3, 2, -5)$. Notice that the components of $\mathbf{n}$ are exactly the coefficients from the equation in standard form. This is true in general, so we can always easily find the normal vector $\mathbf{n}$ directly from the equation in standard form.

12 Linear Transformations