

## Three Viewpoints & Five Questions

We have been motivated by questions that come up naturally when studying the basic geometry of vectors, lines, and planes in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . In particular, we have found ways to describe the lines in  $\mathbb{R}^2$  and  $\mathbb{R}^3$  in two ways: *parametrically* as the images of functions; and *implicitly* as the solutions sets of equations.

We have also encountered questions that ask us to find intersections of lines and planes, which means we want to find points which simultaneously solve several different equations. In any particular case, you can usually find answers using ad hoc methods, as long as you are dealing with a smallish number of variables and a smallish number of equations. Our next goal is to find a systematic way of handling such questions.

To do so, we will engage in a particularly mathematical way of working: we will solve these problems by taking them to be special cases of problems which look harder. It is those harder problems we will solve in one go. Somehow the wider perspective helps us see what is important and get to the bottom of things.

This chapter will introduce the core questions of linear algebra, which we will spend the next few chapters answering. But first we will introduce our most important conceptual tools.

**What are the three geometric viewpoints for understanding linear algebra?**

### SYSTEMS OF EQUATIONS – PROGRESS THROUGH GENERALIZATION

Let us consider an example problem.

### THREE VIEWPOINTS & FIVE QUESTIONS

Given two lines in the plane  $\mathbb{R}^3$ , find their intersection, if it exists.

#### Approach #1: parametric lines

Assume that the lines are given parametrically, as

$$t \mapsto \begin{pmatrix} a \\ b \end{pmatrix} + t \begin{pmatrix} c \\ d \end{pmatrix}$$

and

$$s \mapsto \begin{pmatrix} e \\ f \end{pmatrix} + s \begin{pmatrix} g \\ h \end{pmatrix}.$$

Note that we must choose different letters for the two parameters in this task, because the lines are independent objects. We seek a point  $(x, y)$  which lies on both of these lines.

If such a point exists, then the fact that the point lies on the first line means there is some value of  $t$  so that  $x = a + ct$  and  $y = b + dt$ . Similarly, the fact that the point lies on the second line means that there is some value of  $s$  so that  $x = e + gs$  and  $y = f + hs$ . Taking all of those, bundling them together, and rearranging them, we get this collection of equations:

$$\left\{ \begin{array}{rclcl} x & - & ct & = & a \\ & y & - & dt & = & b \\ x & & & - & gs & = & e \\ & y & & - & hs & = & f \end{array} \right.$$

You might be wondering why we would bother to rewrite things in this way, but it will help us see the unity of our endeavors going forward.

A more serious objection to this approach is that it is *wasteful*. To solve the problem, we only need to find  $t$ , or just  $s$ , or just  $x$  and  $y$ . Our current collection of equations asks us to find all of those. Perhaps we can be more efficient.

#### Approach #2: parametric lines, again

Rather than take apart the vectors into  $x$  and  $y$  components, we leave things as vectors. We still observe that the point we want is on both lines, so it must be the head of a vector which can be

written in two ways, one for each line. Thus we must have some pair of numbers  $s$  and  $t$  so that

$$\begin{pmatrix} a \\ b \end{pmatrix} + t \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + s \begin{pmatrix} g \\ h \end{pmatrix}.$$

Again, we'll do a little bit of re-organizing, and write this in the form below:

$$t \begin{pmatrix} c \\ d \end{pmatrix} + s \begin{pmatrix} -g \\ -h \end{pmatrix} = \begin{pmatrix} e-a \\ f-b \end{pmatrix}$$

That is much better. If we like we can unbundle these 2-vectors to write:

$$\begin{cases} ct & - & gs & = & e-a \\ dt & - & hs & = & f-b \end{cases}.$$

Notice that this set of equations is what you get from the last approach if you eliminate the variables  $x$  and  $y$ . Still, this formulation asks us to find the two numbers  $s$  and  $t$ , when just one of them will do to solve the original problem.

### **Approach #3: lines given by equations**

Suppose that each line is described by an equation, so that the points  $(x, y)$  on the first line are those that satisfy the equation  $ax + by = c$ , and the points on the second line are those that satisfy the equation  $dx + ey = f$ . Then we need to find  $x$  and  $y$  which satisfy both equations:

$$\begin{cases} ax & + & by & = & c \\ dx & + & ey & = & f \end{cases}.$$

And this feels like the perfect fit. We need to find the numbers  $x$  and  $y$ , no more, no less, to solve this problem.

Of course, all three of these have the same information in them, but the set-up and arrangement makes a difference.

What the three set-ups have in common is that they all involve looking for a set of numbers which together are a solution for several different equations all at the same time, and the equations are of a rather simple type.

That is the key insight for our reformulation and generalization. We will try to solve many equations all at the same time, but we will restrict our attention to equations which are simple in form.

### THREE VIEWPOINTS & FIVE QUESTIONS

**Definition 1.** Let  $m$  and  $n$  be counting numbers. A system of  $m$  linear equations in  $n$  unknowns has the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots = \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases},$$

where the unknowns are the variables  $x_1, \dots, x_n$  we are meant to find, and all of the other letters  $a_{ij}$  and  $b_j$  are scalars that we are supposed to know already. The numbers  $a_{ij}$  are commonly called coefficients.

A solution of this system is a single set of values for the  $x_i$ 's which makes all  $m$  of the equations true at the same time.

Each of the three approaches we outlined above involved setting up a system of linear equations. The first had  $m = 4$  equations in  $n = 4$  unknowns, and the second and third each had  $m = 2$  equations in  $n = 2$  unknowns.

If we think carefully about how we handled the second approach of the three, we see that we should be able to reframe a system of linear equations as an equation involving vectors. But this time the vectors are “bigger” than what we have used so far.

**Definition 2.** Let  $m$  be a counting number. We define an  $m$ -vector to be a vertical stack of  $m$  real numbers, like so:

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}.$$

The individual entries  $u_i$  of  $u$  are called its components. The collection of all  $n$ -vectors is called  $n$ -space, and denoted  $\mathbb{R}^n$ .

We may form linear combinations of  $m$ -vectors by doing scalar multiplication and addition in a component-by-component fashion.

Similarly, the notions of dot product, norm, and angle extend in the expected way to  $m$ -vectors.

**Definition 3.** A linear combination of  $m$ -vectors equation with  $n$  terms is an equation of the form

$$x_1 u_1 + x_2 u_2 + \cdots + x_n u_n = v,$$

where all of the  $m$ -vectors  $u_i$  and  $v$  are known, but the scalars  $x_i$  are unknowns we seek.

A solution to such an equation is a collection of scalars  $x_i$  which make the equation true.

**Theorem 4.** A system of  $m$  linear equations in  $n$  unknowns can be re-written as an equivalent linear combination of  $m$ -vectors equation with  $n$  terms which has the same solution set, and vice versa.

*Proof.* We start with a generic system of linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases},$$

and then use each equation as if it were a coordinate position

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

There are now  $n$  different  $m$ -vectors on the left-hand side, and we have the form of a linear combination of  $m$ -vectors equation with  $n$  terms.

Also, this process is reversible. We just redistribute the variables  $x_i$  in the scalar multiplication, and then remove the parentheses.  $\square$

### THREE VIEWPOINTS & FIVE QUESTIONS

#### A FIRST GEOMETRIC VIEWPOINT: HYPERPLANES

We have already seen that the set of solutions to a single linear equation  $ax + by = c$  in  $\mathbb{R}^2$  is the points which lie on a line. Similarly, the set of solutions of a single linear equation  $ax + by + cz = d$  in  $\mathbb{R}^3$  is the collection of points which lie on a plane.

Later, we will make a technical definition of the word *dimension* for our concept, but for now we think of dimension as a naïve concept of counting the number of independent directions of allowed motion. So, in the plane, with two dimensions, we get a one dimensional thing as solution to a linear equation. In 3-space, with three dimensions, we get a two dimensional thing as solution to a linear equation.

What should happen in  $\mathbb{R}^n$ , for  $n > 3$ ? Well, we will work by analogy. We expect some sort of “plane-like” object as the solution to a single linear equation.

**Definition 5.** *The set of points in  $\mathbb{R}^n$  which is described as the solution set to a single linear equation,*

$$\mathcal{H} = \{(x_1, \dots, x_n) | a_1x_1 + a_2x_2 + \dots + a_nx_n = b\},$$

*is called a hyperplane in  $\mathbb{R}^n$ .*

Now, given a system of  $m$  linear equations in  $n$  unknowns, we think of each individual equation as defining a hyperplane in  $\mathbb{R}^n$ . If a point is going to satisfy all of the  $m$  equations at the same time, then it must lie on all of the hyperplanes simultaneously. That is, the point must lie in the common intersection of all  $m$  hyperplanes.

**The Row Picture:** We interpret solving a system of  $m$  linear equations in  $n$  unknowns as finding the common intersection of  $m$  different hyperplanes in  $\mathbb{R}^n$ .

## SECOND GEOMETRIC VIEWPOINT: SPANS

If we write our constraints instead in the form of a linear combination of  $m$ -vectors with  $n$  terms,

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

then the question comes out differently. Our fundamental objects here are  $m$ -vectors, that is, elements of  $\mathbb{R}^m$ . The goal is to figure out if, and how, we may write the vector on the right-hand side of the equation as a linear combination of the vectors

$$u_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad u_2 = \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \dots u_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

**Definition 6.** Let  $\{u_1, \dots, u_n\}$  be a set of vectors in  $\mathbb{R}^m$ . The span of this set is the collection

$$\text{Span}\{u_1, \dots, u_n\} = \{a_1 u_1 + \dots + a_n u_n \mid a_1, \dots, a_n \text{ are scalars}\}$$

of all the possible linear combinations of these vectors.

This new language has us recast the task of solving the linear combination equation this way.

**The Column Picture:** We interpret solving a linear combination of  $m$ -vectors with  $n$  terms as asking if a given  $m$ -vector does or does not lie in the span of a collection of  $n$  other  $m$  vectors, and a solution is a set of coefficients which makes the given linear combination hit the prescribed vector.

Sometimes it can be useful to imagine a linear combination as a type of linkage which has extendible and contractible arms, but for

### THREE VIEWPOINTS & FIVE QUESTIONS

which the angles are frozen. We are allowed to change the lengths of the arms, but not their relative directions. That analogy is imperfect, though, as it is hard to imagine the cases where the lengths of the arms become zero or even negative.

### MATRICES, A NEW KIND OF FUNCTION

Comparing the written version of a system of linear equations with the written version of a linear combination equation, it appears that things get simpler as we go from the system to the vector equation. This is a bit of an illusion, of course, as all of the same information is there. Rather, we have bundled together each of the columns of coefficients into vectors, which abstracts the problem a bit. That is, it packages up several pieces of information (a whole column of independent coefficients) and into one new kind of object which we treat as a single unit (an  $m$ -vector).

The result is that our  $m$  equations in  $n$ -unknowns now looks like a *single* equation in  $n$  unknowns. We have used abstraction to compartmentalize things and make a conceptual simplification. This has so successfully reduced things, we should think about doing it again. What is left to abstract? Well, we combined across the columns, so it is only left to combine across the rows. So we introduce a concept that allows us to bundle together a row's worth of  $m$ -vectors.

**Definition 7.** A matrix with  $m$  rows and  $n$  columns, or an  $m \times n$  matrix for short, is a rectangular array of numbers  $a_{ij}$  enclosed in a large set of parentheses.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Each individual number  $a_{ij}$  is called an entry of the matrix. The numbers  $i$  and  $j$  are called indices. The first,  $i$ , is allowed to take a value between 1 and  $m$  and indicates which row the entry sits



in. The second,  $j$ , is allowed to take a value between 1 and  $n$  and indicates which column the entry sits in. So the  $ij$ -entry is the number which sits at the intersection of row  $i$  and column  $j$ .

Note that an  $m \times n$  matrix must have  $mn$  entries. The whole matrix must be full! We are not allowed to leave any positions empty.

Sometimes, if things are otherwise clear, a matrix is written with just a generic entry label like so:

$$A = (a_{ij})$$

It is often useful to think of an  $m \times n$  matrix as if it is a collection of  $n$  different column vectors, each of which is an  $m$ -vector. Similarly, it is sometimes useful to think of such a matrix as a collection of  $m$  different row vectors, written left-to-right, each of which is an  $n$ -vector.

This perspective with rows dominates the way matrices are read aloud. The convention is to specify the size of the matrix, and then read across the rows, top row first, bottom row last. For example the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is read aloud as “the two-by-two matrix  $a$ ,  $b$ ,  $c$ ,  $d$ .”

Now, to reorganize our linear combination equation, we will bundle the different columns into a matrix. But that leaves out the unknowns? Where should they go? Our approach is to define a new kind of multiplication.

**Definition 8.** Let  $A$  be an  $m \times n$  matrix, and let  $x$  be an  $n$ -vector. Then, if the  $n$  different columns of  $A$  are written as  $m$ -vectors  $u_1, \dots, u_n$ , we define the matrix-vector product of  $A$  and  $x$  to be

### THREE VIEWPOINTS & FIVE QUESTIONS

the  $m$ -vector

$$\begin{aligned} Ax &= \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} | \\ u_1 \\ | \end{pmatrix} + x_2 \begin{pmatrix} | \\ u_2 \\ | \end{pmatrix} + \dots + x_n \begin{pmatrix} | \\ u_n \\ | \end{pmatrix}. \end{aligned}$$

With this definition in place, we can rewrite our system of  $m$  linear equations in  $n$  unknowns, or equivalently, our linear combination of  $m$ -vectors equation with  $n$  terms, as a *matrix-vector equation*:

$$Ax = b$$

In this new version of the problem, we have the matrix  $A$  and the  $m$ -vector  $b$ , and we wish to find the  $n$ -vector  $x$ .

This new set-up is open to the criticism that we have just made up a bunch of new words in order to pack everything into much smaller notation. And that is exactly correct! Mathematicians love to do things like this. The true test is if the new words and notation help us think in a new and creative way about the problems we wish to solve. This particular instance has been effective, so we give up the objection.

How should we think of these matrix objects? Here are some perfectly acceptable ways to think about it:

- A matrix is just a convenient shorthand for a rectangular array of numbers.
- A matrix, like a vector, is a generalized sort of number that we can study for its own sake.
- A matrix is a sort of machine that operates on vectors. It somehow takes an  $n$ -vector as an input, and remixes things to produce an  $m$ -vector as output.

All of these are valid, and we will use them all. The first one is the most obvious and it is that one we will use to actually solve

systems of linear equations. The second one is a mathematician's dream problem: make up something new and see what it is like! We shall come back to this later. For now, let us focus on the third approach.

Given an  $m \times n$  matrix, we can pair it with any  $n$  vector, and produce an  $m$  vector. This is a kind of *function*. (We shall also use the words *transformation* or *mapping*, here.) This is like the ordinary kind of function from a calculus class, in that it is some sort of way to taking some inputs and turning them into outputs which is unambiguously repeatable. But the functions in calculus all have inputs *which are a single number* and outputs *which are a single number*. In our new situation, the matrix takes an input from  $\mathbb{R}^n$  and gives an output in  $\mathbb{R}^m$ . Mathematicians like to write this kind of information down like this:

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

We are supposed to read this as “ $A$  is a transformation of  $n$ -vectors into  $m$ -vectors.”

In calculus, a very useful tool for studying functions is the *graph* of that function. Given a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , you imagine the input numbers as living on a horizontal line, and the output numbers as living on a vertical line, and then plot all of the points  $(x, f(x))$  to make the graph. This will not work in our situation. At least not if we want to make an actual picture. The trouble is that our set of inputs is  $n$ -dimensional, where  $n > 1$ , and our set of outputs is  $m$ -dimensional where  $m > 1$ . So, if we want to think about  $\mathbb{R}^n$  horizontally and  $\mathbb{R}^m$  vertically... We run out of room on a piece of paper.

So we will not use graphs very often. Instead, we will make *transformational pictures* like Figure 1.

In Figure 1, the set of possible inputs, called the *domain*, for our transformation is the  $\mathbb{R}^n$  on the left, and the set of possible outputs, called the *target*, is the  $\mathbb{R}^m$  on the right. The arrow labeled  $A$  is there to remind us that  $A$  is a kind of function which picks somehow maps things from the domain into things in the target.

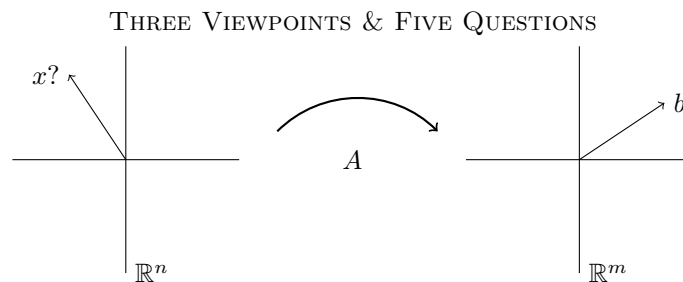


Figure 1: Matrix  $A$  as a transformation

**The Transformation Picture:** We interpret solving an  $m \times n$  matrix equation  $Ax = b$  as finding the input  $n$ -vector  $x$  which will get mapped by  $A$  onto the given output  $m$ -vector  $b$ . The solution is the vector  $x$  which we must find.

### THE FIVE MOTIVATIONAL QUESTIONS

Now we have three different viewpoints on what our equations might represent.

|  |   |
|--|---|
| algebraic formulation                              | interpretation of solution  |
| system of $m$ linear equations in $n$ unknowns     | the common intersection of $m$ hyperplanes in $\mathbb{R}^n$                                      |
| linear combination of $m$ -vectors with $n$ terms  | the coefficients realizing a vector in $\mathbb{R}^m$ as belonging to a span of $n$ other vectors |
| matrix-vector equation with an $m \times n$ matrix | the input vectors from $\mathbb{R}^n$ which are mapped to a given output vector in $\mathbb{R}^m$ |

These three viewpoints are useful in different ways. Certain tasks are easier to think about with a particular model, in a way that often depends on the information you have at hand. So you will want to practice using all three of them, and passing from any one of these to the others, so that you are capable of working

efficiently.

Whichever geometric model you use, there is still the overarching problem: solve the equation. But there are several facets to solving an equation! We will split the problem into these five parts, which will take us some time to study.

1. Is there an effective algorithm for finding solutions?
2. Does a given equation have a solution? Can you tell in advance, without computing the answer?
3. Suppose that a given equation has a solution, how many solutions does it have?
4. Suppose that a given equation has a solution, what shape is the set of solutions? Can it be reasonably understood and described?
5. If an equation has no solution, is there an approximate solution?