

Linear Algebra

Course notes and exercises

Prepared by: Duncan Clark, PhD

Last updated: January 22, 2025

Sections

0	Preface	2
1	Systems of linear equations	3
2	Row operations and echelon form	5
3	Vectors	12
4	Homogeneous systems	17
5	Linear independence	21
6	Linear transformations	23
7	The algebra of matrices	28
8	Invertible matrices	33
9	The determinant	37
10	Vector spaces	44
11	Eigenvalues and eigenvectors	51
12	Diagonalization	55
13	Markov chains	61
14	Inner products and orthogonality	67
15	Orthogonal projection	73
16	Least squares problems	78
17	Singular value decomposition	81
A	Inverse matrix theorem	90
B	Inverses and the adjugate matrix	91
C	Application: Digital image processing	92

0 Preface

0.1 How to use these notes

These course notes and exercises are intended to serve as an aid to the class. Notes for sections contain big picture ideas, definitions, statements of theorems, and worked out examples (when appropriate). They are not meant as a substitution for attending class, but rather as a reference. I *highly encourage* you to work through the exercises and not just look up answers. My intention is to have problems at various difficulty levels (problems marked with an asterisk * are intended to be challenge problems). Keep in mind, the struggle you experience when trying and failing at a problem *is* learning—keep at it and ask me questions as you have them.

Terms in **bold** are typically (parts of) definitions. Definitions are the basis of mathematics, and this course is one in your mathematical journey where they become very important. In this course we also learn a handful of nontrivial theorems. In addition to solving problems, a good place to start with studying is to know the definitions and theorems for the content covered: i.e., know why the definition does what we want it to, and what each of the theorems are, how to use them, and why the conditions are needed.

0.2 Helpful computational tools

- **Online calculator** for quickly checking calculations (such as determinant, inverse, diagonalization, etc.) [WolframAlpha](#)
- **Additional software.**
 - [Numpy](#) is a computational package for Python that includes most of the standard linear algebra algorithms.
 - [Wolfram Mathematica](#) is an extremely powerful language, but with a much higher learning curve. You can get a license through the tech support office
 - [MATLAB](#) can also make really nice pictures for you, but has a similar barrier to entry as Mathematica

0.3 Licensing

This note and exercise packet is licensed under the [Creative Commons CS BY-NC-SA 4.0 License](#).



1 Systems of linear equations

The subject of linear algebra is concerned with **linear equations** and systems thereof.

1.1 Definition. A *linear equation* in the variables x_1, x_2, \dots, x_n is any equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \quad (1.1.1)$$

for constants a_1, a_2, \dots, a_n and b .

This definition is somewhat incomplete though: what does it mean *constant*? what sort of terms are the variables x_1, x_2, \dots, x_n allowed to be? For the purpose of our class, we will primarily concern ourselves with linear equations *of real numbers*—that is, where the constant terms and variables x_i are all real numbers. However, linear algebra as a subject allows for much greater abstraction than this^[1].

1.2 Systems of equations

A **system** of linear equations is a (finite) collection of linear equations. Sometimes we'll write “linear system” as shorthand for “linear system of equations”. It's also sometimes useful to have a name to refer to a system by, typically we'll use a script letter “s”: \mathcal{S} .

1.3 Terminology. Let's say we're given a system of linear equations in the given variables of x_1, x_2, \dots, x_n . A **solution** to this system is an n -tuple $s = (s_1, s_2, \dots, s_n)$ such that substituting $x_1 = s_1, x_2 = s_2, \dots, x_n = s_n$ yields true statements for each equation in the system.

The **solution set** of a given system of linear equations is the collection of all solutions. If a given system has any solutions at all it is called **consistent**. Of note is that not all systems of equations are consistent, if a given system has no solutions it is called **inconsistent**. Two systems which have the same solution set are called **equivalent**.

1.4 Example (“Substitution method”). For instance, the following is a system of linear equations in the variables x, y and z :

$$\begin{aligned} 2x - 3y + z &= 9 \\ x + y &= 2 \\ 2x + 3z &= 6 \end{aligned} \quad (1.4.1)$$

How would find a solution to this system? Most familiar is probably the method of substitution: in this case, rearranging the second and third equations and $y = 2 - x, z = (6 - 2x)/3$, substituting these in for y and z in the first, and solving for x . This yields the sole solution to this system,

$$(x, y, z) = (3, -1, 0)$$

This method is fine for small systems (typically $n \leq 3$ or 4), but becomes increasingly tedious as the number of variables and equations gets large. What's more, is that not all systems of

^[1]We may consider equations with complex number coefficients later on this term. Generally, coefficients to a linear system can come from any *ring*.

equations have just a single solution (or any solution at all). The method of substitution misses a lot of subtlety in the *structure* of solution sets for systems of linear equations.

1.5 Geometric interpretation of systems of equations

It's useful to have a geometric understanding of what's happening when we solve such linear systems. Let's start by recalling the following facts:

1.6 Fact.

- A linear equation in two variables defines a line in the plane
- A linear equation in three variables defines a plane in 3D space

This is true more generally as well: A linear equation in n variables defines an $(n - 1)$ -dimensional “flat region”^[2] of n -dimensional space. This obviously becomes harder to visualize, but the intuition from 2- and 3-dimensional space is often enough to get the picture. A solution to a system of linear equations is then a point s which simultaneously lies on all lines/planes/hyperplanes (etc.) of a given system of equations.

1.7 Remark. As we'll see soon, linear systems of equations have solution sets which are highly structured. This is certainly not the case for non-linear systems of equations (which are in general significantly harder to solve). The upshot is that within the field of linear algebra we are able to completely describe the solution sets of linear systems, it just takes a bit of work.

1.8 Exercises

1. Determine the solution sets to the following linear systems, use either substitution or elementary row operations applied to the augmented matrix of the system.

$$(a) \quad 2s + t = 1$$

$$s - 2t = 2$$

$$(b) \quad x + 4z = -4$$

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

$$3x + 7y + 5z = 6$$

$$(c) \quad x - 3y = 5$$

$$y + 5z - x = 2$$

$$y + z = 0$$

$$(d) \quad u - 6v + 4w = -1$$

$$2v - 7w = 4$$

$$w + 2y = -3$$

$$3w + y = 6$$

^[2]Often (somewhat uncreatively) called a *hyperplane*

2. Determine all possible ways that three planes in 3D space can intersect.
3. True or False. Determine if the following statements are true or false. Justify your answer
 - (a) A linear system can have *exactly* two solutions
 - (b) Any linear system with three equations in two variables is inconsistent.
 - (c) The following linear systems are equivalent

$$\begin{array}{ll} x - 2y + z = 0 & 3y - z = 8 \\ 5y - 2z = 8 & 2y - z = 14 \\ 4x - y + 3z = -6 & 7y - z = -6 \end{array}$$

- (d) There are no values of d such that the following system is consistent:

$$\begin{array}{l} x + 3y = 1 \\ 2x + 6y = d \end{array}$$

Additional suggested exercises: *Lay* section 1.1

2 Row operations and echelon form

While it's always technically possible to enumerate the solutions of a system of linear equations using substitution, once the number of variables gets “large” (typically ≥ 5), working with the equations hands-on becomes somewhat unruly.

2.1 Enter the matrix

Matrices^[3] are an algebraic tool used to abstractify (among other things) the work done in the substitution method for solving linear systems. For us, matrix algebra is the main computational tool we'll be developing this term. It is similar to the algebra of real numbers that you are used to, but with some crucial differences. Since matrices play such a fundamental role in mathematics (and engineering and science), it's worthwhile to spend some time discussing what specifically they are.

2.2 Definition. Let n, m be integers $n, m \geq 1$. An $m \times n$ **matrix** is a rectangular grid of entries:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \quad (2.2.1)$$

^[3]Matrix is the singular, you can write the plural form of this word either as *matrices* or *matrixes*. The etymology of the word matrix in this use is a bit peculiar, the English word *matrix* comes from the Latin *mater*—meaning “mother”.

The terms $a_{i,j}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$ are called the *coefficients* of the matrix A . For now we'll just be considering matrices whose entries are real numbers. An $m \times n$ matrix always has m rows and n columns (i.e., the number of rows is read first). If $m = n$, A is called a **square matrix**.

Generic names for the coefficients of a matrix can be defined by writing $A = \{a_{i,j}\}$, or just writing $A_{i,j}$ to denote the i, j entry of A . When necessary, you can add subscripts to denote what the range of i and j should be: for instance if A is an $m \times n$ matrix you'll sometimes see the (very cluttered) notation

$$A = \{a_{i,j}\}_{1 \leq i \leq m, 1 \leq j \leq n}$$

2.3 Matrices from linear systems

Any linear system of equations has associated to it two matrices: the **coefficient matrix** and **augmented matrix**. The key is that determining the solution set of a system of linear equations can be achieved by “elementary row operations” applied to the system’s associated augmented matrix.

2.4 Definition. *Given a system of equations \mathcal{S}*

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{aligned} \tag{2.4.1}$$

the **coefficient matrix** is the matrix

$$C_{\mathcal{S}} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \tag{2.4.2}$$

and the **augmented matrix** is the matrix

$$A_{\mathcal{S}} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & b_m \end{pmatrix} \tag{2.4.3}$$

2.5 Remark. Note that the augmented matrix contains the coefficient matrix if you ignore the right-most column. It's often customary to typeset the augmented matrix of a linear system in a way that “sets aside” the right-most column (for reasons we'll soon see why) as follows:

$$\left(\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & b_m \end{array} \right) \tag{2.5.1}$$

If you see this notation, you can mentally “ignore” the vertical bar, this is just an $(m+1) \times n$ matrix after all.

2.6 Example. Returning to the system from (1.4.1) we have the following coefficient matrix

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

and augmented matrix

$$\left(\begin{array}{ccc|c} 2 & -3 & 1 & 9 \\ 1 & 1 & 0 & 2 \\ 2 & 0 & 3 & 6 \end{array} \right)$$

2.7 Remark. It’s worth pointing out that any matrix can be viewed as the coefficient matrix of a given linear system. This linear system is not unique, as the constants on the right hand side of the equations would not be specified. Moreover, any matrix with at least two columns can be viewed as the *augmented* matrix to a given linear system, which is unique.

2.8 Row operations

On its own, just writing down the coefficient or augmented matrix to a linear system is not much of a help. However, what we’ll learn now is a process for finding the solution set to a linear system by algebraically manipulating the coefficient/augmented matrix of that system.

Such algebraic manipulations are called **elementary row operations**^[4]. The idea is that we can take a given system \mathcal{S} and replace it by an equivalent system which has the same solution set, with the intention that this system be easier to determine the solution set of.

2.9 Definition (Elementary row operations). *Given a matrix A , any of the following manipulations is called an **elementary row operation** (to A).*

- Add a row to another row
- Swap the positions of any two rows
- Multiply the entries of a row by a given (nonzero) constant

Some useful notation: if we apply an elementary row operation to matrix A to achieve matrix B , let’s write $A \sim B$.

2.10 Fact. *Given a linear system \mathcal{S} , any elementary row operation done to the augmented matrix of \mathcal{S} yields an equivalent linear system.*

^[4]There’s going to be a long standing tradition in linear algebra to overload terms of things with a ton of adjectives, if you just call these “row operations” that’s also fine, the word “elementary” isn’t pulling any real weight here.

Said differently, elementary row operations do not change the solution set of (the linear system associated to) an augmented matrix^[5]. Or even more differently, for augmented matrices A, B if $A \sim B$ then the systems associated to A and B are equivalent.

2.11 Example. In full gory detail, outlined below is (a sequence of) elementary row operations applied to the augmented matrix from (1.4.1) to produce an equivalent system with an “easier to parse” solution. On the right is a description of the row operation applied at each step (writing (n) for the n -th row)

$$\begin{array}{c}
 \left(\begin{array}{ccc|c} 2 & -3 & 1 & 9 \\ 1 & 1 & 0 & 2 \\ 2 & 0 & 3 & 6 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 2 & -3 & 1 & 9 \\ 2 & 0 & 3 & 6 \end{array} \right) & \text{swap (1) and (2)} \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -5 & 1 & 5 \\ 0 & -2 & 3 & 6 \end{array} \right) & \text{subtract } 2 \cdot (1) \text{ from (2) and from (3)} \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -10 & 2 & 10 \\ 0 & -10 & 15 & 10 \end{array} \right) & \text{multiply (2) by 2 and (3) by 5} \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -10 & 2 & 10 \\ 0 & 0 & 13 & 0 \end{array} \right) & \text{subtract (2) from (3)} \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -10 & 2 & 10 \\ 0 & 0 & 1 & 0 \end{array} \right) & \text{multiply (3) by } 1/13 \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -5 & 1 & 5 \\ 0 & 0 & 1 & 0 \end{array} \right) & \text{multiply (2) by } 1/2 \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -5 & 0 & 5 \\ 0 & 0 & 1 & 0 \end{array} \right) & \text{subtract (3) from (2)} \\
 \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right) & \text{multiply (2) by } -1/5 \\
 \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right) & \text{subtract (2) from (1)}
 \end{array}$$

^[5]It’s worth pointing out the algorithmic efficiency of row reduction is $O(n^3)$ (meaning, that the number of steps required to reduce an $n \times n$ matrix is going to grow like a constant multiple of n^3). In particular, this is a problem that can be solved in polynomial time: your computer is very efficient at row reducing very large matrices, which is good because typical applications of linear algebra (e.g., machine learning) involve matrices of thousands (or more) rows or columns

Thus we've turned the original system (1.4.1) into the system \mathcal{S}' given by

$$\begin{aligned} x &= 3 \\ y &= -1 \\ z &= 0 \end{aligned} \tag{2.11.1}$$

From here it's pretty simple to read off the solution, the given triplet $(x, y, z) = (3, -1, 0)$. Moreover, any other solution to \mathcal{S} would *have to* be a solution to \mathcal{S}' , and since it's "clear" that \mathcal{S}' only has the one solution, we've now found the entire solution set to \mathcal{S} .

2.12 Echelon form

The goal of elementary row operations being to "simplify" the reading-off of the solution set of a linear system, it's perhaps unclear what the desired "simplified" form of a given matrix should be. One answer is called *echelon form* as described below (informally, a matrix in echelon form has a bunch of 0s in the lower left hand corner that make the shape of a pixelated triangle).

2.13 Definition. *A matrix is said to be in **echelon form** if the following are satisfied*

- All nonzero rows are above any rows of zeros
- Each leading entry (i.e., first non zero entry) of a row is farther to the right than the leading entry of the row above it
- In a given column, all entries below a leading entry are zero

*A matrix is said to be in **reduced echelon form** if it is in echelon form and further satisfies:*

- Each leading entry is a 1
- Each leading 1 is the only nonzero entry in its column

Reduced echelon form is the easiest to read the solution set off of, as we've seen in example 2.11 and will see more throughout this term. I leave it to you to convince yourself of the following fact:

2.14 Fact. *Given a matrix A , there is a unique matrix B in reduced echelon form such that B is obtained from A by elementary row operations.*

While it's good to get practice doing row operations by hand, it ultimately becomes (extremely) tedious to do by hand. Fortunately, we have very efficient algorithms and very fast computers which are more than happy to do these calculations for us^[6].

^[6]It's not unreasonable to call your graphics card a big linear algebra calculator, or at least certain cores of it.

2.15 Free variables

Suppose you are given a linear system of equations, and as a result of row operations have the following matrix in reduced echelon form:

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 1 \\ 0 & 1 & -1 & -3 \end{array} \right) \quad (2.15.1)$$

How would you describe the solution set? In equations, this would look like:

$$x_1 + 3x_3 = 1, \quad x_2 - x_3 = -3 \quad (2.15.2)$$

The thing is, there is more than one possible solution here: In this case, we can describe variables x_1 and x_2 in terms of the third variable x_3 . In such a situation we call x_3 a **free variable** (also called **independent variable**)—idea being that x_3 is not constricted, any real number can be plugged in for x_3 so long as the other variables are calculated by the equations from (2.15.2). The solution for this system is then described as follows:

$$\begin{aligned} x_1 &= 1 - 3x_3 \\ x_2 &= x_3 - 3 \\ x_3 &\text{ is free} \end{aligned} \quad (2.15.3)$$

The variables x_1, x_2 are not free (often called **bound** or **dependent**). So, one such solution would come from $x_3 = 0$, i.e., $(x_1, x_2, x_3) = (1, -3, 0)$; another comes from $x_3 = 10$ (i.e., $(-29, 7, 10)$), etc. Overall, there are infinitely many solutions, but each solution comes from the “template” described by the solution set (2.15.3).

2.16 Remark. If you’re familiar with parametric curves, this is essentially the same idea: here we’re parametrizing the solution set in terms of the variable x_3 ^[7]. Solution sets can have multiple free variables, or have none at all. This is part of the fun.

2.17 Exercises

- Determine all values x such that the following matrix is the augmented matrix of a consistent linear system, or explain why no such x exists

$$\left(\begin{array}{cc|c} 1 & x & 4 \\ 3 & 6 & 8 \end{array} \right)$$

- Suppose that you have the following augmented matrix for a given linear system \mathcal{S}

$$\left(\begin{array}{ccc|c} 1 & 7 & 3 & -4 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array} \right)$$

Is \mathcal{S} consistent? Explain why or why not.

^[7]And moreover, the solution set is a line given by the parametrization $\vec{r}(t) = (1 - 3t, t - 3, t)$

3. Determine a relationship between x, y and z which makes the following the augmented matrix of a consistent linear system

$$\left(\begin{array}{ccc|c} 1 & -4 & 7 & x \\ 0 & 3 & -5 & y \\ -2 & 5 & -9 & z \end{array} \right)$$

4. Put the following matrices in reduced echelon form.

(a) $\begin{pmatrix} 1 & -3 & 5 \\ 8 & 3 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 7 \\ 7 & 8 & 9 & 10 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & -4 & 2 & 0 \\ -9 & 12 & -6 & 0 \\ -6 & 8 & -4 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 3 & -4 & 2 & 0 \\ -9 & 12 & -6 & 0 \\ -6 & 8 & -4 & 1 \end{pmatrix}$

5. For each matrix from problem 4., assume it is the augmented matrix of a given linear system, and describe the solution set to this system.
6. Suppose you are given a linear system of equations. Your friend Jimbo (a known contrarian) performs row reductions to the augmented matrix, but arrives at the answer

$$\left(\begin{array}{cccc|c} 0 & 0 & 0 & 6 & 4 \\ 0 & -2 & 1 & 0 & 7 \\ 4 & 0 & -1 & 0 & -1 \end{array} \right)$$

- (a) Is his matrix in echelon form?
- (b) Is his matrix still useful for determining the solution set of the original system? Explain why or why not.
7. True or False. Determine if the following statements are true or false. Give a justification of your answer.
- (a) If matrix B is attainable from matrix A by elementary row operations, then A is also attainable from B by elementary row operations
- (b) Any linear system with the following coefficient matrix is consistent (here, * denotes an arbitrary real number)

$$\begin{pmatrix} 1 & * & * \\ 0 & 2 & * \\ 0 & 0 & 3 \end{pmatrix}$$

- (c) Any linear system with the following coefficient matrix is consistent (here, * denotes an arbitrary real number)

$$\begin{pmatrix} 1 & * & 0 \\ * & 2 & * \\ 0 & * & 3 \end{pmatrix}$$

- (d) If a linear system is inconsistent then its coefficient matrix must be equivalent by elementary row operations to a matrix which contains a row of all zeros.

Additional suggested exercises: *Lay*, section 1.2

3 Vectors

We now begin one of the most recurrent themes in all of linear algebra: describing something we already know in a new way. The thing in question now is linear systems; we've already discussed how a linear system can be interpreted as a matrix (coefficient or augmented). In this section we'll describe systems in terms of vectors and vector equations.

You're probably familiar with vectors (at least vectors in the plane, or 3D space, etc.). What may be less familiar is that a *vector* is really just a *matrix*.

3.1 Definition. Let $n \geq 1$ be given. A **vector** \vec{x} in \mathbf{R}^n is an $n \times 1$ matrix, i.e.,

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{3.1.1}$$

for some (real) constants x_1, x_2, \dots, x_n .

3.2 Terminology. The number n of components of a given vector is called that vector's **dimension**. As a symbol, \mathbf{R}^n denotes the set of all vectors of dimension n . We write $\vec{x} \in \mathbf{R}^n$ to say that \vec{x} is a vector in \mathbf{R}^n . For any given n , we say that \mathbf{R}^n is a **real** or **Euclidean**^[8] space.

Though this terminology might be new, you already know and love these objects: **R** is the usual number line $(-\infty, \infty)$, **R**² is the (real) plane^[9], **R**³ is the (real) 3D space, etc. It may be daunting to think geometrically about something like **R**¹⁰⁷, but at least algebraically, a “thing” in this space is just a 107×1 matrix.

3.3 Remark. When necessary, we distinguish vectors as defined in definition 3.1 as **column vectors** (the entries of \vec{x} lie in the same column). It's also possible to write \vec{x} as a **row**

^[8]Named for the ancient Greek mathematician Euclid, who built the axioms for “doing geometry” in the plane a long (long) time ago.

^[9]You might be tempted to call this the *xy*-plane, and though that's “fine” there is benefit to the abstraction of not-yet-specifying the “names” of the variables that makeup the two axes of the plane.

vector, i.e., as a $1 \times n$ matrix as such

$$\vec{x}_{\text{row}} = (x_1 \ x_2 \ \cdots \ x_n)$$

For various reasons, it's more convenient to use column vectors as the standard notion of “vector” in this class, though we will see row vectors occasionally. This does lead to some typesetting woes, I'll sometimes write $x = \{x_i\}$ to denote the column vector x and specify its coordinates, even though this is somewhat unclear notation.

It's also not *entirely* necessary to write the arrow over the top of \vec{x} . Eventually we'll just write x for a vector and *know* what that means, however, it can be mentally helpful for now to distinguish \vec{x} as a vector, rather than just a single real number. For this section, I'll continue to write the arrow, then likely drop it in later sections.

3.4 The algebra of vectors

Vector have similar properties to real numbers, but not *all* properties of the real numbers. It's important to lay down the law on what operations of vectors are defined.

Let's suppose a dimension n is given and that \vec{x} and \vec{y} are vectors in \mathbf{R}^n . The **sum** of \vec{x} and \vec{y} is the vector $\overrightarrow{x+y}$ defined by

$$\overrightarrow{x+y} = \vec{x} + \vec{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad (3.4.1)$$

Similarly, if k is a scalar (i.e., some real number), the **scalar multiple** of \vec{x} by k is the vector \overrightarrow{kx} given by

$$\overrightarrow{kx} = k\vec{x} = \begin{pmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{pmatrix} \quad (3.4.2)$$

Of specific note is that the vector $-\vec{x}$ is just $(-1)\vec{x}$.

It's useful to have notation for the **zero vector**^[10], that is, the vector $\vec{0}$ given by

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.4.3)$$

The notation $\vec{0}$ will typically be used to distinguish the zero vector from the scalar 0, though you may see people just write 0 for both. The following laws describe the *algebra* of vectors. I leave it to you to show these properties all hold.

^[10]Technically, this should be called “the zero vector in \mathbf{R}^n ”, but I think we can live with this level of ambiguity.

3.5 Theorem. For vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbf{R}^n$ and scalars k, ℓ the following properties^[11] hold:

- (i) $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
- (ii) $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$
- (iii) $\vec{x} + \vec{0} = \vec{0} + \vec{x} = \vec{x}$
- (iv) $\vec{x} - \vec{x} = -\vec{x} + \vec{x} = \vec{0}$
- (v) $k(\vec{x} + \vec{y}) = k\vec{x} + k\vec{y}$
- (vi) $(k + \ell)\vec{x} = k\vec{x} + \ell\vec{x}$
- (vii) $1\vec{x} = \vec{x}$
- (viii) $k(\ell\vec{x}) = (k\ell)\vec{x}$

3.6 Span of vectors

As a consequence of Theorem 3.5, we can add and subtract as many scalar multiples of a collection of vectors (of the same dimension) as we desire. As we'll see, it's often useful to decompose a vector into simpler constituent parts (i.e., a sum of scalar multiples of "simpler" vectors). First, some definitions.

3.7 Definition. Given a collection $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$ of vectors in \mathbf{R}^n . We say that a vector \vec{y} is a **linear combination** of $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$ if there are scalars c_1, c_2, \dots, c_k such that

$$\vec{y} = c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_k\vec{x}_k \quad (3.7.1)$$

Similarly, the **span** of vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$ is defined to be the collection of all linear combinations of the vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$, and denote this set as

$$\text{Span}\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k\} \quad (3.7.2)$$

3.8 Example. Let's consider the two pairs of vectors

$$\vec{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \vec{x}_2 = \begin{pmatrix} 4 \\ 1 \\ 4 \end{pmatrix} \quad \vec{y}_1 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \vec{y}_2 = \begin{pmatrix} -4 \\ 2 \\ -6 \end{pmatrix}$$

Both are two vectors in \mathbf{R}^3 , but their spans look different. Specifically $\text{Span}\{\vec{x}_1, \vec{x}_2\}$ is a *plane*^[12] in \mathbf{R}^3 : from the origin, you can "walk" in the directions provided by \vec{x}_1 or \vec{x}_2 , and these are non-parallel directions.

^[11]Abstractly, any set X that contains things called "vectors" that satisfy statements from of this theorem is called a (real) *vector space*. For instance, the set X of continuous functions $f: \mathbf{R} \rightarrow \mathbf{R}$ is a vector space (why?) but not a finite dimensional one.

^[12]Specifically with a normal vector of $\vec{x}_1 \times \vec{x}_2$ through the origin—if this is not familiar to you don't worry too much. If \mathbf{R}^3 has the usual coordinates x, y, z , the equation for this plane is $5x + 8y - 7z = 0$.

On the other hand \vec{y}_1 and \vec{y}_2 are *parallel* (i.e., one is a scalar multiple of the other). So, there is no “second direction” to walk relative to \vec{y}_1 . The set $\text{Span}\{\vec{y}_1, \vec{y}_2\}$ is then a *line* in \mathbf{R}^3 . Note that the main distinguishing feature of these two spans is their dimension: one is 2D (a plane), the other is 1D (a line), despite both being the span of two vectors.

We’ll soon learn more structure that the span of vectors has. For now, let’s call a collection of vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$ of \mathbf{R}^n a **spanning set** (of \mathbf{R}^n) if $\text{Span}\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\} = \mathbf{R}^n$. That is, any vector in \mathbf{R}^n belongs to the span of the given vectors.

3.9 Remark (Useful notation). Some useful notation that we will begin to make frequent use of is to write $x \in X$ to mean that “ x is a thing from the set X ”. For instance, the statement $\vec{x} \in \mathbf{R}^n$ is the same as saying \vec{x} is a vector in \mathbf{R}^n .

3.10 The matrix-vector product

If $A = \{a_{i,j}\}$ is an $m \times n$ matrix, and \vec{x} is a vector in \mathbf{R}^n (given by components x_i for $1 \leq i \leq n$), then the matrix A “acts on” \vec{x} to create a new vector \vec{Ax} ^[13]. The output vector \vec{Ax} is defined as the following

$$\vec{Ax} = x_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{pmatrix} + x_2 \begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{pmatrix} \quad (3.10.1)$$

This may look complicated but, really this is just matrix multiplication as we’ll see later on. Writing the output vector \vec{Ax} in this way tells us a little bit about what’s happening though: the output is a linear combination of the *column* vectors from the matrix A , with weightings given by the components of the vector \vec{x} .

It can be useful to have some notation here, given j with $1 \leq j \leq n$ let’s write let’s write \vec{a}_j for the (column) vector

$$\vec{a}_j = \begin{pmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{pmatrix}$$

We can describe the matrix A by specifying that it has column vectors consisting of the vectors \mathbf{a}_j as follows:

$$A = (\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_m) \quad (3.10.2)$$

(Note that (3.10.2) is still an $m \times n$ matrix, not a row vector.) It then follows that \vec{Ax} is given by

$$\vec{Ax} = x_1 \vec{a}_1 + x_2 \vec{a}_2 + \cdots + x_n \vec{a}_n = \sum_{i=1}^n x_i \vec{a}_i$$

^[13]In fact, these type of operations are *linear transformations* that we’ll be working with throughout this term, but more on that soon

Why bother with matrices and vectors? One reason is that the algebra of working with such matrices and vectors is what underlies linear systems as discussed previously. We're building our way up to working with matrices as our basic objects, the following relation is what allows us to do this.

3.11 Fact. Let $A = \{a_{i,j}\}$ be an $m \times n$ matrix, and $b = \{b_j\}$ a vector in \mathbf{R}^m . Let \mathcal{S} be the linear system of equations

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

Then a vector $x = \{x_i\}$ in \mathbf{R}^n is a solution to the system \mathcal{S} if and only if x is a solution to the matrix-vector equation $Ax = b$.

That is to say, **solving linear systems is the same thing as solving matrix-vector equations**. I leave it to you to convince yourself of the following theorem, which ties together the main new topics of this section so far.

3.12 Theorem. Let A be an $m \times n$ matrix. The following statements are equivalent:

- (i) For any vector $b \in \mathbf{R}^m$, $Ax = b$ has a solution
- (ii) Each vector $b \in \mathbf{R}^m$ is a linear combination of the column vectors of A
- (iii) The column vectors of A are a spanning set of \mathbf{R}^m
- (iv) For each vector $b \in \mathbf{R}^m$, the associated linear system to the augmented matrix is consistent

$$(\vec{a}_1 \quad \vec{a}_2 \quad \cdots \quad \vec{a}_n \mid b)$$

(Here, \vec{a}_i are the column vectors from A)

3.13 Exercises

1. Evaluate the following matrix-vector products, or explain why the product is not defined.

$$(a) \begin{pmatrix} 4 & 1 \\ 5 & -7 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

$$(b) \begin{pmatrix} -1 & 1 & 1 \\ 8 & 6 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$(c) \begin{pmatrix} -1 & 1 & 1 \\ 8 & 6 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$(d) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}$$

2. Determine if \vec{b} is a linear combination of the vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$;

$$\vec{a}_1 = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \vec{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \vec{a}_3 = \begin{pmatrix} 5 \\ -6 \\ 8 \end{pmatrix}, \vec{b} = \begin{pmatrix} 2 \\ -1 \\ 6 \end{pmatrix}$$

If so, is \vec{b} also a linear combination of \vec{a}_1 and \vec{a}_2 *only*?

3. Let $\vec{x} = \begin{pmatrix} 1 \\ 4 \\ -2 \end{pmatrix}, \vec{y} = \begin{pmatrix} -2 \\ -3 \\ 7 \end{pmatrix}$. Solve for h such that $\vec{b} = \begin{pmatrix} 4 \\ 1 \\ h \end{pmatrix}$ is in $\text{Span}\{\vec{x}, \vec{y}\}$ or explain why no such h exists.
4. Let $A = \{a_{i,j}\}$ be an $m \times n$ matrix and \vec{x} a vector in \mathbf{R}^n .

- (a) Show that $Ax = b$ for $b \in \mathbf{R}^m$ given by^[14]

$$b_i = \sum_{j=1}^n a_{i,j} x_j \quad (1 \leq i \leq m)$$

- (b) Use part a. to explain why Fact 3.11 is true.
5. Prove theorem 3.5
6. True or False. Determine if the following statements are true or false. Justify your answers.
- (a) Any spanning set of \mathbf{R}^n must have at least n vectors.
 - (b) Any spanning set of \mathbf{R}^n must have exactly n vectors
 - (c) The zero vector, $\vec{0}$, is in the span of any collection of vectors
 - (d) If \vec{x} and \vec{y} are vectors in \mathbf{R}^3 such that $\text{Span}\{\vec{x}, \vec{y}\}$ is a plane, then \vec{x} and \vec{y} cannot be parallel.

Additional suggested exercises: *Lay*, sections 1.3

4 Describing the solution set of a matrix-vector equation

Now that we understand how a system of linear equations can be phrased in terms of matrices and vectors, we turn to giving a description of the solution set to a given system. Suppose

^[14]If you know the *dot product*, this is just the dot product of $\vec{a}_i \cdot x$. We'll explore dot products more later.

we're given a system $Ax = b$ of linear equations. The easiest way to read off the solution set is to put the augmented matrix

$$(A \mid b)$$

into (reduced) echelon form.

4.1 Example. Let's describe the solution set of $Ax = b$ for A and b as follows:

$$A = \begin{pmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{pmatrix}, \quad b = \begin{pmatrix} 7 \\ -1 \\ -4 \end{pmatrix}$$

First, we find the augmented matrix, and put it in reduced echelon form as follows

$$(A \mid b) = \left(\begin{array}{ccc|c} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -4/3 & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

This yields the system of equations

$$x_1 - 4/3x_3 = -1, \quad x_2 = 2, \quad 0 = 0$$

The last of these is clearly useless (it is satisfied trivially), the first two can be described in terms of the free variable x_3 as

$$x = \begin{pmatrix} -1 + 4/3x_3 \\ 2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 4/3 \\ 0 \\ 1 \end{pmatrix}$$

Let's break down the components of this solution. First, some terminology.

4.2 Terminology. A system of linear equations is called **homogeneous** if it can be written in the form

$$Ax = \vec{0} \tag{4.2.1}$$

where $A = \{a_{i,j}\}$ is some $m \times n$ matrix and x a vector in \mathbf{R}^n . It is not hard to convince ourselves that $x = \vec{0}$ is always a solution to any homogeneous system $Ax = 0$. A solution x to a homogeneous linear system $Ax = \vec{0}$ is called **nontrivial** if $x \neq \vec{0}$.

In example 4.1, anything in the span of the vector $\begin{pmatrix} 4/3 \\ 0 \\ 1 \end{pmatrix}$ is a solution to the homogeneous equation $Ax = \vec{0}$ (you should check this). The solution to the equation $Ax = b$ is then this span *shifted* by the vector $p = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$.

4.3 Theorem. Let A be an $m \times n$ matrix, and b a vector in \mathbf{R}^m . Then, all solutions to the equation $x \in \mathbf{R}^n$ to $Ax = b$ are of the form

$$x = p + h \quad (4.3.1)$$

where p is any solution to $Ax = b$ and h is a solution to the homogeneous equation $Ax = \vec{0}$. Moreover, the collection of all vectors h will always be the span of some (potentially empty) collection of vectors.

4.4 Terminology. Given a matrix A , let's call the set of all solutions to the homogeneous equation $Ax = \vec{0}$ the **nullspace** of A ^[15]. Note then that the nullspace of a matrix is either (i) just the zero vector, or (ii) the span of some finite (nonempty) collection of vectors^[16]. In either condition, the nullspace is given as the span of a collection of vectors, and the easiest way to figure out this collection is to put A in reduced echelon form.

The previous theorem then is just the statement that the solution set to any *non-homogeneous* equation $Ax = b$ is a shift^[17] of the nullspace of A by a *particular* solution to the non-homogeneous equation. If for instance the nullspace of a matrix A consists of the span of two vectors $u, v \in \mathbf{R}^3$, then the nullspace is the plane through the origin that these generate, and for $b \neq \vec{0}$, the solution set to $Ax = b$ is the plane through the point p (corresponding to the particular solution $Ap = b$) given by directions u, v .

4.5 Exercises

1. Write an analogous matrix-vector equation for the linear systems, then describe the general solution to the given system

$$3x + y - 5z = 9$$

$$(a) \quad y + 4z = -3$$

$$x - 10z = 1$$

$$(b) \quad \begin{aligned} -5x + 7y + 9z &= 0 \\ x - 2y + 6z &= 0 \end{aligned}$$

$$x + 3y - 5z = 4$$

$$(c) \quad x + 4y - 8z = 7$$

$$-3x - 7y + 9z = -6$$

2. Determine if $Ax = b$ has a solution, for A, b as follows:

$$A = \begin{pmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ -7 \\ -3 \end{pmatrix}$$

3. Describe the set of all solutions to the homogeneous system $Ax = \vec{0}$ for A the following matrix

^[15]Sometimes you'll see this also called the *kernel*

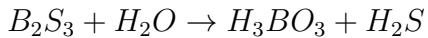
^[16]We can just say that the span of “no vectors” is just the zero vector to combine these two

^[17]In fancy words, you'll see this sometimes called a *coset*

(a) $A = \begin{pmatrix} 3 & -9 & 6 \\ -1 & 3 & -2 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{pmatrix}$

4. Write a 3×3 nonzero matrix A such that $x = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ is a solution to $Ax = 0$. Can you do the same with A a 5×3 matrix?
5. A neat application of linear systems is in balancing chemical equations. Consider a reaction between boron sulfide and water to form boric acid and hydrogen sulfide gas. A chemist would write the (somewhat confusing) unbalanced equation



to represent this. The aim of this problem is to figure out a ratio of the different chemical compounds necessary for this reaction to take place. The idea is to let variables x_1, x_2, x_3, x_4 represent the number of each compound

$$x_1(B_2S_3) + x_2(H_2O) = x_3(H_3BO_3) + x_4(H_2S) \quad (4.5.1)$$

- (a) Explain how the linear system $Ax = \vec{0}$ with
- $$A = \begin{pmatrix} 2 & 0 & -1 & 0 \\ 3 & 0 & 0 & -1 \\ 0 & 2 & -3 & -2 \\ 0 & 1 & -3 & 0 \end{pmatrix}$$
- describes the linear system (4.5.1) (Hint: the rows are the elements
- B, S, H, O
-)
- (b) Describe the solution set to $Ax = \vec{0}$, then find one ratio of chemical compounds which balances this reaction.
6. True or False. Determine if the following statements are true or false. Justify your answer.
- (a) There is a 2×3 matrix A such that $Ax = b$ has a solution for all vectors $b \in \mathbf{R}^2$.
 - (b) There is a 3×2 matrix A such that $Ax = b$ has a solution for all vectors $b \in \mathbf{R}^3$.
 - (c) If $x = \vec{0}$ is a solution to $Ax = b$ then $b = \vec{0}$.
 - (d) Homogeneous systems are always consistent
 - (e) The matrix $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ has only the trivial solution to $Ax = \vec{0}$

Additional suggested exercises: *Lay*, sections 1.4, 1.5

5 Linear independence

You might be noticing some patterns so far, in the form of vague rules like “a linear system with four variable and three equations has at least one free variable (if it is consistent)”. We’ll eventually turn these into more precise statements; the focus of this section is to make sense of the operative phrase “..at least one..” in the before quote. The key here is to know that the rows of the augmented matrix are independent, we begin with a definition.

5.1 Definition (Linear independence). *A collection x_1, x_2, \dots, x_n of vectors is said to be **linearly independent** if given scalar values c_1, c_2, \dots, c_n such that*

$$c_1x_1 + c_2x_2 + \dots + c_nx_n = 0$$

then $c_1 = c_2 = \dots = c_n = 0$.

A set $\{x_1, x_2, \dots, x_n\}$ of vectors is called **linearly dependent** if it is not linearly independent^[18]. Equivalently, a collection of vectors is linearly dependent if one of the vectors from it can be written as a linear combination of the others.

The relation between linear independence and matrices comes from the following theorem.

5.2 Theorem. *Let $A = \{a_{i,j}\}$ be an $m \times n$ matrix and $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n$ be the column vectors of A . Then $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$ is a linearly independent set of vectors if and only if the system $Ax = \vec{0}$ has only the trivial solution.*

Proof. Suppose that $\vec{a}_1, \dots, \vec{a}_n$ are linearly independent and write $A = (\vec{a}_1 \ \dots \ \vec{a}_n)$. Let $x \in \mathbf{R}^n$ be a vector such that $Ax = \vec{0}$. Expanding the left hand side of this equation, we then have

$$x_1\vec{a}_1 + \dots + x_n\vec{a}_n = \vec{0}$$

which is only possible if $x_1 = \dots = x_n = 0$. Therefore, $x = \vec{0}$.

The reverse argument works exactly the same way: if $A = (\vec{a}_1 \ \dots \ \vec{a}_n)$ such that $Ax = \vec{0}$ only has the trivial solution, then since any linear combination

$$c_1\vec{a}_1 + \dots + c_n\vec{a}_n = \vec{0}$$

can be written as $Ac = \vec{0}$ for $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, the vectors $\vec{a}_1, \dots, \vec{a}_n$ must be linearly independent. □

5.3 Remark. The proof above tells us that the solution set $Ax = \vec{0}$ can be interpreted as the collection of linear dependencies of the columns of A .

^[18]This somewhat clunky way of defining things is unfortunately quite common in mathematics

5.4 Dimension and bases

Linear independence is a necessary tool to understand the dimension of the span of a collection of vectors. Suppose we have a collection of vectors v_1, v_2, \dots, v_k . If this is a linearly independent collection, this means that each vector is a “genuinely new direction” when compared to the span of the remaining vectors. Said differently, this means that $\text{Span}\{v_1, v_2, \dots, v_k\}$ is an k -dimensional “space” (inside whatever \mathbf{R}^n the vectors are in).

Given a span of a collection of vectors, we say that span has **dimension k** if there are at most k linearly independent vectors in the span. Spanning sets of \mathbf{R}^n which consist of linearly independent vectors are called *bases* as in the following definition:

5.5 Definition. A collection $v_1, \dots, v_k \in \mathbf{R}^n$ of vectors is said to be a **basis** for \mathbf{R}^n if the following two conditions are satisfied

- v_1, \dots, v_k are linearly independent
- $\text{Span}\{v_1, \dots, v_k\} = \mathbf{R}^n$.

Note that any basis of \mathbf{R}^n consist of exactly n vectors.

5.6 Exercises

1. Determine if the following sets of vectors are linearly independent or not

$$(a) \left\{ \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\}$$

$$(b) \left\{ \begin{pmatrix} 5 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ -2 \\ 3 \end{pmatrix} \right\}$$

$$(c) \left\{ \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 \\ -2 \end{pmatrix} \right\}$$

$$(d) \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

2. Express a linear *dependence* among the following vectors (Hint: find a nontrivial solution to $Ax = \vec{0}$ for $A = (v_1 \ v_2 \ v_3)$).

$$v_1 = \begin{pmatrix} 4 \\ 1 \\ 6 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -7 \\ 5 \\ 3 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 9 \\ -3 \\ 3 \end{pmatrix}$$

3. Solve for b such that the following vectors are linearly independent, or explain why no such b exists

$$\begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 1 \end{pmatrix}, \begin{pmatrix} b \\ 0 \\ 3 \end{pmatrix}$$

4. Suppose A is an matrix of dimension 4×8 . How many linearly independent columns does A need to have 5 free variables in its solution set for $Ax = \vec{0}$. Is it possible for the solution set of A to have no free variables?
5. Show if v_1, v_2, v_3, v_4 form a linearly independent set of vectors then v_1, v_2, v_3 also forms a linearly independent set of vectors.
6. True or False. Determine if the following statements are true or false. Justify your answer.
 - (a) If A has two linearly dependent rows, then A is equivalent to a matrix with a row of all 0s.
 - (b) Any set of four vectors in \mathbf{R}^3 must be linearly dependent.
 - (c) Any set of three vectors in \mathbf{R}^3 is linearly independent.
 - (d) If $X = \{x, y, z\}$ is a collection of three vectors in \mathbf{R}^5 such that each pair $x, y; y, z;$ and x, z is linearly independent, then X is a linearly independent set.

Additional suggested exercises: *Lay* section 1.7

6 Linear transformations

Here's some notation: Let's write $f: D \rightarrow R$ to denote that f is a function with domain D and range R ^[19]. This notation does not specify the names of any input variables, if we wish to define a function by its action on a generic input $x \in D$ we can write $x \mapsto f(x)$. Note that $f(x)$ is **not** the function, it is the output of the function f at point $x \in D$.

Our main subject of focus now is on functions $\mathbf{R}^n \rightarrow \mathbf{R}^m$. There are *a lot* of such functions, but as far as linear algebra is concerned, we're only concerned with functions that "play nicely" with respect to the vector algebra^[20] from Theorem 3.5. These functions are called *linear transformations* and are defined as follows.

6.1 Definition. A function $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a **linear transformation** if for all vectors $v, w \in \mathbf{R}^n$ and scalars k the following are satisfied:

- $T(v + w) = T(v) + T(w)$
- $T(kv) = kT(v)$

^[19]Note: our use of the term range here is that the range is a set which contains the outputs of f . This is not to be confused with the set of outputs of f .

^[20]Compare this with the definition of a *continuous function* from Calculus as one which satisfies $\lim_{x \rightarrow a} f(x) = f(a)$ for all a in its domain: rewriting a as $\lim_{x \rightarrow a} x$, a continuous function is just one which "commutes" with limits: i.e.,

$$\lim_{x \rightarrow a} f(x) = f\left(\lim_{x \rightarrow a} x\right)$$

Note that these two properties mean that **all** linear transformations must satisfy $T(\vec{0}) = \vec{0}$ ^[21]. Notation-wise, it's customary to sometimes drop the parentheses from a linear transformation and write Tx for $T(x)$.

6.2 Remark. Intuitively, linear transformations are just the functions which “only know” about the vector operations on \mathbf{R}^n . Since the vector structure of \mathbf{R}^n is well understood, this means in turn that linear transformations will be able to be understood (more or less) in complete detail. We start with the association of a matrix to a given linear transformation. First some terminology

6.3 Terminology. Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a linear transformation. We say that T is **one-to-one** (also called **injective**) if given vectors $x, y \in \mathbf{R}^m$ such that $T(x) = T(y)$ then $x = y$. Similarly, T is **onto** (also called **surjective**) if given $b \in \mathbf{R}^m$ there is $x \in \mathbf{R}^n$ such that $T(x) = b$. If a linear transformation is both one-to-one and onto it's called an **isomorphism** (also, **bijection**). Some of these words may be new, but the terms *one-to-one*, etc. should be familiar from previous classes (and also mean the same thing here).

Let us define the **kernel** of a transformation T to be the set of all inputs $x \in \mathbf{R}^n$ such that $T(x) = \vec{0}$. We also need a word for the possible “stuff” to get out of T . We'll call the **image** of T to be the collection of all outputs of T , i.e., $\{T(x) : x \in \mathbf{R}^n\}$. If a transformation T is onto then the image and the range are the same. In general, however, the image is just a subset of the range.

Some propositions:

6.4 Proposition. *A linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is one-to-one if and only if the equation $T(x) = \vec{0}$ only has the trivial solution $x = \vec{0}$.*

Proof. If T is one-to-one, then since any linear transformation satisfies $T(\vec{0}) = \vec{0}$ we're done. Suppose that $T(\text{anything}) = \vec{0}$ only has the trivial solution and that x and y are vectors in \mathbf{R}^n such that $T(x) = T(y)$. Then

$$\vec{0} = T(x) - T(y) = T(x - y)$$

so that $x - y = \vec{0}$. Therefore $x = y$ and so T is one-to-one. \square

6.5 Proposition. *If a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is both one-to-one and onto, then $n = m$.*

Proof. This follows from Theorem 6.10. Suppose that T is represented by an $m \times n$ matrix A . If the columns of A span \mathbf{R}^m then we must have $n \geq m$. If the columns of A are linearly independent then we must have $m \leq n$. So, $m = n$. \square

^[21]The word *linear* here can be a bit confusing if you're trying to tie this to “lines” in the xy -plane like $y = mx + b$. For a function $f(x) = mx + b$ to be a *linear transformation* from \mathbf{R} to \mathbf{R} , it must be that $b = 0$. Therefore we're in the somewhat awkward position of saying that not all linear functions are linear transformations. This is mostly just a matter of context though: as far as linear algebra goes things that are linear offset by a constant are called *affine*.

6.6 Matrix representation of a linear transformation

Let $n \geq 1$ be given and for $i = 1, \dots, n$ write e_i for the vector

$$e_i = \begin{pmatrix} & & 0 \\ & \vdots & \\ 1 & (i\text{-th row}) & \\ & \vdots & \\ & & 0 \end{pmatrix} \quad (6.6.1)$$

i.e., e_i is the vector made up of all 0s except for a 1 in row i . The collection of these vectors $\{e_1, \dots, e_n\}$ is often called the **standard basis**^[22] of \mathbf{R}^n . It's straightforward but worthwhile to convince yourself that $\{e_1, \dots, e_n\}$ is a spanning set of \mathbf{R}^n .

Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a linear transformation and set A to be the matrix with column vectors given by $T(e_1), \dots, T(e_n)$ as follows

$$A = (T(e_1) \ T(e_2) \ \cdots \ T(e_n)) \quad (6.6.2)$$

Then, for any vector $x \in \mathbf{R}^n$, the output vector $T(x)$ is the same as the matrix-vector multiplication Ax . The matrix A from (6.6.2) is a *matrix representation* of the linear transformation T . Since this representation will depend on the choice of basis (more on this soon) for \mathbf{R}^n , let's say that the matrix A defined by 6.6.2 is the **standard matrix representation** of T .

That is to say, understanding linear transformations is the same as understanding the matrices which represent them.

6.7 Example (Identity matrix). Let's denote by I_n the $n \times n$ matrix

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \quad (6.7.1)$$

with 1s along the main diagonal^[23] and 0s everywhere else. This matrix I_n is called the **identity matrix** on \mathbf{R}^n . It's straightforward to check that for any $x \in \mathbf{R}^n$ that $I_n x = x$, therefore I_n is the standard representation of the identity function $\mathbf{R}^n \rightarrow \mathbf{R}^n$.

6.8 Example (Linear transformations $\mathbf{R}^2 \rightarrow \mathbf{R}^2$). Let's consider some linear transformations of the plane by carefully examining their representing matrices. I encourage you to play around with them using a matrix visualization tool (such as [this one](#)).

^[22]See definition 5.5

^[23]Top left to bottom right

- Dilation. Let $k > 0$ be some positive number, any matrix of the form is a **dilation** of space

$$A(k) = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$$

- Reflection. The matrices F_x and F_y defined by

$$F_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad F_y = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

represent the **reflection** over the x - and y -axes respectively. Similarly, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is a **reflection** over the line $y = x$. This can be seen in that

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

- Rotation. For θ a real number, any matrix of the form

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

is a **rotation** (about the origin) of the xy -plane counterclockwise by angle θ .

- Shear. For k a nonzero real number, any matrix of the form

$$S_x(k) = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad S_y(k) = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$$

is called a **shear** matrix.

6.9 Relating a transformation to its standard matrix representation

We finish this section by relating some features of a linear transformation T to features of a matrix A which represents it. The idea is really that transformations are the “abstract” notion that we want to work it; an associated matrix is the concrete instantiation of it which we can “do stuff” to (to prove things about transformations).

6.10 Theorem. *Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a linear transformation and A the standard matrix associated to T . Then:*

- T is onto if and only if the columns of A span \mathbf{R}^m*
- T is one-to-one if and only if the columns of A are linearly independent.*

Proof. Write A for the standard matrix representation of T . Note that T is onto if and only if $Ax = b$ has a solution for all $b \in \mathbf{R}^m$. But this is to say exactly that b is in the span of the columns of A , and so the columns of A span \mathbf{R}^m (Theorem 3.12).

T is one-to-one if and only if the nullspace of A contains only the zero vector (Proposition 6.4). But this is exactly to say that the columns of A are linearly independent (Theorem 5.2). \square

6.11 Exercises

1. Let $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be a linear transformation such that

$$T \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \text{and} \quad T \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

Determine the values of the following:

(a) $T \left(2 \begin{pmatrix} 5 \\ 2 \end{pmatrix} - 4 \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right)$

(b) $T \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ Hint: write $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as a linear combination of $\begin{pmatrix} 5 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$

2. Write a matrix representation for a linear transformation $T: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ such that

$$Te_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad Te_2 = \begin{pmatrix} 5 \\ -2 \end{pmatrix}, \quad Te_3 = \begin{pmatrix} -7 \\ 4 \end{pmatrix}$$

then describe the set of all vectors in $x \in \mathbf{R}^3$ such that $T(x) = \vec{0}$.

3. Explain why $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ given by

$$T(x, y) = (2x + y - 1, y - x)$$

is not a linear transformation

4. Write a matrix A such that

$$A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - 3x_3 \\ 4x_1 \\ x_1 - x_2 + x_3 \end{pmatrix}$$

or explain why no such matrix exists.

5. Prove Theorem 6.10

6. Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be one-to-one and suppose that $\{u, v, w\}$ is a linearly independent set of vectors in \mathbf{R}^n . Is $\{Tu, Tv, Tw\}$ linearly independent in \mathbf{R}^m ? Explain why or why not. Is this still true if T is not one-to-one?

7. Describe geometrically the linear transformation \mathbf{R}^2 given by $x \mapsto Ax$ for

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

Do the same for the linear transformation of \mathbf{R}^3 given by $x \mapsto Bx$ for

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

8. True or False. Determine if the following statements are true or false. Justify your answer.

- (a) Let A be an $m \times n$ matrix and define a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ by $x \mapsto Ax$. Then the image of T is the span of column vectors of A .
 - (b) A 17×21 matrix defines a linear transformation $\mathbf{R}^{17} \rightarrow \mathbf{R}^{21}$
 - (c) Any rotation matrix $R_\theta: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is both one-to-one and onto.
 - (d) If A is a 2×3 matrix then the linear transformation $x \mapsto Ax$ is never onto.
 - (e) If A is a 2×3 matrix then the linear transformation $x \mapsto Ax$ is never one-to-one.
9. Let $R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ be a rotation matrix. Show that R_θ is both one to one and onto. Is the same true of the reflection matrices F and shear matrices S_x, S_y from example 6.8?

Additional suggested exercises: Lay sections 1.8, 1.9

7 The algebra of matrices

If matrices are representatives of linear transformations (which are just types of functions), then it reasons that we should be able to do “function operations” (i.e., composition) with linear transformations on the level of the matrices. In this section we will define exactly that: the algebra of matrices, and how it relates to algebraic manipulations of functions that we know and love.

If $A = \{a_{i,j}\}$ and $B = \{b_{i,j}\}$ are $m \times n$ matrices and $c \in \mathbf{R}$ some scalar, then there are $m \times n$ matrices (i) $A + B$ defined by adding the entries of A and B component-wise, and (ii) cA defined by scalar multiplication by c on the entires of A . In formulas:

$$A + B = \{a_{i,j} + b_{i,j}\} \quad cA = \{ca_{i,j}\} \quad (7.0.1)$$

(Note: that this means $m \times n$ themselves are elements of the vector space \mathbf{R}^{mn}). *Multiplication* of matrices is defined in certain circumstances, as demonstrated in the next section.

7.1 Matrix multiplication

Suppose that $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ and $S: \mathbf{R}^m \rightarrow \mathbf{R}^k$ are linear transformations. There is then a linear transformation $S \circ T: \mathbf{R}^n \rightarrow \mathbf{R}^k$ defined by the composite

$$\begin{array}{ccc} \mathbf{R}^n & \xrightarrow{T} & \mathbf{R}^m \\ & \searrow S \circ T & \downarrow S \\ & & \mathbf{R}^k \end{array}$$

That is, $(S \circ T)(x) = S(T(x))$. Let’s just write ST for the composite $S \circ T$. Suppose that T and S have standard representing matrices A and B respectively. Then

$$(ST)(x) = S(T(x)) = B(Ax)$$

(note that Ax is a vector in \mathbf{R}^m) which suggests there should be an $n \times k$ matrix “ BA ” such that $B(Ax) = (BA)x$. Indeed, this is the “multiplication of matrices” as defined as follows:

7.2 Definition. If A is an $m \times n$ matrix with columns \vec{a}_i and B an $k \times m$ matrix then there is an $k \times n$ matrix BA defined as

$$BA = (B\vec{a}_1 \ B\vec{a}_2 \ \cdots \ B\vec{a}_n) \quad (7.2.1)$$

This gives us an easy definition to write down, but it's not really how matrix multiplication is most often thought of. In “pictures”, BA is the matrix defined by the following

$$\underbrace{\begin{pmatrix} & \text{column } j \\ & \vdots \\ \text{row } i & \cdots & \sum_{\ell=1}^m b_{i,\ell} a_{\ell,j} & \cdots \\ & \vdots \end{pmatrix}}_{BA} = \underbrace{\begin{pmatrix} \vdots & \vdots & & \vdots \\ b_{i,1} & b_{i,2} & \cdots & b_{i,m} \\ \vdots & \vdots & & \vdots \\ \cdots & \cdots & & \cdots \\ \cdots & a_{m,j} & \cdots \end{pmatrix}}_B \underbrace{\begin{pmatrix} \cdots & a_{1,j} & \cdots \\ \cdots & a_{2,j} & \cdots \\ \vdots & & \vdots \\ \cdots & a_{m,j} & \cdots \end{pmatrix}}_A \quad (7.2.2)$$

Or said even differently still BA is given componentwise by the sum

$$(BA)_{i,j} = a_{1,j}b_{j,1} + a_{2,j}b_{j,2} + \cdots + a_{m,j}b_{j,m} \quad (7.2.3)$$

7.3 Remark. it's worth convincing yourself that BA is the matrix which represents the linear transformation ST , and that this definition of matrix multiplication is really nothing other than composition of functions. Note that in the same way that function composition is rarely **commutative** (i.e., $ST \neq TS$), the same then applies for matrices.

In many cases only one order of multiplication will be defined, but even if both AB and BA are defined then it's still rarely the case that $AB = BA$.

7.4 Matrix algebra

So far we've defined addition, scalar multiplication, and “composition multiplication” of matrices. It remains to show how these operations relate to each other. I leave it to you to convince yourself of the following:

7.5 Theorem (Laws of matrix algebra). Suppose that A is an $m \times n$ matrix, $k \in \mathbf{R}$ a scalar, and B, C are matrices such that the following matrix products are all defined. Then:

- (a) $C(BA) = (CB)A$
- (b) $C(B + A) = CB + CA$
- (c) $(C + B)A = CA + BA$
- (d) $k(BA) = (kB)A = B(kA)$
- (e) $I_m A = A I_n$

7.6 Terminology. The first law from Theorem 7.5 tells us that matrix multiplication is **associative**, i.e., that a three-fold product of matrices such as CBA is defined unambiguously. Properties b and c tell us that matrix multiplication is **distributive**. The last law tells us that the identity matrices (of appropriate size) are the **identities** of matrix multiplication. Note if A is a **square matrix** (i.e., $n \times n$ for some n), this means that (positive integer) powers of A can be defined as

$$A^k = \underbrace{AA \cdots A}_{k \text{ times}} \quad (7.6.1)$$

where by convention we take $A^0 = I_n$ ^[24]. In general it is not the case that $AB = BA$ (even when both defined). If A, B are square matrices such that $AB = BA$ then we say that A and B **commute**.

7.7 Transposition

Let $A = \{a_{i,j}\}$ be an $m \times n$ matrix. The **transpose** of A is defined to be the $n \times m$ matrix A^T given by exchanging the rows and columns of A : that is,

$$A^T = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & & \vdots \\ a_{1,m} & a_{2,m} & \cdots & a_{n,m} \end{pmatrix} \quad (7.7.1)$$

or in terms of coefficients $A_{i,j}^T = A_{j,i}$. If $x \in \mathbf{R}^n$ is a (column) vector $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, then its transpose is the row vector $x^T = (x_1 \ x_2 \ \cdots \ x_n)$.

Transpose may seem silly at first, but it's useful in describing orthogonality conditions as we'll see soon enough. More vaguely, the transpose allows you to interchange the roles of "rows" and "columns" in most all of what we've learned so far. For now, we can think of it as an additional algebraic tool on the set of matrices. I leave it to you to convince yourself of the following properties.

7.8 Theorem. Let A, B be matrices such that the product BA is defined. Then:

- (a) $(A^T)^T = A$
- (b) $(A + B)^T = A^T + B^T$
- (c) $(BA)^T = A^T B^T$

7.9 Exercises

1. Prove Theorems 7.5 and 7.8

^[24]In fancy words, the collection of $n \times n$ matrices is an *Algebra* in the "capital A"-algebra sense

2. Let T and S be transformations $\mathbf{R}^2 \rightarrow \mathbf{R}^2$ given by:

$$T(x, y) = (2x + y, -x + 3y), \quad S(x, y) = (5x - 2y, x + y)$$

- (a) Find constants a, b, c, d such that $(S \circ T)(x, y) = (ax + by, cx + dy)$.
(b) Show that these constants are the same a, b, c, d such that

$$\begin{pmatrix} 5 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

(Note: This gives an example of the general fact that matrix multiplication encodes composition of linear transformations.)

3. Let A, B, C be the following matrices

$$A = \begin{pmatrix} 1 & 5 & 2 \\ 0 & 4 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 6 \\ -1 & 4 \\ 9 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ -1 & 4 \end{pmatrix}$$

For each of the following, determine if the matrix product is defined, and if so, evaluate the product

- (a) AB
(b) BA
(c) A^2
(d) ACB
(e) BCA
(f) AC^2
4. Let $A = \begin{pmatrix} 2 & 5 \\ -1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 4 & -5 \\ 3 & k \end{pmatrix}$. Determine if there is any value of k such that $AB = BA$.
5. Show that if $x, y \in \mathbf{R}^n$ are column vectors that $x^T y$ is the usual *dot product* of vectors $x \cdot y$ given by

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

What type of object is xy^T ?

6. Collections of matrices are a canonical way of expressing types of algebraic structures more complicated than the real numbers. You may be aware of the *complex numbers* \mathbf{C} , the collection of numbers $a + bi$ where $a, b \in \mathbf{R}$ and i is a “new” number which satisfies $i^2 = -1$. A complex number $a + bi$ is encoded by a 2×2 matrix as

$$a + bi = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

(a) Use that $i^2 = -1$ to write $(a + bi)(c + di)$ as $x + iy$ where $x, y \in \mathbf{R}$

(b) Show that these terms x, y are the terms x, y such that

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

(Congratulations, this is an example of an *Algebra*)

(c) Show that multiplication of complex numbers is commutative, i.e., that

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

(d) The *complex conjugate* of $a + bi$ is the complex number $a - bi$. Show that $a - bi$ is encoded by the matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}^T$$

7. An $n \times n$ matrix is called *nilpotent* if there is an integer $k \geq 0$ such that A^k is the matrix of all 0s.

(a) Show that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ are nilpotent.

(b) Is $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ nilpotent? What about $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$? Explain why or why not

8. Examples of (non)commutativity of 2×2 matrices.

(a) For $k \in \mathbf{R}$, let $S_x(k) = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ be the shear matrix along the x -axis. Show that the composition of shear matrices $S_x(k)S_x(\ell)$ is another shear matrix given by

$$S_x(k)S_x(\ell) = S_x(k + \ell)$$

Conclude that shear matrices (along the same axis) commute.

(b) For $\theta \in \mathbf{R}$, let $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. Show that

$$R_\theta R_\phi = R_{\theta+\phi}$$

and that therefore rotation matrices commute.

(c) Recall that $F = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ is a reflection about the y -axis. For R_θ the rotation matrix from part b., show that

$$R_\theta F = FR_{-\theta}$$

and so these two matrices do not commute.

9. Let $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ be the “double the length” dilation matrix. Show that $AB = BA$ for any 2×2 matrix B .
10. For a A an $n \times n$ matrix, determine if $A^2 - I_n = (A + I_n)(A - I_n)$. Then, determine if $A^2 - B^2 = (A - B)(A + B)$ where B is also an $n \times n$ matrix.

Additional suggested exercises: *Lay* section 2.1

8 Invertible matrices

A quick bit of notation: in the previous section we wrote I_n to denote the $n \times n$ matrix which represents the identity transformation $\mathbf{R}^n \rightarrow \mathbf{R}^n$. In general, we will drop the subscript n and just write I for this matrix, reincluding the n when only it is necessary to distinguish the dimension from some other amount.

8.1 Definition. A matrix A is called **invertible** if there is another matrix B such that $AB = BA = I$.

In notation we write $B = A^{-1}$ and call A^{-1} the **inverse matrix** of A . The use of the definite article *the* here implies the correct statement that the inverse of a matrix is unique (if it exists). If A has an inverse matrix we call A an **invertible** matrix.

8.2 Relation to linear transformations

If A is the standard matrix representation of a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ then A^{-1} represents the inverse function of T . In general, a function has an inverse if and only if it is both one-to-one and onto. Using the characterizations from Theorem 6.10, we have the following characterization of an invertible matrix A

8.3 Proposition. Let A be an $m \times n$ matrix. Then, A is invertible if and only if both of the following conditions are met^[25]

- The columns of A span \mathbf{R}^m
- The columns of A are linearly independent.

That is to say, the n columns of A are a linearly independent set which span \mathbf{R}^m , and so $n = m$. Therefore a necessary (but not sufficient) condition for A to be invertible is that A be a square matrix.

8.4 Technique for computing the inverse of a matrix

We will learn later a slightly easier to use technique for determining if a matrix is invertible^[26]. However, finding the inverse of a matrix only requires the row operations that we know and love so far.

^[25]This is actually a bit misleading: n vectors in \mathbf{R}^n span \mathbf{R}^n if and only if they are independent, so these two conditions are equivalent

^[26]Using the determinant, though this is not necessarily constructive of the inverse matrix

8.5 Proposition. Suppose that A is an $n \times n$ matrix such that the $n \times 2n$ augmented matrix

$$(A \mid I)$$

is equivalent by row operations to an augmented matrix of the form

$$(I \mid B)$$

for some $n \times n$ matrix B . Then $B = A^{-1}$.

Moreover, if row operations applied to $(A \mid I)$ yield an inconsistent system, then A is not invertible.

8.6 Theorem (Inverse of a 2×2 matrix). Suppose that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2×2 matrix such that $ad - bc \neq 0$. Then the inverse of A exists and is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad (8.6.1)$$

8.7 Remark. The term $ad - bc$ for such a matrix A is called the **determinant**. We will spend more time with this later, but for now think of it as follows: If $ad - bc = 0$ then $ad = bc$ and therefore (so long as we don't divide by 0), $a/c = d/b$. That is to say the ratios of the components of $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ are equal, and therefore these vectors are parallel. But if the columns of A are parallel, then they are not linearly independent, and so A is not invertible.

If $ad - bc \neq 0$ then we avoid this problem entirely. Checking that the proposed form of A^{-1} is correct is just a matter of a quick computation, i.e., that

$$A^{-1}A = AA^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

8.8 Adjugates

The matrix $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ is called the **adjugate** of A . The adjugate, and an inverse formula akin to (8.6.1) does exist for larger $n \times n$ matrices, more information can be found in Appendix B. Though it can be useful for small matrices, the general expression for A^{-1} in terms of the adjugate is actually *less* computationally efficient than using row operations on A .

8.9 Proposition (Inverse of the transpose matrix). If A is an invertible matrix, then so is A^T . Moreover, $(A^T)^{-1} = (A^{-1})^T$

8.10 A very large list of things which are equivalent

8.11 Theorem ((Part of) the invertible matrix theorem). Let A be an $n \times n$ matrix. The following are equivalent

- (a) A is invertible

- (b) A is equivalent by row operations to the identity matrix
- (c) A has n leading 1s in its reduced echelon form
- (d) The equation $Ax = \vec{0}$ has only the trivial solution
- (e) The nullspace of A is $\{\vec{0}\}$.
- (f) The columns of A are linearly independent
- (g) The linear transformation $x \mapsto Ax$ is one-to-one
- (h) The equation $Ax = b$ has a solution for each $b \in \mathbf{R}^n$
- (i) The columns of A span \mathbf{R}^n
- (j) The linear transformation $x \mapsto Ax$ is onto
- (k) There is an $n \times n$ matrix C such that $CA = I$
- (l) There is an $n \times n$ matrix D such that $AD = I$
- (m) A^T is invertible

You might scoff at a list of m things which are all equivalent, but there's even more to add to this list later in the course (a more complete list can be found in Appendix A). This list is purposefully obtuse in the number of things in its length: the invertible matrix theorem really is just the collection of different ways of interpreting the following statement for linear transformations $\mathbf{R}^n \rightarrow \mathbf{R}^n$

8.12 Fact. *A linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ is one-to-one if and only if it is onto.*

8.13 Exercises

1. Prove Proposition 8.9
2. Determine if A invertible and, if so, describe A^{-1}
 - (a) $\begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix}$
 - (b) $\begin{pmatrix} 5 & 10 \\ 4 & 7 \end{pmatrix}$
 - (c) $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$
 - (d) $\begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix}$
3. Show that if A, B are invertible matrices then $(AB)^{-1} = B^{-1}A^{-1}$ and $(A^{-1})^{-1} = A$

4. Show that if A, B are matrices and C is invertible then $AC = BC$ implies $A = B$. Is this still true if C is not invertible?
5. Let $A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$. Construct a 4×2 matrix B such that $AB = I_2$, or explain why such a matrix is not possible.
6. Suppose that A is an invertible $n \times n$ matrix and that $b \in \mathbf{R}^n$. Show that $Ax = b$ has a unique solution given by $x = A^{-1}b$.
7. Suppose that A is an $n \times n$ matrix such that there is another $n \times n$ matrix C with $CA = I$. Show that A is invertible. Show as well that A is invertible if there is an $n \times n$ matrix D with $AD = I$.
8. *Recall that a complex number $a + bi$ is represented by the matrix $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.
- (a) Show that A is invertible if and only if $a \neq 0$ and $b \neq 0$.
 - (b) Determine the inverse matrix of A
 - (c) If a, b are real numbers not both equal to 0, determine a complex number $x + yi$ such that $(a + bi)(x + yi) = 1$.
- Congratulations, you've just described one of the three nontrivial division algebras over \mathbf{R} .
9. *Applying a row operation to a matrix A is the same as multiplying A by a certain invertible matrix. These are called **elementary matrices** and they generate the collection of invertible matrices.
- (a) Let $E_{i,j}(k)$ be the sum of the identity matrix I and the $n \times n$ matrix with nonzero entry k in position i, j . Show that the matrix product $A' = E_{i,j}A$ is the result of applying “add k times row j to row i ” in A .
 - (b) For $k \neq 0$, let $R(i, k)$ be the $n \times n$ identity matrix, with the 1 in the i -th row replaced by k . Show that rescaling the i -th row of a matrix A is the same as the matrix product $R(i, k)A$.
 - (c) Let $S(i, j)$ be the $n \times n$ identity matrix with rows i and j exchanged. Show that exchanging rows i and j in a matrix A is the same as the matrix product $S(i, j)A$.
 - (d) Show that all of these matrices $E_{i,j}(k), R(i, k), S(i, j)$ are invertible.

Since a matrix A is invertible if and only if it can be row reduced to the identity matrix, this then tells you that A can be “factored” into a large product of such matrices, and so these elementary matrices generate the collection of all invertible matrices.

Additional suggested exercises: *Lay* sections 2.2, 2.3

9 The determinant

While Proposition 8.5 gives us a tool for computing the inverse of a matrix, it's often tempting to just know *if* a matrix is invertible, without necessarily needing to know explicitly its inverse. We've been necessary and sufficient conditions for invertibility, but all the conditions we've met ultimately rely on row reduction as a computational tool so far. It's time we learn a *new* thing.

9.1 Determinants of 2×2 matrices

That thing is the **determinant**. It will take a bit of setting up to describe the determinant of an $n \times n$ matrix, but we've already come across it in an example so far:

Recall from Theorem 8.6 that a 2×2 matrix

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

is invertible if and only if $ad - bc = 0$. We can view this a function, which we will call “det”

$$\det: \{2 \times 2 \text{ matrices}\} \rightarrow \mathbf{R} \quad (9.1.1)$$

given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \quad (9.1.2)$$

such that $\det A \neq 0$ if and only if A is invertible. Said differently, $\det A$ is a number which represents (among other things) the “invertibility” of A .

Where does this expression come from? There is a deep geometric reasoning behind the determinant and (signed) areas in \mathbf{R}^n . Let's take a closer look for a 2×2 matrix. Consider vectors $u = \begin{pmatrix} a \\ c \end{pmatrix}$ and $v = \begin{pmatrix} b \\ d \end{pmatrix}$ in \mathbf{R}^2 . These vectors form a parallelogram in \mathbf{R}^2 with one vertex at the origin. The (signed) area of this parallelogram is then given by the determinant of A .

In Figure 9.1.3, the area of the red parallelogram is given by

$$(a + b)(c + d) - 2bc - ac - bd = ad - bc$$

The *sign* of $\det A$ is a bit mysterious, but what it does is tell us whether the area created by u, v has been mirrored or not^[27]. Specifically, if $\det A$ is negative, then the volume has been mirrored, and if $\det A > 0$ then it has not. Taking absolute values $|\det A|$ tells you *just* the area of this parallelogram^[28].

^[27]With regards to our section about transformations of the plane, this means has the region been reflected in such a way that text is now backwards

^[28]If you've taken Calc 3 this interpretation of the determinant should be familiar: Given a “change of variables” function $r = (x(u, v), y(u, v))$ the appropriate rescaling function for the double integral is the

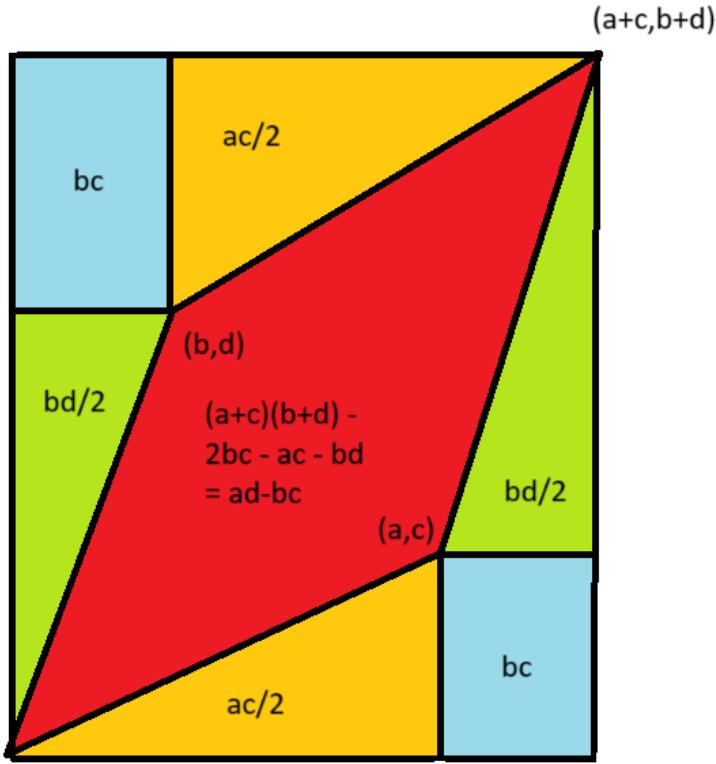


Figure 9.1.3: Visual proof of the area of a parallelogram made by vectors $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$

9.2 Determinants of 3×3 matrices

You may recall the following formula for computing the determinant of a 3×3 matrix^[29]

$$\begin{aligned} \det \begin{pmatrix} a & b & c \\ d & e & f \\ h & i & j \end{pmatrix} &= a \det \begin{pmatrix} e & f \\ i & j \end{pmatrix} - b \det \begin{pmatrix} d & f \\ h & j \end{pmatrix} + c \det \begin{pmatrix} d & e \\ h & i \end{pmatrix} \\ &= a(ej - fi) - b(dj - fh) + c(di - eh) \\ &= aej - afi + bfh - bdj + cdi - ceh \end{aligned} \quad (9.2.1)$$

This funny looking expression tells us the same information for this 3×3 matrix as it does for a 2×2 matrix. More specifically, (9.2.1) computes the signed volume of the formed by

(absolute value of the determinant of the) Jacobian matrix

$$J = \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix}$$

$|\det J|$ gives the amount that area has been skewed from u, v coordinates into x, y coordinate—for instance, the polar coordinate transformation $x(r, \theta) = r \cos \theta, y(r, \theta) = r \sin \theta$ has Jacobian determinant of r

^[29]Perhaps from one interpretation of the calculation of the cross product of vectors $x, y \in \mathbf{R}^3$

the column vectors

$$u = \begin{pmatrix} a \\ d \\ h \end{pmatrix}, v = \begin{pmatrix} b \\ e \\ i \end{pmatrix}, w = \begin{pmatrix} c \\ f \\ j \end{pmatrix}$$

in \mathbf{R}^3 . If this volume is not 0, then the vectors u, v, w must be linearly independent, and therefore the matrix A represents a one-to-one function. Moreover, three linearly independent vectors must span \mathbf{R}^3 , therefore the transformation represented by A is also onto, and so A must be invertible.

9.3 Terminology. There's a funny word lurking in the previous example—**parallelepiped**. This is a real word, meaning “sheared box of dimension n ”. More specifically, a parallelepiped is the n -dimensional region created by “closing up” the box formed by n vectors in \mathbf{R}^n . A specific example of a parallelepiped in \mathbf{R}^3 is given in Figure 9.3.1.

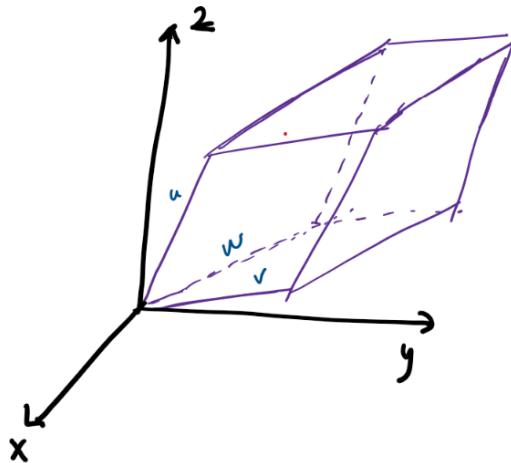


Figure 9.3.1: The parallelepiped in \mathbf{R}^3 formed by vectors u, v, w

9.4 Determinants in general

Though there are some speedier algorithms for special cases, the most typical construction of the determinant of an $n \times n$ matrix is by an inductive process. That is to say, you can compute the determinant of an $n \times n$ in terms of determinants of its $(n - 1) \times (n - 1)$ submatrices, whose determinants are in turn computed by the determinants of their $(n - 2) \times (n - 2)$ matrices and so on^[30]. Eventually you get down to the 2×2 submatrices, and just use equation (9.1.2).

Let's start with the following geometric *definition* of the determinant.

9.5 Definition. Let A be an $n \times n$ matrix and a_1, \dots, a_n be the column vectors of A . Then $\det A$ is the real number equal to the signed hypervolume of the parallelopiped formed by vectors $a_1, \dots, a_n \in \mathbf{R}^n$.

[30]If you care about such things, the number of individual calculations needed to take the determinant is on the order of n^3 for an $n \times n$ matrix, the same as the number of calculations needed for row reduction to reduced echelon form.

9.6 Remark. \det then defines a function

$$\det: \{n \times n \text{ matrices}\} \rightarrow \mathbf{R} \quad (9.6.1)$$

You may also see people write $|A|$ for the determinant of A . Since these kind of vertical bars are already vastly overused, we'll stick with \det .

Note that this definition doesn't necessarily give us the means to compute the determinant just yet. That comes into play in the following section. I leave it to you to convince yourself of the following:

9.7 Theorem. *Let A be an $n \times n$ matrix. Then A is invertible if and only if $\det A \neq 0$*

9.8 Remark. This is another thing to add to the list from Theorem 8.11.

9.9 Computing the determinant

The most common algorithm for computing the determinant of a matrix is called *cofactor expansion*. We need some definitions:

9.10 Definition. *For an $n \times n$ matrix A , let $A_{i,j}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j from A . This matrix $A_{i,j}$ is called the **cofactor** to the entry $a_{i,j}$ in A .*

Let's let S be the $n \times n$ “matrix of signs” as follows

$$S = \begin{pmatrix} + & - & + & - & \cdots \\ - & + & + & \ddots \\ + & - & \ddots \\ - & \ddots \\ \vdots \end{pmatrix} \quad (9.10.1)$$

Note that the i, j entry of S is just $(-1)^{i+j}$. The **cofactor expansion** of $\det A$ along row^[31] i is then the sum

$$\sum_{j=1}^n (-1)^{i+j} a_{i,j} \det A_{i,j} \quad (9.10.2)$$

9.11 Theorem. *For A an $n \times n$ matrix, $\det A$ is computed by cofactor expansion (9.10.2) along any row of A .*

A good strategy for computing the determinant of a big matrix A is to use cofactor expansion along the row which contains the most zeros.

Note that if $n = 3$ this exactly the computation from (9.2.1). Moreover, computing the determinant of a matrix (in this way at least) is an inductive process: you repeatedly apply

^[31]You can also calculate the determinant by cofactor expansion along a *column* of A

cofactor expansion until you reduce your calculation to 2×2 matrices, then use equation (9.1.2). There are other ways of calculating the determinant, but this is the simplest to describe and needs the least additional terminology^[32].

9.12 Properties of the determinant

We've now introduced a new tool, the determinant, to our world of matrices. It remains to see how this interacts with the operations on matrices we've already seen. A useful proposition is the following:

9.13 Proposition. *Let A, B be $n \times n$ matrices. Then the following hold:*

- (a) $\det(AB) = \det(A)\det(B)$
- (b) $\det(A^T) = \det A$
- (c) *If A is invertible, $\det(A^{-1}) = 1/\det A$*

“Proof”. Part a. is proved by showing that elementary row operations are (i) represented by matrices and (ii) have the multiplicative property on determinants. This is what is shown in Proposition 9.14, to really show this rigorously requires a technique called *mathematical induction*. Part b. follows in that cofactor expansion along a *column* also calculates the determinant (you can check this). For part c., note that $AA^{-1} = I$. Using part a we have

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

so that $\det(A^{-1}) = 1/\det A$. □

In general, you do not expect that the determinant should respect addition of matrices, that is:

$$\det(A + B) \neq \det A + \det B$$

A good thing to check is how elementary row operations interact with the determinant. I leave it to you to check the following (there are some exercises that show this in special cases):

9.14 Proposition (Row operations and the determinant). *Let A be a square matrix. Then the following hold:*

- (a) *If B is obtained by multiplying a row of A by k then $\det B = k\det A$*
- (b) *If B is obtained by adding a row to another row within A then $\det B = \det A$*
- (c) *If B is obtained by permuting ℓ rows of A then $\det B = (-1)^\ell \det A$.*

^[32]One such way to calculate the determinant is as a sum over the elements of the n -th symmetric group (the collection of all permutation of the set $\{1, 2, 3, \dots, n\}$), another (much more sophisticated way) is as an alternating n form, they all do the same thing.

9.15 Exercises

1. Compute the determinant of the following matrices

$$(a) \begin{pmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{pmatrix}$$

$$(b) \begin{pmatrix} 1 & -2 & 4 & 2 \\ 0 & 0 & 3 & 0 \\ 2 & -4 & -3 & 5 \\ 2 & 0 & 3 & 5 \end{pmatrix}$$

$$(c) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 7 & -2 & 0 & 0 \\ 2 & 6 & 3 & 0 \\ 3 & -8 & 4 & -3 \end{pmatrix}$$

$$(d) \begin{pmatrix} 6 & 0 & 2 & 4 & 0 \\ 9 & 0 & -4 & 1 & 0 \\ 8 & -5 & 6 & 7 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 3 & 2 & 0 \end{pmatrix}$$

2. An (upper^[33]) **triangular matrix** is any $n \times n$ matrix of the form

$$A = \begin{pmatrix} a_{1,1} & * & * & \cdots & * \\ 0 & a_{2,2} & * & \cdots & * \\ 0 & 0 & a_{3,3} & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n,n} \end{pmatrix}$$

where the diagonal entries $a_{i,i} \neq 0$, 0s below the main diagonal, and the entries marked * are any real number. Show that the determinant of a triangular matrix is given by

$$\det A = a_{1,1}a_{2,2} \cdots a_{n,n}$$

Are triangular matrices invertible?

3. Let $R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$. Show that $\det R_\theta = 1$.
4. Let $B = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix}$, compute $\det(B^4)$.
5. Show that $\det(kA) = k^n \det A$ for A an $n \times n$ matrix and k a scalar. Is \det a linear transformation $\mathbf{R}^{n^2} \rightarrow \mathbf{R}$?

^[33]There are also lower triangular matrices, a matrix is *lower triangular* if its transpose is upper triangular.

6. Show that $\det(A + B) \neq \det A + \det B$ by producing a pair of matrices for which the equation does not hold.
7. A matrix A is called **symmetric** if $A^T A = I$. Show that $\det A = \pm 1$ for any symmetric matrix A .
8. Compute the determinants of the following 3×3 matrices in terms of k

(a) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

9. *True or False.* Determine if the following statements are true or false. Justify your answer

(a) If $A = (a_1 \cdots a_n)$ such that the collection of vectors $a_1, \dots, a_n \in \mathbf{R}^n$ is linearly dependent, then $\det A = 0$

(b) If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that a, b, c, d are all real numbers in the interval $[-1, 1]$, then it is not possible that $\det A = 3$.

(c) If A is an $n \times n$ matrix such that $\det(A^3) = 0$ then A is not invertible

(d) For any invertible matrix P , $\det(PAP^{-1}) = \det(A)$.

(e) If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $a, b, c, d \geq 0$ then $\det A \geq 0$.

10. *The notation $\text{GL}(n)$ is often used to denote the set^[34] of $n \times n$ invertible matrices (with real coefficients), and $\text{SL}(n)$ is used to denote the subset of such matrices $A \in \text{GL}(n)$ with $\det A = 1$. Show that if $A, B \in \text{SL}(n)$ then $A^{-1} \in \text{SL}(n)$ and $AB \in \text{SL}(n)$.

Some terminology: “GL” means *general linear*, and “SL” means *special linear*.

11. *Recall the matrix representations of the elementary row operations from

(a) Show that $\det E_{i,j} = 1$, $\det R(i, k) = k$ and $\det S(i, j) = (-1)^{i+j}$

^[34]It's actually a *group*, actually it's a *Lie group*

- (b) Conclude that an elementary row operation applied to A does not affect whether or not A is invertible
- (c) If the rows of A are linearly independent then $\det(A) \neq 0$.

Additional suggested exercises: *Lay* sections 3.1, 3.2

10 Vector spaces

So far, we've been using our intuition of vectors as "arrows" in n -dimensional (Euclidean) space. However, math is built on abstraction, and linear algebra is a great example of a subject where the tools we build apply in a much more general setting.

10.1 Definition. A *vector space*^[35] is a set X whose elements are things called *vectors* and which act like vectors that we've been working with so far. That is to say, the following axioms are all satisfied

- (a) If $x, y \in X$ then there is a vector $x + y \in X$
- (b) For all $x, y, z \in X$; $x + y = y + x$ and $(x + y) + z = x + (y + z)$
- (c) If $x \in X$, $k \in \mathbf{R}$ a scalar there is a vector $kx \in X$
- (d) There is a vector $\vec{0} \in X$ such that $x + \vec{0} = \vec{0} + x = x$ for all $x \in X$
- (e) Given $x \in X$ there is a vector $-x \in X$ such that $x + (-x) = \vec{0}$
- (f) For any vectors $x, y \in X$, scalars $k \in \mathbf{R}$; $k(x + y) = kx + ky$
- (g) For any vector $x \in X$ and scalar $k, \ell \in \mathbf{R}$; $(k + \ell)x = kx + \ell x$ and $(k\ell)x = k(\ell x)$
- (h) For any vector $x \in X$, $0x = \vec{0}$.

10.2 Example. Examples are always necessary:

- The space $\mathbf{0}$ consisting of a single vector $\vec{0}$ (the zero vector) is a vector space. We call this the **trivial vector space**. You could also define \mathbf{R}^0 to be this space.
- An assuring fact is that the sets \mathbf{R}^n for $n \geq 1$ we've been working with so far are vector spaces. You can think of these as the "geometric" vectors, in that elements in \mathbf{R}^n can (but don't have to) be interpreted as arrows in n -dimensional space.
- Let $C(\mathbf{R})$ denote the set of continuous functions $f: \mathbf{R} \rightarrow \mathbf{R}$. Then $C(\mathbf{R})$ is a vector space, with operations defined as

$$(f + g)(x) = f(x) + g(x), \quad (kf)(x) = kf(x)$$

^[35]In this class, all of our vector spaces are over \mathbf{R} , but more correct is to say "vector space over \mathbf{R} " here. This means the scalars come from the *field* \mathbf{R} . Sometimes it's necessary to consider vector spaces over \mathbf{C} , the field of complex numbers, \mathbf{Q} the field of rational numbers, or *finite fields* like the integers modulo a prime number p

for $f, g \in C(\mathbf{R})$ and k a scalar. The difference with this function space is that it is not a finite-dimensional vector space, therefore understanding elements of this vector space is more nuanced^[36].

- Let $M(m, n)$ be the set of $m \times n$ matrices. Then $M(m, n)$ is a vector space with operations defined as in (7.0.1). $M(m, n)$ is an mn dimensional vector space, a choice of basis would be the collection of all matrices with just one entry of 1, and all other entries 0^[37]. Note that as a set $M(m, n)$ is just \mathbf{R}^{mn} , the difference is in how we choose to interpret such objects.

10.3 Subspaces and bases

We've used words like *span* and *spanning set* a lot so far. More general is to talk about subspaces and bases.

10.4 Definition. Given a vector space X a **subspace** is a subset $V \subset X$ such that:

- (a) If $u, v \in V$ then $u + v \in V$
- (b) If $v \in V$ and k a scalar, then $kv \in V$

Said differently, a subspace $V \subset X$ is just another vector space which lives naturally inside the bigger space X . Any span of vectors in \mathbf{R}^n is a subspace of \mathbf{R}^n . Note as well that this means the zero vector $\vec{0} = 0v$ must be contained in any subspace of X . Any vector space X has the two $\{\vec{0}\}$ and the whole space X . We call $\{\vec{0}\}$ the **trivial** subspace.

The main fact about subspaces is that they must always be given as the span of a (finite) collection of vectors. Given a subspace $V \subset \mathbf{R}^n$ we use the term *basis* to refer to any linearly independent spanning set of V , as follows:

10.5 Definition. Let X be a vector space and $V \subset X$ be a subspace of X . A collection of vectors $v_1, \dots, v_k \in V$ is called a **basis** of V if following two conditions are met

- (a) The vectors v_1, \dots, v_k are linearly independent
- (b) $\text{Span } \{v_1, \dots, v_k\} = V$

Note that a subspace $V \subset X$ may have many bases, but they must all consist of the same number of vectors. A subspace V is said to have **dimension** k if it has a basis consisting of k vectors. In particular, since \mathbf{R}^n has the standard basis e_1, \dots, e_n , \mathbf{R}^n is a vector space of dimension n . For a vector space X let's write $\dim X$ for the dimension of X .

10.6 Canonical subspaces of a matrix/linear transformation

Given a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ represented by $m \times n$ matrix A , there are many canonical subspaces of the domain and range associated to T .

^[36]For instance, if you recall Taylor series from calculus, this is essentially asking what functions can be represented in terms of the basis $\{x^k\}_{k \geq 0}$ of monomials. A similar idea underlies Fourier analysis where you seek to represent (periodic) functions in terms of $\sin(kx)$ and $\cos(kx)$. Not to mention that linear transformations are no longer necessarily given by matrices.

^[37]Something worth checking is that such a matrix is the *external product* of basis vectors $e_i e_j^T$

- The **nullspace** of A , written $\text{null}(A)$, is the set of vectors $x \in \mathbf{R}^n$ such that $Ax = \vec{0}$. This is the same as the **kernel** of T , i.e., the set $\ker T$ of vectors $x \in \mathbf{R}^n$ such that $T(x) = \vec{0}$.
- The **column space** of A , written $\text{col}(A)$, is the set of vectors $b \in \mathbf{R}^m$ such that $Ax = b$ for some $x \in \mathbf{R}^n$. Note that $\text{col}(A)$ is the same as the span of the columns of A , and so is the same as the image of the transformation T .
- The **row space** of A , written $\text{row}(A)$, is the span of the rows of A .

10.7 Proposition. *Given $m \times n$ matrix A , the sets $\text{null}(A)$ and $\text{row}(A)$ are subspaces of \mathbf{R}^n , the set $\text{col}(A)$ is a subspace of \mathbf{R}^m . Moreover, $\text{col}(A^T) = \text{row}(A)$ and $\text{row}(A^T) = \text{col}(A)$.*

Proof. We'll show that $\text{null}(A) \subset \mathbf{R}^n$ is a subspace. Suppose that $x, y \in \text{null}(A)$, then $Ax = Ay = \vec{0}$. So, $A(x + y) = Ax + Ay = \vec{0} + \vec{0} = \vec{0}$. Similarly, if $x \in \text{null}(A)$ and k a scalar, then $A(kx) = k(Ax) = k\vec{0} = \vec{0}$. Therefore $\text{null}(A)$ is a subspace.

Next let's show that $\text{col}(A)$ is a subspace of \mathbf{R}^m . Suppose that $u, v \in \text{col}(A)$. Then there are $x, y \in \mathbf{R}^n$ with $Ax = u$ and $Ay = v$. Therefore $u + v = Ax + Ay = A(x + y)$, and so $u + v \in \text{col}(A)$. Further, if k a scalar then $ku = k(Ax) = A(kx)$ so that $ku \in \text{col}(A)$.

The remainder is left as an exercise. □

10.8 Example (A basis for the nullspace of A). Using the techniques of Section 4 we can produce a canonical basis for the nullspace of A . First, find the matrix B such that $A \sim B$ and B is in reduced echelon form. Then you can read off the vectors associated to any free variable in the solution set of $Bx = \vec{0}$. Since $A \sim B$ if and only if $A = EB$ for E some product of elementary matrices. Since E is then invertible, E^{-1} exists and so the solution sets to $Ax = \vec{0}$ and $Bx = E^{-1}\vec{0} = \vec{0}$ must be the same.

10.9 Terminology. The dimensions of these canonical spaces also have names associated to them:

- The dimension of $\text{null}(A)$ is called the **nullity** of A .
- The dimension of $\text{col}(A)$ is called the **column rank** of A . We denote by $\text{rank}(A)$ the *column* rank of matrix A . The unadjectivized word "rank" typically refers to column rank of a matrix.
- The dimension of $\text{row}(A)$ is called the **row rank** of A .

If A is a matrix representation of a linear transformation T then these words just as well apply to T : for instance, the nullity of T is the dimension of $\text{null}(A)$.

10.10 Rank-nullity theorem

The big theorem for this section is something that we've seen before but haven't yet written down fully.

10.11 Theorem (Rank-nullity theorem). *Let A be an $m \times n$ matrix. Then*

$$\dim(\text{col}(A)) + \dim(\text{null}(A)) = n \quad (10.11.1)$$

Said differently, the dimension of the domain of a linear transformation represented by an $m \times n$ matrix A is always the sum of the dimension of the nullspace of A plus the dimension of the column space of A .

Proof. This theorem follows from the general fact that any matrix is row equivalent to one of reduced echelon form: If A is in reduced echelon form then a column of A

- Contains a leading 1, in which case it counts toward the dimension of the column space
- Does not contain a leading 1, in which case it will count toward a free variable when parametrizing the nullspace of A .

Here's the trick though: Suppose A not in echelon form and write B for its reduced echelon form. Then the column spaces of A and B need not be the same. They are, however, isomorphic^[38], and so $\dim \text{col}(A) = \dim \text{col}(B)$. \square

In terms of the transformation T between (finite dimensional) vector spaces $T: X \rightarrow Y$, that

$$\dim X = \dim(\ker T) + \dim(\text{img } T)$$

10.12 Isomorphisms

Consider the subspace $V = \text{Span} \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} \right\}$ of \mathbf{R}^3 . Since geometrically, V is a plane through the origin, it's tempting to say that V is " \mathbf{R}^2 " sitting inside of \mathbf{R}^3 —but that's not correct. When talking about subspaces (or vector spaces in general) that functionally are the same as something more canonical, the right term to use is *isomorphic*^[39].

10.13 Definition. *Let X, Y be vector spaces. A linear transformation $T: X \rightarrow Y$ is called an **isomorphism** if it is both one-to-one and onto.*

If X, Y are vector spaces such that there is an isomorphism $T: X \rightarrow Y$ then X and Y are called **isomorphic**. In notation we write $X \cong Y$.

10.14 Fact. *If $V \subset \mathbf{R}^n$ is a subspace then $V \cong \mathbf{R}^k$ for some $k \in \{0, 1, 2, \dots, n\}$ ^[40].*

^[38]If $A = EB$ for E a product of elementary matrices, then $Ax = b$ if and only if $Bx = E^{-1}b$. Since E is invertible, the map $b \mapsto E^{-1}b$ from $\text{col}(A) \rightarrow \text{col}(B)$ is an isomorphism.

^[39]*Iso-* from Greek meaning "same", *morp* from Greek meaning "shape"

^[40]An even more general fact is that if V is *any* vector space then $V \cong \mathbf{R}^\alpha$ for some cardinal number α . For instance if the vector space $C(\mathbf{R})$ of continuous functions is isomorphic to \mathbf{R}^∞ where ∞ denotes the smallest infinite cardinal number (sometimes called "countable infinity"). There's an (easy) proof of this using the most controversial axiom in all of mathematics—the *axiom of choice*. While this result seems underwhelming it basically tells you that you are free to think of vector spaces as always coming equipped with a choice of basis, sometimes we need to change this basis, but existence of a basis can always be assumed

10.15 Remark. Using the term *isomorphic* is an important distinction for vector spaces. Addressing the question from earlier, the subspace

$$V = \text{Span} \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} \right\}$$

is not *equal* to \mathbf{R}^2 because the elements in the two are different. However, from the perspective of linear algebra, the distinction between them is purely semantic: they behave as the same object (you can think of V as a “copy” of \mathbf{R}^2 inside \mathbf{R}^3 with coordinates given by the basis). The way to express this mathematically is to say that V is isomorphic to \mathbf{R}^2 . Said differently, isomorphisms are the right notion of “equality” for vector spaces^[41].

In fact, any finite dimensional vector space is isomorphic to \mathbf{R}^n for some $n \geq 0$, therefore, we’re really not losing any generality by just looking at such vector spaces.

10.16 Exercises

1. Let A be the following matrix

$$A = \begin{pmatrix} 1 & 4 & 8 & -3 & -7 \\ -1 & 2 & 7 & 3 & 4 \\ -2 & 2 & 9 & 5 & 5 \\ 3 & 6 & 9 & -5 & -2 \end{pmatrix}$$

Write a basis for $\text{null}(A)$ and for $\text{col}(A)$.

2. Write a basis for the subspace of \mathbf{R}^n

$$(a) \text{Span} \left\{ \begin{pmatrix} 1 \\ -1 \\ -2 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ -1 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -6 \\ 8 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \\ -7 \\ 7 \end{pmatrix}, \begin{pmatrix} 3 \\ -8 \\ 9 \\ -5 \end{pmatrix} \right\}$$

- (b) The set of vectors $v \in \mathbf{R}^7$ such that $v_1 = 0$

- (c) The set of vectors $v \in \mathbf{R}^5$ such that $Av = \vec{0}$ for A the matrix

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

3. Let $V \subset \mathbf{R}^3$ be the subset of vectors $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ such that $x + y + z = 0$

^[41]This might feel uncomfortable at first but you have some experience with this type of language already: for instance, are the numbers $\frac{1}{2}$ and $\frac{3}{6}$ the same? They can both be reduced to the same fraction so the answer must be yes, but they are not literally the same fraction. You can say that fractions $\frac{p}{q}$ and $\frac{r}{s}$ are “equal” if and only if $sp = qr$.

- (a) Show that V is a subspace of \mathbf{R}^3 .
- (b) Show that $\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$ is a basis for V
- (c) Let W be the subset of vectors of \mathbf{R}^3 such that $x+y+z=1$. Is W still a subspace of \mathbf{R}^3 ? Explain why or why not.
4. Determine if the following is a subspace of the given vector space
- The set of matrices $A \in M(m, n)$ such that $\det A = 0$.
 - The set of continuous functions $f \in C(\mathbf{R})$ such that $f(0) = 0$.
 - The set of continuous functions $f \in C(\mathbf{R})$ such that $f(0) = 1$
 - The set of continuous functions $f \in C(\mathbf{R})$ such that $\lim_{x \rightarrow \infty} f$ exists.
5. The **trace** of an $n \times n$ matrix A , denoted $\text{tr}(A)$, is the sum of the diagonal entries from A . That is,
- $$\text{tr}(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$
- Show that the set of matrices $A \in M(n, n)$ of trace 0 is a subspace of $M(n, n)$.
6. Here are even *more* entries to the invertible matrix theorem (Theorem 8.11). Show that the following are equivalent to a matrix A being invertible
- The columns of A form a basis for \mathbf{R}^n
 - $\text{col}(A) \cong \mathbf{R}^n$.
 - $\text{rank}(A) = n$
 - $\dim \text{null}(A) = 0$
 - $\text{null}(A) \cong \mathbf{0}$.
7. Let $u = \begin{pmatrix} 1 \\ -4 \end{pmatrix}, v = \begin{pmatrix} -2 \\ 7 \end{pmatrix}$.
- Show that $\{u, v\}$ is a basis for \mathbf{R}^2 .
 - Let $B = (u \ v)$. Determine B^{-1} .
 - Let T be the linear transformation given by $Te_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, Te_2 = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$. Then the standard matrix representation of T is given by $A = \begin{pmatrix} 1 & 0 \\ 3 & -3 \end{pmatrix}$. The matrix BAB^{-1} is called the *conjugate* of A by B , and represents the transformation T with respect to the basis $\{u, v\}$. Use this to evaluate $T(3u + 2v)$
8. Give an example of a 4×3 matrix A with $\text{rank}(A) = 1$, or explain why no such matrix exists.

9. Is the following a basis for $M(2, 2)$? Explain why or why not (if not, describe the subspace of $M(2, 2)$ formed by \mathcal{B})

$$\mathcal{B} = \left\{ T = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, L = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, R = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$$

10. True or False. Determine if the following statements are true or false. Justify your answer.

- (a) If A is in reduced echelon form $\text{rank}(A)$ is equal to the number of columns with leading 1s
 - (b) If $A \sim B$ by row operations then $\text{null}(A) \cong \text{null}(B)$ and $\text{col}(A) \cong \text{col}(B)$
 - (c) There is a 3×4 matrix A with both nullity and (column) rank 2
 - (d) If A is a 5×7 matrix then $\text{null}(A)$ can have dimension 0
 - (e) If $\{v_1, \dots, v_n\}$ is a basis of \mathbf{R}^n and $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ an invertible linear transformation then $\{Tv_1, \dots, Tv_n\}$ is also a basis of \mathbf{R}^n .
 - (f) If $\{v_1, \dots, v_n\}$ is a basis of \mathbf{R}^n and $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ any linear transformation then $\{Tv_1, \dots, Tv_n\}$ is a basis of \mathbf{R}^m .
 - (g) *For A an $m \times n$ matrix, $\dim(\text{row}(A)) = \dim(A)$
11. Let $T: M(2, 2) \rightarrow M(2, 2)$ (the collection of 2×2 matrices) be given by $T(A) = A + A^T$.
- (a) Show that T is a linear transformation
 - (b) Describe the kernel and image of T
12. *Define $T: C(\mathbf{R}) \rightarrow C(\mathbf{R})$ by T_f is the function given by

$$T_f(x) = \int_0^x f(t) dt$$

- (a) Show that T is a linear transformation
 - (b) Describe the kernel of T
13. *Let P_n denote the set of (real) polynomials of degree at most n
- (a) Show that P_n is a subspace of $C(\mathbf{R})$.
 - (b) Show that $P_n \cong \mathbf{R}^{n+1}$
 - (c) Is $\{1+x^2, 1-x^2\}$ a basis for P_2 ? Explain why or why not (if not, give an example of another element that must be added to make it a basis).
 - (d) Show that $D: P_n \rightarrow P_{n-1}$ given by $D(f) = f'$ (the derivative) is a linear transformation, and describe $\ker D$.

Suggested additional exercises: Lay sections 4.1, 4.2, 4.3

11 Eigenvalues and eigenvectors

To understand a linear transformation T on \mathbf{R}^n , it's often useful to understand what (if any) features of the domain are unchanged by T . Historically, this idea comes (in part) from studying rotations of a solid object in \mathbf{R}^3 . Any such rotation must have a fixed axis^[42], which is also the axis of rotation. Knowing this fixed axis gives information about the rotation, and vice versa. More generally, linear transformations have what are called *eigenvalues* and *eigenvectors*, which characterize the transformation independent of choice of basis.

11.1 Definition (Eigenvalue and eigenvector). *Let T be a linear transformation $\mathbf{R}^n \rightarrow \mathbf{R}^n$. A number $\lambda \in \mathbf{R}$ is called an **eigenvalue**^[43] of T if there is a nonzero vector $x \in \mathbf{R}^n$ such that*

$$T(x) = \lambda x \quad (11.1.1)$$

*Given an eigenvalue λ of T , any vector x such that equation (11.1.1) holds is called an **eigenvector** (for that eigenvalue) of T .*

11.2 The characteristic polynomial

Let's determine how to find eigenvalues of a linear transformation. Note that if λ is an eigenvalue of T then there is nonzero vector x such that

$$T(x) - \lambda x = (T - \lambda I)x = \vec{0} \quad (11.2.1)$$

(recall that I is the identity function on \mathbf{R}^n .) That is to say, that λ is a number such the linear transformation $T - \lambda I$ has a nonzero vector in its kernel, therefore is not one-to-one, and so is not invertible. This is the same as any matrix representation of $T - \lambda I$ not being an invertible matrix. Suppose that $T(x) = Ax$ for some $n \times n$ matrix, A . If λ is an eigenvalue of T , then the matrix $A - \lambda I$ must not be invertible. Since one (of many) equivalent condition for a matrix to be noninvertible is to have determinant of zero, we have the following observation:

11.3 Fact. *If $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ is given by $T(x) = Ax$ for some $n \times n$ matrix A . Then λ is an eigenvalue for T if and only if $\det(A - \lambda I) = 0$.*

11.4 Example. Let's consider the transformation $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ given by $T(x) = Ax$ for A the matrix

$$A = \begin{pmatrix} 3 & 1 \\ 6 & -2 \end{pmatrix} \quad (11.4.1)$$

Calculating, we have

$$\det(A - \lambda I) = \det \begin{pmatrix} 3 - \lambda & 1 \\ 6 & -2 - \lambda \end{pmatrix} = (3 - \lambda)(-2 - \lambda) - 6 = \lambda^2 - \lambda - 12$$

Since $\lambda^2 - \lambda - 12 = (\lambda - 4)(\lambda + 3)$ we've found that $A - \lambda I$ is not invertible if and only if $\lambda = 4$ or $\lambda = -3$.

^[42]This is sometimes called Euler's rotation theorem

^[43]This somewhat funny looking word comes from the root *eigen-* in German, which means “own” or “self”.

In the above example, we found that $\det(A - \lambda I)$ was a polynomial in the variable λ . It's worth convincing yourself that this will generally be true. Since it contains information about eigenvalues, let's give this polynomial a name.

11.5 Definition. Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ be a linear transformation with standard matrix representation given by $n \times n$ matrix A . The **characteristic polynomial** of T , denoted χ is the polynomial given by

$$\chi(\lambda) = \det(A - \lambda I) \quad (11.5.1)$$

It's appropriate to call χ the characteristic polynomial of the matrix A as well, and often more useful since this is how we calculate χ . If possible unclear, we'll write χ_A to denote that χ is the characteristic polynomial of the $n \times n$ matrix A . Note that since λ is a eigenvalue of T if and only if $\det(A - \lambda I) = 0$, we have the following fact:

11.6 Fact. λ is an eigenvalue of the transformation $T: x \mapsto Ax$ if and only if λ is a root of the characteristic polynomial of A .

11.7 Remark. Note that $\det(A - \lambda I)$ is a polynomial in λ of degree at most n . A fact you may recall from algebra is that any degree n polynomial (with coefficients that are real numbers) has at most n distinct roots^[44]. In particular, any such polynomial will factor as a product of (i) linear factors, or (ii) irreducible quadratic factors. That is to say, a linear transformation on \mathbf{R}^n can have at most n different eigenvalues.

11.8 Eigenspaces

Note that if λ is an eigenvalue for T , and x a corresponding eigenvector, then T “fixes” the subspace of \mathbf{R}^n given by the span of x . That is, since T acts by scalar multiplication on this subspace, the action of T on this subspace is particularly simple.

11.9 Proposition. Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ be a linear transformation and λ an eigenvalue of T . Then the collection of vectors x such that $T(x) = \lambda x$ is a (nontrivial) subspace of \mathbf{R}^n .

Proof. This is doable^[45]. Let's call the set of such vectors E_λ . If $x, y \in E_\lambda$ then $T(x + y) = T(x) + T(y) = \lambda x + \lambda y = \lambda(x + y)$ so $x + y \in E_\lambda$. Similarly, if k is a scalar then $T(kx) = \lambda(kx) = k(\lambda x)$ so $kx \in E_\lambda$ as well. Therefore E_λ is a subspace of \mathbf{R}^n . \square

This subspace E_λ is called the **eigenspace** of the eigenvalue λ (of T). Said differently, the eigenspace of λ is the subspace of all eigenvectors associated to λ .

11.10 Example. With the same transformation from Example 11.4, we'll now describe the eigenspaces of the eigenvalues $\lambda = -3$ and $\lambda = 4$. To find the corresponding eigenvectors, we need only describe the nullspaces of the matrices $A - 4I$ and $A + 3I$. We have

$$A - 4I = \begin{pmatrix} -1 & 1 \\ 6 & -6 \end{pmatrix}, \quad A + 3I = \begin{pmatrix} 6 & 1 \\ 6 & 1 \end{pmatrix}$$

^[44]This is a corollary of the [Fundamental theorem of algebra](#)

^[45]Actually it's trivial, since this just the nullspace of $T - \lambda I$

which have nullspaces spanned by vector $x_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $x_{-3} = \begin{pmatrix} 1 \\ -6 \end{pmatrix}$ respectively.

Therefore, we have

$$E_4 = \text{Span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad E_{-3} = \text{Span} \left\{ \begin{pmatrix} 1 \\ -6 \end{pmatrix} \right\}$$

are the two eigenspaces of T . Note further that these two vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -6 \end{pmatrix}$ are linearly independent, and so form a basis for \mathbf{R}^2 , the domain of T .

11.11 Exercises

1. For each of the following matrices, (i) find all eigenvalues, (ii) describe the eigenspace for each eigenvalue as a span of linearly independent eigenvectors

$$(a) \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$$

$$(b) \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

$$(c) \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$(d) \begin{pmatrix} 5 & -2 & 3 \\ 0 & 1 & 0 \\ 6 & 7 & -2 \end{pmatrix}$$

$$(e) \begin{pmatrix} 4 & -7 & 0 & 2 \\ 0 & 3 & -4 & 6 \\ 0 & 0 & 3 & -8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. Recall that an (upper) triangular matrix is anything of the form

$$A = \begin{pmatrix} a_{1,1} & * & * & \cdots & * \\ 0 & a_{2,2} & * & \cdots & * \\ 0 & 0 & a_{3,3} & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n,n} \end{pmatrix}$$

where $*$ denotes some arbitrary real number. Show that if A is an upper triangular matrix, then the eigenvalues of A are the entries $a_{1,1}, a_{2,2}, \dots, a_{n,n}$ from the main diagonal.

3. Let $S_x(k)$ be the shear matrix along the x -axis given by

$$S_x(k) = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

- (a) Show that 1 is the only eigenvalue of $S_x(k)$.
- (b) Show that eigenspace E_1 has dimension 1, find a vector which spans this space
4. Given an angle θ , the following matrix A represents the linear transformation which rotates the xy -plane by an angle of θ counterclockwise:

$$A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Show that if θ is not a multiple of π , then A has no (real) eigenvalues

5. Show that any linear transformation of \mathbf{R}^3 has at least one (real) eigenvalue. (*Hint: any real polynomial of degree 3 must have a root, why is this?*)
6. Let $A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Show that A has exactly one eigenvalue $\lambda = -1$. Describe the eigenspace of this eigenvalue. Can you describe the effect of this matrix on the space \mathbf{R}^3 ?
7. Let A be an $n \times n$ matrix and P an invertible matrix. Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of PAP^{-1} .
8. Let A be a square matrix. Show that A and A^T , the transpose of A , have the same set of eigenvalues.
9. Let A be a square matrix. Show that if $A^2 = 0$ (the zero matrix), then A has no nonzero eigenvalues.
10. Let T be a linear transformation of \mathbf{R}^n and λ and λ' be distinct eigenvalues of T . Let v and v' be eigenvectors for λ and λ' respectively. Show that v and v' are linearly independent.
11. *True or False.* Determine if the following statements are true or false. Give a brief justification of your answer.
- (a) If an $n \times n$ matrix A has 0 as an eigenvalue, then A is not invertible
 - (b) If A is an $n \times n$ which is not invertible, then 0 is an eigenvalue of A
 - (c) Any invertible 2×2 matrix has two linearly independent eigenvectors.
 - (d) If A is any matrix of the form

$$A = \begin{pmatrix} a & x & y \\ 0 & b & 0 \\ z & 0 & c \end{pmatrix}$$

for constants a, b, c, x, y, z , then a, b and c are all eigenvalues of A .

- (e) If T is the linear transformation $\mathbf{R}^2 \rightarrow \mathbf{R}^2$ given by reflection about the line $y = 2x$ then $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is an eigenvalue of T .

12. *Let $D(\mathbf{R})$ be the vector space of (infinitely differentiable) functions $f: \mathbf{R} \rightarrow \mathbf{R}$, and $T: D(\mathbf{R}) \rightarrow D(\mathbf{R})$ the linear transformation given by $T(f) = f'$, the derivative of f . Show that any real number $k \in \mathbf{R}$ is an eigenvalue for T and has corresponding eigenvector (sometimes called *eigenfunction*) given by e^{kx} .

(Note: this is why when solving linear ODEs you always get solutions that are linear combinations of exponentials)

Additional suggested exercises: *Lay* sections 5.1, 5.2

12 Change of basis and diagonalization

At its core, a basis for \mathbf{R}^n is just a coordinate system. So far we've been working with the *standard basis* $\mathcal{E} = \{e_1, \dots, e_n\}$ as our starting point for a basis of \mathbf{R}^n . This is built into our notation for vectors, where writing

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$$

means exactly that

$$x = x_1 e_1 + \dots + x_n e_n$$

and that if $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a linear transformation then T is represented by the standard matrix

$$A = (Te_1 \ \dots \ Te_n)$$

which acts on such vectors.

However, in many instances, this is not the most natural basis in which to view some operation. The first topic of this section concerns the algebra of translating between different bases of \mathbf{R}^n , which will then be applied to discuss diagonalization of matrices.

12.1 Change of basis

Recall that a basis \mathcal{B} for \mathbf{R}^n is just a collection of n linearly independent vectors in \mathbf{R}^n . Let's fix a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ and let P be the matrix

$$P = (b_1 \ \dots \ b_n)$$

(Note that since the vectors b_i came from a basis, the matrix $(b_1 \ \dots \ b_n)$ must be invertible). We call such a matrix P a **transition matrix**^[46]. Now, given $x \in \mathbf{R}^n$, we have

$$Px = x_1 b_1 + \dots + x_n b_n$$

That is, Px is a linear combination of the basis elements from \mathcal{B} , with weights given by the vector x .

^[46]This is potentially contentious: both P and P^{-1} are transition matrices here. In discussing changing bases as we'll see soon, you want to specify the bases you're transitioning "from" and "to"

12.2 Notation. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$ and $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis for \mathbf{R}^n . We write $x_{\mathcal{B}}$

to denote the $n \times 1$ matrix representing the coordinate of the vector x with respect to the basis \mathcal{B} . Said differently, writing

$$x_{\mathcal{B}} = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}} \quad (12.2.1)$$

means precisely that $x'_1 b_1 + \dots + x'_n b_n = x = x_1 e_1 + \dots + x_n e_n$.

To rewrite a vector x from the standard basis to the \mathcal{B} basis, we must solve for these coefficients x'_i , i.e., we must solve for x'_i in the equation

$$P \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Since P is invertible, we must then have $\begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}_{\mathcal{B}} = P^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Said differently, The transition between these coordinate systems $x_{\mathcal{B}} = P^{-1}x$ and $x = Px_{\mathcal{B}}$. That is,

$$\begin{array}{ccc} \mathbf{R}^n & \mathbf{R}^n & \text{Standard basis} \\ \downarrow P^{-1} & \uparrow P & \\ \mathbf{R}^n & \mathbf{R}^n & \mathcal{B}\text{-basis} \end{array} \quad (12.2.2)$$

12.3 Remark. Given bases $\mathcal{A} = \{a_1, \dots, a_n\}$ and $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$ the transition from a vector $x \in \mathbf{R}^n$ written in coordinates with respect to \mathcal{A} (i.e., $x_{\mathcal{A}}$) to a vector written with respect to \mathcal{B} (i.e., $x_{\mathcal{B}}$) is given by

$$x_{\mathcal{B}} = P^{-1}Qx_{\mathcal{A}} \quad (12.3.1)$$

where Q is the transformation matrix $Q = (a_1 \ \dots \ a_n)$ and $P = (b_1 \ \dots \ b_n)$. Seen differently, the following give the transformations between \mathbf{R}^n with respect to different bases.

$$\begin{array}{ccc} \mathbf{R}^n & & \mathcal{A}\text{-basis} \\ Q^{-1} \left(\begin{array}{c} \nearrow \\ \searrow \end{array} \right) Q & & \\ \mathbf{R}^n & & \text{Standard basis} \\ P \left(\begin{array}{c} \nearrow \\ \searrow \end{array} \right) P^{-1} & & \\ \mathbf{R}^n & & \mathcal{B}\text{-basis} \end{array} \quad (12.3.2)$$

12.4 Conjugate matrices

Matrices A and B are called **conjugate**^[47] (also called *similar*) if there is an invertible matrix P such that $PAP^{-1} = B$. The term PAP^{-1} is called the **conjugation of A by P** . Note as well that $PAP^{-1} = B$ is the same thing as $A = P^{-1}BP$.

12.5 Fact. *If T is a linear transformation of \mathbf{R}^n with standard matrix representation A , and $\mathcal{B} = \{b_1, \dots, b_n\}$ a basis for \mathbf{R}^n . Then the matrix representation of T with respect to the basis \mathcal{B} is given by $P^{-1}AP$, where P is the transition matrix $P = (b_1 \ \cdots \ b_n)$*

Since $P = (P^{-1})^{-1}$, this is the same as saying that the matrix representation of T with respect to the basis \mathcal{B} is the conjugation of A by P^{-1} (note that P^{-1} represents the transformation that takes vectors in the standard basis and outputs vectors in the basis \mathcal{B})

$$\begin{array}{ccc} \mathbf{R}^n & \xrightarrow{A} & \mathbf{R}^n \\ P \uparrow & & \downarrow P^{-1} \\ \mathbf{R}^n & \xrightarrow{\text{Mat}_{\mathcal{B}} T} & \mathbf{R}^n \end{array} \quad \begin{array}{c} \text{Standard basis} \\ \mathcal{B}\text{-basis} \end{array} \quad (12.5.1)$$

12.6 Notation. In notation, we'll write $\text{Mat}_{\mathcal{B}} T$ for the matrix representation of T with respect to the basis \mathcal{B} .

Another corollary from Fact 12.5 is the following.

12.7 Fact. *Let A, B be $n \times n$ matrices. A and B are conjugate if and only if there is a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ and bases \mathcal{A} and \mathcal{B} of \mathbf{R}^n such that $A = \text{Mat}_{\mathcal{A}} T$ and $B = \text{Mat}_{\mathcal{B}} T$. Moreover, if A, B are conjugate then using the notation from (12.3.1) B is the conjugation of A by $P^{-1}Q$.*

This is why it's important to specify the basis when referring to a matrix representation of T : different bases produce different matrices, but represent the same transformation.

12.8 Diagonalization

An $n \times n$ matrix is called a **diagonal matrix** if $a_{i,j} = 0$ for all $i \neq j$. Said differently, a **diagonal matrix** is one of the form

$$A = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{pmatrix} \quad (12.8.1)$$

where the terms $a_{i,i} \in \mathbf{R}$ (including 0) for $i = 1, \dots, n$. We will very tactfully use D to denote a generic diagonal matrix.

^[47]You may know of the conjugate of a *surd* such as $a + \sqrt{b}$ being $a - \sqrt{b}$, or the conjugate of a complex number $a + bi$ being $a - bi$. This is actually the same use of the word conjugate. Generally, the word conjugate in algebra applies to things which are the same up to some underlying choice, that choice here is a basis for \mathbf{R}^n .

12.9 Theorem (Diagonalization theorem). *Let A be an $n \times n$ matrix. Then there is a diagonal matrix D and invertible matrix P such that $A = PDP^{-1}$ if and only if A has n linearly independent eigenvectors. Moreover, if $A = PDP^{-1}$ then the entries of D are the eigenvalues of A and the columns of P are corresponding eigenvectors for those eigenvalues.*

That is, if D is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

and $P = (v_1 \ v_2 \ \cdots \ v_n)$ such that $A = PDP^{-1}$, then for $i = 1, \dots, n$ λ_i is an eigenvalue of A (note: you can have repeated entries of the same eigenvalue) and v_i is an eigenvector for λ_i . Since eigenvalues of a matrix are uniquely determined, the matrices D and P are determined up to the enumeration of the eigenvalues of A and choice of eigenvectors.

Note as well that if $A = PDP^{-1}$, then $D = P^{-1}AP$. So D the matrix for the transformation $x \mapsto Ax$ written with respect to a basis of \mathbf{R}^n consisting of the eigenvectors of A .

12.10 Terminology. The process of starting with a matrix A and producing a diagonal matrix D and invertible matrix P such that $A = PDP^{-1}$ is called **diagonalization**. A matrix A for which this process creates a diagonal matrix D such that $A = PDP^{-1}$ is called **diagonalizable**. It is straightforward but time consuming: you first find all eigenvalues of A , then produce bases for each of the eigenspaces of those eigenvalues. This process is reversible, which essentially gives a proof of Theorem 12.9 (see also problem 11.).

12.11 Remark. Depending on the matrix A , it's possible that the diagonalization process fails for a few different reasons. You may have a matrix which does not have any (real) eigenvalues. Or, your matrix may have some (real) eigenvalues, but the dimensions of the eigenspaces don't add up to the number of columns (and so no invertible matrix P can be made). To better understand these outcomes we need a bit more terminology.

12.12 Algebraic multiplicity

Recall that for an $n \times n$ matrix A , the *characteristic polynomial* of A , written χ_A , is given by

$$\chi_A(\lambda) = \det(A - \lambda I)$$

This polynomial χ_A has degree n , and thus therefore has at most n roots^[48] (i.e., values of λ for which $\chi_A(\lambda) = 0$). Suppose that λ_0 is a root of χ_A , then $(\lambda - \lambda_0)$ must divide $\chi_A(\lambda)$. The **multiplicity** (also called *algebraic multiplicity*) of λ is the largest power p such that

$$\chi_A(\lambda) = (\lambda - \lambda_0)^p q(\lambda) \tag{12.12.1}$$

where $q(\lambda)$ is some other polynomial such that $q(\lambda_0) \neq 0$ (i.e., λ_0 is not a root of q).

^[48]For now we are only considering *real* roots of χ_A , but it could of course have complex roots. We'll discuss this later.

12.13 Fact (Some general facts about diagonalization and eigenvectors). *Let A be an $n \times n$ matrix and $\lambda_1, \dots, \lambda_k$ be the unique eigenvectors of A . Then*

- For $i = 1, \dots, k$ let E_{λ_i} denote the eigenspace for λ_i . Then $\dim(E_{\lambda_i})^{[49]}$ is less than or equal to the multiplicity of λ_i (in the characteristic polynomials of A).
- A is diagonalizable if and only if

$$\sum_{i=1}^k \dim(E_{\lambda_i}) = n$$

The latter occur precisely when either (i) $k = n$ and all eigenvalues of A are distinct, or (ii) the dimension of E_{λ_i} equals the multiplicity of λ_i for all i .

12.14 Remark. If the geometric multiplicity of any eigenvalue is strictly less than its algebraic multiplicity, then A is not strictly diagonalizable. It's not game over though, every matrix is what's called *almost diagonalizable*: that is, every (square) matrix has a basis of *generalized eigenvectors*, and is conjugate to an *almost diagonal* matrix^[50]. We may do more with this later.

12.15 Exercises

1. Diagonalize the following matrices, if possible.

(a) $\begin{pmatrix} 1 & 0 \\ 6 & -1 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} -1 & 4 & -2 \\ -3 & 4 & 0 \\ -3 & 1 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} 4 & 0 & 2 \\ 2 & 3 & 4 \\ 0 & 0 & 3 \end{pmatrix}$

(e) $\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix}$

^[49]This is sometimes called the *geometric multiplicity* of λ_i . This fact then says that algebraic multiplicity is always larger than (or equal to) geometric multiplicity

^[50]That is, a diagonal matrix which may have an entry of 1 above any diagonal entry. This is related to the [Jordan decomposition](#) of matrices.

2. Let $A = \begin{pmatrix} 3 & 4 \\ -1 & -1 \end{pmatrix}$ and \mathcal{B} the basis of \mathbf{R}^2

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

Write a matrix B which represents the linear transformation $x \mapsto Ax$ in the basis \mathcal{B} .

3. Let's write $A \simeq B$ to mean that matrix A is conjugate to B , i.e., that there exists an invertible matrix P such that $A = PBP^{-1}$. Show that the following hold
- (a) If $A \simeq B$ then $B \simeq A$
 - (b) For any A , $A \simeq A$
 - (c) If $A \simeq B$ and $B \simeq C$ then $A \simeq C$

Note: these properties show that conjugacy of matrices forms an *equivalence relation*.

4. Construct a 2×2 matrix which is invertible but not diagonalizable, or explain why no such matrix exists. Do the same for a 2×2 matrix which is diagonalizable but not invertible.
5. Show that if A is an $n \times n$ diagonalizable matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ that $\det A = \lambda_1 \lambda_2 \dots \lambda_n$ (the product of the eigenvalues).
6. Let A be an $n \times n$ matrix and P be an invertible matrix.
- (a) For $\lambda \in \mathbf{R}$, explain why $P(A - \lambda I)P^{-1} = PAP^{-1} - \lambda I$
 - (b) Use part a. to show that the characteristic polynomials χ_A and $\chi_{PAP^{-1}}$ are the same.

That is, the characteristic polynomial is a property of a *linear transformation*, not just of any particular matrix representation of that transformation.

7. *True or False.* Determine if the following statements are true or false. Justify your answer
- (a) If A is diagonalizable then A must be invertible
 - (b) If A and B are conjugate then any eigenvector of A is an eigenvector of B
 - (c) If A is conjugate to B , then A^2 is conjugate to B^2 .
 - (d) If A is invertible and A is conjugate to B , then A^{-1} is conjugate to B^{-1} .
 - (e) If A and B are conjugate then the rank of A equals the rank of B
8. Recall that the *trace*, $\text{tr}(A)$, of a matrix A is the sum of its diagonal entries. Show that if A is a diagonalizable matrix and A is conjugate to B then $\text{tr}(A) = \text{tr}(B)$. (Harder: show that this is still true even if A is not diagonalizable).

9. Let \mathbf{P}_3 be the vector space of cubic polynomials and define $T: \mathbf{P}_3 \rightarrow \mathbf{P}_3$ by

$$T(p) = p(0) - p(1)t - p(1)t^2 + p(0)t^3$$

Determine $T(1 + t + t^2 + t^3)$ and $T(1 + t^2)$. Are either of these inputs eigenvectors for T ?

10. Let \mathbf{P}_2 be the vector space of quadratic polynomials and $T: \mathbf{P}_2 \rightarrow \mathbf{P}_2$ given by

$$T(a + bt + ct^2) = 3a + (5a - c)t + (3a - b + 2c)t^2$$

Write a matrix which represents T with respect to the basis $\mathcal{B} = \{1, t, t^2\}$ of \mathbf{P}_2 .

11. Let A be a 2×2 matrix such that $A = PDP^{-1}$ for D and P of the form:

$$D = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \quad P = \begin{pmatrix} u & v \end{pmatrix}$$

(here, $\lambda, \mu \in \mathbf{R}$ and $u, v \in \mathbf{R}^2$ such that P is invertible). Show that λ and μ are eigenvalues of A with eigenvectors u, v respectively.

Note: this same argument can be made for $n \times n$ matrices, completing the proof of Theorem 12.9.

Suggested additional exercises: *Lay* sections 5.3, 5.4

13 Markov chains

Markov chains are extremely useful modeling tools for predicting long term behavior of a discrete time probabilistic system. Named in honor of the Russian mathematician Andrey Markov, Markov chains have found uses all over fields of mathematics, statistics, physics, economics, chemistry and beyond. One notable example is Google's PageRank algorithm^[51] which was a novel use of Markov chains in improving search results in the early days of the internet.

13.1 Stochastic matrices

The word *stochastic* is used in mathematics to refer a system which is probabilistically determined. No real expertise in probability theory is required here, but we do need some terminology:

13.2 Definition. A vector $x \in \mathbf{R}^n$ is called a **probability vector** if the components of x are nonnegative and sum to 1.

^[51]Created by Google founders Sergey Brin and Larry Page while they were graduate students at Stanford. A good writeup from a course taught at Cornell can be found [here](#), we will very likely spend some time in class working through this as an example.

13.3 Definition. A **stochastic matrix** is an $n \times n$ matrix A such that each column^[52] of A is a probability vector.

Think of a stochastic matrices are matrix of probabilities. That is, given an “experiment” with n states which operates in discrete time intervals. The entry $A_{i,j}$ is the probability of transitioning from state i to state j in one time period.

13.4 Example. Each week, a student receives a dozen bottles of a beverage of their choosing: diet Red Bull, spring water, or lime seltzer. A student who chooses diet Red Bull has a 30% of choosing diet Red Bull the next week, a 50% chance of switching to spring water, and a 20% chance of switching to lime seltzer. Similarly, spring water drinkers are 15% likely to switch to diet Red Bull, stick with spring water 60% of the time, and have a 25% chance of switching to the lime seltzer. Last, lime seltzer drinkers switch to diet Red Bull 10% of the time, spring water 20% of the time, and remain lime seltzer fans 70% of the time.

This is modeled by the stochastic matrix

$$A = \begin{pmatrix} 0.3 & 0.15 & 0.1 \\ 0.5 & 0.6 & 0.2 \\ 0.2 & 0.25 & 0.7 \end{pmatrix} \quad (13.4.1)$$

where the first, second, and third columns respectively are diet Red Bull, spring water, and lime seltzer; and the entry in position i, j is the probability of switching from beverage i to beverage j .

13.5 Proposition. If x is a probability vector and A a stochastic matrix, then Ax is a probability vector.

Proof. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ be a probability vector and A a stochastic matrix. Since the i -th coordinate of Ax is given by $\sum_{j=1}^n A_{i,j}x_j$, the sum of all coordinates of Ax must be

$$\sum_{i=1}^n \sum_{j=1}^n A_{i,j}x_j = \sum_{j=1}^n \sum_{i=1}^n A_{i,j}x_j = \sum_{j=1}^n x_j \left(\sum_{i=1}^n A_{i,j} \right) = \sum_{j=1}^n x_j = 1$$

(note the first equation holds since the order of summation for a *finite* sum does not change the result) \square

Since the columns of a stochastic matrix are probability vectors the following corollary is immediate.

13.6 Corollary. If A and B are stochastic matrices then so is AB .

^[52]It's evidently more common in the literature of Markov chains to discuss how stochastic matrices act on *row* vectors instead of *column* vectors (as is typical in our course). That to say, if you look up the definition of stochastic matrices you may see that each *row* is required to add to 1 without restriction on the columns. These two definitions are equivalent, it's just a matter of convention (and transposition).

In particular, the sequence A^2, A^3, \dots of all positive integer powers of A are all stochastic matrices.

13.7 Markov chains

A **Markov chain** is a sequence of probability vectors obtained by repeated application of a stochastic matrix A to an “initial state” vector x_0 . That is, a Markov chain is a sequence $\{x_k\}_{k \geq 0}$ such that for all $k \geq 0$

- $x_k \in \mathbf{R}^n$ is a probability vector
- $x_{k+1} = Ax_k$

(Note that here we’re using subscripts to denote a sequence of vectors, not the components of a vector). Moreover, by the above, if $\{x_k\}$ is a Markov chain based on stochastic matrix A then we must have

$$x_k = Ax_{k-1} = A^2x_{k-2} = \dots = A^k x_0$$

for all k . Each vector x_k is called a **state vector** of the Markov chain. Given a Markov chain, we say $\{x_k\}$ **converges** to a vector $s \in \mathbf{R}^n$ if

$$\lim_{k \rightarrow \infty} x_k = s \tag{13.7.1}$$

(which is to say, the limiting value of each coordinate of x_k equals the corresponding coordinate of s .)

13.8 Steady-state solutions

Suppose that A is a stochastic matrix. A **steady-state vector** is a probability vector $s \in \mathbf{R}^n$ such that $As = s$. That is, it is an arrangement of probabilities unaffected by A . However, this is just to say that s is an eigenvector for A with eigenvalue 1. The following proposition tells us that every stochastic matrix has such an eigenvector.

13.9 Proposition. *If A is a stochastic matrix, then 1 is an eigenvalue of A .*

Proof. Recall that A and A^T have the same set of eigenvalues (see section 11.11 exercise 8.). So, it will be enough if we show that A^T has an eigenvalue of 1. Let $\vec{1}$ denote the vector

$$\vec{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

We claim that $\vec{1}$ is an eigenvector for A^T with eigenvalue 1. Note that the rows of A^T are the columns of A , and since A is stochastic, the rows of A^T must therefore be probability vectors. Let’s look at the first row of A^T and call the elements of this row a_1, \dots, a_n . Then the first entry of $A^T \vec{1}$ is given by

$$(a_1 \ \dots \ a_n) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = a_1 + \dots + a_n = 1$$

By a similar argument, all other entries of $A^T \vec{1}$ must also be 1. But that is to say, $A^T \vec{1} = \vec{1}$. So, $\vec{1}$ is an eigenvector of A^T with eigenvalue 1. \square

Since 1 is an eigenvalue of any stochastic matrix it follows that the eigenspace E_1 must have dimension at least one. A general fact is that all “nice enough” stochastic matrices have a unique steady-state solution (Theorem 13.10). Let’s call a stochastic matrix **regular** if all of its entries are strictly positive.

13.10 Theorem (Perron-Frobenius Theorem). *Let A be a regular^[53] stochastic matrix. Then*

- 1 is an eigenvalue of A
- The eigenspace E_1 has dimension 1, and any nonzero vector in this space has components which all have the same sign, and so there is a unique probability vector $s \in E_1$
- For any initial state probability vector x_0 , the Markov chain $\{x_k\}_{k \geq 0}$ given by $x_k = A^k x_0$ converges to s

Proof. The first claim comes from Proposition 13.9. To show that $\dim E_1 = 1$ and convergence of any Markov chain is a bit beyond the scope of our class^[54].

It’s also worth pointing out that $\dim E_1 = 1$ does not necessarily mean there is a probability vector $s \in E_1$. This comes from the fact that the components of any nonzero vector in E_1 have all the same sign (which again is a bit beyond the scope of our class to prove). Given this, however, the desired steady state solution s is the *unique* probability vector in E_1 , i.e., the one whose components add to 1. \square

Let’s return to the silly beverage example and see how Theorem 13.10 helps us.

13.11 Remark (Example 13.4 continued). Returning to our scenario from before, let’s suppose due to a supply issue the lime seltzer does not arrive, and students split the supply of diet Red Bull and spring water evenly. We then have an initial state vector of

$$x_0 = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

(here the entries reflect the proportions of students selecting each beverage). Following the probabilities from A , the next weeks selections will be

$$x_1 = Ax_0 = \begin{pmatrix} 0.3 & 0.15 & 0.1 \\ 0.5 & 0.6 & 0.2 \\ 0.2 & 0.25 & 0.7 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.225 \\ 0.55 \\ 0.225 \end{pmatrix}$$

^[53]This is actually overkill, you just need that *some* power A^k of A has all strictly positive entries, see exercise 7.

^[54]The idea is that every other eigenvalue (real or not) has size strictly less than one (for complex numbers this means complex norm). So, by writing a state vector with respect to this basis of eigenvectors, repeatedly application of A will “shrink” all components not coming from 1 down to 0, leaving just the steady state solution.

If we assume a student population of 3,000, this would mean in the next week that 1,650 are drinking spring water, and 675 each drinking diet Red Bull and lime seltzer.

The long term behavior of this system is then determined by finding the unique probability vector in the E_1 eigenspace. That is, we find a vector v with $(A - I)v = \vec{0}$ and “normalize” it to be a probability vector. The matrix

$$A - I = \begin{pmatrix} -0.7 & 0.15 & 0.1 \\ 0.5 & -0.4 & 0.2 \\ 0.2 & 0.25 & -0.3 \end{pmatrix}$$

has nullspace spanned by the probability vector

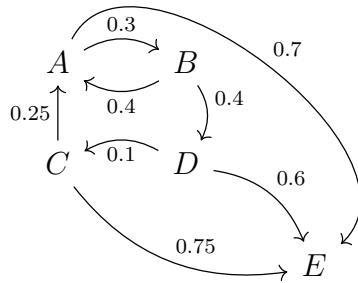
$$s = \begin{pmatrix} 0.150538 \\ 0.408602 \\ 0.44086 \end{pmatrix}$$

which therefore is the steady state solution. That is to say, with the same student population of 3000, we'd have (approximately) 452 diet Red Bull drinkers, 1226 spring water drinkers, and 1323 lime seltzer drinkers (these are rounded so they may not add up to 3000).

13.12 Exercises

- Find the steady state vectors for the following stochastic matrices (you can use a calculator)
 - $\begin{pmatrix} 0.1 & 0.6 \\ 0.9 & 0.4 \end{pmatrix}$
 - $\begin{pmatrix} 0.7 & 0.1 & 0.1 \\ 0.2 & 0.8 & 0.2 \\ 0.1 & 0.1 & 0.7 \end{pmatrix}$
 - $\begin{pmatrix} 0.3355 & 0.3682 & 0.3067 & 0.0389 \\ 0.2663 & 0.2723 & 0.3277 & 0.5451 \\ 0.1935 & 0.1502 & 0.1589 & 0.2395 \\ 0.2047 & 0.2093 & 0.2067 & 0.1765 \end{pmatrix}$
- On any given day, the weather on Planet X is either (i) dangerously windy, (ii) frighteningly cold, or (iii) sunny and pleasant. On a dangerously windy day, there is a 30% chance tomorrow will also be dangerously windy, a 40% chance tomorrow will be frighteningly cold, and a 30% chance that tomorrow will be sunny and pleasant. On a frighteningly cold day, there's a 10% chance tomorrow will be dangerously windy, an 85% chance tomorrow will be frighteningly cold, and a paltry 5% chance that tomorrow will be sunny and pleasant. On a sunny and pleasant day, there's a 45% chance tomorrow will be dangerously windy, a 45% chance that tomorrow will be frighteningly cold, and a 10% chance that tomorrow will be sunny and pleasant.
 - Write a stochastic matrix A that models the weather on Planet X

- (b) A week on Planet X has 5 days: Onesday, Twosday, Threesday, Foursday, and Partyday. Today is Onesday and it's sunny and pleasant. What is the probability it will be sunny and pleasant on Partyday? What about the *following* Partyday?
- (c) A year on Planet X last exactly 1000 days. In an average year, how many days will it be frighteningly cold, dangerously windy, and sunny and pleasant.
3. The following diagram models a system with 5 states A, B, C, D, E . An arrow from state $i \rightarrow j$ is marked with the probability that in one time interval you move from position i to j (the probability of remaining at position i is 1 minus the probability of moving to any of the other states).



- (a) Write a stochastic matrix P which models the system described by the diagram above
- (b) If you start at position D what is the probability that you will end up at position B after 10 moves have passed?
- (c) Describe what a steady-state solution to P models in terms of the system given by this diagram.
4. Let A be the stochastic matrix
- $$A = \begin{pmatrix} 0.9 & 0.01 & 0.09 \\ 0.01 & 0.9 & 0.01 \\ 0.09 & 0.09 & 0.9 \end{pmatrix}$$
- (a) Show that A has three real eigenvalues $\lambda_1 = 1$ and λ_2, λ_3 which both are less than 1 in absolute value.
- (b) Determine the steady-state solution s for A
- (c) Diagonalize A
- (d) Let $x_0 = \begin{pmatrix} 0.8 \\ 0 \\ 0.2 \end{pmatrix}$. Write x with respect to the basis of probability eigenvectors of A and explain why the Markov chain $x_k = A^k x_0$ must converge to s .

5. In general, a 2×2 regular stochastic matrix has the following form

$$A = \begin{pmatrix} \alpha & 1-\beta \\ 1-\alpha & \beta \end{pmatrix} \quad (13.12.1)$$

for α, β real numbers in the interval $(0, 1)$. Determine the steady state solution of A , your answer will be in terms of α, β .

6. The requirement that A be regular in Theorem 13.10 is critical to determine a unique steady state solution. Let A be the stochastic matrix

$$A = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.8 & 0.3 \\ 0 & 0 & 0.2 & 0.7 \end{pmatrix}$$

Show that A is stochastic, but that there is not a unique steady state solution for A .

7. Let's call a stochastic matrix *almost regular* if there is a positive integer $k \geq 1$ such that A^k has all strictly positive entries. Determine if the following are almost regular or not

$$(a) \begin{pmatrix} 0.2 & 1 \\ 0.8 & 0 \end{pmatrix}$$

$$(b) \begin{pmatrix} 1 & 0.2 \\ 0 & 0.8 \end{pmatrix}$$

Suggested additional exercises: *Lay* section 5.9

14 Inner products and orthogonality

So far we've dealt with vectors and vector spaces and learned about transformations between vector spaces. This is purely "algebraic" in nature. But, when thinking about vectors in \mathbf{R}^n we know that vectors (i.e., arrows) also have *length*, and that pairs of arrows which start at the same point must form an *angle* between them.

This additional structure (i.e., the geometric structure of length and angles) must be imposed on vector spaces: the tool that allows us to do this is called an *inner product* (more familiarly, one instance of an inner product is the usual dot product of vectors).

14.1 Definition. Let V be a vector space. An **inner product** is a number $\langle u, v \rangle \in \mathbf{R}$ assigned to each pair of vectors $u, v \in V$ such that the following properties hold: For all vectors $u, v, w \in V$ and scalars $c \in \mathbf{R}$

- (a) $\langle u, v \rangle = \langle v, u \rangle$
- (b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
- (c) $\langle cu, v \rangle = \langle u, cv \rangle = c\langle u, v \rangle$
- (d) $\langle u, u \rangle \geq 0$ with $\langle u, u \rangle = 0$ if and only if $u = \vec{0}$.

Said differently, an inner product is a function $V \times V \rightarrow \mathbf{R}$ given by $(u, v) \mapsto \langle u, v \rangle$ that satisfies the above axioms. A vector space V equipped with an inner product $\langle \cdot, \cdot \rangle$ on V is called an **inner product space**.

14.2 Example. For vectors $x, y \in \mathbf{R}^n$ let $\langle x, y \rangle = x^T y$. This is the usual *dot product* of vectors, also written $x \cdot y$. In coordinates,

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \quad (14.2.1)$$

It is a good exercise to check that this dot product satisfies all the axioms of being an inner product. When referring to the dot product of vectors $x, y \in \mathbf{R}^n$ we'll use $x \cdot y$ instead of $\langle x, y \rangle$.

In fact, given any positive numbers $c_1, \dots, c_n > 0$ there is an inner product given by^[55]

$$\langle x, y \rangle_c = c_1 x_1 y_1 + \cdots + c_n x_n y_n$$

14.3 What an inner product allows

You're probably familiar with the dot product: it shows up as a useful tool in calculus, physics and more. What maybe is not so clear is that the existence of an inner product on a vector space is what's needed to "create geometry" on that space. That is, an inner product gives us a notion of **length** of vectors and **angles** between vectors, as follows:

14.4 Definition. Let V be a vector space and $\langle \cdot, \cdot \rangle$ an inner product on V . Then

- For any $v \in V$, the **length** of v is given by $\sqrt{\langle v, v \rangle}$. We write $|v|$ for the length of v ^[56].
- For any vectors $u, v \in V$ the **angle** (in radians) between u and v is the real number $\theta \in [0, \pi]$ given by

$$\cos \theta = \frac{\langle u, v \rangle}{|u||v|} \quad (14.4.1)$$

14.5 Terminology. A vector $v \in V$ is called a **unit vector** if $|v| = 1$. A pair of vectors $u, v \in V$ are called **orthogonal** (alternatively, **perpendicular**) if $\langle u, v \rangle = 0$. Note from (14.4.1) that if u, v are nonzero vectors then they are orthogonal if and only the angle between them is $\pi/2$.

14.6 Proposition (Law of cosines). Let $u, v \in V$ be vectors. Then

$$|u - v|^2 = |u|^2 + |v|^2 - 2\langle u, v \rangle \quad (14.6.1)$$

^[55]Each of these examples is an inner product represented by a diagonal matrix. In general, if A is an $n \times n$ symmetric matrix (i.e., $A^T = A$) with n positive eigenvalues (this is sometimes called *positive definite*), then there is an inner product on \mathbf{R}^n given by $\langle x, y \rangle_A = x^T A y$, and all inner products on \mathbf{R}^n must have this form. The dot product is represented by the identity matrix, and the "c-weighted" dot product is represented by the diagonal matrix with entries c_i down the diagonal.

^[56]Sometimes you'll also see $\|v\|$ for the length of v , use this if you like, but the single bars are just fine

Proof. Expand $|u - v|^2 = \langle u - v, u - v \rangle$:

$$\langle u - v, u - v \rangle = \langle u, u \rangle - \langle u, v \rangle - \langle v, u \rangle + \langle v, v \rangle = |u|^2 - 2\langle u, v \rangle + |v|^2$$

picture here □

Using the above, we also find that $|u + v|^2 = |u|^2 + |v|^2 + 2\langle u, v \rangle$. Therefore, vectors u, v are orthogonal if and only if $|u + v|^2 = |u|^2 + |v|^2$ (this is essentially the Pythagorean theorem).

14.7 Orthogonal complement

Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$, and suppose that $W \subset V$ is a subspace of V . Let's define the **orthogonal complement of W in V** to be the set W^\perp given by

$$W^\perp = \{x \in V : \langle x, v \rangle = 0 \text{ for all } v \in W\} \quad (14.7.1)$$

14.8 Proposition. *If W is a subspace of V , then W^\perp is also a subspace of V*

Proof. Let $x, y \in W^\perp$ and $v \in W$. Then

$$\langle x + y, v \rangle = \langle x, v \rangle + \langle y, v \rangle = 0 + 0 = 0$$

and for any scalar $k \in \mathbf{R}$ we also have $\langle kx, v \rangle = k\langle x, v \rangle = k0 = 0$ □

14.9 Proposition. *Let A be an $n \times n$ matrix. Then $(\text{row}(A))^\perp = \text{null}(A)$ and $(\text{col}(A))^\perp = \text{null}(A^\perp)$*

Proof. We'll show the first one. Since the i -th entry of Ax is given by the dot product of the i -th row of A and x , it immediately follows that $x \in \text{null}(A)$ if and only each row of A and the vector x have a dot product of 0. That is to say, $x \in \text{row}(A)^\perp$. The other case is an exercise. □

14.10 Orthogonal sets in \mathbf{R}^n

Since our main focus will be on real vectors spaces with finite dimension, we'll now write everything in terms of vectors in \mathbf{R}^n and use the dot product for the inner product on \mathbf{R}^n .

Let's say that a (finite) set of vectors $S = \{v_1, \dots, v_k\}$ in \mathbf{R}^n is called an **orthogonal set** if

$$v_i \cdot v_j = 0 \quad \text{whenever } i \neq j \quad (14.10.1)$$

Orthogonal sets (of nonzero vectors) are always linearly independent, as shown in the following.

14.11 Proposition. *If $S = \{v_1, \dots, v_k\}$ is an orthogonal set of nonzero vectors in \mathbf{R}^n then S is a linearly independent collection of vectors.*

Proof. Suppose that we're given coefficients c_i such that

$$\vec{0} = c_1v_1 + c_2v_2 + \cdots + c_kv_k$$

Take this equation and dot with vector v_i , this gives:

$$0 = \vec{0} \cdot v_i = (c_1v_1 + \cdots + c_kv_k) \cdot v_i = c_i v_i \cdot v_i.$$

Therefore, $c_i|v_i|^2 = 0$, and since $|v_i| \neq 0$ this must mean that $c_i = 0$. Since i was arbitrary, this shows $c_i = 0$ for all $i = 1, \dots, n$; that is that S a linearly independent set of vectors. \square

14.12 Terminology. An **orthogonal basis** of \mathbf{R}^n is a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ which is also an orthogonal set (note from Prop 14.11 that any orthogonal set of n vectors in \mathbf{R}^n is an orthogonal basis). Furthermore, an orthogonal basis is called an **orthonormal basis** if each vector $b_i \in \mathcal{B}$ satisfies $|b_i| = 1$ (i.e., if each basis vector is also a unit vector).

14.13 Proposition. Let $V \subset \mathbf{R}^n$ be a subspace and $\mathcal{B} = \{b_1, \dots, b_k\}$ be an orthogonal basis for V . Let $v \in V$ and suppose that

$$v = c_1b_1 + \cdots + c_kb_k$$

for some scalars c_1, \dots, c_k . Then

$$c_i = \frac{v \cdot b_i}{b_i \cdot b_i}, \quad \text{for } i = 1, \dots, k \tag{14.13.1}$$

In particular, if \mathcal{B} is an orthonormal basis for V , then $c_i = v \cdot b_i$.

Said differently, it's very easy to find the coefficients of a vector written with respect to an orthogonal/orthonormal basis.

14.14 Orthogonal matrices

14.15 Definition. An $n \times n$ matrix A is called **orthogonal** if $A^T A = I$.

Note since $I^T = I$, this is the same as requiring that $AA^T = (A^T A)^T = I$. Geometrically, any rotation matrix is an orthogonal matrix, as are reflection matrices (infact, this describes all orthogonal matrices). One thing to note about orthogonal matrices is the following

14.16 Proposition. An $n \times n$ matrix A is orthogonal if and only if the columns of A form an orthonormal basis of \mathbf{R}^n .

Proof. Suppose that $A = (v_1 \ \cdots \ v_n)$. Then

$$A^T A = \begin{pmatrix} v_1 \cdot v_1 & v_2 \cdot v_1 & \cdots & v_n \cdot v_1 \\ v_1 \cdot v_2 & v_2 \cdot v_2 & \cdots & v_n \cdot v_2 \\ \vdots & \vdots & \ddots & \vdots \\ v_1 \cdot v_n & v_2 \cdot v_n & \cdots & v_n \cdot v_n \end{pmatrix}$$

Therefore, $A^T A = I$ if and only if (i) for each $i = 1, \dots, n$, $v_i \cdot v_i = 1$ (and so $|v_i| = 1$), and (ii) for each $i \neq j$, $v_i \cdot v_j = 0$. That is so say, $S = \{v_1, \dots, v_n\}$ is an orthogonal set of unit vectors, which must span \mathbf{R}^n , and so is an orthonormal basis. \square

One important role of orthogonal matrices is that they represent *isometries*^[57] of \mathbf{R}^n . That is to say, an orthogonal matrix always preserves lengths of vectors, and angles between vectors. Even differentlier, linear transformations represented by orthogonal matrices can only act on \mathbf{R}^n by rotation and reflection.

14.17 Exercises

- Determine if the following is an orthogonal set of vectors (with respect to the usual dot product of vectors in \mathbf{R}^n).

$$(a) \begin{pmatrix} -1 \\ 4 \\ -3 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -4 \\ -7 \end{pmatrix}$$

$$(b) \begin{pmatrix} 2 \\ -7 \\ 1 \end{pmatrix}, \begin{pmatrix} -6 \\ -3 \\ 9 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$$

$$(c) \begin{pmatrix} 5 \\ -4 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} -4 \\ 1 \\ -3 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 5 \\ -1 \end{pmatrix}$$

- Let A be the matrix

$$A = \begin{pmatrix} 5 & -2 & 0 \\ -2 & 3 & 2 \\ 0 & 2 & 3 \end{pmatrix}$$

and for $x, y \in \mathbf{R}^3$ define an inner product by: $\langle x, y \rangle_A = x^T A y$

$$(a) \text{ Let } x = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, y = \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}. \text{ Compute } \langle x, y \rangle_A.$$

$$(b) \text{ Show that } \langle x, y \rangle_A = \langle y, x \rangle_A$$

$$(c) \text{ Give an example of vectors } x, y \in \mathbf{R}^3 \text{ such that } \langle x, y \rangle_A \neq x \cdot y \text{ (the dot product)}$$

$$(d) \text{ Show that } \langle x, x \rangle_A \geq 0 \text{ for all vectors } x \in \mathbf{R}^3 \text{ (Hint: write a vector } x \text{ with respect to a basis of eigenvectors of } A).$$

- Let \mathcal{B} be the collection of vectors

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \right\}$$

$$(a) \text{ Show that } \mathcal{B} \text{ is an orthogonal basis for } \mathbf{R}^3.$$

^[57]*Iso-* meaning “same”, and *metry* coming from “measure”: both come from Greek

- (b) Determine the coordinates of $x = \begin{pmatrix} 8 \\ -4 \\ -3 \end{pmatrix}$ with respect to \mathcal{B} .
4. Find a third vector w so that $\{u, v, w\}$ forms an orthonormal basis of \mathbf{R}^3 for
- $$u = \begin{pmatrix} -2/3 \\ 1/3 \\ 2/3 \end{pmatrix}, \quad v = \begin{pmatrix} 1/3 \\ 2/3 \\ 0 \end{pmatrix}$$
5. Let $V \subset \mathbf{R}^4$ be the subspace with basis $\begin{pmatrix} 1 \\ 3 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \\ 1 \\ 0 \end{pmatrix}$. Determine a basis for V^\perp .
6. Let V be a subspace of \mathbf{R}^n .
- Show that if $x \in V$ and $x \in V^\perp$ then $x = \vec{0}$.
 - Show that if V has dimension k then V^\perp has dimension $n - k$.
7. Let A be an $n \times n$ orthogonal matrix and $x, y \in \mathbf{R}^n$
- Show that $(Ax) \cdot (Ay) = x \cdot y$
 - Use part a. to show that $|Ax| = |x|$. That is to say, orthogonal matrices preserve lengths.
8. Show that if A, B are orthogonal matrices that AB is orthogonal.
9. Let $x, y \in \mathbf{R}^n$, show that $|x - y|^2 + |x + y|^2 = 2|x|^2 + 2|y|^2$
10. Determine if the following statements are true or false. Justify your answer
- If $S = \{v_1, \dots, v_n\}$ is a linearly independent set of vectors then S is an orthogonal set
 - If $S = \{u, v, w\}$ is an orthogonal set of vectors in \mathbf{R}^5 then S is a linearly independent set of vectors
 - If A, B are orthogonal 2×2 matrices then $AB = BA$
 - Any orthogonal set of vectors in \mathbf{R}^n must have finitely many elements
 - If U is an orthogonal matrix then $U^{-1} = U^T$
11. Let $O(n)$ be the collection of $n \times n$ orthogonal matrices, and let $SO(n)$ be the collection of matrices $A \in O(n)$ such that $\det A = 1$. The set $SO(n)$ is called the **special orthogonal group**, matrices $A \in SO(n)$ are rotations of n -dimensional space.
- Show that for any $\theta \in \mathbf{R}$, that $R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ is in $SO(2)$.
 - Produce a matrix $B \in O(n)$ which is not in $SO(n)$.

- (c) *Prove that $\text{SO}(2) = \{R_\theta : \theta \in \mathbf{R}\}$
12. *Let $x \in \mathbf{R}^n$. Use the Pythagorean theorem to show that $|x| = \sqrt{x \cdot x}$ is equal the length of x as a line segment from the origin to point $x \in \mathbf{R}^n$.
13. Let $C([-\pi, \pi])$ denote the vector space of continuous functions $f: [-\pi, \pi] \rightarrow \mathbf{R}$. For $f, g \in C([-\pi, \pi])$ define $\langle f, g \rangle$ to be the real number given by

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x) dx$$

- (a) *Show that $\langle f, g \rangle$ is an inner product
- (b) Show that $f(x) = \sin(x)$ and $g(x) = \cos(x)$ are orthogonal, as vectors in $C([-\pi, \pi])$.
- (c) Are $f(x) = \sin(x)$ and $g(x) = \sin(2x)$ orthogonal?
- (d) Let $I: [-\pi, \pi] \rightarrow [-\pi, \pi]$ be given by $I(x) = x$ and $f_k(x) = \sin(kx)$. Compute $\langle I, f_k \rangle$. (*Note: these are the components of the “pure sine waves” of frequency 1, 2, 3, 4, … in the sawtooth wave I*)

Additional suggested exercises: *Lay* sections 6.1, 6.2

15 Orthogonal projection and Gram-Schmidt

Let's say you have a vector $x \in \mathbf{R}^n$ and want to know some component(s) of x with respect to a particular subspace^[58]. One solution would be to write a basis for V , extend this basis to all of \mathbf{R}^n , write x with respect to this basis, and select only the components that come from basis elements of V . This is doable, but fairly time consuming. A more efficient calculation relies on *projections*, which come from geometrical information gained by the dot product on \mathbf{R}^n (or more generally, an inner product).

15.1 Definition. Given a subspace $V \subset \mathbf{R}^n$, orthogonal basis $\{v_1, \dots, v_k\}$ of V , and vector $x \in \mathbf{R}^n$, the **orthogonal projection of x onto V** is the vector $\text{proj}_V(x) \in V$ given by

$$\text{proj}_V(x) = \left(\frac{x \cdot v_1}{v_1 \cdot v_1} \right) v_1 + \dots + \left(\frac{x \cdot v_k}{v_k \cdot v_k} \right) v_k \quad (15.1.1)$$

In particular, any vector $x \in \mathbf{R}^n$ then has a unique decomposition as

$$x = x_V + x_{V^\perp} \quad (15.1.2)$$

where $x_V \in V$ and $x_{V^\perp} \in V^\perp$. In this case, $x_V = \text{proj}_V(x)$ and $x_{V^\perp} = x - \text{proj}_V(x)$. It should be clear that if $x \in V$ then $\text{proj}_V(x) = x$.

The benefit to orthogonal projections is that it only relies on knowing an orthogonal basis for V , the calculations needed to determine $\text{proj}_V(x)$ then are relatively simple dot product calculations.

[58]This is no so far-fetched of a problem: for instance your vector may be a waveform and you want to understand some particular harmonic content of this waveform (i.e., the component of the it with respect to a particular frequency).

15.2 Remark. Furthermore, if $\{v_1, \dots, v_k\}$ is an orthonormal basis, then each vector v_i has $|v_i| = 1$ and so the projection of x onto V is given by

$$\text{proj}_V(x) = (x \cdot v_1)v_1 + \dots + (x \cdot v_k)v_k \quad (15.2.1)$$

That is to say, the coordinate of $\text{proj}_V(x)$ in basis direction v_i is just $x \cdot v_i$, for $i = 1, \dots, k$. A useful calculation is the following:

15.3 Proposition. Let V a subspace of \mathbf{R}^n , $\{v_1, \dots, v_k\}$ an orthonormal basis of V , and U is the matrix of column vectors $U = (v_1 \ \dots \ v_k)$, then:

$$\text{proj}_V(x) = UU^T x \quad (15.3.1)$$

15.4 The Gram-Schmidt algorithm

The formulas above for projecting a vector onto a subspace all required that the subspace had an orthogonal or orthonormal basis. Since not as bases are orthogonal, it's useful to have a general approach algorithm for turning a basis into an orthogonal one. This is called the *Gram-Schmidt* process.

15.5 Theorem. Let V be a subspace of \mathbf{R}^n and $\mathcal{B} = \{v_1, \dots, v_k\}$ be any basis for V . Then \mathcal{B} can be made into an orthogonal basis $\{u_1, \dots, u_k\}$ given as follows

$$\begin{aligned} u_1 &= v_1 \\ u_2 &= v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 \\ u_3 &= v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 \\ &\vdots \\ u_k &= v_k - \frac{v_k \cdot u_1}{u_1 \cdot u_1} u_1 - \dots - \frac{v_k \cdot u_{k-1}}{u_{k-1} \cdot u_{k-1}} u_{k-1} \end{aligned} \quad (15.5.1)$$

Note that this is the same as writing

$$\begin{aligned} u_1 &= v_1 \\ u_2 &= v_2 - \text{proj}_{u_1}(v_2) \\ u_3 &= v_3 - \text{proj}_{u_1}(v_3) - \text{proj}_{u_2}(v_3) \\ &\vdots \\ u_k &= v_k - \text{proj}_{u_1}(v_k) - \dots - \text{proj}_{u_{k-1}}(v_{k-1}) \end{aligned} \quad (15.5.2)$$

or starting with $u_1 = v_1$ and inductively defining $V_i = \text{Span}\{u_1, \dots, u_i\}$ and defining $u_{i+1} = v_{i+1} - \text{proj}_{V_i}(v_{i+1})$

15.6 Remark. It's immediate that any orthogonal basis $\{u_1, \dots, u_k\}$ can be further refined to an orthonormal basis $\mathcal{U} = \{u'_1, \dots, u'_k\}$ of V by setting $u'_i = u_i / |u_i|$ for each $i = 1, \dots, k$.

15.7 QR factorization

A useful application of the Gram-Schmidt process is what's called the *QR*-factorization of a square matrix A . This matrix factorization underlies several algorithmic solutions to common problems in linear algebra, in particular, it can be used to compute the eigenvalues of A (Example 15.9) and to solve Least squares problems (see Section 16).

15.8 Theorem. *Let A be an $m \times n$ matrix with linearly independent columns. Then A can be written as*

$$A = QR \quad (15.8.1)$$

where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\text{col}(A)$ and R is an $n \times n$ upper triangular matrix with positive entries along the diagonal.

Proof. This is constructive: let's let $A = (a_1 \ a_2 \ \cdots \ a_n)$ be given as a matrix of column vectors, $a_i \in \mathbf{R}^m$. Applying the Gram-Schmidt to the collection $\{a_1, \dots, a_n\}$ we produce an orthonormal basis $\{u_1, \dots, u_n\}$ (note: the columns of A were linearly independent so their span has dimension n) for $\text{col}(A)$ such that

$$\begin{aligned} a_1 &= (u_1 \cdot a_1) u_1 \\ a_2 &= (u_1 \cdot a_2) u_1 + (u_2 \cdot a_2) u_2 \\ &\vdots \\ a_n &= (u_1 \cdot a_n) u_1 + (u_2 \cdot a_n) u_2 + \cdots + (u_n \cdot a_n) u_n \end{aligned} \quad (15.8.2)$$

Let $Q = (u_1 \ u_2 \ \cdots \ u_n)$ and R be the triangular matrix given by

$$\begin{pmatrix} u_1 \cdot a_1 & u_1 \cdot a_2 & u_1 \cdot a_3 & \cdots & u_1 \cdot a_n \\ 0 & u_2 \cdot a_2 & u_2 \cdot a_3 & \cdots & u_2 \cdot a_n \\ 0 & 0 & u_3 \cdot a_3 & \cdots & u_3 \cdot a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_n \cdot a_n \end{pmatrix} \quad (15.8.3)$$

Without loss of generality, it can be arranged so that R has positive entries along the diagonal (if not and, say, $u_i \cdot a_i \leq 0$ then go back and replace u_i with $-u_i$). It then follows that $A = QR$ \square

15.9 Example (*QR*-algorithm for computing eigenvalues). Let A be a square matrix. Then the eigenvalues of A are just the roots of the characteristic polynomial χ_A . In general, computing these roots relies on factoring a degree n polynomial, a problem which gets prohibitively difficult as the size of n gets large. One useful application of the *QR* factorization is to compute (or even approximate) the eigenvalues of a matrix A , without having to factor χ_A .

Here's the idea: let Q_1 be an orthogonal matrix and R_1 upper triangular so that $A = Q_1 R_1$. Set $A_1 = R_1 Q_1$. Since Q_1 is orthogonal, $Q_1^T = Q_1^{-1}$ and therefore

$$Q_1 A_1 Q_1^{-1} = Q_1 R_1 Q_1 Q_1^T = Q_1 R_1 = A$$

But that is to say, that A and A_1 are conjugate, and so have the same eigenvalues. Here's the key: we can repeat this process, and for each $k \geq 2$ find an orthogonal matrix Q_k and upper triangular matrix R_k such that $A_{k-1} = Q_k R_k$. Then set $A_k = R_k Q_k$ and continue. This produces A_1, A_2, A_3, \dots of matrices which are all conjugate to A .

Moreover, under some very mild conditions on A , the sequence of matrices $\{A_k\}$ converges to a diagonal matrix S whose diagonal entries are the eigenvalues of A . The eigenvalues of A can then be approximated by taking the diagonal entries of A_k for some "large enough" value of k .

15.10 Exercises

1. Let $\mathcal{V} = \{v_1, v_2, v_3, v_4\}$ be the basis of \mathbf{R}^4 given by

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ -4 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 3 \\ 5 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -4 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 5 \\ -3 \\ -1 \\ 1 \end{pmatrix}$$

- (a) Show that \mathcal{V} is an orthogonal basis
- (b) Let $x = \begin{pmatrix} 10 \\ -8 \\ 2 \\ 0 \end{pmatrix}$. Write x in the basis \mathcal{V} .
- (c) Determine an orthonormal basis $\mathcal{U} = \{u_1, u_2, u_3, u_4\}$ such that u_i is parallel to v_i for each $i = 1, 2, 3, 4$.
2. Let $u_1 = \begin{pmatrix} 2/3 \\ 1/3 \\ 2/3 \end{pmatrix}$, $u_2 = \begin{pmatrix} -2/3 \\ 2/3 \\ 1/3 \end{pmatrix}$ and $x = \begin{pmatrix} 4 \\ 8 \\ 1 \end{pmatrix}$.
- (a) Let $U = (u_1 \ u_2)$. Compute $U^T U$ and $U U^T$
- (b) Compute $\text{proj}_{\text{Span}\{u_1, u_2\}}(x)$ and $U U^T x$. Should these be the same vector?
3. Apply the Gram-Schmidt process to produce an orthogonal basis from the following bases of the given vector space

(a) $\begin{pmatrix} 2 \\ -5 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}$

(b) $\begin{pmatrix} 3 \\ -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -5 \\ 9 \\ -9 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

$$(c) \begin{pmatrix} -10 \\ 2 \\ -6 \\ 16 \\ 2 \end{pmatrix}, \begin{pmatrix} 13 \\ 1 \\ 3 \\ -16 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ -5 \\ 13 \\ -2 \\ -5 \end{pmatrix}, \begin{pmatrix} -11 \\ 3 \\ -3 \\ 5 \\ -7 \end{pmatrix}$$

4. Determine the QR factorization of the following matrices:

$$(a) A = \begin{pmatrix} 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$(b) B = \begin{pmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{pmatrix}$$

$$(c) C = \begin{pmatrix} -1 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{pmatrix}$$

5. Use the method from Example 15.9 to approximate the eigenvalues of the matrix

$$A = \begin{pmatrix} 1 & 5 & 3 & -2 \\ 7 & 2 & 4 & 3 \\ -8 & 3 & 1 & 9 \\ 10 & 9 & -5 & -2 \end{pmatrix}$$

6. Let A be an $m \times n$ matrix. Show that every $x \in \mathbf{R}^n$ can be uniquely written as $x = p + u$ where $p \in \text{row}(A)$ and $u \in \text{null}(A)$.
7. *True or false.* Determine if the following statements are true or false. Justify your answers.
- (a) If $V \subset \mathbf{R}^n$ is a subspace and $x \in V$ then $\text{proj}_{V^\perp}(x) = \vec{0}$.
 - (b) If $V \subset \mathbf{R}^n$ is a subspace and $x \in \mathbf{R}^n$ such that $\text{proj}_{V^\perp}(x) = \vec{0}$ then $x \in V$.
 - (c) If U is an $m \times n$ matrix with orthogonal columns then $UU^T x = x$ for all $x \in \mathbf{R}^n$.
 - (d) If $V \subset \mathbf{R}^n$ is a subspace, \mathcal{B}_1 is a basis of V and \mathcal{B}_2 is a basis of V^\perp then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a basis of \mathbf{R}^n .

Additional suggested exercises: Lay sections 6.3, 6.4

16 Least squares problems

Recall that the existence of an inner product $\langle \cdot, \cdot \rangle$ allows us to define a notion of *length* of vectors via

$$|v| = \sqrt{\langle v, v \rangle} \quad (16.0.1)$$

Similarly to length, is the notion of *distance* between vectors (i.e., points in \mathbf{R}^n). Let $x, y \in \mathbf{R}^n$ be vectors (thought of as points in \mathbf{R}^n). The **distance** from x to y is given by

$$\text{dist}(x, y) = |x - y| \quad (16.0.2)$$

Note that $\text{dist}(x, y) = \text{dist}(y, x)$ and that $\text{dist}(x, y) = 0$ if and only if $x = y$ ^[59]. As far as this section goes, we'll only be considering the dot product as our inner product, but everything works just as well in general.

16.1 Curves of best fit

One common problem that linear algebra techniques can help in solving is that of minimizing distance in \mathbf{R}^n , in particular in minimizing distance to a given subspace of interest. A common (and useful) example is finding a line of best fit as follows:

16.2 Example (Linear regression). Here is a [Desmos graph](#) for this problem. Consider the points $(1, 1), (5, 2), (7, 3)$ and $(8, 3)$ in the xy -plane. Let's determine how to find a line of best fit for this collection of points. We need to be careful in what this means. Specifically, we're going to find real numbers m_0, m_1 such that

$$y = m_0 + m_1 x \quad (16.2.1)$$

“best predicts” the data given. Think of the term $m_1 x + m_0$ as “predicting” the y value, as follows:

Predicted	Actual
$m_0 + m_1 \cdot 1$	1
$m_0 + m_1 \cdot 5$	2
$m_0 + m_1 \cdot 7$	3
$m_0 + m_1 \cdot 8$	3

The above table can be turned into an $Ax = b$ problem as follows: Set A and b to be the following:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}$$

What we're looking for is a vector $m = \begin{pmatrix} m_0 \\ m_1 \end{pmatrix} \in \mathbf{R}^2$ such that Am is closest to b , i.e., which minimizes the value of $|Am - b|$. Such a vector is called a *least squares*^[60] solution to the $Ax = b$ problem.

^[59]The function dist also satisfies $\text{dist}(x, z) \leq \text{dist}(x, y) + \text{dist}(y, z)$ for all $x, y, z \in \mathbf{R}^n$ making it into a metric

^[60]Least squares here comes from the fact that when using the ordinary dot product of vectors the term $|Ax - b|$ is a sum of squares

16.3 Definition. Let A be an $m \times n$ matrix and $b \in \mathbf{R}^m$. A vector $\bar{x} \in \mathbf{R}^n$ is called a **least squares solution** to the $Ax = b$ problem if

$$|A\bar{x} - b| \leq |Ax - b| \quad (\text{for all } x \in \mathbf{R}^n) \quad (16.3.1)$$

16.4 General solution for least-squares problems

Let A , an $m \times n$ matrix, be given along with a vector $b \in \mathbf{R}^m$. The idea in finding a least squares solution is that \bar{x} yields the closest point $A\bar{x}$ in $\text{col}(A)$ (which is a subspace of \mathbf{R}^m) to a given point $b \in \mathbf{R}^m$. If $b \in \text{col}(A)$, then \bar{x} should simply be a vector such that $A\bar{x} = b$. The goal here to produce a solution analytically when b is a vector outside of $\text{col}(A)$.

Set $\bar{b} = \text{proj}_{\text{col}(A)}(b)$. Then since $\bar{b} \in \text{col}(A)$ there must be at least one vector $x \in \mathbf{R}^n$ such that $Ax = \bar{b}$, let \bar{x} be such a vector. Then, since $b - \bar{b}$ is orthogonal to the column space of A , it must be that $A^T(b - \bar{b}) = \vec{0}$ (Why?). That is to say, $A^Tb = A^T\bar{b}$, or more specifically:

$$A^Tb = A^TA\bar{x} \quad (16.4.1)$$

Note that A is not necessarily invertible (it's not even necessarily a square matrix), so that the A^T 's on the left hand side of (16.4.1) cannot necessarily be canceled. The first claim is the following:

16.5 Proposition. Let A be an $m \times n$ matrix and $b \in \mathbf{R}^m$. Any vector $\bar{x} \in \mathbf{R}^n$ that satisfies (16.4.1) is a least-squares solution to the $Ax = b$ problem.

Proof. With the right picture this should be geometrically clear: Given a vector \bar{x} which satisfies (16.4.1), $A\bar{x}$ is a vector in $\text{col}(A)$ such that $b - A\bar{x}$ meets $\text{col}(A)$ at a right angle. Then, by the Pythagorean theorem, this must be a point in $\text{col}(A)$ which minimizes the distance to b (see picture). \square

The problem then becomes determining when the (16.4.1) can be solved, and how. There's a few ways to express this, but one observation is that if, for instance, A^TA were an invertible matrix then (16.4.1) would result in $\bar{x} = (A^TA)^{-1}A^Tb$, which produces an explicit least-squares solution. A sufficient condition for this to occur is that the columns of A be linearly independent. This is summarized in the following (I leave the proof to you).

16.6 Theorem. Let A be an $m \times n$ matrix. The following are equivalent

- There is a unique least-squares solution \bar{x} for any $b \in \mathbf{R}^m$.
- The columns of A are linearly independent
- The matrix A^TA is invertible.

Furthermore, in any of the cases above the unique least squares solution \bar{x} to the $Ax = b$ problem is given by

$$\bar{x} = (A^TA)^{-1}A^Tb \quad (16.6.1)$$

Proof. From the discussion before if $A^T A$ is invertible then (16.6.1) must be the unique least squares solution. What's maybe not clear is that $A^T A$ is invertible if and only if the columns of A are independent. This follows from the claim that $Ax = \vec{0}$ if and only if $A^T A x = \vec{0}$. Indeed, if $Ax = \vec{0}$ then obviously $A^T(Ax) = A^T\vec{0} = \vec{0}$. What's less obvious is the other way around. If $A^T A x = \vec{0}$, then it must be that $(Ax) \cdot (Ax) = 0$ since

$$(Ax) \cdot (Ax) = x^T A^T A x = x^T \vec{0} = x \cdot \vec{0} = 0$$

But then $|Ax|^2 = 0$, which is to say that $|Ax| = 0$, and so $Ax = \vec{0}$ (since only $\vec{0}$ has 0 length). \square

16.7 Example 16.2 continued

Continuing with the example from above we have

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix}$$

It's clear that the columns of A are linearly independent, so there must be a unique line of best-fit to the given data. We can compute the least-squares solution \bar{m} as follows: First, note $A^T A = \begin{pmatrix} 4 & 21 \\ 21 & 139 \end{pmatrix}$. So

$$\bar{m} = (A^T A)^{-1} A^T b = \begin{pmatrix} 4 & 21 \\ 21 & 139 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 5 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 15/23 \\ 7/23 \end{pmatrix}$$

That is to say, the line of best fit for the data is given by $y = \frac{15 + 7x}{23}$.

Similar ideas can be used to fine curves of best fit to data using other “base shapes” than just lines (i.e., quadratics, sine or cosine curves, etc.)

16.8 QR-factorization and least squares

16.9 Theorem. *Given an $m \times n$ matrix A with linearly independent columns let $A = QR$ be a QR-factorization of A (see Theorem 15.8). Then for each $b \in \mathbf{R}^m$ there is a unique least squares solution to the $Ax = b$ problem given by*

$$\bar{x} = R^{-1} Q^T b \tag{16.9.1}$$

Proof. If $\bar{x} = R^{-1} Q^T b$ then

$$A\bar{x} = QRR^{-1}Q^T b = QQ^T b$$

But Q has columns which form an orthonormal basis of $\text{col}(A)$ and so QQ^T is the projection of b onto $\text{col}(A)$ (See Proposition 15.3). That is, $A\bar{x} = \text{proj}_{\text{col}(A)}(b)$ and so \bar{x} must be the least-squares solution (the solution is unique since A has linearly independent columns by assumption). \square

16.10 Remark. In practice it's much easier to compute $Q^T b$ and then back-solve for \bar{x} by solving the equation

$$R\bar{x} = Q^T b$$

(this is easy to do since R is triangular)

16.11 Exercises

1. Determine a least squares solution to the $Ax = b$ problem for the A and b the following. Is your solution unique?

$$(a) A = \begin{pmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}$$

$$(b) A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}, b = \begin{pmatrix} 5 \\ 1 \\ 0 \end{pmatrix}$$

$$(c) A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{pmatrix}, b = \begin{pmatrix} 2 \\ 5 \\ 6 \\ 6 \end{pmatrix}$$

2. Use the same idea from Example 16.2 to produce the line of best fit for the data $(-1, 0), (0, 1), (1, 2)$ and $(2, 4)$
3. Determine the best curve of the form $y = m_1x + m_2x^2$ which fits the data

$$(1, 1.8), (2, 2.7), (3, 3.4), (4, 3.8), (5, 3.9)$$

4. Determine the sinusoid of the form $y = A \cos x + B \sin x$ which best fits the data $(1, 7.9), (2, 5.4), (3, -0.9)$
5. Let A be an $m \times n$ matrix. Show that $\text{rank}(A) = \text{rank}(A^T A)$

6. *Let $(x_1, y_1), \dots, (x_n, y_n)$ be a collection of points in the xy -plane and suppose that m_0, m_1 are such that $y = m_0 + m_1x$ is the line of best fit to the given points. Set

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}, \quad \bar{y} = \frac{y_1 + \dots + y_n}{n}$$

Show that $\bar{y} = m_0 + m_1\bar{x}$, that is, the average of a collection of points always passes through the line of best fit.

Additional suggested exercises: *Lay* sections 6.5, 6.6

17 Singular value decomposition

This rather dense section concerns the *Singular value decomposition* of an arbitrary matrix A . We'll have to start with some features of symmetric matrices. Recall first that an $n \times n$ matrix A is called **symmetric** if $A = A^T$.

17.1 Spectral decomposition of symmetric matrices

First, some features of symmetric matrices.

17.2 Proposition. *Let A be a symmetric matrix. Then:*

- (a) *If λ, μ are distinct eigenvalues of A with eigenvectors x and y respectively, then x and y are orthogonal.*
- (b) *All eigenvalues of A are real numbers*

Proof. Let's start with a. Let λ, μ be distinct eigenvalues of A with eigenvectors x, y respectively. Then

$$\lambda x \cdot y = (\lambda x)^T y = (Ax)^T y = x^T A^T y = x^T A y = x^T (\mu y) = \mu x^T y = \mu x \cdot y \quad (17.2.1)$$

That is, $(\lambda - \mu)x \cdot y = 0$. Since λ and μ were assumed to be distinct, $\lambda - \mu \neq 0$ and so $x \cdot y = 0$, which is to say that x and y are orthogonal.

For b. suppose that λ is such that $Ax = \lambda x$ for some nonzero x . Then

$$(Ax) \cdot (Ax) = (Ax)^T Ax = x^T A^T Ax = x^T A^T \lambda x = \lambda x^T Ax = \lambda^2 x^T x = \lambda^2 x \cdot x$$

Therefore, $\lambda^2 = |Ax|^2/|x|^2$, which is to say that λ^2 is a positive (possibly 0) real number. Thus $\lambda \in \mathbf{R}$. \square

In addition to the above, a symmetric matrix always has “enough” eigenvectors in that if A is an $n \times n$ symmetric matrix then there is a basis of \mathbf{R}^n consisting of eigenvectors of A ^[61]. From the above, we know this basis can be arranged to be orthogonal (by, for instance, applying Gram-Schmit to each eigenspace). Therefore, if A is an $n \times n$ symmetric matrix, then there is an *orthonormal* basis of \mathbf{R}^n consisting of eigenvectors for A (we'll call this feature that is **orthogonally diagonalizable**). The following theorem summarizes the above and tells us that in fact *only* symmetric matrices have this property.

17.3 Theorem (Spectral theorem for symmetric matrices). *Let A be an $n \times n$ symmetric matrix. Then A has n (real) eigenvalues (perhaps including repetitions) and there is an orthonormal basis of \mathbf{R}^n consisting of eigenvectors of A . Moreover, any $n \times n$ matrix which admits an orthonormal basis of \mathbf{R}^n consisting of its eigenvectors must be symmetric.*

In particular, let A be a given $n \times n$ symmetric matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ then we can find an orthonormal basis $\{u_1, \dots, u_n\}$ of \mathbf{R}^n such that $A = UDU^{-1}$, where D is the

^[61]This is a bit harder to show, but the idea is the following. If A is symmetric and λ an eigenvalue with eigenvector v , then A restricts to a transformation on $\mathbf{R}^n \setminus \text{Span}\{v\}$. Said differently, if u is such that $u \cdot v = 0$ then $Au \cdot v = 0$ still since

$$(Au) \cdot v = u^T A^T v = u^T Av = u^T \lambda v = \lambda u^T v = \lambda u \cdot v =$$

Thus, if you have an eigenvalue λ of A with algebraic multiplicity $m > 1$, you can find vector $v_1 \in E_\lambda$, then $v_2 \in E_\lambda \setminus \text{Span}\{v_1\}$, $v_3 \in E_\lambda \setminus \text{Span}\{v_1, v_2\}$, etc. until you have a basis of eigenvectors $\{v_1, \dots, v_m\}$ of E_λ .

diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

and $U = (u_1 \ \cdots \ u_n)$ is orthogonal. But if U is orthogonal, then $U^{-1} = U^T$ and so:

$$A = UDU^T = (u_1 \ \cdots \ u_n) \begin{pmatrix} \lambda_1 u_1^T \\ \vdots \\ \lambda_n u_n^T \end{pmatrix} = \lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T \quad (17.3.1)$$

17.4 Terminology. The collection of eigenvalues for a matrix A is called the **spectrum** of A and the expression $\lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T$ ^[62] in (17.3.1) is called the **spectral decomposition**^[63] of the matrix A .

If we agree to enumerate the eigenvalues of A in order of decreasing magnitude (i.e., $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$) then the equation

$$A = \lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T \quad (17.4.1)$$

expresses the matrix is A in “order of importance”. In particular, if it’s known that some of the eigenvalues are very small, then terms corresponding to them can be “ignored” if only an approximation of A is desired.

17.5 Singular values

Let A be an $m \times n$ matrix. Since A is not necessarily square things like eigenvalues and diagonalization are not necessarily defined for A . However, we can still get similar type of information from such a matrix in the form of *singular values*

The first observation is the following.

17.6 Proposition. *If A is any $m \times n$ matrix, then $A^T A$ is a symmetric matrix*

Proof. This is basically trivial: $(A^T A)^T = A^T (A^T)^T = A^T A$. □

That is to say that given any matrix A , even if A is not square, the matrix $A^T A$ is orthogonally diagonalizable (see Theorem 17.3). Let $\{v_1, \dots, v_n\}$ be an orthonormal basis of \mathbf{R}^n consisting of eigenvectors of $A^T A$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Note that for each $i = 1, \dots, n$

$$|Av_i|^2 = (Av_i)^T (Av_i) = v_i^T (A^T A v_i) = v_i^T (\lambda_i v_i) = \lambda_i v_i^T v_i = \lambda_i \quad (17.6.1)$$

^[62]It’s worth here emphasizing that if $v \in \mathbf{R}^n$ is a (column) vector that vv^T is an $n \times n$ matrix. Given vectors $u, v \in \mathbf{R}^n$ this term uv^T is sometimes called the *outer product* of u, v . As opposed to the *inner product* $u^T v$ (which is a scalar quantity) the outer product is an $n \times n$ matrix

^[63]Recall that the collection of $n \times n$ matrices is an n^2 dimensional (real) vector space. You can think of this spectral decomposition as writing A with respect to the “basis” of vectors ($n \times n$ matrices) $u_i u_i^T$ with coordinates given by the eigenvalues of A

Therefore, each eigenvalue is positive, and without loss of generality we can arrange them as follows $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. The *singular values* of A are defined as the square roots of these values as follows:

17.7 Definition. For an $m \times n$ matrix A such that $A^T A$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, the *singular values* of a matrix A are given by

$$\sigma_i = \sqrt{\lambda_i} \quad (i = 1, \dots, n) \quad (17.7.1)$$

Note that if $\sigma = \sqrt{\lambda}$ is a singular value of A then there is unit vector v such that $|Av|^2 = \lambda$ and so $|Av| = \sigma$. Such a vector v behaves like a “principle axis” for the domain of A as we’ll see. First a proposition

17.8 Proposition. Let A be an $m \times n$ matrix and $\{v_1, \dots, v_n\}$ be an orthonormal basis of \mathbf{R}^n of eigenvectors for $A^T A$ so that the corresponding eigenvalues satisfy $\lambda_1 \geq \dots \geq \lambda_n \geq 0$. Let $r = \text{rank}(A)$. Then, $\sigma_1 = \sqrt{\lambda_1}, \dots, \sigma_r = \sqrt{\lambda_r}$ are the only nonzero singular values of A and $\{Av_1, \dots, Av_r\}$ is an orthogonal basis for $\text{col}(A)$.

Proof. Left as an exercise □

The main idea is that for *any* matrix A , the nonzero singular values of A essentially characterize A , in that A can be “diagonalized” by its nonsingular values as follows.

17.9 Theorem (Singular value decomposition). Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ and set r be the integer such that $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$ (if all singular values are strictly positive then $r = n$). Let D be the diagonal matrix of nonzero singular values of A ,

$$D = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

and define Σ to be the $m \times n$ matrix $\Sigma = \begin{pmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ (here $\mathbf{0}$ denotes the matrix of the appropriate size with all entries of 0). Then there are orthogonal matrices U ($m \times m$) and V ($n \times n$) such that

$$A = U\Sigma V^T \quad (17.9.1)$$

17.10 Remark (Computing U and V from the singular value decomposition). The matrices from (17.9.1) are computed as follows: Set $V = (v_1, \dots, v_n)$, where v_1, \dots, v_m is an orthonormal basis for $A^T A$ corresponding to the eigenvalues $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ of $A^T A$. U is found as follows. Suppose A has rank r , then the first r entries of U are computed as

$$u_i = \frac{Av_i}{|Av_i|} \quad 1 \leq i \leq r \quad (17.10.1)$$

If $r = m$ then we’re done. Otherwise there is an $m - r$ dimensional subspace of \mathbf{R}^m left over, pick *any* orthonormal basis u_{r+1}, \dots, u_m of this space and set $U = (u_1 \ \dots \ u_m)$.

Note that since V is orthogonal, $V^T = V^{-1}$, and so (17.9.1) should bear resemblance to the usual diagonalization of A (i.e., $A = PDP^{-1}$) albeit with a different matrix on the left and right.

17.11 Terminology. In the diagonalization case, the columns of P are the eigenvectors of A : in this case we call columns of U are called the **left singular vectors** of A and the columns of V are called the **right singular vectors**. The idea is that these vectors “characterize” the transformation A in the same way that eigenvectors do for a square matrix.

17.12 Remark. There is an analogous formula to (17.3.1) using singular values: If A has nonzero singular values $\sigma_1, \dots, \sigma_r$ and singular value decomposition $A = U\Sigma V^T$ with $U = (u_1 \ \dots \ u_m)$, $V = (v_1 \ \dots \ v_n)$ the matrices of left- and right- singular vectors of A ; then

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T \quad (17.12.1)$$

17.13 Example of computing singular value decomposition

An example feels welcome here. Let

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}.$$

In this problem we will determine the singular value decomposition of A . First, $A^T A$ is given by

$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$$

which has eigenvalues (arranged in decreasing order) of $\lambda_1 = 360$, $\lambda_2 = 90$ and $\lambda_3 = 0$ (note since A does not have linearly independent columns that $A^T A$ cannot be invertible, so much have eigenvalue of 0). Therefore, the singular values of A are $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$ and $\sigma_3 = 0$.

The right singular vectors for A come from the corresponding eigenvectors for the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of $A^T A$. These are given by the eigenspaces

$$E_{360} = \text{Span} \left\{ \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix} \right\}, \quad E_{90} = \text{Span} \left\{ \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix} \right\}, \quad E_0 = \text{Span} \left\{ \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix} \right\}$$

The bases above have already been normalized to have length one (and are orthogonal), so the orthonormal matrix V guaranteed by the singular value decomposition of A is

$$V = \begin{pmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}$$

The matrix Σ has the same dimensions as A and the nonzero singular values of A down its diagonal, so is given by

$$\Sigma = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}$$

Last we find U . The vectors in U from the normalized output of A on the eigenvectors from the nonzero eigenvalues of $A^T A$. Let's write $v_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}$ for the bases of E_{360} and E_{90} above. Then U has columns u_1, u_2 defined by

$$u_1 = \frac{Av_1}{|Av_1|} = \frac{1}{6\sqrt{10}} \begin{pmatrix} 18 \\ 6 \end{pmatrix}, \quad u_2 = \frac{Av_2}{|Av_2|} = \frac{1}{3\sqrt{10}} \begin{pmatrix} 3 \\ -9 \end{pmatrix}$$

Note that $|Av_1| = \sigma_1$ and $|Av_2| = \sigma_2$. Therefore,

$$U = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}$$

Putting it all together we get the desired singular value decomposition:

$$\underbrace{\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}}_A = \underbrace{\begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}}_U \underbrace{\begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}}_\Sigma \underbrace{\begin{pmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}}_{V^T}^T$$

17.14 Interpreting singular value decomposition

A useful question is ponder what it is we just accomplished.

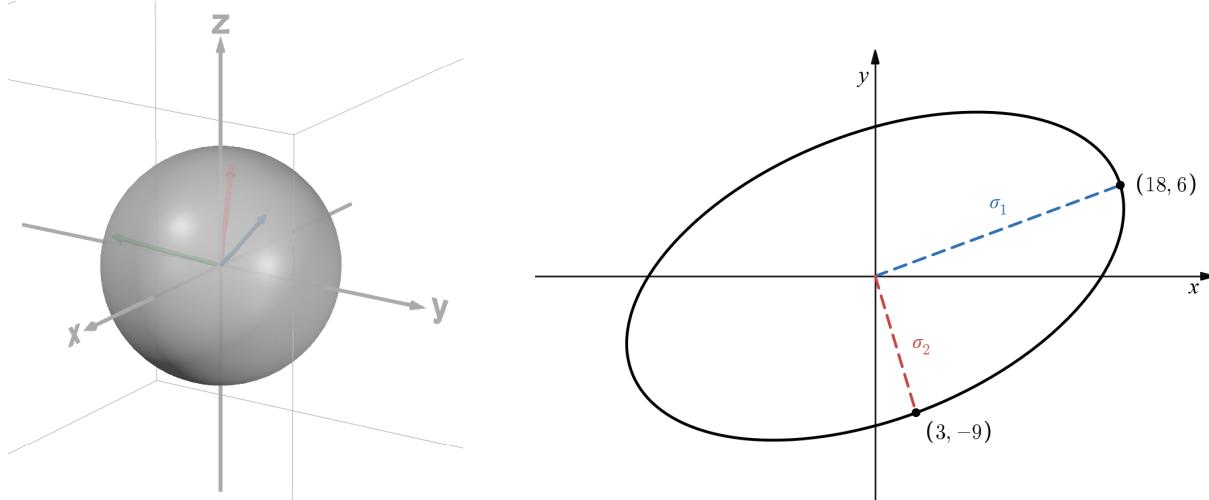


Figure 17.14.1: Left: Right singular vectors of A as seen as vectors landing on the unit sphere S in \mathbf{R}^3 (see also [this Desmos page](#)). Red and blue vectors go to the red and blue axes in the picture to the right, green axis is sent to 0. Right: the image AS of the unit sphere under A . The two principle directions of the ellipse are shown, the blue axis has $\sigma_1 = 6\sqrt{10}$ and points in direction u_1 , red has length $\sigma_2 = 3\sqrt{10}$ and points in direction u_2 .

Suppose A is the 3×2 matrix

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$$

from our previous example. Then A represents a linear transformation $\mathbf{R}^3 \rightarrow \mathbf{R}^2$. Let S be the unit sphere in \mathbf{R}^3 (i.e., $x^2 + y^2 + z^2 = 1$) and

$$AS = \{Av : v \in S\}$$

be the image of the unit sphere under A (as a region in \mathbf{R}^2). In this case, AS is an ellipse: the major and minor radii are $6\sqrt{10}$ and $3\sqrt{10}$ respectively (the singular values of A), and occur along the axes in \mathbf{R}^2 given by the corresponding left singular values of A , the columns of the matrix U . The right singular values of A denote the axes in the domain which get sent to each of these axes in \mathbf{R}^2 : the first two columns of V go to the major and minor axes of AS , respectively, and the third column (corresponding to singular value $\sigma_3 = 0$) gets “crushed” (note that a linear transformation $\mathbf{R}^3 \rightarrow \mathbf{R}^2$ cannot be injective). This is all summarized in Figure 17.14.1.

In general, when finding the singular value decomposition of an $m \times n$ matrix A you determine:

- The image of the unit “sphere” S in \mathbf{R}^n under the matrix A . This will always be an “ellipsoid” AS of dimension $k = \text{rank}(A)$ living in \mathbf{R}^m
- The principle axes of AS along with the “radii” of the ellipse along these axes
- The axes in \mathbf{R}^n which get sent to each principle axis of AS , along with all axes which get crushed by the transformation A

That’s it. Since singular value decomposition can be applied to *any* matrix A , this gives a canonical way to understand any given linear transformation. They takes spheres to ellipsoids, that’s about it. See Appendix C for an example of how singular value decomposition can be used in digital image processing.

17.15 Exercises

1. For each of the following symmetric matrices (i) compute an orthonormal basis of eigenvectors for the matrix, and (ii) determine the spectral decomposition of that matrix

$$(a) \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

$$(b) \begin{pmatrix} 1 & -6 & 4 \\ -6 & 2 & -2 \\ 4 & -2 & -3 \end{pmatrix}$$

$$(c) \begin{pmatrix} 5 & 8 & -4 \\ 8 & 5 & -4 \\ -4 & -4 & -1 \end{pmatrix}$$

2. Determine the singular value decomposition of the following matrices

$$(a) \begin{pmatrix} -1 & 1 \\ 6 & 2 \\ 6 & 2 \end{pmatrix}$$

(b) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$

(d) * $\begin{pmatrix} 6 & -8 & -4 & 5 & -4 \\ 2 & 7 & -5 & -6 & 4 \\ 0 & -1 & -8 & 2 & 2 \\ -1 & -2 & 4 & 4 & -8 \end{pmatrix}$

3. Prove proposition 17.8
4. Let $u \in \mathbf{R}^n$ be a unit vector and set $B = uu^T$.
 - (a) Show that $Bx = \text{proj}_u(x)$ for any $x \in \mathbf{R}^n$.
 - (b) Show that B is symmetric and $B^2 = B$
 - (c) Show that u is an eigenvector of B and determine the corresponding eigenvalue
5. Show that a matrix A is symmetric if and only if $(Ax) \cdot y = x \cdot (Ay)$ for all vectors $x, y \in \mathbf{R}^n$ ^[64].
6. Show that if A is an $n \times n$ matrix, then the product of the singular values of A is equal to $|\det A|$.
7. Show that if P is orthogonal, then PA and A have the same set of singular values
8. Let A have singular value decomposition $A = U\Sigma V^T$. Show that the columns of V are eigenvectors of $A^T A$, and that the columns of U are eigenvectors of AA^T .
9. This problem concerns some computations with exterior products of vectors, in particular it will show in more detail why the spectral decomposition (17.3.1) comes about.
 - (a) Let $\{e_1, \dots, e_n\}$ be the standard basis of \mathbf{R}^n and define $E_{i,j} = e_i e_j^T$. Show that $\{E_{i,j} : 1 \leq i \leq n, 1 \leq j \leq n\}$ is a basis for the vector space of $n \times n$ matrices.
 - (b) Let $A = (a_1 \ \dots \ a_n)$ and $B = (b_1 \ \dots \ b_n)$ be $n \times n$ matrices. Show that

$$AB^T = a_1 b_1^T + \dots + a_n b_n^T \quad (17.15.1)$$

(Hint: Write $A = (a_1 \ 0 \ \dots \ a_n) + (0 \ a_2 \ \dots \ 0) + \dots + (0 \ 0 \ \dots \ a_n)$ and then show for $i = 1, \dots, n$ that

$$(0 \ \dots \ 0 \ a_i \ 0 \ \dots \ 0) B^T = a_i b_i^T$$

^[64]If a linear transformation T on \mathbf{R}^n satisfies $\langle Tx, y \rangle = \langle x, Ty \rangle$ (here, \langle , \rangle is some inner product on \mathbf{R}^n) for all $x, y \in \mathbf{R}^n$, then T is called *self-adjoint*. This problem shows that self-adjoint operators on \mathbf{R}^n are represented by symmetric matrices, at least with the inner product of the dot product

- (c) Let A be an $m \times n$ matrix with nonzero singular values $\sigma_1, \dots, \sigma_r$ such that A has singular value decomposition $A = U\Sigma V^T$ with $U = (u_1 \ \dots \ u_m), V = (v_1 \ \dots \ v_n)$ the matrices of left- and right- singular values of A . Show that

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

as proposed in (17.12.1).

Additional suggested exercises: *Lay* sections 7.1, 7.4, 7.5

A Inverse matrix theorem

A.1 Theorem (The invertible matrix theorem). *Let A be an $n \times n$ matrix. The following are equivalent*

- (a) *A is invertible*
- (b) *A is equivalent by row operations to the identity matrix*
- (c) *A has n leading 1s in its reduced echelon form*
- (d) *The equation $Ax = \vec{0}$ has only the trivial solution*
- (e) *The nullspace of A is $\{\vec{0}\}$.*
- (f) *The columns of A are linearly independent*
- (g) *The linear transformation $x \mapsto Ax$ is one-to-one*
- (h) *The equation $Ax = b$ has a solution for each $b \in \mathbf{R}^n$*
- (i) *The columns of A span \mathbf{R}^n*
- (j) *The linear transformation $x \mapsto Ax$ is onto*
- (k) *There is an $n \times n$ matrix C such that $CA = I$*
- (l) *There is an $n \times n$ matrix D such that $AD = I$*
- (m) *A^T is invertible*
- (n) *$\det A \neq 0$*
- (o) *The columns of A form a basis of \mathbf{R}^n*
- (p) *The rows of A form a basis of \mathbf{R}^n*
- (q) *$\text{col}(A) \cong \mathbf{R}^n$.*
- (r) *$\text{rank}(A) = n$*
- (s) *$\dim \text{null}(A) = 0$*
- (t) *The nullity of A is 0*
- (u) *0 is not an eigenvalue of A*
- (v) *$(\text{col}(A))^\perp = \{\vec{0}\}$*
- (w) *$(\text{null}(A))^\perp = \mathbf{R}^n$*
- (x) *$\text{row}(A) = \mathbf{R}^n$*
- (y) *A has n nonzero singular values*

B Inverses and the adjugate matrix

B.1 A general form of inverse matrices

Let A be an $n \times n$ matrix and assume that $\det A \neq 0$. Then we know that A has an inverse. Recall that $A_{i,j}$ denotes the i,j cofactor matrix of A —this is an $(n-1) \times (n-1)$ submatrix of A . The **cofactor matrix** of A is the matrix C_A given by

$$(C_A)_{i,j} = (-1)^{i+j} \det(A_{i,j}) \quad (\text{B.1.1})$$

For instance, A is the 3×3 matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ then

$$C_A = \begin{pmatrix} \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} & -\det \begin{pmatrix} d & f \\ g & i \end{pmatrix} & \det \begin{pmatrix} d & e \\ g & h \end{pmatrix} \\ -\det \begin{pmatrix} b & c \\ h & i \end{pmatrix} & \det \begin{pmatrix} a & c \\ g & i \end{pmatrix} & -\det \begin{pmatrix} a & b \\ g & h \end{pmatrix} \\ \det \begin{pmatrix} b & c \\ e & f \end{pmatrix} & -\det \begin{pmatrix} a & c \\ d & f \end{pmatrix} & \det \begin{pmatrix} a & b \\ d & e \end{pmatrix} \end{pmatrix}$$

B.2 Definition. Given an $n \times n$ matrix A , the **adjugate** of A , written $\text{adj}(A)$, is the transpose of the cofactor matrix of A . That is, $\text{adj}(A) = C_A^T$, or, in components

$$\text{adj}(A)_{i,j} = (-1)^{i+j} \det A_{j,i} \quad (\text{B.2.1})$$

B.3 Example. For the 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the adjugate of A is given by:

$$\text{adj}(A) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

You may recall that if $\det A \neq 0$, then $\frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ gives a description of the inverse of A . This motivates the following theorem

B.4 Theorem. Let A be an $n \times n$ invertible matrix. Then A^{-1} is given explicitly by

$$A^{-1} = \frac{1}{\det A} \text{adj}(A) \quad (\text{B.4.1})$$

B.5 Remark. While Theorem B.4 does give an explicit method for calculating the inverse of A , it's often more efficient still to use row reduction to solve for A^{-1} (in general, the adjugate computation takes twice the number of computations as row reduction). For small matrices, you may find it's easier to use the adjugate to calculate the inverse, however^[65]

^[65]Admittedly, I could never remember what the adjugate was and made my way through undergrad and grad school only ever using row operations

C Application: Digital image processing

One nifty application of singular value decomposition is to digital image processing. Consider the following picture of my cat Leo sleeping on (one of) his favorite blankets.



Figure C.0.1: Leo enjoying a nap

This image is made from a 719×592 array of pixels, each determined by red, green, and blue values (usually integers between 0 and 255). These can be turned into three 719×592 matrices R, G, B whose entries are the intensities of each of these colors. Suppose we look at the red matrix R . R has 592 singular values $\sigma_1, \dots, \sigma_{592}$, some of which might be 0, which can be computed. If we arrange these in descending magnitude, we can compress the image by retaining the largest singular values while throwing away any that fall beneath some intensity threshold. This is a bit computationally taxing, the singular value decomposition needs to be applied three times (once to each matrix R, G and B), but is in some sense the “most optimal” image compression you could hope for^[66].

In figure C.0.2 you can find several different compressed photos of leo, along with the number of singular values used and the compression ratio. In fact, the singular values “importance” generally follows an exponential decay, by using 50 singular values most of the information is retained and a pretty high compression ratio is achieved.

^[66]JPEG compression works similarly, but the “basis” matrices are set at the beginning, rather than “dictated” by the matrix you’re trying to compress. Still neat.

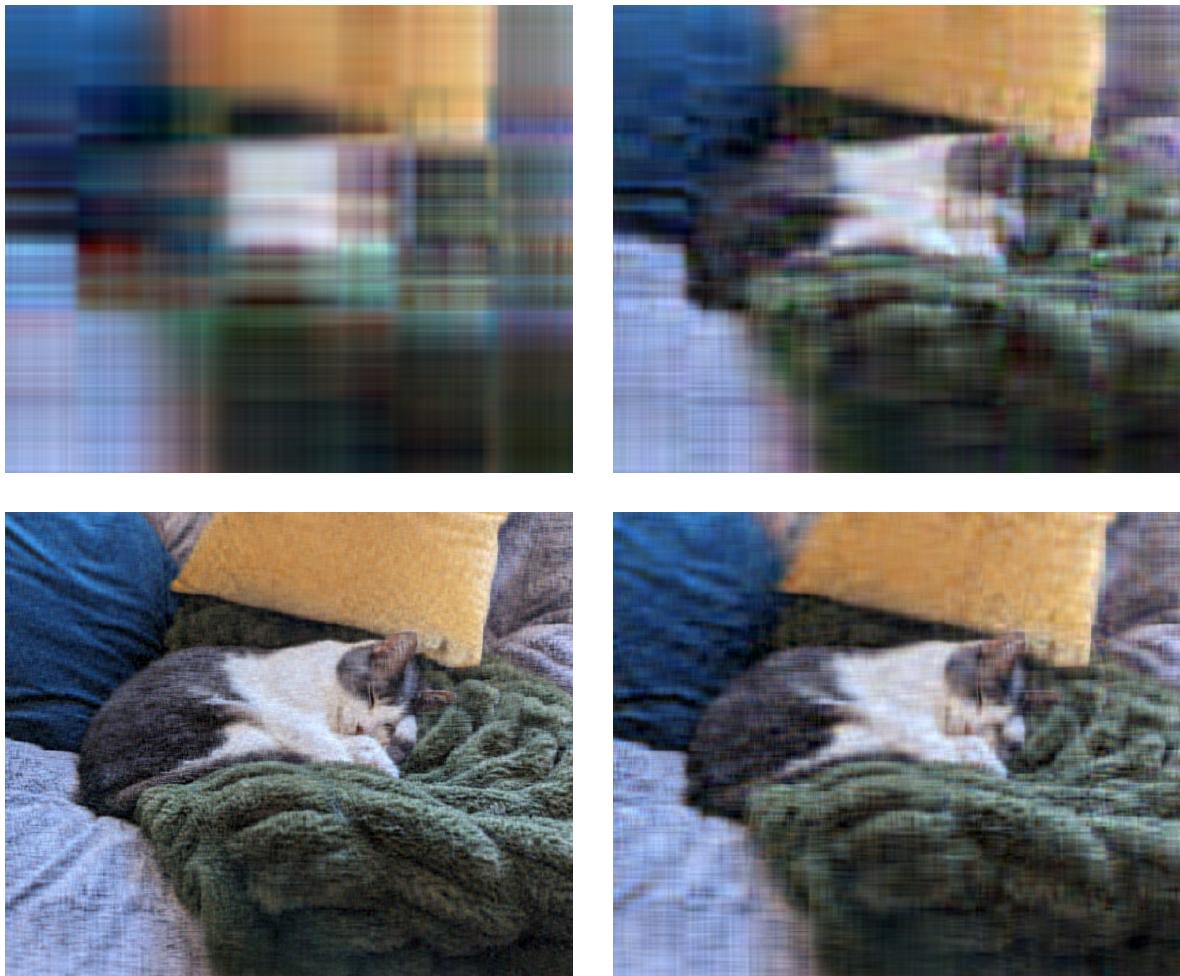


Figure C.0.2: Starting top left, moving clockwise: 3 singular values used, compression ratio 81.11; 10 singular values, compression ratio 32.44; 20 singular values, compression ratio 16.22; 50 singular values, compression ratio 6.49. Credit to [Tim Baumann](#) for the tools to make these photos.