Common Notation

Symbol	Meaning	Reference
$\mathbb{R}$	the set of real numbers	1.1
$\in$	is an element of	1.1
$\mathbb{R}^n$	the set of ordered n-tuples	1.1
$\{a, b, d\}$	a set containing a, b, and d	1.1
$\{3x : x \in \mathbb{R}\}$	the set of all $3x$ such that $x \in \mathbb{R}$	1.1
$\mathbb{R}^2$	the set of ordered pairs of real numbers; the Cartesian plane	1.1
$\subseteq$	is a subset of	1.1
$\varnothing$	the empty set	1.1
M_n	Set of (square) $n \times n$ matrices	1.2
$\overrightarrow{AB}$	the vector from the point A to the point B	1.4.1
O	the point at the origin	1.4.1
$\vec{v}$ or $\mathbf{v}$	a vector	1.4.1
$\vec{0}$ or $\mathbf{0}$	the zero vector	1.4.1
$\mathbb{R}^3$	Euclidean threespace	1.4.2
A^T	Transpose of A	1.5.3
$N(A)$ or $\ker(A)$	Nullspace or kernel of matrix A	1.6
WLOG	Without Loss of Generality	1.7
I_n	Identity matrix in M_n	2.2
$\mathbf{e}_i$ or $\vec{e}_i$	Standard basis vectors for $\mathbb{R}^n$	2.5
V	vector space	3.1
$\mathcal{P}(x)$	space of polynomials in x	3.1
$\mathcal{F}(\mathbb{R}, \mathbb{R})$	the space of functions from $\mathbb{R}$ to $\mathbb{R}$	3.1
$\mathbb{Z}$	the set of integers	3.1
$\exists$	There exists	

Symbol	**Meaning**	**Reference**
$\cong$	Is isomorphic to	3.6
$\sim$	Is similar to	5.2
λ	Eigenvalue of an operator	4.1
E_λ	Eigenspace corresponding to the eigenvalue λ	4.1
$\det A$	Determinant of A	4.2
M_{ij}	The i, j minor matrix of a matrix A	4.2.1
A_{ij}	The i, j cofactor of a matrix A	4.2.1
$\chi_A(\lambda)$	Characteristic polynomial of A	4.3
$\mathrm{Tr}(A)$	Trace of A	5.3
$\mathbf{u} \cdot \mathbf{v}$	dot product of $\mathbf{u}$ and $\mathbf{v}$	6.1
$\\|\mathbf{v}\\|$	magnitude of $\mathbf{v}$	6.1
$d(\mathbf{x}, \mathbf{y})$	distance between $\mathbf{x}$ and $\mathbf{y}$	6.1
$\mathrm{proj}_{\mathbf{v}}\mathbf{u}$	The projection of $\mathbf{u}$ onto $\mathbf{v}$	6.1
$\langle \mathbf{u}, \mathbf{v} \rangle$	The inner product of $\mathbf{u}$ and $\mathbf{v}$	6.2
$U^\perp$	Orthogonal complement to U	6.4
$\mathbf{v}_U, \mathbf{v}_{U^\perp}$	Orthogonal decomposition	6.4

Contents

Introduction: Changing Perspectives

In this course we want to study "high-dimensional spaces" and "vectors". That's not very specific, though, until we explain exactly what we mean by those things.

An important idea of this course is that it is helpful to study the same things from more than one perspective; sometimes a question that is difficult from one perspective is easy from another, so the ability to have multiple viewpoints and translate between them is extremely useful.

In this course we will take three different perspectives, which I am calling "geometric", "algebraic", and "formal". The first involves spatial reasoning and pictures; the second involves arithmetic and algebraic computations; the third involves formal definitions and properties. The formal perspective is the most abstract and sometimes the most confusing, but often the most fruitful: the formal perspective allows us to take problems that don't look like they involve anything we would call "vectors", and apply the techniques of linear algebra to them anyway.

A common definition of a vector is "something that has size and direction." This is a *geometric* viewpoint, since it calls to mind a picture. We can also view it from an *algebraic* point of view by giving it a set of coordinates. For instance, we can specify a two-dimensional vector by giving a pair of real numbers (x, y), which tells us where the vector points from the origin at $(0, 0)$. From the formal perspective we just have "a vector", which must satisfy certain conditions we'll state later.

In the table below I have several concepts, and ways of thinking about them in each perspective. It's fine if you don't know what some of these things mean, especially in the "formal" column; if you knew all of this already you wouldn't need to take this course.

Geometric	Algebraic	Formal
size and direction	n-tuples	vectors
consecutive motion	pointwise addition	vector addition
stretching, rotations, reflections	matrices	linear functions
number of independent directions	number of coordinates	dimension
plane	system of linear equations	subspace
angle	dot product	inner product
Length	magnitude	norm

1 Systems of Linear Equations

We're going to start this course with a very concrete, very algebraic problem: solving equations. As the course progresses, we will see how this problem relates to geometric and formal ideas. We will bring in ideas from geometric and formal perspectives to help us approach this problem, and see how we can use our equation-solving techniques to answer questions that arise in geometric and formal settings.

1.1 Basics of Linear Equations

A *linear equation* is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b \tag{1}$$

where $a_1, \dots, a_n$, and b are all real numbers, and $x_1, \dots, x_n$ are *unknowns* or *variables.* (We might write $a_1, \dots, a_n, b \in \mathbb{R}$; the symbol $\mathbb{R}$ stands for the real numbers, and the symbol $\in$ means "is an element of" or just "in"). We say that this equation has n unknowns.

A *system of linear equations* is a system of the form

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

with the a_{ij} and b_is all real numbers. We say this is a system of m equations in n unknowns.

Importantly, these equations are restricted to be relatively simple. In each equation we multiply each variable by some constant real number, add them together, and set that equal to some constant real number. We aren't allowed to multiply variables together, or do anything else fancy with them. This means the equations can't get too complicated, and are relatively easy to work with.

Example 1.1. A system of two linear equations in two variables is

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

A system of two equations in three variables is

$$\begin{aligned} 5x + 2y + z &= 7 \\ 3x + 2y + z &= 6. \end{aligned}$$

A system of three equations in one variable is

$$\begin{aligned} 3x &= 3 \\ 5x &= 5 \\ x &= 2. \end{aligned}$$

We want to find *solutions* to this system of equations. Since there are n variables, a solution much be a list of n real numbers. We write $\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}\}$ for the set of ordered lists of n real numbers. (We sometimes call these "ordered n-tuples" or "vectors"). Thus $\mathbb{R}^1 = \mathbb{R}$ is just the set of real numbers; $\mathbb{R}^2$ is the set of ordered pairs that makes up the Cartesian plane.

An element $(x_1, \ldots, x_n) \in \mathbb{R}^n$ is a *solution* to a system of linear equations if all of the equalities hold for that collection of x_i. The *solution set* of a system of linear equations is the set of all solutions, and we say two systems are equivalent if they have the same set of solutions.

Example 1.2. The system

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

has $(2, -1) \in \mathbb{R}^2$ as a solution. We will see later that this is the only solution, and thus the set of solutions is $\{(2, -1)\}$.

The system

$$\begin{aligned} 4x + 2y + 2z &= 8 \\ 3x + 2y + z &= 6 \end{aligned}$$

has $(1, 1, 1)$ as a solution. This is not the only solution; in fact, the set of solutions is $\{(x, 2 - x, 2 - x) : x \in \mathbb{R}\}$. (This means that for each real number x, the ordered triple $(x, 2 - x, 2 - x)$ is a solution to our system). We say this is a *subset* of $\mathbb{R}^3$, since it is a collection of elements of $\mathbb{R}^3$, and write $\{(x, 2 - x, 2 - x) : x \in \mathbb{R}\} \subset \mathbb{R}^3$.

The system

$$\begin{aligned} 3x &= 3 \\ 5x &= 5 \\ x &= 2 \end{aligned}$$

clearly has no solutions, since the first equation implies that $x = 1$ but the third equation implies that $x = 2$. Thus the set of solutions is the *empty set* $\{\} = \varnothing$.

We say that two systems of equations are *equivalent* if they have the same set of solutions. Thus the process of solving a system of equations is mostly the process of converting a system into an equivalent system that is simpler.

There are three basic operations we can perform on a system of equations to get an equivalent system:

1. We can write the equations in a different order.
2. We can multiply any equation by a nonzero scalar.
3. We can add a multiple of one equation to another.

All three of these operations are guaranteed not to change the solution set; proving this is a reasonable exercise. Our goal now is to find an efficient way to use these rules to get a useful solution to our system.

Example 1.3. The system

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

is equivalent to

$$\begin{aligned} 2x + y &= 3 \\ -2x + -10y &= 6 \end{aligned}$$

and then

$$\begin{aligned} 0x + -9y &= 9 \\ -2x + -10y &= 6 \end{aligned}$$

then

$$\begin{aligned} 0x + y &= -1 \\ -2x + -10y &= 6 \end{aligned}$$

$$\begin{aligned} 0x + y &= -1 \\ -2x + 0y &= -4 \end{aligned}$$

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

which give us our solution of $x = 2, y = -1$ or $(x, y) = (2, -1)$.

This takes up a really awkward amount of space on the page, though, and we'd like to find a better and more systematic way of approaching this process.

Remark 1.4. There's another possible approach to solving these systems, called the method of substitution. We could observe that if $2x + y = 3$ then $y = 3 - 2x$, and substitute that into our other equation to give

$$\begin{aligned} x + 5(3 - 2x) &= -3 \\ 15 - 9x &= -3 \\ 9x &= 18 \\ x &= 2 \end{aligned}$$

and from here we can see that $y = 3 - 2(2) = -1$.

This is often much simpler to do in your head for small systems. But it scales up really poorly to systems with more than two or three equations and variables, so we'll want to learn something more effective.

1.2 The matrix of a system

Looking at a system of linear equations, we notice that it can be described by an array of real numbers. These numbers are naturally laid out in a rectangular grid, so we wnat to find an efficient way to represent them.

Definition 1.5. A *(real) matrix* is a rectangular array of (real) numbers. A matrix with m rows and n columns is *a* $m \times n$ *matrix*, and we notate the set of all such matrices by $M_{m \times n}$.

A $m \times n$ matrix is *square* if $m = n$, that is, it has the same number of rows as columns. We will sometimes represent the set of $n \times n$ square matrices by M_n.

We will generally describe the elements of a matrix with the notation

$$(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

We can now take the information from a system of linear equations and encode it in a matrix. Right now, we will just use this as a convenient notational shortcut; we will see later on in the course that this has a number of theoretical and practical advantages.

Definition 1.6. The *coefficient matrix* of a system of linear equations given by

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is the matrix

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

and the *augmented coefficient matrix* is

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

Example 1.7. Suppose we have a system

$$\begin{aligned} 4x + 2y + 2z &= 8 \\ 3x + 2y + z &= 6. \end{aligned}$$

Then the coefficient matrix is

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

and the augmented coefficient matrix is

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 8 \\ 3 & 2 & 1 & 6. \end{array}\right]$$

Earlier we listed three operations we can perform on a system of equations without changing the solution set: we can reorder the equations, multiply an equation by a nonzero scalar, or add a multiple of one equation to another. We can do analogous things to the coefficient matrix.

Definition 1.8. The three *elementary row operations* on a matrix are

I Interchange two rows.

II Multiply a row by a nonzero real number.

III Replace a row by its sum with a multiple of another row.

Example 1.9. What can we do with our previous matrix? We can

$$\begin{bmatrix} 4 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} \overset{I}{\to} \begin{bmatrix} 3 & 2 & 1 \\ 4 & 2 & 2 \end{bmatrix} \overset{II}{\to} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \overset{III}{\to} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}.$$

So how do we use this to solve a system of equations? The basic idea is to remove variables from successive equations until we get one equation that contains only one variable—at which point we can substitute for that variable, and then the others. To do that with this matrix, we have

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 8 \\ 3 & 2 & 1 & 6 \end{array}\right] \overset{III}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 6 \end{array}\right] \overset{III}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 2 & -2 & 0 \end{array}\right] \overset{II}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 0 \end{array}\right].$$

What does this tell us? That our system of equations is equivalent to the system

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

This gives us the answer I stated earlier: $z = 2 - x$ and $y = z = 2 - x$.

Example 1.10. Solve the system of equations

$$x + 2y + z = 3$$
$$3x - y - 3z = -1$$
$$2x + 3y + z = 4.$$

This system has augmented coefficient matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 2 & 3 & 1 & 4 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & -1 & -1 & -2 \end{array}\right]$$

$$\stackrel{II}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & 1 & 1 & 2 \end{array}\right] \stackrel{I}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & -7 & -6 & -10 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

which gives us the system

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

The last equation tells us $z = 4$, which then gives $y = -2$ and $x = 3$. We can check that this solves the system.

1.3 Row Echelon Form

We want to solve systems of linear equations, using these matrix operations. We want to be somewhat more concrete about our goals: what exactly would it look like for a system to be solved?

Definition 1.11. A matrix is in *row echelon form* if

- Every row containing nonzero elements is above every row containing only zeroes; and
- The first (leftmost) nonzero entry of each row is to the right of the first nonzero entry of the above row.

Remark 1.12. Some people require the first nonzero entry in each nonzero row to be 1. This is really a matter of taste and doesn't matter much, but you should do it to be safe; it's an easy extra step to take by simply dividing each row by its leading coefficient.

Example 1.13. The following matrices are all in Row Echelon Form:

$$\begin{bmatrix} 1 & 3 & 2 & 5 \\ 0 & 3 & -1 & 4 \\ 0 & 0 & -2 & 3 \end{bmatrix} \quad \begin{bmatrix} 5 & 1 & 3 & 2 & 8 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -7 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 5 \\ 0 & -2 & 3 \\ 0 & 0 & 7 \end{bmatrix}.$$

The following matrices are not in Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 & 5 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 5 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Definition 1.14. The process of using elementary row operations to transform a system into row echelon form is *Gaussian elimination.*

A system of equations sometimes has a solution, but does not always. We say a system is *inconsistent* if there is no solution; we say a system is *consistent* if there is at least one solution.

Example 1.15. Consider the system of equations given by

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\ -1x_1 + -1x_2 + x_5 &= -1 \\ -2x_1 + -2x_2 + 3x_5 &= 1 \\ x_3 + x_4 + 3x_5 &= -1 \\ x_1 + x_2 + 2x_3 + 2x_4 + 4x_5 &= 1. \end{aligned}$$

This translates into the augmented matrix

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

$$\to \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right].$$

We see that the final two equations are now $0 = -4$ and $0 = -3$,s othe system is inconsistent.

Example 1.16. Let's look at another system that is almost the same.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\ -1x_1 + -1x_2 + x_5 &= -1 \\ -2x_1 + -2x_2 + 3x_5 &= 1 \\ x_3 + x_4 + 3x_5 &= 3 \\ x_1 + x_2 + 2x_3 + 2x_4 + 4x_5 &= 4. \end{aligned}$$

This translates into the augmented matrix

$$\left[\begin{array}{rrrrr|r} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & 1 & -1 \\ -2 & -2 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 & 3 \\ 1 & 1 & 2 & 2 & 4 & 4 \end{array}\right] \to \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & 1 & 1 & 3 & 3 \end{array}\right]$$
$$\to \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right] \to \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see this system is now consistent. Our three equations are

$$x_1 + x_2 + x_3 + x_4 + x_5 = 1 \qquad x_3 + x_4 + 2x_5 = 0 \qquad x_5 = 3.$$

Via back-substitution we see that we have

$$x_5 = 3 \qquad x_3 + x_4 = -6 \qquad x_1 + x_2 = 4.$$

Thus we could say the set of solutions is $\{(\alpha, 4-\alpha, \beta, -6-\beta, 3)\} \subseteq \mathbb{R}^5$.

What we were just doing definitely worked, but even after we finished transforming the matrix we still needed to do some more work. So we'd like to reduce the matrix even further until we can just read the answer off from it.

Definition 1.17. A matrix is in *reduced row echelon form* if it is in row echelon form, and the first nonzero entry in each row is the only entry in its column.

This means that we will have some number of columns that each have a bunch of zeroes and one 1. Other than that we may or may not have more columns, which can contain basically anything; we've used up all our degrees of freedom to fix those columns that contain the leading term of some row.

Note that the columns we have fixed are not necessarily the first columns, as the next example shows.

Example 1.18. The following matrices are all in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 17 & 0 & 2 & 8 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The following matrices are not in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 2 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 15 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 1.19. Let's solve the following system by putting the matrix in reduced row echelon form.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 2 \\ x_1 + x_2 + x_3 + 2x_4 + 2x_5 &= 3 \\ x_1 + x_2 + x_3 + 2x_4 + 3x_5 &= 2 \end{aligned}$$

We have

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

From this we can read off the solution $x_1 + x_2 + x_3 = 1, x_4 = 2, x_5 = -1$. Thus the set of solutions is $\{(1 - \alpha - \beta, \alpha, \beta, 2, -1)\}$.

We say some systems of equations are "overdetermined", which means that there are more equations than varaibles. Overdetermined equations are "usually" inconsistent, but not always—they can be consistent when some of the equations are redundant.

Example 1.20. The system

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1 \\ 2x_1 - x_2 + x_3 &= 2 \\ 4x_1 + 3x_2 + 3x_3 &= 4 \\ 2x_1 - x_2 + 3x_3 &= 5 \end{aligned}$$

gives the matrix

$$\left[\begin{array}{rrr|r} 1 & 2 & 1 & 1 \\ 2 & -1 & 1 & 2 \\ 4 & 3 & 3 & 4 \\ 2 & -1 & 3 & 5 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 2 & 1 & 1 \\ 0 & -5 & -1 & 0 \\ 0 & -5 & -1 & 0 \\ 0 & -5 & 1 & 3 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 2 & 1 & 1 \\ 0 & -5 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 3 \end{array}\right]$$

$$\to \left[\begin{array}{rrr|r} 1 & 2 & 1 & 1 \\ 0 & 1 & 1/5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3/2 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 0 & 3/5 & 1 \\ 0 & 1 & 1/5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3/2 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 0 & 0 & 1/10 \\ 0 & 1 & 0 & -3/10 \\ 0 & 0 & 1 & 3/2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This gives us the solution $x_1 = 1/10, x_2 = -3/10, x_3 = 3/2$, which you can go back and check solves the original system.

This overdetermined system does have a solution, but only because two of the equations were redundant, as we could see in the second matrix where two lines are identical. In fact we can go back to the original set of equations, and see that if we add two times the first equation to the second equation, we get the third—which is the redundancy.

Other systems of equations are "underdetermined", which means there are more variables than equations. These systems are usually but not always consistent.

Example 1.21. Let's consider the system

$$\begin{aligned} -x_1 + x_2 - x_3 + 3x_4 &= 0 \\ 3x_1 + x_2 - x_3 - x_4 &= 0 \\ 2x_1 + x_2 - 2x_3 - x_4 &= 0. \end{aligned}$$

This gives us the matrix

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

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

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

We see that we can't "simplify" the fourth column in any way; we don't have any degrees of freedom after we fix the first three columns. This means that we can pick x_4 to be anything we want, and the other variables are given by $x_1 - x_4 = 0, x_2 - 3x_4 = 0, x_3 + x_4 = 0$. Thus the set of solutions is $\{(\alpha, 3\alpha, -\alpha, \alpha)\}$.

Remark 1.22. A system of any size can be either consistent or inconsistent. $0 = 1$ is an inconsistent system with one equation, and

$$\begin{aligned} x_1 + \cdots + x_{100} &= 0 \\ x_1 + \cdots + x_{100} &= 1 \end{aligned}$$

is an inconsistent system with a hundred variables and only two equations. In contrast,

$$\begin{aligned} x_1 &= 1 \\ x_1 &= 1 \\ &\vdots \\ x_1 &= 1 \end{aligned}$$

has only one variable, and many equations, and is still consistent.

1.4 Vectors and Spans

1.4.1 The Cartesian Plane

We'll start by considering the "Cartesian plane", (named after the French mathematician René Descartes, who is credited with inventing the idea of putting numbered cordinates on the plane).

As probably looks familiar from high school geometry, given two points A and B in the plane, we can write $\overrightarrow{AB}$ for the vector with *initial point* A and *terminal point* B.

Since a vector is just a length and a direction, the vector is "the same" if both the initial and terminal points are shifted by the same amount. If we fix an *origin* point O, then any point A gives us a vector $\overrightarrow{OA}$. Any vector can be shifted until its initial point is O, so each vector corresponds to exactly one point. We call this *standard position.*

We represent points algebraically with pairs of real numbers, since points in the plane are determined by two coordinates. Thus we use $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ to denote the set of all ordered pairs of real numbers, giving us an algebraic description of the Cartesian plane. We define the origin O to be the "zero" point $(0, 0)$.

Definition 1.23. If $A = (x, y)$ is a point in $\mathbb{R}^2$, then we denote the vector $\overrightarrow{OA}$ by $\begin{bmatrix} x \\ y \end{bmatrix}$.

We can also denote this vector $[x, y]^T$, as we did in section 1.5.3. Poole sometimes just writes $[x, y]$, and when we don't particularly care about the geometric distinction between a point and a vector we will often write (x, y).

However, the vertical orientation is very important for a lot of calculations we will want to do, including the sort of matrix multiplication we used in section 1.5.4. Therefore we will use the vertical form when it isn't terribly inconvenient.

If we want to discuss "a vector" without specifying any coordintes, we will use a single letter, generally either boldface ($\mathbf{v}$) or with an arrow on top ($\vec{v}$).

The vector $\overrightarrow{OO}$ can't really be drawn—it's the vector with zero length—but it is very important. We call it the *zero vector* and write it as $\vec{0}$ or $\mathbf{0}$.

Example 1.24. Suppose $A = (2, 3)$ and $B = (1, 5)$. Then the vector $\overrightarrow{AB}$ has displacement in the x direction of $1 - 2 = -1$, and in the y direction of $5 - 3 = 2$. Thus it is the same as the vector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ which begins at $(0, 0)$ and ends at $(-1, 2)$.

If we want to take the same vector $\overrightarrow{AB}$ and put its initial point at $(-1, 2)$, then the terminal point will have x coordinate $-1 - 1 = -2$ and y-coordinate $2 + 2 = 4$, and thus be at the point $(-2, 4)$.

Geometrically, a vector is a direction and a distance. A natural question to ask is "what happens if we go in the same direction, but twice as far?" Or three times, or five times, or π times as far?

Definition 1.25. If $\mathbf{v}$ is a vector and r is a positive real number, we define *scalar multiplication* by setting $r \cdot \mathbf{v}$ to be a vector with the same direction as $\mathbf{v}$, but with its length stretched by a factor of r.

If r is a negative real number then we define $r \cdot \mathbf{v}$ to be the vector with the opposite direction from $\mathbf{v}$, and length equal to $|r|$ times the length of v.

We define $0 \cdot \mathbf{v} = \mathbf{0}$ to be the zero vector.

Remark 1.26. Notice that this means $-1 \cdot \mathbf{v}$ is a vector of the same length, but pointing in the opposite direction. So $(-1) \cdot \overrightarrow{AB} = \overrightarrow{BA}$.

Example 1.27. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Then we see that $2 \cdot \mathbf{v}$ must go twice as far in the same direction, and thus $2 \cdot \mathbf{v} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$. Similarly, $-2 \cdot \mathbf{v} = \begin{bmatrix} -2 \\ -6 \end{bmatrix}$. Of course, we know that $0 \cdot \mathbf{v} = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Looking at these examples suggests an algebraic rule for scalar multiplication:

Definition 1.28. If $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is a vector and r is a real number, then we define *scalar multiplication* by $b \cdot \mathbf{v} = b \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} bv_1 \\ bv_2 \end{bmatrix}$. We sometimes say that scalar multiplication is given by *componentwise* multiplication.

Notice that this is exactly the scalar matrix multiplication of section 1.5.1.

Example 1.29. If $\mathbf{v} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ then $7 \cdot \mathbf{v} = \begin{bmatrix} 21 \\ 35 \end{bmatrix}$ and $\pi \cdot \mathbf{v} = \begin{bmatrix} 3\pi \\ 5\pi \end{bmatrix}$.

Remark 1.30. It is very important that scalar multiplication combines two different types of information. We have a real number r, which is a "size" without direction. We also have a vector $\mathbf{v}$ which is a magnitude and direction, and we multiply these two things together.

We cannot multiply two vectors to get another vector (outside of some very specific circumstances like the cross product). We can, of course, multiply two scalars together to get another scalar; you have been doing that since elementary school.

Another question to ask about geometric vectors is "what happens if we go in this direction for this distance, and then once we get there, go in that direction for that distance?" In our diagram of the plane, this is represented by taking two vectors and placing them "head-to-tail".

Definition 1.31. If $\mathbf{v} = \overrightarrow{AB}$ and $\mathbf{w} = \overrightarrow{BC}$, then we define *vector addition* by $\mathbf{v} + \mathbf{w} = \overrightarrow{AC}$.

Example 1.32. If $A = (1, 2)$, $B = (3, 1)$, $C = (5, -1)$, then we have $\begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$.

Example 1.33. If $\mathbf{v} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$ then we can set $A = (0, 0), B = (5, 2), C = (1, 3)$ and have $\mathbf{v} = \overrightarrow{AB}$ and $\mathbf{w} = \overrightarrow{BC}$. Then $\mathbf{v} + \mathbf{w} = \overline{AC} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Drawing a picture every time we want to add vectors gets tedious very quickly. Fortunately, vector addition is easy algebraically: we can just do *componentwise addition.*

Definition 1.34. Algebraically, we define addition of vectors by $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$.

You can see that this gives the same result as the head-to-tail method. And again, it is the same as the matrix addition of section 1.5.1.

Remark 1.35. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, we can form a parallelogram with those vectors as two of its sides. We call this the *parallelogram determined by* $\mathbf{u}$ *and* $\mathbf{v}$. In this case, we see that $\mathbf{u} + \mathbf{v}$ is the vector corresponding to the diagram of the parallelogram.

1.4.2 Threespace and $\mathbb{R}^n$

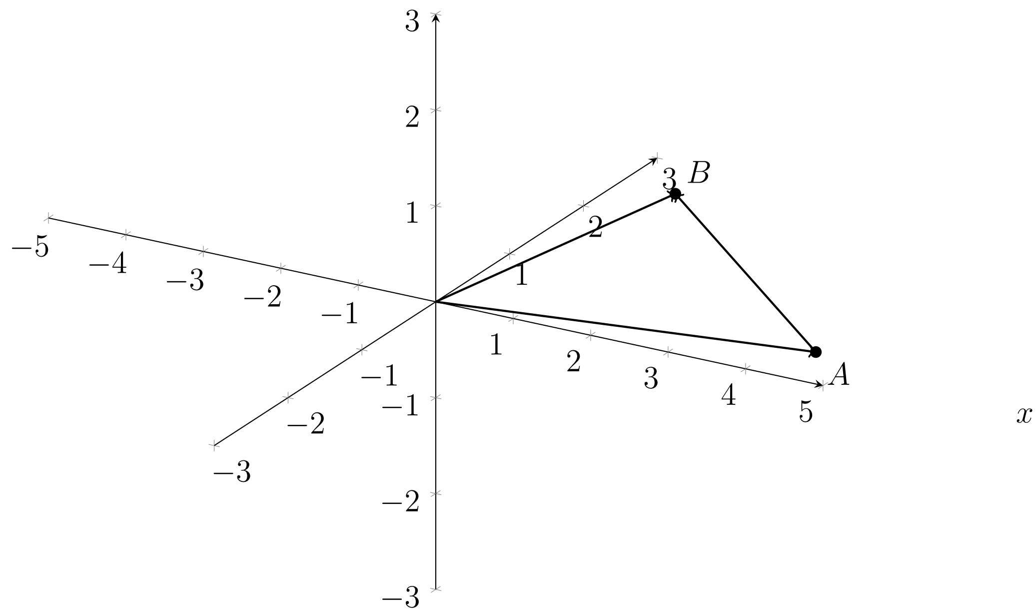
All of the work in section 1.4.1 took place in the "two-dimensional" plane. We can easily extend this work to three-dimensional space. Where each point in the plane requires two coordinates to express, each point in threespace requires three coordinates.

Definition 1.36. We define *Euclidean threespace* to be the three-dimensional space described by three real coordinates. We notate it $\mathbb{R}^3$. The point $(0, 0, 0)$ is called the *origin* and often notated O.

This describes familiar three-dimensional space, in which we all (apparently) live. Just as in the Cartesian plane $\mathbb{R}^2$, we can think about vectors between points.

Example 1.37. Let $A = (3, 2, -1)$ and $B = (5, -2, 3)$. Then we have

$$\overrightarrow{OA} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}, \quad \overrightarrow{OB} = \begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}, \text{ and } \quad \overrightarrow{AB} = \begin{bmatrix} 2 \\ -4 \\ 4 \end{bmatrix}.$$



We can do vector addition and scalar multiplication as before, too.

Example 1.38. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $w = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}$. Then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 5 \\ 0 \\ 6 \end{bmatrix}, \quad 3 \cdot \mathbf{v} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}, \text{ and } \quad (-2) \cdot \mathbf{w} = \begin{bmatrix} -8 \\ 4 \\ -6 \end{bmatrix}.$$

We have so far defined two-dimensional space and three-dimensional space. Geometrically it's hard to go farther, since most of us can't visualize a four- or five-dimensional space. (The Greeks actually argued that while you could raise a number to the second power or the third power, it made no sense to talk about 3^4 since there was no reasonable geometric interpretation. This dispute was only finally resolved in 1637 when René Descartes published a geometric method of taking two line segments, and constructing a line segment whose length was the product of the original lengths; this allowed scholars to multiply two distances and obtain a distance, resolving the philosophical concerns).

But algebraically, there's no difficulty in extending our definitions to higher dimensions and more coordinates in our vectors. (This is probably a large portion of why this course is called "linear algebra" and not "linear geometry").

Definition 1.39. We define *real n-dimensional space* to be the set of n-tuples of real numbers, $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}\}$.

By "abuse of notation" we will also use $\mathbb{R}^n$ to refer to the set of vectors in $\mathbb{R}^n$. We define scalar multiplication and vector addition by

$$r \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} rx_1 \\ rx_2 \\ \vdots \\ rx_n \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Example 1.40. Let $\mathbf{v} = (1, 3, 2, 4)$ and $\mathbf{w} = (5, -1, 2, 8)$ be vectors in $\mathbb{R}^4$. Then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ -1 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 4 \\ 12 \end{bmatrix}, \quad -3 \cdot \mathbf{v} = \begin{bmatrix} -3 \\ -9 \\ -6 \\ -12 \end{bmatrix}.$$

The next question you might ask is "why do we want to talk about $\mathbb{R}^n$?" $\mathbb{R}^2$ and $\mathbb{R}^3$ have obvious geometric interpretations, but it's hard to imagine the geometry of $\mathbb{R}^4$, and far harder to imagine the geometry of $\mathbb{R}^{300}$, or think of what that might describe. I visit very few three hundred dimensional spaces in my life.

And it's true that when we want to talk about "geometry" per se we will find ourselves returning to $\mathbb{R}^2$ and $\mathbb{R}^3$; throughout the course I will be giving low-dimensional examples so you have pictures to mentally reference, and we will do some work on specifically three-dimensional geometry.

But it turns out that a lot of very interesting things we care about "look like" $\mathbb{R}^n$ in a very specific way. We've already seen one example, in the set of solutions to a system of linear equations. Even if a four-dimensional space doesn't make much sense, a set of equations with four variables certainly does!

But there are many other examples, and in section 3.1 we will talk about what it means to look like $\mathbb{R}^n$ in this way.

1.4.3 Linear Combinations and Spans

Definition 1.41. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a list of vectors in $\mathbb{R}^n$, then a *linear combination* of of the vectors in S is a vector of the form

$$\sum_{i=1}^{n} a_i \mathbf{v}_i = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

where $a_i \in \mathbb{R}$ are real numbers, called the *weights* of the linear combination.

Geometrically, a linear combination of vectors represents some destination you can reach only going in the directions of your chosen vectors (for any distance. So if I can go north or west, any distance"northwest" will be a linear combination of those vectors. And "southeast" will as well, since we can always go in the "opposite" direction. But "up" will not be.

Remark 1.42. This is a "linear" combination because it combines the vectors in the same way a line or plane does—adding all the vectors together, but with some coefficient. We will revisit this terminology in the next section when we discuss linear functions.

It's totally possible to have a linear combination of infinitely many vectors. But studying these requires some sense of convergence, and thus calculus/analysis. So we won't talk about it in *this* class, except for the occasional aside.

Example 1.43. Let $S = \{(1,0,0),(0,1,0)\}$. Then we see that

$$(3,2,0) = 3(1,0,0) + 2(0,1,0) \quad \text{and} \quad (-5, 3\pi, 0) = -5(1,0,0) + 3\pi(0,1,0)$$

are linear combinations of vectors in S.

However, $(1,1,1)$ is *not* a linear combination of vectors in S. If it were, we would have

$$a(1,0,0) + b(0,1,0) = (1,1,1)$$

and thus $(a,b,0) = (1,1,1)$ which cannot happen for any $a, b \in \mathbb{R}$.

If we have a set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and another vector $\mathbf{b}$, then asking if $\mathbf{b}$ is in the span of S is the same as asking whether the equation

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{b}$$

has a solution. And this is just a system of linear equations, so we can answer this with a row reduction.

We can also ask for the set of *all* linear combinations of a set of vectors. Geometrically, this is the set of places we can get to only by going in the direction of vectors in S.

Definition 1.44. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. We say the *span* of S is the set of all linear combinations of vectors in S, and write it $\operatorname{span}(S)$ or $\operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$.

For notational consistency, we define the span of the empty set $\operatorname{span}(\{\})$ to be the set $\{\mathbf{0}\}$ that contains only the zero vector.

Example 1.45. As before, take $S = \{(1,0,0), (0,1,0)\}$. Then

$$\operatorname{span}(S) = \{a(1,0,0) + b(0,1,0)\} = \{(a,b,0)\}.$$

Now let $T = \{(3,2,0), (13,7,0)\}$. Then

$$\operatorname{span}(T) = \{a(3,2,0) + b(13,7,0)\} = \{(3a+13b, 2a+7b, 0)\}.$$

Notice that these are actually the same set! The first spanning set "looks" nicer, but it's hard to make this sense of "nice" mathematically precise. We'll do our best, but won't really get there until section 6.

1.5 Matrix Algebra

So far we've treated matrices as just being a convenient way to write down a bunch of numbers. But matrices are interesting mathematical objects in their own right, and we can do a lot of useful calculations with them.

1.5.1 Simple Operations

We want to start with a couple of simple operations. Neither of these operations really depend on the structure of the matrix; they treat the matrix as a list of numbers.

Definition 1.46. If $A = (a_{ij})$ is an $m \times n$ matrix, and $r \in \mathbb{R}$ is a real number, then we can multiply each entry of the matrix A by the real number R. This is called *scalar multiplication* and we say that r is a *scalar*.

$$rA = (ra_{ij}) = \begin{bmatrix} ra_{11} & ra_{12} & \dots & ra_{1n} \\ ra_{21} & ra_{22} & \dots & ra_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ra_{m1} & ra_{m2} & \dots & ra_{mn} \end{bmatrix}.$$

Definition 1.47. If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices, we can add the two matrices by adding each individual pair of coordinates together.

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

Example 1.48.

$$3\begin{bmatrix} 2 & 5 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 15 \\ -3 & 12 \end{bmatrix} \qquad \begin{bmatrix} 4 & 1 & 3 \\ -2 & 5 & -1 \end{bmatrix} + \begin{bmatrix} -2 & 7 & 5 \\ 1 & -6 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 8 & 8 \\ -1 & -1 & 3 \end{bmatrix}$$

1.5.2 Matrix Multiplication

Definition 1.49. If $A \in M_{\ell \times m}$ and $B \in M_{m \times n}$, then there is a matrix $AB \in M_{\ell \times n}$ whose ij element is

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

If you're familiar with the dot product, you can think that the ij element of AB is the dot product of the ith row of A with the jth column of b.

Note that A and B don't have to have the same dimension! Instead, A has the same number of columns that B has rows. The new matrix will have the same number of rows as A and the same number of columns as B.

Example 1.50.

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & -1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 3 \cdot 3 & 1 \cdot (-1) + 3 \cdot 2 \\ 2 \cdot 5 + 4 \cdot 3 & 2 \cdot (-1) + 4 \cdot 2 \end{bmatrix} = \begin{bmatrix} 14 & 5 \\ 22 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 6 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 5 \\ 4 & 1 & 6 \end{bmatrix} = \begin{bmatrix} 4 \cdot 3 + 6 \cdot 4 & 4 \cdot 1 + 6 \cdot 1 & 4 \cdot 5 + 6 \cdot 6 \\ 2 \cdot 3 + 1 \cdot 4 & 2 \cdot 1 + 1 \cdot 1 & 2 \cdot 5 + 1 \cdot 6 \end{bmatrix} = \begin{bmatrix} 36 & 10 & 56 \\ 10 & 3 & 16 \end{bmatrix}.$$

Matrix multiplication is *associative*, by which we mean that $(AB)C = A(BC)$.

Matrix multiplication is not commutative: in general, it's not even the case that AB and BA both make sense. If $A \in M_{3 \times 4}$ and $B \in M_{4 \times 2}$ then AB is a 3×2 matrix, but BA isn't a thing we can compute. But even if AB and BA are both well-defined, they are not equal.

Example 1.51.

$$\begin{bmatrix} 3 & 5 & 1 \\ -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 5 \cdot 1 + 1 \cdot 4 & 3 \cdot 1 + 5 \cdot 3 + 1 \cdot 1 \\ -2 \cdot 2 + 0 \cdot 1 + 2 \cdot 4 & -2 \cdot 1 + 0 \cdot 3 + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 15 & 19 \\ 4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 3 & 5 & 1 \\ -2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 + 1 \cdot (-2) & 2 \cdot 5 + 1 \cdot 0 & 2 \cdot 1 + 1 \cdot 2 \\ 1 \cdot 3 + 3 \cdot (-2) & 1 \cdot 5 + 3 \cdot 0 & 1 \cdot 1 + 3 \cdot 2 \\ 4 \cdot 3 + 1 \cdot (-2) & 4 \cdot 5 + 1 \cdot 0 & 4 \cdot 1 + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 4 \\ -3 & 5 & 7 \\ 10 & 20 & 6 \end{bmatrix}.$$

Particularly nice things happen when our matrices are square. Any time we have two $n \times n$ matrices we can multiply them by each other in either order (though we will still get different things each way!).

Example 1.52.

$$\begin{bmatrix} 4 & 1 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 8 & -13 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ -3 & 5 \end{bmatrix} = \begin{bmatrix} -7 & 4 \\ 10 & -9 \end{bmatrix}.$$

However, matrix multiplication does satisfy the *distributive* and *associative* properties.

Fact 1.53. *If $A \in M_{\ell \times m}$ and $B, C \in M_{m \times n}$ then $A(B + C) = AB + AC$.*

If $A \in M_{\ell \times m}, B \in M_{m \times n}, C \in M_{n \times p}$ then $(AB)C = A(BC)$.

Example 1.54. Let

$$A = \begin{bmatrix} 4 & 1 \\ -3 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} \qquad C = \begin{bmatrix} 3 & 2 \\ 1 & -5 \end{bmatrix}.$$

Then we have

$$AB = \begin{bmatrix} -3 & 2 \\ 8 & -13 \end{bmatrix} \qquad AC = \begin{bmatrix} 13 & 3 \\ -4 & -31 \end{bmatrix} \qquad AB + AC = \begin{bmatrix} 10 & 5 \\ 4 & -44 \end{bmatrix}$$

$$B + C = \begin{bmatrix} 2 & 3 \\ 2 & -7 \end{bmatrix} \qquad A(B + C) = \begin{bmatrix} 10 & 5 \\ 4 & -44 \end{bmatrix}.$$

Thus we see $AB + AC = A(B + C)$.

We can similarly compute

$$AB = \begin{bmatrix} -3 & 2 \\ 8 & -13 \end{bmatrix} \qquad (AB)C = \begin{bmatrix} -3 & 2 \\ 8 & -13 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & -5 \end{bmatrix} = \begin{bmatrix} -7 & -16 \\ 11 & 81 \end{bmatrix}$$

$$BC = \begin{bmatrix} -2 & -7 \\ 1 & 12 \end{bmatrix} \qquad A(BC) = \begin{bmatrix} 4 & 1 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} -2 & -7 \\ 1 & 12 \end{bmatrix} = \begin{bmatrix} -7 & -16 \\ 11 & 81 \end{bmatrix}$$

1.5.3 Transposes

Definition 1.55. If A is a $m \times n$ matrix, then we can form a $n \times m$ matrix B by flipping A across its diagonal, so that $b_{ij} = a_{ji}$. We say that B is the *transpose* of A, and write $B = A^T$.

If $A = A^T$ we say that A is *symmetric*. (Symmetric matrices must always be square).

Example 1.56.

$$\text{If } A = \begin{bmatrix} 1 & 3 & 5 \\ -1 & 4 & 2 \end{bmatrix} \text{ then } A^T = \begin{bmatrix} 1 & -1 \\ 3 & 4 \\ 5 & 2 \end{bmatrix}.$$

$$\text{If } B = \begin{bmatrix} 5 & 3 \\ 3 & -2 \end{bmatrix} \text{ then } B^T = \begin{bmatrix} 5 & 3 \\ 3 & -2 \end{bmatrix}$$

and thus B is symmetric.

Fact 1.57.
- $(A^T)^T = A$.
- $(A + B)^T = A^T + B^T$.
- $(rA)^T = rA^T$.
- *If* $A \in M_{\ell \times m}$ *and* $B \in M_{m \times n}$ *then* $(AB)^T = B^T A^T$.

1.5.4 Matrices and Systems of Equations

We will do a lot with matrices in the future (a linear algebra class that doesn't cover general vector spaces is often called a matrix algebra class). In the current context we mostly want it to make it easier to talk about systems of equations.

Let

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

be a system of linear equations. Then $A = (a_{ij}) \in M_{m \times n}$ is its coefficient matrix, and $\mathbf{b} = (b_1, \dots, b_m)$ is an element of $\mathbb{R}^m$, but we can also think of it as a $m \times 1$ matrix $b = [b_1, \dots, b_m]^T$. If we take $\mathbf{x} = [x_1, \dots, x_n]^T$ to be a $n \times 1$ matrix, we can rewrite our linear system as the equation

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

which is certainly much easier to write down.

Example 1.58. If $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $\mathbf{b} = [4, 6]^T$, then the equation $A\mathbf{x} = \mathbf{b}$ is

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$
$$\begin{bmatrix} x + 3y \\ 2x + 4y \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$
$$x + 3y = 4$$
$$2x + 4y = 6$$

1.6 Homogeneous systems and sets of solutions

There's one particular category of systems of linear equations that's especially important to us, and will lead into the main subject matter of the course.

Definition 1.59. The $n \times 1$ matrix $\mathbf{0} = [0, \ldots, 0]^T$ whose entries are all zero is called the *zero vector*.

A system of linear equations $A\mathbf{x} = \mathbf{b}$ is called *homogeneous* if $\mathbf{b} = \mathbf{0}$, that is, if the constant term in each equation is zero. Otherwise, it is *non-homogeneous*.

It's pretty clear that every homogeneous system has at least one solution: the solution where every variable is equal to zero. It may have many more solutions than that.

Definition 1.60. For a given matrix A, the subspace of solutions to the equation $A\mathbf{x} = \mathbf{0}$ is called the *nullspace* $N(A)$ or the *kernel* $\ker(A)$ of the matrix A.

Example 1.61. Find the null space of $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix}$.

We row reduce the matrix

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

We see that x_3 and x_4 are fixed variables, and x_1, x_2 are determined by x_3 and x_4. (You could of course do this the other way around). Then we have $x_1 = x_3 - x_4$ and $x_2 = x_4 - 2x_3$.

Thus $N(A) = \{(\alpha - \beta, \beta - 2\alpha, \alpha, \beta)\}$, where α and β are parameters for the free variables. (You could just as well write $\{(x_3 - x_4, x_4 - 2x_3, x_3, x_4)\}$, and it would mean the same thing.)

We can also represent this in parametric vector form $\{\alpha(1,-2,1,0)+\beta(-1,1,0,1)\}$. This is just another way of representing the same information, but it makes it clear that we have two different direcitons we can move in: the solution set is the span of $\{(1,-2,1,0),(-1,1,0,1)\}$.

We can see that if we add together two solutions to this system of equations, we will get another. In fact, this must be true of any homogeneous system.

Proposition 1.62 (Homogeneity). *Suppose $A\mathbf{x}=\mathbf{0}$ is a homogeneous sytem of linear equations. Then:*

1. *$\mathbf{0}$ is a solution to the system.*
2. *If $\mathbf{x}_1$ and $\mathbf{x}_2$ are solutions to this system, then $\mathbf{x}_1+\mathbf{x}_2$ is a solution.*
3. *If $\mathbf{x}$ is a solution to this system, and r is a real number, then $r\mathbf{x}$ is a solution.*

Remark 1.63. We can rephrase this result: for any matrix A, we have

1. $\mathbf{0} \in N(A)$
2. If $\mathbf{x}_1, \mathbf{x}_2 \in N(A)$ then $\mathbf{x}_1+\mathbf{x}_1 \in N(A)$
3. If $r \in \mathbb{R}$ and $\mathbf{x} \in N(A)$ then $r\mathbf{x} \in N(A)$.

This says exactly the same thing, but puts the emphasis on the matrix A rather than on the equation $A\mathbf{x}=\mathbf{0}$.

Proof. 1. Calculation confirms that $A\mathbf{0}=\mathbf{0}$.

2. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are solutions, then $A\mathbf{x}_1=\mathbf{0}$ and $A\mathbf{x}_2=0$, so we have

$$A(\mathbf{x}_1+\mathbf{x}_2)=A\mathbf{x}_1+A\mathbf{x}_2=\mathbf{0}+\mathbf{0}=\mathbf{0}.$$

Thus $\mathbf{x}_1+\mathbf{x}_2$ is a solution.

3. If $\mathbf{x}$ is a solution and $r \in \mathbb{R}$, then

$$A(r\mathbf{x})=rA\mathbf{x}=r\mathbf{0}=\mathbf{0}.$$

Thus $r\mathbf{x}$ is a solution.

□

In contrast, the set of solutions to a non-homogeneous system $A\mathbf{x}=\mathbf{b}$ where $\mathbf{b}\neq\mathbf{0}$ never has these nice properties.

1. The zero vector is never a solution, since $A\mathbf{0} = \mathbf{0} \neq \mathbf{b}$.

2. Adding two solutions doesn't give you another solution: $A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{b} + \mathbf{b} = 2\mathbf{b} \neq \mathbf{b}$.

3. Multiplying a solution by a scalar doesn't give another solution: $Ar\mathbf{x} = r\mathbf{b} \neq \mathbf{b}$ unless $r = 1$.

So there's something special about homogeneous systems, which we will discuss in more detail in 3.1.

But even though the set of solutions to a non-homogeneous system doesn't have the nice properties of proposition 1.62, we can still say a lot about what it looks like.

Proposition 1.64. *Suppose $A\mathbf{x} = \mathbf{b}$ is a non-homogeneous linear system.*

If $U = N(A))$ and $\mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{b}$, then the set of solutions to the system $A\mathbf{x} = \mathbf{b}$ is the set

$$N(A) + \mathbf{x}_0 = \{\mathbf{y} + \mathbf{x}_0 : \mathbf{y} \in N(A)\}.$$

Remark 1.65. This looks much clearer in the context of parametric equations. If $N(A)$ is given by $\{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2\}$, and $\mathbf{x}_0$ is one solution to $A\mathbf{x} = \mathbf{b}$, then the entire solution set is $\{x_0 + \alpha_2\mathbf{v}_1 + \alpha_2\mathbf{v}_2\}$. We have the same parametric vectors, but rather than adding them onto the origin, we add them on to our test solution $\mathbf{x}_0$.

Geometrically, this corresponds to having a solution set that is *shaped* the same, but shifted over in space.

Proof. We want to show that two sets are equal, so we show that each is a subset of the other.

First, suppose that $\mathbf{x}_1$ is a solution to $A\mathbf{x}_1 = \mathbf{b}$. Then we have

$$\begin{aligned} b &= A\mathbf{x}_0 \\ b &= A\mathbf{x}_1 \\ b - b &= A\mathbf{x}_1 - A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_0) \\ \mathbf{0} &= A(\mathbf{x}_1 - \mathbf{x}_0). \end{aligned}$$

Thus $\mathbf{y} = \mathbf{x}_1 - \mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{0}$, and then $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y}$ for some $\mathbf{y} \in U$.

Conversely, suppose $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y}$ for some $\mathbf{y} \in U$. Then

$$A\mathbf{x}_1 = A(\mathbf{x}_0 + \mathbf{y}) = A\mathbf{x}_0 + A\mathbf{y} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Thus x_1 is a solution to $A\mathbf{x} = \mathbf{b}$. □

Remark 1.66. Notice this did not depend on the specific matricx, or even really the fact that A is a matrix at all; it only depends on the ability to distribute matrix multiplication across sums of vectors. Operations with this property are called "linear" and we will discuss them in much more detail in section 1.8.

Definition 1.67. Suppose $A\mathbf{x} = \mathbf{b}$ is a system of linear equations. We call the equation $A\mathbf{x} = \mathbf{0}$ the associated homogeneous system of linear equations. That is, the associated homogeneous system has the same coefficients for all the variables, but the constants are all zero.

Thus proposition 1.64 lets us understand the set of solutions to a non-homogeneous system based on the solutions to the associated homogeneous system.

Example 1.68. Let's find a set of solutions to the system

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 + 3x_3 &= 6 \\ 2x_1 + 3x_2 + 4x_2 &= 9. \end{aligned}$$

Gaussian elimination gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & 3 & 6 \\ 2 & 3 & 4 & 9 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Taking $x_3 = \alpha$ as a free variable, our solution set is $\{(\alpha, 3-2\alpha, \alpha)\} = \{(0,3,0)+\alpha(1,-2,1)\}$. Indeed, we see that this set corresponds to elements of the vector space spanned by $\{(1,-2,1)\}$, plus a specific solution $(0,3,0)$.

Alternatively, we could have solved the homogeneous system first, and seen that the solution was $x_1 - x_3 = 0, x_2 + 2x_3 = 0$, telling us that $N(A) = \{\alpha(1,-2,1)\}$. Then we just need to find a solution; to my eyes the obvious solution is $(1,1,1)$. So our theorem tells us that the solution set is $\{(1,1,1) + \alpha(1,-2,1)\}$. This may not *look* like the solution we got before, but it is in fact the same set, since $(1,1,1) = (0,3,0) + (1,-2,1)$.

Example 1.69. Now consider the system

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

It's easy enough to see that this system has no solutions, since the sum of the first two equations should be the third.

This at first might seem concerning, since $N(A)$ is never empty. But our proposition assumed that there was at least one solution to the non-homogeneous system; when there are no solutions, the proposition doesn't actually say anything. But *if* any solution exists, proposition 1.64 tells us that the set of solutions is just the nullspace of A, plus an offset.

Example 1.70. Let's find the set of solutions to

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

We form the matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & -2 & 2 & 4 \\ 1 & 2 & -1 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -3 & 1 & 4 \\ 0 & 1 & -2 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & -3 & 1 & 4 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & -5 & 10 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & 0 & 3 & -2 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

giving us the sole solution $x_1 = 4, x_2 = -2, x_3 = -2$.

If we look at the corresponding homogeneous system, we see that we can reduce the matrix to $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and thus the sole solution to the homogeneous system of equations is $x_1 = x_2 = x_3 = 0$. Then every solution to our non-homogeneous system is a solution to our homogeneous system plus some vector in $\{\vec{0}\}$. Since there is only one vector in that set, there is only one solution to our system.

1.7 Linear Independence

When we write down a vector equation, we can think of it as telling us something about the solution set, but we can also ask what it tells us about the vectors.

We've already seen one example of this process in section 1.4.3 where we talked about spanning sets. We know that $\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ if and only if there's a solution to the vector equation $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{b}$. So this tells us something about the set of vectors.

But sometimes we want to know if our vectors have any redundancies. If we are using the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \right\}$$

then we have a lot of vectors, but they're not really adding any extra information, or increasing the span, or letting us reach new points that we couldn't reach before. We'd like to have a way to detect whether this is happening.

Definition 1.71. We say a set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is *linearly independent* if the only scalars solving the equation

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

are $a_1 = \cdots = a_n = 0$.

If a set of vectors is not linearly independent, we call it *linearly dependent* and there is a *linear dependence* relationship among the vectors.

Example 1.72.

1. The set $S = \{(1,0,0), (0,1,0), (0,0,1)\}$ is linearly independent: suppose

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

   Then we have the system of equations $a = 0, b = 0, c = 0$ and thus all the scalars are zero.

2. The set $S = \{(1,0,0), (0,1,0)\}$ is linearly independent. Suppose

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}.$$

   Then we have the system of equations $a = 0, b = 0$ and thus all the scalars are zero.

3. The set $S = \{(1,0,0),(0,1,0),(1,1,0)\}$ is not linearly independent, since

$$1\begin{bmatrix}1\\0\\0\end{bmatrix}+1\begin{bmatrix}0\\1\\0\end{bmatrix}+(-1)\begin{bmatrix}1\\1\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}=\mathbf{0}.$$

4. Any set containing the zero vector is linearly dependent, since $1\cdot\mathbf{0}=\mathbf{0}$ but $1\neq 0$.

As before, we see that each problem is really asking for the solution to a system of equations. But we don't need to find a solution for any possible constants. Instead, we just want to see if there's more than one solution to the equation $A\mathbf{x}=\mathbf{0}$.

Example 1.73. Let $S=\{(3,5,1),(3,5,3),(1,1,1)\}$. Then we want to know if there exist a,b,c such that

$$a\begin{bmatrix}3\\5\\1\end{bmatrix}+b\begin{bmatrix}3\\5\\3\end{bmatrix}+c\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

The associated matrix is $A=\begin{bmatrix}3&2&1\\5&4&1\\1&3&1\end{bmatrix}$ and we want to solve the homogeneous system, so we just have to reduce the matrix: we get

$$\begin{bmatrix}3&3&1\\5&5&1\\1&3&1\end{bmatrix}\to\begin{bmatrix}0&-6&-2\\0&-10&-4\\1&3&1\end{bmatrix}\to\begin{bmatrix}1&3&1\\0&3&1\\0&2&0\end{bmatrix}\to\begin{bmatrix}1&0&0\\0&3&1\\0&1&0\end{bmatrix}\to\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

And thus we have the unique solution $a=0,b=0,c=0$. Thus S is linearly independent.

There are a few ways we can think about linear independence. One is that a linearly independent set is one where the zero vector can be expressed uniquely—$\mathbf{0}$ is in the span of any set, but it is only in the span of a linearly independent set in one way. In fact, this is enough to make *every* vector expressed uniquely.

We can interpret linear independence geometrically. Generally, any one vector defines a line containing it and the origin. Two vectors in general define a plane, three vectors a threespace, and so on. A set is linearly independent if the linear space it defines is as big as you would expect. A set is linearly dependent if the set is smaller—if, say, you have points but they're all on the same line through the origin, so you don't actually get a whole plane.

Lemma 1.74. *A set* $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ *is linearly dependent if and only if some element can be written as a linear combination of the others.*

Proof. Without loss of generality, assume $\mathbf{v}_1$ can be written as a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_n$. (That is, we're assuming one of the vectors can be written as a linear combination of the others, and since order doesn't matter we can assume that it's $\mathbf{v}_1$ to keep the notation simple). Then we have

$$\begin{aligned} \mathbf{v}_1 &= a_2\mathbf{v}_2 + \dots a_n\mathbf{v}_n \\ \mathbf{v}_1 - \mathbf{v}_1 &= a_2\mathbf{v}_2 + \dots a_n\mathbf{v}_n - \mathbf{v}_1 \\ \mathbf{0} &= (-1)\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n. \end{aligned}$$

Then we have written $\mathbf{0}$ as a nontrivial linear combination of elements of S, and thus S is linearly dependent.

Conversely, suppose S is linearly dependent. Then there are a_i not all zero such that

$$\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n.$$

We know not all the a_i are zero, so assume without loss of generality that $a_1 \neq 0$. Then we have

$$\begin{aligned} -a_1\mathbf{v}_1 &= a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \\ \mathbf{v}_1 &= \frac{-a_2}{a_1}\mathbf{v}_2 + \cdots + \frac{-a_n}{a_1}\mathbf{v}_n \end{aligned}$$

and thus we can write $\mathbf{v}_1$ as a linear combination of the other vectors in S. $\square$

Corollary 1.75. $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ *is linearly dependent if and only if there is some* $\mathbf{v}_i$ *such that* $\operatorname{span}(S) = \operatorname{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_n\}$.

In practice this is how we prefer to test for linear independence: we try to write one vector as a linear combination of the others. Sometimes this is easy and we're done. Other times this is difficult, or we become convinced it's not possible, and then we have to go back to solving linear equations.

Example 1.76.

1. Let $S = \{(1,1,1), (1,1,0), (0,0,1)\}$. We see that $(1,1,1) = (1,1,0) + (0,0,1)$ so this set is linearly dependent.

2. Let $S = \{(1,1,1), (1,1,0), (0,1,1)\}$. It might look like this is similar, and we could write $(1,1,1)$ somehow as a combination of the other two. But we see that's not

actually possible. In fact we write

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a+b \\ a+b+c \\ b+c \end{bmatrix}$$

and this gives us the system

$$0 = a + b \qquad 0 = a + b + c \qquad 0 = b + c$$

with associated matrix

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

Thus we get the unique solution $a = b = c = 0$ and so the set is linearly independent.

3. Let $S = \{(1,1,1), (1,1,0), (2,3,1), (0,1,1)\}$. To show linear dependence, we might want to show that one vector is a sum of the others. In fact we cannot write $(1,1,1)$ as a linear combination of the other vectors:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a+2b \\ a+3b+c \\ b+c \end{bmatrix}$$

gives the system

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

and the last row is a contradiction $0 = 1$. Thus there is no solution to this system; we cannot write $(1,1,1)$ as a linear combination of the other vectors.

But this doesn't mean that the vectors are linearly independent. Corollary 1.75 says that the vectors are independent if and only if *some* vector is a linear combination of the others. There is in fact a vector that is a sum of some of the others: we see that

$$2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

and thus set S is linearly dependent.

If we didn't see this directly, we could set up the matrix associated to all four vectors:

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

We have the fourth column as a free variable, so there is more than one solution. Thus the set is not linearly independent. In particular, if we have a linear combination of these vectors that sums to $\mathbf{0}$, it satisfies $a = 0, b = 2d, c = -d$.

Again we see we're looking at a matrix whose columns are the vectors in the set in question. We want to see if any non-trivial linear combination of our vectors is equal to the zero vector, and that leads to solving a homogeneous system of equations. Thus we get the following result:

Proposition 1.77. *Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbb{R}^m$. Let A be the matrix whose columns are the vectors $\mathbf{v}_i$. Then S is linearly independent if and only if $N(A) = \{\mathbf{0}\}$.*

Proof. By definition, S is linearly independent if and only if the only solution to $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ is $(0, 0, \dots, 0)$. But this is a homogeneous system of linear equations whose matrix is A, so $N(A)$ is the set of solutions to this system of equations. Thus S is linearly independent if and only if the only element of $N(A)$ is the zero vector. □

We again conclude with a fact about linear independence.

Proposition 1.78. *If $S \subseteq T$ and T is linearly independent, then S is also linearly independent.*

Proof. Suppose $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq \{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1, \dots, \mathbf{u}_m\} = T$, and T is linearly independent. Now suppose their are scalars a_i such that

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}.$$

Then we have

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n + 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_m = 0$$

and since T is linearly independent, we have $a_i = 0$ for every a_i. Thus we see that S is linearly independent.

□

1.8 Linear Transformations

Definition 1.79. Let let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a function with domain that takes in vectors in $\mathbb{R}^n$ and outputs vectors in $\mathbb{R}^m$. We say L is a *linear transformation* if:

1. Whenever $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n$, then $L(\mathbf{u}_1 + \mathbf{u}_2) = L(\mathbf{u}_1) + L(\mathbf{u}_2)$.

2. Whenever $\mathbf{u} \in \mathbb{R}^n$ and $r \in \mathbb{R}$, then $L(r\mathbf{u}) = rL(\mathbf{u})$.

Example 1.80. If A is a $m \times n$ matrix, then A gives us a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$, given by $A(\mathbf{x}) = A\mathbf{x}$.

If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ then it gives a transformation $L : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + 2y \\ 3x + 4y \\ 5x + 6y \end{bmatrix}.$$

Geometrically, a linear transformation can stretch, rotate, and reflect, but it cannot bend or shift.

Example 1.81. Consider the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by a rotation of ninety degrees counterclockwise. We can see by drawing pictures that the sum of two rotated vectors is the rotation of the sum of the vectors, and that the rotation of a streched vector is the same as the strech of a rotated vector. So this is a linear transformation.

Example 1.82. A *translation* is a function defined by $f(\mathbf{x}) = \mathbf{x} + \mathbf{u}$ for some fixed vector $\mathbf{u}$. (Geometrically, it corresponds to sliding or translating your input in the direction and distance of the vector $\mathbf{u}$).

This is *not* a linear transformation. For instance, $f(r\mathbf{x}) = r\mathbf{x} + \mathbf{u} \neq r(\mathbf{x} + \mathbf{u}) = rf(\mathbf{x})$ unless $\mathbf{u} = \mathbf{0}$.

Example 1.83. The function $f(x) = x^2$ is not a linear transformation from $\mathbb{R}$ to $\mathbb{R}$, since $f(2x) = (2x)^2 = 4x^2 \neq 2x^2 = 2f(x)$.

Example 1.84. Define a function $L : \mathbb{R}^3 \to \mathbb{R}^2$ by $L(x, y, z) = (x + y, 2z - x)$. We check:

$$\begin{aligned}
L((x_1, y_1, z_1) + (x_2, y_2, z_2)) &= L(x_1 + x_2, y_1 + y_2, z_1 + z_2) \\
&= (x_1 + x_2 + y_1 + y_2, 2z_1 + 2z_2 - x_1 - x_2) \\
&= (x_1 + y_1, 2z_1 - x_1) + (x_2 + y_2, 2z_2 - x_2) \\
&= L(x_1, y_1, z_1) + L(x_2, y_2, z_2). \\
L(r(x, y, z)) &= L(rx, ry, rz) = (rx + ry, 2rz - rx) = \\
&= r(x + y, 2z - x) = rL(x, y, z).
\end{aligned}$$

Thus L is a linear transformation by definition.

1.8.1 Representing linear transformations as matrices

We can describe linear transformations in a bunch of ways, but we always want a formula. It turns out that *all* linear transformations can be represented by matrices.

In essence, we can represent a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ with a matrix because linear transformations are so regular. If we know what happens to each coordinate, then we know what happens to any vector input at all.

Example 1.85. Let $A = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix}$ be a matrix, and thus a linear transformation $\mathbb{R}^3 \to \mathbb{R}^2$. Let's see what happens to each element of the standard basis for $\mathbb{R}^3$.

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

We notice that the image of the standard basis elements are just the columns of the matrix! This isn't a coincidence; the columns of our matrix are telling us exactly where our basis vectors go.

Proposition 1.86. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is an $m \times n$ matrix A such that $L(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^n$.*

In particular, the jth column vector of A is given by $\mathbf{c}_j = L(\mathbf{e}_j)$.

Proof. According to the theorem statement, we know that $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$. So we just need to check that this matrix gives us the linear transformation L.

First we show that our matrix does the right things on the standard basis vectors. We see that

$$A\mathbf{e}_j = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_j & \dots & \mathbf{c}_n \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{c}_j = L(\mathbf{e}_j).$$

Now let $\mathbf{u} \in \mathbb{R}^n$ be any vector. Then we know we can write $\mathbf{u} = \sum_{i=1}^n u_i \mathbf{e}_i$ since every element is some linear combination of basis vectors. Thus we have

$$\begin{aligned} A\mathbf{u} = A\left(\sum_{i=1}^n u_i\mathbf{e}_i\right) &= \sum_{i=1}^n Au_i\mathbf{e}_i = \sum_{i=1}^n u_i A\mathbf{e}_i = \sum_{i=1}^n u_i L(\mathbf{e}_i) && \text{by the previous computation} \\ &= \sum_{i=1}^n L(u_i\mathbf{e}_i) && \text{scalar multiplication} \\ &= L\left(\sum_{i=1}^n u_i\mathbf{e}_i\right) && \text{additivity} \\ &= L(\mathbf{u}). \end{aligned}$$

□

Example 1.87. Let's look at the linear transformation from earlier, of a 90 degree rotation counterclockwise. This is a transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$, so we can find a 2×2 matrix representing it. Let's call the map $R_{\pi/2}$.

By geometry, we see that $R_{\pi/2}(\mathbf{e}_1) = (0, 1) = \mathbf{e}_2$, and that $R_{\pi/2}(\mathbf{e}_2) = (-1, 0) = -\mathbf{e}_1$. Thus the matrix is $\begin{bmatrix} 0 & -1 \\ 1 & 0. \end{bmatrix}$

Let's generalize to any rotation; let R_θ be the rotation counterclockwise by θ. To see what happens we have to draw the unit circle; we compute that $R_\theta(\mathbf{e}_1) = (\cos\theta, \sin\theta)$, and $R_\theta(\mathbf{e}_2) = (\cos(\theta + \pi/2), \sin(\theta + \pi/2) = (-\sin(\theta), \cos(\theta)$. Thus the matrix of R_θ is $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Example 1.88. Define a linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^3$ by $L(x, y) = (x + y, x - y, 2x)$. First we should check that this is in fact a linear transformation, but I won't do that here.

We need to check the image of $\mathbf{e}_1$ and $\mathbf{e}_2$. We see that

$$L(\mathbf{e}_1) = L(1, 0) = (1, 1, 2)$$
$$L(\mathbf{e}_2) = L(0, 1) = (1, -1, 0).$$

Thus the matrix of L is

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

We can check this by computing

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

which is exactly what we should get.

1.8.2 One-to-one and Onto transformations

We want to know what sort of outputs we can get from a given transformation.

Definition 1.89. Let $L\mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say that $\mathbf{v} \in \mathbb{R}^m$ is in the *image* of L if there is a $\mathbf{u} \in \mathbb{R}^n$ such that $L(\mathbf{u}) = \mathbf{v}$. The *image of* L is the set $\{L(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^n\}$.

We say the *kernel* of L is the set $\{\mathbf{u} \in \mathbb{R}^n : L(\mathbf{u}) = \mathbf{0}\}$.

If we know the kernel and image of a transformation, we understand a lot of its properties. These are both easy to find if we have the matrix of our transformation.

Example 1.90. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation given by the matrix

$$A = \begin{bmatrix} 5 & 8 & 7 \\ 0 & 1 & -1 \\ 1 & 3 & 0 \end{bmatrix}.$$

What do the kernel and image look like?

The kernel is the set of vectors $\mathbf{u}$ so that $A\mathbf{u} = \mathbf{0}$. This is just the nullspace of A! So we can row reduce A to get

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

Thus the kernel $\ker(A) = \{(-3x_3, x_3, x_3)\}$.

The image is a little tricker; it's the set of vectors $\mathbf{v}$ for which $A\mathbf{u} = \mathbf{v}$ has a solution. This is precisely the span of the columns of A. From the same row reduction, we can see that the columns are linearly dependent; the third is redundant. So the image is the span of $\{(5,0,1),(8,1,3)\}$.

We're especially interested in the cases where these solutions have to exist and are unique.

Definition 1.91.

- A function f is *one-to-one* or *injective* if it has the property that: if $f(x) = f(y)$ then $x = y$. This tells us that anything in the image of f is only in the image once.
- A function $f : A \to B$ is *onto* or *surjective* if the image of f is B. That is, f is onto if for every $b \in B$ there is an $a \in A$ with $f(a) = b$. This tells us we can reach every element of the codomain from some element of the domain.
- A function f is *bijective* if it is both one-to-one and onto.

Thus a one-to-one function has unique solutions for $L(\mathbf{u}) = \mathbf{v}$ when solutions exist; an onto function is one where a solution always exists. If L is bijective, then there is always a unique solution.

Lemma 1.92. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$. Then*

1. *L is one-to-one if and only if $\ker(L) = \mathbf{0}$, if and only if the columns of A are linearly independent.*
2. *L is onto if and only if the columns of A span $\mathbb{R}^m$.*

Example 1.93. Let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be given by the matrix

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

Is this transformation one-to-one? Is it onto?

Row-reducing A gives us

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

Thus we can see there is no kernel: the nullspace is $\{\mathbf{0}\}$. So the transformation is one-to-one. But the two columns clearly don't span $\mathbb{R}^3$. In particular, we can see that $(1, 1, 1)$ is not in the span of the columns. So this transformation is not one-to-one.

1.9 Applications of Linear Functions

Linear functions can seem really restrictive. Most functions aren't linear, so how important can these linear functions be?

It turns out they have a lot of applications, precisely because they are so restricted and easy to work with. In particular, linear functions have the property that changing the input in the same way changes the output in the same way, no matter what the output starts out as.

1.9.1 Vector Calculus

We've seen linear functions in calculus. The point of (single-variable) differential calculus is to take a (single-variable) function and approximate it as a line. And lines are nice because they have constant *slope*. (This is just a rephrasing of what we said about linear functions: if you change the input by a certain amount, the output always changes by the same amount.)

If you do calculus of multivariable functions, the exact same thing happens. If you've taken multi, you might have looked at finding the equation for the tangent plane to a 2-variable function. This is essentially trying to find a linear function $T : \mathbb{R}^2 \to \mathbb{R}$ that approximates your original function.

(Technically it's an "affine" function, which is a linear function plus a constant, since linear functions have to pass through the origin.)

We can generalize this even further. If we have any function $f : \mathbb{R}^n \to \mathbb{R}^m$, we can find a linear function $T_f : \mathbb{R}^n \to \mathbb{R}^m$, or in other words a $n \times m$ matrix, that approximates it. This lets us do a good job of approximating these functions just like we can use the tangent line to approximate single-varible functions. We can also use this to get variants of Newton's method, to solve equations, and Euler's method, to solve differential equations.

1.9.2 Linear Regressions

More concretely, if you've studied economics or any statistical science or social science, you may have encountered a linear regression. In a linear regression, we have some set of independent variables $u_1, \dots, u_n$ and a set of dependent variables $v_1, \dots, v_m$. (You can imagine the independent variables being things like race, gender, parents' income, and the dependent variables being things like adult income, and health.) The goal of linear regression is to try to figure out what effects the independent variables on the dependent variables.

Specifically, you want a formula that does a pretty good job of predicting $v_1, \dots, v_m$ given $u_1, \dots, u_n$. You don't expect this prediction to be perfect, because real-world data are noisy. But you want to keep your formula simple, for two reasons. First, a simpler formula is easier to interpret. Second, a simpler formula avoids "overfitting", where you wind up predicting the random noise in your sample rather than the actual underlying relationship.

Linear regression approaches this by finding the *linear* (again, technically affine) function that best predicts the relationship. This avoids overfitting by being about as simple as possible. And it's easy to interpret, because you wind up with coefficients that say things like "each \$1000 of extra parental income leads to an extra \$500 of adult income in the children." (Note: specific numbers are entirely made up here.)

1.9.3 Markov Chains

We can also use linear function to model certain types of probabilistic behavior. The key requirement is that the process be random and memoryless: your future behavior can depend on your status now, but not your past status.

Example 1.94. Suppose at a certain time. 70% of the population lives in the suburbs, and 30% lives in the city. But each year, 6% of the people living in the suburbs move to the city, and 2% of the people living in the city move to the suburbs. What happens after a year? Five years?

Because these rates of transition are *constant*, we can model this with a matrix. If s is the number of people in the suburbs, and c is the number in the city, then next year we'll have $.94s + .02c$ people in the suburbs, and $.06s + .98c$ people in the city. With our numbers, that gives 66.4% in the suburbs, and 33.6% in the city.

We could repeat this calculation to find out what happens in two years, and then three, et cetera. But it's simpler, first, if we turn this into a matrix. If we think of (s, c) as a vector in $\mathbb{R}^2$, then the population changes according to the following matrix:

$$A = \begin{bmatrix} .94 & .02 \\ .06 & .98 \end{bmatrix}.$$

Thus after one year the population distribution will be $A \begin{bmatrix} .7 \\ .3 \end{bmatrix}$ and after five years it will be

$$A^5 \begin{bmatrix} .7 \\ .3 \end{bmatrix} = \begin{bmatrix} .744 & .085 \\ .256 & .915 \end{bmatrix} \begin{bmatrix} .7 \\ .3 \end{bmatrix} \approx \begin{bmatrix} .55 \\ .45 \end{bmatrix}.$$

This matrix A is sometimes called a *transition matrix*, although it has nothing to do with the change of basis matrices we discussed in section 5.1. Instead, it measures what fraction of a population transitions from one state to another—in this case, from the suburbs to the city or vice versa. Notice that every column sums up to 1. This isn't an accident; exactly 100% of a population has to go somewhere.

This entire process is very flexible. Any time the probability of transitioning from one state to another is constant, and only depends on which state you start in, we can model our system with a matrix like A, which we call a *Markov process*. In that case, the sequence of vectors $\mathbf{v}_1 = A\mathbf{v}, \mathbf{v}_2 = A^2\mathbf{v}, \ldots$ is called a *Markov chain*.

Each column of A is a *probability vector*, which is a vector of non-negative numbers that add up to one. Each row and each column corresponds to a particular possible state, and each entry tells us the probability of moving into the column-state if we start out in the row-state.

We can use this system to find the projected state after a finite number of steps, but even more usefully we can use it to project the equilibrium state.

Example 1.95. Suppose you run a car dealership that does long-term car leases. You lease sedans, sports cars, minivans, and SUVs. At the end of each year, your clients have the option to trade in to a different style of car. Empirically, you find that you get the following transition matrix:

$$C = \begin{bmatrix} .80 & .10 & .05 & .05 \\ .10 & .80 & .05 & .05 \\ .05 & .05 & .80 & .10 \\ .05 & .05 & .10 & .80 \end{bmatrix}$$

Thus if someone has a sedan this year, they are 80% likely to take a sedan next year, 10% likely to take a sports car, and 5% each likely to take a minivan or an SUV.

Using techniques we'll learn much later in section 4, we'll be able to calculate the steady state result as time goes to infinity. In particular, in a steady state, equal numbers of customers will lease each type of car, no matter what the distribution is right now.

Example 1.96. As a final, non-numerical example, this is how the Google PageRank algorithm works.

They treat web browsing as a random process: given that you are currently on one web page, you have some (small) probability of winding up on any other web page. Of course, this probability is higher if the page you're on links to the new page prominently, so the probability of winding up on any given page is not equal.

Then they build a giant $n \times n$ matrix, where n is the number of web pages they have analyzed. Each column corresponds to a particular web page, and the entries tell you how likely you are to go to any other web page next.

They use linear algebra to compute an equilibrium probability: if you browse the web for an arbitrarily long period of time, how likely are you to land on this page?

And that is, roughly speaking, the page rank. The more likely you are to land on a given web page, from this Markov chain model, the more highly ranked the page is.

1.9.4 Machine Learning

As a final example, linear algebra is heavily involved in most machine learning algorithms, including basically all deep learning, neural networks, and any exciting AI stuff you've read about in the news.

In essence, neural nets *want* to be linear regressions, but linear functions aren't quite flexible enough to do everything we want to do. So instead, we chain a bunch of linear functions together, with very simple non-linear functions in between them so they don't all collapse down into one linear function. But at the core, we're using a chain of matrices, and the "learning" process is an algorithm for finding the best possible coefficients to put in those matrices.

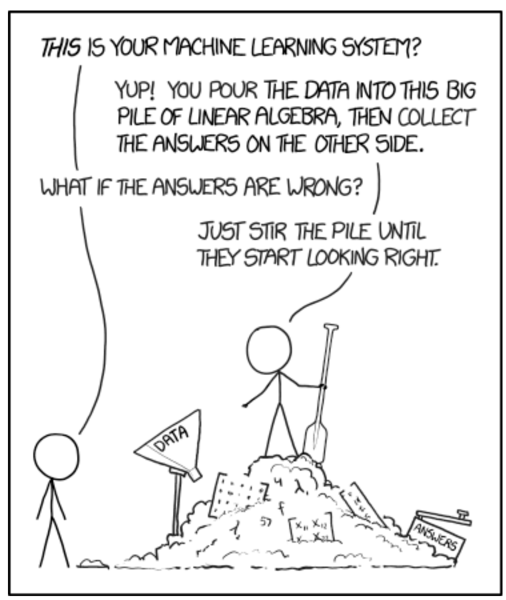


Figure 1.1: `https://xkcd.com/1838/`

2 Matrix Math

In large part, in this course we want to be studying linear functions. We saw in section 1.8.1 that linear functions are all secretly matrices, so this means we want to understand matrices.

2.1 Matrix Multiplication and Composition

We defined matrix multiplication in section 1.5.2, but now we can look at why it behaves as it does. Recall:

Definition 2.1. If $A \in M_{\ell \times m}$ and $B \in M_{m \times n}$, then there is a matrix $AB \in M_{\ell \times n}$ whose ij element is

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

If you're familiar with the dot product, you can think that the ij element of AB is the dot product of the ith row of A with the jth column of b.

Note that A and B don't have to have the same dimension! Instead, A has the same number of columns that B has rows. The new matrix will have the same number of rows as A and the same number of columns as B.

So what's going on here? Well, the fundamental operation we have is multiplying a vector by a matrix.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

But we often want to apply multiple transformations in a row, so instead of just calculating $A\mathbf{x}$ we want to calculuate $BA\mathbf{x}$. The definition of matrix multiplication is exactly what we need so that $(BA) \cdot \mathbf{x} = B \cdot (A\mathbf{x})$.

2.2 The identity matrix and matrix inverses

We just saw that any system of linear equations can be written $A\mathbf{x} = \mathbf{b}$, which reminds us of the single-variable linear equation $ax = b$. In the single-variable case we can just divide both sides of the equation by a, as long as $a \neq 0$; it would be nice if we can do the same thing for any system of linear equations.

But what does it mean to divide by a matrix? When we define division, we often start by understanding reciprocals $\frac{1}{a}$. So we start by asking what matrix is the equivalent of the number 1.

Definition 2.2. For any n we define the *identity matrix* to be $I_n \in M_n$ to have a 1 on every diagonal entry, and a zero everywhere else. For example,

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If $A \in M_n$ then $I_n A = A = A I_n$. Thus it is a *multiplicative identity* in the ring of $n \times n$ matrices.

The identity matrix is symmetric (that is, $I_n^T = I_n$).

Now we want to define multiplicative inverses, the equivalent of reciprocals. The definition is not difficult to invent:

Definition 2.3. Let A and B be $n \times n$ matrices, such that $AB = I_n = BA$. Then we say that B is the *inverse* (or *multiplicative inverse*) of A, and write $B = A^{-1}$.

If such a matrix exists, we say that A is *invertible* or *nonsingular*. If no such matrix exists, we say that A is *singular*.

Example 2.4. The identity matrix I_n is its own inverse, and thus invertible.

The matrices

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1/10 & 2/5 \\ 3/10 & -1/5 \end{bmatrix}$$

are inverses to each other, as you can check.

Example 2.5. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ has no inverse, since

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$$

won't be the identity for any a, b, c, d. Thus this matrix is singular.

Remark 2.6. If $AB = I_n$ then $BA = I_n$. This isn't really trivial but we won't prove it.

As the last example shows, finding the inverse to a matrix is a matter of solving a big pile of linear equations at the same time (one for each coefficient of the inverse matrix). Fortunately, we just got good at solving linear equations. Even more fortunately, there's an easy way to organize the work for these problems.

Proposition 2.7. *Let A be a $n \times n$ matrix. Then if we form the augmented matrix $\begin{bmatrix} A & I_n \end{bmatrix}$, then A is invertible if and only if the reduced row echelon form of this augmented matrix is $\begin{bmatrix} I_n & B \end{bmatrix}$ for some matrix B, and furthermore $B = A^{-1}$.*

Proof. Let X be a $n \times n$ matrix of unknowns, and set up the system of equations implied by $AX = I_n$. This will be the same set of equations we are solving with this row reduction, and thus a matrix X exists if and only if this system has a solution, which happens if and only if the reduced row echelon form of $\begin{bmatrix} A & I_n \end{bmatrix}$ has no all-zero rows on the A side. □

Example 2.8. Let's find an inverse for $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$.

We form and reduce the augmented matrix

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

Thus $A^{-1} = \begin{bmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$. We can check this by multiplying the matrices back together.

Example 2.9. Find the inverse of $B = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 1 & 6 \\ -3 & 0 & -10 \end{bmatrix}$.

We form and reduce the augmented matrix

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 1 & 1 & 6 & 0 & 1 & 0 \\ -3 & 0 & -10 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 2 & 3 & 0 & 1 \end{array}\right]$$

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

Thus $B^{-1} = \begin{bmatrix} -5 & 0 & -2 \\ -4 & 1 & -1 \\ 3/2 & 0 & 1/2 \end{bmatrix}$.

Example 2.10. What happens if we try to find an inverse for $C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$? We start with

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

but then there is no way to make the left-side block of the matrix into the identity I_2. Thus this matrix C is not invertible.

We can look at matrix invertibility as one more way of asking the questions we've already asked. In particular, if A is an invertible matrix, then the associated system of linear equations $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$, since we can write $\mathbf{x} = A^{-1}\mathbf{b}$. In general:

Proposition 2.11. *If A is a $n \times n$ matrix, then the following statements are equivalent:*

1. *The matrix A is invertible;*
2. *The system of equations $A\mathbf{x} = \mathbf{b}$ has a unique solution for any vector $\mathbf{b} \in \mathbb{R}^n$;*
3. *If the columns of A are $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then the vector equation $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{b}$ has a unique solution for any $b \in \mathbb{R}^n$;*
4. *The nullspace $N(A)$ is trivial and the columns span $\mathbb{R}^n$;*
5. *The associated linear transformation is one-to-one and onto;*
6. *The matrix A is row-equivalent to I_n.*

Most of these follow from realizing that they're restating the same question in a different way; each is a question about existence and uniqueness of solutions to some system of equations. The only really interesting one is the last one.

We can prove this by thinking in terms of equations again; the system of equations will have a unique solution if and only if the matrix row-reduces to the identity. But we can also phrase this in terms of so-called *elementary matrices.*

Definition 2.12. We say E is an *elementary matrix* if it is obtained from I_n by a single row operation.

Example 2.13.

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

are all elementary matrices.

We can see by some experimentation that if A is any matrix, and E is an elementary matrix obtained by performing a single row operation to I_n, then EA is the matrix obtained by performing that *same* row operation on A.

Example 2.14.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 5 \\ 1 & -1 & 4 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 5 \\ 1 & -1 & 4 \\ 4 & -2 & 13 \end{bmatrix}.$$

Thus a matrix A is row-equivalent to I_n if and only if there is some sequence of elementary matrices $E_1, E_2, \dots, E_k$ such that $A = E_1 E_2 \dots E_k I_n$.

Further, we see that each elementary matrix must be invertible, since we can encode the opposite of each row operation as its own elementary matrix. Thus if A is row-equivalent to the identity, we can write

$$A^{-1} = (E_1 E_2 \dots E_k I_n)^{-1} = I_n^{-1} E_k^{-1} \dots E_2^{-1} E_1^{-1}.$$

So any matrix row-equivalent to the identity must be invertible, and if we know the sequence of row operations necessary, we can use that to compute the inverse.

(However, this is rarely a useful or efficient way to compute the inverse; the row reduction algorithm is much more useful most of the time.)

2.3 Matrix Factorizations

We saw in the last section that inverse matrices are really useful for solving a large family of similar equations. If you just want to solve $A\mathbf{x} = \mathbf{b}$, then it's easier just to reduce the augmented matrix $[A|b]$. But if you want to solve each of a collection of systems

$$A\mathbf{x}_1 = \mathbf{b}_1, A\mathbf{x}_2 = \mathbf{b}_2, \dots, A\mathbf{x}_k = \mathbf{b}_k$$

then you can save a tremendous amount of effort by calculating A^{-1} and then computing

$$\mathbf{x}_1 = A^{-1}\mathbf{b}_1, \mathbf{x}_2 = A^{-1}\mathbf{b}_2, \dots, \mathbf{x}_k = A^{-1}\mathbf{b}_k.$$

If A is invertible, this works great. But if A isn't invertible, it's not terribly helpful. And we still want to parallelize these computations. (In particular, sometimes the matrix A isn't square, and thus *definitely* isn't invertible. But we want to do our best!)

A basic solution to this problem is the LU factorization. We will write our matrix A as a product of two matrices, L and U. We want L to be a square lower triangular matrix with only 1s on the diagonal, and U to be in row echelon form. Then if we want to solve $A\mathbf{x} = \mathbf{b}$, we have $LU\mathbf{x} = \mathbf{b}$. L will always be invertible, so we can take $U\mathbf{x} = L^{-1}\mathbf{b}$, and since U is already in row echelon form, the equation $U\mathbf{x} = L^{-1}\mathbf{b}$ is easy to solve for $\mathbf{x}$.

Example 2.15. Suppose we want to solve the system

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_5 &= 1 \\ 3x_1 + 7x_2 + 8x_3 + 2x_4 + 10x_4 &= 2 \\ 2x_1 + 9x_2 + 27x_3 + 11x_4 + 12x_5 &= 3. \end{aligned}$$

We will see how to work out that

$$\begin{bmatrix} 1 & 2 & 1 & 0 & 3 \\ 3 & 7 & 8 & 2 & 10 \\ 2 & 0 & 27 & 11 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 0 & 3 \\ 0 & 1 & 5 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Now we can solve the equation in two parts. We can solve $L\mathbf{y} = \mathbf{b}$, or $\mathbf{y} = L^{-1}\mathbf{b}$. It's not too tedious to work out that

$$\begin{aligned} y_1 &= 1 & & \\ 3y_1 + y_2 &= 2 & \Rightarrow \quad & y_2 = -1 \\ 2y_1 + 5y_2 + y_3 &= 3 & \Rightarrow \quad & y_3 = 6. \end{aligned}$$

So now we just have to solve

$$\begin{bmatrix} 1 & 2 & 1 & 0 & 3 \\ 0 & 1 & 5 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 6 \end{bmatrix}.$$

Now we can row-reduce

$$\begin{bmatrix} 1 & 2 & 1 & 0 & 3 & 1 \\ 0 & 1 & 5 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -9 & -4 & 1 & 3 \\ 0 & 1 & 5 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -9 & 0 & 5 & 27 \\ 0 & 1 & 5 & 0 & -1 & -13 \\ 0 & 0 & 0 & 1 & 1 & 6 \end{bmatrix}$$

Then we have free variables x_3, x_5, and our solution set is

$$\left\{ \begin{bmatrix} 27 + 9x_3 - 5x_5 \\ -13 - 5x_3 + x_5 \\ x_3 \\ 6 - x_5 \\ x_5 \end{bmatrix} \right\}.$$

We can see a factorization like this would be useful. But how do we find one? We use our discussion of elementary matrices from section 2.2. Given A, we know we can reduce it to row echelon form. We can record the row operations it takes to do this, and record those as a product of elementary matrices.

Example 2.16. Suppose we want to factor the matrix

$$\begin{bmatrix} 1 & 0 & 4 & 2 \\ 2 & 1 & 9 & 6 \\ 3 & 1 & 13 & 9 \end{bmatrix}.$$

We can row reduce this matrix by

$$\begin{bmatrix} 1 & 0 & 4 & 2 \\ 2 & 1 & 9 & 6 \\ 3 & 1 & 13 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 4 & 2 \\ 0 & 1 & 1 & 2 \\ 3 & 1 & 13 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 4 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 4 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

There are two ways we can think about this next step. One is to look at the elementary matrices we'd need to *undo* those row operations. Since our first step is subtracting twice Row 1 from Row 2, we need the matrix that will add twice Row 1 to Row 2:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Our second step was subtracting three times Row 1 from Row 3, and our third step was subtracting Row 2 from Row 3. We need to undo both of those:

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Thus we'll have

$$L = E_1E_2E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix}.$$

The other way we can think of this, which is often easier, is that we want L to be a matrix that gets reduced to the identity by the same series of operations. So we're looking for a matrix that reduces to the identity when we subtract $2R_1$ from R_2, and then subtract $3R_1$ from R_3, and then subract R_2 from R_3.

Obviously, we can always construct a matrix that looks like this; that is, we can always write $A = LU$ when U is in row echelon form and L is invertible. When do we get this lower triangular form? That will happen precisely when we didn't need to swap any rows to reduce U.

2.4 Subspaces

We often want to look at important and "nice" subsets of $\mathbb{R}^n$. We've already seen a few of these: the set of solutions to a system of equations, the nullspace of a matrix, and the span of a set of vectors are all subsets of $\mathbb{R}^n$ that we care about.

Let's look back at the nullspace. It had a nice property called "homogeneity":

1. If $\mathbf{x}, \mathbf{y} \in N(A)$, then $\mathbf{x} + \mathbf{y} \in N(A)$
2. If $\mathbf{x} \in N(A)$ and $r \in \mathbb{R}$ then $r\mathbf{x} \in N(A)$
3. $\mathbf{0} \in N(A)$.

Spaces with these homogeneity properties are easy to work with. Formally and algebraically, because we can rescale things however we want without leaving the space. Geometrically, these spaces are rigid and "linear": if the space contains any two points, it will contain the line between those two points. And since the space contains the origin, the space must just be a bunch of lines through the origin. So our space is a line, or plane, or hyperplane, that goes through the origin.

Definition 2.17. Let $U \subseteq \mathbb{R}^n$ such that:

1. $\mathbf{0} \in U$

2. (*closed under addition*) If $\mathbf{x}, \mathbf{y} \in U$, then $\mathbf{x} + \mathbf{y} \in U$

3. (*closed under scalar multiplication*) If $\mathbf{x} \in U$ and $r \in \mathbb{R}$ then $r\mathbf{x} \in U$.

Then we say that U is a *subspace* of $\mathbb{R}^n$.

Example 2.18. Let $V = \mathbb{R}^3$ and let $W = \{(x, y, x + y) \in \mathbb{R}^3\}$. Geometrically, this is a plane (given by $z = x + y$). We could in fact write $W = \{(x, y, z) : z = x + y\}$; this is a more useful way to write it for multivariable calculus, but less useful for lienar algebra. W is certainly a subset of V, so we just need to figure out if W is a subspace.

we only need to check three things. If $(x_1, y_1, x_1 + y_1), (x_2, y_2, x_2 + y_2) \in W$ then

$$\begin{bmatrix} x_1 \\ y_1 \\ x_1 + y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ x_2 + y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ (x_1 + x_2) + (y_1 + y_2) \end{bmatrix} \in W.$$

If $r \in \mathbb{R}$, then

$$r \begin{bmatrix} x \\ y \\ x + y \end{bmatrix} = \begin{bmatrix} rx \\ ry \\ (rx) + (ry) \end{bmatrix} \in W.$$

And the zero vector is

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 + 0 \end{bmatrix} \in W.$$

Thus W is a subspace of V.

Example 2.19. Let $V = \mathbb{R}^2$ and let $W = \{(x, x^2)\} = \{(x, y) : y = x^2\} \subseteq V$. Then W is *not* a subspace.

W does in fact contain the zero vector $(0, 0) = (0, 0^2)$. But we see that $(1, 1) \in W$, and $(1, 1) + (1, 1) = (2, 2) \notin W$. Thus W is not a subspace.

There are two (and a half) major sources of subspaces we'll see.

Corollary 2.20. *If $A\mathbf{x} = \mathbf{0}$ is a homogeneous system of linear equations, and $U = N(A)$ is the set of solutions to this system, then U is a subspace of $\mathbb{R}^n$.*

Proof. This follows from proposition 1.62; the definition of a subspace is precisely the homogeneity result we got for solutions to homogeneous systems of equations. □

Remark 2.21. The converse is also true: every subspace of $\mathbb{R}^n$ is the set of solutions to some homogeneous system of linear equations. We won't prove this until later.

Example 2.22. Let's look at our earlier examples again. We took $W \subset \mathbb{R}^3$ defined by $W = \{x, y, x+y\}$. This is precisely the set of solutions to the linear equation $x + y - z = 0$.

We also had the subspace given by $W = \{x, 0, x)\}$. This is the solution to the system of linear equations

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

In contrast, we can see that $\{(x, 1, x)\}$ is the solution to the system

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

which is not homogeneous. And $W = \{(x, x^2)\}$ is the solution to the equation $x^2 - y = 0$, which *is* homogeneous, but isn't *linear*. We saw that neither of these is a vector space.

The other major example is the span of a set of vectors.

Proposition 2.23. *If we have a collection $S = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ of vectors in $\mathbb{R}^n$, then the span of S is a subspace.*

Proof. 1. We know that $0\mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in V$. So we have

$$0\mathbf{u}_1 + 0\mathbf{u}_2 + \cdots + 0\mathbf{u}_n = \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}.$$

Thus $\mathbf{0} \in \operatorname{span}(S)$.

2. Suppose $\mathbf{v}, \mathbf{w} \in \operatorname{span}(S)$. This implies that we can write

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n \qquad \mathbf{w} = b_1\mathbf{w}_1 + \cdots + b_n\mathbf{w}_n$$

for some $a_i, b_i \in \mathbb{R}$. Thus

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= (a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n) + (b_1\mathbf{w}_1 + \cdots + b_n\mathbf{w}_n) \\ &= (a_1 + b_1)\mathbf{u}_1 + \cdots + (a_n + b_n)\mathbf{u}_n \in \operatorname{span}(S). \end{aligned}$$

3. Suppose $r \in \mathbb{R}$ and $\mathbf{v} \in \text{span}(S)$. Then we can write

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n$$

for some $a_i \in \mathbb{R}$. Then

$$r\mathbf{v} = r\left(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n\right) = (ra_1)\mathbf{u}_1 + \cdots + (ra_n)\mathbf{u}_n \in \text{span}(S).$$

Thus we see that span(S) is a subspace of V.

□

Example 2.24. As before, take $V = \mathbb{R}^3$ and $S = \{(1,0,0),(0,1,0)\}$. Then

$$\text{span}(S) = \{a(1,0,0) + b(0,1,0)\} = \{(a,b,0)\}.$$

This is a subspace of $\mathbb{R}^2$.

Now let $T = \{(3,2,0),(13,7,0)\}$. Then

$$\text{span}(T) = \{a(3,2,0) + b(13,7,0)\} = \{(3a+13b, 2a+7b, 0)\}.$$

This is a subspace of $\mathbb{R}^2$; in fact, it is exactly the same subspace! The first spanning set "looks" nicer, but it's hard to make this sense of "nice" mathematically precise. We'll do our best, but won't really get there until section 6.

Example 2.25. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. What is span$\{\mathbf{u}_1, \mathbf{u}_2\}$?

Geometrically, this will be a plane through the origin, containing the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$.

We can use this new terminology to connect up some ideas we have seen previously.

Definition 2.26. Let A be a $m \times n$ matrix. The *column space* of A, denoted col(A), is the subspace of $\mathbb{R}^m$ spanned by the columns of A. The *rowspace* of A, denoted row(A), is the space spanned by the rows of A.

Remark 2.27. We now have three subspaces attached to a given matrix: the rowspace, columnspace, and nullspace. Some sources will also point to a fourth space $N(A^T)$, but I won't really discuss this until we have the tools to explain why it matters.

There are two facts about these spaces that we can obtain from the work we've already done.

Proposition 2.28. *Let A be a $m \times n$ matrix. Then* $\operatorname{col}(A)$ *is the image of the linear transformation associated to A.*

Proof. We proved this in section 1.8. We observed that $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in the span of the columns of A, which is precisely the columnspace of A. □

Proposition 2.29. *If A and B are row-equivalent matrices, then they have the same rowspace.*

Proof. We need to check that each elementary row operation doesn't change the span of the set of vectors.

I. (Switch two rows) Switching the order of two vectors does not affect the span at all.

II. (Multiply a row by a nonzero scalar) Multiplying a vector by a non-zero scalar won't change the span of the set of vectors, since in any linear combination we can always just multiply the relevant coefficient by the inverse of our non-zero scalar.

III. (Add a multiple of one row to another) This won't add anything to the span, since a linear combination of the new vectors will still be a linear combination of the old vectors.

   This won't lose anything from the span, since we can undo the row operation, and so every old vector is a linear combination of new vectors.

□

Finally, I promised two and a half sources of subspaces. The last one is the simplest.

Example 2.30. The sets $0 = \{\mathbf{0}\} \subset \mathbb{R}^n$ and $\mathbb{R}^n \subset \mathbb{R}^n$ are both subspaces of $\mathbb{R}^n$. When we want to ignore these, we'll refer to "proper" subspaces.

2.5 Basis, Dimension, and Rank

We want tools to really understand what subspaces look like. A subspace essentially always contains infinitely many vectors, which is too many to think about. But we can compress that information to just a small, finite collection of vectors. The basic idea is to list the independent *directions* we can move inside the space.

We can find a spanning set for any space. If nothing else, the set of "all the vectors in the subspace" spans the space, but that's not terribly helpful. What we really want is a *minimal* spanning set, one that represents the entire space with no redundancies.

Definition 2.31. Let U be a subspace of $\mathbb{R}^n$. If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is spans U and is linearly independent, then S is a *basis* for U.

Bases are useful because they provide a coordinate system for our subspace. Every vector can be represented in exactly one way as a linear combination of elements of the basis, which means that we can convert in a one-to-one way between "lists of coefficents" and "vectors in the subspace". Any coordinate system is also a basis:

Example 2.32. Earlier we mentioned the *standard basis*

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \dots \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

for $\mathbb{R}^n$. The standard basis is in fact a basis.

We've seen two different major sources of subspaces. The first is that the span of any set is a subspace. In that case it's easy to find a spanning set for our subspace, but how do we find a basis? It turns out we can always just remove linear dependent vectors from our spanning set, and we will eventually hit a basis.

Lemma 2.33 (Basis Reduction)**.** *Suppose $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a spanning set for V. Then S can be reduced to a basis for V. That is, there is a subset $B \subseteq S$ that is a basis for V.*

Proof. If S is linearly independent, then it is a basis and we're done.

So suppose S is linearly dependent. Then we know at least one vector is redundant, so without loss of generality we can reorder the set so that we can write $\mathbf{v}_n$ as a linear combination of the other vectors in S.

But then $\operatorname{span}(S) = \operatorname{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\})$, and $S_1 = \{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$ is a spanning set for V and a proper subset of S. If S_1 is linearly independent, then it is a basis; if not, we can repeat this process until we reach a linearly independent set, which is our basis B. □

Example 2.34. Let $S = \{(1, 1, 0), (1, 1, 1), (0, 0, 1), (2, 7, 0)\}$ be a spanning set for $\mathbb{R}^3$. Find a basis $B \subseteq S$ for $\mathbb{R}^3$.

We'll take as given that this is a spanning set, which is not difficult to check. We see that we can write $(1, 1, 1) = (1, 1, 0) + (0, 0, 1)$, so we can remove $(1, 1, 1)$ without changing the span, and we have $B = \{(1, 1, 0), (0, 0, 1), (2, 7, 0)\} \subseteq S$ is a basis for $\mathbb{R}^3$.

Example 2.35. Let $S = \{(1,2,3),(1,1,1),(5,-2,1),(-4,3,2)\}$ be a spanning set for $\mathbb{R}^3$. Find a basis $B \subseteq S$ for $\mathbb{R}^3$.

We'll take as given that S is a spanning set. We need to write one vector as a linear combination of the others, which is essentially the same problem as finding a nontrivial linear combination equal to zero. So we set up the equation

$$\begin{bmatrix}0\\0\\0\end{bmatrix} = a\begin{bmatrix}1\\2\\3\end{bmatrix} + b\begin{bmatrix}1\\1\\1\end{bmatrix} + c\begin{bmatrix}5\\-2\\1\end{bmatrix} + d\begin{bmatrix}-4\\3\\2\end{bmatrix}$$

which gives us

$$\begin{bmatrix}1&1&5&-4\\2&1&-2&3\\3&1&1&2\end{bmatrix} \to \begin{bmatrix}1&1&5&-4\\0&-1&-12&11\\0&-2&-14&14\end{bmatrix} \to \begin{bmatrix}1&0&-7&7\\0&1&12&-11\\0&0&10&-8\end{bmatrix} \to \begin{bmatrix}1&0&0&7/5\\0&1&0&-7/5\\0&0&1&-4/5\end{bmatrix}$$

This gives us $a = -7/5d, b = 7/5d, c = 4/5d$, or in other words if we set $d = 1$ we get

$$\begin{bmatrix}-4\\3\\2\end{bmatrix} = \frac{7}{5}\begin{bmatrix}1\\2\\3\end{bmatrix} - \frac{7}{5}\begin{bmatrix}1\\1\\1\end{bmatrix} - \frac{4}{5}\begin{bmatrix}5\\-2\\1\end{bmatrix}.$$

Thus $(-4,3,2)$ can be written as a linear combination of the other vectors, and so we have $B = \{(1,2,3),(1,1,1),(5,-2,1)\}$ is a basis for $\mathbb{R}^3$. We know this is a basis because it is still a spanning set, and has the correct number of elements.

(We could actually have removed any vector from this set and still gotten a basis; each element can be written as a combination of the others, as you can see by rearranging the last equation. But it's sufficient here to find one basis.)

Remark 2.36. We can also do this the other way. If we have a linearly independent subset of a space, we can always add elements to it to get a basis.

Example 2.37. Let $S = \{(1,1,0),(1,-1,0)\}$. Find a basis $B \supseteq S$ for $\mathbb{R}^3$.

We see that S is linearly independent, so we just need to find a vector that isn't in $\operatorname{span}(S)$. It's clear that $(0,0,1) \notin \operatorname{span}(S)$, so we see that $B = \{(1,1,0),(1,-1,0),(0,0,1)\}$ satisfies our requirements.

But there are many choices we could make. It's also the case that $(1,1,1) \notin \operatorname{span}(S)$, so we see that $B_1 = \{(1,1,0),(1,-1,0),(1,1,1)\}$ also satisfies our requirements.

The other major source of subspaces is the spaces associated to a matrix. We actually already know how to find a basis for the nullspace; we just didn't use that language at the time.

Example 2.38. Let $A = \begin{bmatrix} 1 & 3 & 0 & 2 & 0 & 5 \\ 0 & 0 & 1 & 3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$ and suppose we want to find $N(A)$. We know we can write

$$N(A) = \left\{ \begin{bmatrix} -3x_2 - 2x_4 - 5x_6 \\ x_2 \\ -3x_4 - 2x_6 \\ x_4 \\ -x_6 \\ x_6 \end{bmatrix} \right\} = \left\{ x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -5 \\ 0 \\ -2 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

But this set of three vectors certainly spans $N(A)$, and it's easy to see that they're linearly independent (because each one has a 1 in an entry where every other vector has 0s). Thus this set of three vectors is a basis for $N(A)$.

Finding a basis for the columnspace is probably more theoretically complicated but practically easier. If our matrix is in reduced echelon form, it's easy to see that the pivot columns form a basis for the column space: each row that *ever* has a non-zero entry will have a column with a 1 in that row and 0s otherwise. These will be linearly independent, and they will span everything.

Example 2.39. Let $A = \begin{bmatrix} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Then the first, second, and fourth columns form a basis for the columnspace.

If the matrix isn't already in row echelon form, it's less obvious what to do. It's not really clear how we can use row reduction, because row reducing a matrix definitely changes the column space.

However, it turns out that row reducing a matrix doesn't change which columns are linearly dependent on each other. If the columns of our matrix are linearly independent, that means we can find a vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$, and the non-zero entries of $\mathbf{x}$ tell us which columns are redundant. But if B is row-equivalent to A, then $B\mathbf{x} = \mathbf{0}$ if and only if $A\mathbf{x} = \mathbf{0}$, so the columns of B have the same linear dependencies that the columns of A have.

This gives us a simple algorithm for finding a basis for the columnspace of a matrix A. We row reduce the matrix to get a matrix B in RREF. We look at the pivot columns of B, and we take the corresponding columns of A for our basis.

Example 2.40. Find a basis for the column space of $A = \begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix}$

We can row reduce this matrix

$$\begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 5 & -9 & 11 \\ 0 & 1 & -3 & 1 \\ 0 & 2 & -3 & 3 \\ 0 & 7 & -12 & 10 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 6 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 9 & 3 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 6 & 6 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 9 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The first three columns are pivot columns, and that means the first three columns of the *original matrix* form a basis for $\operatorname{col}(A)$. Thus a basis is $\{(1, -2, 3, -1), (5, -9, 17, 2), (-9, 15, -30, -3)\}$.

Finally, the rowspace is the easiest space to find a basis for. Two row-equivalent matrices have the same rowspace, and the non-zero rows of a matrix in echelon form are linearly independent. So we can just reduce the matrix and then read off the columns.

Example 2.41. By the previous row reduction, a basis for the matrix $A = \begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix}$ is $\{(1, 0, 0, 4), (0, 1, 0, 2), (0, 0, 1, 1/3)\}$.

It turns out that for a given subspace, any two bases must have the same number of elements. In particular, it's tedious but possible to prove that every spanning set is at least as big as every linearly independent set; that is, if there is a four-element linearly independent set, there are no three-element spanning sets. This allows us to make the following definition:

Definition 2.42. Let $U \subseteq \mathbb{R}^n$ be a subspace, and let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a basis for U. Then we define the *dimension* of U to be m and write $\dim U = m$.

We say that the dimension of the trivial subspace $\{\mathbf{0}\}$ is 0. You can think of this space as having the empty set $\{\} = \varnothing$ as a basis.

It follows from counting the standard basis that the dimension of $\mathbb{R}^n$ is in fact n, which you would hope is the case. In general, a space with a n-element basis has n independent directions we can move in; that corresponds essentially to what we mean by talking about an n-dimensional space.

We'll revisit dimension a few times throughout the course, but right now the most important dimension we need to be concerned with is the *rank*.

Definition 2.43. Let A be a matrix. Then the *rank* of A is $\text{rk}(A) = \dim(\text{col}(A))$.

We can see that the rank is the dimension of the image of A, since the columnspace and the image are the same. There are a few other important results about the rank.

Proposition 2.44. *Let A be a $m \times n$ matrix. Then* $\dim(\text{row}(A)) = \text{rk}(A)$.

Proof. The rank is the dimension of the columnspace, which is the number of pivot columns in A. The dimension of the rowspace is the number of non-zero rows in the reduced form of A. But there is exactly one non-zero row for each pivot column. □

The most important result here is the rank-nullity theorem.

Theorem 2.45 (Rank-Nullity Theorem)**.** *Let A be a $m \times n$ matrix. Then* $\text{rk}(A) + \dim N(A) = n$.

Proof. The rank of A is the number of pivot columns, and $\dim N(A)$ is the number of free variables. But each column is either a pivot column or a free variable, and there are n total columns. □

This theorem tells us, essentially, that if the domain of your transformation has n dimensions, then the dimension of the kernel plus the dimension of the image is also n. I tend to think of this as telling us that every dimension either gets sent to zero, and so is in the kernel, or winds up being sent somewhere non-zero, and so is in the image.

Finally, this allows us to extend our list of characterizations of invertible matrices that we began in section 2.2 with proposition 2.11.

Proposition 2.46. *Let A be a $n \times n$ matrix. Then the following statements are equivalent:*

1. *A is invertible*
2. $\mathrm{rk}(A) = n$
3. *The columns of A form a basis for* $\mathbb{R}^n$
4. *The rows of A form a basis for* $\mathbb{R}^n$
5. $N(A) = \{\mathbf{0}\}$.

Proof. The rank-nullity theorem tells us that A is one-to-one if and only if it is onto, and that the columns are linearly independent if and only if they span $\mathbb{R}^n$. Then we only need to check one of these conditions. □

3 Vector Spaces

3.1 Definition of a Vector Space

In section 1.6 we saw that the set of solutions to a homogeneous system of linear equations has the following three properties:

1. It contains the trivial solution of all zeroes;

2. The sum of two solutions is a solution;

3. Any scalar multiple of a solution is a solution.

In section 2.4 we saw this set of properties again: any subset of $\mathbb{R}^n$ with these properties is a subspace. There's a lot we can say about these sets: they are all nullspaces of some sort of linear function, and they are all the span of some set of "basis" vectors.

But there are other places we can use these tools: any set that looks "like" $\mathbb{R}^n$ will allow us to do the same things. In this section we'll nail down exactly what properties we need to have for something to look enough like $\mathbb{R}^n$ that the tools we've already developed will apply.

Definition 3.1. Let V be a set together with two operations:

- A *vector addition* which allows you to add two elements of V and get a new element of V. If $\mathbf{v}, \mathbf{w} \in V$ then the sum is denoted $\mathbf{v} + \mathbf{w}$ and must also be an element of V.

- A *scalar multiplication* which allows you to multiply an element of V by a real number (or "scalar") and get a new element of V. If $r \in \mathbb{R}$ and $\mathbf{v} \in V$ then the scalar multiple is denoted $r \cdot \mathbf{v}$ and must also be an element of V.

Further, suppose the following axioms hold for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and any $r, s \in \mathbb{R}$:

1. (Closure under addition) $\mathbf{u} + \mathbf{v} \in V$

2. (Additive commutativity) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

3. (Additive associativity) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

4. (Additive identity) There is an element $\mathbf{0} \in V$ called the "zero vector", such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for every $\mathbf{u}$.

5. (Additive inverses) For each $\mathbf{u} \in V$ there is another element $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

6. (Closure under scalar multiplication) $r\mathbf{u} \in V$

7. (Distributivity) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$

8. (Distributivity) $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$

9. (Multiplicative associativity) $r(s\mathbf{u}) = (rs)\mathbf{u}$

10. (Multiplicative Identity) $1\mathbf{u} = \mathbf{u}$.

Then we say V is a *Vector Space*, and we call its elements *vectors*.

This is a mouthful (or a pageful, in this case). But the important thing to note is that this is nothing new, or especially strange. Points 1, 4, and 6 are precisely the definition of a subspace; the other seven points are things that you would probably *assume* were true even if I hadn't pointed them out.

So the right way to think about this definition isn't that a vector space is some strange, new, exotic thing you've never seen before. Instead, we're just listing the properties of $\mathbb{R}^n$, which we already knew about, that allow us to do exactly the things we've been doing. That means that if we find anything else with this same list of properties, we can reuse all the work we've already done.

Example 3.2. $\mathbb{R}^n$ is a vector space, with the previously defined vector addition and scalar multiplication. We check:

Let $\mathbf{u} = (u_1, \dots, u_n).\mathbf{v} = (v_1, \dots, v_n), \mathbf{w} = (w_1, \dots, w_n) \in \mathbb{R}^n, \quad r, s \in \mathbb{R}$. Then, knowing the usual rules of commutativity and associativity of basic arithmetic, we can compute:

1. $\mathbf{u} + \mathbf{v} = (u_1, \dots, u_n) + (v_1, \dots, v_n) = (u_1 + v_1, \dots, u_n + v_n) \in \mathbb{R}^n$.

2. $$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1, \dots, u_n) + (v_1, \dots, v_n) = (u_1 + v_1, \dots, u_n + v_n) \\ &= (v_1 + u_1, \dots, v_n + u_n) = (v_1, \dots, v_n) + (u_1, \dots, u_n) = \mathbf{v} + \mathbf{u} \end{aligned}$$

3. $$\begin{aligned} (\mathbf{u} + \mathbf{v}) + \mathbf{w} &= (u_1 + v_1, \dots, u_n + v_n) + (w_1, \dots, w_n) = (v_1 + u_1 + w_1, \dots, v_n + u_n + w_n) \\ &= (v_1, \dots, v_n) + (u_1 + w_1, \dots, u_n + w_n) = \mathbf{v} + (\mathbf{u} + \mathbf{w}) \end{aligned}$$

4. We have $\mathbf{0} = (0, \dots, 0)$. Then
$$\mathbf{0} + \mathbf{v} = (0 + v_1, \dots, 0 + v_n) = (v_1, \dots, v_n) = \mathbf{v}.$$

5. Set $-\mathbf{u} = (-u_1, \dots, -u_n)$. Then

$$\mathbf{u} + (-\mathbf{u}) = (u_1 + (-u_1), \dots, u_n + (-u_n)) = (0, \dots, 0) = \mathbf{0}.$$

6.

$$r\mathbf{u} = r(u_1, \dots, u_n) = (ru_1, \dots, ru_n) \in \mathbb{R}.$$

7.

$$\begin{aligned} r(\mathbf{u} + \mathbf{v}) &= r(u_1 + v_1, \dots, u_n + v_n) = (r(u_1 + v_1), \dots, r(u_n + v_n)) \\ &= (ru_1 + rv_1, \dots, ru_n + rv_n) = (ru_1, \dots, ru_n) + (rv_1, \dots, rv_n) = r\mathbf{u} + r\mathbf{v}. \end{aligned}$$

8.

$$\begin{aligned} (r+s)\mathbf{u} &= (r+s)(u_1, \dots, u_n) = ((r+s)u_1, \dots, (r+s)u_n) \\ &= (ru_1 + su_1, \dots, ru_n + su_n) = (ru_1, \dots, ru_n) + (su_1, \dots, su_n) = r\mathbf{u} + s\mathbf{u}. \end{aligned}$$

9.

$$r(s\mathbf{u}) = r(su_1, \dots, su_n) = (rsu_1, \dots, rsu_n) = rs(u_1, \dots, u_n).$$

10.

$$1\mathbf{u} = 1(u_1, \dots, u_n) = (1 \cdot u_1, \dots, 1 \cdot u_n) = (u_1, \dots, u_n) = \mathbf{u}.$$

Remark 3.3. That took forever and was incredibly tedious. (It's not actually *difficult*, just extremely annoying). I won't actually ask you to do this yourself, but I do want you to see how these pieces fit together.

So what else is a vector space and "looks like $\mathbb{R}^n$"?

Example 3.4. Let $\mathcal{P}(x) = \{a_0 + a_1 x + \cdots + a_n x^n : n \in \mathbb{N}, a_i \in \mathbb{R}\}$ be the set of polynomials with real coefficients. Define addition by

$$(a_0 + a_1 x + \cdots + a_n x^n) + (b_0 + b_1 x + \cdots + b_n x^n) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

and define scalar multiplication by

$$r(a_0 + a_1 x + \cdots + a_n x^n) = ra_0 + ra_1 x + \cdots + ra_n x^n.$$

Then $\mathcal{P}(x)$ is a vector space.

Example 3.5. Let S be the space of all doubly infinite sequences $\{y_k\} = \{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots\}$. We call this the space of (discrete) *signals*: it represents a sequence of measurements taken at regular time intervals. These sorts of regular measurements are common in engineering and digital information applications (such as digital music).

We define addition and scalar multiplication on the space of signals componentwise, so that

$$\{\dots, x_{-1}, x_0, x_1, \dots\} + \{\dots, y_{-1}, y_0, y_1, \dots\} = \{\dots x_{-1} + y_{-1}, x_0 + y_0, x_1 + y_1, \dots\}$$

and

$$r\{\dots, y_{-1}, y_0, y_1, \dots\} = \{\dots, ry_{-1}, ry_0, ry_1, \dots\}.$$

(In essence, S is composed of vectors that are infinitely long in both directions). Then S is a vector space.

Example 3.6. Let $\mathcal{F}(\mathbb{R}, \mathbb{R}) = \mathcal{F}$ be the set of functions from $\mathbb{R}$ to $\mathbb{R}$—that is, functions that take in a real number and return a real number, the vanilla functions of single-variable calculus. Define addition by $(f + g)(x) = f(x) + g(x)$ and define scalar multiplication by $(rf)(x) = r \cdot f(x)$. Then $\mathcal{F}$ is a vector space.

1. We have vector addition defined by $(f+g)(x) = f(x)+g(x)$. This does give a function, so the vector space is closed under addition.

2. $(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.

3. $((f + g) + h)(x) = (f + g)(x) + h(x) = f(x) + g(x) + h(x) = f(x) + (g + h)(x) = (f + (g + h))(x)$.

4. Let $\mathbf{0}$ be the zero function defined by $\mathbf{0}(x) = 0$. Then we see that $(f + \mathbf{0})(x) = f(x) + \mathbf{0}(x) = f(x) + 0 = f(x)$.

5. Define $(-f)(x)$ by $(-f)(x) = -f(x)$. Then $(f + (-f))(x) = f(x) + (-f)(x) = f(x) - f(x) = 0 = \mathbf{0}(x)$.

6. Define scalar multiplication by $(rf)(x) = rf(x)$. This does give a function, so the vector space is closed under multiplication.

7. $(r(f + g))(x) = r(f + g)(x) = r(f(x) + g(x)) = rf(x) + rg(x) = (rf)(x) + (rg)(x)$.

8. $((r + s)f)(x) = (r + s)f(x) = rf(x) + sf(x) = (rf)(x) + (sf)(x)$.

9. $(r(sf))(x) = r(sf)(x) = rsf(x) = (rs)f(x) = ((rs)f)(x)$.

10. $(1 \cdot f)(x) = 1f(x) = f(x)$.

Thus $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is a vector space.

Example 3.7. If $A\mathbf{x} = \mathbf{0}$ is a homogeneous system of linear equations, then the set of solutions $N(A)$ is a vector space. I won't prove this now because we will shortly develop techniques to make proving this much faster and less irritating in section 3.2.

Example 3.8. The set $M_{m \times n}$ of $m \times n$ matrices is a vector space under the addition and scalar multiplication defined in section 1.5.1, with zero vector given by

$$\mathbf{0} = (0) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

I'm not going to prove this, but you can see that it should be true for the same reason $\mathbb{R}^{mn}$ is a vector space: they're both just lists of real numbers, but one is arranged in a column and the other in a rectangle. The operations are the same.

Example 3.9. The integers $\mathbb{Z}$ are *not* a vector space (under the usual definitions of addition and multiplication). For instance, $1 \in \mathbb{Z}$ but $.5 \cdot 1 = .5 \notin \mathbb{Z}$.

(We only need to find one axiom that doesn't hold to show that a set is not a vector space, since a vector space must satisfy all the axioms).

Example 3.10. The closed interval $[0, 5]$ is not a vector space (under the usual operations) , since $3, 4 \in [0, 5]$ but $3 + 4 = 7 \notin [0, 5]$.

Example 3.11. Let $V = \mathbb{R}$ with scalar multiplication given by $r \cdot x = rx$ and addition given by $x \oplus y = 2x + y$. Then V is not a vector space, since $x \oplus y = 2x + y \neq 2y + x = y \oplus x$; in particular, we see that $3 \oplus 5 = 11$ but $5 \oplus 3 = 13$.

There are many more examples of vector spaces, but as you can see it's fairly tedious to prove that any particular thing is a vector space. In section 3.2 we'll develop a *much* easier way to establish that something is a vector space, so we won't develop any more examples now.

3.1.1 Properties of Vector Spaces

The great thing about the formal approach is that we can show that anything that satisfies the axioms of a vector space must aso follow some other rules. We'll establish a few of those rules here, though of course, there's a sense in which the entire rest of this course will be spent establishing those rules.

As before, you shouldn't think of these rules as new facts; all of them are "obvious". The point is that if we get the list of properties from definition 3.1, then all of these other things still have to occur. It's a guarantee that vector spaces behave how we expect.

Proposition 3.12 (Cancellation)**.** *Let V be a vector space and suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ are vectors. If $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$, then $\mathbf{u} = \mathbf{v}$.*

Proof. By axiom we know that $\mathbf{w}$ has an additive inverse $-\mathbf{w}$. Then we have

$$\begin{aligned}
\mathbf{u} + \mathbf{w} &= \mathbf{v} + \mathbf{w} \\
(\mathbf{u} + \mathbf{w}) + (-\mathbf{w}) &= (\mathbf{v} + \mathbf{w}) + (-\mathbf{w}) \\
\mathbf{u} + (\mathbf{w} + (-\mathbf{w})) &= \mathbf{v} + (\mathbf{w} + (-\mathbf{w})) && \text{Additive associativity} \\
\mathbf{u} + \mathbf{0} &= \mathbf{v} + \mathbf{0} && \text{Additive inverses} \\
\mathbf{u} &= \mathbf{v} && \text{Additive identity.}
\end{aligned}$$

□

Proposition 3.13. *The additive inverse $-\mathbf{v}$ of a vector $\mathbf{v}$ is unique. That is, if $\mathbf{v} + \mathbf{u} = \mathbf{0}$, then $\mathbf{u} = -\mathbf{v}$.*

Proof. Suppose $\mathbf{v} + \mathbf{u} = \mathbf{0}$. By the additive inverses property we know that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$, and thus $\mathbf{v} + \mathbf{u} = \mathbf{v} + (-\mathbf{v})$. By cancellation we have $\mathbf{u} = -\mathbf{v}$. □

Remark 3.14. In our axioms we asserted that every vector *has* an inverse, but didn't require that there be only one.

Proposition 3.15. *Suppose V is a vector space with $\mathbf{u} \in V$ a vector and $r \in \mathbb{R}$ a scalar. Then:*

1. $0\mathbf{u} = \mathbf{0}$
2. $r\mathbf{0} = \mathbf{0}$
3. $(-1)\mathbf{u} = -\mathbf{u}$.

Remark 3.16. We would actually be pretty sad if any of those statements were false, since it would make our notation look very strange. (Especially the last statement). The fact that these statements *are* true justifies us using the notation we use.

Proof. 1.

$$\begin{aligned} \mathbf{u} = 1 \cdot \mathbf{u} &= (0+1)\mathbf{u} && \text{Multiplicative identity} \\ &= 0\mathbf{u} + 1\mathbf{u} && \text{Distributivity} \\ &= 0\mathbf{u} + \mathbf{u} && \text{Multiplicative identity} \\ \mathbf{0} + \mathbf{u} &= 0\mathbf{u} + \mathbf{u} && \text{Additive identity} \\ \mathbf{0} &= 0\mathbf{u} && \text{Cancellation} \end{aligned}$$

2. We know that $\mathbf{0} = \mathbf{0}+\mathbf{0}$ by additive identity, so $r\mathbf{0} = r(\mathbf{0}+\mathbf{0}) = r\mathbf{0}+r\mathbf{0}$ by distributivity. Then we have

$$\begin{aligned} \mathbf{0} + r\mathbf{0} &= r\mathbf{0} + r\mathbf{0} && \text{additive identity} \\ \mathbf{0} &= r\mathbf{0} && \text{cancellation.} \end{aligned}$$

3. We have

$$\begin{aligned} \mathbf{v} + (-1)\mathbf{v} &= 1\mathbf{1} + (-1)\mathbf{v} && \text{multiplicative inverses} \\ &= (1 + (-1))\mathbf{v} && \text{distributivity} \\ &= 0\mathbf{v} = \mathbf{0}. \end{aligned}$$

Then by uniqueness of additive inverses, we have $(-1)\mathbf{v} = -\mathbf{v}$.

□

Example 3.17. We'll give one last example of a vector space, which is both important and silly.

We define the *zero vector space* to be the set $\{\mathbf{0}\}$ with addition given by $\mathbf{0}+\mathbf{0} = \mathbf{0}$ and scalar multiplication given by $r \cdot \mathbf{0} = \mathbf{0}$. It's easy to check that this is in fact a vector space.

Notice that we didn't ask what "kind" of object this is; we just said it has the zero vector and nothing else. As such, this could be the zero vector of any vector space at all. In section 3.2 we will talk about vector spaces that fit inside other vector spaces, like this one.

3.2 Vector Space Subspaces

In section 2.4 we defined subspaces of $\mathbb{R}^n$. We can also define subspaces of other vector spaces.

Definition 3.18. Let V be a vector space. A subset $W \subseteq V$ is a *subspace* of V if W is also a vector space with the same operations as V.

But we can identify these subspaces in the same way we did before.

Proposition 3.19. *But Let V be a vector space and $W \subseteq V$. Then W is a subspace of V if and only if the following three "subspace" conditions hold:*

1. $\mathbf{0} \in W$ *(zero vector);*
2. *Whenever* $\mathbf{u}, \mathbf{v} \in W$ *then* $\mathbf{u} + \mathbf{v} \in W$ *(Closed under addition); and*
3. *Whenever* $r \in \mathbb{R}$ *and* $\mathbf{u} \in W$ *then* $r\mathbf{u} \in W$ *(Closed under scalar multiplication).*

Proof. Suppose W is a subspace of V. Then W is a vector space, so it contains a zero vector and is closed under addition and multiplication by the definition of vector spaces.

Conversely, suppose $W \subseteq V$ and the three subspace conditions hold. We need to check the ten axioms of a vector space. But most of these properties are inherited from the fact that any element of W is also an element of V, and W has the same operations as V.

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in W$ (and thus $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$), and $r, s \in \mathbb{R}$.

1. W is closed under addition by hypothesis.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ since V is a vector space.
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ since V is a vector space.
4. $\mathbf{0} \in W$ by hypothesis, and $\mathbf{u} + \mathbf{0} = \mathbf{u}$ since V is a vector space.
5. $-\mathbf{u} = (-1)\mathbf{u} \in W$ by closure under scalar multiplication.
6. W is closed under scalar multiplication by hypothesis.
7. $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$ since V is a vector space.
8. $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$ since V is a vector space.
9. $(rs)\mathbf{u} = r(s\mathbf{u})$ since V is a vector space.

10. $1\mathbf{u} = \mathbf{u}$ since V is a vector space.

Thus W satisfies the axioms of a vector space, and is itself a vector space. □

Example 3.20. If V is a vector space, then 0 and V are both subspaces of V. We don't actually need to check anything here, since both are clearly subsets of V, and both are already known to be vector spaces.

(When we want to ignore this possibility we will refer to "proper" subspaces, which are neither the trivial space nor the entire space).

Example 3.21. Let $V = \mathcal{P}(x)$ and let $W = \{a_1 x + \cdots + a_n x^n\} = x\mathcal{P}(x)$ be the set of polynomials with zero constant term. Is W a subspace of V?

1. The zero polynomial $0 + 0x + \cdots + 0x^n = 0$ certainly has zero constant term, so is in W.

2. If $a_1 x + \cdots + a_n x^n$ and $b_1 x + \cdots + b_n x^n \in W$, then

$$(a_1 x + \cdots + a_n x^n) + (b_1 x + \cdots + b_n x^n) = (a_1 + b_1)x + \cdots + (a_n + b_n)x^n \in W.$$

   Alternatively, we can say that if we add two polynomials with zero constant term, their sum will have zero constant term.

3. If $r \in \mathbb{R}$ and $a_1 x + \cdots + a_n x^n \in W$, then

$$r\,(a_1 x + \cdots + a_n x^n) = (ra_1)x + \cdots + (ra_n)x^n$$

   has zero constant term and is in W.

Thus W is a subspace of V.

Example 3.22. Let $V = \mathcal{P}(x)$ and let $W = \{a_0 + a_1 x\}$ be the space of linear polynomials. Then W is a subspace of V.

1. The zero polynomial $0 + 0x \in W$.

2. If $a_0 + a_1 x, b_0 + b_1 x \in W$, then $(a_0 + a_1 x) + (b_0 + b_1 x) = (a_0 + b_0) + (a_1 + b_1)x \in W$.

3. If $r \in \mathbb{R}$ and $a_0 + a_1 x \in W$, then $r(a_0 + a_1 x) = ra_0 + (ra_1)x \in W$.

Example 3.23. Let $V = \mathcal{P}(x)$ and let $W = \{1 + ax\}$ be the space of linear polynomials with constant term 1. Is W a subspace of V?

No, because $0 = 0 + 0x \notin W$.

Exercise 3.24. *Fix a natural number $n \geq 0$. Let $V = \mathcal{P}(x)$ and let $W = \mathcal{P}_n(x) = \{a_0 + a_1 x + \cdots + a_n x^n\}$ be the set of polynomials with degree at most n. Then $\mathcal{P}_n(x)$ is a subspace of $\mathcal{P}(x)$.*

Example 3.25. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ be the space of functions of one real variable, and let $W = \mathcal{D}(\mathbb{R}, \mathbb{R})$ be the space of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. Is W a subspace of V?

1. The zero function is differentiable, so the zero vector is in W.
2. From calculus we know that the derivative of the sums is the sum of the derivatives; thus the sum of differentiable functions is differentiable. That is, $(f+g)'(x) = f'(x) + g'(x)$.

   So if $f, g \in W$, then f and g are differentiable, and thus $f+g$ is differentiable and thus in W.
3. Again we know that $(rf)'(x) = rf'(x)$. If f is in W, then f is differentiable. Thus rf is differentiable and therefore in W.

Example 3.26. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and let $W = \mathcal{F}([a,b], \mathbb{R})$ be the space of functions from the closed interval $[a,b]$ to $\mathbb{R}$. We can view W as a subset of V by, say, looking at all the functions that are zero outside of $[a,b]$. Is W a subspace of V?

1. The zero function is in W.
2. If f and g are functions from $[a,b] \to \mathbb{R}$, then $(f+g)$ is as well.
3. If f is a function from $[a,b] \to \mathbb{R}$, then rf is as well.

Example 3.27. Let $V = \mathcal{F}(\mathbb{R}, R)$. Then $\mathcal{C}(\mathbb{R}, \mathbb{R})$ the space of *continuous* real-valued functions is a subspace of V. So iare $\mathcal{D}(\mathbb{R}, \mathbb{R})$ the space of differentiable functions and $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ the space of infinitely differentiable functions.

Example 3.28. Let $V = \mathcal{F}(\mathbb{R}, R)$ and let $W = \{f : f(x) = f(-x) \forall x \in \mathbb{R}\}$ be the set of *even* real-valued functions, the functions that are symmetric around 0. Then W is a subspace of V.

Example 3.29. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and let $W = \mathcal{F}(\mathbb{R}, [a,b])$ be the space of functions from $\mathbb{R}$ to the closed interval $[a,b]$. Is W a subspace of V?

No! The simplest condition to check is scalar multiplication. Let $f(x) = b$ be a function in V. Let $r = (b+1)/b$. Then $(rf)(x) = fb = b+1$ and thus $rf \notin W$.

Example 3.30. Let $V = S$ be the space of signals, and let W be the space of signals that are eventually zero. That is, $W = \{\{y_k\} : \exists n \text{ such that } y_m = 0 \, \forall \, m > n\}$. Then W is a subspace of V.

The space $\{\{y_k\} : y_0 = 0\}$ is a subspace of V. But the space $\{\{y_k\} : y_0 = 1\}$ is not.

3.3 Linear Transformations

We can also extend the definition of linear functions that we saw in section 1.8 to apply to abstract vector spaces. This will additionally give us a large supply of subspaces, since the kernel of any transformation is a subspace.

Definition 3.31. Let U and V be vector spaces, and let $L : U \to V$ be a function with domain U and codomain V. We say L is a *linear transformation* if:

1. Whenever $\mathbf{u}_1, \mathbf{u}_2 \in U$, then $L(\mathbf{u}_1 + \mathbf{u}_2) = L(\mathbf{u}_1) + L(\mathbf{u}_2)$.
2. Whenever $\mathbf{u} \in U$ and $r \in \mathbb{R}$, then $L(r\mathbf{u}) = rL(\mathbf{u})$.

Definition 3.32. Let $L : U \to V$ be a linear transformation. If $\mathbf{u} \in U$ is a vector, we say the element $L(\mathbf{u}) \in V$ is the *image* of $\mathbf{u}$.

If $S \subset U$ then we define the image of S to be the set $L(S) = \{L(\mathbf{u}) : \mathbf{u} \in S\}$ to be the set of images of elements of S. We say the image of the entire set U is the *image* of the function L.

The *kernel* of L is the set $\ker(L) = \{\mathbf{u} \in U : L(\mathbf{u}) = \mathbf{0}\}$ of elements of U whose image is the zero vector.

Another way of thinking about linear transformations is that they send lines to lines. In particular, the image of a subspace under a linear transformation is always a subspace—thus the image of a line will be either a point or a line.

Proposition 3.33. *Let $L : U \to V$ be a linear transformation, and let $S \subseteq U$ be a subspace of U. Then:*

1. $\ker(L)$ *is a subspace of U.*
2. *The image $L(S)$ of S is a subspace of V.*

Proof.
1. This is exactly the same as proposition 1.62.
2. We use the subspace theorem:

(a) We wish to show that $\mathbf{0} \in L(S)$. We claim in particular that $L(\mathbf{0}) = \mathbf{0}$: that is, the image of the zero vector in U must be the zero vector in V. Recall that $0 \cdot \mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in V$, so we have

$$L(\mathbf{0}) = L(0 \cdot \mathbf{0}) = 0L(\mathbf{0}) = \mathbf{0}.$$

Thus since S is a subspace we have $\mathbf{0} \in S$ and thus $\mathbf{0} \in L(S)$.

(b) Suppose $\mathbf{v} \in L(S)$ and $r \in \mathbb{R}$. Then there is some $\mathbf{u} \in S$ with $L(\mathbf{u}) = \mathbf{v}$, and since S is a subspace we know that $r\mathbf{u} \in S$. Thus

$$r\mathbf{v} = rL(\mathbf{u}) = L(r\mathbf{u}) \in L(S).$$

(c) Suppose $\mathbf{v}_1, \mathbf{v}_2 \in L(S)$. Then there exist $\mathbf{u}_1, \mathbf{u}_2 \in S$ such that $L(\mathbf{u}_1) = \mathbf{v}_1$ and $L(\mathbf{u}_2) = \mathbf{v}_2$. Since S is a subspace we know that $\mathbf{u}_1 + \mathbf{u}_2 \in S$. Then

$$\mathbf{v}_1 + \mathbf{v}_2 = L(\mathbf{u}_1) + L(\mathbf{u}_2) = L(\mathbf{u}_1 + \mathbf{u}_2) \in L(S).$$

□

Corollary 3.34. *If $L : U \to V$ is a linear transformation, then the image of L is a subspace of V.*

Example 3.35. Let $\mathcal{D}([a, b], \mathbb{R})$ be the space of continuously differentiable functions from the closed interval $[a, b]$ to the real line. Define the derivative operator $D : \mathcal{D}([a, b], \mathbb{R}) \to \mathcal{C}([a, b], \mathbb{R})$ by $D(f) = f'$. First we claim that D is a linear operator: we have that $D(f+g) = (f + g)' = f' + g' = D(f) + D(g)$, and $D(rf) = (rf)' + rf' = rD(f)$.

The kernel of D is the space of constant functions, which is a one-dimensional subspace. The image of D is actually a little hard to see, but it's actually the set of all continuous functions on $[a, b]$.

In other contexts we might write $\frac{d}{dx}$ instead of D for this linear transformation.

Generalizing this example gives us a lot of solution sets to differential equations.

Example 3.36. Let $L : \mathcal{D}([a, b], \mathbb{R}) \to \mathcal{C}([a, b], \mathbb{R})$ be defined by $L(f) = f' - f$. Then we can check that L is a linear function. The kernel of L is the set of functions f such that $f' = f$, which is the solutions to the differential equation $y' = y$. From Calculus 2 we know that this is the set $\{Ce^x : C \in \mathbb{R}\}$. (As we'll see in section 3.4, this is a one-dimensional subspace spanned by $\{e^x\}$.

Similarly, if we define $T(f) = f'' + f$, then T is a linear transformation whose kernel is the set of solutions to $y'' = -y$. This is the differential equation for simple harmonic motion; in a physics class you might learn that the solution set is $\{A\sin(x) + B\cos(x) : A, B \in \mathbb{R}\}$. This is a two-dimensional subspace spanned by $\{sin, \cos\}$.

Example 3.37. Let $\mathcal{C}([a,b], \mathbb{R})$ be the set of all continuous functions on the closed interval $[a, b]$. The (indefinite) integral isn't quite a linear transformation, since there's an ambiguity in choice of constant. (This is what we mean when we say something is "not well defined": if I tell you to give me the integral of x^2, you can't give me a specific function back so my question is not precise enough).

But the function $I(f) = \int_a^x f(t)\,dt$ is a linear transformation, since $\int_a^x (f+g)(t)\,dt = \int_a^x f(t)\,dt + \int_a^x g(t)\,dt$ and $\int_a^x rf(t)\,dt = r\int_a^x f(t)\,dt$. In this case the choice of a as the basepoint resolves the earlier ambiguity.

The kernel of I is the trivial vector space containing only the zero function. The image is again a bit hard to see, but works out to be the space of differentiable functions with the property that $F(a) = 0$.

This last example shows an important principle: our derivative and integral linear transformations (almost) undo each other. This is a very important property and we will look at it on its own in 3.6.

Example 3.38. Let S be the space of signals, and define the right-shift operator $R_1 : S \to S$ by $L\big(\{\dots, y_{-1}, y_0, y_1, \dots\}\big) = \{\dots, y_{-2}, y_{-1}, y_0, \dots\}$. Then R_1 is linear, and its kernel is empty. But if we consider the operator $R_1 - I$, this is still linear, but the kernel is precisely the one-dimensional subspace of constant signals, spanned by $\{\dots, 1, 1, 1 \dots\}$.

We can also define, e.g., the operator R_3 that shifts signals to the right by three spaces. This is a linear operator with empty kernel. The kernel of $R_3 - I$ is the three-dimensional subspace of periodic signals with period 3; it is spanned by

$$\big\{\{\dots, 1, 0, 0, 1, 0, 0 \dots\}, \{\dots, 0, 1, 0, 0, 1, 0, \dots\}, \{\dots, 0, 0, 1, 0, 0, 1 \dots\}\big\}.$$

Example 3.39. In quantum mechanics, the way a wave function $\psi(t)$ changes over time is given by the *Schrödinger equation*

$$i\hbar \frac{d}{dt}(\psi(t)) = \hat{H}(\psi(t)).$$

Here $\hat{H}$ is a linear operator called the *Hamiltonian* that encodes the properties of the system. A common example of a Hamiltonian is the operator

$$\frac{\hbar}{2m}\nabla^2 + V(\mathbf{r}, t)$$

where V itself describes the effects of some potential field.

A lot of the mathematics of quantum mechanics is analyzing these Hamiltonian operators to determine the behavior of a system. (In particular, physicists want to identify the "eigenvectors" or "eigenfunctions" of these operators; we'll discuss eigenvectors in the next section.)

3.4 Coordinates and Linear Transformations

We can recapitulate our definition of a basis in this new abstract setting.

Definition 3.40. Let V be a vector space and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ a set of vectors in V. We say that S is *linearly independent* if, whenever $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$, then $a_1 = \cdots = a_n = 0$.

We say that the *span* of the set S is the set $\operatorname{span}(S) = \{a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n\}$ of linear combinations of elements of V. We say that S *spans* V or is a *spanning set* for V if $\operatorname{span}(S) = V$.

If a set B is a spanning set for V that is also linearly independent, we say that B is a *basis* for V.

The fundamental idea here is that a basis is a set of coordinates we can use for our vector space. We've seen that the vectors $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ provide a system of coordinates for $\mathbb{R}^3$, by letting us write any vector as a linear combination of those three vectors. When we have a basis B for a vector space V, then we can write any vector in V uniquely as a linear combination of vectors in B.

Example 3.41. The set $S = \{1, x, x^2, x^3\}$ is a basis for $\mathcal{P}_3(x)$. So is the set $T = \{1 + x + x^2 + x^3, 1 + x + x^2, 1 + x, 1\}$.

Determining whether a set is a basis is sometimes annoying, but doesn't involve anything we haven't already done: a basis is just a set that both spans and is linearly independent, and we can check both properties individually. But we'd like to make things a little simpler.

Further, we want to talk about how "big" a space is, and this should plausibly be determined by how many elements there are in the basis. But since every space has more than one basis, talking about the size of "the" basis is potentially problematic. Fortunately, this is not an actual problem, as we shall see.

Lemma 3.42. *If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ spans a vector space V, and $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is a collection of vectors in V with $m > n$, then T is linearly dependent.*

Proof. There are two possible ways to prove this. One involves simply writing out a bunch of linear equations and solving them; this works, but is more tedious than informative. We'll use a more formal and abstract approach to proving this instead, which, hopefully, will actually explain some of *why* this is true.

We will start with the set S, and one by one we will trade out vectors in S for vectors in T, and show that we always still have a spanning set. We will suppose T is linearly independent, and show that $m \leq n$.

Since S is a spanning set, we know that $\mathbf{u}_1 \in \operatorname{span}(S)$, and thus $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}$ is linearly dependent by lemma 1.74. Then we can rewrite our linear dependence equation to express $\mathbf{v}_1$ (without loss of generality) as a linear combination of $\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = S_1$, and thus

$$\operatorname{span}(S) = \operatorname{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}) = \operatorname{span}(S_1).$$

We can repeat this process: at every step we add the next vector from T to get the set $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_k, \dots, \mathbf{v}_n\}$. Since S_{k-1} is a spanning set, this set is linearly dependent; since the $\mathbf{u}_i$ are linearly independent by hypothesis, we can remove one of the $\mathbf{v}_i$, and without loss of generality we can remove $\mathbf{v}_k$, to obtain the set $S_k = \{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$.

If $m > n$, we can continue until we have replaced every $\mathbf{v}_i$. Then we have $S_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a spanning set, and thus $\mathbf{u}_{n+1} \in \operatorname{span}(S_n)$ and so T is linearly dependent, which contradicts our assumption.

Thus if T is linearly independent, we must have $m \leq n$. Conversely, if $m > n$ then T is linearly dependent, as we stated. □

Corollary 3.43. *$S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ are two bases for a space V, then they are the same size, i.e. $m = n$.*

Proof. S is a spanning set and T is linearly independent, so we can't have $m > n$ by lemma 3.42. But T is a spanning set and S is linearly independent, so we can't have $n > m$ by lemma 3.42. Thus $n = m$. □

Definition 3.44. Let V be a vector space. If V has a basis consisting of n vectors, we say that V has *dimension* n and write $\dim V = n$. The trivial vector space $\{\mathbf{0}\}$ has dimension 0.

We say that V is *finite–dimensional* if there is a finite set of vectors that spans V. (Thus if V is n-dimensional it is finite-dimensional). Otherwise, we say that V is *infinite–dimensional.*

In this course we will primarily discuss finite dimensional vector spaces.

Example 3.45. The set $\{1, x, \dots, x^n\}$ is a basis for $\mathcal{P}_n(x)$. This set has $n + 1$ vectors, so $\dim \mathcal{P}_n(x) = n + 1$.

$\mathcal{P}(x)$ does not have a finite basis. We can see this since the set $S = \{1, x, \dots, x^n\}$ is linearly independent for any n; but every spanning set is at least as big as any linearly independent set, so we can never have a finite spanning set. However, if we allow infinite bases, then $\{1, x, \dots, x^n, \dots\}$ is a basis for $\mathcal{P}(x)$.

Remark 3.46. $\mathcal{C}([a, b], \mathbb{R})$ is infinite–dimensional, but if we allow infinite sums and make convergence arguments it is possible to think of the set $\{1, x, \dots, x^n, \dots\}$ as a sort of ("separable") basis. But this requires analysis and is outside the scope of this course. We can also build a (separable) basis out of the functions $\sin(nx)$ and $\cos(nx)$ for $n \in \mathbb{N}$; this is the foundation of Fourier analysis and Fourier series, which are important in signals analysis and many engineering disciplines.

The set $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is absurdly huge, and does not have a countable basis. If you believe the axiom of choice it has a basis, as all sets do, but you can't possibly write it down. You can think of it has having "coordinates" given by functions like

$$f_r(x) = \begin{cases} 1 & x = r \\ 0 & x \neq r \end{cases}$$

but this isn't a basis because it would require uncountable sums, which you can't really define.

As we saw in section 2.5, if we start with a spanning set or a linearly independent set, we can use it to find a basis.

Lemma 3.47 (Basis Reduction)**.** *Suppose $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a spanning set for V. Then S can be reduced to a basis for V. That is, there is a subset $B \subseteq S$ that is a basis for V.*

Proof. This is exactly the same as lemma 2.33. □

Remark 3.48. The lemma, and its proof, assume that S is finite. The result is still (mostly) true if S is infinite, but if the space is finite-dimensional this isn't interesting, and if the space is infinite-dimensional things get very complicated and we don't want to worry about them here.

Example 3.49. Let $S = \{1 - x, x^2 - x, 1 + x + x^2, x^2 - 1\}$ be a spanning set for $\mathcal{P}_2(x)$. Find a basis $B \subset S$.

We need to remove one vector which depends on the others. We need to find a nontrivial linear combination, so we have the equation

$$a(1 - x) + b(x^2 - x) + c(1 + x + x^2) + d(x^2 - 1) = 0$$

which gives the homogeneous system

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

which tells us that $a = d, b = -d, c = 0$.

Thus we have $(1-x) - (x^2 - x) + (x^2 - 1) = 0$ and so $x^2 - x = (1-x) + (x^2-1)$, and thus the element $x^2 - x$ is redundant and a linear combination of the other vectors. We can remove it, and get a basis $\{1 - x, 1 + x + x^2, x^2 - 1\}$.

Notice here that we could have removed the first element $1-x$ or the fourth element $x^2 - 1$, since we can rearrange our equation to write either of those as a linear combination of the others. But we could *not* have removed the element $1 + x + x^2$, since we didn't find we could write it as a combination of the others; it was in fact necessary for this set to span.

Now that we have a basis, we can express elements of our space $\mathcal{P}_2(x)$ in terms of this basis. We can do this, of course, using a row reduction.

We know that $3 + 2x + 4x^2 \in \mathcal{P}_2(x)$. Then we want to solve

$$\begin{aligned} a(1-x) + b(1+x+x^2) + c(x^2-1) &= 3 + 2x + 4x^2 \\ a + b - c &= 3 \\ (-a+b)x &= 2x \\ (b+c)x^2 &= 4x^2 \end{aligned}$$

which gives us three linear equations:

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

Thus we conclude that

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

We sometimes will say that the coordinates of $3 + 2x + 4x^2$ with respect to this basis are $(1, 3, 1)$.

We mentioned in section 2.5 that we could pad out a linearly independent set to get a basis. For abstract vector spaces this is often quite useful.

Lemma 3.50 (Basis Padding). *Suppose $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent in V. Then if V has any finite spanning set $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$, we can obtain a basis by padding S. That is, there is a basis B for V with $S \subseteq B$.*

Proof. If $T \subset \operatorname{span}(S)$, then $\operatorname{span}(T) \subset \operatorname{span}(S)$, so S is a spanning set for V and thus a basis, so we're done.

So suppose without loss of generality that $\mathbf{u}_1 \notin \operatorname{span}(S)$. Then $S_1 = \{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}$ is linearly independent by lemma 1.74 since we can't write any element as a linear combination of the others.

If S_1 spans V, then it is a basis and we're done. If not, there is some other $\mathbf{u}_i \notin \operatorname{span}(S_1)$, so we can repeat the process, and after at most m steps this process will terminate (since we run out of elements in T). When we reach a spanning set, this is our basis.

□

Example 3.51. Let $S = \{1 + x, x^2 - 3\} \subset \mathcal{P}_2(x)$. Can we find a basis B for $\mathcal{P}_2(x)$ that contains T?

We need to find a vector (or quadratic polynomial) that isn't in S. There are lots of choices here, but it looks to me like 1 is not in the span of S. Then we check: suppose $a(1 + x) + b(x^2 - 3) = 1$. Then we have

$$(a - 3b) + ax + bx^2 = 1$$

which gives the system

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

which has no solution. Thus indeed $1 \notin \operatorname{span}(S)$, so $\{1, 1 + x, x^2 - 3\}$ is a basis for $\mathcal{P}_2(x)$.

We can find the coordinates of $3 + 2x + 4x^2$ in this basis as well. We set up

$$\begin{aligned} a(1) + b(1 + x) + c(x^2 - 3) &= 3 + 2x + 4x^2 \\ a + b - 3c &= 3 \\ bx &= 2x \\ cx^2 &= 4x^2 \end{aligned}$$

so we can set up a matrix to row reduce:

$$\left[\begin{array}{ccc|c} 1 & 1 & -3 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 13 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

and we conclude that

$$13(1) + 2(1+x) + 4(x^2 - 3) = 3 + 2x + 4x^2.$$

Definition 3.52. If U is a vector space and $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for U, and $\mathbf{u} \in U$, we can write $\mathbf{u} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. We define the *coordinate vector* of $\mathbf{u}$ with respect to E by

$$[\mathbf{u}]_E = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

The a_i are called the *coordinates* of $\mathbf{u}$ with respect to the basis E.

We here observe that every $\mathbf{u} \in U$ corresponds to exactly one coordinate vector with respect to E, and vice versa. We will discuss this in more detail in 3.6.

Example 3.53. Let $U = \mathcal{P}_3(x)$. Then $E = \{1, x, x^2, x^3\}$ is a basis for U. Also, $F = \{1, 1+x, 1+x^2, 1+x^3\}$ is a basis for U.

Let $f(x) = 1 + 3x + x^2 - x^3 \in U$. Then

$$[f]_E = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -1 \end{bmatrix} \quad [f]_F = \begin{bmatrix} -2 \\ 3 \\ 1 \\ -1 \end{bmatrix}.$$

These are two different vectors of real numbers, but they represent the *same* element of U, just in different bases.

Example 3.54. Let $U = \mathbb{R}^3$ and let $E = \{(1,0,0), (1,1,0), (1,1,1)\}$. Then if $\mathbf{u} = (1,3,2)$, then

$$[\mathbf{u}]_E = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Remark 3.55. If B is the standard basis for $\mathbb{R}^n$, then any time we write a column vector there's an implicit $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_B$ that we just don't bother to write down.

Lemma 3.56. *If U is a vector space and $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for U, then the function $[\cdot]_E : U \to \mathbb{R}^n$ which sends $\mathbf{u}$ to $[\mathbf{u}]_E$ is a linear function.*

Proof. Let $\mathbf{u}, \mathbf{v} \in U$ and $r \in \mathbb{R}$. We can write

$$\mathbf{u} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$$
$$\mathbf{v} = b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n.$$

Then

$$\begin{aligned} [r\mathbf{u}] &= [ra_1\mathbf{e}_1 + \cdots + ra_n\mathbf{e}_n] = (ra_1, \dots, ra_n) = r(a_1, \dots, a_n) = r[\mathbf{u}]. \\ [\mathbf{u} + \mathbf{v}] &= [(a_1 + b_1)\mathbf{e}_1 + \cdots + (a_n + b_n)\mathbf{e}_n] = (a_1 + b_1, \dots, a_n + b_n) \\ &= (a_1, \dots, a_n) + (b_1, \dots, b_n) = [\mathbf{u}] + [\mathbf{v}]. \end{aligned}$$

Thus by definition, $[\cdot]_E$ is a linear transformation.

□

3.5 The Matrix of an Abstract Linear Transformation

Bases allow us to express our vectors in terms of coordinates, and coordinates let us pretend that our vector space is $\mathbb{R}^n$. In particular, they let us turn a lot of questions into questions about row reducing a matrix, which we already know how to do.

We also want to turn questions about linear transformations into questions about matrices. We saw in section 1.8.1 that we could view linear transformations $L : \mathbb{R}^n \to \mathbb{R}^m$ as matrices. But what about other linear transformations? It turns out that we can view *all* linear transformations (of finite-dimensional vector spaces) as matrices.

Theorem 3.57. *Let U and V be finite-dimensional vector spaces, with $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ a basis for U and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ a basis for V. Let $L : U \to V$ be a linear transformation.*

Then there is a matrix A that represents L with respect to E and F, such that $L\mathbf{u} = \mathbf{v}$ if and only if $A[\mathbf{u}]_E = [\mathbf{v}]_F$. The columns of A are given by $\mathbf{c}_j = [L(\mathbf{e}_j)]_F$.

Remark 3.58. This looks really complicated, but it really just says that any v linear transformation is determined entirely by what it does to the elements of some basis; if you have a basis and you know where your transformation sends each element of that basis, you know what it does to everything in your space.

In particular, if we have coordintes for our vector spaces, we can use a matrix to map one set of coordinates to the other, as if we were working in $\mathbb{R}^n$.

$$\begin{array}{ccc} U & \xrightarrow{L} & V \\ \downarrow{\scriptstyle [\cdot]_E} & & \downarrow{\scriptstyle [\cdot]_F} \\ \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \end{array} \qquad \begin{array}{ccc} \mathbf{u} & \xmapsto{L} & L(\mathbf{u}) \\ \downarrow{\scriptstyle [\cdot]_E} & & \downarrow{\scriptstyle [\cdot]_F} \\ [\mathbf{u}]_E & \xmapsto{A} & A[\mathbf{u}]_E = [L(\mathbf{u})]_F \end{array}$$

Proof. We just want to show that $A[\mathbf{u}]_E = [L(\mathbf{u})]_F$ for any $\mathbf{u} \in U$, where

$$A = [\mathbf{c}_1 \dots \mathbf{c}_n] = [[L(\mathbf{e}_1)]_F \dots [L(\mathbf{e}_n)]_F] .$$

Our proof is essentially the same as the proof of Proposition 1.86. Let $\mathbf{u} \in U$. Since E is a basis for U we can write $u = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. Then we have

$$\begin{aligned} [L(\mathbf{u})]_F &= [a_1L(\mathbf{e}_1) + \cdots + a_nL(\mathbf{e}_n)]_F = a_1 [L(\mathbf{e}_1)]_F + \cdots + a_n [L(\mathbf{e}_n)]_F \\ &= a_1\mathbf{c}_1 + \cdots + a_n\mathbf{c}_n; \\ A[\mathbf{u}]_E &= A [a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n]_E = A(a_1, \dots, a_n) = [\mathbf{c}_1 \dots \mathbf{c}_n] (a_1, \dots, a_n) \\ &= \mathbf{c}_1a_1 + \cdots + \mathbf{c}_na_n. \end{aligned}$$

Thus we have $[L(\mathbf{u})]_F = A [\mathbf{u}]_E$, so the matrix A does in fact represent the linear operator L. □

Example 3.59. Let $F = \{(1, 1), (-1, 1)\}$ be a basis for $\mathbb{R}^2$, and let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $L(x, y, z) = (x - y - z, x + y + z)$. Find a matrix for L with respect to the standard basis in the domain and F in the codomain.

$$\begin{aligned} L(1, 0, 0) &= (1, 1) = \mathbf{f}_1 \\ L(0, 1, 0) &= (-1, 1) = \mathbf{f}_2 \\ L(0, 0, 1) &= (-1, 1) = \mathbf{f}_2 \\ A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Example 3.60. Let S be the subspace of $\mathcal{C}([a, b], \mathbb{R})$ spanned by $\{e^x, xe^x, x^2e^x\}$, and let D be the differentiation operator on S. Find the matrix of D with respect to $\{e^x, xe^x, x^2e^x\}$.

We compute:

$$\begin{aligned} D(e^x) &= e^x = \mathbf{s}_1 \\ D(xe^x) &= e^x + xe^x = \mathbf{s}_1 + \mathbf{s}_2 \\ D(x^2e^x) &= 2xe^x + x^2e^x = 2\mathbf{s}_2 + \mathbf{s}_3 \\ A &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Example 3.61. Let $E = \{(1,1,0),(1,0,1),(0,1,1)\}$ and $F = \{(1,0,0),(1,1,0),(1,1,1)\}$ be bases for $\mathbb{R}^3$, and define $L(x,y,z) = (x+y+z, 2z, -x+y+z)$. We can check this is a linear transformation.

To find the matrix of L with respect to E and the standard basis, we compute

$$\begin{aligned} L(1,1,0) &= (2,0,0) \\ L(1,0,1) &= (2,2,0) \\ L(0,1,1) &= (2,2,2). \end{aligned}$$

Thus the matrix with respect to E and the standard basis is

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

If we want to find the matrix with respect to E and F, we observe that

$$\begin{aligned} L(1,1,0) &= (2,0,0) = 2(1,0,0) = 2\mathbf{f}_1 \\ L(1,0,1) &= (2,2,0) = 2(1,1,0) = 2\mathbf{f}_2 \\ L(0,1,1) &= (2,2,2) = 2(1,1,1) = 2\mathbf{f}_3. \end{aligned}$$

Thus the matrix is

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

We notice that this matrix is really simple; this is a "good" choice of bases for this linear transformation.

In contrast, let's look at the transformation $T(x, y, z) = (x, y, z)$. Then we have

$$\begin{aligned}
T(1,1,0) &= (1,1,0) = (1,1,0) = \mathbf{f}_2 \\
T(1,0,1) &= (1,0,1) = (1,0,0) - (1,1,0) + (1,1,1) = \mathbf{f}_1 - \mathbf{f}_2 + \mathbf{f}_3 \\
T(0,1,1) &= (0,1,1) = -(1,0,0) + (1,1,1) = -\mathbf{f}_1 + \mathbf{f}_3.
\end{aligned}$$

Thus the matrix of T with respect to E and F is

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

Thus this transformation, which is really simple with respect to the standard basis, is much more complicated with respect to these bases.

We'll talk a lot more about this choice of basis idea in section 5, and we'll talk about how to find the best basis for a given linear transformation in section 4.

This allows us to translate most of the results from section 2.5 to tell us about any linear transformation of vector spaces. The image of a transformation corresponds to the columnspace of the associated matrix; the kernel corresponds to the nullspace of the associated matrix.

Then this gives us a generalization of the rank-nullity theorem: the rank is the dimension of the rowspace, which is the dimension of the columnspace, which is the dimension of the image. And the nullity is the dimension of the nullspace, which is the dimension of the kernel. The rank-nullity theorem 2.45 tells us the rank and nullity add up to the number of columns of the associated matrix—which is the dimension of the domain of the linear transformation. All combined, this gives us

Theorem 3.62 (Rank-Nullity for Vector Spaces)**.** *Let U, V be finite-dimensional vector spaces, and $L : U \to V$ be a linear transformation. Then* $\dim \ker(L) + \dim \operatorname{Im}(L) = \dim U$.

Example 3.63. Define $L : \mathcal{P}_3(x) \to \mathcal{P}_3(x)$ be given by $L(f) = (1+x)f'' - f'$. We can take a standard basis $\{(1, x, x^2, x^3)\}$ and compute the matrix:

$$\begin{aligned}
L(1) &= 0 & L(x) &= -1 \\
L(x^2) &= (1+x)2 - 2x = 2 & L(x^3) &= (1+x)6x - 3x^2 = 6x + 3x^2
\end{aligned}$$

so we get the matrix

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

This is already pretty close to row-reduced, but we can finish it off and get the reduction

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

which has rank 2 and nullity 2. Thus the dimension of the image is 2; it's spanned by the second and fourth columns of the original matrix, which correspond to $\{-1, 6x + 3x^2\}$. And the dimension of the kernel is 2. The kernel satisfies $a_1 + a_2 = 0, a_3 = 0$, so it's the set $\{a_0 + a_1 x - a_1 x^2\}$.

As a final result, we will see that we now know about every possible subspace. We know that the kernel of a linear transformation is a subspace; but the converse is true as well, and every subspace is the kernel of some linear transformation.

Proposition 3.64. *Let V be a vector space and $U \subset V$ a subspace. Then U is the kernel of some linear transformation.*

Proof. We'll prove this in the case where U and V are finite-dimensional. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be a basis for U. By basis padding, there is a basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{v}_1, \dots, \mathbf{v}_m\}$ for the vector space V.

Define a linear transformation $L : V \to V$ by setting $L(\mathbf{u}_i) = \mathbf{0}$ and $L(\mathbf{v}_i) = \mathbf{v}_i$. That is, for any $\mathbf{v} \in V$, we can write

$$v = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n + b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m,$$

so we define

$$L(\mathbf{v}) = b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m.$$

Then the kernel of L is the set spanned by $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, which is just U. □

3.6 Isomorphisms

In the previous section, we looked at a lot of linear transformations. In particular, we saw the coordinate map $[\cdot]_E : U \to \mathbb{R}^n$ which sends a vector $\mathbf{u} \in U$ to its coordinates with respect to E. We observed that this mapping in fact goes both ways: if we have a vector we can compute the coordinates, and if we have coordinates we can compute the vectors. Functions like this are very important and have a special name.

Definition 3.65. Let $f : U \to V$ be a function. If there is a $g : V \to U$ such that $g(f(\mathbf{u})) = \mathbf{u}$ for all $u \in U$, and $f(g(\mathbf{v})) = \mathbf{v}$ for all $\mathbf{v} \in V$, then we say that $g = f^{-1}$ is the *inverse* of f, and that f is *invertible*.

If f is an invertible linear transformation, we say that f is an *isomorphism* between U and V.

If U and V are vector spaces, we say they are *isomorphic* if there exists an isomorphism from U to V. We write $U \cong V$.

Remark 3.66. We will see that if two spaces are isomorphic, we can treat them as essentially the same. This does not mean they are the same; $\mathbb{R}^5$ is not the same thing as the space of degree-four polynomials.

But if $U \cong V$ then they are the same *as vector spaces*, because if we want to do something to U, we can instead map it to V, do it there, and then map it back.

Example 3.67. Let U be a vector space with basis $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, and let $f : U \to \mathbb{R}^n$ be defined by $f(\mathbf{u}) = [\mathbf{u}]_E$. Then f is invertible, and the inverse of f is given by the function $g(a_1, \dots, a_n) = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. To prove this, we check two things.

For any $\mathbf{u} \in U$ we can write $\mathbf{u} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. Then

$$g(f(\mathbf{u})) = g(f(a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n)) = g(a_1, \dots, a_n) = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n = \mathbf{u}.$$

Similarly, for any $(a_1, \dots, a_n) \in \mathbb{R}^n$ we have

$$f(g(a_1, \dots, a_n)) = f(a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n) = (a_1, \dots, a_n).$$

Thus g is the inverse of f by definition.

Example 3.68. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x, y) = (x+y, x-y)$. Then f is invertible, with inverse $g(a, b) = \left(\frac{a+b}{2}, \frac{a-b}{2}\right)$. To prove this we check:

$$g(f(x,y)) = g(x+y, x-y) = \left(\frac{(x+y)+(x-y)}{2}, \frac{(x+y)-(x-y)}{2}\right) = (x,y)$$
$$f(g(a,b)) = f\left(\frac{a+b}{2}, \frac{a-b}{2}\right) = \left(\frac{a+b}{2} + \frac{a-b}{2}, \frac{a+b}{2} - \frac{a-b}{2}\right) = (a,b).$$

Thus g is the inverse of f by definition.

So far we can check whether a given g is the inverse of f, but we don't have a good way of determining if a function is invertible. But the concepts of one-to-one and onto from section 1.8.2 comes to the rescue.

Proposition 3.69. *Let $L : U \to V$ be a linear transformation of vector spaces. Then L is invertible if and only if L is bijective, i.e. it is one-to-one and onto.*

Proof. Suppose L is bijective. Define a transformation $T : V \to U$ as follows: let $\mathbf{v} \in V$. Then L is onto, so by definition there is some $\mathbf{u} \in U$ such that $L(\mathbf{u}) = \mathbf{v}$. Since L is one-to-one there is only one such element, since if $L(\mathbf{u}) = L(\mathbf{u}_1)$ then $\mathbf{u} = \mathbf{u}_1$ by definition of one-to-one. Define $T(\mathbf{v}) = \mathbf{u}$.

Then for any $\mathbf{v} \in V$, we have $L(T(\mathbf{v})) = L(\mathbf{u}) = \mathbf{v}$, and for any $\mathbf{u} \in U$ we have $T(L(\mathbf{u})) = T(\mathbf{v}) = \mathbf{u}$. Thus by definition, $T = L^{-1}$.

Conversely, suppose L is invertible, and let $T = L^{-1}$. Suppose $L(\mathbf{u}) = L(\mathbf{v})$. Then $T(L(\mathbf{u})) = T(L(\mathbf{v}))$, but $T(L(\mathbf{u})) = \mathbf{u}$ and $T(L(\mathbf{v})) = \mathbf{v}$, so $\mathbf{u} = \mathbf{v}$, and by definition L is one-to-one.

Let $\mathbf{v} \in V$. Then $T(\mathbf{v}) \in U$, and $L(T(\mathbf{v})) = \mathbf{v}$, so $\mathbf{v} \in L(U)$ for any $\mathbf{v}\ inV$. Thus L is onto by definition, and since it is one-to-one and onto, it is bijective.

□

Remark 3.70. Another way to think of the second result is that "onto" guarantees that every element $\mathbf{v} \in V$ has an inverse, and "one-to-one" guarantees that no element has more than one, so the inverse function is actually well-defined as a function.

We can contrast with, say, the function $f(x) = x^2$. This function is not one-to-one, so when we ask for the inverse or square root of 4, we get two possible answers. (This function isn't linear, but we didn't actually use linearity anywhere in the previous proof, and in fact it works for all functions).

Proposition 3.69 gives us an easy way to check if a linear function is injective, and if we also check that it is surjective we can easily see whether it is invertible. We can always check surjectivity directly, and on your homework you will, but we'd like to make this easier as well. To do that we want to convert the Rank-Nullity Theorem to discuss all linear transformations. First we need to lay some groundwork.

Recall the rank-nullity thoerem:

Theorem 3.71 (Rank-Nullity Theorem)**.** *If U, V are finite-dimensional, then* $\dim U = \dim \ker(L) + \dim L(U)$.

Corollary 3.72. *Let $L : U \to V$ be linear. Then L is one-to-one if and only if $\dim U = \dim L(U)$, and L is onto if and only if $\dim V = \dim(U) - \dim \ker(L)$.*

If $\dim U = \dim V$, L is an isomorphism if and only if $\ker(L) = \{\mathbf{0}\}$.

Now we can easily determine if a transformation is invertible. But how do we find the inverse? Like everything in linear algebra, it's easier to do computations if we change things to be a matrix.

Proposition 3.73. *Let $f : U \to V$ be a linear transformation of finite dimensional vector spaces, and let E, F be bases for U, V respectively. Let A be the matrix of f with respect to E, F. Then f is invertible if and only if A is invertible, and the matrix of f^{-1} is A^{-1}.*

Proof. Suppose f is invertible, and that the matrix of f is A and the Let B be the matrix of f^{-1}. Then for any $\mathbf{u} \in U$,

$$[f^{-1}(f(\mathbf{u}))]_E = B[f(\mathbf{u})]_F = BA[\mathbf{u}]_E$$
$$[f^{-1}(f(\mathbf{u}))]_E = [\mathbf{u}]_E$$

and thus $BA[\mathbf{u}]_E = [\mathbf{u}]_E$ for all $\mathbf{u} \in U$. Thus $BA\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, and thus $BA = I_n$. So by definition $B = A^{-1}$.

Conversely, suppose the matrix of f is A, and A has an inverse A^{-1}. Let g be the function corresponding to A^{-1}, so for all $\mathbf{v} \in V$ we have $[g(\mathbf{v})]_E = A^{-1}[\mathbf{v}]_F$. Then for any $\mathbf{u} \in U, \mathbf{v} \in V$, we compute

$$[g(f(\mathbf{u}))]_E = A^{-1}[f(\mathbf{u})]_F = A^{-1}A[\mathbf{u}]_E = [\mathbf{u}]_E$$
$$[f(g(\mathbf{v}))]_F = A[g(\mathbf{v})]_E = AA^{-1}[\mathbf{v}]_F = [\mathbf{v}]_F.$$

Thus $g(f(\mathbf{u})) = \mathbf{u}$ and $f(g(\mathbf{v})) = \mathbf{v}$, so by definition $g = f^{-1}$. □

Corollary 3.74. *A $n \times n$ matrix is invertible if and only if its nullspace is trivial.*

This gives us a method for finding the inverse of any linear transformation: we find the matrix of the transformation, use Gaussian elimination to invert the matrix, and then return to the corresponding transformation.

Example 3.75. Let $L(x, y, z) = (x + y + z, 2x + 3y + 2z, x + 5y + 4z)$. What is L^{-1}?

The matrix for L (with respect to the standard basis) is $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 1 & 5 & 4 \end{bmatrix}$. So we compute

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & 3 & 2 & 0 & 1 & 0 \\ 1 & 5 & 4 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 4 & 3 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & -1 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 3 & 7 & -4 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & -1 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 7/3 & -4/3 & 1/3 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2/3 & 1/3 & -1/3 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 7/3 & -4/3 & 1/3 \end{array}\right]$$

Since A is invertible, this tells us that L is also invertible. And from A^{-1} we can see that

$$L^{-1}(a,b,c) = (2a/3 + b/3 - c/3, -2a + b, 7a/3 - 4b/3 + c/3).$$

We can check by multiplying the original matrices together and seeing that we get the identity, or by computing $L^{-1}(L(x,y,z))$ and confirming that we get $(x,y,z$.

Example 3.76. Let $E : \mathcal{P}_2(x) \to \mathbb{R}^3$ be given by $E(f(x)) = (f(-1), f(0), f(1))$. Can we find an inverse for this function?

Let $\{1, x, x^2\}$ be the basis for $\mathcal{P}_2(x)$, and use the standard basis for $\mathbb{R}^3$. Then the matrix of this transformation is $A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$. To find the inverse, we compute

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & -1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1/2 & -1 & 1/2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 1/2 & -1 & 1/2 \end{array}\right].$$

We can double-check our work, again, by multiplying out the original matrices.

What have we concluded? If we have a quadratic polynomial such that $f(-1) = a, f(0) = b, f(1) = 3$, then we must have

$$f(x) = b + (c/2 - a/2)x + (a/2 + c/2 - b)x^2.$$

Thus we can use this technique to find the (minimal degree) polynomial that goes through a given set of points.

Example 3.77. Let $D : \mathcal{P}_3(x) \to \mathcal{P}_3(x)$ be given by the derivative map. Is this function invertible?

The function is not invertible, since it has non-trivial kernel. We can also see this by writing down the matrix relative to the obvious basis: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Since there is a row of all zeroes, the rows are not linearly independent, so the matrix is not invertible.

Let's tweak things a bit. Let $Q = \{ax + bx^2 + cx^3\} \subset \mathcal{P}_3(x)$, and let $D : Q \to \mathcal{P}_2(x)$ be given by the derivative. Then if we let $E = \{x, x^2, x^3\}$, $F = \{1, x, x^2\}$ be bases for the domain and codomain, we see the matrix is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ whose inverse is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$. Thus this function is invertible; in fact we see the inverse sends $a + bx + cx^2 \mapsto ax + bx^2/2 + cx^3/3$, which you should recognize as an integral.

Example 3.78. Let $L : \mathcal{P}_2(x) \to \mathcal{P}_2(x)$ given by $L(f(x)) = f(x) + f'(x)$. Find an inverse or prove no inverse exists.

We take the basis $\{1, x, x^2\}$ and we have $L(1) = 1, L(x) = x + 1, L(x^2) = x^2 + 2x$. Thus we get the matrix

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

which has inverse

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

(which we can find through a row reduction). Thus our inverse is given by

$$L^{-1}(a + bx + cx^2) = (a - b + 2c) + (b - 2c)x + cx^2.$$

Example 3.79. Let $T : \mathcal{P}_2(x) \to \mathbb{R}^3$ given by $T(f(x)) = \begin{bmatrix} f(0) \\ f(1) \\ f(2) \end{bmatrix}$. Find an inverse or prove no inverse exists.

With the obvious standard bases, we have $T(1) = (1,1,1), T(x) = (0,1,2), T(x^2) = (0,1,4)$. So we get the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

which has inverse

$$\begin{bmatrix} 1 & 0 & 0 \\ -3/2 & 2 & -1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ -3 & 4 & -1 \\ 1 & -2 & 1 \end{bmatrix}$$

(which we can find through a row reduction). Thus our inverse is given by

$$T^{-1} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a + (-3a/2 + 2b - c/2)x + (a/2 - b + c/2)x^2.$$

As a final coda, we'll note that every vector space is isomorphic to itself. This fact will be really important when we want to change coordinate systems.

Proposition 3.80. *Let V be a vector space. Then $V \cong V$. In particular, the identity map Id_V defined by $Id_V(\mathbf{v}) = \mathbf{v}$ is an isomorphism from V to V.*

But two isomorphic vector spaces can have more than one isomorphism between them. In fact, any non-trivial vector space has infinitely many isomorphisms from itself to itself, and these isomorphisms are extremely useful.

Proposition 3.81. *Let U, V be vector spaces, let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for U, and $L : U \to V$ be a linear map. Then $L(E)$ spans $L(U)$, and L is an isomorphism if and only if $L(E)$ is a basis for V.*

Corollary 3.82. *Let U, V be vector spaces, with $\dim U = \dim V$. Then $U \cong V$.*

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for U, and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ a basis for V; we know they have the same number of elements since $\dim U = \dim V$. Define a linear map $L : U \to V$ by linearly extending $L(\mathbf{e}_i) = \mathbf{f}_i$; that is, we define

$$L(a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n) = a_1\mathbf{f}_1 + \cdots + a_n\mathbf{f}_n.$$

This map sends a basis to a basis, and thus is an isomorphism. □

It's in this sense that every finite-dimensional vector space "is really just $\mathbb{R}^n$". Any vector space of dimension n is isomorphic to $\mathbb{R}^n$, because we can specify a vector by giving n coordinates, which just means n real numbers. So for the purposes of doing linear algebra we can treat them as the same. Thus $\mathcal{P}_3(x)$, $M_{2\times 2}$, and $\operatorname{span}\{\sin(x), \sin(2x), \sin(3x), \sin(4x)\}$ can all be viewed as copies of $\mathbb{R}^4$.

Corollary 3.83. *A linear map from V to V is an isomorphism if and only if it sends a basis to a basis.*

4 Eigenvectors and Eigenvalues

In this section we will study a special type of basis, called an eigenbasis. For (almost) any given operator, we get a specific basis which will make most our computations easier.

4.1 Eigenvectors

Definition 4.1. Let $L : V \to V$ be a linear transformation, and let λ be a scalar. If there is a vector $\mathbf{v} \in V$ such that $L\mathbf{v} = \lambda \mathbf{v}$, then we say that λ is an *eigenvalue* of L, and $\mathbf{v}$ is an *eigenvector* with eigenvalue λ.

Geometrically, an eigenvector corresponds to a direction in which our linear operator purely stretches or shrinks vectors, without rotating or reflecting them at all. It can often be an axis of rotation.

Example 4.2. Let $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$. We can check that if $\mathbf{x} = (2, 1)$, then

$$A\mathbf{x} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

so $\mathbf{x}$ is an eigenvector with eigenvalue 3. Similarly, we can check that if $\mathbf{y} = (1, 1)$, then

$$A\mathbf{y} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus $\mathbf{y}$ is an eigenvector with eigenvalue 2.

Example 4.3. Let $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation map. We can see geometrically that this has no non-trivial eigenvectors, since it changes the direction of any vector. Algebraically, if (x, y) is an eigenvector, then we would have

$$R_{\pi/2}(x, y) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix}$$

and thus weh ave $\lambda y = x, \lambda x = -y$, and the only solution here is $x = y = 0$.

In contrast, if we take the rotation map $R : \mathbb{R}^3 \to \mathbb{R}^3$ that rotates around the z-axis, the vector $(0, 0, 1)$ will be an eigenvector with eigenvalue 1.

Example 4.4. Let $V = \mathcal{D}(\mathbb{R}, \mathbb{R})$ be the space of differentiable real functions, and let $\frac{d}{dx} : V \to V$ be the derivative map. If $f(x) = e^{rx}$, then $\frac{d}{dx} f(x) = re^{rx} = rf(x)$, so f is an eigenvector with eigenvalue r.

Proposition 4.5. *Let V be a vector space and $L : V \to V$ a linear transformation.* $\mathbf{v}$ *is an eigenvector with eigenvalue* λ *if and only if* $\mathbf{v} \in \ker(L - \lambda I)$.

Proof. $\mathbf{v}$ is an eigenvector with eigenvalue λ if and only if $L\mathbf{v} = \lambda \mathbf{v} = \lambda I \mathbf{v}$, if and only if $\mathbf{0} = L\mathbf{v} - \lambda I \mathbf{v} = (L - \lambda I)\mathbf{v}$, if and only if $\mathbf{v} \in \ker(L - \lambda I)$. □

Corollary 4.6. *The set of eigenvectors with eigenvalue λ is a subspace of V, called the* eigenspace *corresponding to λ. We denote this space E_λ.*

Corollary 4.7. *A transformation L is invertible if and only if* 0 *is not an eigenvalue of L.*

Proposition 4.8. *Let $L : V \to V$ be a linear transformation. If $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a set of eigenvectors each with a distinct eigenvalue, then E is linearly independent.*

Proof. Let λ_i be the eigenvalue corresponding to $\mathbf{e}_i$. Suppose (for contradiction) that E is linearly dependent, and let k be the smallest positive integer such that $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is linearly dependent; then we must have $a_k \neq 0$, and we can compute

$$\begin{aligned} \mathbf{e}_k &= \frac{-a_1}{a_k}\mathbf{e}_1 + \cdots + \frac{-a_{k-1}}{a_k}\mathbf{e}_{k-1} \\ L(\mathbf{e}_k) &= L\left(\frac{-a_1}{a_k}\mathbf{e}_1 + \cdots + \frac{-a_{k-1}}{a_k}\mathbf{e}_{k-1}\right) = \frac{-a_1}{a_k}L(\mathbf{e}_1) + \cdots + \frac{-a_{k-1}}{a_k}L(\mathbf{e}_{k-1}) \\ \lambda_k \mathbf{e}_k &= \frac{-a_1}{a_k}\lambda_1\mathbf{e}_1 + \cdots + \frac{-a_{k-1}}{a_k}\lambda_{k-1}\mathbf{e}_{k-1}. \end{aligned}$$

We can multiply the first equation by λ_1 and subtract from the last equation; this gives us

$$\mathbf{0} = \frac{-a_1}{a_k}(\lambda_1 - \lambda_k)\mathbf{e}_1 + \cdots + \frac{-a_{k-1}}{a_k}(\lambda_{k-1} - \lambda_k)\mathbf{e}_{k-1}.$$

But we know by hypothesis that the set $\{\mathbf{e}_1, \dots, \mathbf{e}_{k-1}\}$ is linearly independent, so all these coefficients must be zero. Since the a_i are not all zero, we must have at least some $\lambda_i - \lambda_k = 0$. □

It's straightforward enough to *check* that a vector is an eigenvector if we already have a candidate; but how do we find them? Sometimes this is easy

Example 4.9. Let $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. What are the eigenvalues and eigenspaces of A?

We see that

$$A\mathbf{x} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x \\ 2y \end{bmatrix}.$$

Thus the eigenvalues are 3 and 2; the corresponding eigenspaces are spanned by $(1, 0)$ and $(0, 1)$, respectively.

When things aren't this easy, there is still a fairly straightforward approach we can take:

Example 4.10. Let $B = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix}$. Find the eigenvalues and eigenvectors of B.

If $\mathbf{x} = (x, y)$ is an eigenvector with eigenvalue λ, then we have

$$B\mathbf{x} = \begin{bmatrix} 7x + 2y \\ 3x + 8y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix}$$

so we have the system of equations $7x + 2y = \lambda x, 3x + 8y = \lambda y$. Equivalently, we have $(7 - \lambda)x + 2y = 0$ and $(3x + (8 - \lambda)y = 0$. We row-reduce

$$\begin{bmatrix} 7-\lambda & 2 \\ 3 & 8-\lambda \end{bmatrix} \to \begin{bmatrix} 3 & 8-\lambda \\ 0 & 2 + (8-\lambda)(\lambda-7)/3 \end{bmatrix}$$
$$\to \begin{bmatrix} 3 & 8-\lambda \\ 0 & 6 + (-56 + 15\lambda - \lambda^2) \end{bmatrix} = \begin{bmatrix} 3 & 8-\lambda \\ 0 & -\lambda^2 + 15\lambda - 50 \end{bmatrix}.$$

We first see that this is solveable if and only if $0 = \lambda^2 - 15\lambda + 50 = (\lambda - 5)(\lambda - 10)$, and thus if $\lambda = 5$ or $\lambda = 10$. Thus these are the two eigenvalues for B.

If $\lambda = 5$ then we have $3x + 3y = 0$ so $y = -x$. Any vector $(\alpha, -\alpha)$ will be an eigenvector with eigenvalue 5, so the eigenspace for 5 is the span of $\{(1, -1)\}$. And indeed, we compute

$$B(1, -1) = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ -5 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

If $\lambda = 10$ then we have $3x - 2y = 0$ and $y = 3/2x$. Thus any vector $(2\alpha, 3\alpha)$ will be an eigenvector with eigenvalue 10, and the corresponding eigenspace is spanned by $\{(2, 3)\}$. We check:

$$B(2, 3) = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 20 \\ 30 \end{bmatrix} = 10 \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

As the previous example shows, it is completely possible to find the eigenvectors and eigenvalues with the tools we have already, but it's pretty fiddly even for a small example. We'd like to streamline the process, and this leads us to define the determinant.

4.2 Determinants

Definition 4.11. Let $A \in M_{n \times n}$. If A has n distinct eigenvalues, we say that the *determinant* of A, written $\det A$, is the product of the eigenvalues.

More generally, the determinant of A is the product of the eigenvalues "up to multiplicity". Thus if the eigenspace of $\lambda = 2$ is three-dimensional, we will multiply in λ three times.

Definition 4.12 (Formal definition we won't really use)**.**

$$\det A = \prod_{\lambda} \lambda^{e_\lambda} \qquad \text{where} \quad e_\lambda = \dim \ker(A - \lambda I)^n.$$

The determinant is (roughly) the product of the eigenvalues, so it can tell something about what the eigenvalues are. But this doesn't help if we don't have a way of finding the determinant without already knowing the eigenvalues. Fortunately, there is a simple way to compute it.

Example 4.13. The determinant of $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ is $3 \cdot 2 = 6$.

The determinant of $B = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix}$ is $5 \cdot 10 = 50$.

Geometrically, the determinant represents the volume of the n-dimensional solid that our matrix sends the n-dimensional unit cube to; thus it tells us how much our matrix stretches its inputs.

4.2.1 The Laplace Formula

We first need to develop some notation.

Definition 4.14. Let $A = (a_{ij})$ be a $n \times n$ matrix. We define the i,j*th minor matrix* of A to be the $(n-1) \times (n-1)$ matrix M_{ij} obtained by deleting the row and column containing a_{ij}—that is, deleting the ith row and jth column.

We define the i,j*th minor* of A to be $\det M_{ij}$. We define the i,j*th cofactor* to be $A_{ij} = (-1)^{i+j} \det(M_{ij})$.

Example 4.15. Let

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

Then we have

$$M_{1,1} = \begin{bmatrix} -2 & -1 \\ 3 & 3 \end{bmatrix} \qquad M_{3,2} = \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix}.$$

Fact 4.16 (Cofactor Expansion). *Let A be a $n \times n$ matrix.*

If $A \in M_{1\times 1}$ then $A = \begin{bmatrix} a_{11} \end{bmatrix}$ and $\det A = a_{11}$.

Otherwise, for any k we have

$$\begin{aligned}\det(A) &= \sum_{i=1}^{n} a_{ki}A_{ki} = a_{k1}A_{k1} + a_{k2}A_{k2} + \cdots + a_{kn}A_{kn} \\ &= \sum_{i=1}^{n} a_{ik}A_{ik} = a_{1k}A_{1k} + a_{2k}A_{2k} + \cdots + a_{nk}A_{nk}.\end{aligned}$$

Thus we may compute the determinant of a matrix inductively, using cofactor expansion. We can expand along any row or column; we should pick the one that makes our job easiest.

Remark 4.17. This is usually taken to be the definition of determinant. Feel free to think of it that way, and the fact about eigenvectors as a theorem.

You can also think of the determinant as the unique multilinear map that satisfies certain properties. You probably shouldn't, at the moment. But you can.

Example 4.18. Let $A = \begin{bmatrix} 3 & 2 & 1 \\ 0 & 5 & 1 \\ 0 & 0 & 2 \end{bmatrix}$. If we expand along the last row, we get

$$\begin{aligned}\det A &= 0 \cdot (-1)^{3+1} \det \begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix} + 0 \cdot (-1)^{3+2} \det \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix} + 2 \cdot (-1)^{3+3} \det \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \\ &= 2 \det \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = 2 \left(0 \cdot (-1)^{2+1} \det \begin{bmatrix} 2 \end{bmatrix} + 5 \cdot (-1)^{2+2} \det \begin{bmatrix} 3 \end{bmatrix} \right) \\ &= 2(0 + 5 \cdot 3) = 30.\end{aligned}$$

Example 4.19. Let

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

We'd like to expand along the row or column wiht the most zeros, but we don't have any. I'm going to expand along the bottom row because at least everything is the same.

$$\begin{aligned}\det A &= 3(-1)^{3+1} \det \begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix} + 3(-1)^{3+2} \det \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} + 3(-1)^{3+3} \begin{bmatrix} 3 & 1 \\ 5 & -2 \end{bmatrix} \\ &= 3 \left(1(-1)^{1+1}(-1) + 2(-1)^{1+2}(-2) \right) - 3 \left(3(-1)^{1+1}(-1) + 2(-1)^{1+2}5 \right) \\ &\quad + 3 \left(3(-1)^{1+1}(-2) + 1(-1)^{1+2}(5) \right) \\ &= 3(-1+4) - 3(-3-10) + 3(-6-5) = 9 + 39 - 33 = 15.\end{aligned}$$

Using this method, we can compute the determinant of any size of matrix. But for small matrices we can work out quick formulas that encode all this information.

Proposition 4.20.

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc \qquad \det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - gec - hfa - idb.$$

4.2.2 Properties of Determinants

We'd like to do things to make computing determinants easier, in addition to the formulas I just gave. We can start by proving some simple results.

Proposition 4.21. *If A is a $n \times n$ triangular matrix, then $\det A$ is the product of the diagonal entries of A.*

Proof. We use cofactor expansion; at each step, we have a row or column with only one non-zero entry, on the diagonal. At the end of the cofactor expansion we have simply taken the product of the diagonal entries. □

Proposition 4.22. *If A has a row or column of all zeroes, then $\det A = 0$.*

Proof. Do cofactor expansion along the row of all zeros. □

Proposition 4.23. $\det A^T = \det A$.

Proof. Do a cofactor expansion along the column of A^T that corresponds to the row you expanded along in A, or vice versa. □

Fact 4.24 (Row Operations)**.**

- *Interchanging two rows multiplies the determinant by -1.*
- *Multiplying a row by a scalar multiplies the determinant by that scalar.*
- *Adding a multiple of one row to another row does not change the determinant.*
- $$\det \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{r}_n \end{bmatrix} + \det \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{b}_i \\ \vdots \\ \mathbf{r}_n \end{bmatrix} = \det \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{a}_i + \mathbf{b}_i \\ \vdots \\ \mathbf{r}_n \end{bmatrix}.$$

Proof. The proof is really tedious and just involves a bunch of inductions on cofactor expansions. □

Example 4.25.

$$\det \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} = 1 \qquad \det \begin{bmatrix} 3 & 3 & 3 \\ 0 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} = 3$$

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

Corollary 4.26. $\det A = 0$ *if and only if the rows of A are linearly dependent.*

Proposition 4.27. *A matrix A is invertible if and only if* $\det A \neq 0$.

Proof. We can view this proof in two different ways.

From the eigenvalue perspective: $\det A$ is the product of the eigenvalues. Thus $\det A = 0$ if and only if 0 is an eigenvalue of A. But 0 is an eigenvalue of A if and only if A has non-trivial kernel, and A is invertible if and only if $\ker(A)$ is trivial.

From the cofactor perspective: if A is invertible it is row-equivalent to the identity matrix, which has determinant 1. None of the row operations can change a determinant from zero to non-zero or vice versa, so $\det A$ is nonzero.

Conversely, if A is not invertible, it is row-equivalent to a matrix with a row of all zeros, which has determinant zero. Since row operations cannot change a determinant from non-zero to zero, $\det A = 0$ as well. □

Fact 4.28. *If A, B are $n \times n$ matrices, then* $\det(AB) = \det(A)\det(B)$.

Corollary 4.29. *If A is a nonsingular matrix, then* $\det(A^{-1}) = \frac{1}{\det A}$.

Remark 4.30. This is why the inverse of a matrix so often has the same denominator appearing in most of the entries; it's the reciprocal of the determinant.

Example 4.31. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

We check this by multiplying the two of them:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} ad-bc & -ab+ba \\ cd-dc & -bc+ad \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

4.3 Characteristic Polynomials

Definition 4.32. We say that $\chi_A(\lambda) = \det(A - \lambda I)$ is the *characteristic polynomial* of A. This is a polynomial in one variable, λ. We call the equation $\chi_A(\lambda) = 0$ the *characteristic equation* of A.

Proposition 4.33. *The real number λ is an eigenvalue of A if and only if it is a root of the characteristic polynomial of A. That is, the roots of $\chi_A(\lambda)$ is the set of eigenvalues of A.*

Proof. Recall that $\mathbf{v}$ is an eigenvector with eigenvalue λ if and only if $\mathbf{v} \in \ker(A - \lambda I)$. Thus λ is an eigenvalue if and only if $\ker(A - \lambda I)$ has nontrivial kernel, which occurs if and only if $\det(A - \lambda I) = 0$. □

Example 4.34. Find the eigenvalues and corresponding eigenspaces of $A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$.

The characteristic equation is

$$\begin{aligned} 0 = \chi_A(\lambda) &= \begin{vmatrix} 3-\lambda & 2 \\ 3 & -2-\lambda \end{vmatrix} \\ &= (3-\lambda)(-2-\lambda) - 2 \cdot 3 = -6 - 3\lambda + 2\lambda + \lambda^2 - 6 \\ &= \lambda^2 - \lambda - 12 = (\lambda - 4)(\lambda + 3) \end{aligned}$$

so the eigenvalues are 4 and -3. We compute

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

so $\ker(A - 4I) = \{\alpha(2,1)\}$. Thus the eigenspace corresponding to 4 is $E_4 = \operatorname{span}\{(2,1)\}$. Similarly,

$$A + 3I = \begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix} \to \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix}$$

so $\ker(A + 3I) = \{\alpha(-1,3)\}$. Thus the eigenspace $E_{-3} = \operatorname{span}\{(-1,3)\}$.

Example 4.35. Find the eigenvalues and corresponding eigenspaces of $A = \begin{bmatrix} 5 & 1 \\ 3 & 3 \end{bmatrix}$.

The characteristic equation is

$$\begin{aligned} 0 = \chi_A(\lambda) &= \left| \begin{pmatrix} 5-\lambda & 1 \\ 3 & 3-\lambda \end{pmatrix} \right| \\ &= (3-\lambda)(5-\lambda) - 1 \cdot 3 = 15 - 8\lambda + \lambda^2 - 3 \\ &= \lambda^2 - 8\lambda + 12 = (\lambda - 6)(\lambda - 2) \end{aligned}$$

so the eigenvalues are 6 and 2.

$$A - 6I = \begin{bmatrix} -1 & 1 \\ 3 & -3 \end{bmatrix} \to \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$

has kernel $\{\alpha(1,1)\}$, so the eigenspace $E_6 = \operatorname{span}\{(1,1)\}$.

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

has kernel $\{\alpha(-1,3)\}$, so the eigenspace $E_2 = \operatorname{span}\{(-1,3)\}$.

Example 4.36. Find the eigenvalues and corresponding eigenspaces of $A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$.

The characteristic equation is

$$\begin{aligned} 0 = \chi_A(\lambda) &= \left| \begin{pmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{pmatrix} \right| \\ &= (2-\lambda)(-2-\lambda)(2-\lambda) - 3 - 3 - ((-2-\lambda) - 3(2-\lambda) - 3(2-\lambda)) \\ &= -\lambda^3 + 2\lambda^2 + 4\lambda - 8 - 6 + 2 + \lambda + 12 - 6\lambda \\ &= -\lambda^3 + 2\lambda^2 - \lambda = -\lambda(\lambda-1)^2 \end{aligned}$$

so the eigenvalues are 0 and 1 (twice). We have

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

so $\ker(A) = \{\alpha(1,1,1)\}$, and $E_0 = \operatorname{span}\{(1,1,1)\}$. We also have

$$A - I = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 1 \\ 1 & -3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -3 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so $\ker(A - I) = \{\alpha(3,1,0) + \beta(-1,0,1)\}$, and $E_1 = \operatorname{span}\{(3,1,0), (-1,0,1)\}$.

Proposition 4.37. *If A is a $n \times n$ matrix and n is odd, then A has at least one eigenvalue.*

Proof. Recall that a degree n polynomial always has at least one real root if n is odd. Thus if $A \in M_{n \times n}$, $\chi_A(\lambda)$ is degree n, and has a real root, which is an eigenvalue of A. □

Example 4.38. Find the eigenvalues and corresponding eigenspaces of $B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}$.

Since this matrix is triangluar, we know the eigenvalues are $2, 4, 2$. We solve

$$A - 2I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and $\ker(A - 2I) = \{\alpha(0, 0, 1)\}$, so $E_2 = \operatorname{span}\{(0, 0, 1)\}$. Similarly,

$$A - 4I = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

so $\ker(A - 4I) = \{\alpha(0, 1, 0)\}$ so $E_4 = \operatorname{span}\{(0, 1, 0)\}$.

Notice that in this case, the span of the eigenvectors is only 2-dimensional; the eigenvectors don't span the whole domain.

4.4 When Eigenvectors Don't Work

It would be nice if every matrix had a spanning set of eigenvectors, but we've already seen that isn't always the case. There are two different things that can happen here.

4.4.1 Complex Eigenvalues

Example 4.39. Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. (Recall this is the matrix for a rotation 90 degrees counterclockwise.) Then we can compute the characteristic polynomial

$$\chi_A(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1.$$

Then $\chi_A(\lambda)$ has no real roots, and thus A has no eigenvalues or (nontrivial) eigenvectors.

However, if you're familiar with complex numbers, you might note that $\chi_A(\lambda)$ has two complex roots: $\pm i$. Then we can try to find the corresponding eigenspaces.

For E_i, we have

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \to \begin{bmatrix} 1 & -i \\ -i & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$$

and thus $E_i = \ker(A - iI) = \{(iy, y)\}$. So we have $(i, 1)$ as an eigenvector with eigenvalue i. We can check this:

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

Similarly, we can compute E_{-i}. We have

$$\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \to \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$$

and so $E_{-i} = \{(-iy, y)\}$; we have $(-i, 1)$ as an eigenvector with eigenvalue $-i$.

Why do we want to do this? To me as a mathematician, this is just kind of fun; we have a question we can't quite answer, so we change the question a bit. But from a more practical perspective, we can use these complex or imaginary solutions to tell us things about the original real-valued question.

For instance, in this case, the eigenstructure tells us that this matrix is a rotation. If $\mathbf{v} = (i, 1)$, we know that $A\mathbf{v} = i\mathbf{v}$, so $A^2\mathbf{v} = (-1)\mathbf{v}$ and $A^4\mathbf{v} = (i)^4\mathbf{v} = \mathbf{v}$. We can set $\mathbf{u} = (-i, 1)$ and we'll similarly have $A^4\mathbf{u} = (i)^4\mathbf{u} = \mathbf{u}$. So A^4 will leave both eigenvectors unchanged.

It might be helpful at this point to review some facts about complex numbers.

Definition 4.40. We define $i = \sqrt{-1}$ to be a solution to the equation $x^2 + 1 = 0$. We say that the *complex numbers*, written $\mathbb{C}$, are the set $\{a + bi : a, b \in \mathbb{R}\}$.

If $z = a + bi$ is a complex number, we say its *real part* is $Re(z) = a$ and its *imaginary part* is $Im(z) = b$.

We further give $\mathbb{C}$ the obvious addition and multiplication operators:

- $(a + bi) + (c + di) = (a + c) + (b + b + 2)i$
- $(a + bi)(c + di) = ab + (bc + ad)i + cdi^2 = (ab - cd) + (bc + ad)i^2$.

These are mostly interesting because of the following theorem:

Theorem 4.41 (Fundamental Theorem of Algebra)**.** *Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial with all the $a_i \in \mathbb{C}$. Then f has a* root *over $\mathbb{C}$; that is, there is a complex number $z \in \mathbb{C}$ such that $f(z) = 0$.*

In fact, over the complex numbers we can completely factor any polynomial; we can write $f(x) = \prod_{i=1}^{n}(x - b_i)$ for some set of $b_i \in \mathbb{C}$.

The natural next question is whether i was unique. We defined i to be a solution to the equation $x^2 + 1 = 0$, but in fact there is a second solution: $(-i)^2 + 1 = -1 + 1 = 0$, so $-i$ would also have been a solution. That suggests that replacing i with $-i$ should leave any algebraic statement—one defined by polynomials, unchanged. (This sort of argument is the underpinning of an algebraic field called Galois theory, which we won't really be discussing more in this course). Thus we define:

Definition 4.42. If $z = x + yi$ is a complex number, we say the *complex conjugate* of z is $\overline{z} = x - yi$.

Proposition 4.43.

- $\overline{a} + \overline{b} = \overline{a + b}$
- $\overline{a}\overline{b} = \overline{ab}$
- $\overline{a}^{\alpha} = \overline{a^{\alpha}}$

Remark 4.44. We said earlier that complex conjugation should preserve algebraic statements. We can make this more precise: if f is a real polynomial and c is a complex number, then $f(c) = 0$ if and only if $f(-c) = 0$. Thus every polynomial has an even number of non-real roots.

The last thing we might want is a sense of what the complex numbers "look like." The real numbers looked like a line; we could put them all in an order that is compatible with their other properties. We can't do this with the complex numbers; however, since a complex number $a + bi$ is fundamentally just a pair of real numbers, we can represent it as a point on the plane.

Given a complex number $z = a + ib$, the x coordinate is the real part x and the y coordinate is the imaginary part y. (Probably why we use this notation standardly). In figure 4.1 we see the point $a + ib$ graphed on the complex plane, with real coordinate a and complex coordinate b.

We already have a sense of how far apart two points on the plane are; in particular, the point (a, b) has distance $\sqrt{a^2 + b^2}$ from the origin (think of this as a right triangle, or see 4.1 again). After staring at this for a bit, you may notice that $a^2 + b^2 = (a + bi)(a - bi) = (a + bi)\overline{(a - bi)}$. Thus we define:

Definition 4.45. The *complex modulus* or *complex absolute value* is the function given by $|z| = \sqrt{z\overline{z}}$. In figure 4.1 this is the distance labeled r.

The *argument* of z is the angle between the postive x axis and the line from 0 to z. In figure 4.1 this is the angle labeled ϕ. We'll talk a lot more about the argument when we cover Taylor series.

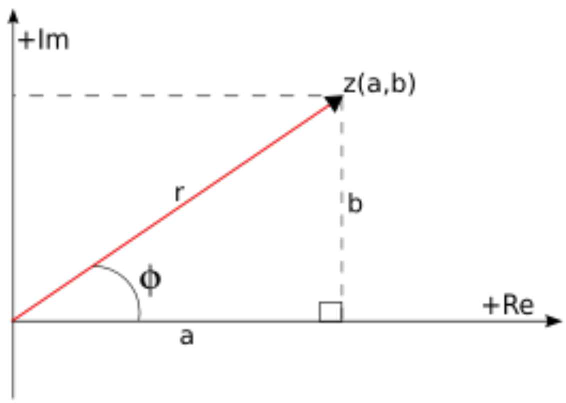


Figure 4.1: the point $a + bi$ graphed on the complex plane

Remark 4.46. The easy way to do division in $\mathbb{C}$ is to note that $1/z = \overline{z}/|z|^2$.

Here's a quick example of some complex computations; make sure you understand them.

Example 4.47.

1. $i^5 + i^{16} = i + 1$
2. $|3 + 4i| = \sqrt{9 + 16} = 5$
3. $1/(1 + i)^2 = 1/(1 + 2i - 1) = 1/2i = -i/2$
4. $\dfrac{1}{i+1} = \dfrac{1-i}{(1+i)(1-i)} = \dfrac{1-i}{1+i-i+1} = \dfrac{1}{2} - \dfrac{i}{2}$.

With that information about complex numbers in mind, we can do another example of matrix eigenspaces, and see what this tells us about them.

Example 4.48. Let $B = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}$. We can work out the characteristic polynomial is

$$\chi_B(\lambda) = \det \begin{bmatrix} .5 - \lambda & -.6 \\ .75 & 1.1 - \lambda \end{bmatrix} = (.5 - \lambda)(1.1 - \lambda) - (-.6)(.75) = \lambda^2 - 1.6\lambda + 1.$$

By the Fundamental Theorem of Algebra, this quadratic must have two complex roots; by the Quadratic Formula, we see that they are $.8 \pm .6i$. Let's set $\lambda = .8 + .6i$ and find the λ eigenspace. We have

$$\begin{bmatrix} .5 - (.8 + .6i) & -.6 \\ .75 & 1.1 - (.8 + .6i) \end{bmatrix} = \begin{bmatrix} -.3 - .6i & -.6 \\ .75 & .3 - .6i \end{bmatrix}$$

This would be a pain to row-reduce, but we don't actually have to; we already know this has non-trivial kernel, so the second row must be some multiple of the first, even if we're too lazy to figure out which multiple. So we can just say that we want to solve $.75x_1+(.3-.6i)x_2 = 0$, and we get $x_1 = (-.4 + .8i)x_2$, so the eigenspace is $E_\lambda = \{((-.4 + .8i)x_2, x_2)\}$. To find an eigenvector, we can take $x_2 = 10$ so $x_1 = -4 + 8i$, and we have that one eigenvector is $(-4 + 8i, 10)$.

We might also want to find the other eigenspace, corresponding to $E_{\overline{\lambda}}$. But we don't have to do any more work here! Our matrix B was defined over the real numbers, so everything is preserved by complex conjugation. We know that $E_\lambda = \left\{\begin{bmatrix}(-4+8i)\alpha \\ 10\alpha\end{bmatrix}\right\}$, so we know that

$$E_{\overline{\lambda}} = \overline{\left\{\begin{bmatrix}(-4+8i)\alpha \\ 10\alpha\end{bmatrix}\right\}} = \left\{\begin{bmatrix}(-4-8i)\alpha \\ 10\alpha\end{bmatrix}\right\}$$

What does this tell us? We see that while neither eigenvalue is 1 or -1, both eigenvalues have *magnitude* 1: $|.8 + .6i| = .64 + .36 = 1$. This tells us that the linear transformation will give some sort of rotation.

In fact, if we begin with the vector $(1, 0)$ and repeatedly apply A to it, the points will trace out an ellipse.

In fact, a matrix with complex eigenvalues always represents a rotation. We will revisit this idea after we've discussed changes of variables.

We end with an observation:

Proposition 4.49. *Every matrix has at least one eigenvalue over the complex numbers.*

Proof. The characteristic polynomial will have at least one root over the complex numbers by the Fundamental Theorem of Alebra. □

4.4.2 Generalized Eigenvectors

There's another way a matrix can fail to have eigenvectors.

Example 4.50. Let $A = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}$. Then the characteristic polynomial is

$$\chi_A(\lambda) = \det\begin{bmatrix}1-\lambda & 1 \\ 0 & 1-\lambda\end{bmatrix} = (1-\lambda)^2.$$

Thus the only root is 1. It has multiplicity 2, so you might hope it has two distinct eigenvectors. To work out the eigenspace E_1 we calculate

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

and so the eigenspace is $E_1 = \{(x, 0)\}$. An eigenvector is $(1, 0)$. There is not actually a second eigenvector.

Thus we see that the only (nontrivial) eigenspace is one-dimensional, but the domain is two-dimensional. There's a second, "missing" dimension. But we can recapture it after a sense.

We saw that $\ker(A - I)$ is one-dimensional. But we can work out $\ker(A - I)^2$:

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

So $\ker(A - I)^2 = \{(x, y)\} = \mathbb{R}^2$.

In particular $(0, 1)$ isn't in the kernel of $A - I$ but it is in the kernel of $(A - I)^2$. We say that $(0, 1)$ is a *generalized eigenvector* of A.

Definition 4.51. Let A be a $n \times n$ matrix, and λ be an eigenvalue of A. If $\mathbf{v} \in \ker(A - \lambda I)^n$, we say that $\mathbf{v}$ is a *generalized eigenvector* of A.

Example 4.52. Let $B = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & 1 \end{bmatrix}$. We can work out that $\chi_B(\lambda) = (1 - \lambda)^3$ and thus the only eigenvalue is 1, with multiplicity 3.

To find E_1, we compute

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

so $E_1 = \left\{ \begin{bmatrix} -2z \\ 0 \\ z \end{bmatrix} \right\}$ is one-dimensional, spanned by $\left\{ \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$.

We can look for generalized eigenvectors. We have

$$(B - I)^2 = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The kernel here is $\left\{ \begin{bmatrix} -2z \\ y \\ z \end{bmatrix} \right\} = \operatorname{span} \left\{ \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$, which is 2-dimensional. $(-2, 0, 1)$ was an eigenvector, but $(0, 1, 0)$ is not an eigenvector; it is a new, generalized eigenvector of rank 2.

Finally, we can compute

$$(B-I)^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Clearly the kernel is $\mathbb{R}^3$. If we take $(-2,0,1)$ and $(0,1,0)$ as our first two vectors, we can take say $(0,0,1)$ as our third, generalized eigenvector of rank 3.

Proposition 4.53. *If A is a $n \times n$ matrix, then (over the complex numbers) we can find n linearly independent generalized eigenvectors. Thus we can find a basis of $\mathbb{C}^n$ that consists of generalized eigenvectors of A.*

We are of course happiest when we can find a basis of regular, real-valued eigenvectors. When we can do that, we can change our coordinates to have a basis of eigenvectors, and that means computing with our matrix becomes easy. We will see how to do this in the next section.

5 Similarity and Change of Basis

We've talked in the past about the power of changing coordinate systems. In particular, we've recently noticed that eigenvectors often give us a coordinate system that's especially well-adapted to working with a given linear transformation. So it's useful to be able to translate between coordinate systems.

As usual, we want to start by talking about matrices, and then see how we can use them to understand linear transformations on arbitrary vector spaces.

5.1 Change of Basis

Recall from section 3.6 that every vector space is isomorphic to itself; a linear transformation $L : V \to V$ is an isomorphism if and only if it sends one basis to another basis.

Definition 5.1. We call such an isomorphism a *change of basis map.* The matrix of such an isomorphism is called a *transition matrix.*

Example 5.2. We know that $E = \{(1,0,0),(0,1,0),(0,0,1)\}$ and $F = \{(1,0,0),(1,1,0),(1,1,1)\}$ are both bases for $\mathbb{R}^3$. Define $L : \mathbb{R}^3 \to \mathbb{R}^3$ by setting

$$L(1,0,0) = (1,0,0)$$
$$L(0,1,0) = (1,1,0)$$
$$L(0,0,1) = (1,1,1)$$
$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then L is an isomorphism sending the basis E to the basis F.

We can find the inverse in one of two ways. One is to use row reduction as usual, but that takes effort. Instead, we can note that the inverse is just the map that sends F to E:

$$L^{-1}(1,0,0) = (1,0,0)$$
$$L^{-1}(1,1,0) = (0,1,0) = (1,1,0) - (1,0,0)$$
$$L^{-1}(1,1,1) = (0,0,1) = (1,1,1) - (1,1,0)$$
$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

By multiplication or substitution we can check that this is definitely an inverse.

Now let's ask a separate question. Suppose we have a vector $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ that is currently expressed in terms of the standard basis, and we would like to find its coordinates in F. This means that we want to write it as a sum of things in F, and so we want to solve the equation

$$x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

If we write this equation in matrix form, we get

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Notice that the matrix in this equation is just the matrix A we found that sends elements of E to elements of F.

We could solve this system of equations by row-reducing. But we can also just compute the inverse matrix: if $A\mathbf{x} = \mathbf{u}$ then $\mathbf{x} = A^{-1}\mathbf{u}$. And as we saw, the matrix A^{-1} is just the matrix that sends elements of F to elements of E. In this context, we call A the transition matrix from F to E, and A^{-1} the transition matrix from E to F.

Example 5.3 (continued)**.** Suppose we'd like to take the vector $\mathbf{u} = 2(1,0,0) + 3(1,1,0) + 5(1,1,1)$ and find its coordinates in the standard basis. We have

$$\begin{aligned} \mathbf{u} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} &= 2(\mathbf{e}_1) + 3(\mathbf{e}_1 + \mathbf{e}_2) + 5(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3) \\ &= (2+3+5)\mathbf{e}_1 + (3+5)\mathbf{e}_2 + 5\mathbf{e}_3 \qquad \text{or} \\ [\mathbf{u}]_E = A[\mathbf{u}]_F &= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ 5 \end{bmatrix}. \end{aligned}$$

In many ways it's more useful to do things the other way. Suppose we have the vector $(5, 2, 7)$ and want to express it as a linear combination of elements of F. Then we need the transition matrix from F to E, which is A^{-1}. So we have

$$A^{-1}\begin{bmatrix}5\\2\\7\end{bmatrix} = \begin{bmatrix}1 & -1 & 0\\0 & 1 & -1\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}5\\2\\7\end{bmatrix} = \begin{bmatrix}3\\-5\\7\end{bmatrix}.$$

Thus

$$\begin{bmatrix}5\\2\\7\end{bmatrix} = 3\begin{bmatrix}1\\0\\0\end{bmatrix} - 5\begin{bmatrix}1\\1\\0\end{bmatrix} + 7\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Example 5.4. Let's represent the polynomial $a + bx + cx^2 \in \mathcal{P}_3(x)$ as a linear combination of $F = \{1, 2x, 4x^2 - 2\}$.

We take $E = \{1, x, x^2\}$ to be the standard basis, and if A is the transition matrix from F to E we have

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

$$\left[\begin{array}{ccc|ccc}1 & 0 & -2 & 1 & 0 & 0\\0 & 2 & 0 & 0 & 1 & 0\\0 & 0 & 4 & 0 & 0 & 1\end{array}\right] \to \left[\begin{array}{ccc|ccc}1 & 0 & -2 & 1 & 0 & 0\\0 & 1 & 0 & 0 & 1/2 & 0\\0 & 0 & 1 & 0 & 0 & 1/4\end{array}\right] \to \left[\begin{array}{ccc|ccc}1 & 0 & 0 & 1 & 0 & 1/2\\0 & 1 & 0 & 0 & 1/2 & 0\\0 & 0 & 1 & 0 & 0 & 1/4\end{array}\right]$$

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

Thus we have

$$[a + bx + cx^2]_F = A^{-1}\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}1 & 0 & 1/2\\0 & 1/2 & 0\\0 & 0 & 1/4\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}a + c/2\\b/2\\c/4\end{bmatrix}.$$

Thus $[a + bx + cx^2]_F = (a + c/2, b/2, c/4)$ and

$$a + bx + cx^2 = (a + c/2)(1) + (b/2)(2x) + (c/4)(4x^2 - 2).$$

I want to mention one last idea here, which is the ability to paste transition matrices together. If A is the transition matrix from F to E, and B is the transition matrix from G to F, then AB is the transition matrix from G to E.

This *sounds* backwards, but really we just write things backwards. Remember that if L is a function from U to V, then we have $\mathbf{v} = L(\mathbf{u})$. We put $\mathbf{u}$ on the *right* of L, and then once we apply L we get $\mathbf{v}$. And when we compose two functions, like $T(L(\mathbf{v}))$, then we do the one on the right first, and the one on the left second.

So we have B going from G to F, and then A going from F to E; when we patch them to gether, we have AB going from G to F and then to E.

(There are mathematicians who "fix" this by writing $(x)f$ when then want to plug x into the function f. That works perfectly well, but looking at it makes me feel deeply uncomfortable.)

This pasting-together is primarily useful once we introduce inverses.

Example 5.5. Let $E = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}, F = \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ be two bases for $\mathbb{R}^3$. Let $\mathbf{u} = 3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$. Let's find the coordinates of $\mathbf{u}$ with respect to F.

We could try to compute the transition matrix directly, but that requires us to do a bunch of equation solving. Instead, we notice that the transition matrix from E to the standard basis is

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

and the transition matrix from F to the standard basis is

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

We want the transition matrix from E to F so that we can convert coordinates from E to F. Thus the matrix we actually want is $B^{-1}A$: A takes us from E to the standard basis, and then B^{-1} takes us from the standard basis to F. We compute

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 2 & 1 & 1 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1/2 & 1/2 & 1/2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/2 & -1/2 & -1/2 \\ 0 & 1 & 0 & 1/2 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1/2 & 1/2 & 1/2 \end{array}\right]$$

and thus

$$B^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$B^{-1}A = \frac{1}{2}\begin{bmatrix} -2 & -1 & 1 \\ 2 & 1 & 3 \\ 4 & 3 & 1 \end{bmatrix}$$
$$[\mathbf{u}]_F = B^{-1}A[\mathbf{u}]_E = \frac{1}{2}\begin{bmatrix} -2 & -1 & 1 \\ 2 & 1 & 3 \\ 4 & 3 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -3 \\ 7 \\ 7 \end{bmatrix} = \begin{bmatrix} -3/2 \\ 7/2 \\ 7/2 \end{bmatrix}.$$

We check that, indeed,

$$\frac{-3}{2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{7}{2}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + \frac{7}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

5.2 Similarity

We now want to return to talking about general linear transformations, but bringing with us our new perspective on bases and changes of bases.

Let $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by a rotation ninety degrees counterclockwise. We saw earlier that with respect to the standard basis, this transformation has matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. But we can also compute the matrix with respect to, say, $F =$

$\{(1,0),(1,1)\}$. Then we have

$$\begin{aligned} R_{\pi/2}(1,0) &= (0,1) = (1,1)-(1,0) \to (-1,1) \\ R_{\pi/2}(1,1) &= (-1,1) = (1,1)-2(1,0) \to (-2,1) \\ B &= \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}. \end{aligned}$$

These two matrices represent the same transformation, with respect to different bases. But they are clearly not the same matrix! What's going on here?

The answer is that we changed the coordinate system, and so our matrix changed. After we account for that, we should get the same matrix. To account for this, we need the change of basis matrix between F and the standard basis E. We have

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

the transition matrix from F to the standard basis, and thus

$$U^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

is the transition matrix from the standard basis to F.

If we want to perform the operation $R_{\pi/2}$ on the vectors of F, we can use the matrix B that we found. Alternatively, we can transform our vectors into E-coordinates, use the matrix A, and then transform back into F-coordinates. This operation would be given by $U^{-1}AU$. We calculate that

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

This is the same as the matrix B, as it should be.

Definition 5.6. If A and B are $n \times n$ matrices, we say they are *similar* if there is some invertible matrix U such that $B = U^{-1}AU$. We write $A \sim B$.

Proposition 5.7. *Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}, F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ be two bases for V, and let $L : V \to V$ be a linear function. Let U be the transition matrix from F to E.*

If A is the matrix representing L with respect to E, and B is the matrix representing L with respect to F, then $B = U^{-1}AU$.

Example 5.8. Let $D : \mathcal{P}_2(x) \to \mathcal{P}_2(x)$ be the differentiation operator. Let's find the matrix of D with respect to $E = \{1, x, x^2\}$ and with respect to $F = \{1, 2x, 4x^2 - 2\}$.

We've already seen that the matrix of D with spect to E is $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

We can work out the matrix with respect to F directly:

$$\begin{aligned} D(1) &= 0 \to (0,0,0) \\ D(2x) &= 2 \to (2,0,0) \\ D(4x^2 - 2) &= 8x \to (0,4,0) \\ B &= \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Alternatively, we could recall that the change of basis matrices between E and F:

$$\begin{aligned} E \to F &: \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} = U^{-1} \\ F \to E &: \begin{bmatrix} 1 & 0 & -2 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} = U. \end{aligned}$$

So we can compute the matrix B for D by saying

$$\begin{aligned} B = U^{-1}AU &= \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Example 5.9. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $L(x, y, z) = (x + 3y + z, 2x - y + 3z, y - z)$. Find the matrix of L with respect to $\{(4, 1, 2), (3, 0, 1), (1, -1, 0)\}$, and show it is similar to the matrix with respect to the standard basis.

We have

$$\begin{aligned} L(1,0,0) &= (1,2,0) \\ L(0,1,0) &= (3,-1,1) \\ L(0,0,1) &= (1,3,-1) \\ A &= \begin{bmatrix} 1 & 3 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & -1 \end{bmatrix}. \end{aligned}$$

We can compute the change of basis matrices. If U is the matrix from F to E, then we have

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

and

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

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

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

so we have

$$U^{-1} = \begin{bmatrix} -1 & -1 & 3 \\ 2 & 2 & -5 \\ -1 & -2 & 3 \end{bmatrix}.$$

Thus to find the matrix with respect to F, we can compute

$$\begin{aligned} B = U^{-1}AU &= \begin{bmatrix} -1 & -1 & 3 \\ 2 & 2 & -5 \\ -1 & -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 4 & 3 & 1 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -3 & 1 & -7 \\ 6 & -1 & 13 \\ -5 & 2 & -10 \end{bmatrix} \begin{bmatrix} 4 & 3 & 1 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -25 & -16 & -4 \\ 49 & 31 & 7 \\ -38 & -25 & -7 \end{bmatrix}. \end{aligned}$$

Example 5.10. We can do one more clever example, to connect back to the work with complex matrices we did in section 4.4.1. Recall we had the matrix $B = \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix}$ with eigenvalues $\lambda = .8 + .6i$ and $\overline{\lambda} = .8 - .6i$, with eigenvectors $(-4 \pm 8i, 10)$. We could try to change basis into this eigenvector basis, but that would be a little messy.

Instead, we can change to a related basis. We can say the *real part* of this vector is $(-4, 10)$, and the *imaginary part* is $(8, 0)$. So we set $U = \begin{bmatrix} -4 & 8 \\ 10 & 0 \end{bmatrix}$ to be the change of basis matrix to the basis formed of the real and complex parts of this matrix. Then we can compute

$$\begin{aligned} U^{-1} &= \frac{1}{80} \begin{bmatrix} 0 & -8 \\ -10 & -4 \end{bmatrix} \\ C = U^{-1}BU &= \frac{1}{80} \begin{bmatrix} 0 & -8 \\ -10 & -4 \end{bmatrix} \begin{bmatrix} .5 & -.6 \\ .75 & 1.1 \end{bmatrix} \begin{bmatrix} -4 & 8 \\ 10 & 0 \end{bmatrix} \\ &= \begin{bmatrix} .8 & .6 \\ -.6 & .8 \end{bmatrix}. \end{aligned}$$

This real matrix is a rotation matrix, rotating counterclockwise by an angle of approximately .64. You might remember from last week that we got an elliptical path; we got an ellipse because the basis we're changing into has vectors of different lengths. So we're changing to that basis, rotating by C, and then changing back to our regular basis.

There's not a really efficient way to determine whether two matrices are similar in general, although we have a few tools that can tell us two matrices are *not* similar.

Proposition 5.11. *Let $A, B \in M_{n \times n}$ with $A \sim B$. Then:*

- *The rank of A is equal to the rank of B.*

- *The nullity of A is equal to the nullity of B.*
- *A is invertible if and only if B is invertible.*

5.3 Determinant and Trace

Because matrices that are similar represent the same underlying transformation, anything that's a property of the transformation, and not of the particular matrix, should remain unchanged. Thus two matrices representing the same linear transformation should have the same eigenvalues. And indeed, this is the case.

Proposition 5.12. *Suppose A and B are similar $n \times n$ matrices, so there exists U such that $B = U^{-1}AU$. Then:*

- $\det(A) = \det(B)$
- $\chi_A(\lambda) = \chi_B(\lambda)$
- *A and B have the same set of eigenvalues.*

Proof. We can prove these two ways. From a formal perspective, we know that A and B must represent the same linear transformation, and since all of these things are properties of the linear transformation, they must be the same for similar matrices.

From a more concrete algebraic perspective, we have:

- $\det(B) = \det(U^{-1}AU) = \det(U^{-1})\det(A)\det(U) = \frac{1}{\det(U)}\det(A)\det(U) = \det(A)$.
- For any λ, we have

  $$U^{-1}(A - \lambda I)U = U^{-1}AU - U^{-1}\lambda IU = B - \lambda I,$$

  so we have $(A - \lambda I) \sim (B - \lambda I)$. By the previous result, $\det(A - \lambda I) = \det(B - \lambda I)$.
- The eigenvalues are the roots of the characteristic polynomial. Since $\chi_A(\lambda) = \chi_B(\lambda)$, the roots are the same and so the eigenvalues are the same.

□

Example 5.13. Let $A = \begin{bmatrix} 2 & 3 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ and let $U = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, so that

$$B = U^{-1}AU = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 3 & 3 \\ 1 & -2 & 1 \end{bmatrix}.$$

Clearly $A \sim B$. We can see immediately that $\chi_A(\lambda) = (2-\lambda)(1-\lambda)(1-\lambda) = 2 - 5\lambda + 4\lambda^2 - \lambda^3$ and $\det(A) = 2$. With a little more work, we have

$$\begin{aligned}
\chi_B(\lambda) = \det \begin{bmatrix} -\lambda & 2 & 0 \\ 2 & 3-\lambda & 3 \\ 1 & -2 & 1-\lambda \end{bmatrix} \\
&= -\lambda\left((3-\lambda)(1-\lambda) - (-2\cdot 3)\right) - 2\left(2(1-\lambda) - 1\cdot 3\right) \\
&= -3\lambda + \lambda^2 + 3\lambda^2 - \lambda^3 - 6\lambda - 4 + 4\lambda + 6 \\
&= 2 - 5\lambda + 4\lambda^2 - \lambda^3 = \chi_A(\lambda).
\end{aligned}$$

Remark 5.14. The converse of this theorem is not true. Similar matrices always have the same characteristic polynomial; but sometimes matrices with the same characteristic polynomial are not similar.

If $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ then $\chi_A(\lambda) = (1-\lambda)^2 = \chi_I(\lambda)$. But clearly A is not similar to the identity, since $U^{-1}IU = I$, and so the only matrix similar to I is itself.

Since the characteristic polynomials of similar matrices are the same, they clearly have all the same coefficients. In fact, we can see that the determinant is just the constant term of the characteristic polynomial, $\chi_A(0)$. There's one other coefficient that's often important. It's not the highest degree coefficient, which is always ± 1; but the second-highest coefficient is often interesting and useful.

Definition 5.15. If $L : V \to V$ is a linear transformation on a n-dimensional vector space, we define the *trace* of L to be $\mathrm{Tr}(L) = (-1)^{n-1}a_{n-1}$ where $\chi_L(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$.

If A is a $n \times n$ matrix, we define the *trace* of A to be the trace of the linear transformation represented by A. Thus $\mathrm{Tr}(A) = (-1)^{n-1}a_{n-1}$ where $\chi_A(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$.

Proposition 5.16.
- *If* $A \sim B$ *then* $\mathrm{Tr}(A) = \mathrm{Tr}(B)$.

- $\mathrm{Tr}(A)$ *is the sum of the eigenvalues of* A *(weighted by multiplicity).*
- $\mathrm{Tr}(A) = \sum_{i=1}^n a_{ii}$ *is the sum of the entries on the main diagonal of* A*.*

Proof.

- This follows from the fact that $\chi_A(\lambda) = \chi_B(\lambda)$.
- If $\chi_A(\lambda) = (\lambda - \lambda_1)^{d_1} \dots (\lambda - \lambda_k)^{d_k}$, then when we multiply this out, the $n-1$ coefficient will be $\sum_{i=1}^k d_k \lambda_k$.
- Proof by induction.

□

Remark 5.17. This tells us that the trace is very easy to compute; unlike the determinant, it doesn't depend on any non-diagonal entries, and just requires some fast, simple addition.

This also tells us that the trace is a *similarity invariant*, meaning that similar matrices have the same trace. Thus we can quickly test whether two matrices might be similar by computing the traces of both.

But notice that, like with the determinant, we can have two matrices which are not similar but have the same trace.

If the matrix A is given as a function of T, there is a specific sense in which the trace is related to the derivative of the determinant. But we're not going to be precise about that in this course.

Proposition 5.18. *The trace is a linear multiplicative map on matrices. That is:*

- $\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
- $\mathrm{Tr}(rA) = r\,\mathrm{Tr}(A)$
- $\mathrm{Tr}(A^T) = \mathrm{Tr}(A)$

Proof. These follow from the characterization of the trace as the sum of the diagonal elements. □

Example 5.19. Let $A = \begin{bmatrix} 3 & 2 & 5 \\ 1 & 4 & 1 \\ 2 & -3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 1 & 1 \\ 4 & -3 & -3 \\ 2 & 1 & 0 \end{bmatrix}$. Then $\mathrm{Tr}(A) = 3+4+2 = 9$ and $\mathrm{Tr}(B) = 5-3+0 = 2$, so we know that $A \not\sim B$.

If $C = \begin{bmatrix} 4 & -2 & 3 \\ 5 & 1 & 7 \\ 1 & 1 & 4 \end{bmatrix}$ then $\mathrm{Tr}(C) = 4+1+4 = 9$, so it's possible that $C \sim A$. But we'd need to do more work to confirm this. On just this evidence, it probably isn't.

In fact we can compute that $\chi_A(\lambda) = -x^3+9x^2-24x+26 \neq -x^3+9x^2-17x-22 = \chi_C(\lambda)$, so the matrices actually aren't similar.

5.4 Diagonalization

We now reach the payoff to all this discussion of changes of coordinates. If we find the eigenvectors of a linear operator and the give us a (eigen)basis for our space, we can always find a matrix representation of our linear operator with a particularly *nice* matrix by working with respect to this eigenbasis. In particular:

Definition 5.20. If D is a $n \times n$ matrix such that $a_{ij} = 0$ whenever $i \neq j$, we say that D is *diagonal*.

Proposition 5.21. *Let* $D = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix}$ *be a diagonal* $n \times n$ *matrix. Then:*

- *Each standard basis vector* $\mathbf{e}_i$ *is an eigenvector of* D *with eigenvalue* d_{ii}.
- $\det(D) = \prod_{i=1}^n d_{ii}$ *is the product of the diagonal entries.*
- $\mathbb{R}^n$ *is spanned by the eigenvectors of* D.

Proof.

- We have

$$D\mathbf{e}_i = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ d_{ii} \\ \vdots \\ 0 \end{bmatrix} = d_{ii}\mathbf{e}_i.$$

- The determinant is the product of the eigenvalues, which are the diagonal entries.
- The standard basis vectors are eigenvalues, and span $\mathbb{R}^n$.

□

Definition 5.22. We say a linear transformation is *diagonalizable* if its matrix in some basis is diagonal.

We say a matrix is *diagonalizable* if its linear transformation is diagonalizable. Thus A is diagonalizable if A is similar to some diagonal matrix.

Proposition 5.23. *Let A be a $n \times n$ matrix. Then:*

1. *A is diagonalizable if and only if the eigenvectors of A span $\mathbb{R}^n$.*
2. *A is diagonalizable if and only if it has n linearly independent eigenvectors.*
3. *If A has n distinct eigenvalues, then A is diagonalizable.*

Proof. 1. Suppose A is diagonalizable, i.e. there is an invertible matrix U and a diagonal matrix D such that $A = U^{-1}DU$. Let F be the image of the standard basis under U^{-1}; then

$$A\mathbf{f}_i = U^{-1}DU\mathbf{f}_i = U^{-1}D\mathbf{e}_i = U^{-1}d_{ii}\mathbf{e}_i = d_{ii}U^{-1}\mathbf{e}_i = d_{ii}\mathbf{f}_i.$$

Thus $\mathbf{f}_i$ is an eigenvector for each i, so we have a basis of eigenvectors.

Conversely Suppose the eigenvectors of A span $\mathbb{R}^n$. Then in particular there is a basis $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ of eigenvectors. Let U be the matrix that sends the standard basis to F. Then for each i we have

$$U^{-1}AU\mathbf{e}_i = U^{-1}A\mathbf{f}_i = U^{-1}\lambda_i\mathbf{f}_i = \lambda_i U^{-1}\mathbf{f}_i = \lambda_i\mathbf{e}_i$$

and thus $U^{-1}AU$ is a diagonal matrix with $d_{ii} = \lambda_i$. Thus A is diagonalizable.

2. A set of n linearly independent vectors is a basis for $\mathbb{R}^n$. Thus A has n linearly indpendent eigenvectors if and only if the eigenvectors span $\mathbb{R}^n$.

3. Let $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ be a set of eigenvectors corresponding to each eigenvalue. Then this set is linearly independent by proposition 4.8, and thus A has n linearly independent eigenvectors.

□

Remark 5.24. Notice that the converse of (3) is not true, by which we mean that it would be false if we said "if and only if". For instance, the identity has only one eigenvalue, but is clearly diagonalizable (and actually diagonal already).

Corollary 5.25. *If A is a $n \times n$ matrix and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ is a basis of eigenvectors, and U is the matrix sending the standard basis to F, then $D = U^{-1}AU$ is a diagonal matrix.*

We say that the matrix U diagonalizes *A.*

Remark 5.26. Diagonalization is not unique; the matrix U depends on the choice of basis. However, since the diagonal entries are the eigenvalues, they will be the same (up to reordering) for any diagonalization.

Example 5.27. Let $A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$. We know (from example 4.34) that the eigenvalues are 4 and -3, so the matrix is diagonalizable; the corresponding eigenvectors are $(2, 1)$ and $(-1, 3)$. So we set

$$\begin{aligned} U &= \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \\ U^{-1} &= \frac{1}{7} \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \\ U^{-1}AU &= \frac{1}{7} \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 8 & 3 \\ 4 & -9 \end{bmatrix} \\ &= \frac{1}{7} \begin{bmatrix} 28 & 0 \\ 0 & -21 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}. \end{aligned}$$

Example 5.28. Let $A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$. We saw in example 4.36 that the eigenvalues are $0, 1, 1$. The eigenvectors are $(1, 1, 1), (3, 1, 0), (-1, 0, 1)$, so we set

$$\begin{aligned} U &= \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ U^{-1} &= \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \\ U^{-1}AU &= \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 0 & 3 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Example 5.29. We saw in example 4.38 that the matrix $B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ had eigenspaces

$E_2 = \text{span}\{(0,0,1)\}$ and $E_4 = \text{span}\{(0,1,0)\}$. The eigenvectors do not span $\mathbb{R}^3$, so A is not diagonalizable.

In general I don't really expect triangular matrices with repeated eigenvalues to be diagonal, but treating this thought fully is beyond the scope of this course.

There are a few major uses for diagonalization. The first is to tell us the basis we "should" be working in, and to allow us to change bases to that basis. The basis in which your operator is diagonal is the basis in which your operator is "really" working; it divides your space up into the dimensions along which your operator really works.

Eigenvectors and diagonalization are often used in various sorts of data analysis. The eigenvector corresponding to the largest eigenvalue is the most significant input, so diagonalization can tell us which components of our data are most important to whatever phenomenon we're studying; this is the idea behind "principal component analysis". If we have time we'll return to this at the end of class.

They are also used in various sorts of approximate computations: if your linear operator has eigenvalues of 5, 3, 1, .1, .1, -.1, .0005, you can get a pretty good approximation of your operator by ignoring the eigenvectors corresponding to the small eigenvalues, and only worrying about the large ones. This is important in a lot of numeric computation.

Finally, we can use diagonalization to simplify many matrix computations. We need to make two observations: one about diagonal matrices, the other about similar matrices.

Proposition 5.30. *Suppose C and D are two diagonal matrices with diagonal entries given by c_{ii}, d_{ii} respectively. Then their product is a diagonal matrix given by*

$$\begin{bmatrix} c_{11} & 0 & 0 & \dots & 0 \\ 0 & c_{22} & 0 & \dots & 0 \\ 0 & 0 & c_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_{nn} \end{bmatrix} \begin{bmatrix} d_{11} & 0 & 0 & \dots & 0 \\ 0 & d_{22} & 0 & \dots & 0 \\ 0 & 0 & d_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_{nn} \end{bmatrix} = \begin{bmatrix} c_{11}d_{11} & 0 & 0 & \dots & 0 \\ 0 & c_{22}d_{22} & 0 & \dots & 0 \\ 0 & 0 & c_{33}d_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_{nn}d_{nn} \end{bmatrix}.$$

Proposition 5.31. *If $A = U^{-1}BU$, then $A^n = U^{-1}B^nU$.*

Proof.

$$\begin{aligned} A^n = (U^{-1}BU)^n &= U^{-1}BUU^{-1}BU \dots U^{-1}BUU^{-1}BU \\ & U^{-1}BI_nB \dots IBIBU = U^{-1}BB \dots BBU = U^{-1}B^nU. \end{aligned}$$

□

Example 5.32. Let Let $A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$. Find A^5.

If $U^{-1}AU = D$, then $UU^{-1}AUU^{-1} = UDU^{-1}$ and thus $A = UDU^{-1}$. So

$$D = \begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} = U^{-1}AU$$

$$A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}\frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} = UDU^{-1}$$

$$\begin{aligned} A^5 = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}^5 &= \left(\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}\frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}\right)^5 \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}^5\frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1024 & 0 \\ 0 & -243 \end{bmatrix}\frac{1}{7}\begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{7}\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 3072 & 1024 \\ 243 & -486 \end{bmatrix} \\ &= \frac{1}{7}\begin{bmatrix} 5901 & 2534 \\ 3801 & -434 \end{bmatrix} = \begin{bmatrix} 843 & 362 \\ 543 & -62 \end{bmatrix}. \end{aligned}$$

Example 5.33. Let $A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$. Find a formula for A^n.

We have

$$\begin{aligned}
\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\
\begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} &= \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \\
\begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}^n &= \left(\begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \right)^n \\
&= \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^n \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}.
\end{aligned}$$

Corollary 5.34. *If A is a diagonalizable matrix whose eigenvalues are only zero or one, then $A^n = A$ for any n.*

We can actually extend this even further, and talk about matrix exponentiation.

At first it's not even clear what it would mean to compute e^A where A is a matrix. But we know from Calculus 2 that we can define $e^x = \exp(x)$ by its Taylor series:

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

We are perfectly capable of plugging a matrix into this formula, so we can define

$$e^A = 1 + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \dots$$

But if D is diagonal, we can compute e^D just by exponentiating each entry; and it turns out that this whole operation plays nicely with transition matrices and changing coordinates. So if $A = U^{-1}DU$ then $e^A = U^{-1}e^D U$ is easy to compute.

This sort of matrix exponentiation is really important in a lot of differential equations modeling problems. It solves multi-dimensional differential equations similar to the single-variable $y' = ky$ that is solved by the regular exponential function.

5.5 Application: Markov Chains

We can combine all these ideas about eigenvalues, similarity, and diagonalization to provide tools to analyze random processes.

Example 5.35. Suppose at a certain time. 70% of the population lives in the suburbs, and 30% lives in the city. But each year, 6% of the people living in the suburbs move to the city, and 2% of the people living in the city move to the suburbs. What happens after a year? Five years? Ten years?

And what is the equilibrium distribution?

Because these rates of transition are *constant*, we can model this with a matrix. If s is the number of people in the suburbs, and c is the number in the city, then next year we'll have $.94s + .02c$ people in the suburbs, and $.06s + .98c$ people in the city. With our numbers, that gives 66.4% in the suburbs, and 33.6% in the city.

We could repeat this calculation to find out what happens in two years, and then three, et cetera. But it's simpler, first, if we turn this into a matrix. If we think of (s, c) as a vector in $\mathbb{R}^2$, then the population changes according to the following matrix:

$$A = \begin{bmatrix} .94 & .02 \\ .06 & .98 \end{bmatrix}.$$

Thus after one year the population distribution will be $A \begin{bmatrix} .7 \\ .3 \end{bmatrix}$ and after five years it will be $A^5 \begin{bmatrix} .7 \\ .3 \end{bmatrix}$. This matrix A is called a *transition matrix*, although it has nothing to do with the change of basis matrices we discussed in section 5.1. Instead, it measures what fraction of a population transitions from one state to another—in this case, from the suburbs to the city or vice versa. Notice that every column sums up to 1. This isn't an accident; exactly 100% of a population has to go somewhere.

So now we can answer our earlier questions, if we can compute A^5 and A^{10}. We saw in 5.4 that this is easy if we *diagonalize* the matrix A. We can compute the eigenvectors, and see that A has an eigenvector $(1, 3)$ with eigenvalue 1, and an eigenvector $(-1, 1)$ with eigenvalue .92. Then we compute

$$\begin{aligned} U &= \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \\ U^{-1} &= \frac{1}{4} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \\ A = UDU^{-1} &= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .92 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix}. \end{aligned}$$

This allows us to compute our exponentials:

$$\begin{aligned} A^5 = UD^5U^{-1} &= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .92^5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \\ &\approx \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .66 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \approx \begin{bmatrix} .744 & .085 \\ .256 & .915 \end{bmatrix} \\ A^5 \begin{bmatrix} .7 \\ .3 \end{bmatrix} \approx \begin{bmatrix} .55 \\ .45 \end{bmatrix}. \end{aligned}$$

Thus after five years we'll have about 55% of people in the suburbs, and 45% in the city.

For ten years, we can do the same computation.

$$\begin{aligned} A^{10} = UD^5U^{-1} &= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .92^{10} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \\ &\approx \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .60 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} \approx \begin{bmatrix} .576 & .141 \\ .424 & .859 \end{bmatrix} \\ A^{10} \begin{bmatrix} .7 \\ .3 \end{bmatrix} \approx \begin{bmatrix} .45 \\ .55 \end{bmatrix}. \end{aligned}$$

So after ten years, we'll have 45% of people in the suburbs, and 55% in the city.

But how do we answer our final question, about the equilibrium? Here we want something like $\lim_{n\to\infty} A^n$. Without diagonalization this would be really hard to compute. But it's easy to see that $\lim_{n\to\infty} D^n = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and thus we get

$$\begin{aligned} \lim_{n\to\infty} A^n = \lim_{n\to\infty} UD^nU^{-1} &= \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 \\ 3/4 & 3/4 \end{bmatrix} \\ \lim_{n\to\infty} A^n \begin{bmatrix} .7 \\ .3 \end{bmatrix} \approx \begin{bmatrix} 1/4 \\ 3/4 \end{bmatrix}. \end{aligned}$$

Thus the equilibrium state is where 25% of people live in the suburbs and 75% live in the cities. Which maybe we could have guessed from the start, since that makes the populations moving each way equal.

This entire process is very flexible. Any time the probability of transitioning from one state to another is constant, and only depends on which state you start in, we can model our system with a matrix like A, which we call a *Markov process*. In that case, the sequence of vectors $\mathbf{v}_1 = A\mathbf{v}, \mathbf{v}_2 = A^2\mathbf{v}, \dots$ is called a *Markov chain*.

Each column of A is a *probability vector*, which is a vector of non-negative numbers that add up to one. Each row and each column corresponds to a particular possible state, and each entry tells us the probability of moving into the column-state if we start out in the row-state.

We can use this system to find the projected state after a finite number of steps, but even more usefully we can use it to project the equilibrium state.

Example 5.36. Suppose our matrix of transition probabilities is $B = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix}$. This matrix has eigenvalues $1/2, 1$ with eigenvectors $(1, -1)$ and $(2, 3)$. Then we can diagonalize:

$$U = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$$
$$U^{-1} = \frac{1}{5}\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}$$

$$\begin{aligned} D = U^{-1}BU &= \frac{1}{5}\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} .5 & 2 \\ -.5 & 3 \end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix} 2.5 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} .5 & 0 \\ 0 & 1 \end{bmatrix} \\ B = UDU^{-1} &= \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} .5 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Then we can easily exponentiate.

$$\begin{aligned}
B^n &= \left(\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} .5 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \right)^n \\
&= \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} .5 & 0 \\ 0 & 1 \end{bmatrix}^n \frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \\
\lim_{n \to \infty} B^n &= \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} .4 & .4 \\ .6 & .6 \end{bmatrix}.
\end{aligned}$$

Thus in the equilibrium, we will have 40% of people in state one, and 60% in state two.

You might have noticed something both of these examples have in common. Both A and B had 1 as an eigenvalue. And our steady-state vectors were in fact eigenvectors with eigenvalue 1. As you might guess, this isn't a coincidence.

Not every Markov process converges to a steady-state vector. But if it converges, the steady state will be an eigenvector with eigenvalue 1.

Further, convergence is guaranteed if all the entries of the matrix are positive (and nonzero). And convergence is also guaranteed if there is only one eigenvector of magnitude 1 (even after taking absolute values).

Example 5.37. Suppose you run a car dealership that does long-term car leases. You lease sedans, sports cars, minivans, and SUVs. At the end of each year, your clients have the option to trade in to a different style of car. Empirically, you find that you get the following transition matrix:

$$C = \begin{bmatrix} .80 & .10 & .05 & .05 \\ .10 & .80 & .05 & .05 \\ .05 & .05 & .80 & .10 \\ .05 & .05 & .10 & .80 \end{bmatrix}$$

Thus if someone has a sedan this year, they are 80% likely to take a sedan next year, 10% likely to take a sports car, and 5% each likely to take a minivan or an SUV.

We find that C has the eigenvalues $1, .8, .7, .7$, with corresponding eigenvalues

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

Thus it will have a steady-state equilibrium. In particular, in a steady state, equal numbers of customers will lease each type of car, no matter what the distribution is right now.

Example 5.38. As a final, non-numerical example, this is how the Google PageRank algorithm works.

They treat web browsing as a random process: given that you are currently on one web page, you have some (small) probability of winding up on any other web page. Of course, this probability is higher if the page you're on links to the new page prominently, so the probability of winding up on any given page is not equal.

Then they build a giant $n \times n$ matrix, where n is the number of web pages they have analyzed. Each column corresponds to a particular web page, and the entries tell you how likely you are to go to any other web page next.

Then they compute the eigenvectors and eigenvalues of this matrix. Or at least, they compute the eigenvector with eigenvalue 1. (Since the matrix is all positive, this is guaranteed to exist, and there are relatively efficient ways to find it.) This gives you an equilibrium probability: if you browse the web for an arbitrarily long period of time, how likely are you to land on this page?

And that is, roughly speaking, the page rank. The more likely you are to land on a given web page, from this Markov chain model, the more highly ranked the page is.

6 Inner Product Spaces and Geometry

In this section we're going to consider vector spaces from a more geometric perspective. In $\mathbb{R}^3$ we have the geometric ideas of "distance" and "angle", but neither of those is necessarily present in an arbitrary vector space. Here we will introduce a new structure called an "Inner Product" that allows us to generalize the angles and distances of $\mathbb{R}^3$ to any vector space with an inner product structure.

6.1 The Dot Product

Definition 6.1. Let $\mathbf{u} = (u_1, \dots, u_n), \mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. We define the *dot product* of $\mathbf{u}$ and $\mathbf{v}$ by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n = \sum_{i=1}^n u_i v_i.$$

This is sometimes also called the *scalar product* on $\mathbb{R}^n$.

Remark 6.2. If we think of $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices, we can think of $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$, the product of a $n \times 1$ matrix with a $1 \times n$ matrix.

The dot product has a number of useful properties. First of all, it allows us to define the length or magnitude of a vector.

Definition 6.3. Let $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. We define the *magnitude* of $\mathbf{v}$ to be

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Notice that this is just the usual definition of distance; in the plane this is

$$\|(x, y)\| = \sqrt{x^2 + y^2},$$

which is just the pythagorean theorem.

Sometimes it's useful to talk about the distance between two points, rather than the length of a vector. But the distance between two points is the length of the vector between them, so we can define the distance between $\mathbf{x}$ and $\mathbf{y}$ to be

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

The dot product has a few important properties:

Proposition 6.4. *Let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*. Then:*

1. *(Positive definite)* $\mathbf{u} \cdot \mathbf{u} \geq 0$, *and if* $\mathbf{u} \cdot \mathbf{u} = 0$ *then* $\mathbf{u} = \mathbf{0}$.

2. *(Symmetric)* $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

3. *(Bilinear) The function defined by* $L(\mathbf{x}) = \mathbf{x} \cdot \mathbf{v}$ *is linear, and the function defined by* $T(\mathbf{y}) = \mathbf{u} \cdot \mathbf{y}$ *is linear.*

Proof.

1. $\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + \cdots u_n^2$. Each term is non-negative since each term is a real square, so the sum is non-negative. The sum is zero if and only if each term is zero, if and only if $\mathbf{u} = (0, \ldots, 0) = \mathbf{0}$.

2. $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n = v_1 u_1 + \cdots + v_n u_n = \mathbf{v} \cdot \mathbf{u}$.

3. We'll prove linearity in the first coordinate; the proof for the second coordinate is identical.

   Fix $\mathbf{v} \in \mathbb{R}^n$ and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $r \in \mathbb{R}$. Define $L(\mathbf{x}) = \mathbf{x} \cdot \mathbf{v}$. Then

$$\begin{aligned} L(r\mathbf{x}) &= (r\mathbf{x}) \cdot \mathbf{v} = (rx_1)v_1 + \cdots + (rx_n)v_n = r(x_1 v_1 + \cdots + x_n v_n) = rL(\mathbf{x}) \\ L(\mathbf{x} + \mathbf{y}) &= (\mathbf{x} + \mathbf{y}) \cdot \mathbf{v} = (x_1 + y_1)v_1 + \cdots + (x_n + y_n)v_n \\ &= (x_1 v_1 + \cdots + x_n v_n) + (y_1 v_1 + \cdots + y_n v_n) = L(\mathbf{x}) + L(\mathbf{y}). \end{aligned}$$

□

The dot product also allows us to compute the angle between two vectors.

Proposition 6.5. *If* $\mathbf{u}, \mathbf{v}$ *are two nonzero vectors in* $\mathbb{R}^n$, *and the angle between them is* θ, *then*

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

Proof. We can form a triangle with sides $\mathbf{u}, \mathbf{v}$, and $\mathbf{u} - \mathbf{v}$. Then by the law of cosines (which I'm sure you all remember from high school trigonometry), we have

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

Then we compute

$$\begin{aligned} \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta &= \frac{1}{2} \left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2 \right) \\ &= \frac{1}{2} \left(\mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - (\mathbf{y} - \mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) \right) \\ &= \frac{1}{2} \left(\mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - (\mathbf{y} \cdot \mathbf{y} - \mathbf{x} \cdot \mathbf{y} - \mathbf{y} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{x}) \right) \\ &= \frac{1}{2} \left(\mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} \right) = \mathbf{x} \cdot \mathbf{y}. \end{aligned}$$

□

Thus the angle between two vectors is given by $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$.

Example 6.6. Let $\mathbf{u} = (3,4)$ and $\mathbf{v} = (-1,7)$. Then $\mathbf{u}\cdot\mathbf{v} = 3\cdot(-1) + 4\cdot 7 = 25$.

We can compute $\|u\| = \sqrt{3^2+4^2} = 25$ and $\|\mathbf{v}\| = \sqrt{(-1)^2+7^2} = 5\sqrt{2}$. The distance between them is $\|\mathbf{u}-\mathbf{v}\| = \|(4,-3)\| = \sqrt{4^2+(-3)^2} = 5$.

The angle between them is given by

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{25}{5\cdot 5\sqrt{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$
$$\theta = \arccos\left(\frac{\sqrt{2}}{2}\right) = \frac{\pi}{4}.$$

We sometimes want to be able to talk about the direction of a vector without worrying about the magnitude. In this case we may wish to comput the *unit vector* given by $\frac{\mathbf{u}}{\|\mathbf{u}\|}$. This vector will clearly have magnitude 1, and point in the same direction that $\mathbf{u}$ does.

If $\mathbf{x},\mathbf{y}$ are unit vectors, then $\cos\theta = \mathbf{x}\cdot\mathbf{y}$.

Example 6.7. The unit vector of $\mathbf{u} = (3,4)$ is $\frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{1}{5}(3,4) = (3/5, 4/5)$. The unit vector of $\mathbf{v} = (-1,7)$ is $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{5\sqrt{2}}(-1,7) = \left(\frac{-1}{5\sqrt{2}}, \frac{7}{5\sqrt{2}}\right)$.

Then the angle between them is given by

$$\cos\theta = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix} \cdot \begin{bmatrix} -1/5\sqrt{2} \\ 7/5\sqrt{2} \end{bmatrix} = \frac{-3}{25\sqrt{2}} + \frac{28}{25\sqrt{2}} = \frac{1}{\sqrt{2}}$$

as before.

There is one more result that is pretty trivial in the case of $\mathbb{R}^n$, but will be very important when we generalize.

Theorem 6.8 (Cauchy-Schwarz Inequality)**.** *If $\mathbf{u},\mathbf{v}$ are vectors in $\mathbb{R}^n$, then*

$$|\mathbf{u}\cdot\mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|. \tag{2}$$

Furthermore, the two sides are equal if and only if either one of the vectors is $\mathbf{0}$, or $\mathbf{u} = r\mathbf{v}$ for some $r \in \mathbb{R}$.

Proof. Recall that $0 \leq |\cos\theta| \leq 1$, with $|\cos\theta| = 1$ if and only if $\theta = n\pi$ for some integer n. Thus

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

Further, the equality holds only if $\|\mathbf{u}\| = 0$, $\|\mathbf{v}\| = 0$, or $\cos\theta = 1$. In the third case this means the angle between the two vectors is an integer multiple of π, so they either point in the same direction, or in opposite directions. ☐

180 degree angles are important, but so are right angles. If two vectors are at a right angle to each other, then we have

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos \pi/2 = \|\mathbf{u}\|\|\mathbf{v}\| \cdot 0 = 0.$$

We give a special name to these vectors:

Definition 6.9. We say that $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$.

Example 6.10. 1. $\mathbf{0}$ is orthogonal to every vector.

2. $(3, 2)$ and $(-4, 6)$ are orthogonal in $\mathbb{R}^2$.

3. Let $\mathbf{u} = (2, 3, 2)$. Can we find a vector orthogonal to it?

   There are lots of them. (They should form an entire plane, if you think about it). One in particular is $(1, 1, -5/2)$.

The last important idea the dot product gives us is the ability to break a vector up into two components. Given $\mathbf{u}$ and $\mathbf{v}$, we can decompose $\mathbf{u}$ into "the part that points in the direction of $\mathbf{v}$" and "the other part."

Suppose we have two vectors $\mathbf{u}$ and $\mathbf{v}$, with angle θ between them. These form two sides of a triangle, with the third side given by $\mathbf{u} - \mathbf{v}$. But we can also draw a line from the endpoint of $\mathbf{u}$ that is perpendicular to $\mathbf{v}$.

We now have a right trangle. The hypotenuse has length $\|\mathbf{u}\|$, so by definition of cosine the length of the adjacent side is $\|\mathbf{u}\| \cos \theta$. But we know that

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

so the length of the adjacent side is $\mathbf{u} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|}$. We sometimes call this number the *scalar projection of* $\mathbf{u}$ *onto* $\mathbf{v}$.

Further, we know the direction that the adjacent side is pointing: it's the same direction as $\mathbf{v}$! So we can find this adjacent side as a vector with the formula

$$\mathbf{p} = \mathbf{u} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}.$$

It is not immediately obvious that this is a vector; but most of the dot products give us scalars, with the final $\mathbf{v}$ giving direction.

Finally, we can write $\mathbf{w} = \mathbf{u} - \mathbf{p}$. We will have that $\mathbf{p} \cdot \mathbf{v} = \|\mathbf{p}\|\|\mathbf{v}\|$ since the two vectors point in the same direction; we will have

$$\begin{aligned}
\mathbf{w} \cdot \mathbf{v} = (\mathbf{u} - \mathbf{p}) \cdot \mathbf{v} &= \mathbf{u} \cdot \mathbf{v} - \mathbf{p} \cdot \mathbf{v} \\
&= \mathbf{u} \cdot \mathbf{v} - \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}(\mathbf{v} \cdot \mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} = \mathbf{0}.
\end{aligned}$$

Thus $\mathbf{w}$ is orthogonal to $\mathbf{v}$. We have written $\mathbf{u} = \mathbf{p} + \mathbf{w}$ so that $\mathbf{w}$ is orthogonal to $\mathbf{v}$, and $\mathbf{p}$ points in the same direction as $\mathbf{v}$.

Definition 6.11. If $\mathbf{u}, \mathbf{v}$ are two vectors in $\mathbb{R}^n$, we define the *projection map onto* $\mathbf{v}$ by

$$\mathrm{proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}.$$

Example 6.12. Let's look back at our earlier vectors $\mathbf{u} = (3, 4)$ and $\mathbf{v} = (-1, 7)$. Then we compute

$$\begin{aligned}
\mathrm{proj}_{\mathbf{v}} \mathbf{u} &= \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v} = \frac{25}{50} \begin{bmatrix} -1 \\ 7 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 7/2 \end{bmatrix} \\
\mathbf{u} - \mathrm{proj}_{\mathbf{v}} \mathbf{u} &= \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} -1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 1/2 \end{bmatrix}.
\end{aligned}$$

6.2 Inner Products

All the ideas of the previous section work in $\mathbb{R}^n$. We want to figure out what the important bits were so that we can use them in other vector spaces. Clearly the most important part was the dot product.

Definition 6.13. An *inner product* on a vector space V is an operation that takes in two vectors $\mathbf{u}, \mathbf{v} \in V$ and returns a real number $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfying the following conditions:

1. (Positive Definite) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.
2. (Symmetric) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
3. (Bilinear) $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \beta\langle \mathbf{v}, \mathbf{w} \rangle$, and $\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{v} \rangle + \beta\langle \mathbf{u}, \mathbf{w} \rangle$.

We write the *norm* of a vector $\mathbf{v}$ as $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

The dot product is clearly an example of an inner product, but there are other important examples we can see.

Example 6.14. Let $V = \mathcal{C}([a,b],\mathbb{R})$ be the space of continuous functions on $[a,b]$, and define an inner product by $\langle f, g\rangle = \int_a^b f(t)g(t)\,dt$. Then

1. $\langle f, f\rangle = \int_a^b f(t)^2\,dt \geq 0$ since $f(t)^2 \geq 0$; and further the integral is zero if and only if $f(t)^2 = 0$ everywhere.
2. $\langle f, g\rangle = \int_a^b f(t)g(t)\,dt = \int_a^b g(t)f(t)\,dt = \langle g, f\rangle$.
3. $\langle \alpha f + \beta g, h\rangle = \int_a^b (\alpha f(t) + \beta g(t))h(t)\,dt = \alpha\int_a^b f(t)h(t)\,dt + \beta\int_a^b g(t)h(t)\,dt = \alpha\langle f, h\rangle + \beta\langle g, h\rangle$.

Thus this is an inner product on $\mathcal{C}([a,b],\mathbb{R})$ by definition.

Example 6.15. Let $V = \mathcal{P}_n(x)$ and fix real numbers $x_0, x_1, \ldots, x_n$ be distinct real numbers. For $f, g \in V$, define

$$\langle f, g\rangle = \sum_{i=0}^n f(x_i)g(x_i).$$

Then we can see $\langle f, f\rangle = \sum_{i=0}^n f(x_i)^2 \geq 0$, and the sum is equal to zero if and only if $f(x_i) = 0$ for all i. But then f is a degree n polynomial with $n+1$ roots, and so must be constantly zero.

You will check the other two conditions on your homework.

We'd like to check that this inner product gives us all the things that the dot product did. In particular we want it to give us distance and angle and projections.

Definition 6.16. Let $\mathbf{u}, \mathbf{v}$ be elements of an inner product space V. If $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, we say that $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal*.

We will eventually see that this corresponds to the two vectors being at a "right angle" to each other. But more immediately, we'll see that this means they are independent in a very specific way.

Definition 6.17. Suppose $\mathbf{u}, \mathbf{v}$ are vectors in an inner product space V, and $\mathbf{v} \neq 0$. We define the projection of $\mathbf{u}$ onto $\mathbf{v}$ by

$$\operatorname{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}.$$

Proposition 6.18. *Let $\mathbf{u}, \mathbf{v}$ be vectors in an inner product space V, with $\mathbf{v} \neq 0$. Let $\mathbf{p} = \operatorname{proj}_{\mathbf{v}}\mathbf{u}$. Then:*

1. *$\langle \mathbf{u} - \mathbf{p}, \mathbf{p}\rangle = 0$—that is, $\mathbf{u} - \mathbf{p}$ is orthogonal to $\mathbf{p}$.*

2. $\mathbf{u} = \mathbf{p}$ *if and only if* $\mathbf{u}$ *is a scalar multiple of* $\mathbf{v}$.

Proof. 1. Excerise.

2. If $\mathbf{u} = \beta\mathbf{v}$, then
$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \beta\mathbf{v}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v} = \beta\mathbf{v} = \mathbf{u}.$$
Conversely, suppose $\mathbf{u} = \text{proj}_{\mathbf{v}}\mathbf{u}$. Then by definition
$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v},$$
so set $\beta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}$ and we have $\mathbf{u} = \beta\mathbf{v}$.

□

Example 6.19. Let $V = \mathcal{C}([-1,1], \mathbb{R})$ be the space of continuous functions on the closed interval $[-1,1]$, with the inner product given as above. Consider the vectors $1, x$. We compute:
$$\begin{aligned}
\|1\| &= \sqrt{\int_{-1}^{1} 1\,dx} = \sqrt{x|_{-1}^{1}} = \sqrt{2} \\
\|x\| &= \sqrt{\int_{-1}^{1} x^2\,dx} = \sqrt{x^3/3|_{-1}^{1}} = \sqrt{2/3} \\
\langle 1, x\rangle &= \int_{-1}^{1} x\,dx = x^2/2|_{-1}^{1} = 0
\end{aligned}$$
so 1 and x are orthogonal. Thus the projection of x onto 1 will give the zero vector: the two vectors have no "direction" in common.

Let's consider now the vector $1 + x$. We have
$$\begin{aligned}
\langle 1 + x, 1\rangle &= \int_{-1}^{1} 1 + x\,dx = x + x^2/2|_{-1}^{1} = 2 \\
\langle 1 + x, x\rangle &= \int_{-1}^{1} x + x^2\,dx = x^2/2 + x^3/3|_{-1}^{1} = 2/3.
\end{aligned}$$
Now we compute
$$\begin{aligned}
\text{proj}_1 1 + x &= \frac{\langle 1 + x, 1\rangle}{\langle 1, 1\rangle}1 = \frac{2}{2}1 = 1 \\
\text{proj}_x 1 + x &= \frac{\langle 1 + x, x\rangle}{\langle x, x\rangle}x = \frac{2/3}{2/3}x = x.
\end{aligned}$$
Thus we can use the inner product to decomopose $1 + x$ into its 1 component and its x component (and the remainder, if there were any).

If two vectors are orthogonal, then they are independent; they don't have any reasonable sub-components pointing in the same direction. This means their lengths are in some sense independent.

Proposition 6.20 (Pythagorean Law)**.** *If* $\mathbf{u}, \mathbf{v}$ *are orthogonal, then*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. Exercise. □

Example 6.21. Returning to our previous example, we can compute that

$$\|1+x\| = \sqrt{\int_{-1}^{1} 1 + 2x + x^2 \, dx} = \sqrt{x + x^2 + x^3/3|_{-1}^{1}} = \sqrt{8/3}.$$

We can confirm that indeed,

$$\|1+x\|^2 = 8/3 = 2 + 2/3 = \|1\| + \|x\|.$$

Using projections we can prove that the Cauchy-Schwarz Inequality, which we saw in theorem 6.8, holds for any inner product.

Theorem 6.22 (Cauchy-Schwarz Inequality)**.** *If* $\mathbf{u}, \mathbf{v}$ *are in an inner product space* V*, then*

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|. \tag{3}$$

Equality holds if and only if $\mathbf{u}$ *and* $\mathbf{v}$ *are linearly dependent.*

Proof. If $\mathbf{v} = \mathbf{0}$, both sides are zero. So assume $\mathbf{v} \neq 0$.

Let $\mathbf{p} = \text{proj}_{\mathbf{v}}\mathbf{u}$. By the Pythagorean law 6.20, we know that

$$\|\mathbf{u}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{u}-\mathbf{p}\|^2.$$

But we know that

$$\|\mathbf{p}\|^2 = \left\| \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}\right\|^2 = \left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\right)^2 \|\mathbf{v}\|^2 = \frac{\langle \mathbf{u}, \mathbf{v}\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle}.$$

Thus we have

$$\begin{aligned}
\frac{\langle \mathbf{u}, \mathbf{v}\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle} &= \|\mathbf{u}\|^2 - \|\mathbf{u}-\mathbf{p}\|^2 \\
\langle \mathbf{u}, \mathbf{v}\rangle^2 &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{p}\|^2\|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2 \\
|\langle \mathbf{u}, \mathbf{v}\rangle| &\leq \|\mathbf{u}\|\|\mathbf{v}\|.
\end{aligned}$$

Further, we can easily see that we get equality if and only if $\mathbf{u} - \mathbf{p} = \mathbf{0}$, if and only if $\mathbf{u} = \mathbf{p}$, if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$. □

Notice that this allows us to define an "angle" between two vectors. Cauchy-Schwarz tells us that

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \leq 1,$$

so we can coherently define:

Definition 6.23. If $\mathbf{u}, \mathbf{v}$ are non-zero vectors in an inner product space, we define the angle between them to be

$$\theta = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|}\right).$$

Finally we'd like to return to the idea of distance, by thinking about what properties a "distance" function *should* have. We get

Definition 6.24. A vector space V together with an operation $\| \cdot \| : V \to \mathbb{R}$ is said to be a *normed linear space* if:

1. $\|\mathbf{v}\| \geq 0$ for any $\mathbf{v} \in V$, with equality if and only if $\mathbf{v} = \mathbf{0}$.
2. $\|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$ for any $\alpha \in \mathbb{R}, \mathbf{v} \in V$.
3. (Triangle inequality) $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$.

Remark 6.25. These three conditions are equivalent to:

1. Nothing has a negative length, and only the zero vector has zero length.
2. Stretching a vector by a scalar multiplies its length by that scalar.
3. The sum of the lengths of two sides of a triangle is greater than the length of the third side. In other words, you can't get somewhere faster by adding a detour in the middle.

A normed linear space is in some sense the right setting in which to do calculus.

Proposition 6.26. *Let V be an inner product space. Then V is a normed linear space with norm given by $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.*

Proof. The first two conditions are easy to prove, so we'll just check the triangle inequality.

We compute

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle \\
&\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\| \|\mathbf{v}\| && \text{Cauchy-Schwarz Inequality} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 .
\end{aligned}$$

Taking the square root of both sides gives

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

as desired. □

Remark 6.27. There are norms that do not come from inner products. A good example is the norm on $\mathbb{R}^n$ given by $\|(a_1, a_2, \dots, a_n\|_1 = |a_1| + |a_2| + \cdots + |a_n|$. We won't worry too much about those in this course, though.

6.3 Orthonormal Bases

Throughout the course, we've been suggesting that we would often like to change from one coordinate system into another which is easier to work with. In this section we'll discuss one particular type of nice basis: one in which all the basis elements are orthogonal.

Definition 6.28. A set $S = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is said to be *orthogonal* if $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ whenever $i \neq j$. We say it is *orthonormal* if every vector has magnitude 1.

Proposition 6.29. *Any orthogonal set of non-zero vectors is linearly independent.*

Proof. Suppose

$$a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n = \mathbf{0}.$$

Then dotting the equation with itself, we get

$$\begin{aligned} \langle a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n, a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n \rangle &= 0 \\ \sum_{i,j=1}^{n} a_i a_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle &= 0 \end{aligned}$$

But since the $\mathbf{u}_i$ are orthogonal, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ when $i \neq j$, so this just gives us

$$\begin{aligned} \sum_{i=1}^{n} a_i^2 \langle \mathbf{u}_i, \mathbf{u}_i \rangle &= 0 \\ a_1^2 \|\mathbf{u}_1\|^2 + \cdots + a_n^2 \|\mathbf{u}_n\|^2 &= 0. \end{aligned}$$

And thus, since $\|\mathbf{u}_i\| > 0$, we must have $a_i = 0$ for each i. □

Thus every orthogonal set is a basis for its span.

Definition 6.30. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V. We say that E is an *orthogonal basis* if $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ whenever $i \neq j$.

We say that E is an *orthonormal basis* if, furthermore, $\|\mathbf{e}_i\| = 1$. Thus E is an orthonormal basis if and only if

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Example 6.31.

- The standard basis for $\mathbb{R}^3$ is orthonormal.
- The basis $\{(1,1,0),(1,-1,0),(0,0,1)\}$ for $\mathbb{R}^3$ is orthogonal but not orthonormal.
  But $\{(\sqrt{2}/2, \sqrt{2}/2, 0), (\sqrt{2}/2, -\sqrt{2}/2, 0), (0,0,1)\}$ is orthornormal.
- Let $V = \mathcal{P}_2(x)$ with inner product given by $\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)\,dt$. The basis $E = \{1, x, 3x^2 - 1\}$ is an orthogonal basis for V, but not orthonormal.
  The basis $F = \left\{ \frac{1}{\sqrt{2}}, \frac{x\sqrt{3}}{\sqrt{2}}, \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1) \right\}$ is orthonormal.
- Let $V = \mathcal{P}_2(x)$ with inner product given by $\langle f, g \rangle = f(-1)g(-1) + f(0)g(0) + f(1)g(1)$. Then $E = \{1, x, x^2 - 2/3\}$ is an orthogonal basis for V.
  An orthonormal basis is $F = \left\{ \frac{\sqrt{3}}{3}, \frac{x\sqrt{2}}{2}, \frac{\sqrt{3}}{\sqrt{2}}\left(x^2 - \frac{2}{3}\right) \right\}$.

Orthonormal bases are particularly nice, for a few reasons.

Proposition 6.32. *Suppose $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is an orthonormal basis for V. Then if $\mathbf{u} = \sum_{i=1}^n a_i \mathbf{e}_i$ and $\mathbf{v} = \sum_{i=1}^n b_i \mathbf{e}_i$, then $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n a_i b_i$.*

Consequently $\|\mathbf{u}\| = |a_1|^2 + \cdots + |a_n|^2$.

Remark 6.33. We use this all the time when we're computing the norm of vectors in $\mathbb{R}^n$. This also gives us our "normal" dot product.

Mor importantly, orthonormal bases make projection, coordinates, and changes of basis very easy.

Proposition 6.34. *Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V, with $\langle e_i, e_j \rangle = 0$ when $i \neq j$. Then if $\mathbf{v} \in V$, we have*

$$\mathbf{v} = \sum_{i=1}^n \left(\operatorname{proj}_{\mathbf{e}_i} \mathbf{v}\right) \mathbf{e}_i = \left(\operatorname{proj}_{\mathbf{e}_1} \mathbf{v}\right) \mathbf{e}_1 + \cdots + \left(\operatorname{proj}_{\mathbf{e}_n} \mathbf{v}\right) \mathbf{e}_n.$$

Proof. Write $\mathbf{v} = a_1 \mathbf{e}_1 + \cdots + a_n \mathbf{e}_n$ and compute each projection. □

Corollary 6.35. *Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be an orthonormal basis for V. Then*

$$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \mathbf{e}_i = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \cdots + \langle \mathbf{v}, \mathbf{e}_n \rangle \mathbf{e}_n.$$

Example 6.36. $E = \{(1,1,0), (1,-1,0), (0,0,1)\}$ is orthogonal. Find the E coordinates of $(6,2,1)$.

We compute:

$$\begin{aligned}
\mathrm{proj}_{\mathbf{e}_1}(6,2,1) &= \frac{(6,2,1)\cdot(1,1,0)}{(1,1,0)\cdot(1,1,0)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{8}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\
\mathrm{proj}_{\mathbf{e}_2}(6,2,1) &= \frac{(6,2,1)\cdot(1,-1,0)}{(1,-1,0)\cdot(1,-1,0)} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{4}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \\
\mathrm{proj}_{\mathbf{e}_3}(6,2,1) &= \frac{(6,2,1)\cdot(0,0,1)}{(0,0,1)\cdot(0,0,1)} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
[(6,2,1)]_E &= (4,2,1)
\end{aligned}$$

Example 6.37. Let $V = \mathcal{P}_2(x)$, with inner product given by $\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Then $E = \{1, x, 3x^2 - 1)$ is orthogonal. Write $3x^2 - 6x + 4$ in E-coordinates.

We compute

$$\begin{aligned}
\mathrm{proj}_{\mathbf{e}_1} 3x^2 - 6x + 4 &= \frac{\langle 3x^2 - 6x + 4, 1 \rangle}{\langle 1, 1 \rangle}(1) = \frac{1}{2}\int_{-1}^{1} 3x^2 - 6x + 4\,dx(1) \\
&= \frac{1}{2}\left(x^3 - 3x^2 + 4x \mid |_{-1}^{1}\right)(1) = 5(1) \\
\mathrm{proj}_{\mathbf{e}_2} 3x^2 - 6x + 4 &= \frac{\langle 3x^2 - 6x + 4, x \rangle}{\langle x, x \rangle}(x) = \frac{3}{2}\int_{-1}^{1} 3x^3 - 6x^2 + 4x\,dx(x) \\
&= \frac{3}{2}\left(\frac{x^4}{4} - 2x^3 + 2x^2 \mid |_{-1}^{1}\right)(x) = -6(x) \\
\mathrm{proj}_{\mathbf{e}_3} 3x^2 - 6x + 4 &= \frac{\langle 3x^2 - 6x + 4, 3x^2 - 1 \rangle}{\langle 3x^2 - 1, 3x^2 - 1 \rangle}(3x^2 - 1) \\
&= \frac{5}{8}\int_{-1}^{1} (3x^2 - 6x + 4)(3x^2 - 1)\,dx(3x^2 - 1) \\
&= \frac{5}{8}\left(9x^5/5 - 9x^4/2 + 3x^3 + 3x^2 - 4x \mid |_{-1}^{1}\right)(3x^2 - 1) = 1(3x^2 - 1) \\
\left[3x^2 - 6x + 4\right]_E &= (5, -6, 1).
\end{aligned}$$

Example 6.38. Let $V = \mathbb{R}^3$ with the usual dot product. Then the standard basis $\{(1,0,0),(0,1,0),(0,0,1)\}$ is an orthonormal basis. Use the dot product to find the coordinates of $(2,3,4)$.

We don't need to use the full projection operator; we just need to compute the inner products, since our basis is orthonormal and not just orthogonal.

$$\operatorname{proj}_{\mathbf{e}_1}(2,3,4) = (2,3,4)\cdot(1,0,0)\begin{bmatrix}1\\0\\0\end{bmatrix} = 2\begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$\operatorname{proj}_{\mathbf{e}_2}(2,3,4) = (2,3,4)\cdot(0,1,0)\begin{bmatrix}0\\1\\0\end{bmatrix} = 3\begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$\operatorname{proj}_{\mathbf{e}_3}(2,3,4) = (2,3,4)\cdot(0,0,1)\begin{bmatrix}0\\0\\1\end{bmatrix} = 4\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$[(2,3,4)]_E = (2,3,4).$$

This isn't a surprise because it was already in coordinates with respect to the standard basis. But this also illustrates a more general principle: if your vector is already written in orthonormal coordinates, your inner product just becomes a dot product.

We'd like a way to generate an orthonormal basis if we don't already have one. This turns out to be straightforward; start with any basis, and one-by-one "fix" elements so that they're orthogonal to all the others.

Proposition 6.39 (Gram-Schmidt Process)**.** *Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V. Then there is an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, where we set:*

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{e}_1 & \mathbf{u}_1 &= \frac{\mathbf{f}_1}{\|\mathbf{f}_1\|}\\
\mathbf{f}_2 &= \mathbf{e}_2 - \operatorname{proj}_{\mathbf{u}_1}\mathbf{e}_2 & \mathbf{u}_2 &= \frac{\mathbf{f}_2}{\|\mathbf{f}_2\|}\\
\mathbf{f}_3 &= \mathbf{e}_3 - \left(\operatorname{proj}_{\mathbf{u}_1}\mathbf{e}_3 + \operatorname{proj}_{\mathbf{u}_2}\mathbf{e}_3\right) & \mathbf{u}_3 &= \frac{\mathbf{f}_3}{\|\mathbf{f}_3\|}\\
&\vdots & &\vdots\\
\mathbf{f}_n &= \mathbf{e}_n - (\operatorname{proj}_{\mathbf{u}_1}\mathbf{e}_n + \dots + \operatorname{proj}_{\mathbf{u}_{n-1}}\mathbf{e}_n) & \mathbf{u}_n &= \frac{\mathbf{f}_n}{\|\mathbf{f}_n\|}.
\end{aligned}$$

Proof. It's clear that each $\mathbf{u}_i$ has norm 1, so we just need to check that they are pairwise orthogonal, which is the same as checking that the $\mathbf{f}_i$ are all orthogonal

But we have constructed the $\mathbf{f}_i$ to be orthogonal by subtracting off the pieces they have in common. For instance, we see that

$$\begin{aligned}\langle \mathbf{f}_1, \mathbf{f}_2\rangle &= \left\langle \mathbf{e}_1, \mathbf{e}_2 - \frac{\langle \mathbf{e}_2, \mathbf{e}_1\rangle}{\langle \mathbf{e}_1, \mathbf{e}_1\rangle}\mathbf{e}_1 \right\rangle = \langle \mathbf{e}_1, \mathbf{e}_2\rangle - \frac{\langle \mathbf{e}_2, \mathbf{e}_1\rangle}{\langle \mathbf{e}_1, \mathbf{e}_1\rangle}\langle \mathbf{e}_1, \mathbf{e}_1\rangle \\ &= \langle \mathbf{e}_1, \mathbf{e}_2\rangle - \langle \mathbf{e}_2, \mathbf{e}_1\rangle = 0.\end{aligned}$$

In general, we see that

$$\langle \mathbf{f}_j, \operatorname{proj}_{\mathbf{f}_j} \mathbf{f}_i\rangle = \left\langle \mathbf{f}_i, \frac{\langle \mathbf{f}_j, \mathbf{f}_j\rangle}{\langle \mathbf{f}_j, \mathbf{f}_j\rangle}\mathbf{f}_j \right\rangle = \langle \mathbf{f}_i, \mathbf{f}_j\rangle$$

and all the other projections will be zero since the $\mathbf{f}_i$ are orthogonal, so each $\mathbf{f}_j$ is orthogonal to all the previous $\mathbf{f}_i$. □

Example 6.40. Let $V = \mathbb{R}^3$ with the usual dot product, and let $E = \{(1, 1, -1), (1, -1, 1), (-1, 1, 1)\}$. Use Gram-Schmidt to orthonormalize this basis.

We take $\mathbf{f}_1 = (1, 1, -1)$, and then we compute $\|\mathbf{f}_1\| = \sqrt{3}$ so $\mathbf{u}_1 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}\right)$.

Then we set

$$\begin{aligned}\mathbf{f}_2 &= \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \operatorname{Proj}_{(1,1,-1)} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{(1,-1,1)\cdot(1,1,-1)}{(1,-1,1)\cdot(1,-1,1)} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{-1}{3}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4/3 \\ -2/3 \\ 2/3 \end{bmatrix} \\ \mathbf{u}_2 &= \frac{\mathbf{f}_2}{\|\mathbf{f}_2\|} = \frac{(4/3, -2/3, 2/3)}{\sqrt{24/9}} = \frac{\sqrt{3}}{2\sqrt{2}}\begin{bmatrix} 4/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} \sqrt{6}/3 \\ -\sqrt{6}/6 \\ \sqrt{6}/6 \end{bmatrix}.\end{aligned}$$

Finally we have

$$\begin{aligned}
\mathbf{f}_3 &= \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} - \operatorname{proj}_{(1,1,-1)} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} - \operatorname{proj}_{(4,-2,2)} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} - \frac{(-1,1,1)\cdot(1,1,-1)}{(1,1,-1)\cdot(1,1,-1)} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \frac{(-1,1,1)\cdot(4,-2,2)}{(4,-2,2)\cdot(4,-2,2)} \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} - \frac{-1}{3} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \frac{-4}{24} \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\
\mathbf{u}_3 &= \frac{\mathbf{f}_3}{\|\mathbf{f}_3\|} = \frac{(0,1,1)}{\sqrt{2}} = \begin{bmatrix} 0 \\ \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}.
\end{aligned}$$

Thus an orthonormal basis for $\mathbb{R}^3$ is

$$\left\{ \begin{bmatrix} \sqrt{3}/3 \\ \sqrt{3}/3 \\ -\sqrt{3}/3 \end{bmatrix}, \begin{bmatrix} \sqrt{6}/3 \\ -\sqrt{6}/6 \\ \sqrt{6}/6 \end{bmatrix}, \begin{bmatrix} 0 \\ \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \right\}.$$

Notice that while this is an orthonormal basis for $\mathbb{R}^3$, it is not the usual one. We will get different orthonormal bases out of the end, depending on which vector we start with.

Example 6.41. Let $V = \mathcal{P}_2(x)$ with the inner product given by $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$. (Note this is a *different* inner product from the one we've been using!) Let's form an orthonormal basis from the set $\{1, x, x^2\}$.

We set $\mathbf{f}_1 = 1$. We compute that

$$\|1\|^2 = \langle 1, 1 \rangle = \int_0^1 1\,dx = 1$$

so this is already a unit vector; we set $\mathbf{u}_1 = 1$.

We take

$$\mathbf{f}_2 = x - \frac{\langle 1, x \rangle}{\langle 1, 1 \rangle}(1) = x - \frac{1}{2}(1) = x - 1/2.$$

We compute

$$\|\mathbf{f}_2\| = \sqrt{\int_0^1 (x - 1/2)^2\,dx} = \sqrt{12} = 2\sqrt{3}$$

so we set

$$\mathbf{u}_2 = \frac{\mathbf{f}_2}{\|\mathbf{f}_2\|} = 2\sqrt{3}(x - 1/2) = \sqrt{3}(2x - 1).$$

Finally, we have

$$\begin{aligned}
\mathbf{f}_3 &= x^2 - \frac{\langle 1, x^2 \rangle}{\langle 1, 1 \rangle}(1) - \frac{\langle \sqrt{3}(2x-1), x^2 \rangle}{\langle \sqrt{3}(2x-1), \sqrt{3}(2x-1) \rangle}\sqrt{3}(2x-1) \\
&= x^2 - \int_0^1 x^2 \, dx(1) - \sqrt{3}\int_0^1 2x^3 - x^2 \, dx(\sqrt{3}(2x-1)) \\
&= x^2 - \frac{1}{3} - \frac{1}{2}(2x-1) = x^2 - x + \frac{1}{6}.
\end{aligned}$$

Then we compute

$$\begin{aligned}
\|\mathbf{f}_3\| &= \sqrt{\int_0^1 (x^2 - x + 1/6)^2 \, dx} = \sqrt{\frac{1}{180}} = \frac{1}{6\sqrt{5}} \\
\mathbf{u}_3 &= \frac{\mathbf{f}_3}{\|\mathbf{f}_3\|} = 6\sqrt{5}x^2 - 6\sqrt{5}x + \sqrt{5}.
\end{aligned}$$

Thus an orthonormal basis for $\mathcal{P}_2(x)$ with this inner product is

$$\{1, \sqrt{3}(2x-1), \sqrt{5}(6x^2 - 6x + 1)\}.$$

6.4 Orthogonal Subspaces

We have used orthogonality to give a vector space a particularly nice basis. We can also break the vector space into two (or more) independent subspaces.

Definition 6.42. If V is an inner product space and U, W are subspaces, we say that U and W are *orthogonal* and write $U \perp W$ if $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ for every $\mathbf{u} \in U, \mathbf{w} \in W$.

If $U \subset V$, we define the *orthogonal complement of* U to be the set of all vectors perpendicular to everything in U:

$$U^\perp = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{u} \rangle = 0 \forall \mathbf{u} \in U\}.$$

Example 6.43.

- In $\mathbb{R}^2$, the orthogonal complement of a line is a line. The orthogonal complement to a set with two points in it is also a line.
- In $\mathbb{R}^3$, the orthogonal complement of a line is a plane, and the orthogonal complement of a plane is a line.

Proposition 6.44. *If U is a subset of V, then $U^\perp$ is a subspace of V.*

Proof. 1. $\mathbf{0}$ is orthogonal to everything, and thus is in $U^\perp$.

2. Suppose $\mathbf{v} \in U^\perp$, and $r \in \mathbb{R}$. Then for any $\mathbf{u} \in U$ we have $\langle r\mathbf{v}, \mathbf{u}\rangle = r\langle \mathbf{v}, \mathbf{u}\rangle = r \cdot 0 = 0$, so $r\mathbf{v} \in U^\perp$ by definition.

3. Suppose $\mathbf{v}, \mathbf{w} \in U^\perp$, and let $\mathbf{u} \in U$. Then

$$\langle \mathbf{v} + \mathbf{w}, \mathbf{u}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{u}\rangle = 0 + 0 = 0.$$

   Thus $\mathbf{v} + \mathbf{w}$ is orthogonal to $\mathbf{u}$ for every $\mathbf{u} \in U$, and so $\mathbf{v} + \mathbf{w} \in U^\perp$.

Thus by the subspace theorem, $\mathbf{U}^\perp$ is a subspace of V. □

Remark 6.45. We will usually consider cases where U is also a subspace of V, but this isn't necessary; nothing above assumes anything about the structure of U.

A basic thing we want to do is, given a subspace, find a basis for the subspace and for its orthogonal complement. As with everything else, we can solve this problem by row-reducing matrices.

Proposition 6.46. *Let A be a matrix. Then* $\ker(A) = (\operatorname{row}(A))^\perp$.

Remark 6.47. In three dimensions, we can use this exact formula to find the normal vector to a plane.

Proof. If $\mathbf{r}_i$ are the rows of the matrix A, then

$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_m \cdot \mathbf{x} \end{bmatrix}$$

and thus $\mathbf{x} \in \ker(A)$ precisely if $\mathbf{x}$ is orthogonal to every row of A. But if $\mathbf{x}$ is orthogonal to every row vector of A, it is orthogonal to every linear combination of them, and thus is orthogonal to their span, which is the rowspace. □

Example 6.48. Suppose we want to find the orthogonal complement to $U = \operatorname{span}\{(1, 4, 2), (1, 1, 1)\}$. Then we write down the matrix

$$A = \begin{bmatrix} 1 & 4 & 2 \\ 1 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2/3 \\ 0 & 3 & 1 \end{bmatrix}$$

so $U^\perp = \ker(A) = \{(-2\alpha, -\alpha, 3\alpha)\} = \operatorname{span}\{(2, 1, -3)\}$. We can check that this is in fact orthogonal to the original two vectors.

There are a couple more useful facts we'd like to know about orthogonal complements, which show that they relate spaces in useful ways.

Proposition 6.49. *If U is a subspace of V and $\mathbf{v} \in V$, then there exist unique $\mathbf{v}_U \in U, \mathbf{v}_{U^\perp} \in U^\perp$ such that $\mathbf{v} = \mathbf{v}_U + \mathbf{v}_{U^\perp}$.*

We say that this is an orthogonal decomposition *of* $\mathbf{v}$.

Proof. Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be an orthogonal basis for U and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ an orthogonal basis for $U^\perp$.

We claim that $E \cup F = \{\mathbf{e}_1, \dots, \mathbf{e}_n, \mathbf{f}_1, \dots, \mathbf{f}_m\}$ is an orthogonal basis for V. It must be orthogonal since E and F are orthogonal sets, and thus it is linearly independent. So we need to show that it spans V.

Suppose $\mathbf{v} \in V$, and consider the element

$$\mathbf{v}' = \mathbf{v} - \sum_{i=1}^{n} \operatorname{proj}_{\mathbf{e}_i} \mathbf{v}.$$

This is an element of V, and by construction it is orthogonal to every $\mathbf{e}_i$ and thus all of U, so $\mathbf{v}' \in U^\perp$. Thus $\mathbf{v}' \in \operatorname{span}(F)$ and so $\mathbf{v} \in \operatorname{span}(E \cup F)$. Thus $E \cup F$ spans V.

Then every element of V can be expressed uniquely as a linear combination of elements of E and F. This gives us a unique representation as a sum of an element of U and an element of $U^\perp$.

□

Corollary 6.50. $\dim U + \dim U^\perp = \dim V$.

Example 6.51. Give the orthogonal decomposition of $(3, -1, 2)$ with respect to the subspace given by $x - y + 2z = 0$ and its complement.

We need to find an orthonormal basis for either $x - y + 2z = 0$ or its orthogonal complement. But we can see that the normal vector to this plane is in the orthogonal complement, so $\{(1, -1, 2)\}$ is a basis for $U^\perp$.

We project $(3, -1, 2)$ onto $\operatorname{span}\{(1, -1, 2)\}$. We have

$$\operatorname{proj}_{(1,-1,2)} \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = \frac{(3,-1,2)\cdot(1,-1,2)}{(1,-1,2)\cdot(1,-1,2)} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \frac{8}{6} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4/3 \\ -4/3 \\ 8/3 \end{bmatrix}$$

So this is the projection into $U^\perp$. The projection into U then is just what's left over: it's

$$\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} - \operatorname{proj}_{(1,-1,2)} \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 - 4/3 \\ -1 + 4/3 \\ 2 - 8/3 \end{bmatrix} = \begin{bmatrix} 5/3 \\ 1/3 \\ -2/3 \end{bmatrix}.$$

(We check that this vector is in fact in the plane U). Then we have an orthogonal decomposition: $(3, -1, 2) = (5/3, 1/3, -2/3) + (4/3, -4/3, 8/3)$.

Example 6.52. Let $V = \mathbb{R}^4$ and let $U = \text{span}\{(1, 2, 3, 4), (2, 1, -1, -2)\}$. Find the orthogonal decomposition of $(1, 1, 1, 1)$ into its components in U and $U^\perp$.

We write a matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & -1 & -2 \end{bmatrix} \to \begin{bmatrix} 3 & 0 & -5 & -8 \\ 0 & 3 & 7 & 10 \end{bmatrix}$$

so $\ker(A) = (5\alpha + 8\beta, -7\alpha - 10\beta, 3\alpha, 3\beta) = \text{span}\{(5, -7, 3, 0), (8, -10, 0, 3)\}$.

We need to find an orthogonal basis for either U or $U^\perp$.We compute

$$\mathbf{f}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$\mathbf{f}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} - \text{proj}_{(1,2,3,4)} \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} - \frac{(2, 1, -1, -2) \cdot (1, 2, 3, 4)}{(1, 2, 3, 4) \cdot (1, 2, 3, 4)} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} - \frac{-7}{30} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \frac{1}{30} \begin{bmatrix} 67 \\ 44 \\ -9 \\ -32 \end{bmatrix}$$

We compute

$$\operatorname{proj}_{\mathbf{f}_1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \frac{(1,2,3,4)\cdot(1,1,1,1)}{(1,2,3,4)\cdot(1,2,3,4)}\begin{bmatrix}1\\2\\3\\4\end{bmatrix} = \frac{10}{30}\begin{bmatrix}1\\2\\3\\4\end{bmatrix} = \begin{bmatrix}1/3\\2/3\\1\\4/3\end{bmatrix}$$

$$\operatorname{proj}_{\mathbf{f}_2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \frac{(1,1,1,1)\cdot(67,44,-9,-32)}{(67,44,-9,-32)\cdot(67,44,-9,-32)}\begin{bmatrix}67\\44\\-9\\-32\end{bmatrix} = \frac{70}{7530}\begin{bmatrix}67\\44\\-9\\-32\end{bmatrix} = \frac{7}{753}\begin{bmatrix}67\\44\\-9\\-32\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}_U = \begin{bmatrix}1/3\\2/3\\1\\4/3\end{bmatrix} + \frac{7}{753}\begin{bmatrix}67\\44\\-9\\-32\end{bmatrix} = \frac{10}{251}\begin{bmatrix}24\\27\\23\\26\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}_{U^\perp} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \begin{bmatrix}1\\1\\1\\1\end{bmatrix}_U = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \frac{10}{251}\begin{bmatrix}24\\27\\23\\26\end{bmatrix} = \frac{1}{251}\begin{bmatrix}11\\-19\\21\\-9\end{bmatrix}.$$

Proposition 6.53. *If U is a subspace of V, then $(U^\perp)^\perp = U$.*

Proof. If $\mathbf{u} \in U$, then $\mathbf{u}$ is orthogonal to every $\mathbf{w} \in U^\perp$ by definition. So $U \subset (U^\perp)^\perp$.

Conversely, suppose $\mathbf{w} \in (U^\perp)^\perp$. We can write $\mathbf{w} = \mathbf{w}_U + \mathbf{w}_{U^\perp}$. Then $\mathbf{w} \in (U^\perp)^\perp$ so we know $\langle \mathbf{w}, \mathbf{w}_{U^\perp}\rangle = \mathbf{0}$.

But $\langle \mathbf{w}, \mathbf{w}_{U^\perp}\rangle = \langle \mathbf{w}_{U^\perp}, \mathbf{w}_{U^\perp}\rangle = 0$ if and only if $\mathbf{w}_{U^\perp} = 0$. Thus $\mathbf{w}_{U^\perp} = \mathbf{0}$, and $\mathbf{w} = \mathbf{w}_U \in U$. ☐

7 Data and Statistics

With the last few days of the course, I want to talk about a major application of linear algebra that will be relevant to many of you in future (or present!) courses, and also is surprisingly relevant to daily life.

7.1 Approximate Linear Equations and Least-Squares Fits

For most of the course, we've been explicitly or implicitly studying equations like $A\mathbf{x} = \mathbf{b}$. We saw in section 1 that not all equations like this have solutions. Specifically, $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in the columnspace of A. In that case, we can do a row reduction to solve for $\mathbf{x}$.

But if $\mathbf{b}$ is not in the columnspace, then the equation doesn't have any solutions. So suppose $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$. We can see that there is no solution to $A\mathbf{x} = \mathbf{b}$, since we can't write $\mathbf{b}$ as a linear combination of $\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

But sometimes we don't need to have an exact solution; a "close enough" solution will do. Sometimes this is because getting close to the right answer is good enough. More often, we're genuinely interested in how close we can get.

Definition 7.1 (Informal definition)**.** In *linear regression*, we have a collection of "dependent" variables that represent outcomes we want to predict, and a collection of "independent" variables that we're using as predictors. We want to find the linear function that's closest to predicting the dependent variables, given the independent variables.

In order to talk about which solutions are "closest", we need some definition of closeness. In section 6.1 we said that we could capture distance with the idea of a norm. There are many norms on $\mathbb{R}^2$, but the simplest one to work with is the Euclidean norm that comes from the dot product.

Remark 7.2. The most common other norm that gets used is called the L^1 norm, and is defined by $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$. This norm puts more emphasis on average behavior and cares less about a couple of bad outliers than the Euclidean L^2 norm. This is called finding the line of least absolute deviation. There are some advantages, but practically speaking it's a lot harder to work with, so it wasn't used much until the advent of computers. In Big Data applications it's pretty common.

So we have a matrix $A \in M_{m\times n}$ and a "target" vector $\mathbf{b} \in \mathbb{R}^m$, and we want to find the vector $\mathbf{x} \in \mathbb{R}^n$ that comes closest to solving $A\mathbf{x} = \mathbf{b}$. That is, we want to minimize $\|A\mathbf{x} - \mathbf{b}\|$.

Definition 7.3. Let $A \in M_{m\times n}$ and $b \in \mathbb{R}^m$. A *least-squares solution* of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for any $\mathbf{x} \in \mathbb{R}^n$.

So for our example, the least-squares solution wants to find the x_1, x_2 that minimizes

$$\left\| x_1 \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \right\|$$

or

$$\left\| \begin{bmatrix} 4x_1 - 2 \\ 2x_2 \\ x_1 + x_2 - 11 \end{bmatrix} \right\| = \sqrt{(4x_1 - 2)^2 + (2x_2)^2 + (x_1 + x_2 - 11)^2}.$$

So far we've been thinking about this algebraically: finding near-solutions to a function, which we then recontextualized as minimizing a function $x \mapsto \|A\hat{\mathbf{x}} - \mathbf{b}\|$. The obvious way to approach this is with calculus: we're minimizing a function, so take some derivatives, set them equal to zero, find the critical points, etc. But I went into math so I wouldn't have to do calculus, so I want to find a better option.

The better answer lies in thinking about this problem geometrically. We want to find the $\hat{\mathbf{x}}$ that makes $A\hat{\mathbf{x}}$ as close as possible to $\mathbf{b}$. But another way to look at that question is that the set of points $\{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$ forms a subspace of $\mathbb{R}^m$, and we want to find the one that's closest to $\mathbf{b}$.

And this is just the projection of $\mathbf{b}$ onto the subspace $A\mathbf{x}$. So if we take $\hat{\mathbf{b}} = \operatorname{Proj}_{\operatorname{col} A} \mathbf{b}$, then there is definitely a solution $\hat{\mathbf{x}}$ to $A\mathbf{x} = \hat{\mathbf{b}}$, and this solution is what we're looking for.

We can now solve our least-squares problem directly from this, if we want to. An orthogonal basis for span$\{(4,0,1),(0,2,1)\}$ is $\{(0,2,1),(4,-2/5,4/5)\}$, and we can project

$$\begin{aligned}
\operatorname{Proj}_{(0,2,1)}(2,0,11) &= \frac{11}{5}\begin{bmatrix}0\\2\\1\end{bmatrix}\\
\operatorname{Proj}_{(4,-2/5,4/5)}(2,0,11) &= \frac{84/5}{84/5}\begin{bmatrix}4\\-2/5\\4/5\end{bmatrix} = \begin{bmatrix}4\\-2/5\\4/5\end{bmatrix}\\
\operatorname{Proj}_{\operatorname{col} A}(2,0,11) &= \begin{bmatrix}4\\4\\3\end{bmatrix}
\end{aligned}$$

and then we just need to row-reduce

$$\left[\begin{array}{cc|c}4&0&4\\0&2&4\\1&1&3\end{array}\right] \to \left[\begin{array}{cc|c}1&0&1\\0&1&2\\0&0&0\end{array}\right]$$

and thus $\hat{\mathbf{x}} = (1,2)$.

But this is still way too complicated, especially if we have hundreds or thousands of variables.

But suppose we have a solution $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Since $\hat{\mathbf{b}}$ is the projection onto $\operatorname{col} A$, we know that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to every column of A. But since dotting with a column of A is the same as multiplying by a matrix with that row, this means that $A^T(\mathbf{b}-\hat{\mathbf{b}}) = \mathbf{0}$. Rearranging gives us

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Definition 7.4. The *normal equations* for the matrix equation $A\mathbf{x} = \mathbf{b}$ is the equation $A^T A\mathbf{x} = A^T\mathbf{b}$. We often write $\hat{\mathbf{x}}$ for a solution to the normal equations.

Proposition 7.5. *The set of least-squares solutions to $A\mathbf{x} = \mathbf{b}$ is exactly the set of solutions to the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.*

Proof. We already showed that every least-squares solution will be a solution to the normal equations. So we just need to show that a solution to the normal equations is a least-squares solution.

Suppose $\mathbf{x}$ is a solution to the normal equations. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, and thus $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\operatorname{col}(A)$. But we can write

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

as a sum of a vector in col A and a vector othogonal to col A. Since orthogonal compositions are unique, that means that $A\hat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto col A, and thus is the least-squares solution to $A\mathbf{x} = \mathbf{b}$. $\square$

Example 7.6. Let's finish off our example from earlier. We wanted to find the least-squares solution to

$$\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

The normal equations are

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

$$\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

This is a set of linear equations we can solve exactly, to get

$$\left[\begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & -84 & -168 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

This again gives us that $\hat{\mathbf{x}} = (1, 2)$.

We can further compute the *least-squares error* of this computation. We have that $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = (4, 4, 3)$, and the difference between the approximate solution and the true solution is $\mathbf{b} - \hat{\mathbf{b}} = (2, 0, 11) - (4, 4, 3) = (-2, -4, 8)$. Then the error is

$$\|\mathbf{b} - \hat{\mathbf{b}}\| = \sqrt{4 + 16 + 64} = \sqrt{84}.$$

Example 7.7. What is a least-squares solution to

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}?$$

We compute

$$A^TA = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 14 \\ -5 \end{bmatrix}$$

So we want to solve

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 14 \\ -5 \end{bmatrix}$$
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1/3 \\ 14/3 \\ -5/3 \end{bmatrix}$$

We can compute that $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = (5, 2, 3, 6)$, and $\mathbf{b} - \hat{\mathbf{b}} = (-3, 3, 3, 0)$. And indeed $A^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}$, so $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to $\operatorname{col} A$ and $\hat{\mathbf{b}}$ is in fact the projection of $\mathbf{b}$ onto $\operatorname{col} A$.

Finally, we can again compute that the error is

$$\|\mathbf{b} - \hat{\mathbf{b}}\| = \|(-3, 3, 3, 0)\| = \sqrt{27}.$$

When is the least-squares solution unique? Conceptually, that happens when A isn't throwing away any information and creating redundancies. This is equivalent to saying that the kernel of A is trivial, or that the columns of A are linearly independent. It turns out in this case that A^TA will always be invertible, and so the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is easy to solve.

Proposition 7.8. *The following statements are equivalent:*

1. *The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in$\mathbb{R}^m$;*

2. *The columns of A are linearly independent;*

3. The matrix A^TA is invertible.

In this case, the least-squares solution is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

7.2 Linear Regression

A particularly important application of least-squares methods is linear regression. We have some collection of input data X and some output data $\mathbf{y}$, and we want to find a linear relationship that allows us to predict the $\mathbf{y}$ from the X. Generally we want to find a parameter vector β that makes $X\beta = \mathbf{y}$ as accurate as possible—a least-squares solution to this equation.

We can start with simple one-dimensional linear regression. Given some collection of experimental data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we want to write a relationship $y = \beta_0 + \beta_1 x$ that predicts y as well as possible.

If these data all form a line, then there are some $\beta_0, beta_1$ such that $y_i = \beta_0 + \beta_1 x_i$ for each i. This is a collection of n linear equations, so we can write this as $X\beta = \mathbf{y}$ where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

But in normal experiments, we don't expect this equation to have an exact solution, because there will be noise—either because of measurement error, because the input data X doesn't fully determine the output data y. So what we really want is the least-squares solution to this equation. This minimizes the difference between the predicted output $\beta_0 + \beta_1 x_i$ and the actual output y_i.

(In fact it minimizes the sum of the squared error terms. If you want to minimize the sum of the error terms you use the L^1 norm, but that doesn't have a nice interpretation in terms of dot products and so is harder to compute.)

Example 7.9. Find the least-squares line that best fits the points $(2, 3), (3, 2), (5, 1), (6, 0)$.

This is almost just a line of slope -1, but it skips a bit between 3 and 5.

We have

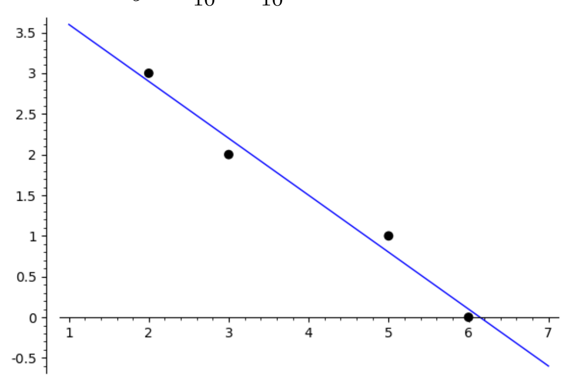
$$X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 16 \\ 16 & 74 \end{bmatrix}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 17 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 4 & 16 & 6 \\ 16 & 74 & 17 \end{array}\right] \to \left[\begin{array}{cc|c} 10 & 0 & 43 \\ 0 & 10 & -7 \end{array}\right]$$

Thus we have a line of best fit $y = \frac{43}{10} - \frac{7}{10}x$.



We can use the same technique when we're fitting a more complicated model. Although we're doing linear algebra, we don't need our *model* to be linear; we just need a linear combination of predictive components.

Example 7.10. A model of a child's systolic blood pressure p, measured in millimeters of mercury, as a function of weight w measured in pounds, is $\beta_0 + \beta_1 \ln(w) = p$. Suppose we have the following series of measurements:

p	91	98	103	110	112
w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88

Let's find the least-squares fit for β_0 and β_1.

Our model is that $\beta_0 \cdot 1 + \beta_1 \cdot \ln w = p$, so we want to take our matrix X with columns 1 and $\ln w$. Thus we have

$$X = \begin{bmatrix} 1 & 3.78 \\ 1 & 4.11 \\ 1 & 4.39 \\ 1 & 4.73 \\ 1 & 4.88 \end{bmatrix} \qquad Y = \begin{bmatrix} 91 \\ 98 \\ 103 \\ 110 \\ 112 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3.78 & 4.11 & 4.39 & 4.73 & 4.88 \end{bmatrix} \begin{bmatrix} 1 & 3.78 \\ 1 & 4.11 \\ 1 & 4.39 \\ 1 & 4.73 \\ 1 & 4.88 \end{bmatrix} = \begin{bmatrix} 5.00 & 21.89 \\ 21.89 & 96.64 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 514 \\ 2265.79 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 5 & 21.89 & 514 \\ 21.89 & 96.64 & 2265.79 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 18.57 \\ 0 & 1 & 19.24 \end{array}\right]$$

so we get the model $p \approx 18.57 + 19.24 \ln(w)$.

7.3 Symmetric Matrices

In the previous section, we often engaged with matrices of the form $X^T X$. These matrices have a special property: they are *symmetric*. Formally, we say a matrix A is symmetric if $A^T = A$; informally, the entries above the diagonal have to match the entries below the diagonal. (The entries on the diagonal don't matter here.)

Symmetric matrices have a number of important properties involving diagonalization.

Example 7.11. Let's look back at the matrix $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$. Its characteristic polynomial is $\chi_A(\lambda) = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$, and the eigenvectors are

$$\mathbf{v}_8 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{v}_6 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

But here we get an extra useful property: this basis is orthogonal!

Let's look at the diagonalization matrix now. We can take

$$U = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \qquad U^{-1} = \frac{1}{6}\begin{bmatrix} -3 & 3 & 0 \\ -1 & -1 & 2 \\ 2 & 2 & 2 \end{bmatrix},$$

and we get $U^{-1}AU = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

But the eigenbasis we got isn't the only possible one. We got an orthogonal set of eigenvectors; what if we normalize to get an orthonormal set?

Our eigenbasis becomes

$$\mathbf{e}_8 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \qquad \mathbf{e}_6 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} \qquad \mathbf{e}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix},$$

and we get diagonalization matrices

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

And what do we notice? $P^{-1} = P^T$! This is itself a general rule.

Proposition 7.12. *1. If A is a symmetric matrix, it has an orthonormal basis of eigenvectors.*

2. If P is a matrix whose columns form an orthonormal set, then $P^{-1} = P^T$.

Proof. 1. We'll prove that eigenvectors are orthogonal. We won't prove that they always form a basis, but they do.

Supppose $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors with eigenvalues λ_1, λ_2. Then we use the fact that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2$.

$$\begin{aligned} \mathbf{v}_1^T A\mathbf{v}_2 &= \mathbf{v}_1^T \lambda_2 \mathbf{v}_2 = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2) \\ \mathbf{v}_1^T A\mathbf{v}_2 &= \mathbf{v}_1^T A^T \mathbf{v}_2 && \text{because } A = A^T \\ &= (A\mathbf{v}_1)^T \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = \lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2. \end{aligned}$$

So we have that $\lambda_1 \neq \lambda_2$, but $\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$. This ipmlies that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, and thus $\mathbf{v}_1 \perp \mathbf{v}_2$.

2. Suppose $P = [\mathbf{c}_1 \dots \mathbf{c}_n]$ is a matrix whose columns are orthonormal vectors. We claim that $P^T P = I$. But

$$P^T P = \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 & \dots & \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^T\mathbf{c}_1 & \mathbf{c}_1^T\mathbf{c}_2 & \dots & \mathbf{c}_1^T\mathbf{c}_n \\ \mathbf{c}_2^T\mathbf{c}_1 & \mathbf{c}_2^T\mathbf{c}_2 & \dots & \mathbf{c}_2^T\mathbf{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_n^T\mathbf{c}_1 & \mathbf{c}_n^T\mathbf{c}_2 & \dots & \mathbf{c}_n^T\mathbf{c}_n \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1\cdot\mathbf{c}_1 & \mathbf{c}_1\cdot\mathbf{c}_2 & \dots & \mathbf{c}_1\cdot\mathbf{c}_n \\ \mathbf{c}_2\cdot\mathbf{c}_1 & \mathbf{c}_2\cdot\mathbf{c}_2 & \dots & \mathbf{c}_2\cdot\mathbf{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_n\cdot\mathbf{c}_1 & \mathbf{c}_n\cdot\mathbf{c}_2 & \dots & \mathbf{c}_n\cdot\mathbf{c}_n \end{bmatrix}.$$

But we know that the $\mathbf{c}_i$ are orthonormal, so $\mathbf{c}_i \cdot \mathbf{c}_i = 1$ and $\mathbf{c}_i \cdot \mathbf{c}_j = 0$ if $i \neq j$, so this is just the identity.

□

This orthogonal diagonalization is useful for a lot of things, but probably the most important and interesting application is principal component analysis.

7.4 Principal Component Analysis

Sometimes we have a large dataset, with hundreds or thousands or millions of dimensions, and we can't possibly understand the whole thing. We want a tool to pick out a few dimensions that we can focus down on.

My favorite example of this is in politics. We would like to compare, say, the votes of various senators to each other in order to compare their records. But this is a lot of information! Each senator in the 115th congress took about six hundred votes, which means that if you just want to record all the votes you have sixty thousand separate numbers, in a 100 by 600 matrix. You could, theoretically, print this out and look at it, but you wouldn't learn very much.

Instead, we want to compres this down into one or two numbers that tell us important information, like which senators are more or less liberal. This process involves a few steps.

We can start by assigning a vector to each senator, recording their votes. so you might have a vector for Mitch McConnell that looks like $[YYYNY]^T$ and one for Chuck Schumer that looks like $[NNYYN]^T$. Then we have 100 distinct vectors in $\mathbb{R}^{600}$, one for each senator. We might code all the Ys with 1s and all the Ns with 0s so that we can do math to this more easily. (The actual coding scheme that gets used is a bit more complex than that.)

In order to make things simpler, we want to *normalize* our vectors, so that they are zero "on average". So we compute a vector mean

$$\mathbf{M} = \frac{1}{100}(\mathbf{X}_1 + \cdots + \mathbf{X}_{100})$$

and compute new vectors $\hat{\mathbf{X}}_{\mathbf{i}} = \mathbf{X}_i - \mathbf{M}$. Each $\hat{\mathbf{X}}_{\mathbf{i}}$ reflects how different $\mathbf{X}_i$ is from the average value.

We can build this collection of vectors into a matrix by having one column for each senator. So our matrix will look like $B = [\hat{\mathbf{X}}_{\mathbf{1}} \hat{\mathbf{X}}_{\mathbf{2}} \dots \hat{\mathbf{X}}_{\mathbf{100}}]$.

Now we want to compute something called the *covariance matrix*. Given a set of data, the *variance* tells us how much variability it has: if all the values are close to their average, we get a small variance, but if some are very large and some are very negative, we get a large variance. If we want to compare two different data sets, we can compute the *covariance*, which is a measure of the extent to which one set has large values at the same time that the other does.

In linear algebra terms, two vectors have large covariance if they produce large values in the same coordinates—and thus if they point roughly in the same direciton. So the covariance between vectors $\mathbf{v}$ and $\mathbf{w}$ is $\mathbf{v} \cdot \mathbf{w}$.

In this case, we want to find the covariance between different *votes*. If the Senate holds a vote on a boring procedural matter and everyone votes "yes", that doesn't tell us very much. But if there's a lot of dispersion, where some people vote yes and other people vote no, that vote should tell us a lot about the ideological alignment of the senators. So we want to compute the dot product with each *row* of B with each row. (And then we want to normalize it for boring statistical reasons.)

When we put this all together, we get the covariance matrix $\frac{1}{n-1} BB^T$. The matrix BB^T is just a 600×600 matrix that tells us, in the ij entry, how similar the pattern of voting on the ith vote is to the pattern of voting on the jth vote. The diagonal $i = j$ entries tell us the variance, which is how much dispersion there is on each issue. The *total variance* is the trace of $\frac{1}{n-1} BB^T$, which is just the sume of the variances.

The off-diagonal entries tell us the covariance how similar two different issues are to each other. If the covariance is zero, then the two variables are called *uncorrelated*.

Example 7.13. Suppose we have a minature "senate" of five people who vote on three different measures:

$$\mathbf{X}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{X}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{X}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{X}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{X}_5 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

We can compute the mean-deviation vectors

$$\hat{\mathbf{X}}_\mathbf{1} = \begin{bmatrix} -3/5 \\ -3/5 \\ -3/5 \end{bmatrix}, \quad \hat{\mathbf{X}}_\mathbf{2} = \begin{bmatrix} 2/5 \\ 2/5 \\ -3/5 \end{bmatrix}, \quad \hat{\mathbf{X}}_\mathbf{3} = \begin{bmatrix} 2/5 \\ -3/5 \\ 2/5 \end{bmatrix}, \quad \hat{\mathbf{X}}_\mathbf{4} = \begin{bmatrix} 2/5 \\ 2/5 \\ 2/5 \end{bmatrix}, \quad \hat{\mathbf{X}}_\mathbf{5} = \begin{bmatrix} -3/5 \\ 2/5 \\ 2/5 \end{bmatrix}.$$

Then we have

$$B = \frac{1}{5} \begin{bmatrix} -3 & 2 & 2 & 2 & -3 \\ -3 & 2 & -3 & 2 & 2 \\ -3 & -3 & 2 & 2 & 2 \end{bmatrix}$$

$$B^T = \frac{1}{5} \begin{bmatrix} -3 & -3 & -3 \\ 2 & 2 & -3 \\ 2 & -3 & 2 \\ 2 & 2 & 2 \\ -3 & 2 & 2 \end{bmatrix}$$

$$S = \frac{1}{2} BB^T = \frac{1}{50} \begin{bmatrix} -3 & 2 & 2 & 2 & -3 \\ -3 & 2 & -3 & 2 & 2 \\ -3 & -3 & 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ 2 & 2 & -3 \\ 2 & -3 & 2 \\ 2 & 2 & 2 \\ -3 & 2 & 2 \end{bmatrix} = \frac{1}{50} \begin{bmatrix} 30 & 5 & 5 \\ 5 & 30 & 5 \\ 5 & 5 & 30 \end{bmatrix}.$$

This matrix S has eigenvectors $\mathbf{u}_1 = (1, 1, 1)$ with eigenvalue $4/5$, $\mathbf{u}_2 = (1, 0, -1)$ with eigenvalue $1/2$, and $\mathbf{u}_3(0, 1, -1)$ also with eigenvalue $1/2$. These eigenvectors form an orthogonal basis for $\mathbb{R}^3$, the space of votes.

So what does this tell us? The greatest amount of variation happens in the "direction" of $(1, 1, 1)$. (This direction explains about $4/9$ of the total variance). If we project each of our vote vectors onto this vector, we get

$$\mathbf{x}_1 = -9/15\mathbf{u}_1 \qquad \mathbf{x}_2 = 1/15\mathbf{u}_1 \qquad \mathbf{x}_3 = 1/15\mathbf{u}_1 \qquad \mathbf{x}_4 = 6/15\mathbf{u}_1 \qquad \mathbf{x}_5 = 1/15\mathbf{u}_1$$

and these points are fairly spread out. (If you wanted to describe this in words, you could say this measures the extent to which someone "wants to vote yes": $\mathbf{X}_1$ is very negative and $\mathbf{X}_4$ is very positive.) If we project onto one of the other two vectors, things will be less spread out. For instance, if we project onto $\mathbf{u}_2$ we get

$$\mathbf{x}_1 = 0\mathbf{u}_2 \qquad \mathbf{x}_2 = 1/2\mathbf{u}_2 \qquad \mathbf{x}_3 = 0\mathbf{u}_2 \qquad \mathbf{x}_4 = 0\mathbf{u}_2 \qquad \mathbf{x}_5 = -1/2\mathbf{u}_2.$$

In a political context, we would expect the "most important" eigenvector to reflect the left-right split.

Example 7.14. We can also use this technique to model things other than voting patterns. Suppose we have a group of people and we take three measurements for each of them:

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad \mathbf{X}_2 = \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix}, \qquad \mathbf{X}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix}, \qquad \mathbf{X}_4 = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix}.$$

We can compute the average is $(5, 4, 5)$, so the mean-deviation form vectors are

$$\hat{\mathbf{X}}_{\mathbf{1}} = \begin{bmatrix} -4 \\ -2 \\ -4 \end{bmatrix}, \qquad \hat{\mathbf{X}}_{\mathbf{2}} = \begin{bmatrix} -1 \\ -2 \\ 8 \end{bmatrix}, \qquad \hat{\mathbf{X}}_{\mathbf{3}} = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix}, \qquad \hat{\mathbf{X}}_{\mathbf{4}} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}.$$

We get

$$B = \begin{bmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{bmatrix}$$

$$S = \frac{1}{3} B^T B = \begin{bmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{bmatrix}.$$

This matrix has eigenvalues that are approximately $34.55, 13.84, 1.61$. Clearly the most important axis of variation is that first eigenvector, which is roughly $(1, 4.09, -12.83)$.

What's the point of this? The main one is that it lets us compress information for representation. If we return to our example of senators, we could imagine each senator as a point in 600-dimensional space, but we can't actually plot that. But we *can* plot a two-dimensional projection of their votes. This technique lets us pick out the two most important dimensions so we can get a handle on what's going on.

It can also tell us how many dimensions are actually important. In most of American history, the two principal components have both been pretty important; in the sixties, for instance, the first component was roughly economically left-to-right, while the second reflected positions on civil rights issues. In the mid nineties the second principal component became much less important relative to the first; this reflects or just describes the increasing polarization, where the left-right axis is almost the only thing that matters.

One other important note: we often try to describe what the principal components "really mean", but this is all done after the fact; the math identifies the most important component, and then we can try to describe it in words. Often for analyzing American politics, for instance, we'll look at the first principal component and see that it lines up with, roughly, the left-right spectrum. But we're not feeding any definition of what counts as "liberal" into the algorithm; we're just getting the output and putting a name on it. And sometimes we'll get out axes that we strugle to describe in any useful way.

Finally, what do the outputs of this look like in practice? [Will insert pictures here later]

7.5 Quadratic Forms and Singular Value Decomposition