Linear Algebra

Ben Woodruff
Revised by Jason Grout

Compiled December 6, 2010

See `http://bitbucket.org/jasongrout/linear-algebra` for the source for the book.

Contents

Chapter 1

Linear Algebra Arithmetic

This chapter covers the following ideas.

1. Be able to use and understand matrix and vector notation, addition, scalar multiplication, the dot product, matrix multiplication, and matrix transposition.

2. Use Gaussian elimination to solve systems of linear equations and write the solution in parametric and vector form. Define and use the words homogeneous, nonhomogeneous, row echelon form, and reduced row echelon form.

3. Find the rank of a matrix. Determine if a collection of vectors is linearly independent. If linearly dependent, be able to write vectors as linear combinations of the preceding vectors.

4. For square matrices, compute determinants, inverses, eigenvalues, and eigenvectors.

5. Illustrate with examples how a nonzero determinant is equivalent to having independent columns, an inverse, and nonzero eigenvalues. Similarly, a zero determinant is equivalent to having dependent columns, no inverse, and a zero eigenvalue.

The next unit will focus on applications of these ideas. The main goal of this unit is to familiarize yourself with the arithmetic involved in linear algebra.

1.1 Basic Notation

Most of linear algebra centers around understanding matrices and vectors. The first chapter of this text contains a brief introduction to the arithmetic involved with matrices and vectors. The second chapter will show you many of the uses of the ideas we are learning. You will be given motivation for all of the ideas learned here, as well as real world applications of these ideas, before the end of the second chapter. For now, work on becoming familiar with the arithmetic of linear algebra so that we can easily discuss the ideas behind linear algebra throughout the course.

1.1.1 Matrices

A *matrix* of size "m by n" ($m \times n$) is a table with m rows and n columns. We   Matrix size is
row by column.

1

normally give matrices names using capital letters, and use the notation

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [a_{jk}],$$

where a_{jk} is the entry in the jth row, kth column.

We say two matrices A and B are equal if corresponding entries are equal (i.e., $a_{jk} = b_{jk}$ for all j and k). We add and subtract matrices of the same size entry-wise. A *scalar* is a single number or variable. We multiply a scalar and a matrix cA by multiplying every entry in the matrix by the scalar c. If the number of rows and columns are equal, then we say the matrix is square. The main *diagonal* of a square $(n \times n)$ matrix consists of the entries $a_{11}, a_{22}, \ldots, a_{nn}$, and the *trace* is the sum of the entries on the main diagonal $(\sum a_{jj})$.

Example 1.1. If $A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -1 \\ 0 & 4 \end{bmatrix}$, then

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

The main diagonal of B consists of the entries $b_{11} = 3$ and $b_{22} = 4$, which means the trace of B is $3 + 4 = 7$ (the sum of the entries on the main diagonal).

The *transpose* of a matrix $A = [a_{jk}]$ is a new matrix $A^T = [a_{kj}]$ formed by interchanging the rows and columns (the first row of A is the first column of B; first column of A is the first row of B, etc.)   transpose

Example 1.2. The transpose of A is illustrated below.

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

1.1.2   Vectors

Vectors represent a magnitude in a given direction. We can use vectors to model forces, acceleration, velocity, probabilities, electronic data, and more. We can use matrices to represent vectors. A row vector is a $1 \times n$ matrix. A column vector is an $m \times 1$ matrix. Textbooks often write vectors using bold face font. By hand (and in this book), we write an arrow above them. The notation $\mathbf{v} = \vec{v} = \langle v_1, v_2, v_3 \rangle$ can represent either a row or column vector. Many different ways to represent vectors are used in different books. For example, you may see the vector $\langle -2, 3 \rangle$ written in any of the following ways.

$$\langle -2, 3 \rangle = -2\mathbf{i} + 3\mathbf{j} = (-2, 3) = \begin{bmatrix} -2 & 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} = (-2 \quad 3) = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$$

In this chapter, we will switch freely between these notations so that you get used to reading different notations. In practice, it is a good idea to not switch notations as often. Also, notice that the notation $(-2, 3)$ (with parentheses) has other meanings as well (like a point in the plane, or an open interval). Please make sure that when you use the notation $(-2, 3)$ to represent a vector, it is clear from the context that you working with a vector.

To draw a vector, pick a starting point (the tail). Then draw a line that
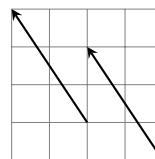
Both vectors represent $\langle -2, 3 \rangle$, regardless of where they start.

moves v_1 units horizontally and v_2 units vertically, and put an arrow at the ending point (the head). Many times we will put the tail of the vector at the origin $(0, 0)$.

The entry-wise methods of adding, subtracting, and multiplying by a scalar that we learned for matrices apply to vectors as well (this is consistent, since a vector can be thought of as a row or column matrix). For example,

$$\langle 1, 3 \rangle - 2\langle -1, 2 \rangle + \langle 4, 0 \rangle = \langle 1 - 2(-1) + 4, 3 - 2(2) + 0 \rangle = \langle 7, -1 \rangle.$$

Sometimes it is easier to see the arithmetic if we write the vectors as single-column matrices:

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 2\begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - 2(-1) + 4 \\ 3 - 2(2) + 0 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}.$$

Vector addition is performed geometrically by placing the tail of the second vector at the head of the first. The resultant vector is the vector which starts at the tail of the first and ends at the head of the second. This is called the parallelogram law of addition. The sum $\vec{u} + \vec{v} = \langle 1, 2 \rangle + \langle 3, 1 \rangle = \langle 4, 3 \rangle$ is illustrated in the margin.

Scalar multiplication $c\vec{u}$ (c is a scalar and $\vec{u}$ is a vector) is equivalent to stretching (scaling) a vector by the scalar without rotating the vector. The product $2\vec{u}$ doubles the length of the vector. If the scalar is negative, then the vector reverses to point in the opposite direction. The product $-\frac{1}{2}\vec{u}$ is a vector half as long as $\vec{u}$ and points in the opposite direction (as illustrated in the margin).

To subtract to vectors $\vec{u} - \vec{v}$, we use scalar multiplication and vector addition. We write $\vec{u} - \vec{v} = \vec{u} + (-\vec{v})$, which means we reverse $\vec{v}$ and then add it to $\vec{u}$. You can visualize this using the parallelogram law of addition. You can also visualize the difference $\vec{u} - \vec{v}$ as the vector connecting the heads of the two vectors when their tails are both placed on the same spot, with the direction pointing towards the head of $\vec{u}$. When you see a vector difference $\vec{u} - \vec{v}$, you might imagine an arrowhead on the subtraction sign like this: $\vec{u} \leftarrow \vec{v}$, or think "head minus tail" to remind you which way the difference vector points. The difference $\vec{u} - \vec{v} = \langle 1, 2 \rangle - \langle 3, 1 \rangle = \langle -2, 1 \rangle$ is illustrated in two vectors in the margin.

Vector Addition

Scalar Multiplication

Vector Subtraction
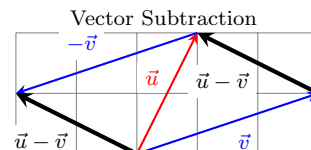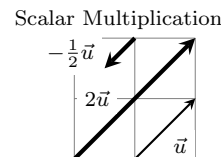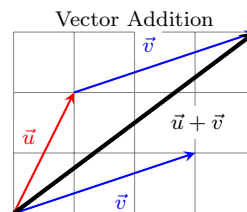
1.1.3 Magnitude and the Dot Product

The length of a vector is found using the Pythagorean theorem: $c^2 = a^2 + b^2$. The *magnitude* (or length) of the vector $\langle 2, 3 \rangle$ is simply the length of the hypotenuse of a right triangle with short side lengths 2 and 3, hence $|\langle 2, 3 \rangle| = \sqrt{2^2 + 3^2} = \sqrt{13}$. We denote magnitude with absolute value symbols, and compute for $\vec{u} = \langle u_1, u_2 \rangle$ the magnitude as $|\vec{u}| = \sqrt{u_1^2 + u_2^2}$. In higher dimensions we extend this as

$$|\vec{u}| = \sqrt{u_1^2 + u_2^2 + u_3^2 + \cdots u_n^2} = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

A unit vector is a vector with length 1. In many books, unit vectors are written with a hat above them, as $\hat{\mathbf{u}}$. Any vector that is not a unit vector can be rewritten as a scalar times a unit vector by dividing the vector by its magnitude. This allows us to write any vector as a magnitude times a direction (where the direction is given by a unit vector).

A unit vector $\hat{\mathbf{u}}$ has length $|\vec{u}| = 1$

Example 1.3. The length of $\vec{u} = \langle 2, 1, -3 \rangle$ is $\sqrt{2^2 + 1^2 + (-3)^2} = \sqrt{4 + 1 + 9} = \sqrt{14}$. A unit vector in the direction of $\vec{u}$ is $\frac{1}{\sqrt{14}} \langle 2, 1, -3 \rangle = \left\langle \frac{2}{\sqrt{14}}, \frac{1}{\sqrt{14}}, \frac{-3}{\sqrt{14}} \right\rangle$.

We can rewrite $\vec{u}$ as a magnitude times a direction by writing

Every vector can be rewritten as a magnitude times direction (unit vector).

$$\vec{u} = |\vec{u}| \frac{\vec{u}}{|\vec{u}|} = \sqrt{14} \left\langle \frac{2}{\sqrt{14}}, \frac{1}{\sqrt{14}}, \frac{-3}{\sqrt{14}} \right\rangle.$$

The *dot product* of two vectors $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is the scalar given by multiplying corresponding components and adding the products together (i.e., $\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum u_i v_i$).

Example 1.4. The dot product of $\begin{bmatrix} 1 & 3 & -2 \end{bmatrix}$ and $\begin{bmatrix} 2 & -1 & 4 \end{bmatrix}$ is

$$\begin{bmatrix} 1 & 3 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 & 4 \end{bmatrix} = (1)(2) + (3)(-1) + (-2)(4) = -9.$$

We use the dot product to find lengths and angles in all dimensions. We will also use it multiply matrices. The rest of this section explains how to use the dot product to find lengths and angles, and is included for completeness. We will revisit the dot product more in later chapters. Notice that if $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$, then we can find the length of a vector using the dot product since

The dot product finds length:
$$\vec{u} \cdot \vec{u} = |\vec{u}|^2$$

$$\vec{u} \cdot \vec{u} = u_1^2 + u_2^2 + u_3^2 + \cdots u_n^2 = \sum_{i=1}^{n} u_i^2 = |\vec{u}|^2.$$

If θ is the angle between two vectors $\vec{u}$ and $\vec{v}$, then we can find the angle between these two vectors using the formula

The dot product finds angles:
$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$$

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta.$$

This follows from the law of cosines $c^2 = a^2 + b^2 - 2ab \cos \theta$, where $a = |\vec{u}|$, $b = |\vec{v}|$, and $c = |\vec{u} - \vec{v}|$. Because the dot product is equal to the square of the magnitude, we have

$$(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}| \cos \theta.$$

Distributing on the left gives

Law of Cosines
$$c^2 = a^2 + b^2 - 2ab \cos \theta$$

$$\vec{u} \cdot \vec{u} - 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}| \cos \theta.$$

Now cancel the common terms on both sides and divide by 2 to obtain our formula $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$. This formula geometrically explains how to connect the dot product to angles in both 2 and 3 dimensions. In higher dimensions, we can define the angle between two vectors using this formula.

$c = |\vec{u} - \vec{v}|$

$a = |\vec{u}|$

$b = |\vec{v}|$

θ

If two vectors meet at a right angle, then since $\cos 90° = 0$, we see that $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}|0 = 0$, so the dot product is zero. Similarly, if the dot product is zero, then $0 = \vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$, so either one of the vectors has zero length, or the angle between them is zero. We say that $\vec{u}$ and $\vec{v}$ are *orthogonal* if $\vec{u} \cdot \vec{v} = 0$. Two lines are perpendicular when the angle between then is 90°. Two vectors are orthogonal when either the angle between them is 90°, or one of the vectors is the zero vector.
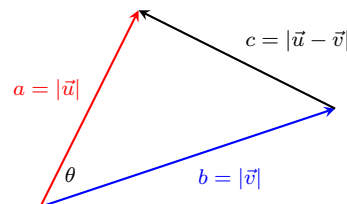
orthogonal = dot product is zero

1.2 Multiplying Matrices

1.2.1 Linear Combinations

The simplest vectors in 2D are (written in several different forms): $\hat{\imath} = \vec{\imath} =$

The standard basis vectors in 3D:
$i = \vec{\imath} = \langle 1, 0, 0 \rangle = (1, 0, 0)$
$j = \vec{\jmath} = \langle 0, 1, 0 \rangle = (0, 1, 0)$
$k = \vec{k} = \langle 0, 0, 1 \rangle = (0, 0, 1)$

$\langle 1,0 \rangle = (1,0)$ and $\hat{\jmath} = \vec{\jmath} = \langle 0,1 \rangle = (0,1)$. We call these the standard basis vectors in 2D. In 3D, we include the vector $\hat{\mathbf{k}} = \vec{k} = \langle 0,0,1 \rangle$ as well as add a zero to both $\vec{\imath}$ and $\vec{\jmath}$ to obtain the standard basis vectors.

A similar idea is used in higher dimensions. The word *basis* suggests that we can base other vectors on these basis vectors, and we typically write other vectors in terms of these standard basis vectors. Using only scalar multiplication and vector addition, we can obtain the other vectors in 2D from the standard basis vectors. For example we can write

$$\langle a,b \rangle = a \langle 1,0 \rangle + b \langle 0,1 \rangle = a\vec{\imath} + b\vec{\jmath}.$$

The entries of a row or column vector are called the coordinates of the vector relative to the standard basis, or just the components of the vector. Notice that we scaled the vectors $\vec{\imath}$ and $\vec{\jmath}$, and then summed the scaled vectors. This is an example of a linear combination.

A *linear combination* of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is an expression of the form $c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$, where each c_i is a constant. A linear combination of vectors is simply a sum of scalar multiples of the vectors. We start with some vectors, stretch each one by some scalar, and then sum the result. Much of what we will do this semester (and in many courses to come) relates directly to understanding linear combinations.

Example 1.5. Let's look at a physical application of linear combinations that you use every day when you drive. When you place your foot on the accelerator of your car, the car exerts a force $\vec{F}_1$ pushing you forward. When you turn the steering wheel left, the car exerts a force $\vec{F}_2$ pushing you left. The total force acting on you is the vector sum of these forces, or a linear combination $\vec{F}_{\text{total}} = \vec{F}_1 + \vec{F}_2$. If $\vec{T}$ represents a unit vector pointing forward, and $\vec{N}$ represents a unit vector pointing left, then we can write the total force as the linear combination

$$\vec{F}_{\text{total}} = |\vec{F}_1|\vec{T} + |\vec{F}_2|\vec{N}.$$

To understand motion, we can now focus on finding just the magnitude of each force and the tangential unit vector $\vec{T}$ and normal unit vector $\vec{N}$. Once we understand these quantities, we can understand any force because it is a linear combination of quantities we understand.

1.2.2 Matrix Multiplication

One of the key applications of linear combinations we will use throughout the semester is matrix multiplication. Let's introduce the idea with an example.

Example 1.6. Consider the three vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$. Let's multiply the first vector by 2, the second by -1, and the third by 4, and then sum the result. This gives us the linear combination

$$2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - 1 \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 9 \end{bmatrix}$$

We define matrix multiplication so that multiplying a matrix and a vector (with the vector on the right) corresponds precisely to creating a linear combination of the columns of A. We now write the linear combination above in matrix form

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 9 \end{bmatrix}.$$

We define the matrix product $A\vec{x}$ (a matrix times a vector) to be the linear combination of columns of A where the components of $\vec{x}$ are the scalars in the linear combination. For this to make sense, notice that the vector $\vec{x}$ must have the same number of entries as there are columns in A. Symbolically, let $\vec{a}_i$ be the ith column of A, so that $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$, and let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Then

Matrix multiplication $A\vec{x}$

the matrix product is the linear combination

The product $A\vec{x}$ gives us linear combinations of the columns of A.

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \vec{a}_1 x_1 + \vec{a}_2 x_2 + \cdots + \vec{a}_n x_n.$$

This should look like the dot product. If you think of A as a vector of vectors, then $A\vec{x}$ is just the dot product of A and $\vec{x}$.

We can now define the product of two matrices A and B. Let $\vec{b}_j$ represent the jth column of B (so $B = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{bmatrix}$). The product AB of two matrices $A_{m \times n}$ and $B_{n \times p}$ is a new matrix $C_{m \times p} = [c_{ij}]$ where the jth column of C is the product $A\vec{b}_j$. To summarize, the matrix product AB is a new matrix whose jth column is a linear combinations of the columns of A using the entries of the jth column of B to perform the linear combinations. In math notation,

Matrix multiplication AB

$$AB = A \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \end{bmatrix}$$

Example 1.7. In this example, the matrix A has 3 rows, so the product should have 3 rows. The matrix B has 2 columns, so we need to find 2 different linear combinations giving us 2 columns in the product.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} (-2) + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} (1) \right) & \left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} (0) + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} (3) \right) \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ -2 & 12 \\ -4 & 18 \end{bmatrix}$$

Example 1.8. A row matrix times a column matrix is equivalent to the dot product of the two when treated as vectors.

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 6 \end{bmatrix} = \begin{bmatrix} 17 \end{bmatrix}$$

Example 1.9. Note that AB and BA are usually different.

$$A = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 4 \end{bmatrix} \quad AB = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix} \quad BA = \begin{bmatrix} 11 \end{bmatrix}$$

Even if the product CD is defined, the product DC does not have to be. In this example D has 3 columns, but C has only 2 rows, and hence the product DC is not defined (as you can't form a linear combination of 3 vectors with only 2 constants).

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix} \quad CD = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 3 & 5 \end{bmatrix} \quad DC \text{ is undefined}$$

The *identity matrix* is a square matrix which has only 1's along the diagonal, and zeros everywhere else. We often use I_n to mean the n by n identity matrix. The identity matrix is like the number 1, in that $AI = A$ and $IA = A$ for any matrix A of appropriate size. If A is a 2 by 3 matrix, then $AI_3 = A$ and $I_2 A = A$ (notice that the size of the identity matrix changes based on the order of multiplication). If A is a square matrix, then $AI = IA = A$.

The identity matrix behaves like the number 1, in that $AI = IA = A$.

Linear combinations of columns of A with scalars from columns of B.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} = \left[\left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} (-2) + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} (1) \right) \quad \left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} (0) + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} (3) \right) \right] = \begin{bmatrix} 0 & 6 \\ -2 & 12 \\ -4 & 18 \end{bmatrix}$$

Rows of A dotted by columns of B.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} & \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 5 & 6 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} & \begin{bmatrix} 5 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ -2 & 12 \\ -4 & 18 \end{bmatrix}$$

Use rows of A to form linear combinations of rows of B.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 \begin{bmatrix} -2 & 0 \end{bmatrix} + (2) \begin{bmatrix} 1 & 3 \end{bmatrix} \\ 3 \begin{bmatrix} -2 & 0 \end{bmatrix} + (4) \begin{bmatrix} 1 & 3 \end{bmatrix} \\ 5 \begin{bmatrix} -2 & 0 \end{bmatrix} + (6) \begin{bmatrix} 1 & 3 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ -2 & 12 \\ -4 & 18 \end{bmatrix}$$

Table 1.1: Three ways to view matrix multiplication.

1.2.3   Alternate Definitions of Matrix Multiplication

Understanding matrix multiplication as linear combinations of vectors is very valuable (and we hope it is your first thought when you see matrix multiplication!) There are many alternate ways to think of matrix multiplication. Here are two additional methods. Table 1.1 illustrates all three.

Row dot column. The product AB of two matrices $A_{m \times n}$ and $B_{n \times p}$ is a new matrix $C_{m \times p} = [c_{ij}]$ where $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ is the dot product of the of the ith row of A and the jth column of B. Wikipedia has an excellent visual illustration of this approach.

Linear combinations of rows of B. The matrix product $\vec{x}B$ (notice the vector is now on the left) is a linear combination of the rows of B using the components of $\vec{x}$ as the scalars. For the product AB, let $\vec{a}_i$ represent the ith row of A. Then the ith row of AB is the product $\vec{a}_i B$, which is a linear combination of rows of B.

We'll use the "linear combinations of columns" view we first introduced because we eventually will want to think of a matrix A as a function, and it is natural to write $A(\vec{x})$ (with the vector on the right) to represent plugging the vector $\vec{x}$ into this function.

1.3   Linear Systems of Equations

A *linear system of equations*, such as

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

x_1, x_2, x_3 system	linear combination vector equation
$\begin{aligned} 2x_1 + x_2 - x_3 &= 2 \\ x_1 - 2x_2 &= 3 \\ 4x_2 + 2x_3 &= 1 \end{aligned}$	$x_1 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$
matrix equation	augmented matrix
$\begin{bmatrix} 2 & 1 & -1 \\ 1 & -2 & 0 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$	$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 2 \\ 1 & -2 & 0 & 3 \\ 0 & 4 & 2 & 1 \end{array} \right]$

Table 1.2: Four equivalent ways to represent a system of equations.

is a system of equations where each variable in the system appears in its own term and is multiplied by a constant called a coefficient (which may be zero or one). Rather than using x, y, z as variables, we'll often use x_1, x_2, x_3, which allows us to easily extend to systems in higher dimensions. If we let $\vec{v}_i$ be a vector whose components are the constants next to x_i, then every linear system can be written in terms of linear combinations, i.e. $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{b}$. Similarly, we can write a linear system in terms of matrix multiplication $A\vec{x} = \vec{b}$, where the columns of A are the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$. The matrix A is called the *coefficient matrix* of the linear system of equations. Adding the column vector $\vec{b}$ coefficient matrix to the right of A gives what we call an *augmented matrix* (where often a dashed augmented matrix or solid line is placed in the matrix to remind us of the equal sign). These four ways of representing a system are illustrated in Table 1.2

Let's add some geometry to the discussion. A linear equation in two variables $ax+by = c$ graphically represents a line. A system of equations with two variables represents multiple lines in the plane. A solution to such a system is a point that lies on all of the lines (i.e., an intersection of all of the lines). We can have one of the following situations:

1. No solution: no point lies on all of the lines (even if some of the lines intersect).

2. One solution: only one point lies on all the lines (i.e., all the lines intersect in one point).

3. Infinitely many solutions: the lines are actually all the same line, so all points on one line are really on all the lines.

A linear equation in 3 variables $ax + by + cz = d$ represents a plane in space. In our 3D system above, a solution graphically represents the intersection of three planes. Three planes can intersect in any of the following ways:

1. No solution: two of the planes are parallel and never intersect, or the planes only intersect in pairs.

2. One solution: the three planes intersect in a single point.

3. Infinitely many solutions:

    (a) the three planes all intersect in a line, so there is a line of solutions.

    (b) the three equations all represent the same plane, so there is a plane of solutions.

In 2D and 3D, we can see geometrically that there will always be either no solution, one solution, or infinitely many solutions. With any system of linear equations, regardless of the number of equations and number of unknowns, this pattern continues. There will always be either no solution, a single solution, or infinitely many solutions. Systems with solutions are called *consistent*, whereas systems without solutions are called inconsistent. A particularly important kind of linear system is of the form $A\vec{x} = \vec{0}$ (i.e., $\vec{b} = \vec{0}$)—this kind of system is called a *homogeneous* system. If $A\vec{x} \neq \vec{0}$ (i.e., $\vec{b} \neq \vec{0}$), we say the system is nonhomogeneous. A homogeneous system is always consistent since $\vec{x} = \vec{0}$ is always a solution. This solution is called the *trivial solution*, and indicates that the graph of every equation passes through the origin. The big question for homogeneous systems is whether they have nontrivial (nonzero) solutions.

A system of equations will have either a unique solution, infinitely many solutions, or no solution.

consistent system: has a solution

homogeneous system: $A\vec{x} = \vec{0}$

1.3.1 Gaussian Elimination

Gaussian elimination is an efficient algorithm we will use to solve systems of linear equations. This algorithm (or a variation) is commonly used in practice. The main idea is to eliminate each variable from all but one equation/row (if possible), using the following three operations. These operations are called *elementary row operations*.

1. Multiply an equation (or row of a matrix) by a nonzero constant,

2. Add a nonzero multiple of any equation (or row) to another equation,

3. Interchange two equations (or rows).

These three operations are the operations learned in college algebra when solving a system using a method of elimination. Gaussian elimination is a way of organizing your work so that you always arrive at a solution (or can determine there is no solution) in a certain number of steps (that's part of the reason it is a good computer algorithm). The process involves a forward reduction and (optionally) a backward reduction. The forward reduction creates zeros in the lower left corner of the matrix. The backward reduction puts zeros in the upper right corner of the matrix. We eliminate the values (i.e., make them zero) in the lower left corner of the matrix, starting with column 1, then column 2, and proceed column by column until all values which can be eliminated have been eliminated. Before formally stating the algorithm, let's look at a few examples.

Example 1.10. Let's start with a system of 2 equations and 2 unknowns. The augmented matrix representing the system is shown as well. To solve

$$\begin{array}{rl} x_1 - 3x_2 & = 4 \\ 2x_1 - 5x_2 & = 1 \end{array} \qquad \left[\begin{array}{cc|c} 1 & -3 & 4 \\ 2 & -5 & 1 \end{array}\right]$$

we eliminate the $2x_1$ in the 2nd row by adding -2 times the first row to the second row.

$$\begin{array}{rl} x_1 - 3x_2 & = 4 \\ x_2 & = -7 \end{array} \qquad \left[\begin{array}{cc|c} 1 & -3 & 4 \\ 0 & 1 & -7 \end{array}\right]$$

Now the augmented matrix is said to be in *row echelon form*. This means that

row echelon form

- in each nonzero row, the first nonzero entry is a 1 (called a "leading 1"),

- the leading 1 in a row occurs further right than the leading 1 in the row above, and

- any rows of all zeros appear at the bottom.

The position in the matrix where the leading 1 occurs is called a pivot. The column containing a pivot is called a *pivot column*. At this point we can use "back-substitution" to get $x_2 = -7$ and $x_1 = 4 + 3x_2 = 4 - 21 = -17$. Alternatively, we can continue the elimination process by eliminating the terms above each pivot, starting with the rightmost (and lowest) pivot and working towards the left and top. This will result in a matrix where the only nonzero entry in any pivot column is the pivot entry (i.e., the leading 1). In our example, if we add 3 times the second row to the first row, we obtain

$$
\begin{array}{ll}
x_1 \;\;\; = -17 \\
x_2 \;\;\; = -7
\end{array}
\qquad
\left[\begin{array}{cc|c}
1 & 0 & -17 \\
0 & 1 & -7
\end{array}\right]
$$

The matrix on the right is said to be in *reduced row echelon form* (or just "rref"). A matrix is in reduced row echelon form if

- the matrix is in echelon form, and

- the only nonzero entry in each pivot column is the pivot entry (i.e., the leading 1).

You can easily read solutions to systems of equations directly from a matrix which is in reduced row echelon form.

We can interpret the solution $x_1 = -17$, $x_2 = -7$ in multiple ways. It is the point $(-17, -7)$ where the two lines $x - 3y = 4$ and $2x - 5y = 1$ intersect. We also know that the only way to obtain a solution to the vector equation $c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -3 \\ -5 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is to let $c_1 = -17$ and $c_2 = -7$, which shows us that $(4, 1)$ is a linear combination of the vectors $(1, 2)$ and $(-3, -5)$. In addition, the only solution to the matrix equation $\begin{bmatrix} 1 & -3 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -17 \\ -7 \end{bmatrix}$. Notice that solving this system of equations tells us information about graphical intersections, linear combination of vectors, and matrix equations.

Example 1.11. Let's now solve a nonhomogeneous (meaning the right side is not zero) system with 3 equations and 3 unknowns:

$$
\begin{array}{rl}
2x_1 + x_2 - x_3 &= 2 \\
x_1 - 2x_2 &= 3 \\
4x_2 + 2x_3 &= 1
\end{array}
\qquad
\left[\begin{array}{ccc|c}
2 & 1 & -1 & 2 \\
1 & -2 & 0 & 3 \\
0 & 4 & 2 & 1
\end{array}\right].
$$

We'll encounter some homogeneous systems later on. To simplify the writing, we'll just use matrices this time. To keep track of each step, we will write the row operation next to the row we will replace. Remember that the three operations are (1) multiply a row by a nonzero constant, (2) add a multiple of one row to another, and (3) interchange any two rows. If we write $R_2 + 3R1$ next to R_2, then this means we will add 3 times row 1 to row 2. If we write $2R_2 - R1$ next to R_2, then we have done two row operations, namely we multiplied R_2 by 2, and then added (-1) times $R1$ to the result (replacing $R2$ with the sum). The steps below read left to right, top to bottom. In order to avoid fractions in the middle of the computation, we wait to divide each pivot entry to make it a 1

until the last step.

$$\Rightarrow^{(1)} \begin{bmatrix} 2 & 1 & -1 & 2 \\ 1 & -2 & 0 & 3 \\ 0 & 4 & 2 & 1 \end{bmatrix} \begin{matrix} \\ 2R_2 - R_1 \\ \end{matrix} \Rightarrow^{(2)} \begin{bmatrix} 2 & 1 & -1 & 2 \\ 0 & -5 & 1 & 4 \\ 0 & 4 & 2 & 1 \end{bmatrix} \begin{matrix} \\ \\ 5R_3 + 4R_2 \end{matrix}$$

$$\Rightarrow^{(3)} \begin{bmatrix} 2 & 1 & -1 & 2 \\ 0 & -5 & 1 & 4 \\ 0 & 0 & 14 & 21 \end{bmatrix} \begin{matrix} \\ \\ R_3/7 \end{matrix} \Rightarrow^{(4)} \begin{bmatrix} 2 & 1 & -1 & 2 \\ 0 & -10 & 2 & 8 \\ 0 & 0 & 2 & 3 \end{bmatrix} \begin{matrix} 2R_1 + R_3 \\ R_2 - R_3 \\ \end{matrix}$$

$$\Rightarrow^{(5)} \begin{bmatrix} 4 & 2 & 0 & 7 \\ 0 & -10 & 0 & 5 \\ 0 & 0 & 2 & 3 \end{bmatrix} \begin{matrix} \\ R_2/5 \\ \end{matrix} \Rightarrow^{(6)} \begin{bmatrix} 4 & 2 & 0 & 7 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 2 & 3 \end{bmatrix} \begin{matrix} R_1 + R_2 \\ \\ \end{matrix}$$

$$\Rightarrow^{(7)} \begin{bmatrix} 4 & 0 & 0 & 8 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 2 & 3 \end{bmatrix} \begin{matrix} R_1/4 \\ R_2/-2 \\ R_3/2 \end{matrix} \Rightarrow^{(8)} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1/2 \\ 0 & 0 & 1 & 3/2 \end{bmatrix}$$

Writing the final matrix in terms of a system of linear equations, we have the solution $x_1 = 2$, $x_2 = -1/2$, $x_3 = 3/2$. Remember that this tells us (1) where three planes intersect, (2) how to write the 4th column $\vec{b}$ in our original augmented matrix as a linear combination of the columns of the coefficient matrix A, and (3) how to solve the matrix equation $A\vec{x} = \vec{b}$ for $\vec{x}$.

The following steps describe the Gaussian elimination algorithm that we used above. Please take a moment to compare what is written below with the example above. Most of the problems in this unit can be solved using Gaussian elimination, so we will practice it as we learn a few new ideas.

1. Forward Phase (getting row echelon form): the following 4 steps should be repeated until you have mentally erased all the rows or all the columns. In step 1 or 4, you will erase a column and/or row from the matrix.

   (a) Consider the first column of your matrix. Start by interchanging rows (if needed) to place a nonzero entry in the first row. If all the elements in the first column are zero, then ignore that column in future computations (mentally erase the column) and begin again with the smaller matrix which is missing this column. If you erase the last column, then stop.

   Computer algorithms place the largest (in absolute value) nonzero entry in the first row. This reduces potential errors due to rounding that can occur in later steps.

   (b) Divide the first row (of your possibly smaller matrix) by its leading entry so that you have a leading 1. This entry is a pivot, and the column is a pivot column. (When doing this by hand, it is often convenient to skip this step and do it at the very end so that you avoid fractional arithmetic. If you can find a common multiple of all the terms in this row, then divide by it to reduce the size of your computations.)

   (c) Use the pivot to eliminate each nonzero entry below the pivot by adding a multiple of the row with the pivot to the nonzero lower row.

   (d) Ignore the row and column containing your new pivot and return to the first step (mentally cover up or erase the row and column containing your pivot). If you erase the last row, then stop.

   Ignoring rows and columns is equivalent to incrementing row and column counters in a computer program.

2. Backward Phase (getting reduced row echelon form): at this point, each row should have a leading 1 (the pivots), and you should have all zeros to the left and below each leading 1. If you skipped step 2 above, then

the pivot entries may not actually be 1, but at the end of this phase you should divide each row by its leading coefficient to make each row have a leading 1.

(a) Starting with the last pivot column. Use the pivot in that column to eliminate all the nonzero entries above it by adding multiples of the row containing the pivot to the nonzero rows above.

(b) Work from right to left, using each pivot in turn to eliminate the nonzero entries above it. Nothing to the left of the current pivot column will change in an elimination step. By working right to left, you methodically work through the matrix and greatly reduce the number of computations needed to fully reduce the matrix.

Example 1.12. As a final example, let's reduce $\begin{bmatrix} 0 & 1 & 1 & -2 & 7 \\ 1 & 3 & 5 & 1 & 6 \\ 2 & 0 & 4 & 3 & -8 \\ -2 & 1 & -3 & 0 & 5 \end{bmatrix}$ to reduced row echelon form (rref). The first step involves swapping two rows. We swap row 1 and row 2 because this places a 1 as the leading entry in row 1.

(1) Get a nonzero entry in upper left

$$\Rightarrow \begin{bmatrix} 0 & 1 & 1 & -2 & 7 \\ 1 & 3 & 5 & 1 & 6 \\ 2 & 0 & 4 & 3 & -8 \\ -2 & 1 & -3 & 0 & 5 \end{bmatrix} \quad R_1 \leftrightarrow R_2$$

(2) Eliminate entries in 1st column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 2 & 0 & 4 & 3 & -8 \\ -2 & 1 & -3 & 0 & 5 \end{bmatrix} \begin{matrix} \\ \\ R_3 - 2R_1 \\ R_4 + 2R_1 \end{matrix}$$

(3) Eliminate entries in 2nd column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 0 & -6 & -6 & 1 & -20 \\ 0 & 7 & 7 & 2 & 17 \end{bmatrix} \begin{matrix} \\ \\ R_3 + 6R_2 \\ R_4 - 7R_2 \end{matrix}$$

(4) Make a leading 1 in 4th column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 0 & 0 & 0 & -11 & 22 \\ 0 & 0 & 0 & 16 & -32 \end{bmatrix} \begin{matrix} \\ \\ R_3/(-11) \\ R_4/16 \end{matrix}$$

(5) Eliminate entries in 4th column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix} \begin{matrix} \\ \\ \\ R_4 - R_3 \end{matrix}$$

(6) Row Echelon Form

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

At this stage, we have found a row echelon form of the matrix. Notice that we eliminated nonzero terms in the lower left of the matrix by starting with the first column and working our way to the right column by column. Columns 1, 2, and 4 are the pivot columns of this matrix. We now use the pivots to eliminate the other nonzero entries in each pivot column (working right to left).

Recall that a matrix is in reduced row echelon (rref) if:

1. Nonzero rows begin with a leading 1.

2. Leadings 1's on subsequent rows appear further right than previous rows.

3. Rows of zeros are at the bottom.

4. Zeros are above and below each pivot.

(7) Eliminate entries in 4th column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 1 & 6 \\ 0 & 1 & 1 & -2 & 7 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} R_1 - R3 \\ R_2 + 2R_3 \\ \\ \end{matrix}$$

(8) Eliminate entries in 2nd column

$$\Rightarrow \begin{bmatrix} 1 & 3 & 5 & 0 & 8 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad R_1 - 3R_2$$

(9) Reduced Row Echelon Form

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

(10) Switch to system form

$$\Rightarrow \begin{matrix} x_1 + 2x_3 & = -1 \\ x_2 + x_3 & = 3 \\ x_4 & = -2 \\ 0 & = 0 \end{matrix}$$

We have obtained the reduced row echelon form. When we write this matrix in the corresponding system form, notice that there is not a unique solution to the system. Because the third column did not contain a pivot column, we can

write every variable in terms of x_3 (the redundant equation $x_3 = x_3$ allows us to write x_3 in terms of x_3). We are free to pick any value we want for x_3 and still obtain a solution. For this reason, we call x_3 a *free variable*, and write our infinitely many solutions in terms of x_3 as

Free variables correspond to non pivot columns. Solutions can be written in terms of free variables.

$$
\begin{array}{ll}
x_1 = -1 - 2x_3 & x_1 = -1 - 2t \\
x_2 = 3 - x_3 & x_2 = 3 - t \\
x_3 = x_3 \quad \text{or by letting } x_3 = t, & x_3 = t \\
x_4 = -2 & x_4 = -2
\end{array}
\quad .
$$

By choosing a value (such as t) for x_3, we can write our solution in what we call "parametric form." We have now given a parametrization of the solution set, where t is an arbitrary real number. We can also write the solution in vector form: $\vec{x} = \langle x_1, x_2, x_3, x_4 \rangle = \langle -1 - 2t, 3 - t, t, -2 \rangle$, where $-\infty < t < \infty$.

parametric form

1.3.2 Reading Reduced Row Echelon Form ("rref")

From reduced row echelon form, you can read the solution to a system immediately from the augmented matrix. Here are some typical examples of what you will see when you reduce an augmented matrix which represents a system that does not have a unique solution. The explanations which follow illustrate how to see the solution immediately from the matrix.

$$
\left[\begin{array}{ccc|c}
1 & 2 & 0 & 1 \\
0 & 0 & 1 & 3 \\
0 & 0 & 0 & 0
\end{array}\right]
\qquad
\left[\begin{array}{ccc|c}
1 & 0 & 4 & -5 \\
0 & 1 & -2 & 3
\end{array}\right]
\qquad
\left[\begin{array}{ccc|c}
1 & 0 & 2 & 0 \\
0 & 1 & -3 & 0 \\
0 & 0 & 0 & 1
\end{array}\right]
$$
$$
\vec{x} = \langle 1 - 2x_2, x_2, 3 \rangle \quad \vec{x} = \langle -5 - 4x_3, 3 + 2x_3, x_3 \rangle \quad \text{no solution since } 0 \neq 1
$$

Example 1.13. In the first example, columns 1 and 3 are the pivot columns. Since column 2 is not a pivot column, we'll let x_2 be a free variable and write the solution in terms of x_2. Rewriting the first matrix in system form yields $x_1 + 2x_2 = 1, x_3 = 3$, or solving for each variable in terms of x_2, we have $x_1 = 1 - x_2, x_3 = 3$. Adding to this system the redundant equation $x_2 = x_2$, we have in vector form $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 - 2x_2 \\ x_2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$. Sometimes it's useful to use a different variable, such as $x_2 = t$, and then write the solution in parametric form as $\langle x_1, x_2, x_3 \rangle = \langle 1 - 2t, t, 3 \rangle$, where t is any real number. This solution is a line in space.

Example 1.14. In the second example, the first and second columns are pivot columns, so x_3 is a free variable. The solution can be written in the form $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5 - 4x_3 \\ 3 + 2x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix}$. Notice that the third and fourth columns of the matrix appear as parts of the vectors in the solution.

Example 1.15. In the third example, the first, second, and fourth columns are pivot columns. Because the last column is a pivot column, we obtain the equation $0 = 1$, which is absurd. Hence the system is inconsistent and has no solution.

Example 1.16. The ideas above generalize to higher dimensions. Here are two large rref matrices and their solutions.

$$
\begin{bmatrix}
0 & 1 & 0 & 2 & 0 & 0 \\
0 & 0 & 1 & 3 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 4 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\qquad
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 4 \end{bmatrix}
+ x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
+ x_4 \begin{bmatrix} 0 \\ -2 \\ -3 \\ 1 \\ 0 \end{bmatrix}
$$

$$
\begin{bmatrix}
0 & 1 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 1 & 3 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\quad
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
+ x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
+ x_4 \begin{bmatrix} 0 \\ -2 \\ -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}
+ x_6 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ -4 \\ 1 \end{bmatrix}
$$

Each non-pivot column corresponds to one of the vectors in the sum above.

1.4   Rank and Linear Independence

The *rank* of a matrix is the number of pivot columns of the matrix. To find the rank of a matrix, reduce the matrix using Gaussian elimination until you discover the pivot columns (you only need to do the forward phase for this).

Recall that a linear combination of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is an expression of the form $c_1 \vec{v}_1 + c_2 \vec{v}_2 + \ldots + c_n \vec{v}_n$, where c_i is a constant scalar for each i. A linear combination of vectors is a sum of scalar multiples of the vectors. Also recall that a linear combination of vectors is really just matrix multiplication, $A\vec{c}$, where the columns of A are the vectors $\vec{v}_i$ and $\vec{c} = \langle c_1, c_2, \ldots, c_n \rangle$. We will try to emphasis this in what follows.

First we need another definition. The *span* of a set of vectors is the set of all possible linear combinations of the vectors. We often write this as $\mathrm{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. In terms of matrices, if A is the matrix with columns $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, then the span of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is the set of all possible vectors $\vec{b}$ such that $A\vec{x} = \vec{b}$ for some $\vec{x}$.

Example 1.17.   1. The span of the vector $\langle 1, 0, 0 \rangle$ is the set of vectors of the form $a \langle 1, 0, 0 \rangle = \langle a, 0, 0 \rangle$, which is the x-axis, a line through the origin.

2. The span of the vectors $\langle 1, 0, 0 \rangle$ and $\langle 0, 1, 0 \rangle$ is the set of vectors in 3D of the form $a \langle 1, 0, 0 \rangle + b \langle 0, 1, 0 \rangle = \langle a, b, 0 \rangle$, which is the xy-plane, a plane through the origin.

3. The span of $\langle 2, 0, 1 \rangle$ and $\langle -1, -2, 3 \rangle$ is the set of vectors

$$
a\langle 2, 0, 1 \rangle + b \langle -1, -2, 3 \rangle = \langle 2a - b, -2b, a + 3b \rangle =
\begin{bmatrix} 2 & -1 \\ 0 & -2 \\ 1 & 3 \end{bmatrix}
\begin{bmatrix} a \\ b \end{bmatrix}
=
\begin{bmatrix} 2a - b \\ -2b \\ a + 3b \end{bmatrix}.
$$

This is again a plane in 3D that passes through the origin.

Geometrically you can obtain the span of vectors by adding together all possible stretches of the given vectors. The span of a set of vectors will be a line, a plane, or some higher-dimensional version of these objects (called a hyperplane) which passes through the origin.

We say that a set of vectors is *linearly independent* if the only solution to the homogeneous system $c_1 \vec{v}_1 + c_2 \vec{v}_2 + \ldots + c_n \vec{v}_n = \vec{0}$ is the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. Otherwise we say the set of vectors is linearly dependent, since one of the vectors depends on (can be obtained as a linear combination of) the others. In terms of spans, we say a set of vectors is linearly dependent when one of the vectors is in the span of the other vectors.

linearly independent

Example 1.18. The set $\{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$ is linearly independent, whereas the set $\{\langle 2, 4 \rangle, \langle 3, 6 \rangle\}$ is linearly dependent (notice that $\langle 3, 6 \rangle = \frac{3}{2} \langle 2, 4 \rangle$ is a linear combination of $\langle 2, 4 \rangle$). The set $\{\langle 1, 2, 0 \rangle, \langle 2, 0, 3 \rangle, \langle 3, -2, 6 \rangle\}$ is linearly dependent because there is a nonzero linear combination of the three vectors that yields the zero vector:

$$-1\langle 1, 2, 0 \rangle + 2\langle 2, 0, 3 \rangle - 1\langle 3, -2, 6 \rangle = \langle 0, 0, 0 \rangle.$$

Alternatively, we know the vectors are dependent because we can write the third vector as a linear combination of the first two:

$$-1\langle 1, 2, 0 \rangle + 2\langle 2, 0, 3 \rangle = \langle 3, -2, 6 \rangle.$$

To find this linear combination, we write the first linear combination as a homogeneous system (in matrix form) $A\vec{c} = \vec{0}$:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 & | & 0 \\ 2 & 0 & -2 & | & 0 \\ 0 & 3 & 6 & | & 0 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

When solving a homogeneous system of equations, the column of zeros at the right contributes no new information, and so is often omitted.

The reduced row echelon form tells us that column 3 is not a pivot column. This is precisely why the third column can be written as a linear combination of the first two, and the numbers in column 3 tell us precisely what coefficients $(-1$ and $2)$ to use writing the third column as a linear combination of the pivot columns.

An easy way to test if vectors are linearly independent is to create a matrix A where each column represents one of the vectors. The vectors are linearly independent if the only solution to $A\vec{c} = \vec{0}$ is $\vec{c} = 0$. Row reduce the matrix to reduced row echelon form. The vectors are linearly independent if and only if each column of A is a pivot column (which forces $\vec{c}$ to be zero). If a column is not a pivot column, then the vector corresponding to that column can be written as a linear combination of the preceding vectors using the coefficients in that column.

Example 1.19. The vectors $[1\ 3\ 5]$, $[-1\ 0\ 1]$, and $[0\ 3\ 1]$ are linearly independent, as the reduced row echelon form of $\begin{bmatrix} 1 & -1 & 0 \\ 3 & 0 & 3 \\ 5 & 1 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and each column is a pivot column.

Example 1.20. The vectors $[1\ 3\ 5]$, $[-1\ 0\ 1]$, and $[1\ 6\ 11]$ are linearly dependent, as the reduced row echelon form of $\begin{bmatrix} 1 & -1 & 1 \\ 3 & 0 & 6 \\ 5 & 1 & 11 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, and column 3 is not a pivot column. This immediately means that the third vector (of the original matrix) is a linear combination of the preceding two. Using the coefficients 2 and 1 from the third column, we can write $[1\ 6\ 11] = 2[1\ 3\ 5] + 1[-1\ 0\ 1]$.

In the preceding example, the coefficients in the linear combination (2 and 1) are called the *coordinates* of $[1\ 6\ 11]$ relative to the vectors $[1\ 3\ 5]$ and $[-1\ 0\ 1]$. *coordinates* We could write down the linear combination because the third column of the matrix, $[1\ 6\ 11]$, was not a pivot column. Notice that we could have read the coordinates right out of the third column of the rref matrix. Every non-pivot column of a matrix can be expressed as a linear combination of the pivot columns, where the coordinates of the non-pivot column come from the same non-pivot column in the rref matrix.

We have used the concept of *coordinates of a vector relative to other vectors* to state that the coordinates of $(2, -3, 4)$ relative to $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ are $2, -3,$ and 4. We will study this concept of coordinates in greater depth later.

1.4.1 Linear Combinations, Spans, and RREF

Solving a system of equations such as $x + 3y = 0$, $2x - y = 5$ is equivalent to solving the vector equation $x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$. This is also equivalent to the following questions.

- Is $\begin{bmatrix} 0 \\ 5 \end{bmatrix}$ a linear combination of the vectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$?

- Is $\begin{bmatrix} 0 \\ 5 \end{bmatrix}$ in the span of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$?

To answer all these questions, first reduce the matrix $\left[\begin{array}{cc|c} 1 & 3 & 0 \\ 2 & -1 & 5 \end{array}\right]$ to reduced row echelon form.

$$\left[\begin{array}{cc|c} 1 & 3 & 0 \\ 2 & -1 & 5 \end{array}\right] \quad R_2 - 2R_1 \quad \Rightarrow \left[\begin{array}{cc|c} 1 & 3 & 0 \\ 0 & -7 & 5 \end{array}\right] \quad 3R_2 + 7R_1$$

$$\Rightarrow \left[\begin{array}{cc|c} 7 & 0 & 15 \\ 0 & -7 & 5 \end{array}\right] \quad \begin{array}{c} R_1/7 \\ R_2/-7 \end{array} \quad \Rightarrow \left[\begin{array}{cc|c} 1 & 0 & 15/7 \\ 0 & 1 & -5/7 \end{array}\right].$$

This means that a solution to our system is $x = 15/7, y = -5/7$, and we can write $\begin{bmatrix} 0 \\ 5 \end{bmatrix} = \frac{15}{7} \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{5}{7} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$, which means that the answer to both questions is "Yes." The example above generalizes to show that a system $A\vec{x} = \vec{b}$ has a solution for $\vec{x}$ if and only if $\vec{b}$ is a linear combination of the columns of A, if and only if $\vec{b}$ is in the span of the columns of A.

The following are equivalent (all are true or all are false):

1. $A\vec{x} = \vec{b}$ has a solution.
2. $\vec{b}$ is a linear combination of the columns of A.
3. $\vec{b}$ is in the span of the columns of A.

1.5 Determinants

For the rest of this unit, we will assume that the matrix A is a square matrix (in other words we are solving a system where the number of equations and number of unknowns are the same). Associated with every square matrix is a number, called the *determinant*, which is related to length, area, and volume. We also use the determinant to generalize volume to higher dimensions. Determinants are only defined for square matrices. The determinant of a 2×2 and 3×3 matrix can be computed as follows:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \det \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \det \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \det \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$$= a(ei - hf) - b(di - gf) + c(dh - ge)$$

We use absolute value signs (vertical bars) around a matrix to as notation for the determinant. Notice the negative sign on the middle term of the 3×3 determinant. Also, notice that we had to compute three determinants of 2 by 2 matrices in order to find the determinant of a 3 by 3.

The determinant in general is defined recursively in terms of minors (determinants of smaller matrices) and cofactors. This method is called the cofactor expansion of the determinant. The minor M_{ij} of a matrix A is the determinant of the the matrix formed by removing row i and column j from A. We define a cofactor to be $C_{ij} = (-1)^{i+j} M_{ij}$. To compute the determinant, first pick a

The determinant is a linear combination of the cofactors.

row or column. We define the determinant to be $\sum_{k=1}^{n} a_{ik} C_{ik}$ (if we chose row i) or alternatively $\sum_{k=1}^{n} a_{kj} C_{kj}$ (if we chose column j). You can pick any row or column you want, and then compute the determinant by multiplying each entry of that row or column by its cofactor, and then summing the results. A sign matrix keeps track of the $(-1)^{j+k}$ term in the cofactor. All you have to do is determine if the first entry of your expansion has a plus or minus, and then alternate the sign as you expand. We will eventually prove why you can pick any row or column. For now, just become comfortable with the arithmetic.

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

sign matrix
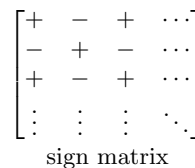
Example 1.21. Let's find the determinant of the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 3 & 4 \\ 2 & -3 & 1 \end{bmatrix}$

in two ways. First we will use a cofactor expansion along the top row (so the first term is positive).

$$\det A = 1 \begin{vmatrix} 3 & 4 \\ -3 & 1 \end{vmatrix} - 2 \begin{vmatrix} -1 & 4 \\ 2 & 1 \end{vmatrix} + 0 \begin{vmatrix} -1 & 3 \\ 2 & -3 \end{vmatrix}$$
$$= 1(3 + 12) - 2(-1 - 8) + 0(3 - 6) = 33$$

Now we'll use the second column (so the first term is negative).

$$|A| = -2 \begin{vmatrix} -1 & 4 \\ 2 & 1 \end{vmatrix} + 3 \begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} - (-3) \begin{vmatrix} 1 & 0 \\ -1 & 4 \end{vmatrix}$$
$$= -(2)(-1 - 8) + (3)(1 - 0) - (-3)(4 - 0) = 33$$

The number 33 is the volume of a three-dimensional parallelepiped created from the columns of A, as described in the next section.

1.5.1   Geometric Interpretation of the determinant

Consider the 2 by 2 matrix $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ whose determinant is $3 \cdot 2 - 0 \cdot 1 = 6$. Draw the column vectors $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ with their base at the origin.   These two vectors give the edges of a parallelogram whose area is the determinant (6). If we swap the order of the two vectors in the matrix, then the determinant of $\begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix}$ is -6. The absolute value of the determinant is still the area, but the sign of the determinant depends on the order of the vectors. Starting at the first vector, if you can turn counterclockwise through an angle smaller than 180° to obtain the second vector, then the determinant is positive. If you have to turn clockwise instead, then the determinant is negative. This is often termed "the right-hand rule," since rotating the fingers of your right hand from the first vector to the second vector will cause your thumb to point up precisely when the determinant is positive.

For a 3 by 3 matrix, the columns give the edges of a three-dimensional parallelepiped and the determinant produces the volume of this object. The sign of the determinant is related to orientation. If, using your right hand, you can place your index finger on the first vector, middle finger on the second vector, and thumb on the third vector, then the determinant is positive. If you can't, then the determinant is negative.

Example 1.22. Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Starting from the origin, each column represents an edge of the rectangular box $0 \le x \le 1$, $0 \le y \le 2$,

Area = 6

$\begin{vmatrix} 3 & 1 \\ 0 & 2 \end{vmatrix} = 6$ and $\begin{vmatrix} 1 & 3 \\ 2 & 0 \end{vmatrix} = -6$
The determinant gives area and direction.
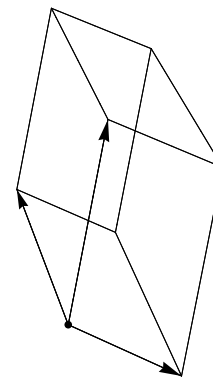
The determinant of a 3 by 3 matrix gives the volume of the parallelepiped created by using the columns of the matrix as the three parallel edges.

$0 \leq z \leq 3$ with volume (and determinant) $V = lwh = (1)(2)(3) = 6$. The sign of the determinant is positive because if you place your index finger pointing in the direction $(1, 0, 0)$ and your middle finger in the direction $(0, 2, 0)$, then your thumb points upwards in the direction $(0, 0, 3)$. If you interchange two of the columns, for example $B = \begin{bmatrix} 0 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, then the volume doesn't change since the shape is still the same. However, the sign of the determinant is negative because if you point your index finger in the direction $(0, 2, 0)$ and your middle finger in the direction $(1, 0, 0)$, then your thumb points down in the direction $(0, 0, -3)$. If you repeat this with your left hand instead of right hand, then your thumb points up.

1.5.2   Zero Determinants and Linear Dependence

Because the determinant helps us find area in 2D, we can use this idea to help us understand when two vectors in 2D are linearly dependent. If two vectors are dependent, then one is a linear combination of the other, hence a multiple of the other. This means that the parallelogram formed by the two vectors has no area (as the two vectors lie on the same line). So if the vectors are dependent, the determinant is zero. Similarly, if the determinant is zero, then the vectors must lie on the same line and hence are linearly dependent. In 3D, three vectors being linearly dependent implies that one of the vectors lies in a plane or line spanned by the other two. Any object in 3D that lies in a plane or line has no volume so the determinant is zero. Similarly, if the determinant of a 3 by 3 matrix is zero, then the column vectors must lie in the same plane and hence are linearly dependent. We now have a geometric interpretation for the key fact:

> The determinant of a matrix is zero if and only if the columns are linearly dependent.

The homeworks asks you to compute determinants of matrices as well as row reduce them so that you can verify this fact in various settings. Notice that the columns of a square matrix are linearly independent if and only if the reduced row echelon form of the matrix is the identity matrix. This shows us that the determinant of a square matrix is nonzero if and only if the reduced row echelon form of the matrix is the identity matrix.

For a square matrix A, the following are equivalent:

1. The determinant is zero.
2. The columns are linearly dependent.
3. The rref of A is not I.

1.6   The Matrix Inverse

Recall that the identity matrix I behaves in matrix multiplication like the number 1 behaves in regular multiplication. When we solve the equation $ax = b$ with numbers, we multiply both sides by a^{-1} to obtain $x = a^{-1}b = \frac{1}{a}b$. The multiplicative inverse of a is simply $1/a$, because $a\frac{1}{a} = \frac{1}{a}a = 1$. We have been studying linear systems of the form $A\vec{x} = \vec{b}$. It would be nice if we could just divide both sides by A, but there is no such thing as division by a matrix in general. If we look only at square matrices, then sometimes it is possible to find a matrix B such that $BA = AB = I$, the identity matrix. If such a matrix B exists, then multiplying both sides of $A\vec{x} = \vec{b}$ on the left by the matrix B yields $BA\vec{x} = B\vec{B}$, or $I\vec{x} = \vec{x} = B\vec{b}$. The matrix B is then called the *inverse* of A, and we write it as A^{-1}, the same symbol we used with regular multiplication. When an inverse exists, the solution to $A\vec{x} = \vec{b}$ is simply $\vec{x} = A^{-1}\vec{b}$.

The solution to $A\vec{x} = \vec{b}$ is $\vec{x} = A^{-1}\vec{b}$, if A^{-1} exists.

To find an inverse, we will start by considering a general 3 by 3 matrix. Once we are done, we will know how to find the inverse of any n by n matrix or deter-

mine it does not exists. If an inverse exists, then write $A^{-1} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$.

Then the equation $AA^{-1} = I$ requires that the 3 matrix equations

$$A \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad A \begin{bmatrix} c_{12} \\ c_{22} \\ c_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and } A \begin{bmatrix} c_{13} \\ c_{23} \\ c_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

each have a solution, or that $(1,0,0), (0,1,0)$, and $(0,0,1)$ are all linear combinations of the columns of A. This requires that we reduce all three augmented matrices

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 1 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \end{array}\right] \quad \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 1 \\ a_{31} & a_{32} & a_{33} & 0 \end{array}\right] \quad \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 1 \end{array}\right].$$

If the first three columns are all pivot columns, then the row operations required to reduce all three matrices will be identical, and none of the 4th columns will be pivot columns. Rather than do three equivalent row reductions, we can solve all three simultaneously by creating the single augmented matrix

$$\left[\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array}\right]$$

or in compact form $\begin{bmatrix} A & | & I \end{bmatrix}$. We now reduce this larger matrix to reduced row echelon form and obtain $\begin{bmatrix} I & | & B \end{bmatrix}$, which tells us the coordinates of $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ relative to the columns of A. The columns of the augmented portion B on the right are the solutions to the three original systems, hence are the columns of A^{-1}. To summarize, an inverse exists precisely if the augmented system $[A|I]$ reduces to $[I|A^{-1}]$. Then the inverse matrix appears on the right in the rref matrix if the identity matrix appears on the left. If the left block of the augmented matrix does not reduce to the identity matrix, then the matrix does not have an inverse. If the left block does not reduce to the identity, then this implies that the original columns of A are dependent. Later we will show that having an inverse is equivalent to having a nonzero determinant and having a rank which equals the number of columns. Remember that this is true for square matrices, or systems where we have the same number of equations as unknowns.

If $A_{n \times n}$ is a square matrix, then the following are equivalent:

1. A has an inverse A^{-1}.
2. The columns of A are linearly independent.
3. The rank of A is n
4. The rref of A is I.
5. $|A| \neq 0$.

The following are also equivalent:

1. A does not have an inverse
2. The columns of A are linearly dependent.
3. The rank is less than n
4. The rref of A is not I (there is a row of zeros along the bottom).
5. $|A| = 0$.

Example 1.23. To find the inverse of $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ we reduce $\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right]$. Reduction gives $\left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{array}\right]$. The inverse is the right block $\begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}$. You can always check your result by computing AA^{-1} to see if it is I, for example $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Using this inverse, the solution to the system $\begin{cases} 1x + 2y = 4 \\ 3x + 4y = 0 \end{cases}$ is $A^{-1}\vec{b} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} -8 \\ 6 \end{bmatrix}$, so $x = -8$ and $y = 6$.

Examples involving larger matrices are in the homework. Remember that you can solve systems by finding an inverse of the coefficient matrix.

Notice that if A has an inverse, then $A\vec{x} = \vec{b}$ always has exactly one solution, $\vec{x} = A^{-1}\vec{b}$. If a square matrix A does not have an inverse, then the system $A\vec{x} = \vec{b}$ can have no solutions or infinitely many solutions.

1.7   Eigenvalues and Eigenvectors

In this section, we also assume that we are dealing with square matrices. We will start by looking at an example to motivate the language we are about to introduce. Consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. When we multiply this matrix by the vector $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, we obtain $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\vec{x}$. Multiplication by the matrix A was miraculously the same as multiplying by the number 3. Symbolically, we have $A\vec{x} = 3\vec{x}$. Not every vector $\vec{x}$ satisfies this property, for example, multiplying by $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ gives $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, which is not a multiple of $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Our main goal in this section is to answer the following two questions:

1. For which nonzero vectors $\vec{x}$ is it possible to write $A\vec{x} = \lambda\vec{x}$ for some scalar λ? (We will call these vectors eigenvectors.)

2. Which scalars λ satisfy $A\vec{x} = \lambda\vec{x}$ for some vector $\vec{x}$? (We will call these scalars eigenvalues.)

Now for some definitions. Let A be a square $n \times n$ matrix. An *eigenvector* is a nonzero vector $\vec{x}$ such that $A\vec{x} = \lambda\vec{x}$ (matrix multiplication is the same as scalar multiplication) for some scalar λ, called the *eigenvalue* associated with the eigenvector $\vec{x}$. Zero vectors are not eigenvectors because $A\vec{0} = \lambda\vec{0}$ no matter what λ is. We can rewrite the equation $A\vec{x} = \lambda\vec{x}$ as $\vec{0} = A\vec{x} - \lambda\vec{x} = A\vec{x} - \lambda I\vec{x}$ and then factor to obtain

$$(A - \lambda I)\vec{x} = \vec{0}.$$

In other words, we need to find a scalar λ so that there are nonzero solutions to the homogeneous system of equations $(A - \lambda I)\vec{x} = \vec{0}$, or equivalently, we need to pick a λ so that the columns of $A - \lambda I$ are linearly dependent. Recall that if $A - \lambda I$ has an inverse, then $(A - \lambda I)\vec{x} = \vec{0}$ has only one solution, $\vec{x} = (A - \lambda I)^{-1}\vec{0} = \vec{0}$. So if $A - \lambda I$ has an inverse, $\vec{x}$ cannot be an eigenvector. With the correct λ, we know that $A - \lambda I$ has no inverse, so $\det(A - \lambda I) = 0$. This last equation is the key equation we will solve since it does not contain the vector $\vec{x}$ anymore. The expression $\det(A - \lambda I)$ is called the *characteristic polynomial* of A. It is a polynomial of degree n with variable λ, and hence there are at most n eigenvalues (which correspond to the roots of the polynomial). In summary, to find the eigenvalues and eigenvectors:

1. we solve $\det(A - \lambda I) = 0$ for λ to obtain the eigenvalues, and then

2. for each eigenvalue λ, we solve $(A - \lambda I)\vec{x} = \vec{0}$ to obtain the eigenvectors associated with λ.

As a way to check your work, the following two facts can help.

- The sum of the eigenvalues equals the trace of the matrix (the sum of the diagonal elements).

- The product of the eigenvalues equals the determinant.

Example 1.24. To find the eigenvalues of $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, first subtract λ from each diagonal entry to get $A - \lambda I = \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix}$, and then find the determinant

(margin notes) eigenvector · eigenvalue · characteristic polynomial · The trace and determinant are equal to the sum and product of the eigenvalues.

$(2 − \lambda)(2 − \lambda) − 1 = \lambda^2 − 4\lambda + 3 = (\lambda − 1)(\lambda − 3)$. The roots are 1 and 3, so the eigenvalues are $\lambda = 1, 3$. As a check, the trace of this matrix is $2 + 2 = 4$ and the sum of the eigenvalues is $1 + 3$. Also, the determinant is $4 − 1 = 3$ which equals the product of the eigenvalues.

When $\lambda = 1$, we compute $A − \lambda I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. We solve the equation $(A −$

$\lambda I)\vec{x} = 0$ by reducing the augmented matrix $\begin{bmatrix} 1 & 1 & | & 0 \\ 1 & 1 & | & 0 \end{bmatrix}$ to its rref $\begin{bmatrix} 1 & 1 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$.

Since x_2 is a free variable, we solve $x_1 + x_2 = 0$ for x_1 to obtain the equations $x_1 = −x_2$, $x_2 = x_2$, and then write our solution in vector form $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} −1 \\ 1 \end{bmatrix}$

for some $x_2 \neq 0$. There are infinitely many eigenvectors corresponding to $\lambda = 1$.

A particular eigenvector is $\begin{bmatrix} −1 \\ 1 \end{bmatrix}$ and all the rest are nonzero linear combinations of this one.

Since we are solving a homogeneous system $A\vec{x} = \vec{0}$, we could avoid writing the last column of zeros.

When $\lambda = 3$, we compute $A − \lambda I = \begin{bmatrix} −1 & 1 \\ 1 & −1 \end{bmatrix}$. Reducing $\begin{bmatrix} −1 & 1 & | & 0 \\ 1 & −1 & | & 0 \end{bmatrix}$

gives $\begin{bmatrix} 1 & −1 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$, which means the eigenvectors are of the form $x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for some

$x_2 \neq 0$. A particular eigenvector corresponding to $\lambda = 3$ is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and all others are nonzero linear combinations of this one.

Finding eigenvalues and eigenvectors requires that we compute determinants, find roots of polynomials, and then solve homogeneous systems of equations. You know you are doing the problem correctly if you get infinitely many solutions to the system $(A − \lambda I)\vec{x} = 0$ for each λ (i.e. there is at least one row of zeros along the bottom of the rref matrix).

Eigenvalues and eigenvectors show up in many different places in engineering, economics, computer science, physics, chemistry, mathematics, and many other fields. We will be exploring how eigenvectors appear throughout linear algebra as the semester progresses. For now we'll just focus on being able to find them.

Example 1.25. Let's find the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 4 \\ 0 & 0 & 1 \end{bmatrix}$. Subtracting λ from the diagonals gives $A − \lambda I = \begin{bmatrix} 2 − \lambda & 1 & 4 \\ 1 & 2 − \lambda & 4 \\ 0 & 0 & 1 − \lambda \end{bmatrix}$.

The determinant of this matrix is easiest to compute if you expand along the third row to get $0 − 0 + (1 − \lambda) \begin{vmatrix} 2 − \lambda & 1 \\ 1 & 2 − \lambda \end{vmatrix}$. Computing the 2 by 2 determinant gives $(1 − \lambda)((2 − \lambda)(2 − \lambda) − 1) = (1 − \lambda)(\lambda − 1)(\lambda − 3)$, so the eigenvalues are $1, 1, 3$ (1 is a repeated root). As a check, their sum is $1 + 1 + 3 = 5$ and the trace of A is $2 + 2 + 1 = 5$ as well. For larger matrices, determinants are not always easy to check, so just checking the trace is a fast way to make sure you are on the right path.

When $\lambda = 1$, we compute $A − \lambda I = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$. Reducing the system

$(A − \lambda I)\vec{x} = 0$ yields $\begin{bmatrix} 1 & 1 & 4 & | & 0 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$, which has two free variables. The solution

is $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} −1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} −4 \\ 0 \\ 1 \end{bmatrix}$. Hence the eigenvectors are nonzero linear

combinations of $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}$. Notice that all the eigenvectors can be obtained as linear combinations of two linearly independent eigenvectors, in part because $\lambda = 1$ is a double root of the characteristic polynomial.

When $\lambda = 3$, we compute $A - \lambda I = \begin{bmatrix} -1 & 1 & 4 \\ 1 & -1 & 4 \\ 0 & 0 & -2 \end{bmatrix}$. Next, reducing

$\begin{bmatrix} -1 & 1 & 4 & 0 \\ 1 & -1 & 4 & 0 \\ 0 & 0 & -2 & 0 \end{bmatrix}$ yields $\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, which has only one free variable .

The solution is $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. The eigenvectors are nonzero linear combina-

tions of $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Because $\lambda = 3$ is a single root of the characteristic polynomial, all

the eigenvectors can be obtained as linear combinations of one eigenvector.

1.8 Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Be able to use and understand matrix and vector notation, addition, scalar multiplication, the dot product, matrix multiplication, and matrix transposition.

2. Use Gaussian elimination to solve systems of linear equations and write the solution in parametric and vector form. Define and use the words homogeneous, nonhomogeneous, row echelon form, and reduced row echelon form.

3. Find the rank of a matrix. Determine if a collection of vectors is linearly independent. If linearly dependent, be able to write vectors as linear combinations of the preceding vectors.

4. For square matrices, compute determinants, inverses, eigenvalues, and eigenvectors.

5. Illustrate with examples how a nonzero determinant is equivalent to having independent columns, an inverse, and nonzero eigenvalues. Similarly, a zero determinant is equivalent to having dependent columns, no inverse, and a zero eigenvalue.

 The next unit will focus on applications of these ideas. The main goal of this unit is to familiarize yourself with the arithmetic involved in linear algebra.

 Here are the preparation problems for this unit. All of these problems come from this book (not Schaum's Outlines). Remember that solutions appear at the end of each chapter.

Preparation Problems (click for solutions)		Webcasts (pdf copy)
Day 1	1c, 2h, 2o, 3b	1, 2, 3
Day 2	4c, 4e, 4f, 5a	4, 5, 6
Day 3	5c, 6a, 6b, 7d	7, 8
Day 4	7h, 7l, 8b, 8j	9, 10
Day 5	9b, 10e, 10g, 10h	11
Day 6	11, Lesson Plan, Quiz	

 Please download and print all the problems. I have tried to keep the text short and readable, and I want you to read the text (reading 3-4 pages for every day of class will get you through the whole book). If you find errors or find some parts hard to read, please email me so that I can improve the text.

 The problems listed below are found in the subsequent pages. The suggested problems are a minimal set of problems that I suggest you complete.

Concept	Suggestions	Relevant Problems
Basic Notation	1bcf,2abehln	1,2
Gaussian Elimination	3all,4acf	3,4
Rank and Independence	5ac,6bd	5,6
Determinants	7adgh	7
Inverses	8ag,9ac	8,9
Eigenvalues	10abdghi	10
Summarize	11(multiple times)	11

1.9   Problems

1. **Vector Arithmetic:** For each pair of vectors, (1) find the length of each vector, (2) compute the dot product of the vectors, (3) find the cosine of the angle between them, and (4) determine if the vectors are orthogonal.

   (a) $(1, 2), (3, 4)$

   (b) $[-2\ 1], [4\ \ -8]$

   (c) $(1, 2, 2), (3, 0, -4)$

   (d) $\langle -2, 3, 5 \rangle, \langle 4, -4, 4 \rangle$

   (e) $[1\ 1\ 1\ 1], [2\ 3\ \ -4\ \ -1]$

   (f) $[1\ \ -4\ 3\ 0\ 2]^T, [2\ 1\ 0\ 10\ 1]^T$

   More problems are in Schaum's Outlines - Chapter 1:1-13, 54-66.

2. **Matrix Arithmetic:** Consider the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ 3 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 4 \end{bmatrix}, D = \begin{bmatrix} 0 & 2 & 1 \\ 3 & 0 & 4 \\ -1 & 1 & 2 \end{bmatrix}, \text{ and } E = \begin{bmatrix} 3 \\ 1 \\ -4 \end{bmatrix}.$$

   Compute each of the following, or explain why it is not possible.

   (a) $A + B$ and $3A - 2B$

   (b) AB and BA

   (c) AC and CA

   (d) BC and CB

   (e) CD and DC

   (f) DE and ED

   (g) CDE

   (h) CC^T and C^TC

   (i) DC^T and CD^T

   (j) AB and B^TA^T

   (k) BA and A^TB^T

   (l) Trace of A, B, D.

   For each of the following, compute the product in three ways (1) using linear combinations of columns, (2) using rows dotted by columns, and (3) using linear combinations of rows (see Table 1.1).

   (m) AB

   (n) BC

   (o) CD

   (p) DE

   More problems are in Schaum's Outlines - Chapter 1:14-26, 67-79; Chapter 3:1-3, 68-70.

3. **Interpreting RREF:** Each of the following augmented matrices requires one row operation to be in reduced row echelon form. Perform the required row operation, and then write the solution to the corresponding system of equations in terms of the free variables.

   (a) $\begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & -2 \end{bmatrix}$

   (b) $\begin{bmatrix} 1 & 2 & 0 & -4 \\ 0 & 0 & 1 & 3 \\ -3 & -6 & 0 & 12 \end{bmatrix}$

   (c) $\begin{bmatrix} 0 & 1 & 0 & 4 \\ 0 & 0 & 5 & 15 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

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

   (e) $\begin{bmatrix} 0 & 1 & 0 & 7 & 0 & 3 \\ 0 & 0 & 1 & 5 & -3 & -10 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

   More problems are in Schaum's Outlines - Chapter 2:37-40

4. **Solving Systems with Gaussian Elimination:** Solve each system of equations using Gaussian elimination by reducing the augmented matrix to reduced row echelon form (rref).

   (a) $\begin{aligned} x - 3y &= 10 \\ 3x + 2y &= 8 \end{aligned}$

   (b) $\begin{aligned} 2x + 6y - z &= 9 \\ x + 3y - 3z &= 17 \end{aligned}$

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

   (d) $\begin{aligned} x_1 + 2x_3 &= -2 \\ 2x_1 - 3x_2 &= -3 \\ 3x_1 + x_2 - x_3 &= 2 \end{aligned}$

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

   (f) $\begin{aligned} x_1 - 2x_2 + x_3 &= 4 \\ -x_1 + 2x_2 + 3x_3 &= 8 \\ 2x_1 - 4x_2 + x_3 &= 5 \end{aligned}$

   (g) $\begin{aligned} x_1 + 2x_3 + 3x_4 &= -7 \\ 2x_1 + x_2 + 4x_4 &= -7 \\ -x_1 + 2x_2 + 3x_3 &= 0 \\ x_2 - 2x_3 - x_4 &= 4 \end{aligned}$

More problems are in Schaum's Outlines - Chapter 2: 44-48, 51-56, 70-72, 75-82, 86-90

5. **Rank and Linear Dependence:** Compute the rank of each matrix. Use this result to determine if the columns are linearly dependent. If the vectors are dependent, write each non-pivot column as a linear combination of the pivot columns.

(a) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & -2 & 1 \\ 3 & 0 & 4 \end{bmatrix}$
(b) $\begin{bmatrix} 1 & -1 & -1 \\ -2 & 3 & 5 \\ 3 & 1 & 9 \end{bmatrix}$
(c) $\begin{bmatrix} 1 & 3 & -1 & 9 \\ -1 & -2 & 0 & -5 \\ 2 & 1 & 3 & -2 \end{bmatrix}$
(d) $\begin{bmatrix} 1 & 3 & 12 & 1 \\ 2 & 0 & -6 & 3 \\ -1 & 1 & 8 & -4 \\ 0 & 2 & 10 & 2 \end{bmatrix}$

More problems are in Schaum's Outlines - Chapter 5: 19, 20, 22, 68.

6. **Linear Independence:** In each problem, determine if the vectors given are linearly independent. If the vectors are linearly dependent, write one of the vectors as a linear combination of the others.

(a) $[2, -1, 0], [1, 0, 3], [3, 2, 4]$

(b) $[1, 2, -3, 4], [3, -2, -1, -2], [5, -2, -3, -1]$

(c) $[0, 1, 1, 1], [1, 0, 1, 1], [1, 1, 0, 1], [1, 1, 1, 0]$

(d) $[1, 0, -1, -2], [1, 2, 3, 0], [0, 1, -1, 2], [2, 0, 1, -5]$

More problems are in Schaum's Outlines - Chapter 5: 1, 3, 4, 50, 53.

7. **Determinants:** Find the determinant of each matrix. For 3 by 3 matrices, compute the determinant in two different ways by using cofactor expansion along different rows or columns.

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

(b) $\begin{bmatrix} 4 & 3 \\ 5 & 9 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 1 & -2 \\ 1 & -1 & 4 \\ 2 & 2 & 1 \end{bmatrix}$

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

(f) $\begin{bmatrix} 5 & 1 & -3 \\ 2 & -1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$

(g) $\begin{bmatrix} 2 & 1 & -6 & 8 \\ 0 & 3 & 5 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & -4 \end{bmatrix}$

(h) $\begin{bmatrix} 3 & 2 & 5 & -1 \\ 0 & 8 & 4 & 2 \\ 0 & -1 & 0 & 0 \\ 0 & -5 & 3 & -1 \end{bmatrix}$

(i) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & -1 & 1 & 1 \\ -1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & -2 & 1 \end{bmatrix}$

For the following matrices, compute the determinant and use your answer to determine if the columns are linearly independent.

(k) Use the matrix from 5a   (l) Use the matrix from 5b   (m) Use the matrix from 5d

More problems are in Schaum's Outlines - Chapter 10: 1-6, 21, 23, 51-55

8. **Inverse:** Find the inverse of each matrix below.

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

(b) $\begin{bmatrix} -2 & 4 \\ 3 & -5 \end{bmatrix}$

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

(d) $\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$

(e) $\begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ -1 & 1 & 4 \end{bmatrix}$

(h) $\begin{bmatrix} 3 & 0 & 3 \\ 0 & -1 & 1 \\ 0 & 3 & -4 \end{bmatrix}$

(i) $\begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & -2 \\ 0 & 3 & 2 \end{bmatrix}$

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

(k) $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

(l) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

More problems are in Schaum's Outlines: Chapter 3: 7-11, 77-80.

9. **Solve systems using an inverse:** Solve each system by using an inverse from above.

(a) $\begin{cases} x_1 + 2x_2 &= 3 \\ 3x_1 + 4x_2 &= 4 \end{cases}$ using 8a.

(c) $\begin{cases} 1x_1 + 2x_3 &= 2 \\ x_2 &= 3 \\ -x_1 + x_2 + 4x_3 &= 1 \end{cases}$ using 8g.

(b) $\begin{cases} -2x_1 + 4x_2 &= 4 \\ 3x_1 - 5x_2 &= -2 \end{cases}$ using 8b.

(d) $\begin{cases} -2x_1 + 5x_3 &= -2 \\ -x_1 + 3x_3 &= 1 \\ 4x_1 + x_2 - x_3 &= 3 \end{cases}$ using 8j.

Now grab any problem from the entire unit that involves solving a system where an inverse applies. Find the inverse of the coefficient matrix and use it to find the solution. (Use the solutions at the back to select a problem with a unique solution).

10. **Eigenvalues and Eigenvectors:** For each matrix, compute the characteristic polynomial and eigenvalues. For each eigenvalue, give all of the corresponding eigenvectors. Remember that you can check your answer by comparing the trace to the sum of the eigenvalues, and the determinant to the product of the eigenvalues.

(a) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

(i) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$

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

(c) $\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$

(f) $\begin{bmatrix} 3 & 1 \\ 4 & 6 \end{bmatrix}$

(h) $\begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

More problems are in Schaum's Outlines - (only worry about the eigenvalues and eigenvectors part of each problem - ignore diagonalization questions) Chapter 11: 9, 10, 11, 17, 18, 20, 57, 58, 59, 60

11. At this point, you are ready to make your own examples. Create your own 2 by 2 or 3 by 3 matrix A.

- Find rref A.
- Find the rank of A.
- Are the columns of A independent?

- Compute $|A|$.
- Compute A^{-1} (or explain why not possible).
- Find the eigenvalues and eigenvectors of A.

For a 3 by 3 matrix, the eigenvalues and eigenvectors may be hard to find by hand (so use technology). Use technology to check your work by hand. The first column of questions applies to all matrices (square or not), whereas the last column only makes sense when the matrix is square. For a 2 by 2 matrix, you can always compute the eigenvalues using the quadratic formula, which may result in irrational or complex eigenvalues.

Repeat this exercise a few times with various types of matrices (e.g., diagonal, triangular, symmetric, skew-symmetric). Do you notice any patterns? If you pick smaller matrices, you can do everything except the eigenvalues and eigenvectors by hand. Once you feel comfortable with the computations by hand, move to a computer and start using larger matrices to find patterns.

1.10 Projects

The following projects require you to use technology to further explore a topic from this unit. Use a computer algebra system (CAS) to perform your computations so you can save your work and quickly modify things as needed.

1. A matrix and its transpose A^T have some common properties. Your job in this project is to explore the relationships between a matrix and its transpose.

   (a) Start by choosing your own 4 by 4 matrix. For both A and A^T, compute the rref, rank, determinant, inverse, eigenvalues, and eigenvectors. Which are the same?

   (b) Change your matrix to some other 4 by 4 or larger matrix and repeat part (a).

   (c) Based on these two examples, which quantities do you think will be always be the same for A and A^T?

   (d) Try creating an ugly large matrix and see if your conjecture is correct. While examples are not proofs that a conjecture is true, examples do provide the foundation upon which all mathematical theorems are built. New ideas often stem from examples.

   (e) What happens if a matrix is not square? Select your own non square matrix (such as 3 by 4). Compute the rref of both A and A^T, as well as the rank.

   (f) Change the matrix and repeat part (e). Make a conjecture about the relationship between the rrefs or ranks of a non-square matrix and its transpose.

1.11 Solutions

1. **Vector Arithmetic:**

   (a) $\sqrt{5}, 5, 11, \cos\theta = \frac{11}{5\sqrt{5}}$, No

   (b) $\sqrt{5}, 2\sqrt{5}, 0, \cos\theta = 0$, Yes

   (c) $3, 5, -5, \cos\theta = -\frac{1}{3}$, No

   (d) $\sqrt{38}, 4\sqrt{3}, 0, \cos\theta = 0$, Yes

   (e) $2, \sqrt{30}, 0, \cos\theta = 0$, Yes

   (f) $\sqrt{30}, \sqrt{106}, 0, \cos\theta = 0$, Yes

2. **Matrix Arithmetic:**

   (a) $\begin{bmatrix} 5 & 1 \\ 6 & 6 \end{bmatrix}, \begin{bmatrix} -5 & 8 \\ 3 & 8 \end{bmatrix}$

   (b) $\begin{bmatrix} 10 & 3 \\ 24 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 4 \\ 9 & 14 \end{bmatrix}$

   (c) $\begin{bmatrix} 5 & 6 & 7 \\ 11 & 12 & 13 \end{bmatrix}$, CA not possible

   (d) $\begin{bmatrix} 2 & -3 & -8 \\ 7 & 6 & 5 \end{bmatrix}$, CB not possible

   (e) $\begin{bmatrix} 1 & 1 & -1 \\ 5 & 8 & 22 \end{bmatrix}$, DC not possible

   (f) $\begin{bmatrix} -2 \\ -7 \\ -10 \end{bmatrix}$, ED not possible

   (g) $\begin{bmatrix} 8 \\ -65 \end{bmatrix}$

   (h) $\begin{bmatrix} 2 & -2 \\ -2 & 29 \end{bmatrix}, \begin{bmatrix} 5 & 6 & 7 \\ 6 & 9 & 12 \\ 7 & 12 & 17 \end{bmatrix}$

   (i) $\begin{bmatrix} -1 & 10 \\ -1 & 22 \\ -3 & 9 \end{bmatrix}, \begin{bmatrix} -1 & -1 & -3 \\ 10 & 22 & 9 \end{bmatrix}$

   (j) $\begin{bmatrix} 10 & 3 \\ 24 & 5 \end{bmatrix}, \begin{bmatrix} 10 & 24 \\ 3 & 5 \end{bmatrix}$

   (k) $\begin{bmatrix} 1 & 4 \\ 9 & 14 \end{bmatrix}, \begin{bmatrix} 1 & 9 \\ 4 & 14 \end{bmatrix}$

   (l) $\operatorname{tr} A = 5$, $\operatorname{tr} B = 6$, $\operatorname{tr} D = 2$.

3. **Interpreting RREF:**

   (a) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \end{array}\right]$, $\begin{array}{l} x_1 = 3 \\ x_2 = -2 \\ x_3 = 1 \end{array}$

   (b) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & -4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$, $\begin{array}{l} x_1 = -2x_2 - 4 \\ x_2 = x_2 \text{ (free variable)} \\ x_3 = 3 \end{array}$

   (c) $\left[\begin{array}{ccc|c} 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$, $\begin{array}{l} x_1 = x_1 \text{ (free)} \\ x_2 = 4 \\ x_3 = 3 \end{array}$

   (d) $\left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$, no solution

   (e) $\left[\begin{array}{ccccc|c} 0 & 1 & 0 & 7 & 0 & 3 \\ 0 & 0 & 1 & 5 & 0 & -4 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$, $\begin{array}{l} x_1 = x_1 \text{ (free)} \\ x_2 = -7x_4 + 3 \\ x_3 = -5x_4 - 4 \\ x_4 = x_4 \text{ (free)} \\ x_5 = 2 \end{array}$

4. **Solving Systems with Gaussian Elimination:**

   (a) $\left[\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & -2 \end{array}\right]$, $\begin{array}{l} x_1 = 4 \\ x_2 = -2 \end{array}$

   (b) $\left[\begin{array}{ccc|c} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & -5 \end{array}\right]$, $\begin{array}{l} x_1 = -3x_2 + 2 \\ x_2 = x_2 \text{ (free)} \\ x_3 = -5 \end{array}$

   (c) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right]$, $\begin{array}{l} x_1 = -2 \\ x_2 = 1 \\ x_3 = 3 \end{array}$

(d) $\begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & -1 \end{bmatrix}$ $\begin{array}{l} x_1 = 0 \\ x_2 = 1 \\ x_3 = -1 \end{array}$

(e) $\begin{bmatrix} 1 & 0 & 1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 1 \end{bmatrix}$, no solution

(f) $\begin{bmatrix} 1 & -2 & 0 & | & 1 \\ 0 & 0 & 1 & | & 3 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$ $\begin{array}{l} x_1 = 2x_2 + 1 \\ x_2 = x_2 \text{ (free)} \\ x_3 = 3 \end{array}$

(g) $\begin{bmatrix} 1 & 0 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & 0 & | & 1 \\ 0 & 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & 1 & | & -3 \end{bmatrix}$ $\begin{array}{l} x_1 = 2 \\ x_2 = 1 \\ x_3 = 0 \\ x_4 = -3 \end{array}$

5. **Rank and Linear Dependence:**

(a) 3, independent

(b) 2, dependent, rref is $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$ so

$$2\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + 3\begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ 9 \end{bmatrix}$$

(c) 2, dependent, rref is $\begin{bmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ so

$$2\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} - 1\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \text{ and}$$

$$-3\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + 4\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ -5 \\ -2 \end{bmatrix}$$

(d) 3, dependent, rref is $\begin{bmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ so

$$-3\begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 3 \\ 0 \\ 1 \\ 2 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 3 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ -6 \\ 8 \\ 10 \end{bmatrix}$$

6. **Linear Independence:**

(a) $\begin{bmatrix} 2 & 1 & 3 \\ -1 & 0 & 2 \\ 0 & 3 & 4 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, independent

(b) $\begin{bmatrix} 1 & 3 & 5 \\ 2 & -2 & -2 \\ -3 & -1 & -3 \\ 4 & -2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, dependent

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

(c) Independent. Placing the vectors in columns of a 4 by 4 matrix reduces to the identity, so there are 4 pivot columns.

(d) $\begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 2 & 1 & 0 \\ -1 & 3 & -1 & 1 \\ -2 & 0 & 2 & -5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & \frac{3}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

dependent

$$\frac{3}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \\ -2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ -5 \end{bmatrix}$$

7. **Determinants:**

(a) -2 (e) -17 (i) 24

(b) 21   (f) -58 (j) 0

(c) 9   (g) -24

(d) -28 (h) -30

8. **Inverse:**

(a) $\begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{5}{2} & 2 \\ \frac{3}{2} & 1 \end{bmatrix}$

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

(d) $\begin{bmatrix} -\frac{5}{2} & \frac{3}{2} \\ 2 & -1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & -3 \\ -2 & 7 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$

(g) $\begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ 0 & 1 & 0 \\ \frac{1}{6} & -\frac{1}{6} & \frac{1}{6} \end{bmatrix}$

(h) $\begin{bmatrix} \frac{1}{3} & 3 & 1 \\ 0 & -4 & -1 \\ 0 & -3 & -1 \end{bmatrix}$

(i) $\begin{bmatrix} -6 & \frac{7}{2} & \frac{1}{2} \\ 2 & -1 & 0 \\ -3 & \frac{3}{2} & \frac{1}{2} \end{bmatrix}$

(j) $\begin{bmatrix} -3 & 5 & 0 \\ 11 & -18 & 1 \\ -1 & 2 & 0 \end{bmatrix}$

(k) $\begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{bmatrix}$

(l) $\dfrac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

9. **Solve systems using an inverse:**

(a) $\begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ \frac{5}{2} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{5}{2} & 2 \\ \frac{3}{2} & 1 \end{bmatrix}\begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$

(c) $\begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ 0 & 1 & 0 \\ \frac{1}{6} & -\frac{1}{6} & \frac{1}{6} \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$

(d) $\begin{bmatrix} -3 & 5 & 0 \\ 11 & -18 & 1 \\ -1 & 2 & 0 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 11 \\ -37 \\ 4 \end{bmatrix}$

10. **Eigenvalues and Eigenvectors:**

(a) $\lambda^2 - 4\lambda + 3$, for $\lambda = 3$ eigenvectors are $\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_2$

where $x_2 \neq 0$, for $\lambda = 1$ eigenvectors are $\begin{bmatrix} 1 \\ 0 \end{bmatrix} x_1$

where $x_1 \neq 0$.

(b) $\lambda^2 + 2\lambda + 1$, for $\lambda = -1$ eigenvectors are $\begin{bmatrix} -1 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$

(c) $\lambda^2 - 4\lambda - 5$, for $\lambda = -1$ eigenvectors are $\begin{bmatrix} -1 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$, for $\lambda = 5$ eigenvectors are $\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$

(d) $\lambda^2 + 1$, for $\lambda = i$ eigenvectors are $\begin{bmatrix} -i \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$, for $\lambda = -i$ eigenvectors are $\begin{bmatrix} i \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$

(e) $\lambda^2 - 4\lambda - 5$, for $\lambda = 5$ eigenvectors are $\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$, for $\lambda = -1$ eigenvectors are $\begin{bmatrix} -2 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$

(f) $\lambda^2 - 9\lambda + 14$, for $\lambda = 7$ eigenvectors are $\begin{bmatrix} 1/4 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$ (alternatively you could write any nonzero multiple of $[1, 4]^T$ if you want to avoid fractions), for $\lambda = 2$ eigenvectors are $\begin{bmatrix} -1 \\ 1 \end{bmatrix} x_2$ where $x_2 \neq 0$

(g) $-\lambda^3 + 3\lambda^2 - 3\lambda + 1 = -(\lambda-1)^3 = -(\lambda-1)^3$, for $\lambda = 1$ eigenvectors are $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x_1$ where $x_1 \neq 0$

(h) $-\lambda^3 + 3\lambda^2 - 3\lambda + 1 = -(\lambda - 1)^3$, for $\lambda = 1$ eigenvectors are $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} x_3$ where x_1 and x_3 cannot simultaneously be zero.

(i) $-\lambda^3 + 7\lambda^2 - 15\lambda + 9 = -(\lambda - 3)^2(\lambda - 1)$, for $\lambda = 3$ eigenvectors are nonzero linear combinations of $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, for $\lambda = 1$ eigenvectors are nonzero linear combinations of $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$.

(j) $-\lambda^3 + 4\lambda^2 - 5\lambda + 2 - (\lambda - 2)(\lambda - 1)^2$, for $\lambda = 1$ eigenvectors are nonzero linear combinations of $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, for $\lambda = 2$ eigenvectors are nonzero linear combinations of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Chapter 2

Linear Algebra Applications

This chapter covers the following ideas. The purpose of this chapter is to show you a variety of applications of linear algebra and provide you with ample opportunities to practice and polish the skills learned in the arithmetic unit.

1. Construct visualizations of matrices related to vector fields. Explain how eigenvalues and eigenvectors relate to vector fields.

2. Explain how to generalize the derivative to a matrix. Use this generalization to locate optimal values of a function using the second derivative test. Explain the role of eigenvalues and eigenvectors in the second derivative test.

3. Describe a Markov process. Explain what a steady-state solution is and how it relates to eigenvectors of the transition matrix.

4. Find the currents in an electrical system involving batteries and resistors, using both Gaussian elimination and Cramer's rule.

5. Find interpolating polynomials. Generalize this curve-fitting process to solve the least squares regression problem of fitting a line to a set of data.

6. Find bases for the column space, row space, and null space of a matrix and its transpose by row reducing A and A^T. Use these bases to find the coordinates of vectors relative to the bases.

2.1   Vector Fields

Vectors help us understand motion, forces, acceleration, and more. Imagine that you were a dust particle floating through the air. Along the way, wind pushes you in various directions, sometimes causing you to swirl around, increasing your speed, or changing your direction. As you move through space, you encounter forces (vectors) pushing you around everywhere you go. This field of vectors provides an example of a vector field in space. Magnetic fields, electrical fields, force fields, velocity fields, flow fields, and more are key ideas studied throughout the sciences. Meteorologists study air flow, physicists study magnetic and electric fields, and engineers study heat flow and fluid flow. In business, economists study the flow of money, which again results in studying vector fields. Pretty much every study of motion or change comes back to studying an underlying

field of vectors. In this section, we will explore some simple two-dimensional vector fields arising from matrices, and then learn to physically observe the eigenvalues and eigenvectors from a plot of the vector field. When studying more complicated vector fields, researchers use high dimensional calculus to simplify the complicated field to the vector fields we will study in this section. Eigenvalues and eigenvectors will play a major role.

Let's start with some notation. When you hear that y is a function of x, you often write $y = f(x)$ to remind yourself that when you put in x you get out y. For most of your lower-level math courses, you have always thought of x and y as real numbers. In this case, the domain of the function (the x-values) is the set of all real numbers, which we denote with the symbol $\mathbb{R}$. The potential y values (the range) are also real numbers. If we want to pay attention to the domain and range of the function, we could write $f\colon D \to R$ where D is the domain and R is the range. For a function such as $y = f(x)$, where both x and y are real numbers (written $x \in \mathbb{R}$ and $y \in \mathbb{R}$), we could write $f\colon \mathbb{R} \to \mathbb{R}$. With this notation, we can generalize the concept of functions to allow us to input vectors and output vectors. If we want to input a 2D vector and get out a 3D vector, then we could write $f\colon \mathbb{R}^2 \to \mathbb{R}^3$. We'll use the notation $\mathbb{R}^n$ to represent the set of n dimensional vectors. Throughout this semester, we'll be studying functions of the form $f\colon \mathbb{R}^n \to \mathbb{R}^m$. In particular, we'll be studying special types of functions called linear transformations which can often be represented by matrices.

A *vector field* is a function where you input a vector and receive back a vector of the same dimension. If $\vec{x}$ is a 2D vector, then a vector field is a function of the form $\vec{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ (we put a vector above $\vec{F}$ because the output is a vector). Any square matrix represents a vector field—if A is a square matrix, then the function $\vec{F}(\vec{x}) = A\vec{x}$ means that the output $\vec{y} = A\vec{x}$. Let's look at an example and a graph.

> $\mathbb{R}$ means all real numbers.
>
> $f\colon D \to R$ notation

> A vector field takes vectors of size n to vectors of size n,
> $$f\colon \mathbb{R}^n \to \mathbb{R}^n.$$

Example 2.1. The matrix $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ gives us the vector field

$$\vec{F}(x,y) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2x + y, x + 2y).$$

We input the point $(x,y) \in \mathbb{R}^2$ and get out the vector $(2x + y, x + 2y) \in \mathbb{R}^2$. For example, we compute $\vec{F}(1,0) = (2(1) + 0, (1) + 2(0)) = (2,1)$. To visualize this input and output, we place the vector $(2,1)$ (the output) in the plane with its base at the point $(1,0)$ (the input). To visualize the function $\vec{F}$, we can pick an assortment of input points and at each input point $\vec{x}$, draw the output vector $\vec{F}(\vec{x})$. The example below shows how to place the vectors corresponding to the 9 points with x or y coordinate $0, \pm 1$.

(x, y)	$\vec{F}(x, y)$
$(0, 0)$	$(0, 0)$
$(1, 0)$	$(2, 1)$
$(1, 1)$	$(3, 3)$ − an eigenvalue is 3
$(0, 1)$	$(1, 2)$
$(-1, 1)$	$(-1, 1)$ − an eigenvalue is 1
$(-1, 0)$	$(-2, 1)$
$(-1, -1)$	$(-3, -3)$
$(0, -1)$	$(-1, -2)$
$(1, -1)$	$(1, -1)$
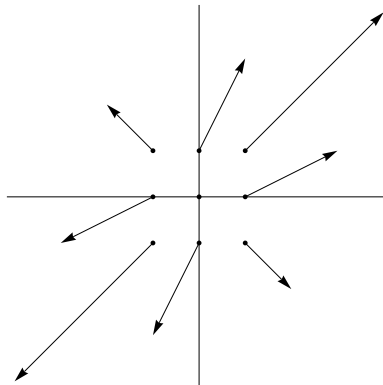
Notice that (1,1) is an eigenvector corresponding to the eigenvalue 3 because $A(1, 1) = 3(1, 1)$. Also, note that $(-1, 1)$ is an eigenvector corresponding to the eigenvalue 1 because $A(-1, 1) = (-1, 1)$. These two eigenvectors represent the directions in which the vector field pushes objects radially outwards. Anywhere along the line spanned by an eigenvector, the vector field pushes radially outward (as shown in the middle figure). The vector field will never push you from one side of an eigenvector to the other side. The eigenvalue for the eigenvector tells you the strength of the push, which explains why the vectors in the $(1, 1)$ direction are three times as long as the vectors in the $(-1, 1)$ direction. The middle plot in Figure 2.1 includes two lines representing the eigenvector directions.

Eigenvectors tell you where vector fields push radially outwards or inwards. Eigenvalues tell the strength and direction (outward or inward) of the push.
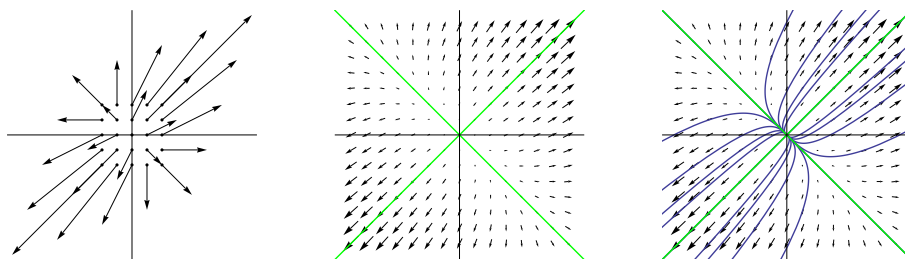
Figure 2.1:   Three views of the vector field $\vec{F}(x, y) = (2x + y, x + 2y)$. The first plot shows the appropriate lengths. The second plot includes two lines which come from the eigenvector directions. The third plot includes flow lines which describe the path of an object flowing through the vector field. Because flow is outward in all directions, this vector field is called a source.

When a computer draws a vector field, it creates hundreds of vectors, takes the longest vector and shrinks its length so that it won't cross any others, and then proportionally shrinks every other vector. This creates lots of tiny vectors which represent the directions (not magnitudes) of the vectors (see the last two plots in Figure 2.1). You can think of a vector field as the flow of water—if you were to drop a particle into the field, then it would follow the vectors (just as drift wood follows the flow of a river). These flow lines are represented in the last plot of Figure 2.1.

Figure 2.2 contains other examples. Again, the eigenvalues and eigenvectors determine how to draw flow lines in the graph.

Sink (inward flow)

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

Eigenvalues: $-2 - \sqrt{2}, \sqrt{2} - 2$
Eigenvectors: $(-\sqrt{2}, 1), (\sqrt{2}, 1)$

Rotational

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

Eigenvalues: $i, -i$
Eigenvectors: $(i, 1), (-1, 1)$

Saddle (both in and out)

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

Eigenvalues: $5, -1$
Eigenvectors: $(1, 1), (-2, 1)$

Outward Spiral

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

Eigenvalues: $1 + i, 1 - i$
Eigenvectors: $(i, 1), (-i, 1)$
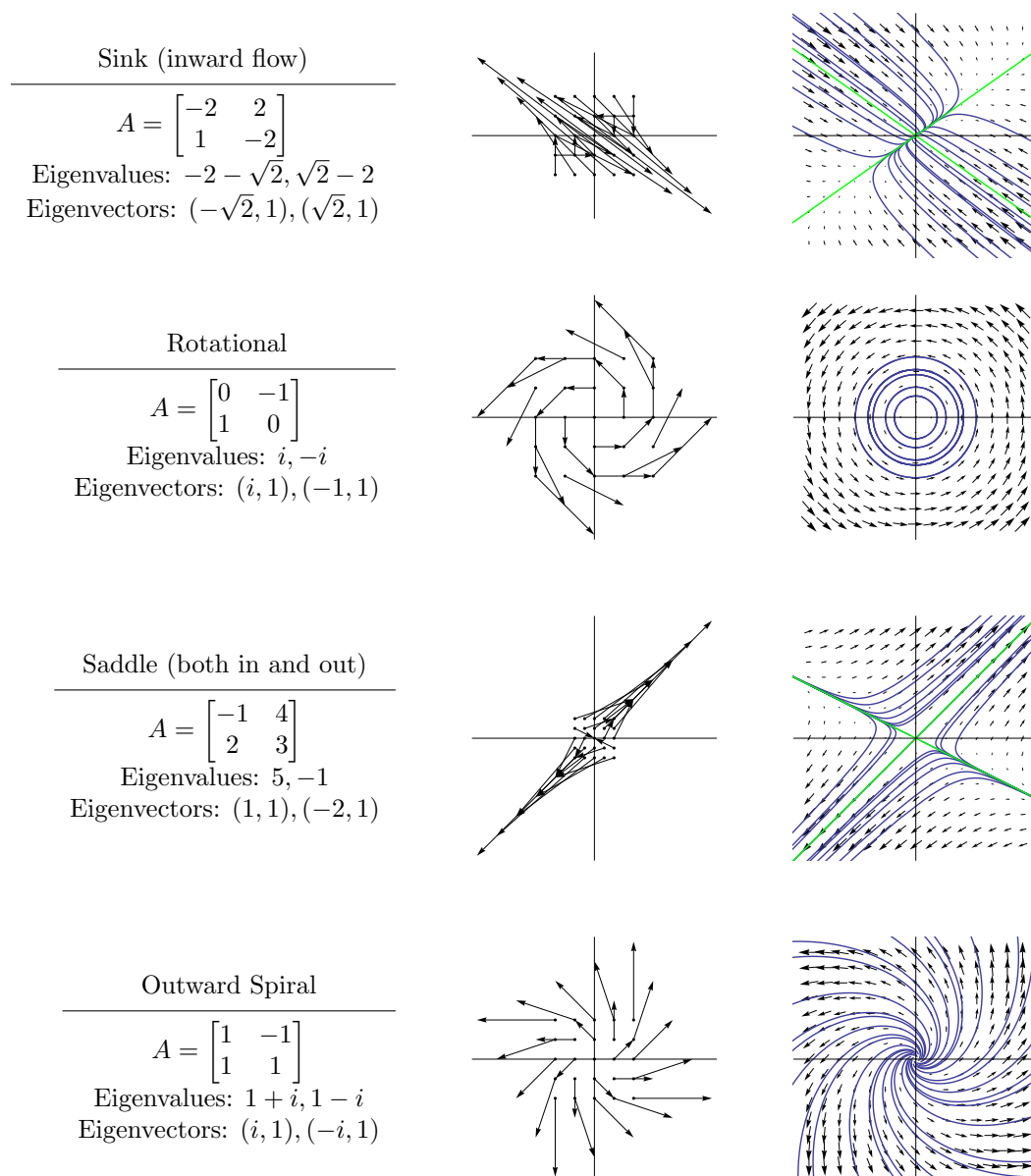
Figure 2.2: Different kinds of vector fields. Every 2 by 2 matrix represents a vector field. When the eigenvalues are real, the eigenvectors help determine direction of flow. Positive eigenvalues result in outward flow along the corresponding eigenvector. Negative eigenvalues result in inward flow. Pure imaginary eigenvalues yield rotational flow. When the eigenvalues involve both real and imaginary parts, the real part determines whether the flow is inward or outward, and the complex part determines rotation.

2.2 The Second Derivative Test

In this section, we'll explore how eigenvalues and eigenvectors of a matrix are the key tool needed to generalize the second derivative test from first semester calculus to all dimensions. The key application is that eigenvalues tell you about concavity of a function in the eigenvector directions. If all eigenvalues are negative, then a function is concave downwards in all directions, which means that a critical point must correspond to a maximum.

Let's start with a review from first semester calculus. If a function $y = f(x)$ has a relative extreme value at $x = c$, then $f'(c) = 0$ or the derivative is undefined. The places where the derivative is either zero or undefined are called critical values of the function. The first derivative test allows you to check the value of the derivative on both sides of the critical value and then interpret whether that point is a maximum or minimum using increasing/decreasing arguments. The second derivative test requires you to compute the second derivative at $x = c$ and then make an observation. If $f''(c) > 0$ (the function is concave upwards), then the function has a minimum at $x = c$. If $f''(c) < 0$ (the function is concave downwards), then the function has a maximum at $x = c$. If $f''(c) = 0$, then the second derivative test fails.

Example 2.2. The function $f(x) = x^3 - 3x$ has derivatives $f' = 3x^2 - 3$ and $f'' = 6x$. The first derivative is zero when $3(x^2 - 1) = 3(x - 1)(x + 1) = 0$, or $x = \pm 1$. The second derivative at $x = 1$ is $f''(1) = 6$ (concave upwards), so there is a minimum at $x = 1$. The second derivative at $x = -1$ is $f''(-1) = -6$ (concave downwards), so there is a maximum at that point. This is all illustrated in the margin.
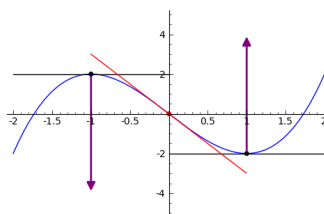
2.2.1 Partial Derivatives

Before extending the second derivative test to all dimensions, we have to talk about taking derivatives. If a function has more than one input variable, say $f(x, y) = 9 - x^2 + 3y$ which has 2 input variables, then we have to take multiple derivatives, one with respect to each variable. These derivatives (called partial derivatives) are found by holding all variables constant except the one you want the derivative with respect to, and then just computing a normal derivative. For the function $f(x, y)$, we define the partial derivative of f with respect to x as

$$f_x(x, y) = f_x = \frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h},$$

and the partial derivative of f with respect to y as

$$f_y(x, y) = f_y = \frac{\partial f}{\partial y} = \lim_{k \to 0} \frac{f(x, y + k) - f(x, y)}{k}.$$

Notice that f_x computes a limit as y is held constant and we vary x.

Example 2.3. For $f(x, y) = 3x^2 + 4xy + \cos(xy) + y^3$, we obtain $f_x = 6x + 4y - y\sin(xy) + 0$ and $f_y = 0 + 4x - x\sin(xy) + 3y^2$. Partial derivatives are found by holding all other variables constant, and then differentiating with respect to the variable in question.

2.2.2 The Derivative

The derivative of a function $\vec{f} \colon \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix written $D\vec{f}(\vec{x})$, where the columns of the matrix are the partial derivatives of the function with respect to an input variable (the first column is the partial derivative

with respect to the first variable, and so on). Some people call this derivative the "total" derivative instead of the derivative, to emphasize that the "total" derivative combines the "partial" derivatives into a matrix. Some examples of functions and their derivative are in Table 2.1. Remember that each column of the matrix corresponds to the partial derivatives with respect to an input variable.

Function	Derivative
$f(x) = x^2$	$Df(x) = [2x]$
$\vec{r}(t) = \langle 3\cos(t), 2\sin(t) \rangle$	$D\vec{r}(t) = \begin{bmatrix} -3\sin t \\ 2\cos t \end{bmatrix}$
$\vec{r}(t) = \langle \cos(t), \sin(t), t \rangle$	$D\vec{r}(t) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix}$
$f(x, y) = 9 - x^2 - y^2$	$Df(x, y) = \begin{bmatrix} -2x & -2y \end{bmatrix}$
$f(x, y, z) = x^2 + y + xz^2$	$Df(x, y, z) = \begin{bmatrix} 2x + z^2 & 1 & 2xz \end{bmatrix}$
$\vec{F}(x, y) = \langle -y, x \rangle$	$D\vec{F}(x, y) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
$\vec{F}(r, \theta, z) = \langle r\cos\theta, r\sin\theta, z \rangle$	$D\vec{F}(r, \theta, z) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$
$\vec{r}(u, v) = \langle u, v, 9 - u^2 - v^2 \rangle$	$D\vec{r}(u, v) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2u & -2v \end{bmatrix}$

Table 2.1: Derivatives of functions are found by creating a matrix whose columns are the partial derivatives of the function. If the function has multiple outputs, then the columns of the matrix will have one row for each output. Hence a function with 2 inputs and 3 outputs will have 2 columns with 3 rows in each column.

2.2.3 The Second Derivative Test

We're now ready to extend the second derivative test to all dimensions. We'll focus on only functions of the form $f \colon \mathbb{R}^2 \to \mathbb{R}$, as this gives you the tools needed to study optimization in all dimensions. Because the output is one-dimensional, it makes sense to talk about a largest or smallest number.

The first derivative test does not work when the domain has more than one dimension because there are infinitely many ways to approach a point—you can't just look at the left and right side to decide if you are at an optimum value. However, at a local extreme value the derivative is still zero, which results in solving a system of equations to find any critical points. In higher dimensions, there are three classifications of critical points: maximum, minimum, or saddle point (a point where the tangent plane is horizontal, but in some directions you increase and in other directions you decrease).

The second derivative test does work when the domain has dimension greater than one. Consider the function $z = f(x, y)$. Its derivative $Df(x, y) = \begin{bmatrix} f_x & f_y \end{bmatrix}$ is a function with two inputs (x, y) and two outputs (f_x, f_y). The second

derivative $D^2 f(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ is a 2×2 square matrix called the Hessian of f. This matrix will always be symmetric, in that the transpose of the matrix equals itself (because $f_{xy} = f_{yx}$ from a theorem in multivariable calculus). At a critical point (where the first derivative is zero), the eigenvalues of $D^2 f$ give the directional second derivative in the direction of a corresponding eigenvector. The largest eigenvalue is the largest possible value of the second derivative in any direction and the smallest eigenvalue is the smallest possible value of the second derivative in any direction.

The **second derivative test** is the following. Start by finding all the critical points (places where the derivative is zero). Then find the eigenvalues of the second derivative. Each eigenvalue represents the directional 2nd derivative in the direction of a corresponding eigenvector. In every other direction, the directional 2nd derivative is between the smallest and largest eigenvalue.

the second derivative test with eigenvalues

1. If the eigenvalues are all positive at a critical point, then in every direction the function is concave upwards. The function has a minimum at that critical point.

2. If the eigenvalues are all negative at a critical point, then in every direction the function is concave downwards. The function has a maximum there.

3. If there is a positive eigenvalue and a negative eigenvalue, the function has a saddle point there.

4. If either the largest or smallest eigenvalue is zero, then the second derivative test fails.

Eigenvalues are the key numbers needed to generalize optimization to all dimensions. A proof of this fact is beyond the scope of this class.

Example 2.4. For the function $f(x, y) = x^2 + xy + y^2$, the derivative is $Df = \begin{bmatrix} 2x + y & x + 2y \end{bmatrix}$, which is zero only at $x = 0, y = 0$ (solve the system of equations $2x + y = 0, x + 2y = 0$). The Hessian is $D^2 f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. The eigenvalues are found by solving $0 = \det \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix} = (2 - \lambda)^2 - 1 = 4 - 4\lambda + \lambda^2 - 1 = (\lambda - 3)(\lambda - 1)$, so $\lambda = 3, 1$ are the eigenvalues. Since both eigenvalues are positive, the function is concave upwards in all directions, so there is a minimum at $(0, 0)$.

The Hessian ($D^2 f$) is always symmetric.

The eigenvectors of the Hessian help us understand more about the graph of the function. An eigenvector corresponding to 3 is $(1, 1)$, and corresponding to 1 is $(-1, 1)$. These vectors are drawn in the graph of the function in the margin, together with two parabolas whose 2nd derivatives are precisely 3 and 1. The parabola which opens upwards the most quickly has a 2nd derivative of 3. The other parabola has a second derivative of 1. In every other direction, the 2nd derivative would be between 1 and 3.
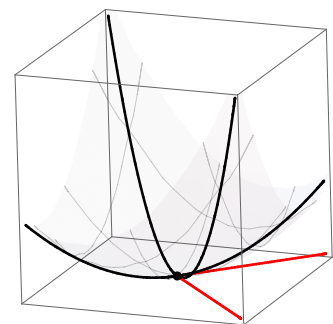
Example 2.5. For the function $f(x, y) = x^3 - 3x + y^2 - 4y$, the derivative is $Df = \begin{bmatrix} 3x^2 - 3 & 2y - 4 \end{bmatrix}$, which is zero at $x = 1, y = 2$ or $x = -1, y = 2$. Hence there are two critical points, so we have to find two sets of eigenvalues. The Hessian is $D^2 f = \begin{bmatrix} 6x & 0 \\ 0 & 2 \end{bmatrix}$. When $x = -1, y = 2$, the eigenvalues of $\begin{bmatrix} -6 & 0 \\ 0 & 2 \end{bmatrix}$ are $\lambda = -6, 2$. Since one is positive and one is negative, there is a saddle point at $(-1, 2)$. When $x = 1, y = 2$, the eigenvalues of $\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$ are $\lambda = 6, 2$. Since both

are positive, there is a minimum at $(-1, 2)$ (as in every direction the function is concave upwards).

Again the eigenvectors help us understand how the function behaves, as illustrated in the margin. At $(1, 2)$, we have an eigenvector $(1, 0)$ corresponding to 6, and $(0, 1)$ corresponding to 2. In both eigenvector directions, the function is concave upwards, but opens more steeply in the $(1, 0)$ direction since $6 > 2$. At $(-1, 2)$, we have an eigenvector $(1, 0)$ corresponding to -6, and $(0, 1)$ corresponding to 2. The functions opens steeply downwards in the $(1, 0)$ direction, and upwards in the $(0, 1)$ direction.
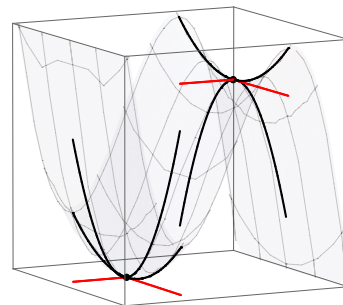
2.3 Markov Processes

Matrices can be used to model a process called a Markov process. To fit this kind of model, a process must have specific states, and the matrix which models the process is a transition matrix which specifies how each state will change through a given transition. An example of a set of states is "open" or "closed" in an electrical circuit, or "working properly" and "working improperly" for operation of machinery at a manufacturing facility. A car rental company which rents vehicles in different locations can use a Markov Process to keep track of where their inventory of cars will be in the future. Stock market analysts use Markov processes and generalizations, liked stochastic processes and hidden Markov models, to make predictions about future stock values. The algorithm that powers Google's internet search is based on the idea that browsing the web is a Markov process.

Example 2.6. Let's illustrate a Markov Process related to classifying land in some region as "Residential," "Commercial," or "Industrial." Suppose in a given region over a 5 year time span that 80% of residential land will remain residential, 10% becomes commercial, and 10% becomes industrial. For commerical land, 70% remains commercial, 20% becomes residential, and 10% becomes industrial. For industrial land, 70% remains industrial, 30% becomes commercial, and 0% becomes residential. To predict what happens to the land zoning at the end of a 5 year period, provided we know the current R, C, and I values, we would compute

$$
\begin{aligned}
R_{\text{new}} &= .8R + .2C + 0I \\
C_{\text{new}} &= .1R + .7C + .3I \\
I_{\text{new}} &= .1R + .1C + .7I
\end{aligned}
\xrightarrow{\text{matrix form}}
\begin{bmatrix} R_{\text{new}} \\ C_{\text{new}} \\ I_{\text{new}} \end{bmatrix}
=
\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix}
\begin{bmatrix} R \\ C \\ I \end{bmatrix}
$$

The matrix on the right above is called the transition matrix of the Markov process. It is a matrix where each column relates to one of the "states," and the numbers in that column are the proportions of the column state that will change to the row state through the transition (the ordering on row and column states is the same). We calculate the next "state" by multiplying our current state by the transition matrix. If current land use is about 50% residential, 30% commercial, and 20% industrial, then 5 years later the land use would be

$$
A\vec{x} =
\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix}
\begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix}
=
\begin{bmatrix} 46 \\ 32 \\ 22 \end{bmatrix}
$$

If the same transitions in land use continue, we can multiply the previous projection (state) by the transition matrix to obtain a 10 and 15 year projection for land use:

Notice that the percentages go in columns. The residential percentages $(.8, .1, .1)$ become the first column. This gives us the transistion matrix

$$
\begin{array}{c}
\begin{array}{ccc} R & C & I \end{array} \\
\begin{array}{c} \text{to } R \\ \text{to } C \\ \text{to } I \end{array}
\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \\
\text{Transition Matrix}
\end{array}
$$

$$A(A\vec{x}) = A^2\vec{x} \qquad\qquad A(A^2\vec{x}) = A^3\vec{x}$$

$$\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 46 \\ 32 \\ 22 \end{bmatrix} = \begin{bmatrix} 43.2 \\ 33.6 \\ 23.2 \end{bmatrix} \qquad \begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 43.2 \\ 33.6 \\ 23.2 \end{bmatrix} = \begin{bmatrix} 41.28 \\ 34.8 \\ 23.92 \end{bmatrix}$$

<div align="center">10 Year Projection   15 Year Projection</div>

Each time we multiply on the left by our transition matrix, we add 5 more years to our projection. This projection is valid as long as the same trends continue.

2.3.1 Steady State Solutions

Consider the land use example from above. Let $\vec{x}_0$ be our initial state. If our transition matrix A remains the same forever, what will eventually be the proportion of land devoted to residential, commercial, or industrial use? We can write each new state as powers of the transition matrix A by writing

$$\vec{x}_1 = A\vec{x}_0, \quad \vec{x}_2 = A\vec{x}_1 = AA\vec{x}_0 = A^2\vec{x}_0, \quad \vec{x}_3 = A^3\vec{x}_0, \quad \vec{x}_n = A^n\vec{x}_0.$$

What happens to the product $A^n\vec{x}_0$ as $n \to \infty$? Can we reach a state $\vec{x} = (R, C, I)$ such that $A\vec{x} = \vec{x}$, where the next state is the same as the current? If this occurs, then any future transitions will not change the state either. This state $\vec{x}$ is called a *steady state*, since it does not change when multiplied by the transition matrix (it remains steady).

Let's look more closely at the steady state equation

$$A\vec{x} = \vec{x} = 1\vec{x}.$$

Do you see the eigenvalue-eigenvector notation $A\vec{x} = \lambda\vec{x}$? The eigenvalue is 1 and the steady state is an eigenvector. Without even thinking that an eigenvalue was important, we reduced finding a steady state to an eigenvalue-eigenvector problem. For any Markov process (the columns of the matrix must sum to 1), the number $\lambda = 1$ will always be an eigenvalue. All we have to do is find the eigenvectors corresponding to the eigenvalue 1. The solution to $\lim\limits_{n\to\infty} A^n\vec{x}_0$ is this steady state, and is an eigenvector of the transition matrix.

> The number $\lambda = 1$ is always an eigenvalue of the transition matrix. Steady states are eigenvectors of the transition matrix.

Example 2.7. For the land use Markov process from the previous example, an eigenvector corresponding to 1 is $\left(\frac{3}{2}, \frac{3}{2}, 1\right)$. Since any nonzero multiple of an eigenvector is again an eigenvector, we can multiply by a constant so that the proportions sum to 100. Multiplying by 2 we have $(3, 3, 2)$, which means that the ratio of land will be 3 acres residential to 3 acres commercial to 2 acres industrial. To write this in terms of percentages, divide each component by 8 (the sum $3 + 3 + 2$) to obtain $3/8 : 3/8 : 2/8$. Multiply by 100 to get the percentages $37.5\% : 37.5\% : 25\%$, the long term percentages of land use.

> Divide each entry by the sum to get percentages.

2.4 Kirchoff's Electrial Laws

Gustav Kirchoff discovered two laws of electricity that pertain to the conservation of charge and energy. To fully describe these laws, we must first discuss voltage, resistance, and current. Current is the flow of electricity; often you can compare it to the flow of water. As a current passes across a conductor, it encounters resistance. Ohm's law states that the product of the resistance R and current I across a conductor equals a drop in voltage V, i.e. $RI = V$. If the voltage remains constant, then a large resistance corresponds to a small current. A resistor is an object with high resistance which is placed in an electrical system to slow down the flow (current) of electricity. Resistors are measured in terms

> Ohm's law:
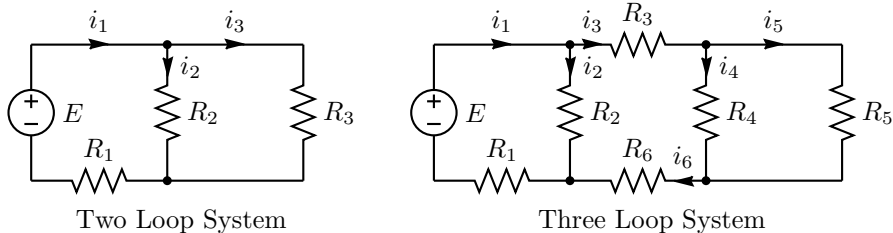> Voltage $=$ Resistance $\times$ Current
> $$V = RI$$

Figure 2.3: Electrical Circuit Diagrams.

of ohms, and the larger the ohms, the smaller the current. Figure 2.3 illustrates two introductory electrical systems.

In this diagram, wires meet at nodes (illustrated with a dot). Batteries and voltage sources (represented by $\ominus$ or other symbols) supply a voltage of E volts. At each node the current may change, so the arrows and letters i represent the different currents in the electrical system. The electrical current on each wire may or may not follow the arrows drawn (a negative current means that the current flows opposite the arrow). Resistors are depicted with the symbol $-\!\!\bigvee\!\!\!\bigwedge\!\!\!-$, and the letter R represents the ohms.

Kirchoff discovered two laws which help us find the currents in a system, provided we know the voltage of any batteries and the resistance of any resistors.

1. Kirchoff's current law states that at every node, the current flowing in equals the current flowing out (at nodes, current in = current out).

2. Kirchoff's voltage law states that on any loop in the system, the directed sum of voltages supplied equals the directed sum of voltage drops (in loops, voltage in = voltage out).

Let's use Kirchoff's laws to generate some equations for the two loop system.

1. First we will examine Kirchoff's current law. At the first node (top middle), current i_1 flows in while i_2 and i_3 flow out. Kirchoff's current law states that

$$i_1 = i_2 + i_3$$

At each node, current in equals current out.

or $i_1 - i_2 - i_3 = 0$. At the second node, both i_2 and i_3 are flowing in while i_1 flows out. This means that $i_2 + i_3 = i_1$ or $-i_1 + i_2 + i_3 = 0$. This second equation is the same as multiplying both sides of the first by -1 (so we say the 2nd equation depends on the first).

2. We now look at Kirchoff's voltage law. Pick a loop and work your way around the loop in a clockwise fashion. Each time you encounter a battery or resistor, include a term for the voltage supplied E on the left side of an equation, and the voltage drop (resistance times current Ri) on the right. If you encounter a battery or resistor as you work against the current, then times that term by -1.

The left loop has a battery with voltage E (voltage in). While moving along the wire containing i_i we encounter a resistor with resistance R_1 which contributes a voltage drop of $R_1 i_1$ volts. Along i_2 we encounter R_2 for a drop of $R_2 i_2$ volts. An equation for the first loop is

$$E = R_1 i_1 + R_2 i_2.$$

On loops, voltage in equals voltage out.

On the right loop we encounter along current i_3 a resistor with resistance R_3 ohms. While working our way against the arrow drawn on i_2, we

encounter an R_2 ohm resistor (hence we have to put a negative sign in front of $R_2 i_2$. There are no batteries on the second loop. The two resistors give us the equation

$$0 = -R_2 i_2 + R_3 i_3.$$

We can now write a system of equations involving the unknowns i_1, i_2, i_3, put it in matrix form, and then row reduce to solve

$$\begin{array}{rl} i_1 - i_2 - i_3 & = 0 \\ -i_1 + i_2 + i_3 & = 0 \\ R_1 i_1 + R_2 i_2 & = E \\ -R_2 i_2 + R_3 i_3 & = 0 \end{array} \xrightarrow{\text{matrix form}} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ -1 & 1 & 1 & 0 \\ R_1 & R_2 & 0 & E \\ 0 & -R_2 & R_3 & 0 \end{array}\right]$$

$$\xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \dfrac{ER_2 + ER_3}{R_1 R_2 + R_1 R_3 + R_2 R_3} \\ 0 & 1 & 0 & \dfrac{ER_3}{R_1 R_2 + R_1 R_3 + R_2 R_3} \\ 0 & 0 & 1 & \dfrac{ER_2}{R_1 R_2 + R_1 R_3 + R_2 R_3} \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The reason we have a row of zeros at the bottom of our system is because the two rows corresponding to the nodes are linearly dependent. When we reduce the matrix, this row dependence results in a row of zeros.

A similar computation can be done for the three loop system. There are 6 unknown currents, 4 nodes, and 3 loops. This will give us 7 equations with 6 unknowns. The 4 equations from the nodes will again contribute rows which are linearly dependent, which means you can always ignore an equation from one of the nodes. In electrical network problem, row reduction will always give a unique solution. In the homework, you are asked to setup systems of equations for various electrical systems, and then solve them.

Example 2.8. Let's look at an example which involves numbers. Suppose $E = 12$ (a 12 volt battery) and the resistors have $R_1 = 2, R_2 = R_3 = 4$ ohms. The top node gives the equation $i_1 = i_2 + i_3$ (remember flow in equals flow out). We'll skip the bottom node, as it contributes another row which is dependent on the first. The left loop gives the equation $12 = 2i_1 + 4i_2$, while the right loop gives the equation $0 = -4r_2 + 4r_3$. We write this in matrix form

$$\begin{array}{rl} i_1 - i_2 - i_3 & = 0 \\ 2i_1 + 4i_2 & = 12 \\ -4i_2 + 4i_3 & = 0 \end{array} \xrightarrow{\text{matrix form}} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 2 & 4 & 0 & 12 \\ 0 & -4 & 4 & 0 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 3/2 \\ 0 & 0 & 1 & 3/2 \end{array}\right]$$

which tells us the currents are $i_1 = 3$, $i_2 = 3/2$, and $i_3 = 3/2$.

2.4.1 Cramer's Rule

Cramer's rule is a theoretical tool which gives the solution to any linear system $A\vec{x} = \vec{b}$ with n equations and n unknowns, provided that there is a unique solution. Let $D = \det(A)$. Let D_i be the determinant of the matrix formed by replacing the ith column of A with $\vec{b}$. Then Cramer's rule states that

$$x_1 = \frac{D_1}{D}, x_2 = \frac{D_2}{D}, \dots, x_n = \frac{D_n}{D}.$$

We may prove it in class with pictures which connect determinants to area (eventually I'll add this to an appendix). This method of solving a system of

equations is quickly doable for 2 by 2 and 3 by 3 systems, but becomes computationally inefficient beyond (as computing determinants is time consuming and numerically unstable on large matrices). For large systems, it is better to use Gaussian elimination. However, Cramer's rule is a powerful theoretical tool, and can simplify symbolic computations for small systems. Let's look at a few examples.

Cramer's rule works great on 2 by 2 and 3 by 3 systems, where determinants are easy to compute.

Example 2.9. Let's solve $\begin{bmatrix} 1 & 2 & 0 \\ -2 & 0 & 1 \\ 0 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ using Cramer's rule.

We compute the determinant of the coefficient matrix first to obtain

$$D = \begin{vmatrix} 1 & 2 & 0 \\ -2 & 0 & 1 \\ 0 & 3 & -2 \end{vmatrix} = -11.$$

Next we replace each column of the coefficient matrix with the right column of our augmented system and compute the three determinants to obtain

$$\begin{vmatrix} 2 & 2 & 0 \\ -2 & 0 & 1 \\ 1 & 3 & -2 \end{vmatrix} = -12, \begin{vmatrix} 1 & 2 & 0 \\ -2 & -2 & 1 \\ 0 & 1 & -2 \end{vmatrix} = -5, \begin{vmatrix} 1 & 2 & 2 \\ -2 & 0 & -2 \\ 0 & 3 & 1 \end{vmatrix} = -2.$$

Cramer's rule requires that we divide each of these determinants by the original determinant, giving the solution $x_1 = 12/11, x_2 = 5/11, x_3 = 2/11$. Using Gaussian Elimination, we obtain the same solution

$$\begin{bmatrix} 1 & 2 & 0 & | & 2 \\ -2 & 0 & 1 & | & -2 \\ 0 & 3 & -2 & | & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & | & 12/11 \\ 0 & 1 & 0 & | & 5/11 \\ 0 & 0 & 1 & | & 2/11 \end{bmatrix},$$

however the arithmetic involved in keeping track of fractions or really large integers becomes much more difficult by hand without Cramer's rule.

Example 2.10. Consider the electrical system in Example 2.8, where $E = 12$, $R_1 = 2$, and $R_2 = R_3 = 4$. The corresponding augmented matrix we used to solve this system was

$$\begin{bmatrix} 1 & -1 & -1 & | & 0 \\ 2 & 4 & 0 & | & 12 \\ 0 & -4 & 4 & | & 0 \end{bmatrix}, A = \begin{bmatrix} 1 & -1 & -1 \\ 2 & 4 & 0 \\ 0 & -4 & 4 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix}, D = \begin{vmatrix} 1 & -1 & -1 \\ 2 & 4 & 0 \\ 0 & -4 & 4 \end{vmatrix} = 32.$$

We now replace each column with $\vec{b}$ and compute the determinant of the corresponding matrix (remember to cofactor along the column which contains $0, 12, 0$ to do this quickly)

$$D_1 = \begin{vmatrix} 0 & -1 & -1 \\ 12 & 4 & 0 \\ 0 & -4 & 4 \end{vmatrix} = 96, D_2 = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 12 & 0 \\ 0 & 0 & 4 \end{vmatrix} = 48, D_3 = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 4 & 12 \\ 0 & -4 & 0 \end{vmatrix} = 48.$$

Dividing each of these by the determinant of the original matrix gives the solution $i_1 = 96/32 = 3$, $i_2 = i_3 = 48/32 = 3/2$, which matches the solution we found using row reduction in the previous section.

2.5 Fitting Curves to Data

Our discussion and examples in this section will involve data in the xy plane ($\mathbb{R}^2$). However, the concepts and formulas generalize to any number of dimensions

2.5.1 Interpolating Polynomials

Through any two points (with different x-values) there is a unique line of the form $y = mx + b$. If you know two points, then you can use them to find the values m and b. Through any three points (with different x-values) there is a unique parabola of the form $y = ax^2 + bx + c$, and you can use the three points to find the values a, b, and c. The pattern holds: for any $n + 1$ points (with different x-values), there is a unique polynomial (called an *interpolating polynomial*) of degree n which passes through those points. In this section, we will illustrate how to find interpolating polynomials from a list of points, and show how the solution requires solving a linear system.

> For any $n + 1$ points (with different x-values), there is a unique polynomial of degree n which passes through those points.

To organize our work, let's first standardize the notation. Rather than writing $y = mx + b$, let's write $y = a_0 + a_1 x$ (where $a_0 = b$ and $a_1 = m$). For a parabola, let's write $y = a_0 + a_1 x + a_2 x^2 = \sum_{k=0}^{2} a_k x^k$. We can now write any polynomial in the form

> Standardizing the notation makes a problem easier to generalize.

$$y = a_0 + a_1 x + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^k.$$

By standardizing the coefficients, we can use summation notation to express a polynomial of any degree by changing the n on the top of the summation sign. Now that our notation is organized, let's use it to find a polynomial through 3 points.

Example 2.11. Let's find a parabola through the three points $(0, 1)$, $(2, 4)$, $(4, 5)$. The polynomial is $y = a_0 + a_1 x + a_2 x^2$ and our job is to find the three constants a_0, a_1, a_2. Since we have three points, we put these points into the equation to obtain a system of three equations:

$$a_0 = 1, \qquad a_0 + 2\, a_1 + 4\, a_2 = 4, \qquad a_0 + 4\, a_1 + 16\, a_2 = 5.$$

This is a linear system with 3 equations and 3 unknowns. We now write the system in matrix form and reduce it

$$\begin{bmatrix} 1 & 0 & 0 & | & 1 \\ 1 & 2 & 4 & | & 4 \\ 1 & 4 & 16 & | & 5 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & | & 1 \\ 0 & 1 & 0 & | & 2 \\ 0 & 0 & 1 & | & -1/4 \end{bmatrix}.$$

The reduced row echelon form tells us that the coefficients are $a_0 = 1$, $a_1 = 2$, $a_2 = -1/4$, which means our parabola is $y = 1 + 2x - \frac{1}{4}x^2$. Cramer's rule gives $D = 16$, $D_1 = 16$, $D_2 = 32$, $D_3 = -4$, so $a_0 = 16/16 = 1$, $a_1 = 32/16 = 2$, and $a_2 = -4/16 = -1/4$, confirming our result from Gaussian elimination.

Once you have obtained the interpolating polynomial, you can always check your work by plugging the points into the interpolating polynomial. When $x = 0$, we have $y = 1 + 0 + 0 = 1$; when $x = 2$, we have $y = 1 + 2 - 1 = 2$; and when $x = 4$, we have $y = 1 + 8 - 4 = 5$. In other words, the parabola passes through the three points $(0, 1)$, $(2, 4)$, and $(4, 5)$, as needed. This parabola is shown in the margin.
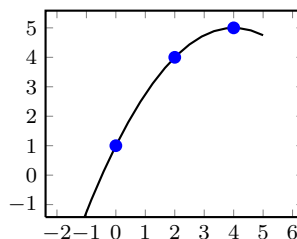
> Check your work by plugging the points back into your solution.

In the example above, notice that powers of x appear as the coefficients in our coefficient matrix, and we augment that matrix by the y values. This is the general pattern for finding an interpolating polynomial. The diagram below shows the general method for finding an interpolating polynomial through three points. Note that each entry of the first column of the matrix is 1 since $x_1^0 = 1$, $x_2^0 = 1$, and $x_3^0 = 1$.

$$(x_1, y_1), (x_2, y_2), (x_3, y_3)$$

$$\begin{bmatrix} 1 & x_1^1 & x_1^2 & y_1 \\ 1 & x_2^1 & x_2^2 & y_2 \\ 1 & x_3^1 & x_3^2 & y_3 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & a_0 \\ 0 & 1 & 0 & a_1 \\ 0 & 0 & 1 & a_2 \end{bmatrix}$$

$$y = a_0 + a_1 x + a_2 x^2$$

Finding an interpolating polynomial through four points is very similar—just add one more row and column to the matrix and repeat the process.

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$$

$$\begin{bmatrix} 1 & x_1^1 & x_1^2 & x_1^3 & y_1 \\ 1 & x_2^1 & x_2^2 & x_2^3 & y_2 \\ 1 & x_3^1 & x_3^2 & x_3^3 & y_3 \\ 1 & x_4^1 & x_4^2 & x_4^3 & y_4 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & a_0 \\ 0 & 1 & 0 & 0 & a_1 \\ 0 & 0 & 1 & 0 & a_2 \\ 0 & 0 & 0 & 1 & a_3 \end{bmatrix}$$

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

This pattern generalizes to all dimensions (and the corresponding algorithm is coded into spreadsheet programs such as Microsoft Excel and OpenOffice Calc, as well as many mathematics and statistics software packages). Remember that the x-values must be different (since a function requires only one output for each input). Once you have obtained your solution, remember that you can easily check if your solution is correct by plugging the points into your polynomial.

2.5.2   Least Squares Regression

Interpolating polynomials give a polynomial which passes through every point in a set of data. While they pass through every point in a set of data, the more points the polynomial must pass through, the more the polynomial may have to make large oscillations in order to pass through each point. Furthermore, often data points from real-life measurements have error associated with them, so we don't necessarily want to mandate that our function go through the data point. Often we just need a simple line or parabola that passes near the points and gives a good approximation of a trend.

Here's an example that saved me money when buying a minivan. When I needed to purchase a minivan for my expanding family, I gathered mileage and price data for about 40 cars from the internet. I plotted this data and discovered an almost linear downward trend (as mileage increased, the price dropped). I created a line to predict the price of a car based on mileage. I then used this data to talk the dealer into dropping the price of the car I wanted by over $1000.

How do you find a line that is "closest" to a set of points, and is the best line to approximate your data? Finding an equation of this line, called the *least squares regression* line, is the content of this section. The least squares regression line is used to find trends in many branches of science and business. Let's introduce the idea with an example.

Example 2.12. Find a line that is closest to passing through the three points $(0, 1)$, $(2, 4)$, and $(4, 5)$. Since the points are not collinear, there is not a line through all three points. Suppose for a moment that there was a line of the form $y = mx + b = a_0 + a_1 x$ that did pass through the points. Plugging our 3 points into the equation $a_0 + a_1 x = y$ gives the system of equations

collinear means lie on the same line

$$\begin{cases} a_0 = 1 \\ a_0 + 2a_1 = 4 \\ a_0 + 4a_1 = 5 \end{cases} \xrightarrow{\text{augmented matrix}} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 5 \end{bmatrix} \xrightarrow{\text{matrix eqn}} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}.$$

Notice that the system can be written in matrix form $A\vec{x} = \vec{b}$, where A contains a column of 1's and x-values, $\vec{x}$ is the vector of coefficients a_i, and $\vec{b}$ is a column of y-values. If you try to reduce this matrix, you will discover the system is inconsistent (has no solution), which should not be a surprise since there is no line which passes through these three points. Notice that there are more equations (3 equations) than variables (a_0 and a_1), which means the system is overdetermined.

Recall that if a system has more equations than variables, we say the system is overdetermined.

While there is no solution, can we still use our matrix equation to find a solution? Is there a way to reduce the number of rows in our system, so that the resulting system has only 2 rows? If we multiply on the left by a 2 by 3 matrix, we would obtain a system with 2 rows instead of 3, and the rows of the new matrix would be linear combinations of the rows of our original matrix. The only 2 by 3 matrix in this problem is the transpose of A. So let's multiply both sides of the matrix equation by the transpose of A, and see what happens:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix},$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix},$$

$$A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 28 \end{bmatrix}.$$

Multiplying both sides of the equation $A\vec{x} = \vec{b}$ on the left by A^T gives us the equation $A^T A \vec{x} = A^T \vec{b}$, or

Multiply both sides by A^T, the transpose of A, to obtain a system that has a solution.

$$\begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 28 \end{bmatrix}.$$

This is a system of 2 equations with 2 unknowns, and it has a unique solution. Reducing $\begin{bmatrix} 3 & 6 & | & 10 \\ 6 & 20 & | & 28 \end{bmatrix}$ to $\begin{bmatrix} 1 & 0 & | & 4/3 \\ 0 & 1 & | & 1 \end{bmatrix}$ means the solution is $y = \frac{4}{3} + x$ (illustrated in the margin).

We will see exactly why this works in Section 2.6 when we start looking at the columns of our matrix as a basis for a vector space and asking questions about linear combinations, spans, and projections. For now, let's just use the transpose to solve least squares regression problems.

In general, a least squares regression problem requires the following:

1. Assume the form of a solution, such as $y = a_0 + a_1 x$ for a line.

2. Plug in values to get a matrix equation $A\vec{x} = \vec{b}$.

3. Multiply both sides by A^T.

4. Solve the simplified system (elimination, Cramer's rule, an inverse, or other more sophisticated methods).

When trying to find a least squares regression line, the simplified system will have a 2 by 2 coefficient matrix, so Cramer's rule provides an extremely quick solution.

Example 2.13. Let's find the least squares regression line which passes nearest the four points $(0,0)$, $(-1,2)$, $(-2,4)$, and $(0,-1)$. We are after an equation of

the form $y = a_0 + a_1 x$. The four points give the equations $a_0 = 0$, $a_0 - a_1 = 2$, $a_0 - 2a_1 = 4$, and $a_0 = -1$. In matrix form, we write

$$\begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ -1 \end{bmatrix}, \text{ where } A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -2 \\ 1 & 0 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ -1 \end{bmatrix},$$

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & 0 \end{bmatrix}, \text{ so } A^T A = \begin{bmatrix} 4 & -3 \\ -3 & 5 \end{bmatrix}, \quad A^T \vec{b} = \begin{bmatrix} 5 \\ -10 \end{bmatrix}.$$

Now solve the matrix equation $A^T A \vec{x} = \begin{bmatrix} 4 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 5 \\ -10 \end{bmatrix} = A^T \vec{b}$.

Cramer's rule gives the solution as

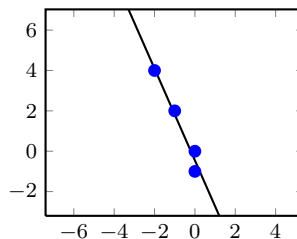$$a_0 = \frac{\begin{vmatrix} 5 & -3 \\ -10 & 5 \end{vmatrix}}{\begin{vmatrix} 4 & -3 \\ -3 & 5 \end{vmatrix}} = \frac{-5}{11}, \qquad a_1 = \frac{\begin{vmatrix} 4 & 5 \\ -3 & -10 \end{vmatrix}}{\begin{vmatrix} 4 & -3 \\ -3 & 5 \end{vmatrix}} = \frac{-25}{11}.$$

The least square regression line is $y = -\frac{5}{11} - \frac{25}{11}x$ (illustrated in the margin).

2.5.3   A Quick 2 by 2 Inverse

Finding the least square regression line requires solving the system $A^T A \vec{x} = A^T \vec{b}$ for $\vec{x}$. Symbolically we can solve this system by multiplying both sides on the left by $(A^T A)^{-1}$ (this inverse will exist), giving the solution

$$\vec{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = (A^T A)^{-1} A^T \vec{b}.$$

For 2 by 2 matrices, there is quick way to find the inverse. To find the inverse of a 2 by 2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we need to reduce $\left[\begin{array}{cc|c} a & b & 1 \\ c & d & 0 \end{array}\right]$ to find the first column of the inverse, and $\left[\begin{array}{cc|c} a & b & 0 \\ c & d & 1 \end{array}\right]$ to find the second column. Cramer's rule gives the formula

$$A^{-1} = \begin{bmatrix} \begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix}/|A| & \begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix}/|A| \\[2mm] \begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix}/|A| & \begin{vmatrix} a & 0 \\ c & 1 \end{vmatrix}/|A| \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

To find the inverse of a 2 by 2,

1. switch the diagonal entries,
2. change the sign on the off-diagonals, and
3. divide by the determinant.

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

To find the inverse of a 2 by 2 matrix, just interchange the diagonal entries, change the sign on the others, and divide by the determinant.

Example 2.14. $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}^{-1} = \frac{1}{-2} \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$, and $\begin{bmatrix} 13 & 3 \\ -7 & 2 \end{bmatrix}^{-1} = \frac{1}{47} \begin{bmatrix} 2 & -3 \\ 7 & 13 \end{bmatrix}$.

Example 2.15. Let's repeat the introductory example (2.12) from the last section by using an inverse to find the line closest to passing through the three

points $(0, 1)$, $(2, 3)$, and $(4, 6)$. The matrix equation and relevant matrices are

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix},$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix},$$

$$A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 10 \\ 28 \end{bmatrix}.$$

The inverse of $(A^T A)$ is $\dfrac{1}{60 - 36} \begin{bmatrix} 20 & -6 \\ -6 & 3 \end{bmatrix}$, which means our solution is

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = (A^T A)^{-1} A^T b = \frac{1}{24} \begin{bmatrix} 20 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ 28 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 32 \\ 24 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 1 \end{bmatrix}.$$

We of course obtained the same solution $y = \frac{4}{3} + x$.

One advantage of using an inverse to solve a least square regression problem is that the product $W = (A^T A)^{-1} A^T$ can be computed once, and then solutions to similar least squares regression problems are found by the simple product $\vec{x} = W \vec{b}$. A different set of data points where the x-values remain the same but the y-values change can then be solved by just changing $\vec{b}$ and using the same results as before.

Example 2.16. Let's find the least square regression line for the two data sets

1. $(0, 2)$, $(2, 1)$, and $(4, 3)$

2. $(0, 6)$, $(2, 3)$, and $(4, -1)$

Notice that both of the data sets have the same x values as the previous example. Hence we can still use

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix}, \quad (A^T A)^{-1} = \frac{1}{24} \begin{bmatrix} 20 & -6 \\ -6 & 3 \end{bmatrix}.$$

The change occurred with the y-values, which means that $\vec{b}$ changed for each problem. Before solving either, we compute

$$W = (A^T A)^{-1} A^T = \frac{1}{24} \begin{bmatrix} 20 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 20 & 8 & -4 \\ -6 & 0 & 6 \end{bmatrix}.$$

We can now solve both problems rapidly by multiplying W and $\vec{b}$.

1. The y values are $2, 1, 3$, so we compute

$$\frac{1}{24} \begin{bmatrix} 20 & 8 & -4 \\ -6 & 0 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 36 \\ 6 \end{bmatrix}$$

which means that our line is (after simplifying fractions) $y = \frac{3}{2} + \frac{1}{4}x$.
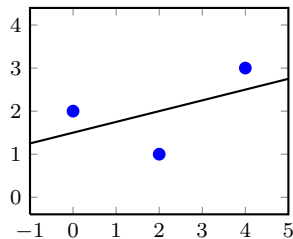
2. The y values are $6, 3, -1$, so we compute

$$\frac{1}{24}\begin{bmatrix} 20 & 8 & -4 \\ -6 & 0 & 6 \end{bmatrix}\begin{bmatrix} 6 \\ 3 \\ -1 \end{bmatrix} = \frac{1}{24}\begin{bmatrix} 148 \\ -42 \end{bmatrix}$$

which means that our line is (after simplifying fractions) $y = \frac{37}{6} - \frac{7}{4}x$.

Least squares regression is not limited to just finding a best fit line. We can use the same ideas to find a best fitting parabola through many points. To do this, we would start by assuming the data fit a parabola $y = a_0 + a_1x + a_2x^2$, plug in the x and y values to obtain a linear system with 3 unknowns, write the system in matrix form (the columns of the matrix would involve ones, x values, and x^2 values), multiply both sides by the transpose to obtain a system that was not overdetermined (3 by 3), and then solve the system. Most spreadsheet programs have this feature built in for any size polynomial you choose.
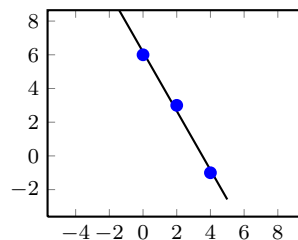
2.6   Vector Spaces

To understand why we need to use the transpose of A when doing regression, we need to introduce some of the most fundamental, powerful, and far-reaching ideas in linear algebra: vector spaces. We will study these ideas more deeply throughout the semester.

In this section, we will define vector spaces and look at particular examples like the column space and null space of a matrix. We will find that the solution to the linear regression problem is very naturally phrased using this language. Along the way, we will obtain geometric interpretations of these ideas in preparation for a more general approach to vectors which is coming in the next chapter.

2.6.1   The Vector Space $\mathbb{R}^n$

We have already been using 2D and 3D vectors quite a bit prior to now. The symbol $\mathbb{R}^n$ represents the set of all vectors in n-dimensional space. What does the word dimension really mean? We'll answer that now by looking at an example.

Example 2.17. Recall that the span of a set of vectors is the set of all possible linear combinations of those vectors. So the span of the vector $(2, 1)$ in $\mathbb{R}^2$ is the set of all vectors of the form $c(2, 1) = (2c, c)$, which represents a line through the origin with slope $1/2$. Since $(0, 1)$ is not a multiple of $(2, 1)$, the vector $(0, 1)$ is not in the span of $(2, 1)$ (i.e., $(0, 1)$ does not lie on the line $y = \frac{1}{2}x$). In other words, the set $\{(2, 1), (0, 1)\}$ is linearly independent (neither vector is in the span of the other). The span of $\{(2, 1), (0, 1)\}$ is larger than the line—it is the entire plane. Every other vector in the plane is a linear combination of these two vectors since they form a linearly independent set. You can check this by trying to find the linear combination to equal an arbitrary vector (x, y):

$$\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \xrightarrow{\text{augmented matrix}} \left[\begin{array}{cc|c} 2 & 0 & x \\ 1 & 1 & y \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{cc|c} 1 & 0 & \frac{x}{2} \\ 0 & 1 & y - \frac{x}{2} \end{array}\right].$$

So, for example, the vector $(x, y) = (3, 5)$ is the linear combination $\frac{x}{2}(2, 1) + (y - \frac{x}{2})(0, 1) = \frac{3}{2}(2, 1) + (5 - \frac{3}{2})(0, 1)$. Notice that since the first two columns are always pivot columns (since the two vectors form a linearly independent set), the system will always have a single solution. We call the scalars in the linear combination ($\frac{x}{2}$ and $y - \frac{x}{2}$) the *coordinates* of the vector relative to the

basis $\{(2,1),(0,1)\}$. For example, the coordinates of $(3,5)$ relative to the basis $\{(2,1),(0,1)\}$ are $\frac{3}{2}$ and $5 - \frac{3}{2}$. A basis gives us a way of writing any vector by writing the coordinates.

Here is the crucial point about dimension: because we were able to obtain the entire plane as the span of *two* linearly independent vectors, we say the dimension of $\mathbb{R}^2$ is *two*.

Notice that the dimension does not depend on the particular two vectors I chose. The only requirement was that the two vectors form a linearly independent set. Then the first two columns in the above rref will always be pivot columns, so I will always be able to write any vector in the plane as a linear combination of my two linearly independent vectors. We call a set of two linearly independent vectors, like $\{(2,1),(0,1)\}$, a *basis* for $\mathbb{R}^2$.

The *dimension* of a vector space (like $\mathbb{R}^2$) is the number of linearly independent vectors needed to span the vector space. A basis for a vector space is a set of vectors that are:

The dimension of a vector space is the number of vectors in a basis.

- linearly independent and

- span the vector space.

If a vector space has dimension n, then any n linearly independent vectors will be a basis for the vector space.

For $\mathbb{R}^n$, the *standard basis vectors* are

standard basis vectors

$$\vec{e}_1 = (1,0,0,\ldots,0),\ \vec{e}_2 = (0,1,0,\ldots,0),\ \ldots,\ \vec{e}_n = (0,0,0,\ldots,1).$$

Every other vector in $\mathbb{R}^n$ is a linear combination of these standard basis vectors. If not specified otherwise, we assume that any coordinates are relative to this standard basis. We will find that using a different basis for our coordinates can simplify various problems. For example, we'll often use eigenvectors as some of our basis vectors.

2.6.2 Vector Subspaces of $\mathbb{R}^n$

The span of the vector (1,2,0) can be represented as a line through the origin in 3D. The span of the vectors $\{(1,2,0),(0,0,1)\}$ represents a plane through the origin. These spans are examples of vector spaces.

A set of vectors V from $\mathbb{R}^n$ is called a *vector subspace* of $\mathbb{R}^n$ if any linear combination of vectors in V is also in V. We can break this definition down into two parts. The set of vectors V is a vector subspace of $\mathbb{R}^n$ if:

vector subspace

1. The sum of any two vectors in the set is still in the set. In notation, we write for any $\vec{u}, \vec{v} \in V$ the sum $\vec{u} + \vec{v} \in V$. We say "V is closed under addition."

2. Any scalar multiple of a vector is still in the set. In notation, we write for any $\vec{v} \in V$ and $c \in \mathbb{R}$, the vector $c\vec{v} \in V$. We say "V is closed under scalar multiplication."

The span of a set of vectors is always a vector subspace. The simplest way to show that a set of vectors is a subspace is to show that it is a span of some vectors.

A basis of vectors for a subspace V is a set of linearly independent vectors whose span is the set V. The dimension of a subspace is the number of vectors in a basis. Lines through the origin are one-dimensional subspaces, planes through the origin are two-dimensional subspaces, and so on.

Example 2.18. Here are three examples of vector subspaces.

1. The set of vectors $V = \{(2c, 3c) \mid$ where c is any real number$\}$ is a vector subspace of $\mathbb{R}^2$. It represents a line through the origin. A basis for V is $\{(2, 3)\}$, since every vector in the subspace is a scalar multiple of $(2, 3)$. Since there is one vector in the basis, V is one-dimensional.

2. The set of vectors $V = \{(2c, c + d, 4d) \mid c, d \in \mathbb{R}\}$ is a two-dimensional vector subspace of $\mathbb{R}^3$. It represents a plane through the origin. We can rewrite any vector in V in the form $c(2, 1, 0) + d(0, 1, 4)$, so $V = \text{span}\{(2, 1, 0), (0, 1, 4)\}$. Since these two vectors are linearly independent, the set $\{(2, 1, 0)\,(0, 1, 4)\}$ is a basis for the subspace, and the space is two-dimensional.

3. The set of vectors $V = \{(c + 2e, 2c + 4e, d - 2e) \mid c, d, e \in \mathbb{R}\}$ is a two-dimensional vector subspace of $\mathbb{R}^3$. You might first think this subspace is three-dimensional because you can choose 3 different numbers (c, d, and e) and we can write any vector in the form $c(1, 2, 0) + d(0, 0, 1) + e(2, 4, -2)$. However, the three vectors are linearly dependent. We can see this from the rref:

$$\begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ 0 & 1 & -2 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Because the third column is not a pivot column, the third vector depends on the first two vectors. Since the first two columns are pivot columns, the first two vectors form a linearly independent set, so a basis for V is simply $\{(1, 2, 0), (0, 0, 1)\}$. That is why the space is actually two-dimensional (it is a plane through the origin).

Example 2.19. Here are three examples of sets that are *NOT* vector subspaces.

1. The set of vectors on the positive x-axis, $V = \{(a, 0) \mid a \geq 0\}$, is not a vector subspace because if you multiply vector in V by -1, the result is not in V.

2. The set of vectors in $\mathbb{R}^2$ whose magnitude is less than one (i.e., any vector from the origin to a point inside the unit circle) is not a vector subspace. One reason why is that $(1/2, 0)$ is in the set, but $4(1/2, 0) = (2, 0)$ is outside the circle and not in the set.

3. The set of vectors in $\mathbb{R}^2$ which are on either the x or y-axis is not a vector subspace. For example, both $(1, 0)$ and $(0, 1)$ are in the set, but the sum $(1, 0) + (0, 1) = (1, 1)$ is not on either axis, so is not in set.

Now let's look at some of the most common vector subspaces that come up in practice.

The column and row space of a matrix

Remember that a matrix can be viewed as a collection of column vectors or as a collection of row vectors. The span of the columns of a matrix is called the *column space* of the matrix, denoted $\text{col}(A)$. The span of the rows of the matrix is called the *row space* of the matrix, denoted $\text{row}(A)$. The reduced row echelon form of a matrix tells us how to find a basis for column space and how to find a basis for the row space.

column space

row space

Example 2.20. Let's find a basis for the column space of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Of course, every vector in the column space is already a linear combination of the three column vectors (by definition, since the column space is the span of the columns). If the columns are linearly independent, then they would form a basis for the column space. However, from the reduced row echelon form, we see that the third column is a linear combination of the first two. A basis for the column space is $\{(1, 2, 0), (2, 0, 3)\}$ (i.e., the pivot columns). The coordinates of $(3, -2, 6)$ relative to this basis are -1 and 2. In other words, we can write $(3, -2, 6)$ as the linear combination

$$\begin{bmatrix} 3 \\ -2 \\ 6 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

We can write all three columns of the matrix in terms of the basis vectors as

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

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

Each column of a matrix is a linear combination of the pivot columns. The coordinates of each column vector are the numbers in the corresponding column of the rref (excluding zero rows).

We see from the example that we can write any matrix as the product of its pivot columns and the nonzero rows of its rref. We can factor any matrix A as $A = CR$, where C is the matrix of pivot columns of A and R is the matrix whose rows are the nonzero rows from $\text{rref}(A)$.

Example 2.21. Let's find a basis for the row space of the same matrix,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

We will see that a basis is the nonzero rows of the rref. Using $A = CR$ as above, and remembering that multiplication can be thought of as linear combinations of rows, we can write

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} + 2 \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \\ 2 \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} + 0 \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \\ 0 \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} + 3 \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \end{bmatrix}.$$

This shows us that every row of the original matrix is a linear combination of the nonzero rows of rref. The row space is the span of the rows, and since the rows of A depend on the nonzero rows of rref, we can span the row space using these nonzero rows. Since the nonzero rows of the rref are linearly independent (placing these rows into columns and reducing won't change the pivots), we know that the nonzero rows of rref form a basis for the row space.

Each row is a linear combination of the nonzero rows of rref. The coordinates of each row are found in the pivot columns.

Since the number of pivots (the rank of the matrix) equals the number of nonzero rows in the rref, we can see that the size of the basis for the column space and the size of the basis for the row space are equal. Therefore the dimensions of the column space and the row space are equal (and are equal to the rank of the matrix).

Let's summarize what we learned in the previous examples.

1. The pivot columns are a basis for the column space.

2. The nonzero rows of the rref are a basis for the row space.

3. Every matrix can be decomposed as $A = CR$, where C is a matrix of the pivot columns and R is the nonzero rows of the rref.

4. The dimension of the column space equals the dimension of the row space, which equals the rank of the matrix.

The transpose of A interchanges the rows and columns, so the row space of A^T is the same as the column space of A, and the column space of A^T is the same as the row space of A.

Example 2.22. Let's use the transpose to find different bases for the column space and row space of the same matrix from the previous examples:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 3 \\ 3 & -2 & 6 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 3/2 \\ 0 & 1 & -3/4 \\ 0 & 0 & 0 \end{bmatrix}.$$

The set of pivot columns of A^T, $\{(1, 2, 3), (2, 0, -2)\}$, is a basis for the column space of A^T and the row space of A. The coordinates of the third column of A^T (the third row of A), $(0, 3, 6)$, relative to this basis are $(3/2, -3/4)$.

The nonzero rows of the rref of A^T, namely $\{(1, 0, 3/2), (0, 1, -3/4)\}$ are a basis for the row space of A^T and the column space of A. Every column of A can be written as a linear combination of these two vectors. The coordinates of $(1, 2, 0)$ relative to this basis are $(1, 2)$. The coordinates of $(2, 0, 3)$ are $(2, 0)$. The coordinates of $(3, -2, 6)$ are $(3, -2)$.

In the last few examples, we have two different ways to obtain a basis for the column space and row space of a matrix. Table 2.2 provides graphs of all 4 bases. You can either reduce A or A^T, and use the pivot columns and/or nonzero rows of rref to obtain a basis. Which basis is preferred? The short answer is "both." Each basis has a different use. We'll revisit this topic more later.

> Recall that the coordinates of a vector relative to a basis are the scalars you need to recreate the vector using a linear combination of the basis vectors. Hence the coordinates of $(0, 3, 6)$ relative to the basis $\{(1, 2, 3), (2, 0, -2)\}$ are $(3/2, -3/4)$ since
> $$\begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}$$

The null space of a matrix

The column and row spaces provide two examples of vector subspaces. These two spaces are defined as the span of a set of vectors, hence are automatically closed under addition and scalar multiplication. We'll now explore another set of vectors, called the null space or kernel of a matrix, where it is not immediately obvious that the set is closed under addition and scalar multiplication.

The *null space* (or *kernel*) of a matrix A is the set of vectors $\vec{x}$ that satisfy the equation $A\vec{x} = \vec{0}$. Notice that the zero vector is always in this set. If there are no other vectors in this set, then we say that null space is the trivial vector space and has dimension zero.

> null space=null$(A)=\{\vec{x} \mid A\vec{x} = \vec{0}\}$

The null space is closed under addition and scalar multiplication. If both $\vec{x}$ and $\vec{y}$ are in the null space, then $A\vec{x} = \vec{0}$ and $A\vec{y} = \vec{0}$. The sum $\vec{x} + \vec{y}$ satisfies

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} \qquad \text{(distribute)}$$
$$= \vec{0} + \vec{0} = \vec{0} \qquad (A\vec{x} = \vec{0} = A\vec{y})$$

This shows that $\vec{x} + \vec{y}$ is also in the null space. To show that the null space is closed under scalar multiplication, we observe that if $\vec{x}$ is in the null space and c is any real number, then

$$A(c\vec{x}) = cA\vec{x} = c\vec{0} = \vec{0},$$

Visualizing two different bases for the column and row space.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 0 & 3 & 6 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$
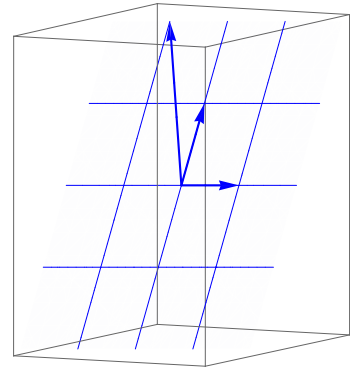
$$A^T = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 3 \\ 3 & -2 & 6 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 3/2 \\ 0 & 1 & -3/4 \\ 0 & 0 & 0 \end{bmatrix}$$

Bases for the column space of A, $\text{col}(A)$

Using Pivot Columns of A	Using nonzero rows of rref(A^T)	Both
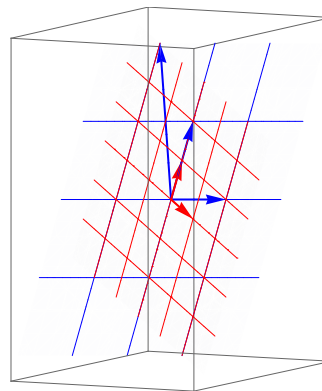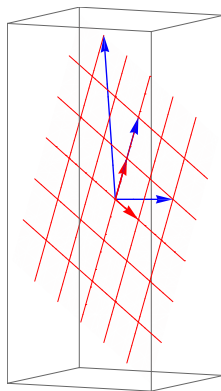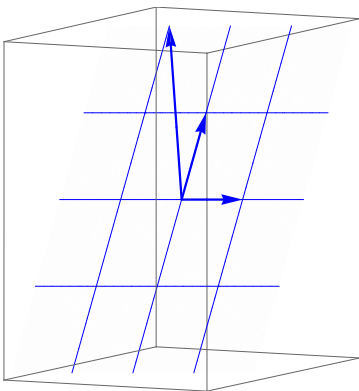
Bases for the row space of A, $\text{row}(A)$

Using Pivot Columns of A^T	Using nonzero rows of rref(A)	Both
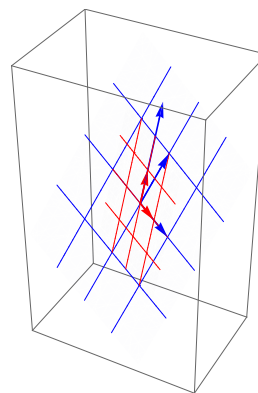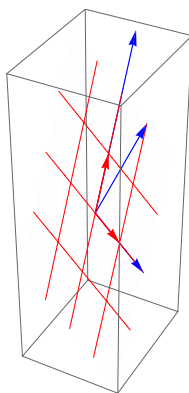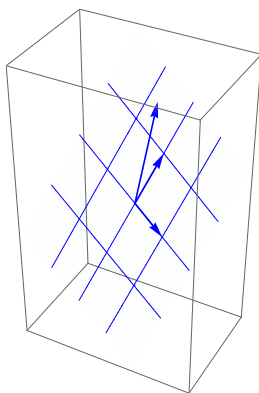
Table 2.2:   In examples 2.20, 2.21, and 2.22, we found two different bases for the column and row space. This table shows graphs to help us visualize the differences between these bases. In all the graphs, the blue vectors represent columns or rows in the matrix and the red vectors represent the vectors in the basis. The top row of graphs shows us visualizations of column space where (1) we use the pivot columns to create our basis, (2) we use the nonzero rows of the rref of A^T, and (3) both bases are drawn on top of each other. The bottom row of graphs repeats this visualization for the row space.

which shows that the null space is closed under scalar multiplication.

We have already seen the null space. Finding eigenvectors requires that we solve $(A - \lambda I)\vec{x} = 0$, so eigenvectors come from the null space of $A - \lambda I$. When we compute the null space of $A - \lambda I$, we call it the *eigenspace* of A for eigenvalue λ.

Example 2.23. Let's redo example 1.25 related to eigenvectors, but this time use the vocabulary of null spaces. The matrix $A = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 4 \\ 0 & 0 & 1 \end{bmatrix}$ has eigenvalues $1, 1, 3$. When $\lambda = 1$, we compute $A - \lambda I = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$. To find the null space, we solve $(A - \lambda I)\vec{x} = 0$. Row reduction (after augmenting the matrix) gives us the rref $\begin{bmatrix} 1 & 1 & 4 & | & 0 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$, which shows we have two free variables. The solution is $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}$. The null space is this entire solution set (called the eigenspace corresponding to $\lambda = 1$), and is a vector subspace of dimension 2. A basis for the null space is $\{(-1, 1, 0), (-4, 0, 1)\}$. Every nonzero vector in this null space is an eigenvector of A with eigenvalue 1. A similar computation (computing the null space for $A - 3I$) shows the eigenspace for $\lambda = 3$ is a one-dimensional vector space with basis $\{(1, 1, 0)\}$. Example 1.25 has all the details.

2.6.3 Why does regression require the transpose?

We finish this unit with a brief explanation of why we used the transpose to solve the least square regression problem. We will use the column space of A and the null space of A^T. We'll also focus on a specific example in 3D so we can draw pictures. Later, we will revisit this idea and develop everything in full generality along with a more powerful vocabulary and theory. Until then, we'll stick to a visual example.

Before getting started, recall that the dot product of two vectors is 0 if and only if the vectors meet at a 90 degree angle. When this occurs, we say that the vectors are orthogonal. How do you find all the vectors that are orthogonal to a set of vectors V? The answer is "Place all the vectors into the columns of A, and then find the null space of A^T." This null space is called the *orthogonal complement* of the set V.

The two vectors $(1, 1, 1)$ and $(-2, 3, 1)$ are orthogonal because their dot product is $-2 + 3 - 1 = 0$.

The orthogonal complement to a vector space V is the set of vectors that are orthogonal to every vector in the space. The orthogonal complement of the column space of A is the null space of A^T.

Example 2.24. Let's find all vectors that are orthogonal to $\vec{v} = (1, 2, -3)$. In other words, we want all vectors $\vec{u} = (u_1, u_2, u_3)$ such that

$$0 = (1, 2, -3) \cdot (u_1, u_2, u_3) = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \vec{v}^T \vec{u}$$

(where $\vec{v}^T$ is the row matrix with the single row $\vec{v}$). This is the same problem as finding the null space of the matrix $v^T = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix}$. Since the free variables are u_2 and u_3, we have the solution

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} -2u_2 + 3u_3 \\ u_2 \\ u_3 \end{bmatrix} = u_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

The set of vectors that are orthogonal to $\vec{v}$ is a vector space of dimension 2 with basis $\{(-2,1,0),(3,0,1)\}$. This is a plane passing through the origin.

If you are given a set of vectors $\vec{v}_1, \vec{v}_2$, how you do find all the vectors that are orthogonal to both? If you let $A = [\vec{v}_1 \ \vec{v}_2]$ (columns of A are the vectors), then a vector $\vec{x}$ is orthogonal to both $\vec{v}_1$ and $\vec{v}_2$ if $A^T\vec{x} = \vec{0}$. Hence the null space of the transpose of A is the set of vectors that are orthogonal to the columns of A.

We're now ready to understand linear regression.

Example 2.25. Recall example 2.12, in which we found the line closest to passing through the three points $(0,1)$, $(2,4)$, and $(4,5)$.

In order to find the closest line to several points, the first question is, "Is there a line that passes through all three points?" If there is a line $y = a_0 + a_1x$ that passes through all of the points, then we have a system of equations

$$\begin{array}{l} a_0 = 1 \\ a_0 + 2a_1 = 4 \\ a_0 + 4a_1 = 5 \end{array} \xrightarrow{\text{augmented matrix}} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 5 \end{array}\right] \xrightarrow{\text{matrix eqn}} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}.$$

Notice that the system can be written in matrix form $A\vec{x} = \vec{b}$, where A contains a column of 1's and x-values, $\vec{x}$ is the vector of coefficients a_i, and $\vec{b}$ is a column of y-values. The line $y = a_0 + a_1x$ passes through the points if there is a solution to $A\vec{x} = \vec{b}$ (i.e., the system of equations is consistent). This will happen if $(1,4,5)$ is in the column space of A. The question "Is there a line that passes through all of the points?" is equivalent to asking "Is $\vec{b}$ in the column space of A?"

Recall that the system $A\vec{x} = b$ is consistent if b is a linear combination of columns of A.

However, in this example, $A\vec{x} = \vec{b}$ is not consistent since the rref is

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 5 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

So the answer is "No, $\vec{b}$ is not in the column space of A." Since $\vec{b}$ is not in the column space of A, we ask for the next best thing: "What vector in the column space of A is closest to $\vec{b}$?" In fact, if $\vec{b}$ is in the column space of A, then the answer to the last question is $\vec{b}$ itself. So really, the second question is the question we want to answer.

The key to linear regression is answering the question

> What vector $\vec{w} = A\vec{x}$ in the column space of A is the closest vector to $\vec{b}$?

If $\vec{b}$ is in the column space of A, then the answer is $\vec{b}$, the system of equation is consistent, we can find a_0 and a_1, and the line passes through all of the points. If $\vec{b}$ is not in the column space of A, then we have to do more work.

To be closest to $\vec{b}$, a vector $\vec{r}$ (the "residual" vector) with base at $\vec{w}$ and head at $\vec{b}$ must meet the column space of A (a plane) at a 90 degree angle. So the vector $\vec{r} = \vec{b} - \vec{w}$ is orthogonal to every vector in the column space, or every column of A, hence is in the null space of A^T (so $A^T\vec{r} = \vec{0}$). Solving for $\vec{w}$ gives us $\vec{w} = \vec{b} - \vec{r}$. Multiplying both sides on the left by A^T and remembering $A\vec{x} = \vec{w}$ gives us the regression equation

$$A^TA\vec{x} = A^T\vec{w} = A^T(\vec{b} - \vec{r}) = A^T\vec{b} - A^T\vec{r} = A^T\vec{b} - \vec{0} = A^T\vec{b}.$$

Because $\vec{b}$ is not in the column space of A, we find the vector $\vec{w}$ closest to $\vec{b}$ in the columns space of A. The difference $\vec{b} - \vec{w}$ is a vector $\vec{r}$ that is orthogonal to the column space.

Solving for $\vec{x}$, as before, gives us $\vec{x} = (a_0, a_1) = (4/3, 1)$. The line closest to the points was $y = \frac{4}{3} + x$.

The vector $\vec{w} = A\vec{x}$ is in the plane spanned by the columns of A. Because we already know the solution, in this example we can compute the vector $\vec{w}$ as

$$\vec{w} = A\vec{x} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 4/3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 10/3 \\ 16/3 \end{bmatrix}.$$

We can also compute the residual vector $\vec{r} = \vec{b} - \vec{w} = (-1/3, 2/3, -1/3)$, and check that $\vec{w} \cdot \vec{r} = (1/9)(-4 + 20 - 16) = 0$. The magnitude of the residual vector tells us how close $\vec{b}$ is to $\vec{w}$, which gives us an idea of how closely the line fits the points. In this case, $|\vec{r}| = \sqrt{2/3} \approx 0.864$.

Note that the vector $\vec{x}$, $(4/3, 1)$, is the vector of coordinates of $\vec{w}$ relative to the pivot columns of A (here we used the pivot columns of A for our basis). In the picture in the margin above, the gray parallelogram (part of the plane that is the column space of A) was formed by stretching $(1,1,1)$ by $4/3$ and $(0,2,4)$ by 1, illustrating the linear combination that gives $\vec{w}$.

You can think of $\vec{w}$ as the projection of $\vec{b}$ onto the column space of A. The solution to the least squares problem, $\vec{x}$, is the just the vector of coordinates of the projection of $\vec{b}$ onto the column space of A. Regression is really a problem about projecting a vector onto a vector space, then writing a vector in coordinates relative to a specific basis.

2.7 Summary

We have used matrix multiplication in all 3 ways:

1. linear combinations of columns (transition matrices and regression)

2. row dot column (projections in regression)

3. linear combination of rows (electrical networks and curve fitting)

We have seen how Gaussian elimination helps solve problems. We have seen how the inverse and determinants (Cramer's rule) can rapidly help us find solutions. We have used eigenvalues and eigenvectors to understand geometric properties like flow in vector fields and maxima and minima of surfaces, as well as find steady state solutions in stochastic processes. This is a small sample of the beginning of the applications of linear algebra. We also started learning the vocabulary of vector spaces, finding bases and dimensions, which will be the key to extending what we have learned to many more places. The next unit will help us learn to discover patterns and show us how to use definitions and theorems to explicitly state the patterns we find.

2.8   Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Construct visualizations of matrices related to vector fields. Explain how eigenvalues and eigenvectors relate to vector fields.

2. Explain how to generalize the derivative to a matrix. Use this generalization to locate optimal values of a function using the second derivative test. Explain the role of eigenvalues and eigenvectors in the second derivative test.

3. Describe a Markov process. Explain what a steady-state solution is and how it relates to eigenvectors of the transition matrix.

4. Find the currents in an electrical system involving batteries and resistors, using both Gaussian elimination and Cramer's rule.

5. Find interpolating polynomials. Generalize this curve-fitting process to solve the least squares regression problem of fitting a line to a set of data.

6. Find bases for the column space, row space, and null space of a matrix and its transpose by row reducing A and A^T. Use these bases to find the coordinates of vectors relative to the bases.

Here are the preparation problems for this unit. All of these problems come from this book (not Schaum's Outlines). Remember that solutions appear at the end of each chapter.

Preparation Problems (click for solutions)		Webcasts (pdf copy)
Day 1	1b, 1g, 2c, 3b	1, 2, 3
Day 2	4b, 5b, 5d, 6b	4, 5, 6
Day 3	6e, 7a, 7b, 8b	7, 8, 9
Day 4	8c, 9b, 10b, 11b	10, 11
Day 6	Lesson Plan, Quiz, Start Project	

The problems listed below are found in this book.

Concept	Suggestions	Relevant Problems
Vector Fields		1
Kirchoff's Laws		2
Cramer's Rule		3
Interpolating Polynomials		4
Least Square Regression		5
2nd Derivative Test		6
Stochastic Processes		7
Column and Row Spaces		8,9
Null Space and Projections		10,11

2.9 Problems

1. **Vector Fields:** For each of the following vector fields, (1) construct a rough plot of the field by graphing the vectors at the 8 points surrounding the origin, (2) find the eigenvalues and eigenvectors, and if real (3) add to your plot the eigenvectors and some flow lines (curves representing the path of motion). Check your work using Sage.

   (a) $F(x, y) = (2x + y, x + 2y)$

   (b) $F(x, y) = (2x + 4y, 4x + 2y)$

   (c) $F(x, y) = (-y, x)$

   (d) $F(x, y) = (2x + 3y, x + 4y)$

   (e) $F(x, y) = (2x + 3y, 4x + 1y)$

   (f) $F(x, y) = (y, -x)$

   (g) $F(x, y) = (x + y, -x + y)$

   (h) $F(x, y) = (x - y, x - y)$

   (i) $F(x, y) = (x - y, 2x - 2y)$

   (j) $F(x, y) = (-x - y, x - y)$

2. **Kirchoff's Laws:** Consider the following electrical systems. Use the given values to find the current in each wire.
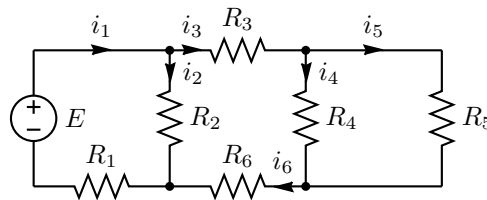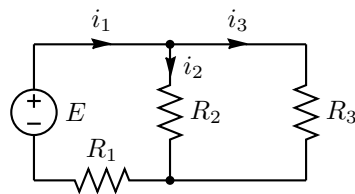
   (a) $E = 12, R_1 = 2, R_2 = 2, R_3 = 2$.

   (b) $E = 12, R_1 = 2, R_2 = 3, R_3 = 3$.

   (c) $E = 12, R_1 = 2, R_2 = 3, R_3 = 6$.

   (d) $E = 12, R_1 = 1, R_2 = 2, R_3 = 2$.

   (e) $E = 9, R_1 = 3, R_2 = 1, R_3 = 2$.

   (f) $E = 6, R_1 = 1, R_2 = 1, R_3 = 2$.

   (g) $E = 12, R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1, R_5 = 1, R_6 = 1$

   (h) $E = 12, R_1 = 2, R_2 = 1, R_3 = 3, R_4 = 4, R_5 = 2, R_6 = 3$

3. **Cramer's Rule:** Use Cramer's rule to find solutions to the following linear systems. If Cramer's rule fails, explain why.

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

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

   (c) $3x + y = 6, x + 3y = 2$

   (d) $2x + y = 1, 4x + 2y = 2$

   (e) $x + y = 0, 2x - y + z = 1, x + z = 0$

   (f) $x + 2y = 3, x - y + z = 0, x + 3y + z = 1$

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

   (h) Re solve any of the problems from problems 2 or 4 using Cramer's rule.

4. **Interpolating Polynomials:** In each of the following scenarios, find a polynomial of least degree which passes through the given points. Verify your result graphically by plotting the points and the polynomial.

   (a) $(1, 2), (3, 3)$

   (b) $(0, 1), (2, 3), (-1, 4)$

   (c) $(1, 1), (2, 2), (-1, 5)$

   (d) $(1, 2), (3, 3), (5, 6)$

   (e) $(1, 2), (3, 3), (5, 5)$

   (f) $(0, 1), (1, 3), (-1, 4), (2, 4)$

   (g) Select any number of points (with varying x values), and find the interpolating polynomial which goes through them. Check your answer by plugging the points into your polynomial.

   (h) Open up a spread sheet program (like Microsoft Excel). Make a column of x values and a column of y values. Create a scatter plot of the data and then on your scatter plot add the interpolating polynomial as well as its equation. You can use this to check your answer on any problem you create yourself.

5. **Regression:** For each set of data points, find an equation of the least squares regression line. Plot the points and the line on the same axes. Does the line appear to be the best fit?

   (a) $(0,0), (1,3), (1,2)$

   (b) $(1,1), (2,1), (3,2)$

   (c) $(1,2), (3,0), (5,1)$

   (d) $(0,0), (1,3), (1,2), (4,5)$

   (e) $(0,0), (1,3), (1,2), (4,-5)$

   (f) $(-1,2), (1,3), (2,5), (3,3)$

   (g) Challenge: Find an equation of the least squares regression parabola $y = a_0 + a_1 x + a_2 x^2$ which passes through the points $(0,0), (1,3), (1,2), (4,5)$. [Hint, you will need a 4 by 3 matrix for A instead of an n by 2. Use the transpose.]

6. **Second Derivative Test:** For each function, find the location of all critical points. Then use the second derivative test to determine if each critical point corresponds to a maximum, minimum, or saddle point. Graph the function in 3D to verify your results, and locate the eigenvectors and eigenvalues in the picture.

   (a) $f(x,y) = x^2 + xy + y^2$

   (b) $f(x,y) = x^2 + 4xy + y^2$

   (c) $f(x,y) = x^2 + 2xy + y^2$

   (d) $f(x,y) = x^2 - 4x + y^2 + 2y + 1$

   (e) $f(x,y) = x^2 - 2x + xy + y^2$

   (f) $f(x,y) = x^2 + xy + 3y^2$

   (g) $f(x,y) = x^3 - 3x + y^2 - 2y$ (2 critical points)

   (h) $f(x,y) = x^3 - 3x + y^3 - 3y^2$ (4 critical points)

   (i) $f(x,y,z) = x^2 + xy + y^2 + z^2 - 4z$ (the Hessian will be a 3 by 3 matrix)

   (j) Create your own function of the form $f(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$. Find the critical points and determine if they correspond to a maximum, minimum, or saddle point.

   (k) For what values of A, B, C, D, E, F will $f(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$ have a critical point that corresponds to a minimum? a maximum? a saddle point? When will the second derivative test fail? How does $B^2 - 4AC$ relate to this problem? [Hint: the quadratic formula will prove useful in trying to find the eigenvalues symbolically. How does the product of the eigenvalues relate to the determinant of the Hessian?]

7. **Markov Processes:** In each scenario, write the transition matrix. If an initial state is given, then find the next two states. Finish by finding a steady state solution and use it to answer the question at the end.

   (a) In a certain town, there are 3 types of land zones: residential, commercial, and industrial. The city has been undergoing growth recently, and the city has noticed the following trends. Every 5 years, 10% of the older residential land gets rezoned as commercial land, while 5% gets rezoned as industrial. The other 85% remains residential. For commercial land, 70% remains commercial, while 10% becomes residential and 20% becomes industrial. For industrial land, 60% remains industrial, while 25% becomes commercial and 15% becomes residential. Currently the percent of land in each zone is 40% residential, 30% commercial, and 30% industrial. What will the percent distribution be in 5 years? In 10 years? If this trend continues indefinitely, what percent of the land will eventually be residential?.

   (b) Suppose we own a car rental company which rents cars in Idaho Falls and Rexburg. The last few weeks have shown a weekly trend that 60% of the cars rented in Rexburg will remain in Rexburg (the other 40% end up in IF), whereas 80% of the cars rented in Idaho Falls will remain in Idaho Falls. If there are currently 60 cars in Rexburg and 140 cars in IF, how many will be in each city next week? In two weeks? In three weeks? If this trend continues indefinitely, about how many cars should you expect to find in Rexburg each week?

   (c) Repeat the previous problem if 40% of the cars rented in Rexburg will remain in Rexburg (the other 60% end up in IF), whereas 80% of the cars rented in Idaho Falls will remain in Idaho Falls.

   (d) Repeat the previous problem if 70% of the cars rented in Rexburg will remain in Rexburg (the other 30% end up in IF), whereas 80% of the cars rented in Idaho Falls will remain in Idaho Falls.

(e) A school cafeteria regularly offers 3 types of meals to its students. One of the meals is always a pizza/salad combo, One is always hamburgers and fries, and one is a daily special which changes daily. In an attempt to understand student preferences, the school discovered the following information. If a student has a hamburger one day, then there is a 30% chance they will try the daily special the next day, and a 40% percent chance they will have the salad bar. If they have the salad bar, then there is a 30% chance they'll switch to the daily special, and a 40% chance they'll switch to the hamburger. If the have the daily special, then there is a 50% chance they'll get the daily special the next day, a 20% chance they'll switch to pizza, and a 30% chance they'll switch to hamburger. If this trend continues, what percent of the students will eat each type of meal?

[While this problem is highly rigged, there is a branch of applied mathematics which is studied by financial analysts call stochastic processes which models such a scenario. This modeling process can help predict how a new restaurant will perform in a city, sometimes predict stock market fluctuations, and more. The study of stochastic processes begins with a Markov chain and then introduces statistics and probability to help predict what happens when trends change.]

8. **Column and Row Spaces:** For each matrix, find the reduced row echelon form of both A and A^T. Use your results to give two bases for the column space. Then give two bases for the row space. If you feel like you have mastered row reduction, then please use a computer (practice using Sage) to do the row reduction for you so that you can focus on just being able to pick out the two different bases.

(a) $\begin{bmatrix} 2 & 3 & 1 \\ -3 & -5 & 3 \end{bmatrix}$

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

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

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

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

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

9. **Coordinates of a vector relative to a basis:** For this problem, use the same matrices as problem 8. From the columns of A, select a basis for the column space of A. Then give the coordinates of $\vec{v}$ and $\vec{w}$ relative to this basis, or explain why the vector is not in the column space of A.

(a) $\vec{v} = (2, 1)$,
    $\vec{w} = (-1, 3)$

(b) $\vec{v} = (6, 3)$,
    $\vec{w} = (-1, 3)$

(c) $\vec{v} = (3, -4, 1)$,
    $\vec{w} = (-1, 3, 2)$

(d) $\vec{v} = (3, -4, 1)$,
    $\vec{w} = (3, -3, 0)$

(e) $\vec{v} = (0, 8, 10)$,
    $\vec{w} = (-13, -9, 2)$

(f) $\vec{v} = (-7, 0, -12)$,
    $\vec{w} = (-8, 11, -18)$

10. **Null Space, Eigenspace, and Orthogonal Complement:** By finding the solutions to either $A\vec{x} = \vec{0}$ (the null space), $(A - \lambda I)\vec{x} = \vec{0}$ (an eigenspace), or $A^T\vec{x} = \vec{0}$ (an orthogonal complement), find a basis and the dimension of the vector subspace requested.

$$A = \begin{bmatrix} 0 & -4 & -4 \\ 9 & 2 & -1 \\ -3 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 4 & -4 & -7 \\ 3 & 6 & 1 & 0 \\ -1 & -2 & 1 & 2 \\ 0 & 0 & 8 & 12 \end{bmatrix}, C = \begin{bmatrix} 2 & 0 & 4 & 2 \\ -1 & -6 & 1 & -4 \\ 2 & -2 & 5 & 1 \end{bmatrix}, D = \begin{bmatrix} 0 & 4 & 4 \\ 3 & -2 & -5 \\ 1 & 4 & 3 \\ -2 & 0 & 2 \end{bmatrix}$$

(a) Find a basis for the null space of A.

(b) Find a basis for the null space of B.

(c) Find a basis for the null space of C.

(d) Find a basis for the null space of D.

(e) Find a basis for the set of vectors that are orthogonal to the columns of A.

(f) Find a basis for the set of vectors that are orthogonal to the columns of B.

(g) Find a basis for the set of vectors that are orthogonal to the columns of C.

(h) Find a basis for the set of vectors that are orthogonal to the columns of D.

(i) For $\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ -2 & -2 & 5 \end{bmatrix}$, find a basis for the eigenspace corresponding to $\lambda = 3$.

(j) For $\begin{bmatrix} -4 & 8 & 6 \\ -9 & 14 & 9 \\ 6 & -8 & -4 \end{bmatrix}$, find a basis for the eigenspace corresponding to $\lambda = 2$.

11. **Projections:** Let $\vec{w}$ be the projection of $\vec{b}$ onto the column space of A. Find the coordinates of $\vec{w}$ relative to the columns of A. Then state the vector $\vec{w}$. Check that you are correct by showing that $\vec{r} = \vec{b} - \vec{w}$ is orthogonal to the columns of A (i.e. show $A^T \vec{r} = \vec{0}$). Compare to problem 5

(a) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

(e) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 0 \\ 3 \\ 2 \\ -5 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$

(d) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 0 \\ 3 \\ 2 \\ 5 \end{bmatrix}$

(f) $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 3 \end{bmatrix}$

2.10   Projects

1. Consider the matrices $A = \begin{bmatrix} 2 & 4 & -4 & -7 \\ 3 & 6 & 1 & 0 \\ -1 & -2 & 1 & 2 \\ 0 & 0 & 8 & 12 \end{bmatrix}$ and $B = \begin{bmatrix} -4 & -7 & 2 & 4 \\ 1 & 0 & 3 & 6 \\ 1 & 2 & -1 & -2 \\ 8 & 12 & 0 & 0 \end{bmatrix}$. Because both matrices have the same columns (just in a different order), the column space of A and the column space of B both represent the same vector subspace. There are infinitely many vectors in this subspace, so there are infinitely many ways to find a basis. Is there a standard basis, one that we can compare any other basis to in an easy manner? The problems below will help you discover this basis.

(a) Find a basis for the column space of A by selecting the pivot columns.

(b) Find a basis for the column space of B by selecting the pivot columns.

(c) Find a basis for the column space of A by reducing A^T.

You should now have 3 different bases. The next 5 problems should all result in the same basis.

(d) Find a basis for the column space of B by reducing B^T.

(e) Let C be a matrix whose columns are the basis vectors obtained in part (a). Row reduce the transpose to obtain a different basis.

(f) Let D be a matrix whose columns are the basis vectors obtained in part(b). Row reduce the transpose to obtain a different basis.

(g) Interchange the columns of A in some other way. Row reduce the transpose of your matrix to obtain a basis for the column space of A.

(h) Joe claims that a basis for the column space is $\{(-5, 3, 1, 12), (0, 7, -1, 8)\}$. Is he correct? To answer this, place his vectors in the columns of a matrix and then row reduce the transpose.

(i) This basis is the basis Sage gives you when you ask for the column space of a matrix. Using Sage, compute the column space of A and B using the `column_space()` method (i.e., `A.column_space()`).

The basis obtained above is called the standard basis for the column space of A. Because the rref of a matrix is unique, the standard basis for a vector space will always be the same, no matter how the vector space is described. Any time someone gives you a basis for a vector subspace of $\mathbb{R}^n$, you can obtain the standard basis using the ideas above. Just place the vectors into the columns of matrix, and then reduce the transpose. Alternatively, just place the vectors into the rows of a matrix, and then reduce the matrix.

2.11 Solutions

Here are the solutions

1. Vector Fields:

   (a) $3, (1,1), 5, (-1,1)$
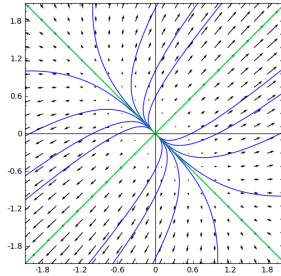
   (b) $6, (1,1), -2, (-1,1)$
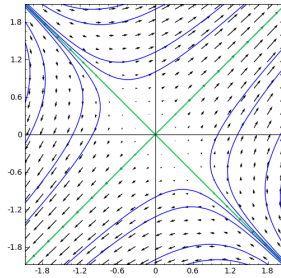
   (c) $i, (i,1), -i, (-i,1)$
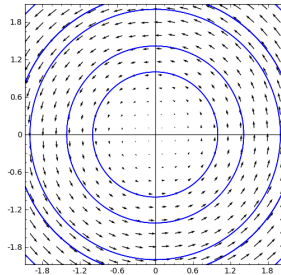
   (d) $5, (1,1), 1, (-3,1)$
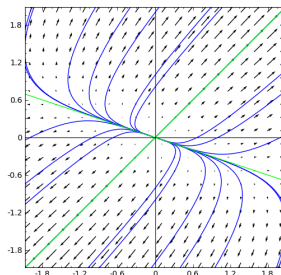
   (e) $5, (1,1), -2, (-3/4,1)$
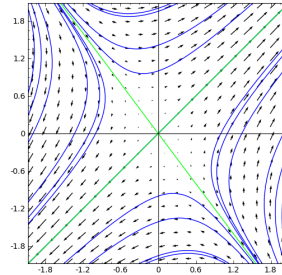
   (f) $i, (-i,1), -i, (i,1)$
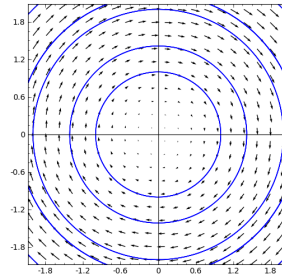
   (g) $1+i, (-i,1), 1-i, (i,1)$
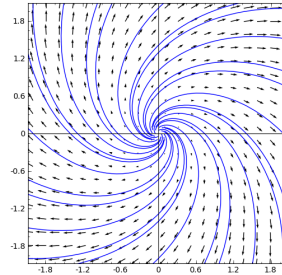
   (h) $0, (1,1)$, there only one independent eigenvectors.
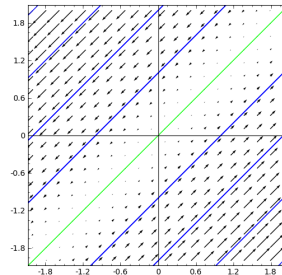
   (i) $0, (1,1), -1, (1,2)$
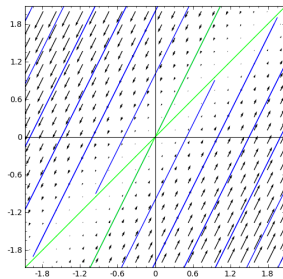
(j) $-1+i, (1,-i), -1-i, (1,i)$

2. Kirchoff's Laws

$$\left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ R_1 & R_2 & 0 & E \\ 0 & -R_2 & R_3 & 0 \end{array}\right]$$
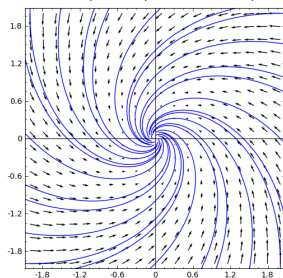
(a) $(4,2,2)$

(b) $(24/7, 12/7, 12/7)$

(c) $(3,2,1)$

(d) $(6,3,3)$

(e) $(27/11, 18/11, 9/11)$

(f) $(18/5, 12/5, 6/5)$

$$\left[\begin{array}{cccccc|c} 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 \\ R_1 & R_2 & 0 & 0 & 0 & 0 & E \\ 0 & -R_2 & R_3 & R_4 & 0 & R_6 & 0 \\ 0 & 0 & 0 & -R_4 & R_5 & 0 & 0 \end{array}\right]$$

(g) $\{7,5,2,1,1,2\}$

(h) $\left\{\frac{25}{6}, \frac{11}{3}, \frac{1}{2}, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right\}$

3. Cramer's rule

(a) $\left\{\frac{3}{7}, -\frac{2}{7}\right\}$

(b) $\left\{\frac{5}{2}, -\frac{1}{2}\right\}$

(c) $\{2,0\}$

(d) Fails. The determinant of the coefficient matrix is zero.

(e) $\left\{\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right\}$

(f) $\left\{\frac{5}{2}, \frac{1}{4}, -\frac{9}{4}\right\}$

(g) Fails. The determinant of the coefficient matrix is zero.

4. Interpolating Polynomials

(a) $1/2x + 3/2$

(b) $(4/3)x^2 - (5/3)x + 1$

(c) $x^2 - 2x + 2$

(d) $(1/4)x^2 - (1/2)x + 9/4$

(e) $(1/8)x^2 + 15/8$

(f) $-x^3 + (5/2)x^2 + (1/2)x + 1$

5. Regression (Make the plots using technology)

(a) $y = \frac{5x}{2}$

(b) $y = \frac{x}{2} + \frac{1}{3}$

(c) $y = \frac{7}{4} - \frac{x}{4}$

(d) $y = \frac{10x}{9} + \frac{5}{6}$

(e) $y = \frac{5}{2} - \frac{5x}{3}$

(f) $y = \frac{3x}{7} + \frac{19}{7}$

(g) $A = \left[\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 16 & 4 & 1 \end{array}\right], B = \left[\begin{array}{c} 0 \\ 3 \\ 2 \\ 5 \end{array}\right], A^T A = $

$\left[\begin{array}{ccc} 258 & 66 & 18 \\ 66 & 18 & 6 \\ 18 & 6 & 4 \end{array}\right], A^T B = \left[\begin{array}{c} 85 \\ 25 \\ 10 \end{array}\right],$

$y = -\frac{5}{12}x^2 + \frac{35}{12}x + 0$

6. Second Derivative Test

(a) At $(0,0)$ eigenvalues are $3,1$ (both positive so min) with eigenvectors $[1,1]^T, [-1,1]^T$.

(b) At $(0,0)$ eigenvalues are $6,-2$ (saddle point) with eigenvectors $[1,1]^T, [-1,1]^T$.

(c) At $(0,0)$ eigenvalues are $4,0$ (test fails) with eigenvectors $[1,1]^T, [-1,1]^T$.

(d) At $(2,-1)$ eigenvalues are $2,2$ (both positive so min) with eigenvectors $[1,0]^T, [0,1]^T$.

(e) At $(4/3, -2/3)$ eigenvalues are $3,1$ (both positive so min) with eigenvectors $[1,1]^T, [-1,1]^T$.

(f) At $(0,0)$ eigenvalues are $6.23607, 1.76393$ (both positive so min) with eigenvectors $[.236, 1]^T, [-4.236, 1]^T$.

(g) At $(-1,1)$ eigenvalues are $-6,2$ (saddle) with eigenvectors $[1,0]^T, [0,1]^T$. At $(1,1)$ eigenvalues are $6,2$ (both positive so min) with eigenvectors $[1,0]^T, [0,1]^T$.

(h) At $(-1,0)$ eigenvalues are $-6,-6$ (both negative so max) with eigenvectors $[1,0]^T, [0,1]^T$. At $(-1,2)$ eigenvalues are $-6,6$ (saddle) with eigenvectors $[1,0]^T, [0,1]^T$. At $(1,0)$ eigenvalues are $6,-6$ (saddle) with eigenvectors $[1,0]^T, [0,1]^T$. At $(1,2)$ eigenvalues are $6,6$ (both positive so min) with eigenvectors $[1,0]^T, [0,1]^T$.

7. Markov Processes

(a) Transition matrix $\left[\begin{array}{ccc} 0.85 & 0.1 & 0.15 \\ 0.1 & 0.7 & 0.25 \\ 0.05 & 0.2 & 0.6 \end{array}\right]$, 5 years

(41.5,32.5,26), 10 years: (42.425,33.4,24.175), Steady state: $[4,3,2]^T$ so 4/9 or 44.4% will be residential.

(b) Transition matrix $\begin{bmatrix} 3/5 & 1/5 \\ 2/5 & 4/5 \end{bmatrix}$, 1 week (64,136), 2 week: (65.6,134.4), 3 week: (66.24,133.76), Steady state: $[1,2]^T$ so 1/3 (33.3%) will be in Rexburg. This means 66 or 67 cars will be in Rexburg.

(c) Transition matrix $\begin{bmatrix} 2/5 & 1/5 \\ 3/5 & 4/5 \end{bmatrix}$, 1 week (52,148), 2 week: (50.4,149.6), 3 week: (50.08,149.92), Steady state: $[1,3]^T$ so 1/4 or 25% will be in Rexburg. This means 50 cars will be in Rexburg.

(d) Transition matrix $\begin{bmatrix} 7/10 & 1/5 \\ 3/10 & 4/5 \end{bmatrix}$, 1 week (70,130), 2 week: (75,125), 3 week: (77.5,122.5), Steady state: $[2,3]^T$ so 2/5 or 40% will be in Rexburg. This means 80 cars will be in Rexburg.

(e) My order is hamburger, pizza/salad, special (your order may vary which means your matrix will be a little different, but the eigenvector will still have the same ratios). Transition matrix $\begin{bmatrix} 3/10 & 2/5 & 3/10 \\ 2/5 & 3/10 & 1/5 \\ 3/10 & 3/10 & 1/2 \end{bmatrix}$, Steady state: $[29, 26, 33]^T$ or $[29/88, 26/88, 33/88]^T$ so Hamburger - 32.9545%, Pizza/Salad - 29.5455%, Special - 37.5%.

8. **Column and Row Spaces:**

(a) Col: $\{(2, -3), (3, -5)\}$, $\{(1,0), (0,1), \}$.
Row: $\{(1,0,14), (0,1,-9)\}$, $\{(2,3,1), (-3,-5,3)\}$.

(b) Col: $\{(2,1), (1,3)\}$, $\{(1,0), (0,1)\}$.
Row: $\{(1,2,0), (0,0,1)\}$, $\{(2,4,1), (1,2,3)\}$.

(c) Col: $\{(1,2,-3), (2,0,-2)\}$, $\{(1,0,-1), (0,1,-1)\}$.
Row: $\{(1,0,1/2), (0,1,5/4)\}$, $\{(1,2,3), (2,0,1)\}$.

(d) Col: $\{(1,2,-3), (3,1,-4)\}$, $\{(1,0,-1), (0,1,-1)\}$.
Row: $\{(1,0,-(2/5)), (0,1,4/5)\}$, $\{(1,3,2), (2,1,0)\}$.

(e) Col: $\{(2,1,-2), (1,3,1), (-2,0,1)\}$, $\{(1,0,0), (0,1,0), (0,0,1)\}$.
Row: $\{(1,0,0,-1), (0,1,0,2), (0,0,1,-2)\}$, $\{(2,1,-2,4), (1,3,0,5), (-2,1,1,2)\}$.

(f) Col: $\{(0,2,-1), (1,-1,2), (3,0,5)\}$, $\{(1,0,0), (0,1,0), (0,0,1)\}$.
Row: $\{(1,0,-1,0), (0,1,2,0), (0,0,0,1)\}$, $\{(0,1,2,3), (2,-1,-4,0)\}$.

9. **Coordinates of a vector relative to a basis:** The basis vectors are the pivot columns of rref. See 8 for the basis vectors. To get the coordinates, augment A by the two vectors, and row reduce.

(a) $\vec{v} : (13, -8)$, $\vec{w} : (4, -3)$

(b) $\vec{v} : (3, 0)$, $\vec{w} : (-(6/5), 7/5)$

(c) $\vec{v} : (-2, 5/2)$, $\vec{w} :$ not in space.

(d) $\vec{v} : (-3, 2)$, $\vec{w} : (-(12/5), 9/5)$

(e) $\vec{v} : (-4, 4, -2)$, $\vec{w} : (0, -3, 5)$

(f) $\vec{v} : (1, 2, -3)$, $\vec{w} : (3, -5, -1)$

10. **Null Space, Eigenspace, and Orthogonal Complement:** There is more than one way to pick a basis. Here possible answers. I obtained these basis by writing solutions in terms of the free variable (rrefing A or A^T) and then multiplying by the least common denominator to eliminate fractions.

(a) $\{(1, -3, 3)\}$

(b) $\{(1, 0, -3, 2), (-2, 1, 0, 0)\}$

(c) $\{(-2, -1, 0, 2), (-4, 1, 2, 0)\}$

(d) $\{(1, -1, 1)\}$

(e) $\{(1, 2, 6)\}$

(f) $\{(12, -8, 0, 7), (2, 1, 7, 0)\}$

(g) $\{(-7, -2, 6)\}$

(h) $\{(1, 2, 0, 3), (-7, -2, 6, 0)\}$

(i) $\{(1, 0, 1), (-1, 1, 0)\}$

(j) $\{(1, 0, 1), (4, 3, 0)\}$

11. **Projections:** The work is identical to problem 5. The only difference is the vocabulary. I'll give the coordinates of $\vec{w}$ as well as $\vec{w}$

(a) Coor: $(0, 5/2)$,
$\vec{w} = 5/2(0, 1, 1)$

(b) Coor: $(1/3, 1/2)$,
$\vec{w} = 1/3(1, 1, 1) + 1/2(1, 2, 3)$

(c) Coor: $(7/4, -1/4)$,
$\vec{w} = 7/4(1, 1, 1) - 1/4(1, 3, 5)$

(d) Coor: $(5/6, 10/9)$,
$\vec{w} = 5/6(1, 1, 1, 1) + 10/9(0, 1, 1, 4)$

(e) Coor: $(5/2, -5/3)$,
$\vec{w} = 5/2(1, 1, 1, 1) - 5/3(0, 1, 1, 4)$

(f) Coor: $(19/7, 3/7)$,
$\vec{w} = 19/7(1, 1, 1, 1) + 3/7(-1, 1, 2, 3)$

Chapter 3

Patterns

The purpose of this chapter is to help you learn to create examples of patterns that others before you have noticed. Vector spaces are the most important pattern we will explore in this chapter.

1. Create examples to illustrate mathematical definitions. Know the definitions in this chapter.

2. Create examples to illustrate theorems. You do not have to memorize all the theorems in this chapter, but when one is given, make sure you can construct examples of it.

3. When a matrix has an inverse, this is equivalent to many other facts about the matrix and corresponding systems of equations. Illustrate this equivalence with examples, and be able to explain why the the ideas are equivalent.

4. Be able to explain why something is or is not a vector space. Give examples of vector spaces and subspaces that you have encountered previously in this class and other classes.

5. Explain how to find a basis for a vector space and the dimension of a vector space. Give the coordinates of vectors relative to a chosen basis.

3.1   Recognizing and Naming Patterns

Have you ever noticed while reading a mathematics or other science textbook that the book is filled with the following four things: (1) examples, (2) definitions, (3) theorems, and (4) proofs? One of the main goals of mathematics is to describe the patterns in the world around us, and these four things are often the ways a pattern is found and communicated.

Most of the time, recognizing a pattern starts with *examples*. After working with lots of examples, we start to notice recurring patterns in the objects we deal with and the relationships between those objects. We create *definitions* to explicitly describe the structures and objects that keep recurring. When the patterns help us understand when and how these objects occur, or describe relationships between these objects, we create *theorems*. *Proofs* provide the logical backbone to the theorems and show why the theorems are true in every case. Everything in this process starts with examples.

Learning mathematics requires learning to work with examples. The main goal of this chapter is to learn how to create examples to understand the

definitions and theorems that others have discovered. If you encounter a new idea in any scientific area, start by looking at examples. After you have studied several simple examples, you will discover patterns that help you understand the new idea. Sometimes, your exploration of simple examples will result in new discoveries (theorems). Examining, understanding, and generalizing from examples is one of the main tools of mathematical and scientific research. Learning to create and look at examples is crucial.

The most important structure in linear algebra is the notion of a vector space. This is an abstract space that came from studying the properties of vectors in $\mathbb{R}^n$, and then realizing that these same properties hold in other places (like polynomials, functions, waves, and more). Before we formally define a vector space in general, we will practice creating examples to understand definition and theorems. Most of what follows is a summary of the ideas introduced in the previous chapters. Once we have done this, we'll be ready to formally define general vector spaces. Then the real power of linear algebra begins.

As you encounter each new definition or theorem, your homework will be to create an example which illustrates the new idea. To illustrate a definition, create several examples which satisfy the definition. Make some of the examples simple examples, and make some examples wild and crazy examples of the definition. Then make several examples that do not quite satisfy the definition (but are close). Making examples of what a term means and what it does not mean will solidify the definition in your mind.

To illustrate a theorem, make a few examples that satisfy the conditions in the theorem and show that the result of the theorem is satisfied. Again, make some examples that are simple and some that are wild and crazy. Then make some examples that do not quite satisfy the conditions in the theorem and do not have the result hold (for example, you might make examples that satisfy all but one of the conditions in the theorem).

3.1.1 Definitions

Whenever you see a new definition, one of the most important things you can do is create a few examples which illustrate that definition. Most of these words are review, but a few are new. Make sure you know these definitions.

This first definition is a formal statement of what we mean when we write $\mathbb{R}^n$. It is the foundation for the formal vector space definition we'll see later.

Definition 3.1. The symbol $\mathbb{R}$ is used to represent the real numbers. The real vector space $\mathbb{R}^n$ is a set together with a way of adding two elements (called *vector addition*) and a way of multiplying an element by a real number (called *scalar multiplication*). The set we use is the set of n-tuples $(v_1, v_2, \ldots, v_n)$ where each v_i is an element of $\mathbb{R}$. We write this symbolically as

$$\{(v_1, v_2, \ldots, v_n) \mid v_i \in \mathbb{R} \text{ for } 1 \leq i \leq n\}.$$

Each n-tuple $\vec{v} = (v_1, v_2, \ldots, v_n)$ is called a vector. The entries of the vector are called the *components* of the vector. Vector addition is performed by adding the components of the vectors. Scalar multiplication is done by multiplying each component by the scalar.

Definition 3.2. The *dot product* of two vectors $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is the scalar $\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum u_i v_i$. We say that two vectors are *orthogonal* if their dot product is zero. See example 1.4.

Definition 3.3. A (real) *matrix* is a rectangular grid of real numbers. We use the notation a_{ij} to represent the entry in the ith row and jth column of

the matrix A. If a matrix A has m rows and n columns, we say the matrix has *size m by n* and write A_{mn}. The *sum* of two matrices of the same size is performed by adding corresponding entries. The *product* of two matrices $A_{mn} = \{a_{ij}\}$ and $B_{np} = \{b_{jk}\}$ is a new matrix $C = AB$ where the ikth entry of C is $c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$ (the dot product of the ith row of A and the kth column of B).

Definition 3.4. We say that two matrices are *row equivalent* if one can be obtained from the other by a finite sequence of these three row operations: (1) interchange two rows, (2) multiply one row by a nonzero constant, and (3) add to a row a multiple of a different row.

    A matrix is said to be in *row echelon form* (ref) if (1) each nonzero row begins with a 1 (called a *leading 1*), (2) the leading 1 in a row occurs further right than a leading 1 in the row above, and (3) any rows of all zeros appear at the bottom. The location in the matrix of each leading 1 is called a *pivot*, and the column containing the pivot is called a *pivot column*. A matrix is said to be in *reduced row echelon form* (rref) if the matrix is in row echelon form, and each pivot column contains all zeros except for the pivot (the leading one).

> See example 1.12. Every matrix we obtain in the row reduction process of A is row equivalent to all the other matrices obtained from A by doing row operations, and to the reduced row echelon form of A. Each of the matrices in example 1.12 is row equivalent to all the other matrices in the example.

Definition 3.5: Homogeneous. A linear system $A\vec{x} = \vec{b}$ is said to be either *homogeneous* if $\vec{b} = \vec{0}$ or *nonhomogeneous* if $\vec{b} \neq \vec{0}$.

Definition 3.6: Consistent. A linear system is said to be *consistent* if it has at least one solution and *inconsistent* if it has no solutions.

Definition 3.7: Linear combination, Span. A *linear combination* of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a sum

> See example 1.17.

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \sum_{i=1}^{n} c_i\vec{v}_i$$

where each scalar $c_1, c_2, \ldots, c_n$ is a real number. The *span* of a set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is the set of all possible linear combinations of those vectors.

Definition 3.8: Linear independence. We say a set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of vectors is *linearly independent* if the only solution to the homogeneous equation $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}$ is the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. If there is a nonzero solution to this homogeneous equation, then we say the set of vectors is *linearly dependent*, or we just say the vectors are linearly dependent (leave off "set of").

> See example 1.18.

Definition 3.9: Basis. A *basis* for $\mathbb{R}^n$ is a set of vectors that are (1) linearly independent and (2) span all of $\mathbb{R}^n$. The *standard basis* for $\mathbb{R}^n$ is the set

$$\{\vec{e}_1 = (1, 0, 0, \ldots, 0, 0),\ \vec{e}_2 = (0, 1, 0, \ldots, 0, 0),\ \ldots,\ \vec{e}_n = (0, 0, 0, \ldots, 0, 1)\}.$$

Definition 3.10: Column space, Rank. The *column space* of a matrix is the span of the column vectors. A *basis for the column space* of a matrix is a set of linearly independent vectors whose span is the column space. The *dimension* of the column space is the number of vectors in a basis for the column space. The *rank* of a matrix is the dimension of the column space.

> See examples 2.20 and 2.22.

Definition 3.11: Row space. The *row space* of a matrix is the span of the row vectors. A *basis for the row space* is a set of linearly independent vectors whose span is the row space. The *dimension* of the row space is the number of vectors in a basis for the row space.

> See examples 2.20 and 2.22.

Definition 3.12: Coordinates. If $\beta = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ is a basis for a vector space, and a vector $\vec{v}$ equals the linear combination

$$\vec{v} = c_1 \vec{b}_1 + c_2 \vec{v}_2 + \cdots c_n \vec{v}_n,$$

then we call the scalars c_1, c_2, ..., c_n the *coordinates* of $\vec{v}$ relative to the basis β. We may write $\vec{v} = (c_1, c_2, \ldots, c_n)_\beta$ to represent the vector $\vec{v}$, or if the basis β is understood, we may just write $\vec{v} = (c_1, c_2, \ldots, c_n)$.

Definition 3.13: Diagonal, Identity, Triangular matrices. A *diagonal matrix* is a square matrix in which the entries off the main diagonal are zero ($a_{ij} = 0$ if $i \neq j$). The *identity matrix* is a diagonal matrix in which each entry on the main diagonal is 1 ($a_{ii} = 1$). We often write I to represent the identity matrix and I_n to represent the $n \times n$ identity matrix when specifying the size is important. An *upper triangular matrix* is a square matrix in which every entry below the main diagonal is zero. A *lower triangular matrix* is a square matrix in which every entry above the main diagonal is zero.

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 8 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 3 & 2 \\ 0 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 3 & 4 & 0 \\ 2 & 0 & -1 \end{bmatrix}$$
$$\text{diagonal} \qquad \text{identity} \qquad \text{upper triangular} \quad \text{lower triangular}$$

Definition 3.14: Determinant. The *determinant* of a 1×1 matrix is the entry a_{11}. A *minor* M_{ij} of an $n \times n$ matrix is the determinant of the $(n-1) \times (n-1)$ submatrix obtained by removing the ith row and jth column. A *cofactor* of a matrix is the product $C_{ij} = (-1)^{i+j} M_{ij}$. The determinant of an $n \times n$ matrix ($n \geq 2$) is found by multiplying each entry a_{1i} on the first row of the matrix by its corresponding cofactor, and then summing the result. Symbolically, we write

$$\det A = |A| = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

Definition 3.15: Inverse. The *inverse* of a square matrix A is a matrix B such that $AB = I$ and $BA = I$, where I is the identity matrix. If a square matrix has an inverse, we say the matrix is *nonsingular* or *invertible*. If a square matrix does not have an inverse, we say the matrix is *singular* or *noninvertible*.

Definition 3.16: Transpose, Symmetric. The *transpose* of a matrix $A_{m \times n} =$ $\{a_{ij}\}$ is a new matrix $A_{n \times m}^T$ in which the ith column of A is the ith row of A^T. A *symmetric matrix* is a matrix A such that $A = A^T$.

Definition 3.17: Eigenvector, Eigenvalue. An *eigenvector* of a square ma- trix A is a nonzero vector $\vec{x}$ with the property that $A\vec{x} = \lambda\vec{x}$ for some scalar λ. We say that $\vec{x}$ is an eigenvector corresponding to the *eigenvalue* λ.

3.1.2 Theorems

The definitions above summarize most of what we have done in the first two chapters. The list of properties below summarize many of the patterns we have already observed, together with many patterns that we will be exploring as the semester progresses. Each is called a theorem because it is a statement that requires justification (a proof). In the homework, your job is to create examples to illustrate these theorems. If a theorem is hard to remember, then create more examples to illustrate the theorem until you recognize the patterns that emerge. Mathematics is all about finding patterns and then precisely stating the pattern

in words. You do not have to memorize all of these theorems, but you should be able to create examples to illustrate them. In the next section, we will prove some of them.

Before stating the theorems, we first need to discuss the phrase "if and only if." Theorem 3.16 states

> The columns are linearly independent if and only if every column is a pivot column.

We could write this statement symbolically by writing "p if and only if q."   The words "if and only if" create two if-then statements, namely (1) if p then q and (2) if q then p. This gives us the two statments:

> The statement "p if and only if q" implies both
> 1. if p then q, and
> 2. if q then p.

1. *If* the columns are linearly independent *then* every column is a pivot column.

2. *If* every column is a pivot column *then* the columns are linearly independent.

Every time you see the words "if and only if," remember that this means you have two if-then statements. That means that you need at least two examples to illustrate such a theorem (one for each if-then statement).

Vector Properties

Theorem 3.1 (Vector space properties of $\mathbb{R}^n$)**.** Vector addition and scalar multiplication in $\mathbb{R}^n$ satisfy the following properties:

(A_1) Vector addition is associative: $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.

(A_2) Vector addition is commutative: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.

(A_3) The zero vector $\vec{0}$ satisfies $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$.

(A_4) The additive inverse of $\vec{v}$, denoted $-\vec{v}$, is the vector which satisfies $\vec{v} + (-\vec{v}) = 0$.

(M_1) Scalar multiplication distributes across vector addition: $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$.

(M_2) Scalar multiplication distributes across scalar addition: $(c + d)\vec{v} = c\vec{v} + d\vec{v}$.

(M_3) Scalar multiplication is associative: $(cd)\vec{v} = c(d\vec{v})$

(M_4) Scalar multiplication by 1 does nothing: $1\vec{v} = \vec{v}$

System Properties

Theorem 3.2. Every matrix has a unique reduced row echelon form (rref). This implies that for any matrix, the location of the pivot columns is unique (the locations do not depend on how the matrix is reduced to rref).

Theorem 3.3. Two matrices are row equivalent if and only if they have the same reduced row echelon form.

Theorem 3.4. A linear system of equations $A\vec{x} = \vec{b}$ has a solution if and only if the number of pivot columns of A equals the number of pivot columns of the augmented matrix $[A \mid b]$.

Theorem 3.5. If a linear system of equations has more than one solution, it has infinitely many. Solving a linear system has three possible outcomes: (1) no solution, (2) exactly one solution, or (3) infinitely many solutions.

Theorem 3.6 (Removing non-pivot columns does not alter the row operations involved in obtaining rref). Let A be a matrix whose rref is R. Let B be the matrix obtained from A by removing the kth column. If the kth column of A is not a pivot column, then the rref of B is found by removing the kth column from R.

Theorem 3.7. If a consistent linear system of equations $A\vec{x} = \vec{b}$ has n variables and k pivot columns, then it has $n - k$ free variables.

Theorem 3.8 (Superposition Principle). For a homogeneous system $A\vec{x} = \vec{0}$, any linear combination of solutions is also a solution.

Theorem 3.9. For a nonhomogeneous system $A\vec{x} = \vec{b}$, the difference $\vec{x}_1 - \vec{x}_2$ between any two solutions is a solution to the homogeneous system $A\vec{x} = \vec{0}$.

Matrix Properties

Theorem 3.10 (Vector space properties of M_{mn}). Let A, B and C be m by n matrices, and c, d be real numbers. Matrix addition and scalar multiplication satisfy the following properties:

(A_1) Matrix addition is associative: $(A + B) + C = A + (B + C)$.

(A_2) Matrix addition is commutative: $A + B = B + A$.

(A_3) The zero matrix 0 satisfies $0 + A = A + 0 = A$.

(A_4) The additive inverse of A is $-A$ which satisfies $A + (-A) = 0$.

(M_1) Scalar multiplication distributes across matrix addition: $c(A + B) = cA + cB$.

(M_2) Scalar multiplication distributes across scalar addition: $(c+d)A = cA+dA$.

(M_3) Scalar multiplication is associative: $(cd)A = c(dA)$.

(M_4) Scalar multiplication by 1 does nothing: $1A = A$.

Theorem 3.11 (Transpose of a product). The transpose of a product is found by multiplying the transposes in the reverse order:

$$(AB)^T = B^T A^T.$$

Theorem 3.12. The product $A\vec{x}$ of a matrix A and column vector $\vec{x}$ is a linear combination of the columns of A.

Theorem 3.13. The product $\vec{x}A$ of a row vector $\vec{x}$ and matrix A is a linear combination of the rows of A.

Theorem 3.14. Consider the matrices $A_{m \times n} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$ and $B_{n \times p} = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_p \end{bmatrix}$, where $\vec{a}_i$ and $\vec{b}_i$ are the columns of A and B. The product AB equals $AB = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \end{bmatrix}$. In other words, every column of AB is a linear combination of the columns of A. Similarly, every row of AB is a linear combination of the rows of B.

Linear Independence Properties

Theorem 3.15. The rank of a matrix is the number of pivot columns, which equals the number of leading 1's.

Theorem 3.16. The columns of a matrix are linearly independent if and only if every column is a pivot column.

Theorem 3.17. A set (of two or more vectors) is linearly dependent if and only if at least one of the vectors can be written as a linear combination of the others. In particular, two vectors are linearly dependent if and only if one is a multiple of the other.

Theorem 3.18. The system $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b}$ is in the column space of A.

Theorem 3.19. The dimension of the column space (the rank) of a matrix equals the dimension of the row space of the matrix.

Theorem 3.20. Consider a matrix A whose columns are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. A basis for the column space is the set of pivot columns of A. Let R be the reduced row echelon form of A with any rows of zeros removed. For each column $\vec{v}_k$ of A, the kth column of R gives the coordinates of $\vec{v}_k$ relative to the basis of pivot columns.

In terms of matrix multiplication, if C is the matrix obtained from A by erasing the non-pivot columns, and $\vec{r}_k$ is the kth column of R, then $C\vec{r}_k = v_k$. The columns of R tell us how to linearly combine the pivot columns of A to obtain all columns of A.

Theorem 3.21. Consider a matrix A whose rows are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$. Let R be the reduced row echelon form of A with any rows of zeros removed. A basis for the row space is the set of nonzero rows of R. Let C be the matrix obtained from A by erasing the non-pivot columns of A. For each row $\vec{v}_k$ of A, the kth row of C gives the coordinates of $\vec{v}_k$ relative to the basis of nonzero rows of R.

In terms of matrix multiplication, if $\vec{r}_k$ is the kth row of C, then $\vec{r}_k R = v_k$. The rows of C tell us how to linearly combine the rows of R to obtain all rows of A.

Inverse Properties

The following theorems hold for invertible (square) matrices.

Theorem 3.22. The inverse of a matrix is unique.

Theorem 3.23. The inverse of A^{-1} is A. This implies that $(A^k)^{-1} = (A^{-1})^k$ and $(cA)^{-1} = \frac{1}{c}A^{-1}$.

Theorem 3.24. The inverse of the transpose is the transpose of the inverse: $(A^T)^{-1} = (A^{-1})^T$.

Theorem 3.25 (Socks-Shoes property)**.** The inverse of a product is $(AB)^{-1} = B^{-1}A^{-1}$.

In the previous theorem, notice that the order of multiplication is reversed. You put on your socks and then your shoes. To undo this, you first take off your shoes, then your socks.

Theorem 3.26. If A is invertible, then the solution to $A\vec{x} = \vec{b}$ is $\vec{x} = A^{-1}\vec{b}$.

Determinant Properties

The following theorems hold for square matrices.

Theorem 3.27. The determinant can be computed by using a cofactor expansion along any row or any column. Symbolically, we can write this as $|A| = \sum_{i=1}^{n} a_{ij} C_{ij}$ for any column j, or $|A| = \sum_{j=1}^{n} a_{ij} C_{ij}$ for any row i.

Theorem 3.28. The determinant of a triangular matrix is the product of the entries along its main diagonal.

Theorem 3.29. The determinant of AB is $|AB| = |A||B|$. This implies that $|cA| = c^n |A|$ and $|A^{-1}| = \frac{1}{|A|}$.

Theorem 3.30. The determinant of A and A^T are the same: $|A^T| = |A|$.

Theorem 3.31. A matrix is invertible if and only if its determinant is not zero.

Theorem 3.32. If one column of a matrix is a multiple of another column, then the determinant of the matrix is zero. If one row of a matrix is a multiple of another row, then the determinant of the matrix is zero.

Theorem 3.33. Let A be a square matrix. Define the adjoint of A, written adj(A), to be the matrix whose ijth entry is the jith cofactor of A (the transpose of matrix of cofactors). The product of the adjoint and the original matrix is adj(A)$A = |A|I$, a multiple of the identity matrix. If the determinant of A is nonzero, this implies that the inverse of A is $A^{-1} = \frac{1}{|A|}$adj(A).

Eigenvalues and Eigenvectors

The following theorems hold for square matrices.

Theorem 3.34 (How to compute eigenvalues and eigenvectors)**.** The nonzero vector $\vec{x}$ is an eigenvector of A corresponding to the eigenvalue λ if and only if $(A - \lambda I)\vec{x} = \vec{0}$ and $\det(A - \lambda I) = 0$.

Theorem 3.35. For an n by n real matrix, the characteristic polynomial is an nth degree polynomial. This means there are n eigenvalues (counting multiplicity and complex eigenvalues).

Theorem 3.36. For a triangular matrix (all zeros either above or below the diagonal), the eigenvalues appear on the main diagonal.

Theorem 3.37. The eigenvalues of A^T are the same as the eigenvalues of A.

Theorem 3.38. If the sum of every row of A is the same, then that sum is an eigenvalue corresponding to the eigenvector $(1, 1, \ldots, 1)$. If the sum of every column of A is the same, then that sum is an eigenvalue of A as well (because it is an eigenvalue of A^T). A stochastic matrix (e.g., a transition matrix for a Markov process) always has 1 as an eigenvalue because the columns sum to 1.

Theorem 3.39. If a matrix is symmetric ($A^T = A$), then the eigenvalues of A are real. In addition, eigenvectors corresponding to different eigenvalues of a symmetric matrix are orthogonal (if $\vec{x}_1$ and $\vec{x}_2$ are eigenvectors corresponding to different eigenvalues, then $\vec{x}_1 \cdot \vec{x}_2 = 0$).

Theorem 3.40. Let A be an n by n matrix. Suppose there are n distinct eigenvalues of A, named $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\vec{x}_i$ be an eigenvector corresponding to λ_i for each i. The set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ is linearly independent, and hence is a basis for $\mathbb{R}^n$.

3.1.3 Proving theorems

In this section I'll illustrate various methods of justifying a theorem (giving a proof). You do not have to be able to prove every one of the theorems from the previous section. If you want a challenge, try to prove all of the theorems below.

Prove by using a definition

To prove theorem 3.25 ($(AB)^{-1} = B^{-1}A^{-1}$), we just use the definition of an inverse. It often helps to rewrite a definition, perhaps changing the names of variables if needed. The inverse of a matrix C is a matrix D such that $CD = DC = I$ (multiplying on either side gives the identity). To show that the inverse of AB is $B^{-1}A^{-1}$, we need to show that multiplying AB on both the left and right by $B^{-1}A^{-1}$ results in the identity. We multiply AB on the left by $B^{-1}A^{-1}$ and compute

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

Similarly, we multiply AB on the right by $B^{-1}A^{-1}$ and compute

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

Because multiplying AB on both the left and right by $B^{-1}A^{-1}$ gives us the identity matrix, we know $B^{-1}A^{-1}$ is the inverse of AB. All we did was use the definition of an inverse.

Proving something is unique

To show something is unique, one common approach is to assume there are two things and then show the two things must be the same. We'll do this to show that inverses are unique (Theorem 3.22). Suppose that A is a square matrix with an inverse B and another inverse C. Using the definition, we know that $AB = BA = I$, and $AC = CA = I$. We now need to show that $B = C$ (show the two things are the same). If we multiply both sides of the equation $AB = I$ on the left by C, then we obtain $C(AB) = CI$. Since matrix multiplication is associative, we can rearrange the parentheses on the left to obtain $(CA)B = C$. Since C is an inverse for A, we know $CA = I$, which means $IB = C$, or $B = C$, which shows that the inverse is unique.

Prove by using another theorem

Sometimes using other theorems can quickly prove a new theorem. Theorem 3.13 says that the product $\vec{x}A$ is a linear combination of the rows of A. The two theorems right before this one (Theorems 3.11 and 3.12) state that $(AB)^T = B^T A^T$ and that $A\vec{x}$ is a linear combination of the columns of A. Rather than prove Theorem 3.13 directly, we are going to use these other two theorems. The transpose of $\vec{x}A$ is $A^T \vec{x}^T$, which is a linear combination of the columns of A^T. However, the columns of A^T are the rows of A, so $A^T \vec{x}^T$ is a linear combination of the rows of A. This means that $\vec{x}A$ (undoing the transpose) is a linear combination of the rows of A. Theorems which give information about columns often immediately give information about rows as well.

If and only if

The words "if and only if" require that an "if-then" statement work in both directions. The following key theorem (Theorem 3.17) has an if and only if statement in it.

A set of two or more vectors is linearly dependent if and only if one of the vectors is a linear combination of others.

To prove this, we need to show two things:

1. If the vectors are linearly dependent, then one is a linear combination of the others.

2. If one of the vectors is a linear combination of the others, then the vectors are linearly dependent.

We'll start by proving the first item. If the vectors are dependent, then there is a nonzero solution to the equation

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

Pick one of the nonzero constants, say c_k. Subtract $c_k\vec{v}_k$ from both sides and divide by $-c_k$ (which isn't zero) to obtain

$$\frac{c_1}{-c_1}\vec{v}_1 + \cdots + \frac{c_{k-1}}{-c_k}\vec{v}_{k+1} + \frac{c_{k+1}}{-c_k}\vec{v}_{k+1} + \cdots + \frac{c_n}{-c_k}\vec{v}_n = \vec{v}_k,$$

which means that v_k is a linear combination of the other vectors.

Now let's prove the second item. Suppose one of the vectors is a linear combination of the others (say the kth). Write

$$v_k = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k+1} + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n.$$

Subtracting v_k from both sides gives a nonzero solution ($c_k = -1$) to the equation $\vec{0} = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k+1} - v_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_n\vec{v}_n$, which means the vectors are linearly dependent.

Proof by contradiction

Sometimes in order to prove a theorem, we make a contradictory assumption and then show that this assumption leads to an error (which means the assumption cannot be correct). We will do this to show that if A is invertible, then $|A| \neq 0$. We will also use the fact that the determinant of a product is the product of the determinants, $|AB| = |A||B|$, a fact we will verify in a later chapter. Let's start with a matrix A which has an inverse and then assume that the determinant is $|A| = 0$. The inverse of A is A^{-1} and the product $AA^{-1} = I$ is the identity. Taking determinants of both sides gives $|AA^{-1}| = |I| = 1$, or $|A||A^{-1}| = 1$ (remember the identity matrix has determinant 1). Since we assumed that $|A| = 0$, we have $0|A^{-1}| = 0 = 1$, which is absurd. Our assumption that $|A| = 0$ must be incorrect. We must have $|A| \neq 0$.

Prove $a = b$ by showing $a \leq b$ and $b \leq a$

If $a \leq b$ and $b \leq a$, then $a = b$. Sometimes it is easier to show an inequality between two numbers than it is to show the numbers are equal. We will use this to show that the rank of a matrix (the dimension of the column space, which is the number of pivot columns) equals the dimension of the row space. Let the rank of A be r, and let the dimension of the row space equal s. Write $A = CR$ where the columns of C are the pivot columns of A and the rows of R are the nonzero rows of the rref of A. We know that the rank r equals the number of columns in C. Since every row of A can be written as a linear combination of the rows of R, the rows of R span the rows space. We know that a basis for the

row space cannot contain more than r vectors since there are r rows in R. This means that the dimension s of the row space must satisfy $s \leq r$.

We now repeat the previous argument on the transpose of A. Write $A^T = C'R'$ where the columns of C' are the pivot columns of A^T and the rows of R' are the nonzero rows of the rref of A^T. The columns of C' are linearly independent rows of A, so C' has s columns, and R' has s rows. Since every row of A^T can be written as a linear combination of the rows of R', a basis for the row space of A^T (i.e. the column space of A) cannot have more than s vectors, or $r \leq s$. Combining both inequalities, since $r \leq s$ and $s \leq r$, we have $r = s$.

3.2   A huge equivalence: Connecting all the ideas

Theorem 3.41. For an n by n matrix A, the following are all equivalent (meaning you can put an if and only if between any two of these statements):

1. The system $A\vec{x} = \vec{0}$ has only the trivial solution.

2. The system $A\vec{x} = \vec{b}$ has a unique solution for every $\vec{b}$.

3. The rref of A is the identity matrix.

4. Every column of A is a pivot column.

5. The columns of A are linearly independent. (The dimension of the column space is n and the span of the columns of A is $\mathbb{R}^n$.)

6. The rank of A is n.

7. The rows of A are linearly independent. (The dimension of the row space is n and the span of the rows of A is $\mathbb{R}^n$.)

8. The vectors $\vec{e}_1 = (1, 0, 0, \ldots, 0)$, $\vec{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\vec{e}_n = (0, 0, 0, \ldots, 1)$ are in the column space of A.

9. A is invertible.

10. $|A| \neq 0$.

11. The eigenvalues of A are all nonzero.

We have been using this theorem periodically without knowing it. Because this theorem is true, there is an if and only if between any pair of items. In addition, because there is an if and only if between every item, we can also negate every single statement and obtain an equivalence as well. In this negated equivalence, we use the symbol $\iff$ to represent the words "if and only if."

Theorem 3.42. There is more than one solution to $A\vec{x} = 0$ $\iff$ the system $A\vec{x} = \vec{b}$ does not have a unique solution (it may not have one at all) $\iff$ rref$(A) \neq I$ $\iff$ at least one column is not a pivot column $\iff$ the columns are linearly dependent (the dimension of the column space is less than n) $\iff$ the rank of A is less than n $\iff$ the rows are linearly dependent $\iff$ A is not invertible $\iff$ $\det(A) = 0$ $\iff$ zero is an eigenvalue.

The results of these two theorems involve two if-then statements (an if and only if) between every pair of statements in the same theorem, making 110 different theorems altogether. We can prove the theorem much more quickly than writing out 110 different proofs, though. For example, if we want to show four statements are equivalent, then we can show that (1) implies (2), (2) implies

(3), (3) implies (4), and (4) implies (1). Then we will have shown that (1) also implies (3) and (4) because we can follow a circle of implications to obtain an implication between any two statements. To prove the theorem above, we need to show a chain of 11 implications that eventually circles back on itself.

Proof. [(1) implies (2)] If the system $A\vec{x} = \vec{0}$ has only the trivial solution, then the reduced row echelon form of $[A \mid \vec{0}]$ is $[I \mid \vec{0}]$. The row operations used to reduce this matrix will be the same as the row operations used to reduce $[A \mid \vec{b}]$. Hence the rref of $[A \mid \vec{b}]$ will be $[I \mid \vec{x}]$, where $A\vec{x} = \vec{b}$, so there will be a unique solution.

[(2) implies (3)] With a unique solution to $A\vec{x} = \vec{b}$ for any $\vec{b}$, the solution to $A\vec{x} = \vec{0}$ is $\vec{x} = \vec{0}$. This means that the rref of $[A|\vec{0}]$ is $[I|\vec{0}]$. Since removing the last column of zeros won't change any row operations, the rref of A is I.

[(3) implies (4)] Since the reduced row echelon form of A is I, then each column contains a leading 1, and hence is a pivot column

[(4) implies (5)] Since each column is a pivot column, the only solution to $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}$ (where $\vec{v}_i$ is the ith column of A) is the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. This is because in order for each column of A to be a pivot column, each row of A must have a leading 1 (since there are n rows as well as n columns). Notice that we have just shown (4) implies (1) as well, which means that (1) through (4) are all equivalent.

[(5) implies (6)] The definition of rank is the dimension of the column space. Since there are n pivot columns, the rank is n.

[(6) implies (7)] This is the result of Theorem 3.19, which also shows that (7) implies (5).

[(7) implies (8)] Since (7) is equivalent to (5), we know that the span of the column space is all of $\mathbb{R}^n$. This means in particular that the vectors listed are in the span of the columns of A.

[(8) implies (9)] To find the inverse of A, we need to solve $AB = I$ or $[A\vec{b}_1\ A\vec{b}_2\ \cdots\ A\vec{b}_n] = [\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n]$. This means we need to solve the equation $A\vec{b}_i = \vec{e}_i$ for $\vec{b}_i$ for each i, and this can be done because each $\vec{e}_i$ is in the column space of A. The inverse of A is thus $[\vec{b}_1\ \vec{b}_2\ \cdots\ \vec{b}_n]$.

[(9) implies (10)] We already showed that if a matrix has an inverse, then the determinant is not zero (otherwise $|A||A^{-1}| = |I|$ gives $0 = 1$).

[(10) implies (11)] If the determinant is not zero, then $|A - 0I| \neq 0$, which means 0 is not an eigenvalue.

[(11) implies (1)] If zero is not an eigenvalue, then the equation $A\vec{x} = 0\vec{x}$ cannot have a nonzero solution. This means that the only solution is the trivial solution. □

At this point, if we want to say that (10) implies (9) (determinant nonzero implies there is an inverse), then the proof above shows that (10) implies (11) which implies (1) which implies (2) and so on until it implies (9) as well. By creating a circle of implications, all of the ideas above are equivalent.

In the homework, I ask you to illustrate this theorem with a few matrices. Your job is to show that your example matrix either satisfies every single one of the statements in Theorem 3.41, or satisfies every single one of the statements in Theorem 3.42.

3.3 Vector Spaces

The notion of a vector space is the foundational tool upon which we will build for the rest of the semester. Vectors in the plane, vectors in space, and planes in space through the origin are all examples of vector spaces. We have looked at

the column space, row space, null space, and eigenspaces of a matrix as other examples of vector spaces. In all of the examples we have looked at, we were able to create the vector space by considering a basis and using the span of that basis to obtain the vector space. This is the fundamental idea we need to abstract vector spaces to a more general setting. Continuous and differentiable functions form vector spaces. Matrices of the same size form a vector space. Radio waves, light waves, electromagnetic waves, and more all form vector spaces. Mathematicians discovered some common properties about all of these kinds of objects, and created a definition of of the term "vector space" to allow them to focus on the key commonalities. The results which sprang from the following definition created linear algebra as we know it today.

Vector spaces are spans of sets of vectors.

Definition 3.18. A (real) *vector space* is a set V of things (which we will call "vectors") together with two operations:

(C1) *vector addition* $(+)$, which assigns to any pair $\vec{u}, \vec{v} \in V$ another vector $\vec{u} + \vec{v} \in V$,

(C2) *scalar multiplication* $(\cdot)$, which assigns to any $\vec{v} \in V$ and $c \in \mathbb{R}$ another vector $c\vec{v} \in V$.

These two operations must satisfy the following axioms ($c, d \in \mathbb{R}$ and $\vec{u}, \vec{v}, \vec{w} \in V$):

We use the word "real" to emphasize that the scalars are real numbers. For those of you who know about fields, you can replace "real" with any field, such as the complex numbers or finite fields used in computers or cryptography. For simplicity, we will stick to scalars being real numbers for now.

We say that the set V is closed under the operations of vector addition and scalar multiplication because adding vectors and scaling vectors produces other vectors in V.

Compare these properties to Theorems 3.1 and 3.10.

(A_1) Vector addition is associative: $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.

(A_2) Vector addition is commutative: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.

(A_3) There is a zero vector $\vec{0}$ in V which satisfies $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$.

(A_4) Every $\vec{v} \in V$ has an additive inverse, called $-\vec{v}$, which satisfies $\vec{v} + (-\vec{v}) = \vec{0}$.

(M_1) Scalar multiplication distributes across vector addition: $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$.

(M_2) Scalar multiplication distributes across scalar addition: $(c + d)\vec{v} = c\vec{v} + d\vec{v}$.

(M_3) Scalar multiplication is associative: $(cd)\vec{v} = c(d\vec{v})$

(M_4) Scalar multiplication by 1 does nothing: $1\vec{v} = \vec{v}$

The definition above is the formal definition of a vector space. The definition generalizes the properties we already use about vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. This definition essentially means you can add and multiply as you would normally expect—no surprises will occur. However, you can work with this definition if you think

> A vector space is a set of things ("vectors") closed under vector addition and scalar multiplication, and these operations obey the usual rules of addition and multiplication.

With vector addition and scalar multiplication properly defined, the span of any set of vectors from the vector space will always be a vector space. In other words, every linear combination of the vectors in that space must also be in the space. This means that lines and planes in $\mathbb{R}^3$ which pass through the origin are examples of vector spaces. As the semester progresses, we will discover that our intuition about what happens in $\mathbb{R}^2$ and $\mathbb{R}^3$ generalizes to all dimensions and can help us understand any kind of vector space. For now, let's look at some examples of vector spaces, and explain why they are vector spaces.

Example 3.1: Examples of vector spaces. Here are some examples of sets (and corresponding vector addition and scalar multiplication operations) that are vector spaces.

1. Both $\mathbb{R}^2$ and $\mathbb{R}^3$ are vector spaces. The set of all lists of real numbers of length n, $\mathbb{R}^n$, is a vector space. Vector addition is defined component-wise. Addition is associative, commutative, the zero vector is $(0, 0, \ldots, 0)$, and $(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = \vec{0}$. Also, scalar multiplication distributes across vector and scalar addition, it is associative, and $1\vec{v} = \vec{v}$. The definition of a vector space was defined to capture the properties of $\mathbb{R}^n$ which are useful in other settings.

2. The set of m by n matrices, written M_{mn}, with "vector addition" being normal matrix addition and "scalar multiplication" being normal matrix-scalar multiplication, is a vector space.

3. The set of all polynomials, written $P(x)$, with "vector addition" being the normal sum of polynomials, and "scalar multiplication" being the normal number-times-polynomial operation, is a vector space.

4. The set of continuous functions on an interval $[a, b]$, written $C[a, b]$, where "vector addition" is just adding the two functions together and "scalar multiplication" is just multiplying the function by a scalar, is a vector space.

5. The set of continuously differentiable functions on an interval (a, b), written $C^1(a, b)$, where "vector addition" is just adding the two functions together and "scalar multiplication" is just multiplying the function by a scalar, is a vector space.

Example 3.2: Examples that are *NOT* vector spaces. Here are some examples of sets that are not vector spaces along with an explanation of one of the properties they do not satisfy.

1. A line that does not pass through the origin. The zero vector is not on the line.

2. Polynomials of degree n. There is no zero vector.

3. Invertible Matrices. There is no zero matrix since it is not invertible.

4. The nonnegative x axis. There is no additive inverse for $(1, 0)$, as $(-1, 0)$ is not in the space.

5. The line segment from $(-1, 0)$ to $(1, 0)$. The product $2(1, 0) = (2, 0)$ is not in the space.

Example 3.3: The set of all polynomials is a vector space. To illustrate how to verify that a set is a vector space, let's formally verify that one of the examples above is indeed a vector space. Let's work with the set of all polynomials $P(x)$. Our vectors are polynomials, which are functions of the form

$$a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i.$$

We must define vector addition and scalar multiplication. If $a(x) = \sum_{i=0}^{n} a_i x^i$ and $b(x) = \sum_{i=0}^{m} b_i x^i$ are two polynomials, define their sum to be a new polynomial

$c(x)$ where $c_i = a_i + b_i$ for each i (where $a_i = 0$ if $i > n$ and $b_i = 0$ if $i > m$). This is just normal polynomial addition. For example, we compute the sum $(1 + 4x + 3x^2) + (x - x^3) = (1 + 0) + (4 + 1)x + (3 + 0)x^2 + (0 - 1)x^3$. If c is a real number, then scalar multiplication $c \cdot a(x)$ is a new polynomial $b(x)$ where $b_i = ca_i$ for each i. This is just the normal number-times-polynomial operation

First, we need to check that the "vector addition" of two polynomials is also in our set (i.e., the result is a polynomial). This is clear from the definition. We also need to check that the "scalar product" of a number and a polynomial is also a polynomial. This again is clear from the definition of scalar multiplication.

Finally, we need to show the 8 axioms hold:

(A_1) Polynomial addition is associative: $(a(x)+b(x))+c(x) = a(x)+(b(x)+c(x))$. To verify this, we look at each coefficient of the sum. The coefficients are $(a_i + b_i) + c_i = a_i + (b_i + c_i)$, and equality holds because addition of real numbers is associative.

(A_2) Polynomial addition is commutative: $a(x)+b(x) = b(x)+a(x)$. Again this follows because addition of real numbers is commutative. In particular, $a_i + b_i = b_i + a_i$ for each i.

(A_3) The zero polynomial $z(x) = 0$ is the zero vector, as $0 + a(x) = a(x) + 0 = a(x)$ for any polynomial.

(A_4) The additive inverse of $a(x) = a_0 + a_1 x + \cdots + a_n x^n$ is $-a(x) = -a_0 - a_1 x - \cdots - a_n x^n$.

(M_1) Scalar multiplication distributes across vector addition: $c(a(x) + b(x)) = ca(x) + cb(x)$. For each i, we compute $c(a_i + b_i) = ca_i + cb_i$.

(M_2) Scalar multiplication distributes across scalar addition: $(c + d)a(x) = ca(x) + da(x)$. For each i, we compute $(c + d)a_i = ca_i + da_i$.

(M_3) Scalar multiplication is associative: $(cd)a(x) = c(da(x))$. For each i, we compute $(cd)a_i = c(da_i)$.

(M_4) Scalar multiplication by 1 does nothing: $1a(x) = a(x)$. For each i, we compute $1a_i = a_i$.

We have now shown that the set of polynomials is a vector space.

3.3.1 Subspaces

The formal definition of a vector space is rather long. Trying to show that a set is a vector space can be a long and tedious process. Once we have taken the time to show that a set is a vector space, we can use this result to find many subspaces inside the vector space that are themselves vector spaces. We have done this already when we looked at the column space, row space, null space, and eigenspaces of a matrix. The notion of a subspace allows us to quickly find new vector spaces inside of vector spaces we already understand.

Definition 3.19. Suppose V is a vector space. If U is a set of vectors in V such that U is a vector space itself (using the addition and multiplication in V), then we say that U is a *vector subspace* of V and write $U \subset V$.

Example 3.4. We know that the vectors in a plane in $\mathbb{R}^3$ form a vector space. Each line through the origin that lies in the plane is a vector subspace of the plane. Each line and plane in $\mathbb{R}^3$ which passes through the origin is a vector subspace of $\mathbb{R}^3$.

We have also seen lots of examples of sets of vectors in a vector space that are *not* vector subspaces. We now state a key theorem which allows us to quickly verify that a set of vectors in a vector space is a actually vector subspace.

Theorem 3.43. Suppose U is a set of vectors in a vector space V and the following three conditions hold:

1. $0 \in U$

2. $\vec{u} + \vec{v} \in U$ for every $\vec{u}, \vec{v} \in U$ (the set is closed under vector addition)

3. $c\vec{u} \in U$ for every $\vec{u} \in U$ and $c \in \mathbb{R}$ (the set is closed under scalar multiplication)

Then U is a vector subspace of V.

The first item requires that the set actually contain at least one element. The second and third items basically state that every linear combination of vectors in U is also in U (the set is closed under linear combinations). A vector subspace is a nonempty set of vectors in which every linear combination of vectors is also in the set. The key value of this theorem is that it allows us to quickly show that many spaces are vector spaces. Let's start with some examples, and then we'll end with a proof.

Example 3.5. Here are some examples of vector subspaces.

1. The span of a set of vectors in a vector space V is always a subspace of V. This is because every linear combination is by definition in the span, so it is in the subspace. Hence, if M is an m by n matrix, the row space (the span of the rows) is a subspace of $\mathbb{R}^n$ and the column space is a subspace of $\mathbb{R}^m$.

2. The set of upper triangular m by n matrices is a subspace of M_{mn}. The zero matrix is in M_{mn}. The sum of two upper triangular matrices is upper triangular. Multiplying each entry of an upper triangular matrix by a scalar will still result in an upper triangular matrix.

3. The set of polynomials of degree less than or equal to n, written $P_n(x)$, is a subspace of $P(x)$. The zero polynomial ($p(x) = 0$, which has degree 0) is in the set. If two polynomials each have degree less than or equal to n, then their sum also has degree less than or equal to n. Also, multiplying a polynomial by a constant cannot increase the degree, so the set is closed under scalar multiplication.

4. More examples are given in Schaum's outlines on page 160 (examples 4.4, 4.5, and 4.6).

5. See Example 2.18 for some examples we've seen before of sets which are vector subspaces of $\mathbb{R}^n$.

Example 3.6. Here are some examples of sets that are *NOT* vector subspaces. To show a set of vectors in a vector space is not a vector subspace, you just need to find one example that shows the set is not closed under addition or scalar multiplication (or show that the set is empty).

1. See Example 2.19 for some examples we've seen before of sets which are not vector subspaces.

2. The set of polynomials of degree two (each polynomial *must* have an x^2 term) is not a vector subspace of the set of polynomials because the zero polynomial is not in this set. Also, the sum of x^2 and $1 - x^2$ is $(x^2) + (1 - x^2) = 1$, which is a polynomial of degree zero instead of degree 2 (so the set is not closed under addition).

Proof. The proof of this theorem is rather quick. Let V be the larger vector space, and U the subset of vectors that satisfy the 3 properties. From the conditions in the theorem, we know that U is nonempty and closed under vector addition and scalar multiplication. Because every vector in U is also a vector in V, we immediately know that vector addition is associative (A_1) and commutative (A_2), scalar multiplication distributes across vector addition (M_1) and distributes across scalar addition (M_2), scalar multiplication is associative (M_3), and multiplication by 1 does nothing (M_4). The first condition in the theorem shows that U has a zero vector (A_3). So all we need to show is that the additive inverse (A_4) of any vector in U is also in U. If $\vec{u} \in U$, then because $c\vec{u} \in U$ for every $c \in \mathbb{R}$, we can pick $c = -1$ to obtain $-1\vec{u} \in U$. This means that $u + (-1\vec{u}) = (1 - 1)\vec{u} = 0\vec{u} = \vec{0}$, which means every $u \in U$ has an additive inverse $-1\vec{u} \in U$ (A_4). We have now verified that U satisfies all the axioms required to be a vector space. $\qquad\square$

The first example of a vector subspace that we examined involved looking at the span of a set of vectors. Because any vector subspace is closed under linear combinations, the span of all vectors in the subspace is the subspace itself. This means that every vector subspace is the span of a set of vectors. This pattern will be used so often that we will state it as a theorem.

Theorem 3.44. If W is a subset of vectors in a vector space V, then W is a subspace of V if and only if W is the span of a set of vectors.

Example 3.7. Let's use the previous theorem to show that some sets are vector subspaces.

1. The set of 2 by 2 diagonal matrices is spanned by $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, and so is a vector subspace of M_{22}.

2. We already know that $P(x)$ is a vector space. The span of $\{1, x, x^2, x^3, x^4\}$ is the set of polynomials of the form $a + bx + cx^2 + dx^3 + ex^4$, which is the set of all polynomials of degree 4 or less, $P_4(x)$. Since $P_4(x)$ is the span of set of vectors, it is a vector subspace of $P(x)$.

3.3.2 Basis and Dimension

Vector spaces are often rather large sets. In the previous chapter, we found basis vectors for the column and row space of a matrix by row reducing a matrix. These basis vectors allow us to talk about the entire space by considering the span of the basis vectors. The coordinates of other vectors can then be given relative to the chosen basis. Let's review these ideas with an example of finding a basis for the column space of a matrix.

Example 3.8. The reduced row echelon form of $\begin{bmatrix} 1 & -1 & 1 \\ 3 & 0 & 6 \\ 5 & 1 & 11 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.
The column space is the vector space spanned by the column vectors $\vec{u}_1 = (1, 3, 5)$, $\vec{u}_2 = (-1, 0, 1)$, and $\vec{u}_3 = (1, 6, 11)$. The reduced row echelon form tells us that the third column is 2 times the first (u_1) plus the second (u_2). Since the third

column is already a linear combination of the first two columns, it does not contribute any new vectors to the span of just the first two vectors. Hence, the column space is the span of $\vec{u}_1$ and $\vec{u}_2$ (we can remove the dependent vector without shrinking the number of vectors in the span). Because the first two columns are linearly independent, we cannot remove either vector without reducing the number of vectors in their span.

A basis for the column space is a linearly independent set of vectors whose span is the the column space. Because $\vec{u}_1$ and $\vec{u}_2$ are linearly independent and they span the column space, they form a basis for the column space. The coordinates of the third vector relative to this basis are $(2, 1)$, the numbers in the third column of the rref. The dimension of the column space is the number of vectors in a basis for the column space, so here the dimension of the column space is 2. The column space is a plane of vectors in $\mathbb{R}^3$.

The example above illustrates the key ideas we need to work with any finite dimensional vector spaces. Let's now give the formal definitions.

Definition 3.20: Basis, Dimension, Coordinates. A *basis* for a vector space is a set of vectors that are (1) linearly independent and (2) span the vector space. The *dimension* of a vector space is the number of vectors in a basis for the vector space. We say the dimension of the zero vector space (the vector space consisting of only the zero vector) is zero. If $\beta = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis and $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$, then we call $(c_1, c_2, \ldots, c_n)_\beta$ the *coordinates* of $\vec{v}$ relative to the basis β. If the basis is understood, we just leave off the subscript β.

Remember, once you have a new definition, you should always try to illustrate that definition with examples. Let's now look at a few more examples.

Example 3.9. Let's start by looking at $\mathbb{R}^n$, a vector space we are familiar with. The standard basis for $\mathbb{R}^2$ is $\vec{e}_1 = (1, 0), \vec{e}_2 = (0, 1)$, so the dimension is 2. Similarly, $\vec{e}_1 = (1, 0, \ldots, 0), \vec{e}_2 = (0, 1, \ldots, 0), \ldots, \vec{e}_n = (0, 0, \ldots, 1)$ is a basis for $\mathbb{R}^n$, so the dimension is n.

Example 3.10. Let's find a basis for a polynomial vector space. Consider the set of polynomials $\{1, x, x^2\}$. The span of these polynomials is $a_0(1) + a_1(x) + a_2(x^2)$, which is every polynomial of degree less than or equal to 2. To show that this set forms a basis for $P_2(x)$, we only need to show that these polynomials are linearly independent. To show this, we must show that

$$a_0(1) + a_1(x) + a_2(x^2) = 0 \text{ implies } a_0 = a_1 = a_2 = 0$$

(the only linear combination which results in the zero polynomial is the trivial combination). Since the equation above holds for every real number x, we can simply plug in 3 different x values to obtain 3 different equations. Letting $x = 0, 1, -1$ gives us the three equations $a_0 = 0$, $a_0 + a_1 + a_2 = 0$, and $a_0 - a_1 + a_2 = 0$. Solving this system yields $a_0 = a_1 = a_2 = 0$. Hence, the three polynomials $1, x, x^2$ are linearly independent. These three polynomials form a basis for $P_2(x)$, and we call this the standard basis for $P_2(x)$. Using this basis, the coordinates of $a_0 + a_1 x + a_2 x^2$ relative to the standard basis are (a_0, a_1, a_2). Notice that dimension of $P_2(x)$ is 3 (one more than the maximum degree of the polynomials).

In general, the set $\{1, x, x^2, x^3, \ldots, x^n\}$ is called the standard basis for $P_n(x)$.

For vector spaces that are not $\mathbb{R}^n$, a simple way to work with vectors is to start by finding a basis. Once you have a basis, you can give the coordinates of every vector in the space relative to that basis. These coordinates can then be thought of as vectors in $\mathbb{R}^n$. Questions about the vector space can then be

reduced to questions about vectors in $\mathbb{R}^n$. As seen in the last example, the vector space $P_2(x)$ consists of all polynomials of the form $a_0 + a_1x + a_2x^2$. If we use the standard basis for polynomials, then the coordinates are simply (a_0, a_1, a_2). We can now use these coordinates to ask other questions about the vectors in $P_2(x)$. Here are two examples.

Example 3.11. Let V be the vector space spanned by $1 + x, x + x^2, 4 + x - 3x^2$. We'll show it is a 2-dimensional vector subspace of $P_2(x)$, and provide a basis for this space in two different ways. The coordinates of these three polynomials relative to the standard basis are $(1, 1, 0)$, $(0, 1, 1)$, and $(4, 1, -3)$. To study the span of the polynomials, we instead look at the column space of

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

Row reduction of A shows us that the first two columns of A are pivot columns, hence basis vectors for the column space of A. This tells us that the corresponding set of polynomials $\{1 + x, x + x^2\}$ is a basis for V. The coordinates of $4 + x - 3x^2$ relative to this basis are $(4, -3)$.

To find a different basis for V, we instead look at the row space of

$$A^T = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 4 & 1 & -3 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Row reduction tell us that $(1, 0, -1)$ and $(0, 1, 1)$ are basis vectors for the row space of A^T (hence the column space of A). The corresponding polynomials $1 - x^2$ and $x + x^2$ are basis vectors for V. The coordinates of $4 + x - 3x^2$ relative to this basis are $(4, 1)$.

Example 3.12. Let's show the vectors $1 = (1, 0, 0), 1 + x = (1, 1, 0), 1 + x + x^2 = (1, 1, 1)$ form a basis for $P_2(x)$, and then write $2 + 3x - 4x^2 = (2, 3, -4)$ as a linear combination of these vectors. To do this, we place the coordinates of each vector (relative to the standard basis) into the columns of a matrix, and then reduce

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & -4 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -4 \end{bmatrix}.$$

Since the first 3 columns are pivot columns, they are linearly independent. This implies that the polynomials represented by these vectors are linearly independent. Since the three sets of coordinates span all of $\mathbb{R}^3$, the three polynomials must span all of $P_2(x)$. The last column of the rref tells us that the coordinates of $(2, 3, -4)$ relative to $\beta = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ are $(-1, 7, -4)_\beta$. In terms of polynomials we write $2 + 3x - 4x^2 = -1(1) + 7(1 + x) - 4(1 + x + x^2)$. To emphasize the vocabulary, we can now state two things:

1. The coordinates of $2 + 3x - 4x^2$ relative to $\{1, x, x^2\}$ are $(2, 3, -4)$.

2. The coordinates of $2 + 3x - 4x^2$ relative to $\{1, 1 + x, 1 + x + x^2\}$ are $(-1, 7, -4)$.

We'll often start by using a standard basis to represent vectors, and then change the basis to a different one if the need arises. Later we will explore this idea in more depth. For now we'll focus on finding a basis and coordinates relative to that basis.

Example 3.13. As a final example, let's introduce the standard basis for the vector space of m by n matrices. We'll focus on 2 by 2 matrices, though this example generalizes to any size matrix. The set of matrices

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

forms a basis for the vector space M_{22} of 2 by 2 matrices (which means this space has dimension 4). To show it is a basis, we must show that (1) the matrices are linearly independent and (2) the matrices span M_{22}. To show linearly independent, we solve

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The only solution is the trivial solution, so the matrices are linearly independent. These 4 matrices clearly span the space, as the coordinates of any matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ relative to this basis are (a, b, c, d). We can now work with 2 by 2 matrices as a vector space by considering vectors in $\mathbb{R}^4$ instead.

Whenever we encounter vector spaces of finite dimension, we will often start by finding a basis and then work with coordinates relative to this basis instead. Then we can work with the coordinates as vectors in $\mathbb{R}^n$ to perform calculations. This means that once we understand what happens in $\mathbb{R}^n$, we can understand what happens in any vector space of dimension n.

3.4 Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Create examples to illustrate mathematical definitions. Know the definitions in this chapter.

2. Create examples to illustrate theorems. You do not have to memorize all the theorems in this chapter, but when one is given, make sure you can construct examples of it.

3. When a matrix has an inverse, this is equivalent to many other facts about the matrix and corresponding systems of equations. Illustrate this equivalence with examples, and be able to explain why the the ideas are equivalent.

4. Be able to explain why something is or is not a vector space. Give examples of vector spaces and subspaces that you have encountered previously in this class and other classes.

5. Explain how to find a basis for a vector space and the dimension of a vector space. Give the coordinates of vectors relative to a chosen basis.

Here are the preparation problems for this unit. Problems that come from Schaum's Outlines "Beginning Linear Algebra" are preceded by a chapter number. From here on out, many of the problems we will be working on come from Schaum's Outlines. Realize that sometimes the method of solving the problem in Schaum's Outlines will differ from how we solve the problem in class. The difference is that in Schaum's Outlines they almost always place vectors in rows prior to row reduction, whereas we will be placing vectors in columns. There are pros and cons to both methods.

Preparation Problems (click for handwritten solutions)	
Day 1	1f, 2a, 2f, 2i
Day 2	2j, 4b, Chp 4:6 Chp 4:13
Day 3	Chp 4:22 Chp 5:18 Chp 6:6 Chp 6:9
Day 4	Lesson Plan, Quiz, Start Project

3.4.1 Webcasts

Click here for webcasts on YouTube[i]. Click here for a pdf copy of the papers in the webcasts.

3.4.2 Suggested Problems

The first three sets of problems are in this book. The remaining problems come from Schaum's Outlines "Beginning Linear Algebra" by Seymour Lipschutz.

---

[i]http://www.youtube.com/user/bmwoodruff#g/c/54D7F4A63AE6AE18

Concept	Suggestions	Relevant Problems
Illustrating Definitions	1cdf	1
Illustrating Theorems	2acdfijl (all are good)	2
The huge equivalence	4abe	4
Vector Spaces	In Schaum's	5
Subspaces	In Schaum's	6
Spans	In Schaum's	7
Basis and Dimension	In Schaum's	**??**
Coordinates	In Schaum's	9
Linear Independence	In Schaum's	10

3.5 Problems

Much of the point to this section is to learn to read mathematics and create examples to illustrate definitions and theorems. For this reason, each problem below requires reading and working with different definitions and theorems. The common theme to the problems is that you are to create examples. The latter half of the problems are in the Schaum's Outlines "Beginning Linear Algebra" book.

1. **Illustrating Definitions:** For this problem, you are creating examples to illustrate a definition. If actual vectors or matrices are given, use them to illustrate the definition. Then create your own example using your own vectors and/or matrices.

   (a) Give an example of a pair of nonzero vectors that are orthogonal. Also give an example of a pair of vectors that are not orthogonal. Draw your vectors to show that the angles meet at a 90 degree angle.

   (b) Give an example of a matrix that is in row echelon form but not reduced row echelon form. Then give the reduced row echelon form of the matrix.

   (c) Show that the set of vectors $\{(1,1,0),(1,0,1),(0,1,1)\}$ is linearly independent by showing that the only solution to $c_1(1,1,0) + c_2(1,0,1) + c_3(0,1,1) = (0,0,0)$ is the trivial solution. Then select your own 3 vectors in $\mathbb{R}^3$ and determine if they are independent or dependent.

   (d) Consider the matrix $A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 0 & 6 \\ -1 & 3 & 2 & 0 \end{bmatrix}$ Find a basis for the column space and row space of A. What are the coordinates of each column relative to your basis vectors for the column space? Now select your own 3 by 4 matrix and repeat this problem.

   (e) Show that the inverse of $A = \begin{bmatrix} -2 & 0 & 5 \\ -1 & 0 & 3 \\ 4 & 1 & -1 \end{bmatrix}$ is $B = \begin{bmatrix} -3 & 5 & 0 \\ 11 & -18 & 1 \\ -1 & 2 & 0 \end{bmatrix}$ by computing both AB and BA. Now select your own 3 by 3 matrix A and use software to find the inverse $B = A^{-1}$ as well as compute AB and BA to verify that B is the inverse.

   (f) Consider the matrix $A = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 4 \\ 0 & 0 & 1 \end{bmatrix}$. Compute $A\vec{x}$ for each of the 4 vectors $\vec{x}_1 = (-1,1,0)$, $\vec{x}_2 = (2,0,3)$, $\vec{x}_3 = (1,1,0)$, and $\vec{x}_4 = (-4,0,1)$ to determine which are eigenvectors. If a vector is an eigenvector, state the corresponding eigenvalue. Now select your own 3 by 3 matrix A and use software to find a basis for the eigenspace (you may have to play with the numbers a little to get simple vectors, or grab an example from earlier in the text where you know the answer). For each vector in the eigenspace, compute $A\vec{x}$ and verify that $A\vec{x} = \lambda\vec{x}$.

2. **Illustrating Theorems:** Follow the instructions for each problem. For most of the problems below, you will start with a matrix and then use that matrix to illustrate several theorems from the theorem section of the textbook. After you have done so with the examples provided, you should select your own matrix and repeat the illustration.

(a) Which of the following 3 matrices are row equivalent?

$$A = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 3 & 13 \\ 1 & 2 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 & -2 \\ 3 & -2 & -3 \\ -1 & 0 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -1 & -1 \\ 4 & -3 & -1 \\ 3 & -1 & 3 \end{bmatrix}.$$

By row reducing each matrix, use Theorem 3.3 to make your decision.

Now try problem 4.22 on page 172 in Schaum's outlines.

(b) Consider the nonhomogeneous system of equations $x + 2y + 4w = 4, 2x + 3y - z = -1, 3x + 5y -$
$z + 4w = 3$, written in matrix form as $\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 3 & -1 & 0 \\ 3 & 5 & -1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix}$. Use Theorem3.4 to decide if
there is a solution to this equation. How many free variables are there based on Theorem3.7? Find two different solutions $\vec{x}_1$ and $\vec{x}_2$ to this system (let $z = 1, w = 0$ and $z = 0, w = 1$). Compute $A(\vec{x}_1 - \vec{x}_2)$ and use your result to verify **??**.

Now create your own 3 by 5 augmented matrix and repeat this problem.

(c) Consider the homogeneous system of equations $x + 2y + 4w = 0, 2x + 3y - z = 0, 3x + 5y - z + 4w = 0$,
written in matrix form as $\begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 3 & -1 & 0 \\ 3 & 5 & -1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. How many free variables are there based
on Theorem3.7? Find two different solutions $\vec{x}_1$ and $\vec{x}_2$ to this system (let $z = 1, w = 0$ and $z = 0, w = 1$). Illustrate Theorem3.8 by computing $A(\vec{x}_1 + \vec{x}_2)$.

Now create your own homogeneous system and repeat this problem.

(d) Consider the two matrices $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, B = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$. Verify Theorem3.11 by computing AB,
$(AB)^T$, $A^T B^T$ and $B^T A^T$ (notice that $A^T B^T$ doesn't even make sense).

Now repeat this problem with two matrices such that the product AB makes sense.

(e) Use the same two matrices from the previous problem to illustrate Theorem3.14. Table 1.1 on page 7 contains another example.

Now select two matrices of your own and repeat this problem.

(f) Consider the matrix $A = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 3 & -1 & 0 \\ 3 & 5 & -1 & 4 \end{bmatrix}$. Find the rref of A. Use Theorem 3.15 to determine
the rank of A. Use Theorem 3.16 to determine if the columns are linearly independent. Use Theorem 3.20 to give a basis for the column space as well as the coordinates of each column of A relative to this basis. Use Theorem 3.21 to give a basis for the row space of A as well as the coordinates of each row of A relative to this basis. Are the dimensions of the column and row space the same (as Theorem 3.19 suggests)?

Now repeat this problem with a matrix of your own.

(g) Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$. Find the inverse of A by row reducing $[A|I]$. Then find the inverse of A^{-1} by
row reducing $[A^{-1}|I]$ to illustrate Theorem 3.23. Now compute $(A^T)^{-1}$ and $(A^{-1})^T$ to illustrate Theorem 3.24.

Repeat this problem with a 2 by 2 or 3 by 3 matrix of your own.

(h) Consider the matrices $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix}$. Illustrate Theorem 3.25 by computing A^{-1}, B^{-1}, $(AB)^{-1}$, and $B^{-1}A^{-1}$.

Repeat this on a pair of 2 by 2 or 3 by 3 matrices of your own.

(i) Consider the triangular 4 by 4 matrix $A = \begin{bmatrix} 1 & 8 & -12 & 0 \\ 0 & 2 & 5 & 16 \\ 0 & 0 & -3 & -4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$. Find the determinant of A (make

sure you use a cofactor expansion along columns that contain lots of zeros). Similarly find the characteristic polynomial of A (the determinant of $A - \lambda I$). The polynomial should already be factored, so find the eigenvalues of A. Compare your results to theorems 3.28 and 3.36.

Now select a 4 by 4 or 5 by 5 triangular matrix and repeat this problem.

(j) Consider the 3 by 3 matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 3 & 4 \\ 2 & -3 & 1 \end{bmatrix}$. Find the determinant of A by using a cofactor

expansion along row 1. Now find the determinant of A^T by using a cofactor expansion along column 1. Why does your work verify Theorem 3.30? Show that the characteristic polynomial of A and A^T is the same by computing the determinants of $A - \lambda I$ and $A^T - \lambda I$ using the first row and first column. Why does this imply Theorem 3.37. Notice that without even computing the eigenvalues, we can conclude that they are the same for A and A^T because the characteristic polynomials are the same.

Repeat this problem on a different 3 by 3 matrix.

(k) Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & -3 \\ 2 & 4 & 6 \end{bmatrix}$. Compute the determinant of A. Now use Theorem 3.32

to explain why the determinant is zero. Use Theorem 3.31 to explain why A has no inverse.

Create a 4 by 4 matrix in which one row is a multiple of another row. Then repeat this problem.

(l) Consider the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 3 & 0 \\ 4 & -1 & 2 \end{bmatrix}$. Compute each of the cofactors of A. Find the adjoint of

A (place the C_{ij} cofactor in the ith column, jth row). Let $B = \frac{1}{|A|}\text{adj}(A)$. Show by computing AB that B is the inverse of A (Theorem 3.33).

Pick another 3 by 3 matrix that has an inverse, and repeat this problem.

(m) Consider the matrix $A = \begin{bmatrix} 1 & 2 & -4 & 3 \\ 1 & 3 & 7 & -9 \\ -4 & -1 & 2 & 5 \\ 0 & 3 & -4 & 3 \end{bmatrix}$. What is the sum of each row of A? Show that

$\vec{x} = (1, 1, 1, 1)$ is an eigenvector of A by computing $A\vec{x}$. What is the corresponding eigenvalue (it should be the sum of each row)? You have essentially showed why Theorem 3.38 is true. Use Theorem 3.37 to decide if $\lambda = 2$ is an eigenvalue of A^T. [This problem is why $\lambda = 1$ is an eigenvalue of the transition matrices in a Markov process.]

Create a 4 by 4 matrix where each row sums to the same number. Repeat the computations above.

(n) Consider the symmetric matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$. Find the 3 distinct eigenvalues of A, and a

basis vector for each corresponding eigenspace. Find the dot product of each pair of vectors and compare your result to Theorem 3.39. Show that the three eigenvectors are linearly independent as Theorem 3.40 suggests (rref a matrix with these eigenvectors as column vectors).

Repeat this problem with another symmetric matrix.

3. **Proving Theorems:** Prove the following theorems. (We will not be focusing on this in our class.)

4. **A Huge Equivalence:** Illustrate Theorem 3.41 (the huge equivalence) with the following matrices. First determine if the matrix is singular or nonsingular, and then verify the 11 items from the theorem. Feel free to use a computer to do your computations.

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

(e) Create your own 3 by 3 or 4 by 4 matrix. Illustrate the theorem with this matrix.

5. **Vector Spaces:** See chapter 4 problems 1–6, 38. Complete at least 6, 38.

(a) Show that the set M_{22} of 2 by 2 matrices using regular matrix addition and regular scalar matrix multiplication is a vector space by showing that it satisfies all 8 axioms.

(b) Let V be the set of real valued functions $y = f(x)$ defined on the entire real number line. Define vector addition as function addition $(f + g)(x) = f(x) + g(x)$, and scalar multiplication as $(cf)(x) = c(f(x))$. Show that V is a vector space with these two operations.
Could you have changed "the entire real number line" to any interval?

(c) This example shows that vector addition and multiplication can be defined in very different ways and still result in a vector space. Let V be the set of vectors (a, b) in the plane such that both $a > 0$ and $b > 0$. If (a, b) and (c, d) are elements of V, define vector addition to be $(a, b) + (c, d) = (ac, bd)$. Define scalar multiplication by a real number x to be $x(a, b) = (a^x, b^x)$. Show that V is a vector space with zero vector $(1, 1)$.

(d) Let V be the set of vectors (a, b) in $\mathbb{R}^2$. If we use regular vector addition but redefine scalar multiplication to be $k(a, b) = (2ka, 2kb)$, does this make V a vector space? (Check the 4 axioms related to scalar multiplication.)

(e) Let V be the set of vectors (a, b) in $\mathbb{R}^2$. Explain why the following ways of defining vector addition and scalar multiplication do not make V a vector space. Find an axiom that is not satisfied.

i. Usual addition $(a, b) + (c, d) = (a+c, b+d)$, but define scalar multiplication as $k(a, b) = (a, kb)$.
ii. Define addition $(a, b) + (c, d) = (a+c, b)$, and use regular scalar multiplication $k(a, b) = (ka, kb)$.

(f) Let V be the set of vectors (a, b) in $\mathbb{R}^2$. Explain why the following ways of defining vector addition and scalar multiplication do not make V a vector space. Find an axiom that is not satisfied.

i. Usual addition $(a, b) + (c, d) = (a+c, b+d)$, but define scalar multiplication as $k(a, b) = (k^2a, kb)$.
ii. Define addition $(a, b) + (c, d) = (a, d)$, and use regular scalar multiplication $k(a, b) = (ka, kb)$.

6. **Vector Subspaces:** See chapter 4 problems 13–15, 45–46. Complete at least 13, 15, 46.

(a) Which of the following subsets of $\mathbb{R}^2$ are subspaces? Explain.

i. The y axis, $\{(0, y)|\ y \in \mathbb{R}\}$.
ii. The positive x-axis: $\{(x, 0)|\ x > 0\}$.
iii. The nonnegative x-axis: $\{(x, 0)|\ x \geq 0\}$.
iv. The line $y = 3x$: $\{(x, 3x)|\ x \in \mathbb{R}\}$.
v. The line $y = 3$: $\{(x, 3)|\ x \in \mathbb{R}\}$.
vi. The x and y axes: $\{(x, y)|\ x = 0 \text{ or } y = 0\}$.
vii. The points inside the unit circle: $\{(x, y)|x^2 + y^2 \leq 1\}$.

(b) Let V be the set of real valued functions $f(x)$ that are continuous on $[0, 1]$. Show that V is a subspace of the vector space of real valued functions defined on $[0, 1]$.

(c) Show that the set $P(x)$ of polynomials is a vector subspace of the vector space of real valued function defined on the entire reals.

(d) Show that the set $P_3(x)$ of polynomials of degree 3 or less together with the zero polynomial is a vector subspace of $P(x)$ in two ways:

    i. Use Theorem 3.43

    ii. Use Theorem 3.44

(e) Let V be the set of 3 by 3 upper triangular matrices. Show that V is a subspace of the vector space M_{33} of 3 by 3 matrices in 2 ways:

    i. Use Theorem 3.43

    ii. Use Theorem 3.44

(f) Which of the following subsets of M_{22}, the vector space of 2 by 2 matrices, are vector subspaces?

    i. The set of 2 by 2 symmetric matrices.

    ii. The set of 2 by 2 invertible matrices.

    iii. The set of 2 by 2 upper triangular matrices.

(g) Which of the following subsets of $P(x)$, the vector space of polynomials, are vector subspaces?

    i. The set of polynomials of degree 2.

    ii. The set of polynomials of degree 2, together with the zero polynomial.

    iii. The set of polynomials of degree 2 or less, together with the zero polynomial.

    iv. The set of polynomials of degree 2 or more, together with the zero polynomial.

7. **Spans:** (solutions link) See chapter 4 problems 18, 19, 21–23, 51–53, 58–61. Complete at least 22, 53.

(a) Which of the following matrices have the same row space?

$$
\text{i. } \begin{bmatrix} 1 & 2 & -1 \\ 0 & -2 & 4 \\ -1 & 2 & -7 \end{bmatrix} \quad
\text{ii. } \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 4 \\ 1 & -1 & 5 \end{bmatrix} \quad
\text{iii. } \begin{bmatrix} 1 & 2 & -3 \\ 0 & 1 & -2 \\ 2 & -1 & 4 \end{bmatrix} \quad
\text{iv. } \begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 8 \\ 0 & 2 & -4 \\ 4 & 1 & 10 \end{bmatrix}
$$

(b) Which of the following matrices have the same column space? [Hint: row reduce the transpose.]

$$
\text{i. } \begin{bmatrix} 1 & 0 & -1 \\ 4 & -2 & 0 \\ 7 & -6 & 5 \end{bmatrix} \quad
\text{ii. } \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 5 & -2 \end{bmatrix} \quad
\text{iii. } \begin{bmatrix} 1 & 0 & 2 \\ 4 & 1 & 3 \\ 5 & 2 & 0 \end{bmatrix} \quad
\text{iv. } \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 3 & 2 & 9 \\ 1 & -1 & 6 & 7 \end{bmatrix}
$$

(c) Which of the following sets of polynomials span $P_2(x)$?

    i. $1 + x, 2x + 3x^2, 2 - x + 4x^2$

    ii. $1 + x, 2x + 3x^2, 3 + x - 3x^2$

    iii. $1 + 2x - x^2, 1 + 2x$

    iv. $1 + 2x - x^2, 1 + 2x, 1 + x^2, 2 + x^2$

    v. $1 + 2x - x^2, 1 + 2x, -x^2, 3 + 6x - x^2$

(d) Which of the following sets of polynomials span $P_3(x)$?

    i. $1 + 2x - x^2, 1 + 2x + 3x^3, x + 3x^2 - 2x^3$

    ii. $1 + 2x - x^2, 1 + 2x + 3x^3, x + 3x^2 - 2x^3, x$

    iii. $1 + 2x - x^2, x^2 + 3x^3, 1 + 2x + 3x^3, x + 3x^2 - 2x^3$

(e) Which of the following sets of matrices span M_{22}?

$$
\text{i. } \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 2 & -3 \end{bmatrix}, \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}
$$

$$
\text{ii. } \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 2 & -3 \end{bmatrix}, \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 13 \\ 7 & 11 \end{bmatrix}
$$

$$
\text{iii. } \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 7 & 0 \\ 1 & 4 \end{bmatrix}
$$

$$
\text{iv. } \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}
$$

8. **Basis and Dimension:**

See chapter 5 problems 9–11, 14–18, 59, 60, 66, 67, 70–72. Complete at least 10, 18, 60, 67, 71.

The rest of the problems in this unit I will have to ask you to do from Schaum's. The main point to each of the remaining problems is to first start by converting each vector to its coordinates relative to a standard basis, and then row reducing. While completing the problem is as simple as row reducing a matrix, coming up with the problems is incredibly time consuming. I'll eventually make sufficient problems here, but for now you need to refer to Schaum's.

(a) Find a basis and dimension for the subspace spanned by the polynomials.

   i. $1 - 3x + 2x^2, -1 + 3x - 2x^2, 2 - 6x + 4x^2$
   ii. $2 + x, 6 - 3x + 2x^2, 3x + x^2$

(b) Find a basis and dimension for the subspace spanned by the matrices.

   i. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}, \begin{bmatrix} -1 & -2 \\ -3 & -4 \end{bmatrix}$

9. **Coordinates of vectors:** See chapter 6 problems 1–7, 22–26. Complete at least 6, 25.

Additional problems related to this idea (but without using the vocabulary "coordinates of a vector") are in chapter 4 problems 8–12, 40–44.

(a) Find the coordinates of the matrix A relative to the given basis:

   i. $A = \begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix}; \left\{ \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right\}$

10. **Linear Independence:** See chapter 6 problems 8–10, 27–28. Complete at least 8,9.

Additional problems related to this topic are in chapter 5 problems 1–5, 7, 48–51. These problems do not utilize the vocabulary of coordinates of a vector relative to a basis, and hence require a more difficult approach. Try one of these problems so that you see how the language of bases and coordinates can greatly simplify our work.

(a) Which sets of matrices are linearly independent?

   i. $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$

3.6   Projects

1. In this project, you will explore different subspaces of polynomials. As there are infinitely many ways to give a basis for a vector space, we need to find a way to determine when two different bases yield the same vector space.

Consider the following sets of polynomials:

$$S_1 = \{1, 2x, 3x^2\}$$
$$S_2 = \{2 + 2x, 2x + 3x^2, 2 + 4x + 3x^2\}$$
$$S_3 = \{1 + x, 2x + 3x^2, 2 - x + 4x^2\}$$
$$S_4 = \{1 + x^2, x + x^2\}$$
$$S_5 = \{-2 + 4x + 9x^2, 2 - 6x - 12x^2\}$$

For each i, let V_i be the vector space spanned by S_i. Your job is to determine which sets are bases, and then to determine which vector spaces are equal. Here are the steps:

(a) For all the polynomials above, start by obtaining the coordinates of each polynomial relative to the standard basis $\{1, x, x^2\}$.

(b) For each of the 5 sets, place the coordinate vectors into the columns of a matrix and row reduce. Show that S_2 is the only set that is linearly dependent. This shows that S_2 is not a basis, and the other sets are a basis for the space they span (V_2, V_4 and V_5 are 2 dimensional vector subspaces of $P_2(x)$).

(c) Are V_2 and V_4 the same vector space? If they are the same vector space, then the vectors in S_4 would be linear combinations of the vectors in S_2. Row reduce the matrix $\begin{bmatrix} 2 & 0 & 2 & 1 & 0 \\ 2 & 2 & 4 & 0 & 1 \\ 0 & 3 & 3 & 1 & 1 \end{bmatrix}$. Use your row reduction to explain why the vectors in S_4 are not linear combinations of the vectors in S_2. This means that V_2 and V_5 are not the same vector space.

(d) Are V_2 and V_5 the same vector space? Row reduce $\begin{bmatrix} 2 & 0 & 2 & -2 & 2 \\ 2 & 2 & 4 & 4 & -6 \\ 0 & 3 & 3 & 9 & -12 \end{bmatrix}$. Show that the coordinates of $(-2, 4, 9)$ and $(2, -6, -12)$ relative to the first two columns of S_2 are $(-1, 3)$ and $(1, -4)$. This shows that the vectors in S_5 are a linear combination of the vectors in S_2 , which means that V_5 is a subspace of V_2.

Now row reduce $\begin{bmatrix} -2 & 2 & 2 & 0 & 2 \\ 4 & -6 & 2 & 2 & 4 \\ 9 & -12 & 0 & 3 & 3 \end{bmatrix}$. Find the coordinates of all three vectors in S_2 relative to the vectors in S_5. We now know that the vectors in S_2 are a linear combination of the vectors in S_5, which means that V_2 is a subspace of V_5.

Since $V_2 \subset V_5$ and $V_5 \subset V_2$, we know know that $V_2 = V_5$.

(e) There is a simpler way to check if two sets span the same space. It boils down to picking the right basis to answer this question. Recall from the previous project that the standard basis for a vector space is found by placing the vectors in rows (instead of columns) and then row reducing.

For each of the 5 sets, place the coordinate vectors into the rows of a matrix and row reduce. You can quickly do this by going back to your code from a) and row reducing the transpose. The nonzero rows in each case give you the standard basis. You should see that $V_1 = V_3$ with standard basis $\{1, x, x^2\}$, and $V_2 = V_5$ with standard basis $\{1 - 3/2x^2, x + 3/2x^2\}$, while S_4 is the standard basis for V_4.

In summary, picking the right basis to solve a problem can make the problem trivial. We did two things in the work above: (1) we found the coordinates of vectors relative to a basis, and (2) we determined if two vector spaces were equal.

- We can quickly finding the coordinates of vectors by placing the vectors into columns and row reducing. The coordinates show up in the non pivot columns of rref.

- We can quickly determine if two sets of vectors span the same space by placing the vectors into the rows of two matrices and row reducing. The standard basis for each space is the nonzero rows of rref. The two vector spaces are equal if and only if these nonzero rows are identical.

Much of what we do throughout the rest of this semester is directly related to learning to pick the right basis in which to solve a problem. Much of upper level mathematics, statistics, engineering, and physics are directly related to picking the right basis in which to solve a problem. With the right basis, problems become trivial.

3.7   Solutions

1. **Definitions:**

   (a) Orthogonal: $(1, 2) \cdot (-4, 2) = 0$. Not: $(1, 2) \cdot (0, 1) \neq 0$. Answers will vary.

   (b) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}$. Row echelon form doesn't require zeros above pivots, whereas reduced row echelon form requires zeros above each pivot. Row echelon form is not

unique.

(c) $\begin{bmatrix} 1 & 1 & 0 & | & 0 \\ 1 & 0 & 1 & | & 0 \\ 0 & 1 & 1 & | & 0 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \end{bmatrix}$, hence

the only solution is $c_1 = c_2 = c_3 = 0$. The only way you'll get a solution other than the trivial solution is if there is a free variable, hence one column is not a pivot column.

(d) A reduces to $\begin{bmatrix} 1 & 0 & -\frac{4}{5} & \frac{9}{5} \\ 0 & 1 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Column space basis: $\{(1, 2, -1), (2, 4, 3)\}$.

Row space basis: $\{(1, 0, -\frac{4}{5}, \frac{9}{5}), (0, 1, \frac{2}{5}, \frac{3}{5})\}$.

Coord. of $(1, 2, -1)$ are $(1, 0)$.

Coord. of $(2, 4, 3)$ are $(0, 1)$.

Coord. of $(0, 0, 2)$ are $(-4/5, 2/5)$.

Coord. of $(3, 6, 0)$ are $(9/5, 4/5)$.

(e) The product is the identity either way. Hence B satisfies the definition of being the identity, $AB = BA = I$.

(f) $A\vec{x}_1 = (-1, 1, 0) = 1\vec{x}_1$ so $\lambda = 1$;

$A\vec{x}_2 = (16, 14, 3) \neq \lambda\vec{x}_2$ so not an eigenvector;

$A\vec{x}_3 = (3, 3, 0) = 3\vec{x}_3$ so $\lambda = 3$;

$A\vec{x}_4 = (-4, 0, 1) = 1\vec{x}_4$ so $\lambda = 1$;

2. **Illustrating Theorems:**

(a) A and C are row equivalent because they have the same rref.

(b) The rref is $\begin{bmatrix} 1 & 0 & -2 & -12 & | & -14 \\ 0 & 1 & 1 & 8 & | & 9 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$. The last column is not a pivot column, so both A and $[A|b]$ have two pivot columns. Hence there is a solution. With 4 variables and 2 pivot columns, there are $4 - 2 = 2$ free variables.

The general solution is $(-14, 9, 0, 0) + z(2, -1, 1, 0) + w(12, -8, 0, 1)$. Two solution are $(-12, 8, 1, 0)$ and $(-2, 1, 0, 1)$. The difference is $\vec{x}_1 - \vec{x}_2 = (-10, 7, 1, -1)$ and $A(-10, 7, 1, -1) = (0, 0, 0)$.

(c) The rref is $\begin{bmatrix} 1 & 0 & -2 & -12 & | & 0 \\ 0 & 1 & 1 & 8 & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$. Unknowns $= 4$. Pivots $= 2$. Free variables $= 4$-2.

The general solution is $z(2, -1, 1, 0) + w(12, -8, 0, 1)$. Two solutions are $\vec{x}_1 = (2, -1, 1, 0)$, $\vec{x}_2 = (12, -8, 0, 1)$, sum is $\vec{w} = (14, -9, 1, 1)$ and $A\vec{w} = (0, 0, 0)$. The sum of any two solutions is again a solution.

(d) $AB = \begin{bmatrix} 9 \\ 21 \end{bmatrix}$, $(AB)^T = \begin{bmatrix} 9 & 21 \end{bmatrix}$, $A^T B^T$ is not defined, $B^T A^T = \begin{bmatrix} 9 & 21 \end{bmatrix}$.

(e) $AB = \left[\begin{pmatrix} 1 \\ 4 \end{pmatrix}(2) + \begin{pmatrix} 2 \\ 5 \end{pmatrix}(-1) + \begin{pmatrix} 3 \\ 6 \end{pmatrix}(3) \right] = \begin{bmatrix} 1(2) + 2(-1) + 3(3) \\ 4(2) + 2(-1) + 6(3) \end{bmatrix}$

(f) rref is $\begin{bmatrix} 1 & 0 & -2 & -12 \\ 0 & 1 & 1 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Two pivot columns means rank is 2. Not every column is a pivot column, so dependent.

Column space basis: $\{(1, 2, 3), (2, 3, 5)\}$.

Coord. of $(1, 2, 3)$ are $(1, 0)$. Coord. of $(2, 3, 5)$ are $(0, 1)$. Coord. of $(0, -1, -1)$ are $(-2, 1)$. Coord. of $(4, 0, 4)$ are $(-12, 8)$. (These coordinates come from the columns of the nonzero rows of rref.)

Row space basis: $\{(1, 0, -2, -12), (0, 1, 1, 8)\}$. Coord. of $(1, 2, 0, 4)$ are $(1, 2)$. Coord. of $(2, 3, -1, 0)$ are $(2, 3)$. Coord. of $(3, 5, -1, 4)$ are $(3, 5)$. (These coordinates come from the rows of the pivot columns.)

(g) $\begin{bmatrix} 1 & 2 & | & 1 & 0 \\ 3 & 5 & | & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & | & -5 & 2 \\ 0 & 1 & | & 3 & -1 \end{bmatrix}$,

$\begin{bmatrix} -5 & 2 & | & 1 & 0 \\ 3 & -1 & | & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & | & 1 & 2 \\ 0 & 1 & | & 3 & 5 \end{bmatrix}$,

$(A^T)^{-1} = (A^{-1})^T = \begin{bmatrix} -5 & 3 \\ 2 & -1 \end{bmatrix}$.

(h) $A^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$, $B^{-1} = \begin{bmatrix} 3 & -2 \\ -4 & 3 \end{bmatrix}$, $(AB)^{-1} = B^{-1}A^{-1} = \begin{bmatrix} -21 & 8 \\ 29 & -11 \end{bmatrix}$.

(i) $|A| = (1)(2)(-3)(4) = -24$. Cofactor along column 1 at each step. Characteristic polynomial is $|A - \lambda I| = (1 - \lambda)(2 - \lambda)(-3 - \lambda)(4 - \lambda)$, which means $\lambda = 1, 2, -3, 4$. The diagonal entries are the eigenvalues.

(j) $|A| = 33 = |A^T|$. $|A - \lambda I| = |A^T - \lambda I| = (1 - \lambda)[(3 - \lambda)(1 - \lambda) + 12] - 2[(-1)(1 - \lambda) - 8]$. Since the characteristic polynomials are the same, the eigenvalues are the same.

(k) $|A| = 0$, row 3 is twice row 1. Since the determinant is zero, there is no inverse.

(l) The matrix of cofactors is $C_{ij} = \begin{bmatrix} 6 & -2 & -13 \\ -4 & 2 & 9 \\ 0 & 0 & 1 \end{bmatrix}$. The adjoint is $\text{adj}(A) = \begin{bmatrix} 6 & -4 & 0 \\ -2 & 2 & 0 \\ -13 & 9 & 1 \end{bmatrix}$. Determinant is $|A| = 2$.

Inverse is $\frac{1}{|A|}\text{adj}(A) = \begin{bmatrix} 3 & -2 & 0 \\ -1 & 1 & 0 \\ -\frac{13}{2} & \frac{9}{2} & \frac{1}{2} \end{bmatrix}$.

(m) Each row sums to 2. Just compute $A(1, 1, 1, 1) = (2, 2, 2, 2)$ to show $\lambda = 2$ is an eigenvalue corresponding to $(1, 1, 1, 1)$.

(n) The eigenvalues are $\lambda = 1, 3, 4$ with corresponding eigenvectors $(1, -1, 0), (1, 1, 0), (0, 0, 1)$. All three dot products are zero. Placing the three vectors into the columns of a matrix and reducing gives the identity, so they are independent.

3. **Proofs:**

4. **A Huge Equivalence:**

(a) (1) only the trivial solution, (2) a unique solution, (3) rref is I, (4) every column is a pivot, (5) columns are independent, (6) rank is 2, (7) rows are independent, (8) $\begin{bmatrix} 2 & 1 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}}$ $\begin{bmatrix} 1 & 0 & 4/5 & -1/5 \\ 0 & 1 & -3/5 & 1/5 \end{bmatrix}$ so $(1, 0)$ and $(0, 1)$ are in column space, (9) $A^{-1} = \begin{bmatrix} 4/5 & -1/5 \\ -3/5 & 1/5 \end{bmatrix}$, (10) $|A| = 5 \neq 0$, (11) $\lambda = 5, 1$ (none are zero).

(b) (1) Infinitely many solutions, (2) not unique solutions, (3) rref is not I, (4) column 2 is not a pivot, (5) columns are dependent, (6) rank is 1 (less than 2), (7) rows are dependent, (8) $\begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 1/2 & 0 & 1/4 \\ 0 & 0 & 1 & -1/2 \end{bmatrix}$ so $(1, 0)$ is not in column space, (9) A^{-1} does not exist, (10) $|A| = 0$, (11) $\lambda = 4, 0$ (zero is an eigenvalue).

(c) The only solution to the homogeneous system is zero, which means all 11 items are true. In particular, $A^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{9} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{5}{9} \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$, $|A| = 9$, $\lambda = 3, 3, 1$.

(d) There is more than one solution to the homogeneous system. So all 11 items are false. $|A| = 0$. The eigenvalues are $\frac{1}{2}(-1 - \sqrt{21}), \frac{1}{2}(-1 + \sqrt{21}), 0$ (zero is an eigenvalue).

5. **Vector Spaces:**

(a) Theorem 3.10 basically states this.

(b) Function addition is associative $(f + g) + h = f + (g + h)$ and commutative $f + g = g + f$ because addition on real numbers is. The zero function $f(x) = 0$ is the additive identity, and additive inverses are $(-f)(x) = -f(x)$. Scalar multiplication distributes across vector addition $c(f + g)(x) = cf(x) + cg(x)$ and scalar addition $((c + d)f)(x) = cf(x) + df(x)$ because of the distributive rules for real numbers. We also know that $c(df(x)) = (cd)f(x)$ and $1(f(x)) = f(x)$ from the properties of real numbers.

You could change "the entire real number line" to any interval, and nothing above would change.

(c) Vector addition is $(a, b) + (c, d) = (ac, bd)$. Scalar multiplication is $x(a, b) = (a^x, b^x)$. Vector addition is closed because multiplication of two positive numbers is positive. Scalar multiplication is closed because $a^x > 0$ as long as $a > 0$.

(A1) $((a, b) + (c, d)) + (e, f) = (ac, bd) + (e, f) = (ace, bdf) = (a, b) + (ce, df) = (a, b) + ((c, d) + (e, f))$.

(A2) $(a, b) + (c, d) = (ac, bd) = (ca, db) = (c, d) + (a, b)$. The zero vector $(1, 1)$ satisfies $(1, 1) + (a, b) = (1a, 1b) = (a, b) + (1, 1)$ as required.

(A3) $(1, 1) + (a, b) = (a, b) = (a, b) + (1, 1)$.

(A4) The additive inverse of (a, b) is $(1/a, 1/b)$, since $(1/a, 1/b) + (a, b) = (1, 1) = (a, b) + (1/a, 1/b)$.

(M1) $k[(a, b) + (c, d)] = k(ac, bd) = ((ac)^k, (bd)^k) = (a^k c^k, b^k d^k) = (a^k, b^k) + (c^k, d^k) = k(a, b) + k(c, d)$.

(M2) $(j + k)(a, b) = (a^{j+k}, b^{j+k}) = (a^j a^k, b^j b^k) = (a^j, b^j) + (a^k, b^k) = j(a, b) + k(a, b)$.

(M3) $j(k(a, b)) = j(a^k, b^k) = (a^{jk}, b^{jk}) = jk(a, b)$.

(M4) $1(a, b) = (a^1, b^1) = (a, b)$.

The following image shows you what 1 dimensional subspace of the logarithmic vector space look like. These "straight" lines through the origin $(1, 1)$ do not appear straight at all when graphed in the regular Cartesian coordinate system.
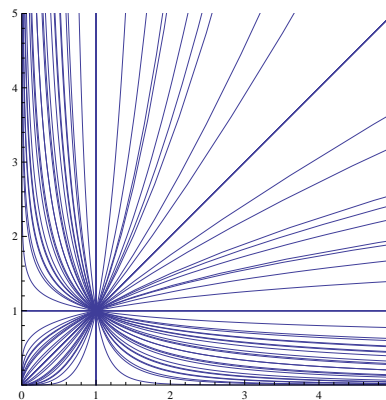
(d) The 3rd and 4th axioms for scalar multiplication fail. (M3) $j(k(a, b)) = j(2ka, 2kb) = (2j2ka, 2j2kb) = jk(2a, 2b)$ instead of $jk(a, b)$. (M4) $1(a, b) = (2a, 2b) \neq (a, b)$.

(e) i. Distribution across scalar multiplication fails: $(j + k)(a, b) = (a, (j + k)b) = (a, jb + kb)$ whereas $j(a, b) + k(a, b) = (a, jb) + (a, kb) = (2a, (j + k)b)$ (notice that a is doubled if we multiply first before adding).

   ii. Vector addition is not commutative. $(a, b) + (c, d) = (a + c, b)$ while $(c, d) +$

$(a, b) = (a + c, d)$ (the second components are not the same, $b \neq d$.

(f) i. Distribution across scalar multiplication fails. (M2) $(j + k)(a, b) = ((j + k)^2 a, (j + k)b)$ whereas $j(a, b) + k(a, b) = (j^2 a, jb) + (k^2 a, b) = ((j^2 + k^2)a, (j + k)b)$, and $j^2 + k^2 \neq (j + k)^2$.

ii. Vector addition is not commutative. $(a, b) + (c, d) = (a, d)$ while $(c, d) + (a, b) = (c, b)$.

6. **Vector Subspaces:** See chapter 4 problems 13-15, 45-46. Complete at least 13, 15, 46.

(a) i. Yes. It contains zero, the sum of any two vectors on the y axis is on the y axis, and any scalar multiple of a vector on the y axis is still on the y axis. The span of $(0, 1)$ is the y axis.

ii. No. The zero vector is not on the positive x axis.

iii. No. You cannot times vectors by negative scalars, so it's not closed under scalar multiplication.

iv. Yes. It is the span of $(1, 3)$. Lines through the origin are vector subspaces.

v. No. The line does not contain the zero vector. It doesn't pass through the origin.

vi. No. It does contain the origin, and is closed under scalar multiplication, but not under vector addition. The sum $(1, 0) + (0, 1) = (1, 1)$ is not on either axis.

vii. No. Not closed under scalar multiplication. The product $4(1/2, 0) = (2, 0)$ is not in the set.

(b) Use theorem 3.43. (1) The zero function is continuous, so zero is in the set. (2) The sum of two continuous functions is a continuous function, so the set is closed under addition. (3) A constant times a continuous function is still continuous, so the set is closed under scalar multiplication.

(c) Use theorem 3.43. (1) The zero function is a polynomial, so zero is in the set. (2) The sum of two polynomials is a polynomial, so the set is closed under addition. (3) A constant times a polynomial is a polynomial, so the set is closed under scalar multiplication.

(d) i. (1) The zero function is said to be in the set, so zero is in the set. (2) The sum of two polynomials of degree three must be a polynomial of degree 3 or less, so the set is closed under addition. (3) Multiplying a polynomial by a nonzero constant does not change the degree, so the set is closed under scalar multiplication.

ii. The span of the polynomials $\{1, x, x^2, x^3\}$ is $P_3(x)$.

(e) Let V be the set of 3 by 3 upper triangular matrices. Show that V is a subspace of the vector space M_{33} of 3 by 3 matrices in 2 ways:

i. (1) The zero matrix is upper triangular. (2) The sum of two upper triangular matrices is upper triangular, so the set is closed under addition. (3) Multiplying an upper triangular matrix by a constant gives an upper triangular matrix, so the set is closed under scalar multiplication.

ii. The span of $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is the set of upper triangular matrices $\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$.

(f) i. Yes. Zero is a symmetric matrix. The sum of two symmetric matrices is symmetric. A constant times a symmetric matrix is symmetric. Alternatively, the symmetric matrices $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ are spanned by $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and hence are a subspace of $M_2 2$.

ii. No. The zero matrix is not invertible.

iii. Yes. Similar to the previous problem related to 3 by 3 matrices.

(g) i. No. The zero polynomial does not have degree 2.

ii. No. The sum of two degree 2 polynomials may be lower in degree. For example $(x^2) + (x - x^2) = x$. So this set is not closed under addition.

iii. Yes. It is spanned by $\{1, x, x^2\}$

iv. No. Similar to part (b).

7. **Spans:** (problems link)

(a) All but the third have the same row space. The rref of these matrices has the same nonzero rows.

(b) All but the third have the same column space. The rref of the transpose of each matrix has the same nonzero rows.

(c) i. Yes. The rref is the identity.

ii. No. Only 2 independent columns. Column 3 has coordinates $(3, -1)$.

iii. No. You need at least 3 vectors to span $P_2(x)$.

iv. Yes.   There are 3 independent vectors. The 4th column has coordinates $(1, -1, 2)$.

v. No. Only 2 independent vectors. The coordinates of column 3 are $(1, -1)$, column 4 $(1, 2)$.

(d) i. No. You need at least 4 vectors.

ii. Yes.

iii. No. The third polynomial is the sum of the first 2. There are only 3 independent vectors.

(e) i. Yes.

ii. No.

iii. No.

iv. Yes.

(f) **Basis and Dimension:**

Chapter 4

Linear Transformations

The study of linear algebra is really the study of special functions (linear transformations) between vector spaces. The purpose of this chapter is to extend the notion of functions between numbers to functions between vector spaces using matrices. We'll see that vector subspaces related to these functions can help us solve many problems we have been looking at in previous chapters. Thinking about these functions as coming from matrices will help us see the real full power of matrices.

This chapter covers these ideas:

1. Construct graphical displays of linear transformations. Graphically explain how to see eigenvalues, eigenvectors, and determinants from these visualizations, as well as when an inverse transformation exists.

2. Know the vocabulary of functions: domain, range, image, inverse image, kernel, injective (one-to-one), surjective (onto), and bijective.

3. Define linear transformation, and give examples involving both finite and infinite dimensional vector spaces. Show that every linear transformation between finite dimensional vector spaces can be represented by a matrix in relation to the standard basis vectors.

4. Describe the column space, null space, and eigenspaces of a matrix. Show that these spaces are vector spaces.

5. Show that every solution to the linear equation $T(\vec{x}) = \vec{b}$ can be written in the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_p$ is a particular solution and $\vec{x}_h$ is a solution to the homogeneous equation $T(\vec{x}) = \vec{0}$.

6. Explain what it means to compose linear transformations (both algebraically and graphically). Then show how to decompose a matrix as the product of elementary matrices, and use this decomposition (through row reduction) to compute the determinants of large matrices.

4.1   Matrix Transformations

In this section, we'll explore how matrices transform (reshape, move) the plane and space. In the next section we will use this geometric idea to define linear transformations. We'll then show that every matrix represents a linear transformation, and that every linear transformation (between finite dimensional vector spaces) can be represented as a matrix. The concept that matrices represent functions between vector spaces exposes the real power of matrices.

4.1.1 Square Matrices

Let's start by illustrating how a square matrix A transforms a vector $\vec{x}$ by multiplication into another vector $\vec{y} = A\vec{x}$. For a 2 by 2 matrix such as $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$, the columns of A tell us what the standard basis vectors $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, 1)$ get transformed into. The product $A\vec{e}_1 = (2, 0)$ is the first column of A and $A\vec{e}_2 = (1, 3)$ is the second column. Every other vector (x, y) in 2D is transformed in a similar fashion by considering the product

$$A \begin{bmatrix} x \\ y \end{bmatrix} = A \left(x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \qquad \text{write } (x, y) \text{ in terms of basis } \{\vec{e}_1, \vec{e}_2\}$$

$$A \begin{bmatrix} x \\ y \end{bmatrix} = A \begin{bmatrix} x \\ 0 \end{bmatrix} + A \begin{bmatrix} 0 \\ y \end{bmatrix} \qquad \text{distribute across addition}$$

$$= x A \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y A \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{pull scalars out front}$$

$$= x A\vec{e}_1 + y A\vec{e}_2.$$

From the previous computation, once we know where $(1, 0)$ and $(0, 1)$ are transformed, we can transform every other vector by just using this information.

This concept of everything depending on $\vec{e}_1$ and $\vec{e}_2$ generalizes to any basis. If $\{\vec{b}_1, \vec{b}_2\}$ is a basis for $\mathbb{R}^2$, then any vector (x, y) can be written as a linear combination of $\vec{b}_1$ and $\vec{b}_2$. By the computation above, once we know what $A\vec{b}_1$ and $A\vec{b}_2$ are, we know what A does to any vector (x, y). Suppose, for example, that $(x, y) = c_1\vec{b}_1 + c_2\vec{b}_2$. Then $A(x, y) = A(c_1\vec{b}_1 + c_2\vec{b}_2) = c_1 A\vec{b}_1 + c_2 A\vec{b}_2$. Everything depends on what A did to the basis vectors $\{\vec{b}_1, \vec{b}_2\}$.

We'll illustrate transformations from 2 by 2 matrices with 2D plots (see Figure 4.1). We draw the vectors $(1, 0), (0, 1)$ and their transformed images $(2, 0), (1, 3)$. The unit square transforms into a parallelogram. The matrix transforms a triangle into a new triangle, a circle into an ellipse, and a heart into a slanted heart. One key observation is that the matrix transforms lines to lines, which is why we call the transformation a *linear* transformation.

The eigenvectors of the matrix A tell us the directions in which things are transformed radially outwards or inwards ($A\vec{x} = \lambda\vec{x}$ means that the vector $\vec{x}$ is just stretched outward or inward by the factor λ). The determinant tells us how much area (or volume for 3 by 3 matrices) is stretched and if the transformation involved a flip. The determinant is the product of the eigenvalues, so an outward stretch of 2 in one direction followed by an outward stretch of 3 in another direction should result in increasing the area by a factor of $(2)(3) = 6$ (see Figure 4.1). If the determinant of our transformation is zero, then in some direction the transformation squashes things flat, so the area (or volume for 3 by 3 matrices) of the transformed image is zero.

Example 4.1. Figure 4.2 illustrates the transformations given by the four matrices

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

The matrix A has two positive eigenvalues. The transformation pushes everything outwards. In the $(1, 1)$ directions, lengths are multiplied by 4. In the $(-1, 1)$ direction, lengths are multiplied by 2. This results in the area increasing by a factor of 8 (i.e., the new area is 8 times the original area). Notice that the eigenvector directions provide the major and minor axes for the ellipse. This is always true when the matrix is symmetric.

The matrix B has positive and negative eigenvalues. The determinant is negative because the image was flipped over in the transformation. Area is increased by a factor of 4. The matrix is not symmetric, which is why the eigenvector directions are not the major and minor axes of the transformed ellipse.
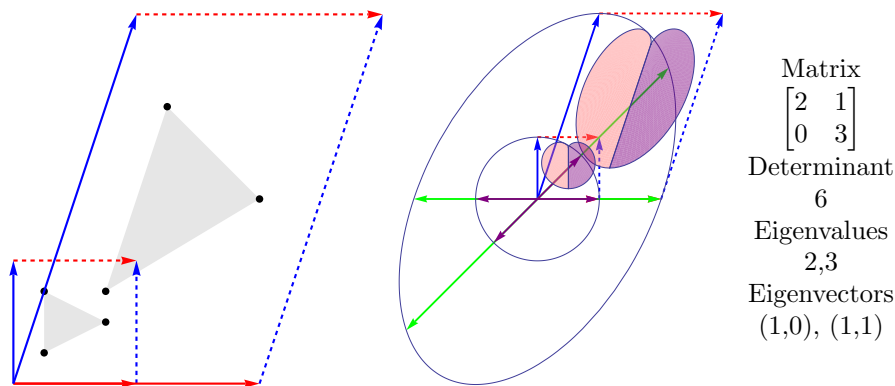
Figure 4.1: Two visualizations of a transformation. Matrix transformations map (see left) lines to lines, triangles to triangles, boxes to parallelograms, and (see right) circles to ellipses. Every other point is transformed in a similar fashion (the heart maps to a slanted heart). The determinant measures the increase in area (the parallelogram has 6 times the area of the box, the ellipse has 6 times the area of the circle). Eigenvectors tell you the directions in which the transformation moves points radially outwards. Eigenvalues tell you how much outward stretching occurs in the eigenvector directions.

We'll look at C and D simultaneously. The difference between these two matrices is that the columns have been interchanged. Interchanging the columns results in a flip, which causes the determinant to switch signs. Area is doubled for each transformation. The eigenvalues of D are complex, which means there is no direction in which points are transformed radially outwards—every point is rotated some amount by this transformation.

Computers are constantly drawing things in 2D and 3D and they use matrices to transform images to get the pictures and text that you see on the screen. For this and other reasons, studying and implementing fast matrix multiplication in computers has been a huge field for decades involving millions of dollars and many companies and people.

3 by 3 Matrices

Let's now look at some visualization of 3D transformations. The ideas are similar to what happens in 2D—we just add an additional dimension. This adds another eigenvector direction, and instead of the determinant telling us about a change in area, it now tells us about a change in volume.

Example 4.2. Figure 4.3 illustrates the transformations resulting from the two matrices $\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 & -1 \\ 0 & -3 & 0 \\ -2 & 0 & 0 \end{bmatrix}$ in two different ways. The left images show how the matrices transforms a 3D object. The right images shows how the matrices transform the unit sphere into an ellipsoid. The images on the right also include the eigenvector directions. You can log on to Sage to view more images and rotate the images to get a better view of what each transformation does.

Both matrices have determinant 6, which means that in both cases the volume of objects is multiplied by 6 through the transformation. The first transformation stretches objects by 1 in the direction $(0, -1, 1)$, by 2 in the direction $(-1, 0, 1)$, and by 3 in the direction $(0, 1, 1)$. The second transformation has two directions in which
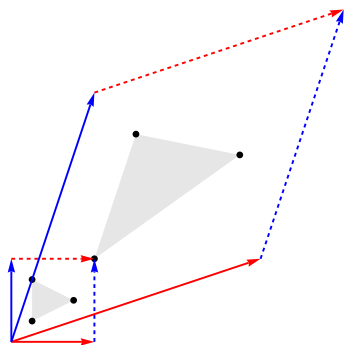
Matrix
$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$
Determinant
8
Eigenvalues
$4, 2$
Eigenvectors
$(1, 1), (-1, 1)$
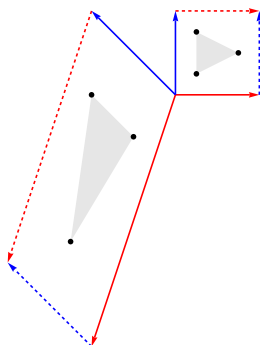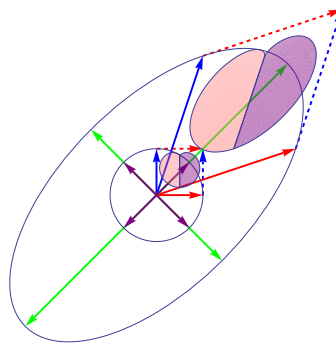
Matrix
$$\begin{bmatrix} -1 & -1 \\ -3 & 1 \end{bmatrix}$$
Determinant
-4
Eigenvalues
$-2, 2$
Eigenvectors
$(1, 1), (-1, 3)$
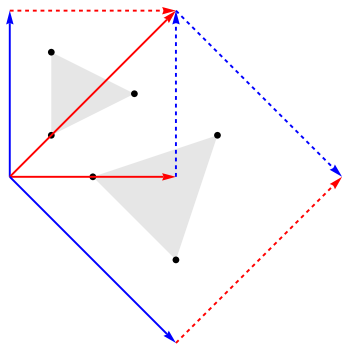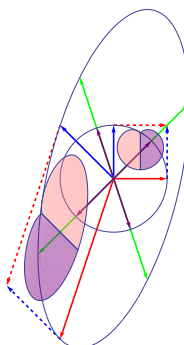
Matrix
$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
Determinant
-2
Eigenvalues
$-\sqrt{2}, \sqrt{2}$
Eigenvectors
$(1 - \sqrt{2}, 1), (1 + \sqrt{2}, 1)$
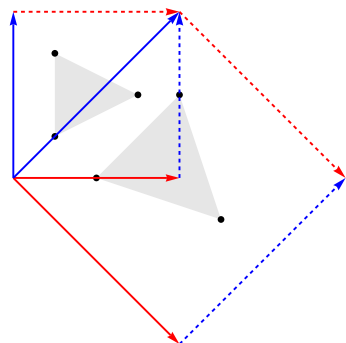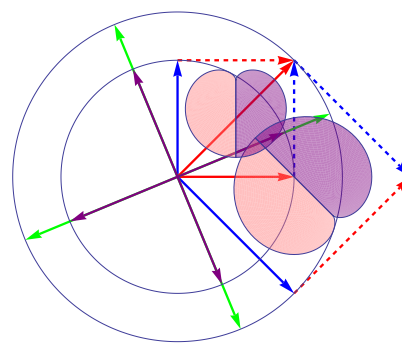
Matrix
$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$
Determinant
2
Eigenvalues
$1 + i, 1 - i$
Eigenvectors
$(-i, 1), (i, 1)$
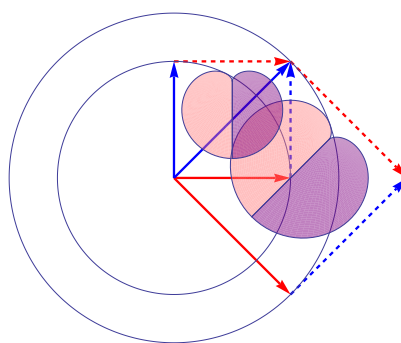
Figure 4.2: Four 2D transformations. See Example 4.1 for details. In the pictures on the right, the purple vectors are eigenvectors of length 1, and the green vectors have been scaled by the eigenvalue.
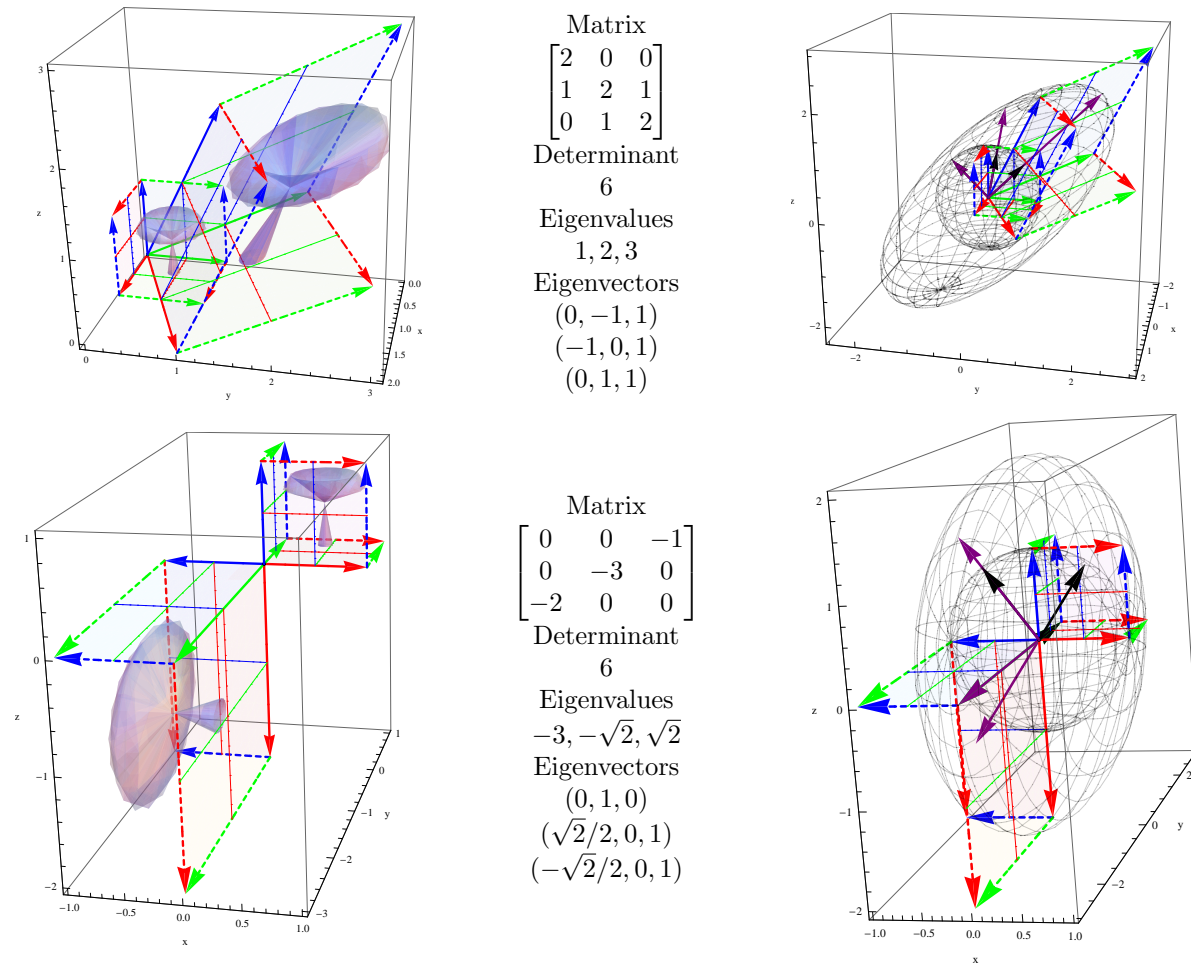
Figure 4.3:   Two 3D transformations. See Example 4.2 for details. In the picture on the right, the black vectors are eigenvectors of length 1, where the purple vectors have been scaled by the eigenvalue.

a reflection occurs (the product of two negatives results in a positive determinant). The amounts of each eigenvalue stretch, together with the eigenvector directions, are listed in the figure.

Singular Matrices—Not Invertible

In all the examples above, we have been looking at square matrices where the determinant is nonzero. When the determinant is zero, the matrix does not have an inverse and we say the matrix is singular. This means that the columns are linearly dependent, which means that when we construct a visualization of the transformation, we should see multiple columns overlapping. The next examples illustrate transformations for singular matrices.

Example 4.3. For the matrix $A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}$, we map $(1, 0)$ to $(1, -1)$ and $(0, 1)$ to $(-2, 2)$. Both vectors lie on the same line. We obtain the two visualizations below.

When the determinant is zero, a 2D transformation takes the entire plane to a line. The column space is the line onto which the transformation maps.
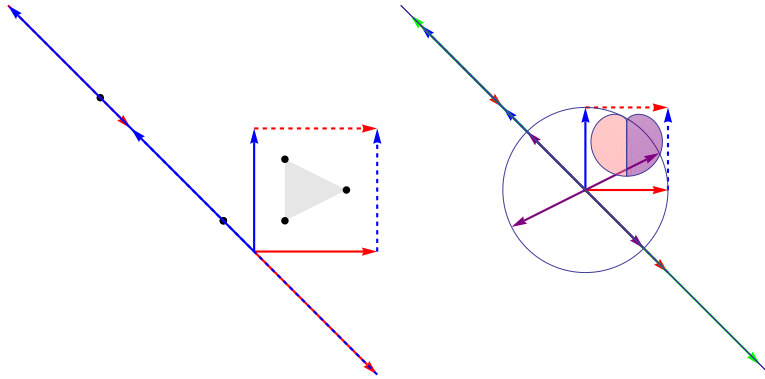
Notice that every vector, in addition to $(1, 0)$ and $(0, 1)$, is mapped onto this same line. The column space of A is precisely the line given by the span of $(-1, 1)$. The column space corresponds directly with the line onto which everything is mapped. Because $\det A = 0$, we know that $\lambda = 0$ is an eigenvalue. The corresponding eigenvector $(2, 1)$ is drawn in the image on the right. This eigenvector is found by solving $A\vec{x} = 0$, which means that the set of solutions is the null space. Every vector on the line containing $(2, 1)$ is transformed to the origin (so in this direction we are squashing an entire line to a point). Because this transformation squashes two dimension onto a line, it is impossible to invert the transformation (hence no inverse matrix exists).

> The null space of A is precisely the set of vectors which get squashed to $\vec{0}$ by A.

Example 4.4. For the matrix $A = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$, we map the vector $(1, 0, 0)$ to $(0, -1, 1)$, the vector $(0, 1, 0)$ to $(1, 0, -1)$, and the vector $(0, 0, 1)$ to $(-1, 1, 0)$. These three images lie on the same plane (since the determinant is zero). We obtain two visualizations below.
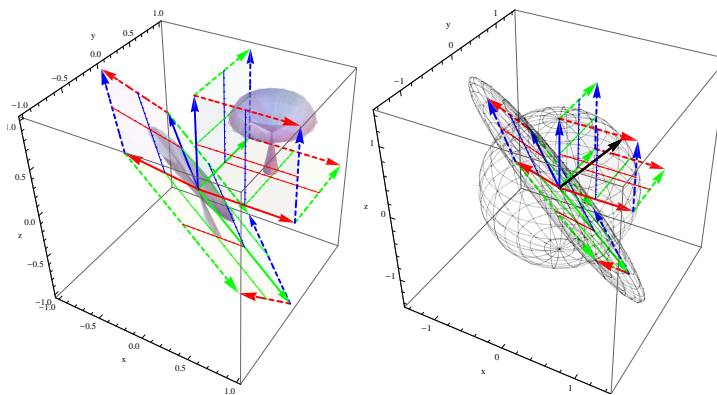
> When the determinant is zero, a 3D transformation takes the all of space to a plane (if the rank is 2) or a line (if the rank is 1). The column space is precisely the subspace of $\mathbb{R}^3$ onto which the transformation squashes things.

Notice that every vector is mapped onto the same plane containing column vectors of A. The column space of A is precisely the plane onto which everything is mapped. The eigenvector corresponding to $\lambda = 0$ is $(1, 1, 1)$, found by solving $A\vec{x} = \vec{0}$, i.e., finding the null space. This eigenvector is drawn in the image on the right. Anything which lies on the line containing $(1, 1, 1)$ is also transformed to the origin, so in this direction we are squashing an entire line to a point. Because this transformation squashes three dimension onto a 2D plane, it is impossible to invert the transformation (hence A has no inverse). Noninvertible square transformations always involve squashing a dimension.

In summary, if the determinant is nonzero, then the transformation maps all of 2D to all of 2D, or all of 3D to all of 3D. If the determinant is zero, then the transformation maps 2D onto a subspace of 2D (a line), or 3D onto a subspace of 3D (a plane or

line). In all cases, the column space of the matrix is precisely the space onto which the transformation maps vectors.

4.1.2 Matrices that are not square

Every example we have looked at in this chapter so far involved either a 2 by 2 or 3 by 3 matrix. In these cases, we were mapping the plane to the plane, or space to space. What kind of transformation does a nonsquare matrix represent, and what use would it be? Any time we see 3D graphics on a computer, or often when we watch telivision, we are viewing a 3D world that is being presented to us on a 2D plane. This is an example of a transformation which takes us from 3D to 2D (often written $\mathbb{R}^3 \to \mathbb{R}^2$). Projecting a 2D image into 3D (written $\mathbb{R}^2 \to \mathbb{R}^3$) requires a transformation from 2D to 3D.

The size of the matrix determines the kind of transformation. For an m by n matrix A, the matrix product $A\vec{x} = \vec{b}$ requires that $\vec{x}$ have n components and $\vec{b}$ have m components. In other words, the transformation associated with the matrix A takes in vectors in $\mathbb{R}^n$ and gives back vectors in $\mathbb{R}^m$. In notation, the matrix A_{mn} gives us a linear transformation $A_{mn} \colon \mathbb{R}^n \to \mathbb{R}^m$. The number of columns of A is the dimension of the inputs, while the number of rows of A is the dimension of the outputs of the transformation.

Let's look at two more examples. Then we'll be ready to generalize the patterns we've observed into a general theory about linear transformations.

Example 4.5. Let's look at the two transformations given by the matrices $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & -1 & -1 \\ 1 & 2 & -1 \end{bmatrix}$. The first transformation requires that we input 2D vectors and returns 3D vectors. To graph this transformation, we notice that multiplying A by the vectors $(1,0)$ and $(0,1)$ returns the two columns of A. So in the xy plane we draw the unit square, and in 3D we graph the parallelogram formed by the two columns. This is illustrated below. Notice that the unit circle becomes an ellipse in the plane spanned by the columns of A. Again we see that all the output vectors are drawn in the column space of A.
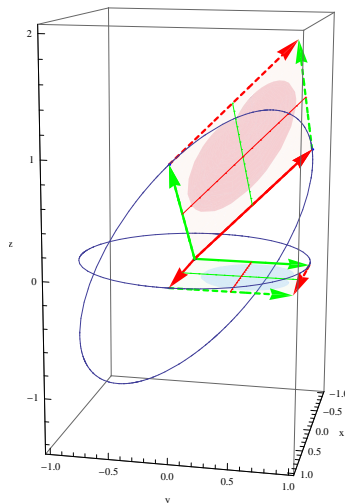
The transformation given by B requires that we input a 3D vector to obtain a 2D vector. We graph the unit cube in 3D and graph the columns of A in the xy plane. In this transformation, we are squashing a dimension out of the problem. This is the kind of transformation used to cast shadows in 3D graphics, as well as to present a 2D image of 3D objects for viewing on a screen.
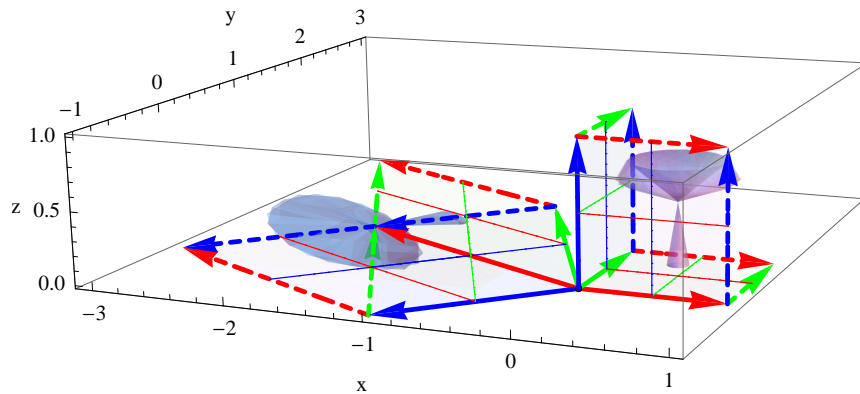
4.2 What is a Linear Transformation?

4.2.1 General Function Notation and Vocabulary

Let's start with a review of function notation from beginning algebra. A function $y = f(x)$ is a rule which assigns to each input x from the domain a value y from the range. Notationally we may write $f : D \to R$ to represent the function as a map from D (the domain) to R (the range). The sets D and R in previous classes have normally been the set of real numbers. Before we proceed to work with functions between vector spaces, we need to build some vocabulary. We now consider functions (maps or transformations) from any set D to any set R.

Definition 4.1.
- Any mapping which assigns to each element of the *domain D* exactly one element of the *range R* is called a *function*. We write $f : D \to R$ to emphasis the domain D and range R. Sometimes the range is called the *codomain* of the function.

- The *image* of an element $x \in D$ is the value $y \in R$ such that $f(x) = y$. We can write $x \mapsto y$ (read "x maps to y") to illustrate the idea that the input x transforms into the output y through this function.

- The *image* of f, written im $f = f(D)$, is the set of $y \in R$ which are the image of some $x \in D$. This means that im f is always a subset (or is equal to) the range of f. In set notation, we write

$$\text{im } f = \{y \in R \mid f(x) = y \text{ for some } x \in D\}.$$

- The *preimage* of $y \in R$, written $f^{-1}(y)$, is the set of all $x \in D$ such that $f(x) = y$. There may be any number of things in D which map to y, so the preimage could be any size.

- We say a function f is *one-to-one*, or *injective*, if every y value in the image has only one x value which maps to it. In other words, for every $y \in \text{im } f$, the preimage $f^{-1}(y)$ has one thing in it. Mathematically, we could also say that if $f(x_1) = f(x_2)$, then $x_1 = x_2$. Another way to say this is that the preimage of any $y \in R$ is always size one or less.

- We say a function f is *onto*, or *surjective*, if im $f = R$. In other words, every $y \in R$ is the image of some $x \in D$. Another way to say this is that the preimage of any $y \in R$ is always size one or more.

- We say a function f is *bijective* if f is both one-to-one and onto (both injective and surjective). In other words, every $y \in R$ has exactly one $x \in D$ in its preimage.

- If a function f is bijective, we define the *inverse* of $f^{-1} \colon R \to D$ to be the function which takes a $y \in R$ and returns the preimage $x \in D$. This means that $f^{-1}(f(x)) = x$, and so f^{-1} inverts f.

- The composition of two functions $f \colon A \to B$ and $g \colon B \to C$ is a new function $g \circ f \colon A \to C$ that takes an element $a \in A$ and returns the element $c \in C$ given by $c = (g \circ f)(a) = g(f(a))$. Notice that in order to compose two maps, the image of f must be a subset of the domain of g.

Example 4.6. Consider the sets $D = \{A, B, C\}$, $R_1 = \{1, 3, 9\}$, and $R_2 = \{1, 3, 7, 9\}$.

1. Let's define a function $f \colon D \to R_1$ by $f(A) = 1$, $f(B) = 9$, $f(C) = 3$. The image of B is 9, so we write $B \mapsto 9$. The image of f is $f(D) = \{1, 3, 9\}$, so f is onto. The preimage of 3 is $f^{-1}(3) = C$. This map $f \colon D \to R_1$ is injective and surjective, so it is bijective and has an inverse f^{-1}. The inverse does $f^{-1}(1) = A$, $f^{-1}(9) = B$, and $f^{-1}(3) = C$.

2. If we define $f \colon D \to R_2$ by the same rules as above, then f is not surjective (onto) because 7 is not the image of any value in D (i.e., the preimage of 7 is empty). You can also see that the function is not surjective since im $f \neq R_2$.

3. If we define $f(A) = 1$, $f(A) = 9$, $f(C) = 3$, then $f \colon D \to R_1$ is not a function. There are two things wrong: $f(B)$ is not defined, and A gets mapped to two different values, $A \mapsto 1, 9$.

4. If we define $f \colon D \to R_1$ by $f(A) = 1$, $f(B) = 3$, $f(C) = 3$, then f is a function. The image of both B and C is 3 so the preimage of 3 is $f^{-1}(3) = \{B, C\}$ and has size 2. This means f is not injective. Another way to think about this is that $f(B) = f(C)$, but $B \neq C$.

5. If we define $f \colon D \to R_1$ by $f(A) = 1$, $f(B) = 3$, $f(C) = 9$, and we define $g \colon R_1 \to R_2$ by $g(1) = 1$, $g(3) = 3$, $g(9) = 7$, then the composite function $g \circ f$ satisfies $g(f(A)) = g(1) = 1$, $g(f(B)) = g(3) = 3$, and $g(f(C)) = g(9) = 7$. The composite is injective, but not surjective (it misses 9 in R_2).

4.2.2 Linear Transformations

One of the main topics in linear algebra is the study of special types of functions, called linear transformations (or linear mappings) where the domain and codomain (range) are both vector spaces. All of the transformations resulting from matrices in the first section are linear transformations. The following definition is a result of noticing the patterns which arose from the examples in the first section.

Definition 4.2: Linear Transformation. We call a function $T \colon V \to W$ between vector spaces V and W linear function or *linear transformation* if both of the following criteria are satisfied:

1. $T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$ for every $\vec{v}_1, \vec{v}_2 \in V$

2. $T(c\vec{v}) = cT(\vec{v})$ for every scalar c and vector $\vec{v} \in V$.

We say that T preserves vector addition and scalar multiplication. We also say that T preserves linear combinations.

A linear transformation is a function between vector spaces in which you can perform vector addition and scalar multiplication either before or after applying the function. If a function is linear, then repeated application of the definition shows that

$$T(c_1 \vec{v}_1 + \cdots + c_n \vec{v}_n) = c_1 T(\vec{v}_1) + \cdots + c_n T(\vec{v}_n).$$

Linear combinations of vectors becomes linear combinations of the images of each vector when using a linear transformation.

If T is linear, then the transform of a linear combination is the linear combination of the transforms:

$$T\left(\sum_{i=1}^{n} c_i \vec{v}_i\right) = \sum_{i=1}^{n} c_i T(\vec{v}_i).$$

In particular, this means that if you know the image of each vector in a basis for V, you know the image of every vector in V, since every vector in V is a linear combination of basis vectors.

Example 4.7. Let's start by illustrating what a linear transformation is not. We'll show that the functions $f(x) = 2x + 1$, $g(x) = x^2$, and $h(x, y) = (x + y, xy)$ are not linear transformations.

1. The function $f(x) = 2x + 1$ from $\mathbb{R}$ to $\mathbb{R}$ (whose graph is a line which does not pass through the origin) is not linear because

$$f(x + y) = 2(x + y) + 1 = 2x + 2y + 1$$

whereas if we transform each point and then sum we obtain

$$f(x) + f(y) = 2x + 1 + 2y + 1 = 2x + 2y + 2.$$

Because $f(x + y) \neq f(x) + f(y)$, this function is not linear.

Notice that if T is a linear transformation, then

$$T(x) = T(0 + x) = T(0) + T(x).$$

Thus $T(0)$ must be zero. Linear transformations always map the zero vector in the domain to the zero vector in the range.

2. The function $g(x) = x^2$ from $\mathbb{R}$ to $\mathbb{R}$ (a parabola) is not linear because

$$g(x + y) = (x + y)^2 = x^2 + 2xy + y^2 = g(x) + g(y) + 2xy \neq g(x) + g(y).$$

Alternatively, we could also look at scalar multiplication to show this function is not linear: $g(2x) = 4x^2$ whereas $2g(x) = 2x^2$, so $g(2x) \neq 2g(x)$. Raising variables to powers other than 1 results in nonlinear maps.

3. The function $h(x, y) = (x + y, xy)$ from $\mathbb{R}^2$ to $\mathbb{R}^2$ is not linear. We can show this by looking at an example rather than using variables. We compute $f(1, 1) = (2, 1)$ and $f(2(1, 1)) = f(2, 2) = (4, 4)$, but $2f(1, 1) = (4, 2)$. We have given an example where $f(2\vec{x}) \neq 2f(\vec{x})$, so the map does not always preserve scalar multiplication. The component xy is the problem here. In general, multiplying variables together creates nonlinear functions.

Now let's look at some examples of linear transformations. In each example, we will formally show that the function is linear by showing that it preserves both vector addition and scalar multiplication. The first three examples are crucial, while the last three represent ideas explored more in future classes where vector spaces of infinite dimension are studied. Many practical applications (like communication, audio and video compression, for example) require knowledge about infinite dimensional vector spaces. In this class we will focus on linear transformations between finite dimensional vector spaces so that we can represent the transformations using matrices.

Example 4.8. The function $T(x) = 4x$ is linear because

1. $T(x + y) = 4(x + y) = 4x + 4y = T(x) + T(y)$, so T preserves vector addition.
2. $T(cx) = 4(cx) = c(4x) = cT(x)$, so T preserves scalar multiplication.

The same reasoning shows that any line through the origin, $y = mx$ for any m, is a linear transformation. In the next two examples, we'll show that any function of the form $\vec{y} = M\vec{x}$ is a linear transformation as well for any matrix M. Linear transformations are really just extensions of the simple line $y = mx$ to higher dimensions. The slope becomes a matrix, while the variables become vectors.

Example 4.9. The function $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (2x + 3y, x - y)$ is linear. To show this, we check that T preserves both vector addition and scalar multiplication.

1. $T((x, y) + (s, t)) = T((x + s, y + t)) = (2(x + s) + 3(y + t), (x + s) - (y + t)) = (2x + 3y, x - y) + (2s + 3t, s - t) = T(x, y) + T(s, t)$, hence the map preserves vector addition.

2. $T(c(x, y)) = T(cx, cy) = (2cx + 3cy, cx - cy) = c(2x + 3y, x - y) = cT(x, y)$, hence the map preserves scalar multiplication.

We can write this function in matrix form as

$$T(x, y) = (2x + 3y, x - y) = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Notice that $T(1, 0) = (2, 1)$ is the first column and $T(0, 1) = (3, -1)$ is the second column.

> The columns of the matrix are the images of the standard basis vectors (1,0) and (0,1).

Example 4.10. Now let's work with a vector space other than $\mathbb{R}^n$. Let $V = P_2(x)$ be the set of polynomials of degree 2 or less, together with the zero polynomial. The derivative function $D \colon P_2(x) \to P_2(x)$ defined by $D(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x$ is a linear transformation.

1. The derivative preserves vector addition because $(f + g)' = f' + g'$ by the sum rule for derivatives.

2. The derivative preserves scalar multiplication becuse $(cf)' = c(f)'$ by the constant multiple rule for derivatives.

The vectors $\{1, x, x^2\}$ are the standard basis vectors for $P_2(x)$. Relative to this standard basis, the coordinates of $a_0 + a_1 x + a_2 x^2 = a_0(1, 0, 0) + a_1(0, 1, 0) + a_2(0, 0, 1)$ are just (a_0, a_1, a_2). We can then rewrite the derivative transformation in coordinate form as $D(a, b, c) = (b, 2c, 0)$. In matrix form, we write

$$D(a_0, a_1, a_2) = (a_1, 2a_2, 0) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

Notice that $D(1, 0, 0) = (0, 0, 0)$ is the first column, $D(0, 1, 0) = (1, 0, 0)$ is the second column, and $D(0, 0, 1) = (0, 2, 0)$ is the third column. In general, the columns of our matrix will be the coordinates of the images of the standard basis vectors.

> The columns of our matrix are the coordinates of the images of our basis vectors

Example 4.11. Let $V = P(x)$ be the set of polynomials, an infinite dimensional vector space with basis $\{1, x, x^2, x^3, \ldots\}$. The derivative transformation $D \colon P(x) \to P(x)$ is a linear transformation because of the sum rule and constant multiple rule for derivatives (just as in the last example). For those of you who have had calculus 2, Taylor polynomials and Taylor series are all related to this vector space and transformations on it.

Example 4.12. Let V be the set of functions that are infinitely differentiable on the entire real number line (so e^x, $\sin x$, $\cos x$, and x^2 are in V, but $\ln x$, $\sqrt{x}$, and $\tan x$ are not in V). The derivative transformation $D \colon V \to V$ defined by $D(f) = f'$ is a linear transformation on this space (by the same arguments as the previous example). Notice that you input a function f and get out a function f', so this map takes vectors in V to vectors in V. This space is infinite dimensional, and a basis for this space is so large that we say it is uncountable[i].

Let $T \colon V \to \mathbb{R}$ be the integral transformation, defined by $T(f) = \int_0^1 f(x) dx$. This transformation takes a function f and returns a real number by integrating over $[0, 1]$. To show this function is linear, we check

1. $\int_0^1 f(x) + g(x) dx = \int_0^1 f(x) dx + \int_0^1 g(x) dx$, so $T(f + g) = T(f) + T(g)$ (vector addition is preserved).

---

[i]The difference between countably infinite and uncountably infinite is studied in later math classes

2. $\int_0^1 cf(x)dx = c\int_0^1 f(x)dx$, so $T(cf) = cT(f)$ (scalar multipilcation is preserved).

The derivative and integral are both examples of linear transformations that you have studied before (though you probably didn't use the language of linear transformations on vector spaces!). In real analysis, these ideas are explored in greater depth and become foundational tools for understanding finance, economics, and more.

Example 4.13. Let V be the set of infinitely differentiable functions defined on all of $\mathbb{R}$. For any $y(x)$ in V, let

$$L(y) = y'' - 2xy' + 3x^2y$$

be the differential operator which takes various derivatives of y and combines them together by multiplying by functions of x. This is a linear transformation $L: V \to V$ because

1. $L(y_1 + y_2) = (y_1 + y_2)'' - 2x(y_1 + y_2)' + 3x^2(y_1 + y_2) = (y_1'' - 2xy_1' + 3x^2y_1) + (y_2'' - 2xy_2' + 3x^2y_2) = L(y_1) + L(y_2)$, so vector addition is preserved.

2. $T(cy) = (cy)'' - 2x(cy)' + 3x^2(cy) = c(y'' - 2xy' + 3x^2y) = cT(y)$, so scalar multiplication is preserved.

We use transformations like L to find solutions to differential equations and to show that the set of solutions is a vector space. These ideas are studied more in differential equations, engineering, and physics, and become the foundational tool for understanding mechanical systems, electrical circuits, radio waves, signal processing, and more.

4.2.3 Standard Matrix Representation

Every matrix A we have studied this semester represents a linear transformation $f(\vec{v}) = A\vec{v}$. We can see that the function is linear since

1. $f(\vec{v}_1 + \vec{v}_2) = A(\vec{v}_1 + \vec{v}_2) = A\vec{v}_1 + A\vec{v}_2 = f(\vec{v}_1) + f(\vec{v}_2)$, so f preserves vector addition.

2. $f(c\vec{v}) = A(c\vec{v}) = c(A\vec{v}) = cf(\vec{v})$, so f preserves scalar multiplication.

Conversely, every linear transformation (between finite dimensional vector spaces) can be expressed in matrix form $T(\vec{x}) = A\vec{x}$ for some matrix A. Examples 4.9 and 4.10 illustrate how to find the matrix A relative to the standard basis. To find this matrix, we start by computing the image of each standard basis vector, i.e.

$$T(1, 0, \ldots, 0), \, T(0, 1, \ldots, 0), \, \ldots, \, T(0, 0, \ldots, 1).$$

We then place these vectors in the columns of a matrix. This matrix A is called the standard matrix representation of the linear transformation. We use the word "standard" because are using the standard basis vectors to give the coordinates of each vector. The next two units explore using different basis vectors to describe linear transformations.

Example 4.14. Let's find the standard matrix representation of the linear transformation

$$T(x, y, z, w) = (2x + 3y - 4z, \, x + y - w, \, z + 2x - 5w).$$

We start by computing images of the four standard basis vectors

$$T(1, 0, 0, 0) = (2, 1, 2), \qquad T(0, 1, 0, 0) = (3, 1, 0),$$
$$T(0, 0, 1, 0) = (-4, 0, 1), \qquad T(0, 0, 0, 1) = (0, -1, -5).$$

We now place these vectors into the columns of a matrix. This matrix gives the standard matrix representation of T. We could write T as

$$T(x, y, z, w) = A\vec{x} = \begin{bmatrix} 2 & 3 & -4 & 0 \\ 1 & 1 & 0 & -1 \\ 2 & 0 & 1 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}.$$

The matrix A is called the standard matrix representation of T. Notice that we input 4D vectors (A has 4 columns) and get out 3D vectors (A has 3 rows).

> The number of rows is the dimension of the range. The number of columns is the dimension of the domain.

Sometimes a linear transformation is described in words. Examples include things like "rotate everything 30 degrees" or "find the shadow of this object if the sun is located at ..." In such a case, you can create the standard matrix representation of the transformation by asking where each standard basis vector is sent. Here is an example.

Example 4.15. Let's rotate an object in the plane counterclockwise by 90 degrees. How do we obtain the standard matrix representation of this transformation? We find what happens to the standard basis vectors. Let $T(x, y)$ be the linear transformation which rotates counterclockwise by 90 degrees. Then $T(1, 0) = (0, 1)$ and $T(0, 1) = (-1, 0)$. This means the standard matrix representation of T is

$$\begin{bmatrix} \cos \pi/2 & \cos \pi \\ \sin \pi/2 & \sin \pi \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \text{(a rotation of 90 degrees)}.$$

If instead we want to rotate counterclockwise by 30 degrees ($\pi/6$ radians) then we compute $T(1, 0) = (\cos \pi/6, \sin \pi/6) = (\sqrt{3}/2, 1/2)$ (the values on the unit circle at $\pi/6$) and $T(0, 1) = (\cos 2\pi/3, \sin 2\pi/3) = (-1/2, \sqrt{3}/2)$ (the values on the unit circle at $\pi/2 + \pi/6 = 2\pi/3$). This means the standard matrix representation of T is

$$\begin{bmatrix} \cos \pi/6 & \cos(\pi/2 + \pi/6) \\ \sin \pi/6 & \sin(\pi/2 + \pi/6) \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \qquad \text{(a rotation of 30 degrees)}.$$
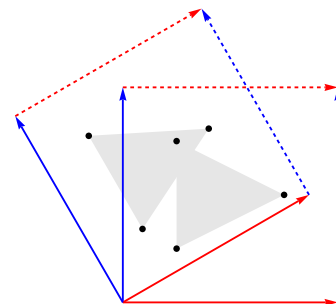
A 30 degree rotation.

Let's generalize the pattern we've seen above to rotations through any angle. We can write the image of the second basis vector above in terms of trig function at θ, instead of at $\pi/2 + \theta$, if we recall that $\cos(\pi/2 + \theta) = -\sin(\theta)$ and $\sin(\pi/2 + \theta) = \cos(\theta)$ (these identities just say that shifting cosine left $\pi/2$ gives negative sine, and shifting sine left $\pi/2$ gives cosine). We can now write the standard matrix representation of a rotation through θ degrees as

$$\begin{bmatrix} \cos \theta & \cos(\pi/2 + \theta) \\ \sin \theta & \sin(\pi/2 + \theta) \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \qquad \text{(a rotation of } \theta \text{ radians)}.$$

Sometimes we are able to extract from a problem information about how a transformation changes vectors other than the standard basis vectors. When this occurs, provided we know what the transformation does to a basis of the domain, we can still find the standard representation. It just requires a little more work. Let's look at an example.

Example 4.16. Suppose we know that $T(x, y)$ is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ and we know that $T(1, 1) = (3, 0, 5)$ and $T(2, 0) = (1, -2, 0)$. We don't know A yet, but we do know that

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix} \text{ and } A \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \qquad \text{so} \qquad A \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 0 & -2 \\ 5 & 0 \end{bmatrix}.$$

The vectors $(1, 1)$ and $(2, 0)$ are linearly independent, so they form a basis for $\mathbb{R}^2$. The matrix formed by placing these vectors in columns, $\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$, is an invertible matrix

since the columns are independent. Since it is invertible, we can multiply both sides of our equation on the right by this inverse to obtain

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

We found the standard matrix representation by using the inverse of the matrix of basis vectors that we had. We can check that our matrix A is correct by computing the matrix product

$$\begin{bmatrix} 1/2 & 5/2 \\ -1 & 1 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 0 & -2 \\ 5 & 0 \end{bmatrix}$$

to show that $T(1,1) = (3,0,5)$ and $T(2,0) = (1,-2,0)$, as required. We also know now that $T(1,0) = (1/2,-1,0)$ and $T(0,1) = (5/2,1,5)$, so we know the images of the standard basis vectors.

In general, when you know what a transformation does to any basis of the domain, you can use this to find the standard matrix by multiplying on the right by an inverse. Schaum's outlines writes out the equations and then solves them (which is equivalent to using an inverse). In this book, we'll focus on the fact that the inverse matrix is a key tool needed to understand a linear transformation.

4.3 Subspaces From Linear Transformations

In the first section where we graphed matrix transformations, we noticed that the column space of the matrix was related to the possible outputs of the transformation. We also saw that the null space of the matrix told us precisely which vectors in the domain were mapped to the origin. These two spaces associated with matrices are important subspaces related to general linear transformations. Because a linear transformation could have a domain or range which is infinite dimensional, we need a new word to talk about these subspaces of transformations, even when no matrix exists to represent the transformation. The new words are image and kernel. Here are the formal definitions.

Definition 4.3. Let $T\colon V \to W$ be a linear transformation.

1. The *image* of T (written im T) is the set of all values in W which are the image of some $\vec{v}$ in V. Symbolically we write

$$\text{im } T = \{\vec{y} \in W \mid T(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in V\}.$$

2. The *kernel* of T (written ker T) is the set of all values in V whose image is $\vec{0}$ in W. Symbolically we write

$$\text{ker } T = \{\vec{x} \in V \mid T(\vec{x}) = \vec{0}\}.$$

If a linear transformation can be represented by a matrix (i.e., when the domain and range have finite dimension), then from the definitions we can see that the image and kernel of the transformation are the same as the column space and null space of the matrix. Let's formalize this pattern into a theorem.

Theorem 4.1. If $T\colon V \to W$ is a linear transformation between finite dimensional vector spaces and the standard matrix representation of T is A,

1. the column space of A is the image of T and

2. the null space of A is the kernel of T.

We already know that the column space and null space of a matrix are vector subspaces. This is also true of the image and kernel of the transformation.

Theorem 4.2. Let $T\colon V \to W$ be a linear transformation.

1. The image of T is a vector subspace of the range (codomain) W.

2. The kernel of T is a vector subspace of domain V.

Notice that the image is a subspace of the range and the kernel is a subspace of the domain. The image is the set of outputs, and the kernel is the set of inputs that map to zero.

Example 4.17. Let's show that the image of T is a vector subspace of W. In the homework, we ask you to show that the kernel is a vector subspace of V. Remember from Theorem 3.43 that we need to show three things: (1) the zero vector in W is in the image, (2) the image is closed under addition, and (3) the image is closed under scalar multiplication.

1. We know that linear transformations map the zero vector in V to the zero vector in W. Thus the image of the zero vector $\vec{0}_V$ in V is $T(\vec{0}_V) = \vec{0}_W$, the zero vector in W. Hence $\vec{0}_W$ is in the image.

2. We need to show that if $\vec{y}_1$ and $\vec{y}_2$ are in the image, then so is their sum. Well, since $\vec{y}_1$ belongs to W, we know that $\vec{y}_1 = T(\vec{x}_1)$ for some $\vec{x}_1 \in V$. Similarly $\vec{y}_2 = T(\vec{x}_2)$ for some $\vec{x}_2 \in V$. Since V is a vector space, we know that $\vec{x}_1 + \vec{x}_2$ is also in V. We can now compute (since T is linear and preserves addition)

$$T(\vec{x}_1 + \vec{x}_2) = T(\vec{x}_1) + T(\vec{x}_2) = \vec{y}_1 + \vec{y}_2$$

Linear transformations preserve addition. This is why the image is closed under addition.

which means that the sum $\vec{y}_1 + \vec{y}_2$ is the image of $\vec{x}_1 + \vec{x}_2$. This shows that the image is closed under vector addition.

3. The only thing that remains is to show that if $\vec{y}$ is in W, then so is any scalar multiple of $\vec{y}$. Since $\vec{y} \in W$, we know there is some $\vec{x} \in V$ such that $T(\vec{x}) = \vec{y}$. Since V is a vector space, we know that $c\vec{x}$ is also in V for any c. Now we can compute (since T is linear and preserves scalar multiplication)

$$T(c\vec{x}) = cT(\vec{x}) = c\vec{y}$$

Linear transformations preserve scalar multiplication. This is why the image is closed under scalar multiplication.

which means that $c\vec{y}$ is the image of $c\vec{x}$. This shows that the image is closed under scalar multiplication.

Your job in the homework will be to show that the kernel of a linear transformation is a vector subspace of the domain V.

Now that we know the image and kernel of a transformation are vector subspaces, we can look for spanning sets, bases, and find the dimensions of these spaces. We used the word "rank" for the dimension of the column space, but have not yet given a name to the dimension of the null space. Here are some more definitions.

Definition 4.4. Let $T\colon V \to W$ be a linear transformation. The dimension of the image of T is called the *rank* of T (just as the rank is the dimension of the column space). The dimension of the kernel of T is called the *nullity* of T. Similarly, the *nullity* of a matrix is the dimension of the null space of the matrix.

Now that we have some words to describe the dimensions of the image and kernel, we can state one of the key theorems which relates these two subspaces. A basis for the column space of a matrix A is simply the pivot columns of A. We find a basis for the null space by solving for all the variables in terms of the free variables, which means that each free variable contributes a vector to the basis of the null space. Since the free variables correspond to nonpivot columns, and the column space corresponds to pivot columns, we see that the sum of the dimensions of the column space and null space must always equal the number of columns.

Theorem 4.3 (The Rank-Nullity Theorem)**.** Let $T\colon V \to W$ be a linear transformation between finite dimensional vector spaces, whose standard matrix representation is A. Then

1. The rank of T is the number of pivot columns of A.

2. The nullity of T is the number of nonpivot columns of A.

3. The rank of T plus the nullity of T is always equal to the dimension of V.

Notice that the third item does not refer at all to the matrix A. The third fact only relies on V having finite dimension. Even though our reasoning above only proves it is true when W also has finite dimension, it is still true even if W has infinite dimension (in which case T does not have a standard matrix representation).

Example 4.18. Consider the linear transformation $T\colon \mathbb{R}^5 \to \mathbb{R}^6$ with standard matrix

$$
A = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & -3 & -7 & -1 & 13 \\ 0 & 1 & 1 & 1 & -1 \\ 4 & 0 & -8 & 4 & 20 \\ 3 & 0 & -6 & 3 & 15 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & -2 & 0 & 4 \\ 0 & 1 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.
$$

The column space of A is the span of the columns of A. The image T is all vectors of the form $A\vec{x}$, which is also the span of the columns of A. This means that the column space of A and the image of T are the same vector subspace of the range of T, $\mathbb{R}^6$. A basis for this vector subspace is the set of pivot columns

$$\{(1, 2, 0, 4, 3, 0),\ (2, -3, 1, 0, 0, 0),\ (1, -1, 1, 4, 3, 0)\},$$

so the column space has dimension 3. The ranks of both A and T are 3.

To find the null space of A, and hence the kernel of T, we solve the equation $A\vec{x} = 0$. The 3rd and 5th columns correspond to the free variables. We can write the solution as

$$
\begin{array}{ll}
x_1 - 2x_3 + 4x_5 = 0 \quad & x_1 = 2x_3 - 4x_5 \\
x_2 + 1x_3 - 2x_5 = 0 & x_2 = -1x_3 + 2x_5 \\
\quad x_3 = x_3 \quad \Rightarrow & \quad x_3 = x_3 \\
x_4 + x_5 = 0 & x_4 = -x_5 \\
\quad x_5 = x_5 & \quad x_5 = x_5
\end{array}
\Rightarrow
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}.
$$

We see that the null space is the span of the vectors

$$\{(2, -1, 1, 0, 0),\ (-4, 2, 0, -1, 1)\}.$$

The kernel of T is a vector subspace of the domain of T, $\mathbb{R}^5$. The numbers in these vectors are the negatives of the numbers in the nonpivot columns of rref, with a 1 or 0 placed in the spots representing the free variables. The reason we negate all the numbers is that we have to subtract these numbers from both sides of the first equation to solve for each variable in terms of the the free variables. The nullity of T is 2 (the number of free variables). The rank plus the nullity equals $3 + 2 = 5$, the dimension of the domain (or number of columns).

4.4 Applications of the kernel

4.4.1 Solutions to Systems and the Kernel

The kernel arises from solving the equation $T(\vec{x}) = \vec{0}$. Can we use the kernel to help us solve the equation $T(\vec{x}) = \vec{b}$? The answer is amazingly yes. The key lies in the fact that if both $\vec{x}_1$ and $\vec{x}_2$ are solutions to $T(\vec{x}) = \vec{b}$, then

$$T(\vec{x}_1 - \vec{x}_2) = T(\vec{x}_1) - T(\vec{x}_2) = \vec{b} - \vec{b} = \vec{0},$$

so $x_1 - x_2$ is in the kernel of T (did you see how we used the properties from T being a linear transformation?). Solving for $\vec{x}_1$ gives us the fact that $\vec{x}_1 = \vec{x}_2 + \vec{x}_h$ where $\vec{x}_h$ is some vector in the kernel of T. On the other hand, if $\vec{x}_h$ is any vector in the kernel, then

$$T(\vec{x}_1 + \vec{x}_h) = T(\vec{x}_1) + T(\vec{x}_h) = \vec{b} + \vec{0} = \vec{b},$$

so adding any vector in the kernel of T to $\vec{x}_1$ gives anothers solution to $T(\vec{x}) = \vec{b}$. Let's formalize this into a theorem.

Theorem 4.4. Let T be a linear transformation. Consider the nonhomogeneous equation $T(\vec{x}) = \vec{b}$. If $\vec{x}_p$ is one particular solution to the equation $T(\vec{x}) = \vec{b}$ (i.e., $T(\vec{x}_p) = \vec{b}$), then every solution can be written in the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_h$ is a solution to the corresponding homogeneous equation $T(\vec{x}) = \vec{0}$. In other words, every solution to the nonhomogeneous equation is found by first finding a particular solution $\vec{x}_p$ and then adding $\vec{x}_p$ to every vector in the kernel.

This theorem is a key tool used to solve differential equations. Once you know one solution, you can find every other solution by just adding the solution to every vector in the entire kernel. This makes the kernel an extremely useful subspace to understand.

Example 4.19. Suppose that a system of equations gives rise to the augmented rref matrix on the left below. Then the solution is on the right.

$$\begin{bmatrix} 0 & 1 & 0 & 2 & 0 & | & 0 \\ 0 & 0 & 1 & 3 & 0 & | & 1 \\ 0 & 0 & 0 & 0 & 1 & | & 4 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 4 \end{bmatrix} + x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ -2 \\ -3 \\ 1 \\ 0 \end{bmatrix}$$

A particular solution to this system is $\vec{x}_p = (0, 0, 1, 0, 4)$ (obtained by letting the free variables x_1 and x_4 both be zero). The null space is the span of the vectors $(1, 0, 0, 0, 0)$ and $(0, -2, -3, 1, 0)$. Notice that every solution to the system can be written in the form $\vec{x}_p + \vec{x}_h$, where $\vec{x}_h$ is a vector in the null space.

4.4.2 Eigenspaces are Null Spaces

We have already noticed in the first section that the eigenvectors of a matrix tell us the directions in which a transformation pushes points radially outwards or inwards. These directions provide us a line (vector subspace) in which all transforming is done radially outwards (or inwards if the eigenvalues are less than one in magnitude). Recall that corresponding to each eigenvalue is an infinite collection of eigenvectors. By definition, the zero vector is not an eigenvector, hence the set of eigenvectors cannot be a vector subspace. However, if we include the zero vector, we obtain a vector space. We now make some formal definitions.

Definition 4.5: Eigenspaces. If λ is an eigenvalue of a matrix A, then the set of vectors $E_\lambda = \{\vec{x} \mid A\vec{x} = \lambda\vec{x}\}$ (the eigenvectors together with zero) is called the eigenspace of A corresponding to λ.

We say that λ is an eigenvalue of a linear transformation $T \colon V \to V$ (notice the domain and range are the same) if $T(\vec{x}) = \lambda\vec{x}$ for some nonzero $\vec{x}$, called an eigenvector of T corresponding to λ. If λ is an eigenvalue of T, then the set of vectors $E_\lambda = \{\vec{x} \mid T(\vec{x}) = \lambda\vec{x}\}$ (the eigenvectors together with zero) is called the eigenspace of T corresponding to λ.

We find eigenvectors of a matrix by solving $(A - \lambda I)\vec{x} = \vec{0}$ which means that the eigenspace E_λ is precisely the null space of $(A - \lambda I)$. Because the eigenspace is a null space, we know that it is a vector space.

4.4.3 Squashing in transformations

The dimension of the kernel of a linear transformation can help us understand how much squashing goes on in the transformation. If the kernel is trivial (only the zero vector), then no vectors will be mapped to zero except for zero itself. This means that the transformation will map lines to line, planes to planes, 3D volumes to 3D volumes, etc. In this case, we say the linear transformation is nonsingular. If the domain and range have the same dimension, then the transformation will have an inverse.

However, if the transformation has a nontrivial kernel (i.e., not just the zero vector), then the dimension is squashed in some direction. We call the transformation singular, and it cannot have an inverse. For example, an entire line could be mapped to a point, or an entire plane could be mapped to a line.

4.4.4 Casting Shadows using the Null Space

The null space of a matrix tells us the directions which get mapped to zero under the linear transformation corresponding to the matrix. We know how to compute the null space if we are given a matrix. What if we are given the null space and asked for a matrix with that null space? Can we find a matrix with a specific null space? Let's explore this issue by figuring out a linear transformation that tells how shadows are cast on the ground.

When the sun shines down on you during the afternoon, it casts a shadow on the ground. The function which takes a point (a, b, c) and gives back the (x, y) coordinate of the shadow is a linear transformation (notice that straight things cast straight shadows!). Suppose you are positioned at the origin standing with your head pointing up along the z axis. Imagine putting a stick at your feet pointed directly at the the sun (so it casts no shadow). This stick gives a vector that points directly at the sun. Any object along that direction (between your feet and the sun) will cast a shadow at your feet. Since anything along the line between your feet at the origin and the sun casts a shadow at your feet, that entire line is the null space of our linear transformation (i.e., all of those vectors get mapped to $(0, 0)$). Can we find a matrix for this shadowy linear transformation? That will let us calculate the shadow of any object.

Since points already on the ground do not cast a shadow, we know that the vector $(x, y, 0)$ gets mapped to (x, y). In particular, $(1, 0, 0) \mapsto (1, 0)$ and $(0, 1, 0) \mapsto (0, 1)$. As discussed above, we also know that a vector pointing directly at the sun gets mapped to $(0, 0)$.

Example 4.20. Suppose that the sun is positioned on a line containing the vector $(2, 1, 4)$ (i.e., the stick at your feet pointing at the sun is the vector $\vec{v} = (2, 1, 4)$). We want transformation that has this vector as a basis for the kernel. Because we're mapping from 3D onto the xy plane, we need a 2 by 3 matrix. Divide every component in $\vec{v}$ by 4 to get a 1 in the z component. The vector $(1/2, 1/4, 1)$ looks a lot like the vectors we obtain by having a free variable z. The null space is all linear combinations of this vector, which we can express as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \begin{bmatrix} 1/2 \\ 1/4 \\ 1 \end{bmatrix}.$$

We now have two equations: $x = 1/2z$ and $y = 1/4z$, or $x - \frac{1}{2}z = 0$ and $y - \frac{1}{4}z = 0$. Now we just write down the matrix associated with this system of linear equations:

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

Notice that since made sure that the left two columns formed an identity matrix, we get the other mappings that we wanted: $(1, 0, 0) \mapsto (1, 0)$ and $(0, 1, 0) \mapsto (0, 1)$. The image in the margin illustrates this projection.
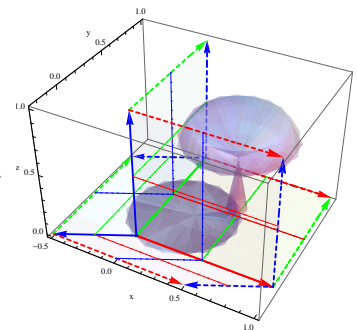
Knowing the null space is sufficient to find a transformation to cast shadows onto the xy plane.

4.5 Composition of Linear Transformations

The composition of two functions f and g is a new function $(g \circ f)(x)$ defined by $(g \circ f)(x) = g(f(x))$. The idea is that you input an x value into f, and then input $f(x)$ into g. The same principle applies to linear transformations. If $S \colon U \to V$ and $T \colon V \to W$ are two linear transformations, then the composition $T \circ S \colon U \to W$ is defined by $T(S(\vec{u}))$ (where $S(\vec{u})$ is a vector in V which serves as an input to T). Let's look at an application of function composition.

Example 4.21. If we want to rotate an object 90 degrees counterclockwise, and then mirror-image the object about the y-axis, we can achieve this by composing two linear transformations. The linear transformation S which rotates and object in the plane 90 degrees has standard matrix representation $S_A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, since $S(1,0) = (0,1)$ and $S(0,1) = (-1,0)$. The linear transformation T which flips objects about the y axis has standard matrix representation $T_A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, since $T(1,0) = (1,0)$ and $T(0,1) = (0,-1)$. The composition $T(S(x,y))$ will first perform the rotation $S(x,y)$ and then reflect the result about the y axis. We compute $S(T(x,y))$ by computing

$$(S \circ T)(x,y) = S_A \left(T_A \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

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

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

Notice that function composition is precisely matrix multiplication.

> When multiple linear transformations are involved, we often use the notation S_A or T_A to represent the standard matrix representations.

The composition of two linear transformations has as its standard matrix representation the product of the standard matrix representations of each transformation. In other words, if $S \colon U \to V$ and $T \colon V \to W$ are both linear transformations with standard matrices S_A and T_A, then $T \circ S \colon U \to V$ (sometimes just written TS) has standard matrix $(TS)_A = T_A S_A$. Notice that in order for this composition to make sense, the range of S and domain of T must match, which means that the number of rows of S_A must equal the number of columns of T_A (in other words you have to be able to compute the matrix product $T_A S_A$). One of the primary purposes for matrix multiplication is to compose linear transformations.

When the domain and range of a linear transformation have the same dimension, the standard matrix is a square matrix. In such cases, we can compute the determinant of a transformation, find the eigenvalues and eigenvectors of a transformation, and ask if the transformation is invertible. If the transformation T is invertible with standard matrix A, then the standard matrix of the inverse is A^{-1}. Symbolically we write $T^{-1}(\vec{x}) = A^{-1}\vec{x}$.

4.6 Elementary Matrices and Determinants

To quickly understand a complicated linear transformation or complicated matrix, it is often easiest to break the transformation down into a series of very simple transformations, or break the complicated matrix down into a product of much simpler matrices. Understanding each part because easy, and then combining what you know about the parts tells you things about the complicated transformation or structure. In this section, we will examine one way to break a matrix up into a product of much simpler matrices that leads us to a very fast method used in industry to calculate the determinant, solve systems of equations, and much more.

The goal here is to factor the matrix A into the product of so-called "elementary matrices," which are simple matrices directly related to row reduction. Whenever we write down a product of matrices, we are also writing down a composition of linear transformations, so we are also exploring how to make a take apart a complicated linear transformation into a series of much simpler ones.

Definition 4.6. An *elementary matrix* is a matrix obtained by performing one of the following row operations to the identity matrix.

1. Switch any two rows.

2. Multiply a row by a nonzero constant.

3. Add a multiple of one row to another.

The row operations above are precisely the operations we use to perform Gaussian elimination. You can also undo each row operation above:

1. The inverse of switching two rows is to switch the same rows again.

2. The inverse of multiply a row by k is to multiply the same row by $1/k$.

3. The inverse of adding k times row i to row j is to subtract k times row i from row j.

Three examples of elementary matrices follow, along with their inverses.

$$\text{Switch } R_1 \text{ and } R_2 \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{Multiply } R_1 \text{ by } 2 \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{Add } 4R_3 \text{ to } R_1 \quad \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Determinants of elementary matrices are simple as well.

1. Switching two rows of the identity matrix results in a matrix with determinant -1. You are just swapping two vectors, which means that you are changing the orientation, not stretching space.

2. Multiplying a row of the identity matrix by a nonzero constant results in a determinant equal to the constant. Since the identity matrix is a triangular matrix, the determinant is just the product of the diagonal entries. Visually this can be seen as stretching one dimension of space by a constant (e.g., you double the widths of everything, so areas and volumes double).

3. Adding a multiple of one row of the identity matrix to another row results in a determinant of 1. Again, the elementary matrix is a triangular matrix, but this time the diagonal is still all ones, so the determinant is 1. Visually, the transformation is a shear transformation in which you are pushing a part of an object in a direction without changing the area or volume.

With elementary matrices, finding inverses and determinants is easy. We now state a key theorem which allows us to connect elementary matrices to Gaussian elimination.

Theorem 4.5. Multiplying the matrix A on the left by an elementary matrix E is the same as performing on A the row operation used to obtain E.

The reason this works is because multiplying a matrix A on the left by B is equivalent to computing linear combinations of the rows of A using the scalars in the rows of B. (Before proceeding, try making your own example to illustrate this theorem. This is the key way to learn to read technical documents.)

Example 4.22. Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 2 & -4 & 0 \end{bmatrix}$. If we multiply the matrix

A on the left by $\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, we obtain $\begin{bmatrix} 9 & -14 & 3 \\ 0 & -1 & 3 \\ 2 & -4 & 0 \end{bmatrix}$. This is the same as adding 4

times row 3 to row 1 of A.

4.6.1 Factoring a matrix into elementary matrices

We will factor a matrix into the product of elementary matrices (which is the same thing as writing the transformation as a composition of the much simpler elementary transforms). This process uses row reduction and requires keeping track of each step in the forward phase of Gaussian elimination. This factorization will lead to an efficient algorithm used in practice for finding the determinant of large matrices, computing inverses, or solving systems of equations. We'll focus on how elementary matrices simplify the process of computing determinants.

We know that multiplying a matrix A on the left by an elementary matrix E is equivalent to performing a row operation on A. If A reduces to the identity matrix I, we can repeatedly multiply by elementary matrices on the left (one matrix for each row operation) to obtain an equation

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

Recall that the inverse of a product of matrices is found by inverting each matrix and reversing the order of the product, $(AB)^{-1} = B^{-1}A^{-1}$. We now multiply both sides of the equation by

$$(E_n \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_n^{-1},$$

and then solve for A to obtain

$$A = E_1^{-1} E_2^{-1} \cdots E_n^{-1}.$$

If A row reduces to the identity matrix, then A is the product of elementary matrices.

Since the determinant of a product is the product of the determinants, we now have

$$|A| = |E_1^{-1}||E_2^{-1}| \cdots |E_n^{-1}|.$$

Each of these determinants is very easy to calculate (see the list above), and their product is the determinant of A. Basically, all we had to do was row reduce the matrix, and then we get the determinant from that for practically free. This is much, *much* less work than our previous way of calculating determinants.

There is even a faster way to calculate the determinant by factoring the matrix. We don't have to finish the row reduction to calculate the determinant easily. If we stop the row reduction once we have obtained an upper triangular matrix U (after the forward phase of row reduction, when we have put zeros in the lower left corner), then we could write

$$A = E_1^{-1} E_2^{-1} \cdots E_n^{-1} U.$$

The determinant of A is simply the product of the determinants of each elementary matrix multiplied by the determinant of the triangular matrix U, i.e.

$$|A| = |E_1^{-1}||E_2^{-1}| \cdots |E_n^{-1}||U|.$$

Since the determinant of U is easy to calculate (the product of its diagonal entries), we get the determinant easily after only doing the forward phase of Gaussian elimination.

Let's illustrate how to find the determinant using row reduction with an example, and then state some obervations as a theorem.

Example 4.23. Consider the matrix $A = \begin{bmatrix} 0 & -1 & 3 \\ 1 & 2 & 3 \\ 2 & -4 & 0 \end{bmatrix}$. Let's find the determinant of A using row reduction. We reduce A to an upper triangular matrix as follows:

$$\begin{bmatrix} 0 & -1 & 3 \\ 1 & 2 & 3 \\ 2 & -4 & 0 \end{bmatrix} \xrightarrow[E_1]{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 2 & -4 & 0 \end{bmatrix} \xrightarrow[E_2]{-2R_1 + R_3 \to R_3} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & -8 & -6 \end{bmatrix}$$

$$\xrightarrow[E_3]{-8R_2 + R_3 \to R_3} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & 0 & -30 \end{bmatrix}$$

So far we can write this reduction process as the following product of elementary matrices.

$$E_3 E_2 E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -8 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & 0 & -30 \end{bmatrix}.$$

Solving for A we obtain (remember to invert each elementary matrix and reverse the order)

$$A = E_1^{-1} E_2^{-1} E_3^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & 0 & -30 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ +2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & +8 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & 0 & -30 \end{bmatrix}.$$

The determinant of A is the product of the determinants of the elementary matrices and the upper triangular matrix,

$$|A| = |E_1^{-1}||E_2^{-1}||E_3^{-1}||U| = (-1)(1)(1)(30) = -30,$$

where the 30 is simply the product of the diagonal entries of U. Notice that if we just kept track of the fact that we only switched rows once, and that we never multiplied any row by a nonzero constant, then we could have found this determinant by just multiplying the determinant of our upper triangular matrix by (-1) (one row swap).

Theorem 4.6 (Determinants through Row Reduction). Let A be a square matrix. Row reduce A to an upper triangular matrix U, and let $E_1, E_2, \ldots, E_n$ be the corresponding elementary matrices obtained from each row operation. Then $E_n \cdots E_2 E_1 A = U$, or equivalently $A = E_1^{-1} E_2^{-1} \cdots E_n^{-1} U$, which means the determinant of A is $|A| = |E_1|^{-1} |E_2|^{-1} \cdots |E_n|^{-1} |U|$.

Computers use this algorithm (and other algorithms too) to find determinants of large matrices. The cofactor expansion method we learned earlier requires approximately $n!$ (n factorial) computations, which is computationally impossible for large values of n. When $n = 30$ (a fairly small matrix compared to what industry uses), the determinant would take around $30! = 265252859812191058636308480000000 = 2.65 \times 10^{32}$ computations using the cofactor expansion method. It would take the fastest supercomputer in the world (in 2010, having around 22,000 processors) over 2 billion years to compute this determinant using cofactor expansion. Using the algorithm from this section, a 2010 laptop can calculate the determinant in about 0.0003 seconds.

Example 4.24. Consider the matrix $A = \begin{bmatrix} 0 & 3 & 4 & -2 \\ -1 & 2 & -2 & 1 \\ 2 & 1 & 5 & 3 \\ 1 & 4 & 2 & -1 \end{bmatrix}$. We now reduce A to

an upper triangular matrix and then find the determinant.

$$
\begin{bmatrix} 0 & 3 & 4 & -2 \\ -1 & 2 & -2 & 1 \\ 2 & 1 & 5 & 3 \\ 1 & 4 & 2 & -1 \end{bmatrix}
\xrightarrow[E_1]{R_1 \leftrightarrow R_4}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ -1 & 2 & -2 & 1 \\ 2 & 1 & 5 & 3 \\ 0 & 3 & 4 & -2 \end{bmatrix}
\xrightarrow[E_2]{R_1 + R_2 \to R_2}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 6 & 0 & 0 \\ 2 & 1 & 5 & 3 \\ 0 & 3 & 4 & -2 \end{bmatrix}
$$

$$
\xrightarrow[E_3]{-2R_1 + R_3 \to R_3}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 6 & 0 & 0 \\ 0 & -7 & 1 & 5 \\ 0 & 3 & 4 & -2 \end{bmatrix}
\xrightarrow[E_4]{R_2 \frac{1}{6} \to R_2}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & -7 & 1 & 5 \\ 0 & 3 & 4 & -2 \end{bmatrix}
\xrightarrow[E_5]{7R_2 + R_3 \to R_3}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 5 \\ 0 & 3 & 4 & -2 \end{bmatrix}
$$

$$
\xrightarrow[E_6]{-3R_2 + R_4 \to R_4}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 4 & -2 \end{bmatrix}
\xrightarrow[E_7]{-4R_3 + R_4 \to R_4}
\begin{bmatrix} 1 & 4 & 2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & -22 \end{bmatrix}
$$

The determinants of the elementary matrices are all 1 except for $|E_1| = -1$ and $|E_4| = \frac{1}{6}$. The determinant of the upper triangular matrix at the last stage is -22 (the product of the diagonals). Since this upper triangular matrix is $U = E_7 E_6 E_5 E_4 E_3 E_2 E_1 A$, we know that $(1)(1)(1)(\frac{1}{6})(1)(1)(-1)|A| = -22$. Solving for $|A|$ gives us $|A| = (-6)(-22) = 132$.

To find a determinant using row reduction, all you have to do is reduce the matrix to an upper triangular matrix. Then find the determinant of this upper triangular matrix, and modify it as follows:

1. Each time you you interchanged rows, change the sign of the determinant.

2. Each time you multiply a row by a nonzero constant, divide the determinant of your upper triangular matrix by this constant.

This will give you the determinant of your original matrix. Notice that adding a multiple of one row to another does not change the determinant, since the corresponding elementary matrix has determinant 1.

4.7   Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Construct graphical displays of linear transformations. Graphically explain how to see eigenvalues, eigenvectors, and determinants from these visualizations, as well as when an inverse transformation exists.

2. Know the vocabulary of functions: domain, range, image, inverse image, kernel, injective (one-to-one), surjective (onto), and bijective.

3. Define linear transformation, and give examples involving both finite and infinite dimensional vector spaces. Show that every linear transformation between finite dimensional vector spaces can be represented by a matrix in relation to the standard basis vectors.

4. Describe the column space, null space, and eigenspaces of a matrix. Show that these spaces are vector spaces.

5. Show that every solution to the linear equation $T(\vec{x}) = \vec{b}$ can be written in the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_p$ is a particular solution and $\vec{x}_h$ is a solution to the homogeneous equation $T(\vec{x}) = \vec{0}$.

6. Explain what it means to compose linear transformations (both algebraically and graphically). Then show how to decompose a matrix as the product of elementary matrices, and use this decomposition (through row reduction) to compute the determinants of large matrices.

Here are the preparation problems for this unit. Problems that come from Schaum's Outlines "Beginning Linear Algebra" are preceded by a chapter number. From here on out, many of the problems we will be working on come from Schaum's Outlines. Realize that sometimes the method of solving the problem in Schaum's Outlines will differ from how we solve the problem in class. The difference is that in Schaum's Outlines they almost always place vectors in rows prior to row reduction, whereas we will be placing vectors in columns. There are pros and cons to both methods. Using columns helps us focus on understanding coordinates of vectors relative to bases, whereas using rows helps us determine when two vector spaces are the same.

<div align="center">Preparation Problems (click for handwritten solutions)</div>

Day 1	1a, 2a, Schaum's 8.4, Schaum's 8.14
Day 2	Schaum's 8.17, Schaum's 9.7, Schaum's 8.21, Schaum's 8.29
Day 3	Schaum's 9.4, Schaum's 8.38, Schaum's 5.24, Schaum's 3.15,
Day 4	Everyone do 4
Day 5	Lesson Plan, Quiz, Start Project

Concept	Where	Suggestions	Relevant Problems
Visualizations	Here	1ac,2ab,3ab	All
Mappings	Schaum's Ch 8	1,4,7,9,53	1-10,12,51-55
Linear Transformations	Schaum's Ch 8	14,15,17	13-20, 57-66
Standard Matrix Representation	Schaum's Ch 9	4,7,10	1a,4,7,10,27a,29,30,34,42,
Important subspaces	Schaum's Ch 8	21,23,25	21-26, 70-78
Types of Transformations	Schaum's Ch 8	29,31,35,38	29-32,34-39,79-80,85,89-91
Homogeneous Equations	Schaum's Ch 5	24,26	24-26,62-65
Elementary Matrices	Schaum's Ch 3	15,82	12-17,55-56,81-83
Determinants	Here	4	4 (multiple times)

4.8 Problems

Most of the homework from this unit comes from Schaum's Oultines. Here are some additional problems related to the topics in this unit.

1. 2D Transformations: Use Sage to visualize how each of the following matrices transforms the plane. Before drawing each one, first construct your own 2D graph where you transform a 1 by 1 box to the appropriate parallelogram (so show where you send the vectors $(1,0)$, $(0,1)$, and $(1,1)$). Then use Sage to see if your graph is correct. Notice how the determinant of the transformation is precisely the area of the transformed region (where a negative sign means the image was flipped). Look for eigenvectors and convince yourself that the transformation stretches the image radially outwards in this direction, where the scaling factor is the eigenvalue. If the eigenvalues are complex, look for a clockwise or counterclockwise rotation.

   (a) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$, $\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}$

   (b) $\begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix}$, $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$

   (c) $\begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$, $\begin{bmatrix} -2 & 4 \\ 4 & -8 \end{bmatrix}$, $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$

   (e) Type in 3 more matrices of your own choice, trying to obtain examples with positive, negative, and complex eigenvalues.

   (f) Use a rotation matrix to visualize rotations in the plane using the angles $\pi/4, \pi/2, \pi, -\pi/2, 3\pi/5, 3\pi$, do any result in real (instead of complex) eigenvalues? Why or why not?

2. 3D Transformations: For each matrix below, start by determining where the unit cube is mapped. Then use Sage to visualize the 3D transformation. Identify the determinant and eigenvectors in the diagrams, and relate them to the concepts from the 2D problems above.

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

   (b) $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, $\begin{bmatrix} 2 & 4 & 1 \\ 4 & 2 & 0 \\ 0 & 6 & 2 \end{bmatrix}$, $\begin{bmatrix} 2 & 0 & 4 \\ 0 & 2 & 0 \\ 4 & 0 & 2 \end{bmatrix}$

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

3. Non Square Transformations: In order to visualize a transformation which results from a non square matrix, extend the matrix to be a square matrix by adding rows or columns of zeros. Before using the computer, construct a rough sketch of the linear transformation by determining where each basis vector is mapped. Then use Sage to visualize the transformation. Notice how in each case 0 is an eigenvalue, and the eigenvector corresponding to 0 is the direction in which space was smashed so that you obtain the appropriate transformation.

   (a) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 2 \end{bmatrix}$, $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$, $\begin{bmatrix} 3 & -2 \end{bmatrix}$

   (b) $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 1 \end{bmatrix}$

4. Elementary Matrices and Determinants: Select a large matrix (4 by 4 or bigger) and compute the determinant using software. Then use row reduction by hand to verify that this is the determinant of the matrix. If the determinant is too large to work with, then change your matrix until you get a smaller determinant. Repeat this problem as needed.

4.9   Projects

4.10   Solutions

Please use Sage to check all your work.

Chapter 5

Changing Bases

This chapter covers the following ideas. The purpose of this chapter is to help you understand different bases can provide useful ways of viewing a linear transformation.

1. Explain how to describe vectors using different coordinate systems. Be able to change between different coordinate systems, and use these ideas to construct matrix representations of linear transformations.

2. Show that the null space (kernel), column space (image), and eigenspaces of a matrix (linear transformation) are vector subspaces. Be able to extend a basis for these spaces to a basis for the domain or codomain.

3. For square matrices, explain why similar matrices $B = P^{-1}AP$ represent the same linear transformation (just using a different basis). Explain why the determinant, eigenvalues, rank, and nullity of a linear transformation are not dependent upon the basis chosen.

4. Explain how to diagonalize a matrix, as well as explain when it is or is not possible.

5.1 Coordinate Systems

5.1.1 Coordinates relative to a basis

When we write the vector $\vec{u} = (2, 4, 7)$, we are giving the coordinates of the vector relative to the standard basis $E = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. We could write $(2, 4, 7)$ as a linear combination of these basis vectors by writing $(2, 4, 7) = 2\vec{e}_1 + 4\vec{e}_2 + 7\vec{e}_3$. The numbers $2, 4, 7$ are called the *coordinates* of $\vec{u}$ relative to the standard basis vectors, and they form a vector $[\vec{u}]_E = (2, 4, 7)_E$ called the coordinate vector of $\vec{u}$ relative to the basis E. We use the subscript E to emphasize that the vector $(2, 4, 7)$ is really just the coordinate vector whose components are the coordinates relative to E.

If instead we use the basis vectors $S = \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}$, then since $(2, 4, 7) = 2(1, 1, 1) + 2(0, 1, 1) + 3(0, 0, 1)$, we write $[\vec{u}]_S = (2, 2, 3)_S$ as the coordinate vector of $\vec{u}$ relative to S. We will write $[(2, 4, 7)]_S = (2, 2, 3)_S$ to say that the coordinates of $(2, 4, 7)$ relative to S are $(2, 2, 3)_S$. The subscript S reminds us which basis vectors we are using for the coordinates.

When we work with polynomials in $P_3(x)$ with standard basis $E = \{1, x, x^2, x^3\}$, the coordinates of $2 + 4x - 5x^2 + 3x^3$ relative to this basis are $[2 + 4x - 5x^2 + 3x^3]_E = (2, 4, -5, 3)_E$. If the basis is understood from context, then often we will omit the

subscript. The coordinates of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & -2 & 0 \end{bmatrix}$ relative to the standard basis

$$S = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

are simply $\left[\begin{bmatrix} 1 & 2 & 3 \\ 4 & -2 & 0 \end{bmatrix} \right]_S = (1, 2, 3, 4, -2, 0)$ (we didn't put a subscript on the last coordinate vector because the basis was clear from context).

How do you find the coordinates of a vector relative to a basis? We've been doing it in every chapter up till now every time we find reduced row echelon form. The coordinates of a vector relative to pivot columns are found in the non-pivot columns of rref. The new thing so far in this chapter is the notation $[\vec{u}]_S$ for the coordinate vector relative to the basis S. Let's remind ourselves of why row reducing gives the coordinate vector.

Example 5.1. To find the coordinates of $(1, 0)$ relative to the basis $S = \{(1, 2), (3, 5)\}$, we need to find coefficients (x, y) such that

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{or in matrix form} \quad \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

To solve this system we simply reduce the matrix $\begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 0 \end{bmatrix}$ to $\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \end{bmatrix}$. This shows us that the coordinates of $(1, 0)$ relative to S are simply $[(1, 0)]_S = (-5, 2)_S$.

Notice that the matrix form in the last example showed that if B is a matrix with the basis vectors as columns, then $B[\vec{x}]_S = \vec{x}$. $B[\vec{x}]_s = \vec{x}$

Example 5.2. Consider the basis $S = \{(1, 2), (3, 5)\}$.

1. Find the coordinates of $(1, 0)$, $(0, 1)$, $(3, 6)$, $(10, -7)$, and (a, b) relative to this basis.

2. What vector has coordinates $(3, -2)_S$?

3. How does the inverse of $B = \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$ help find the coordinates of (a, b)?

To answer the first question, we could do five different row reductions. However, we can do them all simultaneously by inserting the basis vectors as the first two columns of a matrix, augmenting by all of the other vectors, and then reducing:

$$\begin{bmatrix} 1 & 3 & 1 & 0 & 3 & 10 & a \\ 2 & 5 & 0 & 1 & 6 & -7 & b \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -5 & 3 & 3 & -71 & -5a + 3b \\ 0 & 1 & 2 & -1 & 0 & 27 & 2a - b \end{bmatrix}.$$

The coordinates of each vector relative to S are $[(1, 0)]_S = (-5, 2)_S$, $[(0, 1)]_S = (3, -1)_S$, $[(3, 6)]_S = (3, 0)_S$, $[(10, -7)]_S = (-71, 27)_S$, and $[(a, b)]_S = (-5a + 3b, 2a - b)_S$. The columns of the rref give us the coordinates of each vector relative to our basis.

The coordinates $(3, -2)_S$ represent the vector $3(1, 2) - 2(3, 5) = (-3, -4)$. We can write this fact as $[(-3, -4)]_S = (3, -2)_S$.

Notice that in our answer for the first question, we already augmented by $(1, 0)$ and $(0, 1)$ prior to reducing. This means that columns 3 and 4 of the rref contain the inverse $B^{-1} = \begin{bmatrix} -5 & 3 \\ 2 & -1 \end{bmatrix}$. Also notice that

$$B^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -5 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -5a + 3b \\ 2a - b \end{bmatrix}$$

which is precisely the coordinates of (a, b) relative to S.

Notice in the last answer in this example, we see that $B^{-1}\vec{x}$ gives the coordinates of $\vec{x}$ relative to S, i.e., $[\vec{x}]_S = B^{-1}\vec{x}$.

$[\vec{x}]_s = B^{-1}\vec{x}$

The two methods above (using row reduction or inverses) will always give you the coordinates of $\vec{u}$ relative to the basis $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ (where S is a basis for $\mathbb{R}^n$). We will focus on using inverse matrices for the remainder of this chapter, as we need inverse matrices in what follows. Much of the work in this chapter involves using inverse matrices to discover new ideas.

Example 5.3. Given the basis $S = \{(1, 2), (3, 4)\}$, we will find the coordinate of $(3, 6)$, $(10, -7)$, and (a, b) relative to this basis by using an inverse matrix. First, the matrix B and B^{-1} are

Remember the trick from Section 2.5.3 to easily find the inverse of a 2 by 2 matrix:

$$B = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad B^{-1} = \frac{1}{-2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}.$$

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

The coordinates relative to S are

$$[(3, 6)]_S = B^{-1}\begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}_S,$$

$$[(10, -7)]_S = B^{-1}\begin{bmatrix} 10 \\ -7 \end{bmatrix} = \begin{bmatrix} -61/2 \\ 27/2 \end{bmatrix}_S,$$

$$[(a, b)]_S = B^{-1}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -2a + \frac{3b}{2} \\ a - \frac{b}{2} \end{bmatrix}_S.$$

5.1.2 Change of Basis

If we are given the coordinates $[\vec{u}]_S$ of a vector $\vec{u}$ relative to a basis S, how do we find the coordinates $[\vec{u}]_R$ of $\vec{u}$ relative to a different basis R? We can start by writing everything in terms of a single basis (often the standard basis). Let B be the matrix whose columns are the basis vectors of S (relative to the standard basis E). Let C be the matrix whose columns are the basis vectors of R (relative to the standard basis E). For any vector $\vec{u}$, whose coordinates relative to the standard basis are $[\vec{u}]_E$, we write the two equations

$$B[\vec{u}]_S = [\vec{u}]_E \quad \text{and} \quad C[\vec{u}]_R = [\vec{u}]_E.$$

We equate the left side of each equation to obtain the matrix equation

$$B[\vec{u}]_S = C[\vec{u}]_R.$$

Multiplication on the left by C^{-1} gives

$$C^{-1}B[\vec{u}]_S = P[\vec{u}]_S = [\vec{u}]_R.$$

You can think of this transformation P as a composition of transformations, $P[\vec{u}]_S = C^{-1}(B[\vec{u}]_S) = C^{-1}[\vec{u}]_E = [\vec{u}]_R$, or in words, first convert $[\vec{u}]_S$ to standard coordinates $[\vec{u}]_E$, and then convert to coordinates relative to R, $[\vec{u}]_R$.

The matrix $P = C^{-1}B$ is often called the "change of basis matrix from S to R," or the "change of coordinate matrix from S to R." To transform from the old basis S to the new basis R, we just multiply by P: $P[\vec{u}]_S = [\vec{u}]_R$. The change of basis matrix Q from R to S inverts this process, so it is $Q = P^{-1} = B^{-1}C$.

Example 5.4. Consider the two bases for the plane $\mathbb{R}^2$:

$$S = \{(2, 1), (3, -2)\} \quad \text{and} \quad R = \{(1, 0), (2, -1)\}.$$

1. Find the coordinates of $(-1, 3)_S$ relative to the basis R by finding the change of basis matrix P from S to R,

2. Find the coordinates of $(-1, 3)_R$ relative to the basis S by finding the change of basis matrix Q from R to S, and

3. Verify that $Q = P^{-1}$.

1. Let $B = \begin{bmatrix} 2 & 3 \\ 1 & -2 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$ (the matrices whose columns are the basis vectors S and R, respectively). We know that $B \begin{bmatrix} -1 \\ 3 \end{bmatrix}_S = \begin{bmatrix} 7 \\ -7 \end{bmatrix} = \vec{u}$ is the actual vector we are trying to represent in terms of the basis R. To find the coordinates relative to R, we write $B \begin{bmatrix} -1 \\ 3 \end{bmatrix}_S = C[\vec{u}]_R$, or $C^{-1} B \begin{bmatrix} -1 \\ 3 \end{bmatrix}_S = [\vec{u}]_R$, where $C^{-1} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$. The change of basis matrix P from S to R is $C^{-1} B = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$ which means that $[\vec{u}]_R = P \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -7 \\ 7 \end{bmatrix}_R$. So the vector with coordinates $(-1, 3)_S$ relative to S is $(7, -7)$ which has coordinates $(-7, 7)_R$ relative to R.

2. We know that $C \begin{bmatrix} -1 \\ 3 \end{bmatrix}_R = \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \vec{u}$ is the actual vector we are trying to represent in terms of the basis S. To find the coordinates relative to R, we write $C \begin{bmatrix} -1 \\ 3 \end{bmatrix}_R = B[\vec{u}]_S$, or $B^{-1} C \begin{bmatrix} -1 \\ 3 \end{bmatrix}_R = [\vec{u}]_S$, where $B^{-1} = \begin{bmatrix} \frac{2}{7} & \frac{3}{7} \\ \frac{1}{7} & -\frac{2}{7} \end{bmatrix}$. The change of basis matrix Q from R to S is $B^{-1} C = \begin{bmatrix} \frac{2}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{4}{7} \end{bmatrix}$ which means that $[\vec{u}]_R = Q \begin{bmatrix} -1 \\ 3 \end{bmatrix}_S = \begin{bmatrix} \frac{1}{7} \\ \frac{11}{7} \end{bmatrix}_S$. So the vector with coordinates $(-1, 3)_R$ relative to R is $(5, -3)$ which has coordinates $(1/7, 11/7)_S$ relative to S.

3. Because $P = C^{-1} B$ and $Q = B^{-1} C$, we know that $P^{-1} = (C^{-1} B)^{-1} = B^{-1}(C^{-1})^{-1} = B^{-1} C = Q$ (remember that inverting matrices requires that we invert each matrix and reverse the order). Hence $P^{-1} = Q$.

Picking the right coordinates can help us simplify a problem tremendously. Changing coordinates lets us express a problem in many different ways. Often, we want to think about a problem in a simple way (using coordinates that are natural for the problem), but then in order to do more work on the problem, we must change coordinates to express the problem differently. The next example illustrates this concept.

Example 5.5. Our goal in this example is to find the equation of an ellipse that is rotated 45 degrees, centered at the origin, and with major axis length 2 and minor axis length 1. Looking at the picture in the margin, we can see that it would be very natural to think of this ellipse using the dotted axes rather than the normal coordinate axes.

This problem involves looking at two different coordinate systems for the plane. One coordinate system has the standard basis $S = \{(1, 0), (0, 1)\}$, which we will call the xy coordinate system. The other coordinate system involves the directions along the dotted lines. To preserve distances along the dotted lines, we'll take as our rotated basis the unit vectors in these directions: $R = \{(1/\sqrt{2}, 1/\sqrt{2}), (-1/\sqrt{2}, 1/\sqrt{2})\}$.

In summary, we have the following coordinate systems now:

1. The standard xy coordinate system with basis

$$S = \{(1, 0), (0, 1)\}$$

represented by the x and y axes. Let B be the matrix with the vectors of S as columns. Consider the vector $\vec{u} = (a, b)$. The coordinates $(x, y)_S$ of $\vec{u}$ relative to S satisfy the matrix equation

$$B \begin{bmatrix} x \\ y \end{bmatrix}_S = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{or substituting in } B, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_S = \begin{bmatrix} a \\ b \end{bmatrix}.$$
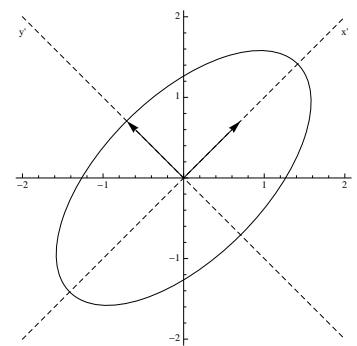
2. The alternate $x'y'$ coordinate system with basis

$$R = \{(1/\sqrt{2}, 1/\sqrt{2}), \, (-1/\sqrt{2}, 1/\sqrt{2})\}$$

represented by the x' ($y = x$) and y' ($y = -x$) axes. Let C be the matrix with the vectors of R as columns. Consider again the vector $\vec{u} = (a, b)$. The coordinates $(x', y')_R$ of (a, b) relative to R (so $[\vec{u}]_R = (x', y')_S$) satisfy the matrix equation

$$C \begin{bmatrix} x' \\ y' \end{bmatrix}_R = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{or substituting in } C, \quad \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}_R = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Our goal is to find an equation of the ellipse in the xy coordinate system. We know that in the $x'y'$ coordinate system an equation is $\dfrac{(x')^2}{4} + (y')^2 = 1$. To transform an equation from using R coordinates to using the normal S coordinates, we need to find a matrix P so that $[\vec{u}]_R = P[\vec{u}]_S$. In other words, we need to find the change of basis matrix from S to R (notice the order reversed).

For every vector $\vec{u} = (a, b)$, we already have the equations

$$B[\vec{u}]_S = [\vec{u}]_E = C[\vec{u}]_R \qquad \text{or} \qquad B \begin{bmatrix} x \\ y \end{bmatrix} = C \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

To find P, we simply solve for $[\vec{u}]_R$, which gives us $C^{-1}B[\vec{u}]_S = [\vec{u}]_R$, or $P = C^{-1}B$, or more explicitly,

$$P = C^{-1}B = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

We now have the transformation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = P \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x/\sqrt{2} + y/\sqrt{2} \\ -x/\sqrt{2} + y/\sqrt{2} \end{bmatrix},$$

or $x' = x/\sqrt{2} + y/\sqrt{2}, y' = -x/\sqrt{2} + y/\sqrt{2}$.

We can now answer the original question. The rotated ellipse had equation $\frac{(x')^2}{4} + (y')^2 = 1$ in the $x'y'$ coordinate system. We replace each x' and y' with $x' = x/\sqrt{2} + y/\sqrt{2}$ and $y' = -x/\sqrt{2} + y/\sqrt{2}$ to obtain the equation in terms of the normal xy coordinates:

$$\frac{(x/\sqrt{2} + y/\sqrt{2})^2}{4} + (-x/\sqrt{2} + y/\sqrt{2})^2 = 1.$$

Any time you encounter a problem that would be simpler in a new coordinate system, there is a change of basis matrix that can simplify the problem. [i]

---

[i]Different authors disagree on the use of the term "change of basis matrix." We call P the change of basis matrix *from S to R*, referring to the fact that multiplying $[\vec{u}]_S$ by P transforms coordinates relative to S into coordinates relative to R: $P[\vec{u}]_S = [\vec{u}]_R$. When working with an equation, for example, another point of view is that we can quickly convert an equation using R coordinates $[\vec{u}]_R$ to an equation using S coordinates $[\vec{u}]_S$ by replacing each instance of $[\vec{u}]_R$ with $P[\vec{u}]_S$. For example, if we know that $[\vec{u}]_R + [\vec{v}]_R - 2[\vec{w}]_R = [\vec{x}]_R$, then we can transform this equation to an equation involving coordinates relative to S by multiplying each coordinate vector by P: $P[\vec{u}]_S + P[\vec{v}]_S - 2P[\vec{w}]_S = P[\vec{x}]_S$. Hence some people call P the change of basis matrix *from R to S* (instead of *from S to R*). Both phrases appear in different places and both can be considered correct. The difference lies in the context of the problem and how you plan to use the matrix. Are you focusing on what the transformation does to individual vectors, or what the transformation does to equations?

In the equation $P[\vec{u}]_S = [\vec{u}]_R$, should we say P is a change of basis matrix from S to R or from R to S? Both can be considered correct, so make sure you know what the author means by the words "change of basis matrix."

5.2 Matrix Representations of Linear Transformations

We showed in the last chapter that every linear transformation between two finite dimensional vector spaces has a standard matrix representation. As a reminder, let's work a small example.

Example 5.6. The linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x,y) = (2x + y, 3x + 4y)$ we represented in matrix form by first computing the transformation at each standard basis vector, i.e., $T(1,0) = (2,3), T(0,1) = (1,4)$, and then putting these vectors in the columns a matrix $A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$.

We compute $T(\vec{x})$ by computing the matrix product $A\vec{x}$. We replaced the linear transformation with matrix multiplication. We now examine how the matrix A changes if we use a different basis to represent the domain and range.

Consider a linear transformation $T\colon V \to V'$ (where both V and V' are finite dimensional) with corresponding bases S and R. How do we construct the matrix representation of T relative to these bases? We need a matrix A so that we can input coordinates relative to S and get out coordinates relative to R. Notationally this matrix is written as $[T]_{S,R}$ and satisfies the equation

$$[T]_{S,R}[\vec{v}]_S = [T(\vec{v})]_R.$$

To find our matrix, we first consider everything relative to a standard basis.

- Let A be the standard matrix representation of the transformation. So

$$A\vec{v} = T(\vec{v}).$$

- Let B be the matrix whose columns are the basis vectors of S. So

$$B[\vec{v}]_S = [\vec{v}]_E = \vec{v}.$$

- Let C be the matrix whose columns are the basis vectors of R. So

$$C[\vec{u}]_R = [\vec{u}]_E = \vec{u}, \quad \text{or} \quad C[T(\vec{v})]_R = [T(\vec{v})]_E = T(\vec{v}).$$

On the left of $A\vec{v} = T(\vec{v})$ we replace $\vec{v}$ with $B[\vec{v}]_S$, and on the right we replace $T(\vec{v})$ with $C[T(\vec{v})]_R$. We obtain the key equation

$$AB[\vec{v}]_S = C[T(\vec{v})]_R.$$

Multiplying both sides on the left by C^{-1} gives

$$[T]_{S,R}[\vec{v}]_S = C^{-1}AB[\vec{v}]_S = [T(\vec{v})]_R.$$

This last equation requires that we input the coordinates of a vector relative to S, and then we get out the coordinates of the transformed vector relative to the basis R. This is precisely what we were looking for, which means that the matrix representation relative to S and R is $[T]_{S,R} = C^{-1}AB$.

Example 5.7. Consider the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x,y) = (x+y, x-y, 2x+3y)$. Find a matrix representation using the bases $S = \{(2,3), (-1,1)\}$ and $R = \{(1,0,0), (1,1,0), (3,2,1)\}$. To start with, the standard matrix representation and basis matrices are

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

Any vector $\vec{u}$ in the domain can be written as coordinates relative to S using the product $B[\vec{u}]_S = \vec{u}$. Any vector in the image can be written as coordinates relative to R using the product $C[T(\vec{u})]_R = T(\vec{u})$. Since $A\vec{u} = T(\vec{u})$, we replace $\vec{u}$ with $B[\vec{u}]_S$, and we replace $T(\vec{u})$ with $C[T(\vec{u})]_R$ to obtain $AB[\vec{u}]_S = C[T(\vec{u})]_R$ or $C^{-1}AB[\vec{u}]_S = [T(\vec{u})]_R$. The inverse of C is $C^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$. We can compute our matrix representation as

$$C^{-1}AB = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -7 & 1 \\ -27 & -4 \\ 13 & 1 \end{bmatrix} = [T]_{S,R}.$$

Just as changing coordinates helps simplify problems, changing coordinates can also help us simplify transformations. Often, we can find "natural" coordinates for the domain and range so that the transformation matrix $[T]_{S,R}$ is a very simple matrix. When the matrix is very simple, it is easier to intuitively understand the transformation.

The rest of this section focuses on learning to pick a good basis for the domain and the range so that the matrix representation of T, $[T]_{S,R}$, is as simple as possible, in particular so that the matrix consists of the identity matrix in the upper left corner and zeroes everywhere else:

$$[T]_{S,R} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

To obtain this representation, we need to learn how to extend bases for the kernel and image to bases for the entire domain and range. We'll start with this, and then apply our knowledge to matrix representations of transformations.

5.2.1 The Kernel, Image, and Eigenspaces

We have already spent time studying the three most important subspaces related to a linear transformation. These subspaces are the kernel (null space), the image (column space), and the eigenspaces. Before proceeding, let's review their definitions and why each is a subspace.

Definition 5.1. Consider a linear transformation $T\colon V \to U$.

1. The kernel (null space) is the set of all vectors in the domain V whose image is zero: $\{\vec{x} \mid T(\vec{x}) = \vec{0}\}$.

2. The image (column space) is the set of all vectors in the range U whose preimage is not empty: $\{\vec{y} \mid T(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in V\}$.

3. If $V = U$, for each eigenvalue λ, the eigenspace corresponding to λ is the set of all vectors satisfying $T(\vec{x}) = \lambda\vec{x}$. Note that eigenvalues and eigenspaces are only defined if $V = U$, or in other words, if the domain and range of T are the same.

Example 5.8. Let's verify that each of these spaces is indeed a vector subspace using Theorem 3.43. In what follows, we'll assume that T is a linear transformation whose standard matrix representation is A, i.e., $T(\vec{x}) = A\vec{x}$.

1. Let's show that the kernel is a vector subspace of the domain. Since all the vectors in the kernel are vectors in the domain that get sent to zero, we know it is a subset of the domain. We now show it satisfies the three requirements to be a subspace. We will represent the zero vector of V as $\vec{0}_V$ and the zero vector of U as $\vec{0}_U$.

    (a) Every linear transformation satisfies $T(\vec{0}_V) = \vec{0}_U$, so the zero vector is in the kernel because $T(\vec{0}_V)$ is sent to the zero vector in U.

(b) Is the sum of two vectors in the kernel still in the kernel? If $T(\vec{x}_1) = \vec{0}_U$ and $T(\vec{x}_2) = \vec{0}_U$ (so $\vec{x}_1$ and $\vec{x}_2$ are in the kernel), then we simply compute $T(\vec{x}_1 + \vec{x}_2) = T(\vec{x}_1) + T(\vec{x}_2) = \vec{0}_U + \vec{0}_U = \vec{0}_U$. This shows that the sum $\vec{x}_1 + \vec{x}_2$ is in the kernel.

(c) Is a scalar multiple of something in the kernel still in the kernel? If $T(\vec{x}) = \vec{0}_U$ (so $\vec{x}$ is in the kernel), then we compute $T(c\vec{x}) = cT(\vec{x}) = c\vec{0}_U = \vec{0}_U$. This shows that the scalar multiple $c\vec{x}$ is in the kernel.

2. Now let's show that the image is a vector subspace of the range. This proof is easiest if we utilize the fact that the span of any set of vectors is always a subspace. Since the image of T is the same as the column space of its standard matrix representation A, then immediately the image of T is the span of the columns of A, hence a vector subspace.

3. Now we'll show that if $V = U$, then for each eigenvalue λ of T, the set of vectors that satify $T(\vec{x}) = \lambda\vec{x}$ is a vector subspace of both the domain and range. In order to compute eigenvalues, the domain and range must be the same, so we only need to show it satisfies the three requirements to be a subspace.

(a) The zero vector satifies $T(\vec{0}) = \vec{0} = \lambda\vec{0}$, so it is in the eigenspace. Note that $\vec{0}$ is always in the eigenspace, but $\vec{0}$ is not an eigenvector. The eigenvectors are the other vectors in the eigenspace.

(b) If $\vec{x}_1$ and $\vec{x}_2$ are in the eigenspace, then $T(\vec{x}_1) = \lambda\vec{x}_1$ and $T(\vec{x}_2) = \lambda\vec{x}_2$. We compute

$$T(\vec{x}_1 + \vec{x}_2) = T(\vec{x}_1) + T(\vec{x}_2) = \lambda\vec{x}_1 + \lambda\vec{x}_2 = \lambda(\vec{x}_1 + \vec{x}_2),$$

which shows that $\vec{x}_1 + \vec{x}_2$ is in the eigenspace.

(c) If $\vec{x}$ is in the eigenspace, then $T(\vec{x}) = \lambda\vec{x}$. We then compute $T(c\vec{x}) = cT(\vec{x}) = c\lambda\vec{x} = \lambda(c\vec{x})$. We've shown that $c\vec{x}$ is also in the eigenspace.

We have just shown that the kernel (null space), image (column space), and eigenspaces are vector spaces. You should take the time to work through the proofs above on your own. The fact that these sets of vectors are actually vector spaces means we can find a basis for each of them and talk about the coordinates of vectors in the domain or range relative to these bases.

5.2.2 Extending a basis for a subspace to the whole space

The three vector spaces above (kernel, image, and eigenspaces) play an important role in understanding a linear transformation. When creating a matrix representation of a linear transformation, we need a basis for the entire domain and the entire range. The kernel provides us with a basis for part of the domain. The image provides us with a basis for part of the range. To use these vector spaces, we must first extend a basis of a subspace to a basis for the whole space. That is our current goal.

Extending Bases

First, we need some key theorems which discuss the relationships between linearly independent sets of vectors, spanning sets of vectors, and sets of vectors that are both linearly independent and span (i.e., bases). We will briefly discuss these in class, and you can find more rigorous proofs in Jim Heffron's free book on linear algebra[ii] or Rob Beezer's free book on linear algebra[iii].

[ii]`http://joshua.smcvt.edu/linearalgebra/book.pdf`, section III.2, page 118

[iii]`http://linear.ups.edu/jsmath/latest/fcla-jsmath-latest.html`, section D

Theorem 5.1. In a vector space, a linearly independent set of vectors never has more vectors than a spanning set.

This means that if you know the size of the smallest spanning set, any set of vectors with bigger size is guaranteed to be linearly dependent.

Theorem 5.2. In a vector space, a spanning set of vectors never has less vectors than a linearly independent set.

This means that if you know the size of the largest linearly independent set, any set of vectors with smaller size is guaranteed to not span the space.

Theorem 5.3. In a vector space, all sets of vectors which are bases have the same number of vectors. We call this number the *dimension* of the vector space.

This refers to the fact that a basis is a set of vectors which both span the space and are linearly independent.

Furthermore, we can get always create a basis from a spanning or linearly independent set.

Theorem 5.4. If S is a spanning set of vectors for a vector space V, then we can make a basis for V by deleting appropriate vectors in S.

In the last theorem, "appropriate vectors" to delete are vectors which do not change $\text{span}(S)$ (i.e., we can delete a vector as long as the new set is still a spanning set).

Theorem 5.5. If S is a linearly independent set of vectors in a vector space V, we can make a basis for V by adding appropriate vectors to S (so that the basis contains the vectors in our initial set S).

In this last theorem, "appropriate vectors" we add are vectors which ensure that S is still linearly independent.

This last theorem leads to a way to construct a basis for a vector space from a basis for a subspace.

Theorem 5.6. Let V be a vector space, and U be a subspace of V. If B is a basis for U, then we can add vectors to B to make a basis C for the vector space V. In other words, we can make a basis for V which contains the vectors in B.

Since B is a basis for U, B is a linearly independent set of vectors in V. We just need to add appropriate vectors until we have a spanning set for V.

Now, we will take the kernel (a subspace of the domain) and add vectors to make it a basis for the domain. Similarly, we will take the image (a subspace of the range), and add vectors to make it a basis for the range.

Extending a basis for the kernel to a basis for the domain

The kernel of $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a vector subspace of the domain $\mathbb{R}^n$, so we can select a basis for the kernel. If the kernel has dimension k, we can pick k vectors $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ as the basis. This isn't a basis for the entire domain, so how can we pick $n - k$ other vectors $\{\vec{u}_{k+1}, \ldots, \vec{u}_n\}$ so that $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k, \vec{u}_{k+1}, \ldots, \vec{u}_n\}$ is a basis for the domain $\mathbb{R}^n$? There are many ways to do this, I'll just illustrate one. Place your k basis vectors in a matrix and augment by the n by n identity matrix. This produces and n by $n + k$ matrix. Reduce this matrix, and select as your basis the n pivot columns of the resulting matrix. The first k vectors, the basis for the null space, will always be selected, followed by $n - k$ of the standard basis vectors.

Example 5.9. The reduced row echelon form of $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. A basis for the null space is $\{(-2, 1, 0)\}$. Putting this vector in the first column of a matrix and augmenting by the identity gives us $\begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

A basis for the domain is the pivot columns, namely the first, second and fourth vectors $\{(-2, 1, 0), (1, 0, 0), (0, 0, 1)\}$.

Example 5.10. The reduced row echelon form of $A = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$ is $\begin{bmatrix} 1 & \frac{4}{3} \\ 0 & 0 \end{bmatrix}$. A basis for the null space is $\{(-4/3, 1)\}$. Putting this vector in the first column and augmenting by the identity gives us

$$\begin{bmatrix} -\frac{4}{3} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{4}{3} \end{bmatrix}.$$

A basis for the domain is the first and second columns, namely $\{(-4/3, 1), (1, 0)\}$. Are there other correct answers? Of course there are, as there are infinitely many ways to choose a basis.

Extending a basis for the image to a basis for the range

The image of $T \colon \mathbb{R}^n \to \mathbb{R}^m$ is a vector subspace of the range $\mathbb{R}^m$. If the image is r dimensional (recall the rank is the dimension) then we can select a basis for the image consisting of r vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$. How do we pick $m - r$ other vectors $\{\vec{u}_{k+1}, \ldots, \vec{u}_m\}$ so that $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{v}_{r+1}, \ldots, \vec{v}_m\}$ is a basis for the range $\mathbb{R}^m$? Again augment these r basis vectors by the m by m identity matrix, and row reduce. The m pivot columns of the resulting matrix provide your basis. This basis will always include the first r vectors, and then you just add in some of the standard basis vectors.

Example 5.11. The reduced row echelon form of $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. A basis for the column space is $\{(1, 2), (3, 1)\}$. Since it already has 2 dimensions, there is no need to extend it to a basis for the range, we already have a basis.

Example 5.12. The reduced row echelon form of $A = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$ is $\begin{bmatrix} 1 & \frac{4}{3} \\ 0 & 0 \end{bmatrix}$. A basis for the column space is $\{(3, 6)\}$. Putting this vector in the first column of a matrix with 2 rows and augmenting by the identity gives us $\begin{bmatrix} 3 & 1 & 0 \\ 6 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & \frac{1}{6} \\ 0 & 1 & -\frac{1}{2} \end{bmatrix}$. This means that a basis for the range is the first and second columns, namely $\{(3, 6), (1, 0)\}$.

Example 5.13. The reduced row echelon form of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ -2 & -4 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$. A basis for the column space is $\{(1, 3, -2)\}$. Putting this vector in the first column of a matrix with 3 rows and augmenting by the identity gives us $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}}$

$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} \end{bmatrix}$. This means that a basis for the range is the first three columns, namely $\{(1, 3, -2), (1, 0, 0), (0, 1, 0)\}$.

Obtaining a basis for the domain and range using eigenspaces

The last fact we will develop in this section relates to choosing a basis for both the domain and range using eigenvectors. The first fact is that eigenvectors corresponding to different eigenvectors are always linearly independent.

Theorem 5.7. If $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are n eigenvectors corresponding to n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are linearly independent.

For a proof, see problem 11.44. The key reason we want to use this theorem is that if the domain V and range V of a linear transformation $T \colon V \to V$ has dimension n, and there are n distinct eigenvalues, then we can form a basis for V by selecting an eigenvector corresponding to each eigenvalue. This creates a basis for V by using eigenvectors. A basis chosen in this fashion will always contain a basis for the null space (as vectors in the null space are eigenvectors corresponding to the eigenvalue $\lambda = 0$). Choosing your basis vectors to be eigenvectors will greatly simplify how we view a linear transformation.

Example 5.14. The matrix $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$ has eigenvalues $5, -2$. Corresponding eigenvectors are $(1,1)^T$ and $(-4,3)^T$ (verify this by hand or by computer). A basis for the domain and range can be chosen to be $\{(1,1), (-4,3)\}$. These vectors are linearly independent.

Example 5.15. The matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ has eigenvalues $3, 3, 1$. Notice that 3 appears with multiplicity 2. To find the eigenvectors, we subtract 3 from the diagonal and reduce $\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ to obtain $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Since there are two free variables, we can find 2 linearly independent eigenvectors corresponding to 3, namely $(1,1,0)$ and $(0,0,1)$. An eigenvector corresponding to 1 is $(-1,1,0)$. These three vectors are linearly independent, and so a basis for the domain and range is $\{(1,1,0), (0,0,1), (-1,1,0)\}$.

Example 5.16. Similar computations can be done for the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ which has eigenvalues $3, 3, 1$. To find the eigenvectors corresponding to 3, we subtract 3 from the diagonal and reduce $\begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ to obtain $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. There is only one free variable, so only one linearly independent eigenvector corresponding to 3, namely $(1,1,0)$. There is also only one linearly independent eigenvector corresponding to 1, which is still $(-1,1,0)$. [iv] We can only obtain two linearly independent eigenvectors, and so a basis for the domain and range cannot be obtained just from one eigenvectors alone. For now, we augment these two vectors by the identity and row reduce to obtain
$$\begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$
A basis for the domain and range which contains eigenvectors (and one other) is the first, second, and fifth columns, namely $\{(1,1,0), (-1,1,0), (0,0,1)\}$. The next chapter on Jordan form focuses on what to do when you cannot obtain a basis consisting of only eigenvectors.

5.2.3 Simplifying Matrix Representations

We have already seen there are many different ways of using a matrix to represent the same transformation. If B and C are invertible matrices, and A is the standard matrix representation of T, then $C^{-1}AB$ is a matrix representation of T relative to the bases S and R formed from the columns of B and C. Which bases S and R provide the best bases, or the simplest matrix representation? There is not really a correct answer here, as simplest is a relative term, but there are some nice bases you can choose to simplify

---

[iv]Some software packages return $(0,0,0)$ as a third eigenvector which the software expects you to forget, as the zero vector is by definition never an eigenvector.

computations. The kernel and image of a transformation provide us with a starting point to obtain key basis vectors needed to represent a transformation in an extremely simple way.

If you start with a basis for the kernel (null space) and extend it to a basis for the domain, then the matrix representation relative to this basis will contain columns of zeros (one for every basis vector in the null space). In other words, you can arrange your basis vectors to purposefully squash out of the matrix everything with gets mapped to zero. Purposefully filling a matrix with lots of zeros prior to multiplication can greatly simplify computations.

Example 5.17. Consider the linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x+y-z, 2x+3y+4z)$. The standard matrix representation is $A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 4 \end{bmatrix}$ whose rref is $\begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & 6 \end{bmatrix}$. A basis for the kernel (null space) is $(7, -6, 1)$. We extend this to a basis for the domain by augmenting by the identity and reducing, $\begin{bmatrix} 7 & 1 & 0 & 0 \\ -6 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}}$

$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 6 \end{bmatrix}$, which means that $S = \{(7, -6, 1), (1, 0, 0), (0, 1, 0)\}$ is a basis for the domain. Using $E = \{(1, 0), (0, 1)\}$ as a basis for the range, we compute the matrix representation of T relative to the bases S and E by noting

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = C^{-1}, A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, C^{-1}AB = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 2 & 3 \end{bmatrix} = [T]_{S,E}.$$

Notice how this made the entire first column of the matrix zero. If we rearrange the basis vectors of S so that the vectors in the null space are at the end, $S = \{(1, 0, 0), (0, 1, 0), (7, -6, 1)\}$, then our matrix is simply $[T]_{S,E} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix}$. We just moved the column of zeros to the end of the matrix instead of the beginning. Putting the column of zeros at the end is done in practice to make it easy to remember that we can ignore the last column (or columns) of the matrix (instead of the first).

After picking a basis for the domain which contains the kernel, let's carefully construct a basis for the range to make our matrix representation extremely nice. If k is the dimension of the kernel and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for the domain (whose last k vectors are in the kernel—they were moved to the end), then consider the $n - k$ vectors $\{T(\vec{v}_1), \ldots, T(\vec{v}_{n-k})\}$ which are nonzero vectors in the image. Since the image has to have dimension $n - k$, these $n - k$ vectors form a basis for the image. We can extend them to a basis R for the range (augmenting by the identity and reducing). The matrix representation of T relative to the bases S and R will always be a matrix which consists of all zeros and $n - k$ 1's (what can be simpler than zeros and 1's). This representation of T illustrates how the kernel and image can be use to obtain an extremely nice form for our matrix representation[v].

Example 5.18. Consider the linear transformation from example 5.17, $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x+y-z, 2x+3y+4z)$. Using the bases $S = \{(1, 0, 0), (0, 1, 0), (7, -6, 1)\}$ and $E = \{(1, 0), (0, 1)\}$ (the standard basis for the plane), the matrix representation of T relative to these bases is $[T]_{S,E} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix}$. Let's pick a different basis for the range. A basis for the image is $R = \{(1, 2), (1, 3)\}$, the nonzero columns of $[T]_{S,E}$. The

---

[v]Mathematicians who work with maps in high dimensions often start by placing things in "general position" which is very similar to picking nice bases so that their maps consist of identity and zero transformations, precisely what we are doing now.

matrix representation of T relative to the bases S and R is then found using

$$C = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, C^{-1} = \frac{1}{1}\begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\text{and so} \quad [T]_{S,R} = C^{-1}AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Not only have we made the last column zero, but we have also placed the identity matrix in the upper left corner of the matrix. This will happen in general, even with bigger matrices.

Example 5.19. Let's look at a much larger example. Perform the computations on the computer to verify the work below. Consider the transformation $T(x, y, z, w) = (2x - y, y + z, 2x + z, x + w)$. The standard matrix representation of T is

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

A basis for the kernel is $(-1, -2, 2, 1)$. We extend this to a basis for the domain by augmenting by the identity and reducing,

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix},$$

. A basis for the domain is $S = \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (-1, -2, 2, 1)\}$, where we placed the vector in the kernel at the end. Using the standard basis for the range, we compute the matrix representation of T relative to the bases S and E by noting

$$C = I = C^{-1}, A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\text{and so} \quad [T]_{S,E} = C^{-1}AB = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

The last column now contains only zeros. Now that we've got a basis for the domain that contain a basis for the kernel, we pick a different basis for the range. The vectors $\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ have images $\{(2, 0, 2, 1), (-1, 1, 0, 0), (0, 1, 1, 0)\}$ (the nonzero columns of the previous matrix). These nonzero columns provide a basis for the image. Extend this set to a basis for the range by augmenting by the identity and reducing

$$\begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 \end{bmatrix}.$$

We select the pivot columns $R = \{(2, 0, 2, 1), (-1, 1, 0, 0), (0, 1, 1, 0), (1, 0, 0, 0)\}$ as a basis for the range. The matrix representation of T relative to the bases S and R is then found using

$$C = \begin{bmatrix} 2 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, C^{-1} = \frac{1}{1}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 1 & -2 \\ 1 & 1 & -1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$[T]_{S,R} = C^{-1}AB = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In the first step we made the last column zero. In this step we put the identity matrix in the upper left corner of the matrix, with a row of zeros at the bottom. This is perhaps the simplest form you can obtain.

Every linear transformation $T\colon V \to V'$ of rank r can be represented by a matrix of the form $[T]_{S,R} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$, where the only nonzero entries are r 1's, located in the upper left corner. To obtain this basis require the following steps:

1. Find a basis for the kernel.

2. Extend the basis for the kernel to a basis S for the domain, placing the vectors in the kernel at the end.

3. Find the images each basis vector for the domain. The nonzero vectors form a basis for the image.

4. Extend the basis for the image to a basis R for the range, placing the vectors in the image at the front.

Along the way, there are infinitely many choices you can make at each step. The interesting thing is that regardless of the choices we make, we will always obtain the exact same form $[T]_{S,R} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. In the next section we will focus on what happens if $T\colon V \to V$ has the same domain and range, and we use the same basis S for both the domain and range.

5.3   Similar Matrices

If a linear transformation $T\colon V \to V$ has the same domain as range, then we often want to pick the same basis S for both the domain and range. The coordinates of $\vec{u}$ relative to S are written $P[\vec{u}]_S = [\vec{u}]_E$, where P is the matrix whose columns are the basis vectors in S. If A is the standard matrix representation of the linear transformation, then we write

$$A[\vec{u}]_E = [T(\vec{u})]_E.$$

Replacing $[\vec{u}]_E$ on the left with $P[\vec{u}]_S = [\vec{u}]_E$, and replacing $[T(\vec{u})]_E$ on the right with $P[T(\vec{u})]_S = [T(\vec{u})]_E$ gives the matrix equation

$$AP[\vec{u}]_S = P[T(\vec{u})]_S \quad \text{or} \quad P^{-1}AP[\vec{u}]_S = [T(\vec{u})]_S.$$

The matrix representation of T relative to the basis S is $[T]_{S,S} = [T]_S = P^{-1}AP$ (we only write one S). This new matrix is found by multiplying on the right by P, and on the left by P^{-1}. Pictorially, we see

$$S \xrightarrow{P} E \xrightarrow{A} E \xrightarrow{P^{-1}} S$$

Definition 5.2. We say A and B are similar matrices, and write $A \sim B$ if there exists an invertible matrix P such that $B = P^{-1}AP$.

Similar matrices have many properties that we'll put into a theorem.

Theorem 5.8. Similarity has the following properties[vi].

[vi]Any relationship between objects which satisfies the first three properties in this theorem is said to be an *equivalence relation*. Equivalence relations are studied more in later math classes.

1. Every matrix is similar to itself: $A \sim A$. (We say similarity is reflexive.)

2. If $A \sim B$, then $B \sim A$. (We say similarity is symmetric.)

3. If $A \sim B$ and $B \sim C$, then $A \sim C$. (We say similarity is transitive.)

4. Similar matrices have the same determinant, eigenvalues, rank, and nullity.

Proof. Let's take a moment to justify why each of the facts above is true.

1. Every matrix A is similar to itself because letting $P = I$, the identity matrix, gives $A = I^{-1}AI$.

2. If A is similar to B, then there is some P such that $B = P^{-1}AP$. If we solve for A by multiplying on the left by P and the right by P^{-1}, we get $A = PBP^{-1}$, or $A = (P^{-1})^{-1}BP^{-1} = Q^{-1}BQ$, where $Q = P^{-1}$. Hence B is similar to A.

3. If A is similar to B and B is similar to C, then $B = P^{-1}AP$ and $C = Q^{-1}BQ$ for some matrices P and Q. We replace B in the second equation to obtain $C = Q^{-1}BQ = Q^{-1}P^{-1}APQ = R^{-1}AR$ where $R = PQ$ (recall that $(PQ)^{-1} = Q^{-1}P^{-1}$). This shows that A is similar to C.

4. If A and B are similar, then $B = P^{-1}AP$. Taking determinants gives $|B| = |P^{-1}AP| = |P^{-1}||A||P| = \frac{1}{|P|}|A||P| = |A|$, hence similar matrices have the same determinant.

   To show they have the same eigenvalues, we compute

   $$\begin{aligned}
   |B - \lambda I| &= |P^{-1}AP - \lambda I| \\
   &= |P^{-1}AP - \lambda P^{-1}IP| \\
   &= |P^{-1}(A - \lambda I)P| \\
   &= |P^{-1}||(A - \lambda I)||P|,
   \end{aligned}$$

   which means $|B - \lambda I| = 0$ if and only if $|P^{-1}||(A - \lambda I)||P| = 0$. The latter equals zero if and only if $|(A - \lambda I)| = 0$. Hence B and A have the same eigenvalues.

   We now show the rank and nullity are the same. First, we'll show that for every vector in the null space of A, there is a corresponding vector in the null space of B (so the nullity of A cannot be larger than the nullity of B). If $\vec{x}$ is in the null space of A, then $A\vec{x} = \vec{0}$. The vector $\vec{y} = P^{-1}\vec{x}$ is in the null space of B since $B\vec{y} = P^{-1}AP(P^{-1}\vec{x}) = PA\vec{x} = P\vec{0} = \vec{0}$. Because P^{-1} is invertible, the linear transformation $T(\vec{x}) = P^{-1}\vec{x}$ has the property that $\ker T = \{\vec{0}\}$, which means that T is one-to-one. This means that every null vector $\vec{y}$ in the nullspace of B came from a different vector $\vec{x}$ in the nullspace of A. This means that the nullity of A has to be less than or equal to the nullity of B. If we apply the same reasoning the other way, we see that the nullity of B has to be less than or equal to the nullity of A. Therefore the nullity of A and the nullity of B are the same. Since rank plus nullity equals the number of columns for both matrices, we know that the ranks are the same as well.

   □

Similar matrices are just different ways of viewing a linear transformation by looking at it in terms of different bases. The facts above show that the determinant, eigenvalues, rank, and nullity of a linear transformation do not depend on the basis chosen. They are important numbers which are properties of the linear transformation, and do not depend on the specific basis or matrix chosen to represent the linear transformation.

Example 5.20. Consider the matrix $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$. Let $P = \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix}$ which means $P^{-1} = \begin{bmatrix} -4 & 9 \\ 1 & -2 \end{bmatrix}$. The matrix representation of T relative to S (the column vectors

of P) is the similar matrix

$$P^{-1}AP = \begin{bmatrix} -4 & 9 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 39 & 170 \\ -8 & -35 \end{bmatrix}.$$

Now consider the same matrix A, but pick a different matrix $P = \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}$ which means $P^{-1} = \dfrac{1}{3} \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$. The matrix representation of T relative to S (the column vectors of P) is the similar matrix

$$P^{-1}AP = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}.$$

Notice how this choice of basis resulted in a diagonal matrix. The next section shows how you can find the right matrix P so that A is similar to a diagonal matrix.

5.3.1 Diagonalizing a Matrix: Picking the "Best Basis"

Recall the following: we say that A and D are similar matrices if $D = P^{-1}AP$ for some matrix P. In this section, we will show that if we want D to be a diagonal matrix, then the columns of P must be linearly independent eigenvectors of A. This means that there is a basis S (the columns of P) so that the transformation T associated with A will have the very simple matrix $[T]_S = D$. If $D = P^{-1}AP$, then we say that the matrix P *diagonalizes* the matrix A.

Let A be an n by n matrix and suppose that $D = P^{-1}AP$ for some diagonal matrix $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, or in other words, $AP = PD$ for some invertible matrix $P = [\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n]$. Computing both AP and PD gives us

$$
\begin{aligned}
AP &= A[\vec{x}_1 \ \vec{x}_2 \ \vec{x}_3 \ \cdots \ \vec{x}_n] \\
&= [A\vec{x}_1 \ A\vec{x}_2 \ A\vec{x}_3 \ \cdots \ A\vec{x}_n], \text{ and} \\
PD &= [\vec{x}_1 \ \vec{x}_2 \ \vec{x}_3 \ \cdots \ \vec{x}_n] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\
&= [\lambda_1 \vec{x}_1 \ \lambda_2 \vec{x}_2 \ \lambda_3 \vec{x}_3 \ \cdots \ \lambda_n \vec{x}_n].
\end{aligned}
$$

Comparing columns between AP and PD, we see that $A\vec{x}_1 = \lambda_1 \vec{x}_1$, $A\vec{x}_2 = \lambda_2 \vec{x}_2$, and so on. This means that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A with corresponding eigenvectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$. Furthermore, since P is invertible, we know that the eigenvectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ must be linearly independent. Let's formalize this understanding in the following key theorem.

Theorem 5.9. A matrix A can be diagonalized (i.e., $D = P^{-1}AP$, where $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $P = [\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n]$) if and only if A has n linearly independent eigenvectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Example 5.21. Consider the linear transformation $T(x, y) = (x + 4y, 2x + 3y)$ with standard matrix representation $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$. The eigenvalues are 5 (with eigenvector $(1, 1)$) and -1 (with eigenvector $(-2, 1)$). Let the eigenvectors be a basis for the domain and range $\mathbb{R}^2$: $S = \{(1, 1), (-2, 1)\}$. Form the matrix $P = \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}$ whose columns are the eigenvectors. The inverse of P is $P^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$. The matrix representation

of T relative to S is

$$P^{-1}AP = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix},$$

a diagonal matrix. We have diagonalized A. In other words, we have found a basis S so that $[T]_S$ is a diagonal matrix.

Example 5.22. Let's now repeat with a 3 by 3 matrix. Consider the linear transformation defined by the matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The eigenvalues are 3 (with multiplicity 2) and 1. The eigenspace for the eigenvalue 3 (i.e., the nullspace of $A - 3I$) has dimension 2 and has a basis $\{(0,0,1), (1,1,0)\}$. The eigenspace for the eigenvalue 1 (i.e., the nullspace of $A - I$) is dimension 1 and has a basis $\{(-1,1,0)\}$ (verify each of these facts). So we see that we can get 3 linearly independent eigenvectors, two from the eigenspace for the eigenvalue 3 and one from the eigenspace for the eigenvalue 1: $S = \{(0,0,1), (1,1,0), (-1,1,0)\}$. S is then a basis for our domain and range $\mathbb{R}^3$. Form the matrix $P = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ whose columns are the eigenvectors. The inverse of P is $P^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$. The matrix representation of T relative to S is

$$[T]_S = P^{-1}AP = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

a diagonal matrix. Again, we have diagonalized A (or found a diagonal representation $[T]_S$ of the linear transformation T).

The next example shows that not every matrix can be diagonalized. Remember that in order to diagonalize a matrix, you must be able to find n linearly independent eigenvectors.

Example 5.23. Consider the linear transformation defined by the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The eigenvalues are 3 (multiplicity 2) and 1. This time, however, the eigenspace for the eigenvalue 3 has dimension only 1 (a basis is $\{(0,0,1)\}$). The eigenspace for the eigenvalue 1 is also dimension 1 (a basis is $\{(-1,1,0)\}$). So we only have two linearly independent eigenvectors for A, one from each eigenspace. In this case, there are not enough linearly independent eigenvectors to construct the matrix P and diagonalize this matrix.

5.3.2 Algebraic and Geometric Multiplicity

Is every n by n matrix diagonalizable? Does every n by n matrix have n linearly independent eigenvectors? Example 5.23 shows that the answer is no. What does it take for a matrix to not be diagonalizable? The matrix must not have enough linearly independent eigenvectors. We now make some definitions to simplify talking about this issue.

Definition 5.3. The *algebraic multiplicity* of an eigenvalue is the number of times that eigenvalues appears as a root of the characteristic polynomial.

The *geometric multiplicity* of an eigenvalue is the dimension of the corresponding eigenspace.

If the geometric multiplicity of an eigenvalue is less than the algebraic multiplicity, then we say the eigenvalue is *deficient* (there were not enough eigenvectors in the eigenspace).

For a sketch of a proof for the next theorem, see problem 11.46 in Schaum's.

Theorem 5.10. The geometric multiplicity is always less than or equal to the algebraic multiplicity.

Theorem 5.11. A matrix is diagonalizable if and only if the algebraic multiplicity of each eigenvalue equals the geometric multiplicity of each eigenvalue.

Theorem 5.12. If all the eigenvalues are distinct, then A is diagonalizable.

Example 5.24. Let's revisit examples 5.22 and 5.23 to compute the algebraic and geometric multiplicity of each eigenvalue. For both matrices, the characteristic polynomial has zeros $3, 3, 1$, which means that the algebraic multiplicity of $\lambda = 3$ is 2, and the algebraic multiplicity of $\lambda = 1$ is 1.

For example 5.22, a basis for the E_3 (the eigenspace corresponding to $\lambda = 3$) is $\{(0, 0, 1), (1, 1, 0)\}$, and a basis for E_1 is $\{(-1, 1, 0)\}$. The geometric multiplicity of $\lambda = 3$ is 2, and the geometric multiplicity of $\lambda = 1$ is 1. Because the geometric and algebraic multiplicities match, this matrix is diagonalizable.

For example 5.23, a basis for E_3 is $\{(0, 0, 1)\}$. The geometric multiplicity of $\lambda = 3$ is only 1, whereas the algebraic multiplicity is 2. The eigenvalue $\lambda = 3$ is deficient; it's missing an eigenvector. For this reason, the matrix is not diagonalizable.

5.3.3   Physical Applications of Diagonalization

Suppose we have a vector field that pushes objects straight outwards in the direction $(1, 1)$ with a force that is triple the distance from the origin, and pulls objects towards the origin in the direction $(-1, 1)$ with a force that is equal to the distance to the origin. In other words, our vector field has eigenvectors $(1, 1)$ and $(-1, 1)$ with corresponding eigenvalues 3 and -1. Is this enough information to determine the entire vector field? We seek a matrix A with eigenvectors $(1, 1)$ and $(-1, 1)$ with eigenvalues 3 and -1. The principle of diagonalization comes to our aid. Because A has two linearly independent eigenvectors, we know that it is diagonalizable where

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, P^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

We solve the equation $D = P^{-1}AP$ for A to obtain

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

We just obtained the matrix which represents our vector field. All we needed to know was the eigenvalues and eigenvectors, something than can be seen and measured physically in a vector field. So we now have the ability to take a physical moving force and use diagonalization to connect it back to matrices.

5.4   Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Explain how to describe vectors using different coordinate systems. Be able to change between different coordinate systems, and use these ideas to construct matrix representations of linear transformations.

2. Show that the null space (kernel), column space (image), and eigenspaces of a matrix (linear transformation) are vector subspaces. Be able to extend a basis for these spaces to a basis for the domain or codomain.

3. For square matrices, explain why similar matrices $B = P^{-1}AP$ represent the same linear transformation (just using a different basis). Explain why the determinant, eigenvalues, rank, and nullity of a linear transformation are not dependent upon the basis chosen.

4. Explain how to diagonalize a matrix, as well as explain when it is or is not possible.

Here are the preparation problems for this unit.

<div align="center">

Preparation Problems (click for handwritten solutions)

Day 1	Schaum's 6.3, Schaum's 6.6, Schaum's 6.12, Schaum's 6.16
Day 2	Schaum's 9.1, Schaum's 9.6, Schaum's 9.18, 2b
Day 3	Schaum's 11.13, Schaum's 11.17, Schaum's 11.18, Schaum's 11.19
Day 4	Lesson Plan, Quiz, Start Project

</div>

The following homework relates to what we are learning in this unit.

Concept	Where	Suggestions	Relevant Problems
Coordinates	Schaum's 6	1,3,4,6,9	1-10,22-28
Change of Basis	Schaum's 6	12,15,16,31	12-16,30-33
Matrix Representations (different bases)	Schaum's 9	16,18,19,44	16-20,42-45
Matrix Representations (same basis)	Schaum's 9	1,2,5,6,9,11	1-12,27-34,36,39
General Position	Here	1,2	1,2
Diagonalization	Schaum's 11	9,11,13,14,17,18	9,11-22,57,60-63
Visual illustrations	Here	TBA	TBA

Note that Schaum's does not use inverse matrices. Please take advantage of the fact that you can use inverse matrices to change coordinates. I have done by hand all the problems in Schaum's so that you can see how an inverse matrix yields the same solution. Please use these solutions. Also, remember that once you leave 2 by 2 matrices, please use your computer to find eigenvalues and inverse matrices. You should practice finding the eigenvectors of some of the 3 by 3 matrices by hand.

5.5   Problems

Most of the homework from this unit comes from Schaum's Oultines. Here are some additional problems related to the topics in this unit.

1. **Key Subspaces:** Show the following by checking the three requirements to be a subspace.

   (a) The kernel of a linear transformation is a subspace of the domain.

   (b) The image of a linear transformation is a subspace of the range.

(c) For each eigenvalue of a linear transformation, the corresponding eigenspace is a subspace of both the domain and range.

2. **General Position:** For each of the following linear transformations, find a basis S for the domain and a basis S' for the range so that the matrix representation relative to S and S' consists of all zeros and 1's, where the 1's occur on the diagonal of a sub matrix located in the upper left corner of the matrix. In other words, the matrix representation consists of adding rows and/or columns of zeros to the identity matrix (where the number of 1's equals the rank of the matrix).

(a) $T(x, y, z) = (x + 2y - z, y + z)$

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

(c) $T(x, y, z) = (x + 2y - z, 2x + 4y - 2z)$

(d) $T(x, y) = (x + 2y, y, 3x - y)$

(e) $T(x, y) = (x + 2y, 2x + 4y, 3x + 6y)$

(f) Now select your own linear transformation and repeat this problem.

3. Consider the linear transformation, vectors, and bases

$$T(x, y) = (x + 4y, 2x + 3y), \quad \vec{v} = (0, 5), \quad S = \{(1, 1), (-2, 0)\}, \quad S' = \{(1, 2), (3, 1)\}.$$

Find the following, and write the name given to this quantity:

(a) $[T]_E$ (b) $[\vec{v}]_S$ (c) $T(\vec{v})$ (d) $[T(\vec{v})]_{S'}$ (e) $[T]_{S,S'}$ (f) $[T]_{S',S}$ (g) $[T]_S$ (h) $[T]_{S'}$

(i) If $(a, b)_S = [(x, y)]_S$, find $[T(x, y)]_{S'}$ in terms of a and b.

(j) If $(a, b)_S = [(x, y)]_S$, find $[T(x, y)]_S$ in terms of a and b.

(k) Find a basis S that diagonalizes T. What is $[T]_S$ using this basis?

4. Consider the linear transformation $T(x, y, z) = (x + 2y - z, 2x + 4y + 4z, -x - 2y)$.

(a) Find a basis for the kernel of T.

(b) Find a basis S for the domain of T that includes the basis vectors for the kernel as the last vectors.

(c) What is $[T]_{S,E}$?

(d) Find a basis S' for the range of T that includes the nonzero columns of $[T]_{S,E}$ as the first vectors.

(e) What is $[T]_{S,S'}$? Does it have the identity matrix in the upper left corner as it should?

5. Which matrices are diagonalizable? Find the algebraic and geometric multiplicity of each eigenvalue and use this to answer the question. If the matrix is diagonalizable, what similarity matrix P diagonalizes the matrix?

(a) $\begin{bmatrix} 8 & -2 \\ 15 & -3 \end{bmatrix}$

(b) $\begin{bmatrix} 5 & -1 \\ 9 & -1 \end{bmatrix}$

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

(d) $\begin{bmatrix} 6 & -2 & 3 \\ 4 & 0 & 4 \\ -2 & 1 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

(f) $\begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

(g) $\begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

(h) $\begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$

5.6 Projects

1. This project explores the use of coordinates relative to a basis to send and receive radio communications. We'll start with a brief introduction to the idea, and then do some computations. Make sure you have a CAS like SAGE nearby so that you can rapidly perform the computations required.

   Radio waves form a vector space, which means the sum of two radio waves is again a radio wave, and radio waves can be scaled by any real number. We will represent radio waves in this example by using trigonometric functions. The telecommunications industry (radio, TV, internet, cell phone, etc.) utilizes these facts to send and receive messages. Here is the basis idea:

   - The government assigns each radio station a different frequency to broadcast at (they state a basis and require everyone to use this basis). We'll see what happens later if someone doesn't follow the rules.

   - The radio station first converts their message into numbers (vectors). They then creates and transmit radio waves whose amplitude matches the numbers being sent. (Here is where we use the ability to scale vectors by any amount.)

   - Every radio station simultaneously transmits radio waves at their frequency, so these radios waves add together. This total combined radio wave looks very different than any of the individually transmitted radio waves. (Here is where we use the ability to sum any collection of vectors.)

   - When you turn the dial on your radio tuner to FM 91.5, an electrical circuit in your radio called an "integrator" computes an integral (yes it actually computes an integral). This integral is a linear transformation that takes radio waves to numbers. The numbers obtained from this integral are the numbers sent by FM 91.5, or the coordinates of some vectors, relative to a basis. When you change the dial to a different frequency, then the integral returns the coordinates transmitted by the new station.

   The rest of this project asks you to explore the idea above, using some simple examples. First, we need a basis for the space of radio waves (the Federal Communications Commission (FCC) provides the basis for the telecommunications industry). We'll represent radio waves using sines and cosine functions, so let's use the simplified basis

   $$S = \{\cos(\pi x), \sin(\pi x), \cos(2\pi x), \sin(2\pi x), \cos(3\pi x), \sin(3\pi x), \cos(4\pi x), \sin(4\pi x), \dots.$$

   This is an infinite dimensional vector space of functions. When you purchase the right to broadcast, you purchase the right to use certain basis vectors. You basically are buying the rights to a subspace. You can then broadcast using any linear combination of these basis vectors.

   (a) Suppose we have two radio stations at frequencies 1 and 2. The radio station with frequency 1 wants to send the numbers $(2, -1)$, so they transmit the vector $v_1 = 2\cos(\pi x) - \sin(\pi x)$. The radio station with frequency 2 wants to send the numbers $(-3, 1)$, so they transmit the vector $v_2 = -3\cos(2\pi x) + 1\sin(2\pi x)$. Notice how both stations used the numbers they want to send as the coordinates of their vector. Use your CAS to plot v_1 and v_2 on the same set of axes for $-2 \le x \le 2$. In Sage you could type

   ```
   v1=2 *cos(pi*x)-sin(pi*x)
   v2=-3 *cos(2*pi*x)+sin(2*pi*x)
   p1=plot(v1,-2,2,color='red')
   p2=plot(v2,-2,2,color='blue')
   p1+p2
   ```

   (b) Add to your graph the radio waves from three more stations:

   $$v_3 = 7\cos(3\pi x) - 2\sin(3\pi x), v_4 = 0\cos(4\pi x) + 5\sin(4\pi x), v_5 = -2\cos(5\pi x) + 6\sin(5\pi x).$$

   (c) When all 5 stations broadcast simultaneously, their waves sum together. Graph the sum $v = v_1 + v_2 + v_3 + v_4 + v_5$ for for $-2 \le x \le 2$. Notice the graph is periodic. This is the wave that your antenna receives.

   (d) We now use our tuner to obtain the numbers sent by station 1. Use your CAS to compute the following two integrals

   $$\int_{-1}^{1} v\cos(\pi x)dx \quad \text{and} \quad \int_{-1}^{1} v\sin(\pi x)dx.$$

   You should obtain the numbers (2,-1). You can use the following Sage command:

```
integrate( (v)*cos(1*pi*x),(x,-1,1)).n()
```

The .n() gives you a decimal approximation (try removing it and see what you get).

(e) How do we obtain the numbers sent by station 2? Use your CAS to compute

$$\int_{-1}^{1} v \cos(2\pi x)dx \quad \text{and} \quad \int_{-1}^{1} v \sin(2\pi x)dx.$$

Did you obtain the coefficients (-3,1)? These are the coordinates of $v_2 = -3\cos(2\pi x) + 1\sin(2\pi x)$ relative to the basis vectors $\{\cos(2\pi x), \sin(2\pi x)\}$ purchased by station 2.

(f) In the previous two parts, we first multiplied our vector v by $\cos(n\pi x)$ or $\sin(n\pi x)$ where n is the frequency of the station, and then integrated from -1 to 1. Electrical circuits do all this inside the radio. Repeat these integrals for $n = 3, 4, 5$ to verify that you get the coordinates $(7, -2)$, $(0, 5)$, and $(-2, 6)$.

(g) As long as everyone broadcasts at their assigned frequency, mandated by the FCC, the integrals provide the correct coordinates. What happens if someone broadcasts at a frequency not approved by the FCC? Their radio wave can jam the other stations. Let's broadcast the radio wave $u = 20\cos(1.5\pi x) + 3\sin(1.5\pi x)$ and see what happens.

- Let $w = v + u = v_1 + v_2 + v_3 + v_4 + v_5 + u$ and graph w. Notice that it is no longer periodic.
- Compute $\int_{-1}^{1} w \cos(1\pi x)dx$ and $\int_{-1}^{1} w \sin(1\pi x)dx$, and compare to $(2, -1)$. Use a decimal approximation (so .n() in Sage).
- Compute $\int_{-1}^{1} w \cos(5\pi x)dx$ and $\int_{-1}^{1} w \sin(5\pi x)dx$, and compare to $(-2, 6)$.
- Notice the jamming frequency 1.5 caused a larger problem to frequency 1 than it did to frequency 5. You should have obtained $(17.279, 0.528)$ instead of $(2, -1)$ and $(-2.839, 5.581)$ instead of $(-2, 6)$.
- Change the jamming vector to $u = 20\cos(4.5\pi x) + 3\sin(4.5\pi x)$ and repeat the integrals above.

Jamming frequencies can greatly jam the frequencies close to them. This is one reason why you are asked to turn off your cell phone and laptop during takeoff on an air plane. You don't want cellular or wireless emissions to accidentally jam the air traffic control communications, resulting in two planes crashing on the runway.

Using a radio jammer is illegal without a permit, and could result in huge fines and/or prison time. The electronic devices we use must comply with the FCC rules which state:

> This device complies with Part 15 of the FCC rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference, and (2) this device must accept any interference received, including interference that may cause undesired operation.

You could interpret this as: (1) your device should not jam other communications, and (2) your device must allow others to jam it. The latter rule allows law enforcement to jam communications if needed.

(h) Why does all of this work? It has to do with the fact that certain integrals are zero. If you select any two different basis vectors f, g in S, and compute $\int_{-1}^{1} fg \, dx$, then the result is zero. In other words, we have $\int_{-1}^{1} \cos(m\pi x)\sin(n\pi x)dx = 0$, and provided $m \neq n$ we also have $\int_{-1}^{1} \cos(m\pi x)\cos(n\pi x)dx = 0$ and $\int_{-1}^{1} \sin(m\pi x)\sin(n\pi x)dx = 0$. These trig functions (vectors) are said to be orthogonal because this integral is zero. Your final job in this assignment is to compute a few of these integrals to verify this result.

- Compute $\int_{-1}^{1} \cos(2\pi x)\cos(n\pi x)dx$ for $n = 1, 2, 3$. One is nonzero.
- Compute $\int_{-1}^{1} \cos(2\pi x)\sin(n\pi x)dx$ for $n = 1, 2, 3$. All are zero.
- Compute $\int_{-1}^{1} \sin(3\pi x)\cos(n\pi x)dx$ for $n = 1, 2, 3$. All are zero.
- Compute $\int_{-1}^{1} \sin(3\pi x)\sin(n\pi x)dx$ for $n = 1, 2, 3$. One is nonzero.

As a challenge (optional), use the product-to-sum trig identities to prove the general cases. We'll be revisiting this topic when we study inner products. The integral $\int_{-1}^{1} fg\,dx$ is called an inner product. Because the nonzero integrals are 1, we say that S is an orthonormal basis.

5.7 Solutions

Use the links below to download handwritten solutions to the relevant problems in Schaum's. I have provided these solutions to illustrate how the matrix inverse provides the key tool to all the problems in the unit. Schaum's shows how these problems can be solved without an inverse. The solutions below emphasize the use of the inverse.

- Chapter 6 Solutions
- Chapter 9 Solutions
- Chapter 11 Solutions
- All Solutions

As time permits, I will add more solutions to the problems from this book.

Please use Sage to check all your work. As time permits, I will add in solutions to the problems I have done.

1. **Key Subspaces:** All 3 proofs are given in example 5.8.

2. **General Position:** Once you have choses B and B', you know you are correct if $B'^{-1}AB$ has the identity matrix in the upper left corner, with possible extra rows or columns of zeros at the bottom or right. See problem 4 for a step-by-step way to do each of these.

3. (a) $[T]_E = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ is the standard matrix representation or T (relative to the standard bases).

   (b) $[\vec{v}]_S = (7, 7/2)$ is the coordinates of $\vec{v}$ relative to S.

   (c) $T(\vec{v}) = (28, 21)$ is the image of $\vec{v}$.

   (d) $[T(\vec{v})]_{S'} = (7, 7)$ is the coordinates of $T(\vec{v})$ relative to S'.

   (e) $[T]_{S,S'} = \begin{bmatrix} 2 & -2 \\ 1 & 0 \end{bmatrix}$ is the matrix representation of T relative to S and S'.

   (f) $[T]_{S',S} = \begin{bmatrix} 8 & 9 \\ -\frac{1}{2} & 1 \end{bmatrix}$ is the matrix representation of T relative to S' and S.

   (g) $[T]_S = \begin{bmatrix} 5 & -4 \\ 0 & -1 \end{bmatrix}$ is the matrix representation of T relative to S.

   (h) $[T]_{S'} = \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix}$ is the matrix representation of T relative to S'.

   (i) Interpret $(a, b)_S = [(x, y)]_S$ as "the vector (a, b) is the coordinates of (x, y) relative to S". To find the coordinates of $T(x, y)$ relative to S', we compute $[T(x, y)]_{S'} = [T]_{S,S'}[(x, y)]_S = \begin{bmatrix} 2 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = (2a - 2b, a)_{S'}$.

   (j) Interpret $(a, b)_S = [(x, y)]_S$ as "the vector (a, b) is the coordinates of (x, y) relative to S". To find the coordinates of $T(x, y)$ relative to S, we compute $[T(x, y)]_S = [T]_S[(x, y)]_S = \begin{bmatrix} 5 & -4 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = (5a - 4b, -b)_S$.

   (k) You just find the eigenvalues $\lambda = 5, -1$ and eigenvectors $S = \{(1, 1), (-2, 1)\}$. With this order $[T]_S = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$. If you reverse the order of $S = \{(-2, 1), (1, 1)\}$, then $[T]_S = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$.

4. (a) $\{(-2, 1, 0)\}$

   (b) $S = \{(1, 0, 0), (0, 0, 1), (-2, 1, 0)\}$

   (c) The matrix representation of T relative to S and E is $[T]_{S,E} = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 4 & 0 \\ -1 & 0 & 0 \end{bmatrix}$.

   (d) $S' = \{(1, 2, -1), (-1, 4, 0), (1, 0, 0)\}$.

   (e) The matrix representation of T relative to S and S' is $[T]_{S,S'} =$
   $$\begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & 0 \\ -1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & 4 \\ -1 & -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} =$$
   $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

5. (a) Yes. $\lambda = 2$ has algebraic and geometric multiplicity equal to 1. So does $\lambda = 3$. One possible choice of matrices is $P = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

   (b) No. $\lambda = 2$ has algebraic multiplicity 2, but only 1 corresponding eigenvector in a basis for the eigenspace so geometric multiplicity is 1.

   (c) For $\lambda = 2$ both algebraic and geometric multiplicities are 2. For $\lambda = 3$ the geometric and algebraic multiplicities are both 1. Hence it is diagonalizable. $P = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

   (d) No. The algebraic multiplicity of $\lambda = 2$ is 2, while the geometric multiplicity is 1.

(e) No. The algebraic multiplicity of $\lambda = 2$ is 3, while the geometric multiplicity is 2.

(f) No. The algebraic multiplicity of $\lambda = 2$ is 3, while the geometric multiplicity is 2.

(g) No. The algebraic multiplicity of $\lambda = 2$ is 3, while the geometric multiplicity is 2.

(h) No. The algebraic multiplicity of $\lambda = 2$ is 3, while the geometric multiplicity is 1.

Chapter 6

Inner Products

This chapter covers the following ideas.

1. Explain the dot product and inner product. Use them to define length and angles of vectors. Normalize vectors, obtain vectors of a given length in a given direction, and explain how to tell if two vectors are orthogonal.

2. Obtain the orthogonal complement of a set of vectors (relate it to the null space of the transpose). Use the orthogonal complement to project vectors onto vector spaces.

3. By considering the projection of a vector onto a vector subspace, explain why solving $A^T A \vec{x} = A^T \vec{b}$ gives the solution to the least squares regression problem. This is why statistics works.

4. Use the Gram-Schmidt orthogonalization process to obtain an orthonormal basis of vectors. Show how to compute components of vectors relative to an orthogonal basis (called Fourier coefficients).

5. Illustrate by examples that when a matrix is symmetric, eigenvectors corresponding to different eigenvalues are orthogonal, which means you can find an orthonormal basis to diagonalize the matrix.

6.1   The Dot Product

Recall the length of a vector $\vec{u} = (u_1, u_2)$ in $\mathbb{R}^2$ is (from the Pythagorean theorem)

$$\|\vec{u}\| = |\vec{u}| = \sqrt{u_1^2 + u_2^2}.$$

The notation $|\vec{u}|$ and $\|\vec{u}\|$ are both used to denote the length or **norm** or a vector, depending on the context and author. Repeated use of the Pythagorean theorem gives

The norm of a vector is a synonym for the magnitude or length.

$$|(u_1, u_2, u_3)| = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$

It seems reasonable to generalize this to $\mathbb{R}^n$ by letting

$$|\vec{u}| = |(u_1, \ldots, u_n)| = \sqrt{u_1^2 + \cdots u_n^2}.$$

Now that we have spent a semester looking for matrix patterns, notice that the length can be written as a matrix product $\sqrt{\vec{u}^T \vec{u}}$ So the square of the length of a vector is precisely the quantity

$$\|\vec{u}\|^2 = \vec{u}^T \vec{u}.$$

Mathematicians discovered that sometimes the quantity $\vec{u}^T \vec{v}$ appears where $\vec{u}$ and $\vec{v}$ are different vectors. This quantity appeared so often that they gave it a name, the dot product. We'll see some applications in just a bit.

Definition 6.1. Given vectors $\vec{u} = (u_1, u_2, \ldots, u_n)$ and $\vec{v} = (v_1, v_2, \ldots, v_n)$, define the dot product $\vec{u} \cdot \vec{v}$ to be the scalar

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = u_1 v_1 + u_2 v_2 + \cdots u_n v_n = \sum_{i=1}^{n} u_i v_i.$$

If $\vec{u} = \vec{v}$, then the dot product $\vec{u} \cdot \vec{u} = \|\vec{u}\|^2$ equals the square of the length of the vector. If the length of a vector is 1, we say the vector is a **unit vector**. The standard basis vectors for $\mathbb{R}^n$ are unit vectors. If a vector is not a unit vector, we can make it a unit vector by dividing by its length. Unit vectors give us a way to find vectors of any length in a given direction, all we have to do is multiply a unit vector by the length we want. As result, we can write every vector as a product of its norm (length) and a unit vector in the same direction, i.e. we will write

Unit vectors have length 1.

$$\vec{v} = c\vec{u} \quad \text{where} \quad c = |\vec{v}| \quad \text{and} \quad \vec{u} = \frac{\vec{v}}{|\vec{v}|} \quad \text{is a unit vector.}$$

Much of what we do in this unit will involve decomposing a vectors into the product of its length and direction.

Example 6.1. Let's find a vector of length 7 in the direction of $\vec{v} = (1, 3, -2)$.

The vector $\vec{v} = (1, 3, -2)$ has length $|\vec{v}| = \sqrt{1 + 9 + 4} = \sqrt{14}$, so a unit vector in the same direction as $(1, 3, -2)$ is

$$\vec{u} = \frac{\vec{v}}{|\vec{v}|} = \frac{1}{\sqrt{14}}(1, 3, -2) = \left(\frac{1}{\sqrt{14}}, \frac{3}{\sqrt{14}}, -\frac{2}{\sqrt{14}} \right).$$

Since the vector $\vec{u}$ is 1 unit long (its a unit vector), we multiply by 7 to obtain a vector 7 units long in the same direction. A vector 7 units long in the same direction as $(1, 3, -2)$ is

$$\vec{w} = 7\vec{u} = (7) \left(\frac{1}{\sqrt{14}}(1, 3, -2) \right) = \left(\frac{7}{\sqrt{14}}, \frac{21}{\sqrt{14}}, -\frac{14}{\sqrt{14}} \right).$$

6.1.1 Properties of the Dot Product

Before proceeding to the definition of an inner product, let's pinpoint some properties of the dot product. We'll be using these properties to define inner products on abstract vector spaces.

Theorem 6.1. The dot product $\vec{u} \cdot \vec{v}$ on $\mathbb{R}^n$ satisfies the following properties:

1. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$ (The dot product preserves vector addition.)

2. $(a\vec{u}) \cdot \vec{v} = a(\vec{u} \cdot \vec{v})$ (The dot product preserves scalar multiplication.)

3. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ (The dot product is commutative)

4. $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$ (The dot product is positive definite.)

The first two properties imply that the dot product is linear. You can pull out constants and perform the dot product either before or after you multiply. The third properties shows that it does not matter in which order you perform the dot product. In particular, this implies that the first two properties apply to the right side of the dot product (for example $\vec{u} \cdot (b\vec{v}) = b(\vec{u} \cdot \vec{v})$). The last property just requires that the only way to get out zero when dotting a vector with itself is to use the zero vector. Essentially, we need this property to show that length $|\vec{u}| = \sqrt{\vec{u} \cdot \vec{u}} = 0$ is well defined (as you can't take the square root of negative numbers) and that the only vector whose length is zero is the zero vector itself.

The dot product provides a nice way to compute angles between vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$. If θ is the angle between 2 vectors $\vec{u}$ and $\vec{v}$, then we can find the angle between these two vectors using

The dot product finds angles:

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$$

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta.$$

This follows from the law of cosines $c^2 = a^2 + b^2 - 2ab \cos \theta$, where $a = |\vec{u}|$, $b = |\vec{v}|$, and $c = |\vec{u} - \vec{v}|$. Because the dot product is equal to the square of the magnitude, we have

$$(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}| \cos \theta.$$

Distributing on the left gives

$$\vec{u} \cdot \vec{u} - 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}| \cos \theta.$$

Now cancel the common terms on both sides and divide by 2 to obtain our formula $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$. This formula geometrically explains how to connect the dot product to angles in both 2 and 3 dimensions. In higher dimensions we now define the angle between two vectors using this formula.

Law of Cosines

$$c^2 = a^2 + b^2 - 2ab \cos \theta$$

Definition 6.2: Angle between two vectors. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. We define the angle θ between $\vec{u}$ and $\vec{v}$ to be the unique angle between 0 and 180 degrees such that

$$cos\theta = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|} = \frac{\vec{u} \cdot \vec{v}}{\sqrt{\vec{u} \cdot \vec{u}}\sqrt{\vec{v} \cdot \vec{v}}}.$$

We can define both lengths and angles entirely in terms of the dot product.

In order to make this generalization, we need to know that

$$-1 \leq \frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|} \leq 1$$

so that it will always be the cosine of some angle. This is equivalent to showing $|\vec{u} \cdot \vec{v}| \leq |\vec{u}||\vec{v}|$, which is called the Cauchy-Schwartz inequality. A proof of the Cauchy Swartz inequality is on page 275 (problem 7.35) in Schaum's.

When two vectors meet at a 90 degree angle, then $\cos 90° = 0$. This means that the dot product is zero since

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos 90 = |\vec{u}||\vec{v}|0 = 0.$$

If the dot product is zero, then we know $0 = \vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$. Since $|\vec{u}||\vec{v}| \cos \theta = 0$, either one of the vectors has zero length or the cosine of the angle between them is zero (hence the angle is 90 degrees). The word perpendicular suggests that two objects meet at a 90 degree angle. If the dot product of two vectors is zero, then they may meet perpendicularly, or one of the vectors might be zero. We create a new word "orthogonal" to mean precisely this: the vectors meet at a 90 degree angle or one of the vectors is zero.

Definition 6.3. We say that $\vec{u}$ and $\vec{v}$ are orthogonal if their dot product is zero.

orthogonal = dot product is zero

Example 6.2. The dot product of the two vectors $\vec{u} = (1, 0, 3, -4)$ and $\vec{v} = (3, -1, 5, 0)$ is

$$\vec{u} \cdot \vec{v} = (1)(3) + (0)(-2) + (3)(5) + (-4)(0) = 18.$$

The length of each vector is

$$|\vec{u}| = \sqrt{9 + 1 + 16} = \sqrt{26} \quad \text{and} \quad |\vec{v}| = \sqrt{9 + 1 + 25} = \sqrt{35}.$$

The angle between $\vec{u}$ and $\vec{v}$ is

$$\theta = \cos^{-1} \left(\frac{18}{\sqrt{26}\sqrt{35}} \right),$$

which you can obtain in radians or degrees using a calculator.

The dot product of the two vectors $\vec{u} = (-2, 1, 3)$ and $\vec{v} = (1, 5, -1)$ is $-2 + 5 - 3 = 0$, so $\cos \theta = 0$ (without finding the lengths). This means that the angle between the vectors is 90 degrees, and the vectors are orthogonal.

When the dot product of two vectors is zero, we say the vectors are orthogonal. Provided neither vector is zero, this means that the angle between the two vectors is 90 degrees. In summary, dot product helps us

1. Find the length of a vector: $\vec{u} \cdot \vec{u} = |\vec{u}|^2$,

2. Find the angle between two vectors: $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$,

3. Determine if two vectors are orthogonal: $\vec{u} \cdot \vec{v} = 0$.

6.2 The Inner Product - A General Dot product

In the last section we showed that the dot product is really just the matrix product

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}.$$

This product satisfies the following properties:

1. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$ (The dot product preserves vector addition.)

2. $(a\vec{u}) \cdot \vec{v} = a(\vec{u} \cdot \vec{v})$ (The dot product preserves scalar multiplication.)

3. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ (The dot product is commutative)

4. $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$ (The dot product is positive definite.)

We also use the dot product to obtain lengths using $|\vec{u}|^2 = \vec{u} \cdot \vec{u}$ and angles using $|\vec{u}||\vec{v}| \cos \theta = \vec{u}\vec{v}$. All of the computations are done in the vector space $\mathbb{R}^n$. We now ask, is there a similar kind of product in other vector spaces, so that we can talk about lengths and angles between abstract vectors in any vector space? This question led to the definition of an inner product.

Definition 6.4. An (real) **inner product** on a vector space V is a function (written $\langle \vec{u}, \vec{v} \rangle$) which takes two vectors $\vec{u}$ and $\vec{v}$ and returns a numbers $\langle \vec{u}, \vec{v} \rangle$. This function satisfies the following axioms:

1. $\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$ (The inner product preserves vector addition.)

2. $\langle a\vec{u}, \vec{v} \rangle = a \langle \vec{u}, \vec{v} \rangle$ (The inner product preserves scalar multiplication.)

3. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$ (The inner product is commutative)

4. $\langle \vec{u}, \vec{u} \rangle \geq 0$ and $\langle \vec{u}, \vec{u} \rangle = 0$ if and only if $\vec{u} = \vec{0}$ (The inner product is positive definite.)

A vector space together with an inner product is called an **inner product space**.

The axioms above allow us to treat an inner product in the same way we treat the dot product. The first two properties imply that the inner product is linear, which basically means we can distribute as in

$$\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = a \langle \vec{u}, \vec{w} \rangle + b \langle \vec{v}, \vec{w} \rangle.$$

If $\langle \vec{u}, \vec{v} \rangle$ is an inner product on a vector space V, then relative to this inner product we can

1. Find the norm (length) of a vector: $\langle \vec{u}, \vec{u} \rangle = \|\vec{u}\|^2$,

2. Find the angle between two vectors :$\langle \vec{u}, \vec{v} \rangle = \|\vec{u}\|\|\vec{v}\| \cos \theta$,

3. Determine if two vectors are orthogonal: $\langle \vec{u}, \vec{v} \rangle = 0$.

When working with an abstract inner product, we normally use double bars $\|\vec{u}\|$ instead of single bar $|\vec{u}|$ to talk about the norm of a vector.

In order to define angles, we need to know that $-1 \leq \cos \theta \leq 1$, or $-1 \leq \dfrac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \ \|\vec{v}\|} \leq 1$. The Cauchy-Swartz inequality (pg 247) verifies that $|\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\|\|\vec{v}\|$, which provides the required justification. Always remember: the inner product is just a generalization of the dot product which allows us to work with vector spaces and define lengths and angles just as we did with the dot product. It's time for some examples of inner products.

Example 6.3. The dot product is itself an inner product, as it satisfies the 4 axioms. We will often call the usual inner product on $\mathbb{R}^n$.

Example 6.4. The Federal Communications Commission (FCC) creates regulations about the types of signals that can be broadcast. Because of these regulations, signals form a vector space (a vector space of periodic function). In this example, we'll show how to define inner products on a vector space of functions.

Consider the vector space of continuous functions on the interval $[0, 1]$. Define an inner product on this space by $\langle f, g \rangle = \int_0^1 fg\,dx$. If you think of a function as an infinitely long list of numbers, then essentially this inner product requires that you multiply each component $f(x)$ by its corresponding component $g(x)$ and then add up the values from 0 to 1 using an integral. It's basically the same as the dot product, just now we have infinitely many values to sum so we use an integral instead of a sum.

We will now show that the inner product $\langle f, g \rangle = \int_0^1 fg\,dx$ is an inner product by showing it satisfies the axioms.

1. We need to show that you can add either before or after computing the inner product (we say the inner product distributes across addition or it preserves addition). We compute

$$\langle f + g, h \rangle = \int_0^1 (f + g)h\,dx = \int_0^1 (fh + gh)\,dx$$
$$= \int_0^1 fh\,dx + \int_0^1 gh\,dx = \langle f, h \rangle + \langle g, h \rangle.$$

   Since $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$, this product preserves vector addition.

2. We need to show that you can multiply by a scalar either before or after computing the inner product (we say the inner product distributes across scalar multiplication or it preserves scalar multiplication). We compute

$$\langle af, g \rangle = \int_0^1 (af)g\,dx = a \int_0^1 fg\,dx = a \langle f, g \rangle.$$

   Since $\langle af, g \rangle = a \langle f, g \rangle$, this product preserves scalar multiplication.

3. We need to show that you can interchange the order of the inner product without changing the value (we say the inner product is commutative). We compute

$$\langle f, g \rangle = \int_0^1 fg\,dx = \int_0^1 gf\,dx = \langle g, f \rangle$$

   Since $\langle f, g \rangle = \langle g, f \rangle$, this product is commutative.

4. We need to show that the inner product of a nonzero vector with itself always returns a positive number (so that we can use the inner product to define lengths). We also need to show that the inner product of zero with itself is zero. If both these are satisfied, we say the inner product is positive definite. We compute $\langle f, f \rangle = \int_0^1 f^2 dx$. Because $f^2 \geq 0$, we know that the area under f^2 cannot be negative, so $\int_0^1 f^2 dx \geq 0$. If $f \neq 0$ somewhere, then there will be some small interval on which f is positive (because f is continuous). This means f^2 will have a nonzero amount of area under it. Hence $\langle f, f \rangle = 0$ if and only if $f = 0$ Because $\langle f, f \rangle = \int_0^1 f^2 dx \geq 0$ and equals zero only when $f(x) = 0$, this product is positive definite.

We have shown that the inner product $\langle f, g \rangle = \int_0^1 fg\,dx$ is an inner product on the vector space of continuous functions defined on $[0, 1]$. This inner product is closely related to the inner products used by the FCC to establish radio communication. It is also related to data compression of JPG, MP3, and other electronic formats.

The previous example showed how to define an inner product on an infinite dimensional vector space. Let's now take a look at alternate ways to define inner products on $\mathbb{R}^n$, using symmetric matrices. To introduce this example, recall that the dot product is really just the matrix product

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = \vec{u}^T I \vec{v},$$

where I is the identity matrix inserted between u^T and v. As mathematicians solved various problems, they discovered a pattern that if A is a symmetric matrix (so $A^T = A$) whose eigenvalues are positive, then $\vec{u}^T A \vec{v}$ created an inner product. After further work, they then showed that if $\langle \vec{u}, \vec{v} \rangle$ is an inner product on a finite dimensional vector space, then relative to some basis S you can always find a symmetric matrix A such that

$$\langle \vec{u}, \vec{v} \rangle = [\vec{u}]_S^T A [\vec{v}]_S$$

. In other words, when dealing with finite dimensional vector spaces, symmetric matrices produce inner products, and inner products produce symmetric matrices. Studying symmetric matrices is the same as studying inner products.

Example 6.5. Let A be a symmetric matrix which satisfies

$$\vec{x}^T A \vec{x} > 0 \quad \text{for} \quad \vec{x} \neq 0.$$

Define a new inner product on $\mathbb{R}^n$ by the formula $\langle \vec{u}, \vec{v} \rangle = \vec{u}^T A \vec{v}$. Let's show that this provides an inner product on $\mathbb{R}^n$.

1. We compute

$$\begin{aligned}
\langle \vec{u} + \vec{v}, \vec{w} \rangle &= (\vec{u} + \vec{v}) A \vec{w} \\
&= \vec{u} A \vec{w} + \vec{v} A \vec{w} \\
&= \langle \vec{u}, \vec{w} \rangle + \langle \vec{u}, \vec{w} \rangle
\end{aligned}$$

   which shows it preserves addition.

2. We compute

$$\begin{aligned}
\langle a\vec{u}, \vec{v} \rangle &= (a\vec{u}) A \vec{v} \\
&= a\vec{u} A \vec{v} \\
&= a \langle \vec{u}, \vec{v} \rangle
\end{aligned}$$

   which shows it preserves scalar multiplication.

3. Recall that the inner product produces a number, so $\vec{u}^T A \vec{v}$ is a number. We compute both

$$\langle \vec{u}, \vec{v} \rangle = u^T A \vec{v} \quad \text{and} \quad \langle \vec{v}, \vec{u} \rangle = v^T A \vec{u}$$

   The transpose of a number is itself, so $\vec{u}^T A \vec{v}$ is the same as its transpose $(\vec{u}^T A \vec{v})^T = \vec{v}^T A^T \vec{u}$ (remember you reverse the order and transpose each matrix). Since A is symmetric, we can replace A^T with A. This means

$$\langle \vec{u}, \vec{v} \rangle = \vec{v}^T A^T \vec{u} = \vec{v}^T A \vec{v} = \langle \vec{v}, \vec{u} \rangle .$$

   This shows that the product is commutative.

4. We are assuming $\vec{x}^T A \vec{x} > 0$ for $\vec{x} \neq 0$. So it is immediately positive definite. We say a symmetric matrix is positive definite if it has this property.

In the example above, we assumed that the matrix A satisfied $\vec{x} A \vec{x} > 0$ for all nonzero vectors $\vec{x}$. This property is precisely the property needed to satisfy the last axiom of an inner product. We now state a theorem that will help us connect eigenvalues to this property.

Theorem 6.2. A symmetric matrix A satisfies the property $\vec{x}^T A \vec{x} > 0$ if and only if the eigenvalues of A are all positive. If A is a symmetric matrix with all positive eigenvalues, then we say A is a positive definite matrix.

Let's give a partial explanation of why this theorem works. Start by finding the eigenvalues of A and a set of linearly independent eigenvectors $S = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$. For each eigenvector $\vec{x}_i$ corresponding to λ_i, we compute (remember you can replace matrix multiplication with scalar multiplication)

$$\vec{x}_1^T A \vec{x}_1 = \vec{x}_1^T \lambda_i \vec{x}_1 = \lambda_i \vec{x}_i^T \vec{x}_i = \lambda_i \vec{x}_i \cdot \vec{x}_i$$

where we know $\vec{x}_i \cdot \vec{x}_i > 0$ from the usual inner product (dot product). We've shown

$$\vec{x}_1^T A \vec{x}_1 = \lambda_i \vec{x}_i \cdot \vec{x}_i > 0$$

only if all the eigenvalues are positive. If one of the eigenvalues were zero or negative, then for a corresponding eigenvector $\vec{x}$, we know that $\vec{x} A \vec{x}$ is zero or negative. We have now shown why the eigenvalues must be positive. To prove the reverse implication (that all positive eigenvalues implies $\vec{x} A \vec{x} > 0$ for all nonzero $\vec{x}$, including vectors other than eigenvectors) requires considerable more work. We'll postpone that argument for now.

The usual inner product (the dot product) is obtain by letting $A = I$, the identity matrix. All the eigenvalues of the identity matrix are 1, so $\vec{u} I \vec{v}$ defines an inner product on $\mathbb{R}^n$. If you want to show that a product on a finite dimensional vector space is an inner product, one approach is to show that it is represented by a symmetric matrix with positive eigenvalues. Let's now illustrate this with an example.

Example 6.6. Consider the product on $\mathbb{R}^2$ defined by

$$\langle (u_1, u_2), (v_1, v_2) \rangle = 3u_1 v_1 - 1u_1 v_2 - 1u_2 v_1 + 3u_2 v_2.$$

Does this define an inner product on $\mathbb{R}^2$.

We can answer this question in 2 ways. We could show that this product satisfies the axioms to be an inner product, or we can find a symmetric matrix A with positive eigenvalues such that

$$3u_1 v_1 - 1u_1 v_2 - 1u_2 v_1 + 3u_2 v_2 = \begin{bmatrix} u_1 & u_2 \end{bmatrix} A \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Let's try the latter approach. We write $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and then compute

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{bmatrix} = \begin{bmatrix} au_1 v_1 + bu_1 v_2 + cu_2 v_1 & du_2 v_2 \end{bmatrix}.$$

So letting $a = 3, b = -1, c = -1, d = 3$ allows us to write

$$\langle (u_1, u_2), (v_1, v_2) \rangle = 3u_1 v_1 - 1u_2 v_1 + 3u_2 v_2 = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

This matrix is symmetric, so we now just need to find the eigenvalues. We compute $(3 - \lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2) = 0$, and so the eigenvalues (4 and 2) are both positive. We know that A defines an inner product.

This is an alternate approach to showing a formula is an inner product, but it only works on finite dimensional vector spaces. .

As a final example, let's look at the vector space M_{mn} of m by n matrices. Define an inner product on M_{mn} by computing

$$\langle A, B \rangle = \text{trace}(B^T A).$$

We use the transpose of B so that the product $B^T A$ makes sense. Recall that the trace is the sum of the diagonal entries. Because we want the trace of $B^T A$, we need to compute the diagonal entries of $C = B^T$. For each j, to compute C_{jj} we take row j of B^T and dot it with column j of A. This is the same thing as dotting column j of B with column j of A, so we have

$$C_{jj} = \vec{a}_j \cdot \vec{b}_j = \sum_{i=1}^{n} a_{ij} b_{ij}.$$

The trace is $C_{11} + C_{22} + \cdots + C_{nn} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} b_{ij}$, which is the same thing as multiplying all corresponding entries together and then summing the result. In other words, if you just rewrite the matrices as big long vectors in $\mathbb{R}^{mn}$, then you use the usual dot product of $\mathbb{R}^{mn}$ to define an inner product on M_{mn}. Because this is equivalent to the dot product on $\mathbb{R}^{mn}$, we know it satisfies the axioms to be an inner product.

Example 6.7. To find the inner product of $A = \begin{bmatrix} 2 & 3 & 0 \\ -1 & 4 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 0 & 5 \\ 4 & 1 & 1 \end{bmatrix}$ we compute

$$B^T A = \begin{bmatrix} 2 & 3 & 0 \\ -1 & 4 & 2 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 0 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} -6 & 13 & 8 \\ -1 & 4 & 2 \\ 9 & 19 & 2 \end{bmatrix}.$$

The trace of this matrix is $-6 + 4 + 2 = 0$. This means that these two matrices are orthogonal.

Alternatively, we could rewrite the matrices as vectors in $\mathbb{R}^6$ and then use the usual inner product to obtain

$$(2, 3, 0, -1, 4, 2) \cdot (-1, 4, 0, 1, 5, 1) = (2)(-1) + (3)(0) + (0)(5) + (-1)(4) + (4)(1) + (2)(1) = -2 - 4 + 4 + 2 = 0.$$

See page 247 in Schaum's for related examples.

6.3 Orthogonal Complements

Recall that two vectors in an inner product space are said to be orthogonal if their inner product is zero, $\langle \vec{u}, \vec{v} \rangle = 0$. Given a single vector $\vec{u} = (u_1, \ldots, u_n)$ in $\mathbb{R}^n$, a vector $\vec{v} = (v_1, \ldots, \vec{v}_n)$ is orthogonal to $\vec{u}$ if and only if

$$u_1(v_1) + \cdots + u_n(v_n) = 0.$$

If we think of the elements of $\vec{u}$ as coefficients of a single equation and the elements of $\vec{v}$ as variables, then $\vec{v}$ is orthogonal to $\vec{u}$ if and only if $\vec{v}$ is in the null space of the 1 by n matrix $\vec{u}^T$ (where we think of $\vec{u}$ as a column matrix).

The set of all vectors orthogonal to $\vec{u}$ is the null space of $\vec{u}^T$.

If we want to find all vectors that are orthogonal to both $\vec{u}_1$ and $\vec{u}_2$, we just find the null space of $\begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}^T$. We now make a crucial definition.

Definition 6.5. The orthogonal complement of a set of vectors S in an inner product space V, denoted by $S^\perp$, is the set of all vectors that are orthogonal to every vector in S. In set notation, we write

$$S^\perp = \{\vec{v} \in V \mid \langle \vec{u}, \vec{v} \rangle \text{ for all } \vec{u} \in S\}.$$

Theorem 6.3. Let S be a set of vectors, whose span is the vector space V. The orthogonal complement of S is a vector space. The orthogonal complement of S is the same as the orthogonal complement of V. In particular, if A is any matrix whose columns span V, then the orthogonal complement of S is precisely the null space of A^T. In symbols we write

$$S^\perp = \text{nullspace}(A^T).$$

If the set of vectors S is finite, then we can find the orthogonal complement by just finding the null space of the transpose of the matrix whose columns are the vectors in S. If the vectors are dependent, then row reduction will eliminate the dependent vectors. If the set of vectors is infinite (but we are in a finite vector space), then we can find a basis for the span of the vectors. Then $S^\perp$ is precisely the null space of the transpose of the matrix whose columns are these basis vectors. We will mostly be working with a finite set of vectors.

Example 6.8. Consider the set $S = \{(1, 2, -4)\}$ composed of a single vector. Let's find the orthogonal complement. Let A the matrix whose only column is the vector in S. The reduced row echelon form of A^T is $\begin{bmatrix} 1 & 2 & -4 \end{bmatrix}$. A basis for the null space is $\{(-2, 1, 0), (4, 0, 1)\}$, which means that so $S^\perp = \text{span}\{(-2, 1, 0), (4, 0, 1)\}$.

Notice that S contained only one vector, hence the subspace $V = \text{span}\{(1, 2, -4)\}$ is represented by a line through the origin. The orthogonal complement of a line should be a plane, as there is an entire plane full of vector that are orthogonal to the line. Because the orthogonal complement had 2 vectors in a basis, it is a vector subspace in $\mathbb{R}^3$ represented by a two dimensional plane. This is precisely the plane which meets $(1, 2, -4)$ at a 90 degree angle.

Example 6.9. Consider the set $S = \{(1, 0, 3), (0, 2, 4)\}$. Let's find the orthogonal complement of S. Let A the matrix whose columns are the vectors in S. The reduced row echelon form of A^T is $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$. A basis for the null space is $\{(-3, -2, 1)\}$, so $S^\perp = \text{span}\{(-3, -2, 1)\}$, the one dimensional line which meets the plane spanned by the given vectors at a 90 degree angle.

Why do we use the words "orthogonal complement"? The word "orthogonal refers to the fact that every vector in $S^\perp$ is orthogonal to every vector in S. To understand the word "complement", recall that $40°$ and $50°$ are complementary angles because their sum is complete ($90°$). The word "complement" refers to the fact using vectors in S and $S^\perp$, we can obtain a "complete" basis for all of $\mathbb{R}^n$. The orthogonal complement adds to S enough vectors to form a basis for $\mathbb{R}^n$, it provides a way to extend S to a basis for all of $\mathbb{R}^n$.

6.3.1 The Cross Product

When working with vectors in $\mathbb{R}^3$, there is a special product, called the cross product, that can help you find a vector that is orthogonal to two other vectors. The cross product of two vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ is a new vector that is orthogonal to both $\vec{u}$ and $\vec{v}$, and it is given by the formula

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2, -(u_1 v_3 - u_3 v_2), u_1 v_2 - u_2 v_1) = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix},$$

where $\vec{i} = (1, 0, 0), \vec{j} = (0, 1, 0), \vec{k} = (0, 0, 1)$. The cross product is extremely useful when focusing on computations in $\mathbb{R}^3$. Here are some properties of the cross product. You will spend more time with the cross product in other classes where the cross product is needed. Remember that the cross product is just a specific vector in the orthogonal complement of $\vec{u}$ and $\vec{v}$.

1. The cross product of two vectors is a new vector which is orthogonal to both $\vec{u}$ and $\vec{v}$. This can be verified by performing the dot products $\vec{u} \cdot (\vec{u} \times \vec{v})$ and $\vec{v} \cdot (\vec{u} \times \vec{v})$. Hence, the cross product is a vector in the orthogonal complement of $\vec{u}$ and $\vec{v}$. The cross product is a special vector found in the orthogonal complement of a plane.

2. The magnitude of the cross product is $|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}| \sin \theta$, which is very similar to the dot product $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$. Using this formula, the length of the cross product is the area of the parallelogram formed using the two vectors $\vec{u}$ and $\vec{v}$.

3. Since the cross product is orthogonal to both $\vec{u}$ and $\vec{v}$, it must point in one of two directions. The direction of the cross product is found using the right hand rule. If you place your index finger of your right hand on the first vector $\vec{u}$, and then point your middle finger of your right hand in the direction of $\vec{v}$, then the direction of $\vec{u} \times \vec{v}$ is in the direction of your thumb.

4. Note that $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$. Because of this, we say the cross product is anti-commutative, so be careful not to switch the order on the cross product.

Example 6.10. A vector which is orthogonal to both $(1, -2, 3)$ and $(2, 0, -1)$ is

$$(1, -2, 3) \times (2, 0, -1) = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & -2 & 3 \\ 2 & 0 & -1 \end{bmatrix}$$

$$= \vec{i} \begin{vmatrix} -2 & 3 \\ 0 & -1 \end{vmatrix} - \vec{j} \begin{vmatrix} 1 & 3 \\ 2 & -1 \end{vmatrix} + \vec{k} \begin{vmatrix} 1 & -2 \\ 2 & 0 \end{vmatrix}$$

$$= \vec{i}((-2)(-1) - (0)(3)) - \vec{j}((1)(-1) - (2)(3)) + \vec{k}((1)(0) - (2)(-2))$$

$$= 2\vec{i} + 7\vec{j} + 4\vec{k} = (2, 7, 4).$$

Notice that if we reverse the order, then $(2, 0, -1) \times (1, -2, 3) = (-2, -7, -4)$. This vector is still orthogonal to both $(1, -2, 3)$ and $(2, 0, -1)$, it just points in the opposite direction. In addition, the area of the parallelogram formed by the vectors $(1, -2, 3)$ and $(2, 0, -1)$ is

$$|(2, 7, 4)| = \sqrt{4 + 49 + 16} = \sqrt{69}.$$

6.4 Projections

6.4.1 Projecting a vector onto a vector

The projection of $\vec{v}$ onto the vector $\vec{w}$, written $\text{proj}_{\vec{w}}\vec{v}$ is a vector parallel to $\vec{w}$ whose magnitude is the component of $\vec{v}$ in the direction of $\vec{w}$. If you draw both vectors with their base at the origin, and create a right triangle with $\vec{v}$ as the hypotenuse and the adjacent edge on the line containing $\vec{w}$, then $\text{proj}_{\vec{w}}\vec{v}$ is the vector which starts at the origin and ends where the right triangle ends, as illustrated in the two examples below.
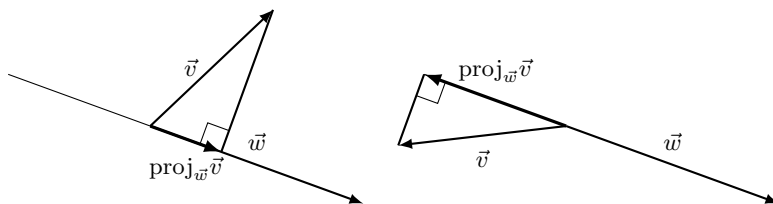
We'll develop a formula for the projection first by using trigonometry, and then second by using facts about inner products and null spaces. We start with trigonometry. The length of the projection is

$$|\vec{v}| \cos \theta = |\vec{v}| \frac{\vec{v} \cdot \vec{w}}{|\vec{v}||\vec{w}|} = \frac{\vec{v} \cdot \vec{w}}{|\vec{w}|}.$$

A unit vector for the direction of the projection is $\dfrac{\vec{w}}{|\vec{w}|}$. Hence we have

$$\text{proj}_{\vec{w}}\vec{v} = (\text{mag})(\text{dir}) = \left(\frac{\vec{v} \cdot \vec{w}}{|\vec{w}|} \right) \frac{\vec{w}}{|\vec{w}|} = \left(\frac{\vec{v} \cdot \vec{w}}{\vec{w} \cdot \vec{w}} \right) \vec{w}.$$

In terms of inner products, the projection of $\vec{v}$ onto $\vec{w}$ is a vector $c\vec{w}$ in the direction of $\vec{w}$ which is "closest" to $\vec{v}$. To be "closest" means we want the distance from the tip of $\vec{v}$ to the line containing $\vec{w}$ to be as small as possible. This occurs precisely when the difference $\vec{v} - c\vec{w}$ has the smallest length, which means that $\vec{v} - c\vec{w}$ meets $\vec{w}$ at a 90 degree angle. We need to pick a value c so that $\vec{v} - c\vec{w}$ and $\vec{w}$ are orthogonal, or have an inner product of zero. Since these two vectors are orthogonal, we need to find c so that

$$\langle \vec{v} - c\vec{w}, \vec{w} \rangle = 0.$$

Since the inner product preserves vector addition and scalar multiplication, we seek c so that

$$\langle \vec{v}, \vec{w} \rangle - c \langle \vec{w}, \vec{w} \rangle = 0$$

Solving for c gives

$$c = \frac{\langle \vec{w}, \vec{w} \rangle}{\langle \vec{v}, \vec{w} \rangle} \quad \text{and so the projection is} \quad \text{proj}_{\vec{w}} \vec{v} = c\vec{w} = \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \vec{w}.$$

The scalar $c = \frac{\langle \vec{v}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle}$ is called the Fourier coefficient of $\vec{v}$ with respect to $\vec{w}$, or the component of $\vec{v}$ along $\vec{w}$.

Example 6.11. The projection of $\vec{v} = (0, 3, 4)$ onto $\vec{w} = (2, 1, -2)$ is

$$\text{proj}_{(2,1,-2)}(0, 3, 4) = \frac{\langle (0, 3, 4), (2, 1, -2) \rangle}{\langle (2, 1, -2), (2, 1, -2) \rangle} (2, 1, -2) = \frac{-5}{9}(2, 1, -2).$$

The number $-5/9$ is the Fourier coefficient of $(0, 3, 4)$ with respect to $(2, 1, -2)$.

The projection of $\vec{w} = (2, 1, -2)$ onto $\vec{v} = (0, 3, 4)$ is

$$\text{proj}_{(0,3,4)}(2, 1, -2) = \frac{\langle (2, 1, -2), (0, 3, 4) \rangle}{\langle (0, 3, 4), (0, 3, 4) \rangle} (0, 3, 4) = \frac{-5}{25}(0, 3, 4).$$

The number $-5/25 = -1/5$ is the Fourier coefficient of $(2, 1, -2)$ with respect to $(0, 3, 4)$.

6.4.2 Projecting a vector onto a subspace

When projecting a vector $\vec{v}$ onto a vector $\vec{w}$, we need to find a vector $\text{proj}_{\vec{w}} \vec{v} = c\vec{w}$ in the space spanned by $\vec{w}$ so that the difference $\vec{n} = \vec{b} - c\vec{w}$ is orthogonal to $\vec{w}$. We wanted to make $\vec{n} = \vec{b} - c\vec{w}$ become a vector in the orthogonal complement of $\vec{w}$ by carefully choosing c. The projection is then simply $\vec{b} - \vec{n}$. Before generalizing this process to larger subspaces, let's follow through this process with a specific example to see how the computations work. We will then repeat this process where we replace $\vec{w}$ with any vector subspace W and learn how to project vectors onto vector subspaces.

Example 6.12. Let's redo the projection example where we project $\vec{v} = (0, 3, 4)$ onto $\vec{w} = (2, 1, -2)$. We want a scalar c so that the vector $\vec{v} - c\vec{w} = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix} - c \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$ is in the null space of $A^T = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}^T = \begin{bmatrix} 2 & 1 & -2 \end{bmatrix}$. This means that $A^T(\vec{v} - c\vec{w}) = 0$ or $A^T c\vec{w} = A^T \vec{v}$. With the numbers from this problem, we need to satisfy the matrix equation

$$\begin{bmatrix} 2 & 1 & -2 \end{bmatrix} c \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix} \quad \text{or} \quad c\vec{w}^T \vec{w} = \vec{w}^T \vec{v}.$$

This is precisely $c(9) = -5$ so $c = -5/9$. Notice that we obtained $c = \vec{w} \cdot \vec{v} / \vec{w} \cdot \vec{w}$.

We are now ready to project an arbitrary vector $\vec{b}$ onto a subspace W. We're using $\vec{b}$ so that the formula we obtain in the end is precisely the forumla we used to perform linear regression. The projection of a vector $\vec{b}$ onto a subspace W is the vector $\vec{w}$ in that space whose head is nearest to the head of $\vec{b}$. To find this vector, we seek a vector $\vec{n}$ in the orthogonal complement of W such that

$$\vec{b} - \vec{n} = \vec{w}$$

is a vector in W. To achieve this, let A be a matrix whose columns form a basis for the space W. The orthogonal complement of W is the null space of A^T. So the solutions to $A^T \vec{n} = \vec{0}$ are precisely the vectors in the orthogonal complement of W. Because the projection $\vec{w} = \vec{b} - \vec{n}$ by definition is in the vector space space W, we know that a linear combination of the columns of A will provide this vector $\vec{w}$. Let $\vec{x}$ be the coordinates of $\vec{w}$ relative to the basis vectors in A, so that

$$A\vec{x} = \vec{w}.$$

Multiplying both sides by A^T gives us

$$A^T A\vec{x} = A^T \vec{w} = A^T \vec{b} - A^T \vec{n} = A^T \vec{b} - \vec{0}.$$

The vectors $\vec{x}$ and $\vec{b}$ satisfy the equation $A^T A\vec{x} = A^T \vec{b}$. We can solve this equation for $\vec{x}$ by inverting the square matrix $A^T A$ to obtain

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}.$$

We can then find the projection of $\vec{b}$ onto W as

$$\text{proj}_W \vec{b} = A\vec{x} = A(A^T A)^{-1} A^T \vec{b}.$$

Does this look familiar? Does it look like our linear regression formula? Linear regression is just projecting a vector onto a subspace, where the coefficients of the line are coordinates $\vec{x}$ of the vector $\vec{w}$ in the space W that is nearest to $\vec{b}$. If S is the columns of A, then the solution to the least square regression problem is simply $\vec{x} = [\vec{w}]_S$ where $\vec{w} = \text{proj}_A \vec{b}$.

Example 6.13. Let's find a line that is closest to passing through the three points $(0,1), (2,3), (4,6)$. Suppose for a moment that there were a line of the form $y = mx + b$ that did pass through the points. Plugging our 3 points into the equation $a_0 + a_1 x = y$ gives the inconsistent system of equations

$$\begin{cases} a_0 = 1 \\ a_0 + 2a_1 = 3 \\ a_0 + 4a_1 = 6 \end{cases} \xrightarrow{\text{augmented matrix}} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 6 \end{array}\right] \xrightarrow{\text{matrix form}} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

Notice that the system can be written in matrix form $A\vec{x} = \vec{b}$ where A contains a column of 1's and x values, and $\vec{b}$ is a column of y values. Because the system above has no solution, we want to find the vector $\vec{w} = \vec{b} - \vec{n}$ in the column space of A (meaning $\vec{w} = A\vec{x}$) where $\vec{w}$ is the projection of $\vec{b}$ onto the 2D space spanned by the columns of A. In other words, find the shortest vector $\vec{n}$ such that $\vec{w} = \vec{b} - \vec{n}$ is a solution to $A\vec{x} = \vec{b} - \vec{n}$. The vector $\vec{n}$ must exist, so multiply both sides by A^T to obtain

$$A^T A\vec{x} = A^T (\vec{b} - \vec{n}) = A^T \vec{b} - A^T \vec{n} = A^T \vec{b},$$

where $A^T \vec{n} = \vec{0}$ because it is in the orthogonal complement of the columns of A (the shortest vector from $\vec{b}$ to the column space of A must be orthogonal to the columns of A). The resulting system has only 2 rows and a symmetric coefficient matrix $A^T A$:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, A^T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \end{bmatrix}, A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix}, A^T \vec{b} = \begin{bmatrix} 10 \\ 30 \end{bmatrix}, \begin{bmatrix} 3 & 6 \\ 6 & 20 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 30 \\ 10 \end{bmatrix}.$$

Reduce $\begin{bmatrix} 3 & 6 & | & 10 \\ 6 & 20 & | & 30 \end{bmatrix}$ to $\begin{bmatrix} 1 & 0 & | & 5/6 \\ 0 & 1 & | & 5/4 \end{bmatrix}$, which means the solution is $y = \frac{5}{6} + \frac{5}{4}x$.
Alternatively, invert $A^T A$ to obtain the solution

$$\vec{x} = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} 5/6 & -1/4 \\ -1/4 & 1/8 \end{bmatrix} \begin{bmatrix} 10 \\ 30 \end{bmatrix} = \begin{bmatrix} 5/6 \\ 5/4 \end{bmatrix}.$$

Let's revisit the previous example but this time use the language of projections and orthogonal complements.

Example 6.14. Let's find the projection of $\vec{b} = (1, 3, 6)$ onto the space

$$W = \text{span}\{(1, 1, 1), (0, 2, 4)\}.$$

If $\vec{w} = \text{proj}_W \vec{b}$, then $\vec{n} = \vec{b} - \vec{w}$ is in the null space of the transpose of $A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}$, so

$A^T(\vec{b} - \vec{w}) = \vec{0}$, or $A^T \vec{w} = A^T \vec{b}$. Since $\vec{w}$ is in the column space of A, we know that $\vec{w} = A\vec{x}$ for some x. We now solve $A^T A\vec{x} = A^T \vec{b}$. The computations are the same as the previous example, and we obtain $\vec{x} = (5/6, 5/4)$. This means

$$\text{proj}_W \vec{b} = A\vec{x} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 5/6 \\ 5/4 \end{bmatrix} = \begin{bmatrix} 5/6 \\ 10/3 \\ 35/6 \end{bmatrix}.$$

The numbers $(5/6, 5/4)$ are the coordinates of the projection relative to the columns of A (a basis for the column space).

6.5 Orthogonal and Orthonormal

The least squares regression problem introduced an idea related to finding the projection of a vector $\vec{b}$ onto a vector subspace W. This requires that we obtain a basis for the space W. Once we have picked a basis, then the projection $A\vec{x} = \vec{w}$ must satisfy $A^T A\vec{x} = A^T \vec{b}$, where $\vec{x}$ is the coordinates of $\vec{w}$ relative to our basis. Solving for $\vec{x}$ we obtain $\vec{x} = (A^T A)^{-1} A^T \vec{b}$. To find the projection $\vec{w} = A\vec{x}$, we multiply by A to obtain

$$\vec{w} = \text{proj}_W \vec{b} = A\vec{x} = A(A^T A)^{-1} A^T \vec{b}.$$

The coordinates of $\vec{w}$ relative to this basis are $\vec{x} = (A^T A)^{-1} A^T \vec{b}$, and are called the Fourier coefficients of $\vec{b}$ relative to the basis chosen for W. In the next section we will show how to obtain a "nice" basis so that obtaining the projection and its coordinates is very simple. As always, this requires that we pick a good basis in which to perform computations.

The standard basis vectors $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ are each orthogonal to the other, so we say they are an **orthogonal set of vectors**. These standard basis vectors are also all unit vectors; they all have norm 1. Because they are all unit vectors and orthogonal, we say they form an **orthonormal** set of vectors. As we proceed throughout this section, we will find that when you choose as your basis an orthonormal set of vectors, computations related to projections and inversion become extremely simple (hence computer programmers prefer to work with orthonormal bases). The telecommunications industry works with an orthonormal basis because computations are extremely fast with this basis.

An orthogonal set of vectors is a set where each vector is orthogonal to all the others.

An orthonormal set of vectors is an orthogonal set where each vector has norm 1.

Example 6.15. The set of vectors $(2, -1, 0), (1, 2, 0), (0, 0, 1)$ forms an orthogonal set because the dot product of any two vectors is zero (check this). Normalizing each vector (dividing by its magnitude) gives the set $(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, 0), (\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, 0), (0, 0, 1)$, which is an orthonormal set of vectors.

Let $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \end{bmatrix}$ be a matrix whose columns form an orthogonal set of vectors. Because the vectors are orthogonal, if $i \neq j$ then $\vec{w}_i \cdot \vec{w}_j = 0$. To compute the matrix product $A^T A$, we take row i of A^T and dot it with column j. But row i of A^T is the same as column i of A, so the ij entry in $A^T A$ is zero if $i \neq i$. This means that $A^T A$ is a diagonal matrix. The diagonal entries are simply $\vec{w}_i \cdot \vec{w}_i = |\vec{w}_i|^2$, the square of the length of the vectors. Diagonal matrices are extremely simple to work with and invert.

In terms of projections, if we choose basis vectors $S = \{\vec{w}_1, \ldots, \vec{w}_n\}$ for W that are orthogonal, then $A^T A$ is a diagonal matrix (the diagonal entries are the square of the lengths of the basis vectors), and the coordinates of our projection relative to this basis (the Fourier coefficients) are (inverting a diagonal matrix is just inverting each diagonal entry)

$$\vec{x} = (A^T A)^{-1} A^T \vec{b} = \left(\begin{bmatrix} \vec{w}_1^T \\ \vec{w}_2^T \\ \vdots \\ \vec{w}_n^T \end{bmatrix} \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \end{bmatrix} \right)^{-1} \begin{bmatrix} \vec{w}_1^T \\ \vec{w}_2^T \\ \vdots \\ \vec{w}_n^T \end{bmatrix} \vec{b}$$

$$= \begin{bmatrix} \vec{w}_1^T \vec{w}_1 & 0 & \cdots & 0 \\ 0 & \vec{w}_2^T \vec{w}_2 & \ddots & 0 \\ \vdots & & \ddots & \vdots \\ \vec{w}_n^T \vec{w}_n & & & \end{bmatrix}^{-1} \begin{bmatrix} \vec{w}_1^T \vec{b} \\ \vec{w}_2^T \vec{b} \\ \vdots \\ \vec{w}_n^T \vec{b} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\langle \vec{w}_1, \vec{w}_1 \rangle} & 0 & \cdots & 0 \\ 0 & \frac{1}{\langle \vec{w}_2, \vec{w}_2 \rangle} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\langle \vec{w}_n, \vec{w}_n \rangle} \end{bmatrix}^{-1} \begin{bmatrix} \langle \vec{w}_1, \vec{b} \rangle \\ \langle \vec{w}_2, \vec{b} \rangle \\ \vdots \\ \langle \vec{w}_n, \vec{b} \rangle \end{bmatrix} = \begin{bmatrix} \frac{\langle \vec{w}_1, \vec{b} \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle} \\ \frac{\langle \vec{w}_2, \vec{b} \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle} \\ \vdots \\ \frac{\langle \vec{w}_n, \vec{b} \rangle}{\langle \vec{w}_n, \vec{w}_n \rangle} \end{bmatrix}.$$

The coordinates $\vec{x} = (c_1, \ldots, c_n)$ are found by computing the inner product of each row with $\vec{b}$ and dividing by the length of that row, or

$$c_i = \frac{\vec{w}_i^T \vec{b}}{\vec{w}_i^T \vec{w}_i} = \frac{\vec{w}_i \cdot \vec{b}}{\vec{w}_i \cdot \vec{w}_i} = \frac{\langle \vec{w}_i, \vec{b} \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle} \qquad \text{(Fourier coefficients)}.$$

If we choose basis vectors S that are orthonormal (so $\langle \vec{w}_i, \vec{w}_i \rangle = 1$), then $A^T A$ is simply the identity matrix. This means $\vec{x} = (A^T A)^{-1} A^T \vec{b} = A^T \vec{b}$ and so the projection of $\vec{b}$ onto W is $\vec{w} = A I A^T \vec{b}$. The coordinates of $\vec{w}$ relative to this basis (called Fourier coefficients) are simply $\vec{x} = A^T \vec{b}$. Notice that this means the coordinates $\vec{x} = (c_1, \ldots, c_n)$ are found by computing the inner product of each row with $\vec{b}$ (no division is needed because each vector has length 1). Using symbols, if $\{\vec{w}_1, \ldots, \vec{w}_n\}$ is an orthonormal basis for W, then the Fourier coefficients of $\vec{b}$ relative to this basis are

$$c_i = \vec{w}_i^T \vec{b} = \vec{w}_i \cdot \vec{b} = \langle \vec{w}_i, \vec{b}_i \rangle.$$

Everything we have developed works in any vector space provide you have an inner product.

Example 6.16. What is the projection of $(3, -2, 5)$ onto the space spanned by the vectors $\vec{v}_1 = (2, -1, 0)$ and $\vec{v}_2 = (1, 2, 0)$? Since the vectors are orthogonal (check that their dot product is 0), we can find the Fourier coefficients by dotting each $\vec{v}_i$ with $\vec{b}$ and dividing by the square of the length. We obtain

$$c_1 = \frac{\langle \vec{v}_1, \vec{b} \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} = \frac{\langle (2, -1, 0), (3, -2, 5) \rangle}{\langle (2, -1, 0), (2, -1, 0) \rangle} = \frac{8}{5},$$

$$c_2 = \frac{\left\langle \vec{v}_2, \vec{b} \right\rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} = \frac{\langle (1,2,0), (3,-2,5) \rangle}{\langle (1,2,0), (1,2,0) \rangle} = \frac{-1}{5}.$$

This means the projection is

$$\vec{w} = c_1 \vec{v}_1 + c_2 \vec{v}_2 = \frac{8}{5}(2,-1,0) + \frac{-1}{5}(1,2,0) = \frac{(16-1, -8-2, 0)}{5} = (3,-2,0).$$

Let's consider the same set of vector, but now use an orthonormal basis to do the projection. These vectors are orthogonal, so normalizing them (divide by their magnitude) yields the orthonormal set of vectors $\vec{w}_1 = \frac{1}{\sqrt{5}}(2,-1,0), \vec{w}_2 = \frac{1}{\sqrt{5}}(1,2,0)$. The coefficients of the projection (the Fourier coefficients relative to the orthonormal basis) are hence

$$c_1 = \vec{w}_1 \cdot \vec{b} = \frac{1}{\sqrt{5}}(2,-1,0) \cdot (3,-2,5) = \frac{8}{\sqrt{5}}, \quad c_2 = \vec{w}_2 \cdot \vec{b} = \frac{1}{\sqrt{5}}(1,2,0) \cdot (3,-2,5) = \frac{-1}{\sqrt{5}}.$$

This means the projection $\vec{w}$ onto the space is

$$\vec{w} = c_1 \vec{w}_1 + c_2 \vec{w}_2 = \frac{8}{\sqrt{5}}\vec{w}_1 + \frac{1-}{\sqrt{5}}\vec{w}_2 = \frac{8}{\sqrt{5}}\frac{1}{\sqrt{5}}(2,-1,0) + \frac{-1}{\sqrt{5}}\frac{1}{\sqrt{5}}(1,2,0)$$
$$= \frac{1}{5}(16-1, -8-2, 0) = (3,-2,0).$$

Computer programs prefer to work with orthonormal bases because with them you can avoid the division by the square of the length of each basis vectors. When working with problems by hand, we'll often just use an orthogonal basis that avoids fractions. This way we can focus on the ideas at hand and then at the very end include the division.

6.5.1   Gram-Schmidt Orthogonalization Process

Given a set of linearly independent vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$, how do I create a set of orthonormal vectors $\{\vec{w}_1, \ldots, \vec{w}_n\}$ whose span is the same as the span of the original vectors? Start with your first vector and let $\vec{w}_1 = \vec{v}_1$. Since $\vec{v}_2$ is not in the space spanned by $\vec{w}_1$, find the projection of $\vec{v}_2$ onto $\vec{w}_1$ and then let $\vec{w}_2$ be $\vec{v}_2$ minus this projection, $w_2 = v_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1$. This means that $\vec{w}_1$ and $\vec{w}_2$ are orthogonal, and they span the same space as $\vec{v}_1$ and $\vec{v}_2$. Since $\vec{v}_3$ is not in the space spanned by $\vec{w}_1$ and $\vec{w}_2$, find the projection of $\vec{v}_3$ onto the plane spanned by $\vec{w}_1$ and $\vec{w}_2$ and then let $\vec{w}_3$ be $\vec{v}_3$ minus this projection, $w_3 = v_3 - \frac{\langle \vec{v}_3, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle}\vec{w}_2$. This means that $\vec{w}_1, \vec{w}_2$, and $\vec{w}_3$ are orthogonal, and span the same space as $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Repeat this process, each time projecting $\vec{v}_k$ onto the subspace generated by the $k-1$ previous orthogonal vectors. Once you are done, you will have an orthogonal set of vectors. Divide each vector by its magnitude to obtain an orthonormal set. Schaum's has a description of this on page 255.

Example 6.17. Example 7.13 in the book (page 256) illustrates this nicely. Due to time constraints, I won't have my own example here.

6.5.2   Orthogonal Matrices

We have seen in chapter 5 that many problems become simpler when we can find a basis that diagonalizes a linear transformation. If P is a similarity matrix that diagonalizes A, then $D = P^{-1}AP$ is the diagonal matrix. As matrices become larger, finding the inverse or P is perhaps the most time intensive part of this computation. This section shows that when A is a symmetric matrix, it is always possible to find a basis such that the inverse of P is extremly simple, namely $P^{-1} = P^T$ (the inverse is the transpose). To introduce this idea, let's start with a simple example.

Example 6.18. The matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues $3, 1$ with eigenvector $(1, 1)$ and $(-1, 1)$. Notice that these two eigenvectors are orthogonal as $(1, 1) \cdot (-1, 1) = 0$. If we let $P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, then $P^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2} P^T$, almost the transpose. The vectors we choose were orthogonal, but not normalized. Since eigenvectors can be multiplied by any scalar and still remain as eigenvectors, let's use the vectors $\frac{1}{\sqrt{2}}(1, 1)$ and $\frac{1}{\sqrt{2}}(-1, 1)$ instead. With these eigenvectors, we have

$$P = \begin{bmatrix} 1\sqrt{2} & -1\sqrt{2} \\ 1\sqrt{2} & 1\sqrt{2} \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 1\sqrt{2} & 1\sqrt{2} \\ -1\sqrt{2} & 1\sqrt{2} \end{bmatrix} = P^T, \quad \text{and} \quad D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = P^T A P.$$

The inverse of P is extremely simple and we can replace $D = P^{-1}AP$ with $D = P^T AP$.

Now it's time for some language.

Definition 6.6. A matrix is said to be orthogonal if $A^T A = AA^T = I$, or in other words $A^T = A^{-1}$ (the transpose is the inverse).

If $\vec{u}_i$ is the ith column of A, then the product $A^T A$ has in the ijth spot the product $\vec{u}_i^T \vec{u}_j$. In order for a matrix to be orthogonal, this requires that $\vec{u}_i^T \vec{u}_j = 0$ if $i \neq j$ (so columns i and j are orthogonal) and that $\vec{u}_i^T \vec{u}_j = 1$ when $i = j$ (so each vector has length 1). We have just proved the following theorem

Schaum's has more examples on pages 113-114.

Theorem 6.4. The following three conditions are equivalent for a square matrix A:

1. A is orthogonal - meaning $A^T A = AA^T = I$.

2. The columns of A form an orthonormal set.

3. The rows of A form an orthonormal set.

The key use we'll make of this theorem is that if a similarity matrix P has orthonormal columns, then its inverse is the transpose, and we can diagonalize A by writing $D = P^T AP$.

Example 6.19. See example 7.14 in Schaum's (page 257). Part b is a rotation matrix. Every rotation in any dimension is represented by an orthogonal matrix, so computer graphic designers use these all the time.

6.5.3 Diagonalizing Symmetric Matrices

If A is a matrix that can be diagonalized by an orthogonal matrix, then what can be said about A? We'll show that A must be symmetric ($A^T = A$). Here's why. If $D = P^T AP$, then multiply on the left by P and on the right by $P^T = P^{-1}$ to obtain $A = PDP^T$. The transpose of A is hence (reverse the order and transpose each)

$$A^T = (PDP^T)^T = P^T D^T (P^T)^T = P^T D P^T = A,$$

since $D^T = D$ (as a diagonal matrix) and $(P^T)^T = P$. Hence, the only kinds of matrices that can be diagonalized with an orthogonal matrix are precisely symmetric matrices.

Symmetric matrices appeared earlier in the semester when we studied optimization and the second derivative test. They also appeared at the beginning of the unit as every inner product can be represented by a symmetric matrix. Much has been learned about symmetric matrices. We now state a key theorem about symmetric matrices.

Theorem 6.5 (Real Spectral Theorem)**.** Let A be a symmetric matrix.

1. All the eigenvalues of A are real.

2. For each eigenvalue, the geometric multiplicity matches the algebraic multiplicity.

3. Eigenvectors corresponding to different eigenvalues are orthogonal.

In particular, these three facts imply that every symmetric matrix can be diagonalized by an orthogonal matrix of real numbers (the Gram-Schmidt process may have to be used if an eigenvalue occurs with algebraic multiplicity more than one).

Example 6.20. See page 404 in Schaum's, example 11.9. When finding the characteristic polynomial, you can either compute the trace and determinant to find the coefficients, or you can subtract λ from the diagonals and then find the determinant.

Now, do you remember the second derivative test we learned at the beginning of the semester? Every matrix involved in the second derivative test will always be symmetric. This means that with every 2nd derivative test (optimization problem), you will always be able to find the eigenvalues and they will be real numbers. In addition, a basis of eigenvectors can be chosen to always meet at right angles. Changing bases is a way of rearranging how we view a problem, and symmetric matrices have extremely nice bases.

6.6 Preparation

This chapter covers the following ideas. When you create your lesson plan, it should contain examples which illustrate these key ideas. Before you take the quiz on this unit, meet with another student out of class and teach each other from the examples on your lesson plan.

1. Explain the dot product and inner product. Use them to define length and angles of vectors. Normalize vectors, obtain vectors of a given length in a given direction, and explain how to tell if two vectors are orthogonal.

2. Obtain the orthogonal complement of a set of vectors (relate it to the null space of the transpose). Use the orthogonal complement to project vectors onto vector spaces.

3. By considering the projection of a vector onto a vector subspace, explain why solving $A^T A \vec{x} = A^T \vec{b}$ gives the solution to the least squares regression problem. This is why statistics works.

4. Use the Gram-Schmidt orthogonalization process to obtain an orthonormal basis of vectors. Show how to compute components of vectors relative to an orthogonal basis (called Fourier coefficients).

5. Illustrate by examples that when a matrix is symmetric, eigenvectors corresponding to different eigenvalues are orthogonal, which means you can find an orthonormal basis to diagonalize the matrix.

Here are the preparation problems for this unit.

Preparation Problems (click for handwritten solutions)

Day 1	7.2, 7.3, 7.4, 7.5	
Day 2	7.6, 7.9, 7.10, 7.12	
Day 3	7.14, 7.19, 7.20, 7.25	
Day 4	7.22, 7.26, 11.25, 11.27	
Day 5	Lesson Plan, Quiz, Prepare to teach each other	

The homework problems in this unit all come from Schaum's Oultines.

Concept	Sec.	Suggestions	Relevant Problems
Dot and Inner Product	7	1-5,7	1-7, 53-56
Orthogonality	7	9-14,	9-14, 60-67
Gram-Schmidt Process	7	18-23	18-23, 70-74
Orthogonal Matrices	7	24-28	24-28, 75-77
Diagonalizing Symmetric Matrices	11	25, 27	25-30, 67-70
More with Orthogonal Matrices	3		32-35, 97-100

Make sure you also review the linear regression problems from the second unit on Applications. In this unit, our main focus is to learn how inner products are used to give us lengths, angles, projections, and then linear regression. Along the way we will pick up some important vocabulary which is used in future generalizations of these ideas to other subjects. Orthonormal bases are extremely nice bases to use, since the corresponding matrix is orthogonal meaning $P^{-1} = P^T$ (inverting is simply transposing).

6.7 Problems

All the problems for this unit are in Schaum's Outlines.

6.8 Projects

1. Currently there is none planned for this unit. Focus on finishing up the project for chapter 5. That will be your last project.

6.9   Solutions

Index

augmented matrix, 8

basis, 48
    $\mathbb{R}^n$, 66
    vector space, 81
bijective, 103

characteristic polynomial, 20
codomain, 103
coefficient matrix, 8
cofactor, 67
column space, 49, 66
consistent, 9, 66
coordinates, 15, 47, 67, 81

determinant, 16, 67
diagonal, 2
dimension, 48, 81
domain, 103
dot product, 65

eigenspace, 53
eigenvalue, 20, 67
eigenvector, 20, 67
elementary matrix, 115
elementary row operations, 9

free variable, 13
function, 103

homogeneous, 9, 66

identity matrix, 6
image, 103, 109
inconsistent, 66
injective, 103
interpolating polynomial, 42
inverse, 67, 104
invertible, 67

kernel, 51, 109

leading 1, 66
least squares regression, 43
linear combination, 5
linear system of equations, 7

linear transformation, 104
linearly dependent, 66
linearly independent, 14, 66

matrix, 1, 65
    inverse, 18
    product, 66
    row equivalent, 66
    size, 66
    sum, 66
    transpose, 67
matrix types
    diagonal, 67
    identity, 67
    lower triangular, 67
    symmetric, 67
    upper triangular, 67
minor, 67

nonhomogeneous, 66
noninvertible, 67
nonsingular, 67
null space, 51
nullity, 110

one-to-one, 103
onto, 103
orthogonal, 4, 65
orthogonal complement, 53

pivot, 66
pivot column, 10, 66
preimage, 103

range, 103
rank, 14, 66
    linear transformation, 110
reduced row echelon form, 10, 66
row echelon form, 9, 66
row space, 49, 66

scalar, 2
scalar multiplication, 65, 76
singular, 67
span, 14, 66