LINEAR ALGEBRA REVIEW

ALLAN YASHINSKI

ABSTRACT. We review the basics of linear algebra.

1. LINEAR SYSTEMS

The study of linear algebra traditionally begins with systems of linear equations. A *linear equation* is an equation such as

$$2x - 5y = 4, \qquad \text{or} \qquad 6x_1 + x_2 + 2x_3 - 3x_4 = 0.$$

The first linear equation has two variables (or unknowns), and the second has four. In general, we may have a large number of variables, so our convention will be to use $x_1, x_2, \ldots, x_n$ as the variables. Formally, a linear equation in n variables $x_1, x_2, \ldots, x_n$ is an equation which can be written in the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

where $a_1, \ldots, a_n$ and b are constants.

We shall study *systems of linear equations*, such as

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 2 \\ 3x_1 + 5x_2 + 5x_3 = 1 \\ 2x_1 + x_2 + 2x_3 = 5 \end{cases}$$

A *solution* of such a system is a triple (x_1, x_2, x_3) of numbers that satisfy all equations simultaneously. To *solve* such a system is to find all solutions. Linear systems are solved by the *Row Reduction Algorithm*, which is also called *Gaussian Elimination.* We arrange all constants of the linear system in an *augmented matrix*

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

and then we apply *elementary row operations.* There are three types of elementary row operations:

(1) (Replacement) Replace one row by the sum of itself and a multiple of another row.
(2) (Interchange/Swap) Swap two rows.
(3) (Scaling) Multiply all entries in a row by the same nonzero scalar.

Note that we can always go back and forth between the linear system and its augmented matrix. Manipulations to the augemented matrix are really manipulations to the linear system. The most important property of these row operations is that **they do not alter the solution set of the underlying linear system**. This means that we can perform a sequence of row operations to the matrix until the solutions are staring us in the face. This is the idea behind the row reduction algorithm.

We return to the example above to illustrate row reduction. Note that row replacement operations are the ones we do most frequently. Every time we perform a row operation, we use the symbol $\sim$, which formally means "is row equivalent to".

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 3 & 5 & 5 & 1 \\ 2 & 1 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & -4 & -5 \\ 2 & 1 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & -4 & -5 \\ 0 & -3 & -4 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & -4 & -5 \\ 0 & 0 & 8 & 16 \end{bmatrix}$$

At this point, we have achieved something called *row echelon form (REF)*. The system is practically solved. Translating it back into a linear system gives

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 2 \\ \quad - \ x_2 - 4x_3 = -5 \\ \qquad\qquad 8x_3 = 16 \end{cases}$$

One can do *back substitution*, which means the following. Notice that we can solve the third equation to obtain the value of x_3. Using it, we could then solve the second equation for x_2. Then plugging x_2 and x_3 into the first equation would allow us to solve for x_1. Even though we haven't done it explicitly, this illustrates the fact that this linear system must have a unique solution (x_1, x_2, x_3). Rather than doing back substitution, let's continue doing more row operations.

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & -4 & -5 \\ 0 & 0 & 8 & 16 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & -4 & -5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & -4 \\ 0 & -1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & 2 & 0 & -4 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

The matrix is now in *reduced row echelon form (RREF)*. If we translate it back into a linear system, the solution is transparent:

$$\begin{cases} x_1 \qquad\quad = 2 \\ \quad x_2 \quad = -3 \\ \qquad x_3 = 2 \end{cases}$$

We see that $(2, -3, 2)$ is the unique solution.

Let's be a little more precise about the echelon forms above. The *leading entry* of a row in a matrix is the first (that is, leftmost) nonzero entry in that row. A matrix is in *row echelon form (REF)* if

(1) Each leading entry is to the right of the leading entry in the row above it,
(2) Any row of all 0's is on the bottom,
(3) All entries below a leading entry are 0.

The third condition is actually implied by the first two, but it was included because it is descriptive of a key feature of row echelon form. A matrix is in *reduced row echelon form (RREF)* if it is in row echelon form, and additionally

(4) All leading entries are 1.
(5) All entries above each leading entry are 0.

For example, the matrices

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

are both row echelon forms, but the second is actually in reduced row echelon form.

Theorem 1.1. *Every matrix has a unique reduced row echelon form.*

This means two things. First, it is always possible to do row reduction until a matrix is in RREF. Second, the final RREF is unique, which means that it doesn't matter what sequence of row operations you used to get there. So while the row operations done in the row reduction algorithm are often described in a very specific order, it actually doesn't matter if you follow that order as long as you end up in RREF. However, the row reduction algorithm is efficient. If you do things your own way, you may end up doing more row operations than are necessary.

For contrast, note that REF is **not** unique. You can take any matrix in REF and multiply a nonzero row by 2 to get a different REF.

The leading entries in a row echelon form are called *pivots*. The pivots, specifically their locations, reveal a lot of important information in various contexts. Note that the locations of the pivots are seen in REF. They are seen in RREF as well, but if you only need to know the locations of the pivots, you can stop early in REF. For example, to solve the system

$$\begin{cases} \phantom{-3x_1 - {}} 5x_2 + 10x_3 = 7 \\ -3x_1 - 2x_2 - 5x_3 = -1 \\ 6x_1 + x_2 + 4x_3 = 2 \end{cases}$$

we make an augmented matrix and proceed to REF (fill in the missing computation yourself):

$$\begin{bmatrix} 0 & 5 & 10 & 7 \\ -3 & -2 & -5 & -1 \\ 6 & 1 & 4 & 2 \end{bmatrix} \sim \ldots \sim \begin{bmatrix} -3 & -2 & -5 & -1 \\ 0 & 5 & 10 & 7 \\ 0 & 0 & 0 & 21/5 \end{bmatrix}$$

As a linear system, this reads

$$\begin{cases} -3x_1 - 2x_2 - 5x_3 &= -1 \\ \phantom{-3x_1 - {}} 5x_2 + 10x_3 &= 7 \\ 0 &= 21/5 \end{cases}$$

There is a bit of a problem with the last equation, namely that it is false. This assertion of a false statement is a contradiction. It is contradicting an assumption we made above. By working with the equations in the original linear system, we were implicitly assuming that they were all true. So we were actually assuming that there is a triple (x_1, x_2, x_3) that can satisfy all equations at once. Since we arrived at a contradiction, what we assumed must be false. So there are no solutions to this linear system. A linear system with no solutions is called *inconsistent*, and a linear system that has at least one solution is called *consistent*.

In terms of pivots, the problem with this linear system is that its augmented matrix has a pivot in the final column. In fact, a linear system is inconsistent if and only if its final column contains a pivot. So once we see the pivot in the final column of the REF, we can stop and conclude there are no solutions. If we proceeded to RREF,

$$\begin{bmatrix} -3 & -2 & -5 & -1 \\ 0 & 5 & 10 & 7 \\ 0 & 0 & 0 & 21/5 \end{bmatrix} \sim \ldots \sim \begin{bmatrix} 1 & 0 & 1/3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

we still see the contradiction $0 = 1$, but there is nothing worthwhile gained from the extra computations.

Suppose we are solving a linear system with 3 equations and 5 unknowns, and its augmented matrix has the following RREF:

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

The system is consistent because the final column does not have a pivot. The feature about this example that is interesting is that some of the other columns do not have pivots. Note that each of the columns, except for the last one, corresponds to one of the variables x_1, x_2, x_3, x_4, x_5. We say that a variable is a *basic variable* if its column contains a pivot, and it is a *free variable* if its column does not contain a pivot. In this example, x_1, x_3 and x_4 are the basic variables, whereas x_2 and x_5 are the free variables. As a linear system, we have

$$\begin{cases} x_1 + 2x_2 + x_5 = 3 \\ x_3 + 6x_5 = 4 \\ x_4 - x_5 = 5. \end{cases}$$

Each basic variable appears in exactly one equation, and the free variables can appear in many equations. So we can solve for each basic variable to obtain

$$\begin{cases} x_1 = 3 - 2x_2 - x_5 \\ x_3 = 4 - 6x_5 \\ x_4 = 5 + x_5 \\ x_2, x_5 \text{ are free} \end{cases}$$

As their name suggests, free variables are free to be anything. We can choose any value for each of x_2 and x_5, and it will determine values for the basic variables. Each time we do this, we obtain one solution. For example, if we choose $x_2 = 1$ and $x_5 = 2$, then

$$\begin{aligned} x_1 &= 3 - 2(1) - 2 = -1 \\ x_3 &= 4 - 6(2) = -8 \\ x_4 &= 5 + 2 = 7. \end{aligned}$$

This gives the solution $(-1, 1, -8, 7, 2)$. However, if we make a difference choice, say $x_2 = 0$ and $x_5 = -3$, we obtain the solution $(6, 0, 22, 2, -3)$. Each choice of values for the free variables leads to a new solution. Since we have infinitely many choices for the free variables, we have infinitely many solutions to the linear system. In fact, a consistent linear system has infinitely many solutions if and only if there is a free variable.

Theorem 1.2. *Every linear system has either zero solutions, exactly one solution, or infinitely many solutions.*

To summarize, there are no solutions if and only if there is a pivot in the final column of the augmented matrix. If it is consistent and there is a free variable, then there are infinitely many solutions. If it is consistent and there is no free variable (all variables are basic), then there is a unique solution. Note that we can determine which of the three cases we are in based on the locations of the pivots, which can be seen in REF. However, if we want to describe the solution(s) of a linear system, we need the RREF.

Exercise 1.3. Consider the linear system whose augmented matrix has the form

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix}$$

for certain constants b_1, b_2, b_3. Find a condition that the values b_1, b_2, b_3 must satisfy so that this system is consistent. In the case where it is consistent, determine how many solutions exist.

2. Matrices and Vectors

2.1. **Matrix notation.** An $m \times n$ *matrix* is a rectangular array of numbers that has m rows and n columns. By convention, we always put the number of rows before the number of columns. For example

$$\begin{bmatrix} 2 & 3 \\ 6 & -1 \end{bmatrix}, \qquad \begin{bmatrix} 3/2 \\ 0 \\ \pi \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

are, respectively, 2×2, 3×1, and 4×2 matrices. We traditionally use capital letters such as A, B, C, etc. to refer to matrices. A generic $m \times n$ matrix A is sometimes denoted

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix},$$

where a_{ij} represents the entry in row i and column j. A matrix with a single column is called a *column vector*. For example,

$$\mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

are column vectors. A matrix with a single row is called a *row vector*. We won't use them as much as column vectors. When we simply say *vector*, we usually mean column vector. A 1×1 matrix is just a number. We usually write it without matrix bracket notation.

The *main diagonal* of a matrix A consists of the entries $a_{11}, a_{22}, a_{33}, \dots$ etc. A matrix is called *diagonal* if all of its entries that are not on the main diagonal are zero. Here are some examples of diagonal matrices:

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \qquad \begin{bmatrix}3 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & 11\end{bmatrix}, \qquad \begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}.$$

The term diagonal typically refers to square (that is, $n\times n$) matrices. But sometimes it is important to consider matrices such as

$$\begin{bmatrix}4 & 0 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & 1/2 & 0\end{bmatrix}, \qquad \begin{bmatrix}8 & 0\\ 0 & 5\\ 0 & 0\\ 0 & 0\end{bmatrix},$$

and these are sometimes referred to as diagonal.

An $n \times n$ matrix is called *upper triangular* if all entries below its main diagonal are zero. It is *lower triangular* if all entries above its main diagonal are zero. For example,

$$\begin{bmatrix}2 & 1\\ 0 & 4\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}1 & 0 & 0\\ 5 & 0 & 0\\ 0 & 4 & -1\end{bmatrix}$$

are upper triangular and lower triangular, respectively. Note that an $n \times n$ matrix is diagonal if and only if it is both upper and lower triangular.

2.2. **Matrix arithmetic.** Matrices can be added together, provided they have the same size. Addition is done entry-wise. For example

$$\begin{bmatrix}1 & 2 & 4\\ -1 & 0 & 1\end{bmatrix} + \begin{bmatrix}3 & -5 & 3\\ 2 & 1 & 0\end{bmatrix} = \begin{bmatrix}4 & -3 & 7\\ 1 & 1 & 1\end{bmatrix}, \qquad \begin{bmatrix}2\\ 1\\ 2\end{bmatrix} + \begin{bmatrix}-1\\ 0\\ 1\end{bmatrix} = \begin{bmatrix}1\\ 1\\ 3\end{bmatrix}.$$

There is also a *scalar multiplication* operation (note that scalar is just a fancy name for number) in which we multiply each entry of a matrix by the same scalar:

$$3\begin{bmatrix}2 & 1\\ -1 & -3\end{bmatrix} = \begin{bmatrix}6 & 3\\ -3 & -9\end{bmatrix}, \qquad (-2)\begin{bmatrix}3\\ -1\\ 0\end{bmatrix} = \begin{bmatrix}-6\\ 2\\ 0\end{bmatrix}.$$

Matrix multiplication is not as straightforward. The matrix product AB is only defined if the number of columns of A equals the number of row of B. So AB is defined when A is $m \times n$ and B is $n \times p$. The resulting product AB is an $m \times p$ matrix. The row i and column j entry of the product AB is given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj},$$

where lower case a's and b's are the entries of A and B (as described above). In other words, to compute the entry of AB is row i and column j, we take row i from A and column j from B, and compute the sum of the products of corresponding entries. For example,

$$\begin{bmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{bmatrix}\begin{bmatrix}7 & 8\\ 9 & 10\end{bmatrix} = \begin{bmatrix}1\cdot 7 + 2\cdot 9 & 1\cdot 8 + 2\cdot 10\\ 3\cdot 7 + 4\cdot 9 & 3\cdot 8 + 4\cdot 10\\ 5\cdot 7 + 6\cdot 9 & 5\cdot 8 + 6\cdot 10\end{bmatrix} = \begin{bmatrix}25 & 28\\ 57 & 64\\ 89 & 100\end{bmatrix}.$$

Note that

$$\begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

is undefined. In gereral, matrix multiplication is *noncommutative*. This means that we can (and typically do) have $AB \neq BA$, even when both sides are defined. Note that it is possible to have $AB = BA$ in certain examples, as illustrated by the following exercise.

Exercise 2.1. Let $A = \begin{bmatrix} 2 & 1 \\ 0 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 3 & 4 \\ 8 & 11 \end{bmatrix}$. Show that $AB \neq BA$, but $BC = CB$.

One common occurence of matrix multiplication is when A is $m \times n$ and $\mathbf{x}$ is an $n \times 1$ column vector. Then the product $A\mathbf{x}$ is an $m \times 1$ column vector. For example:

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

2.3. **Zero vector and identity matrix.** The *zero vector* is the column vector of all 0's. Well really there are several zero vectors, depending on the size of the column. Let $\mathbf{0}_n$ denote the $n \times 1$ zero vector:

$$\mathbf{0}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathbf{0}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{0}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{etc.}$$

We will often denote the zero vector simply by $\mathbf{0}$. The number of entries should be clear from the context. For any $\mathbf{v} \in \mathbb{R}^n$, the zero vector $\mathbf{0}$ $(= \mathbf{0}_n)$ satisfies

$$\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}.$$

Also, for any $m \times n$ matrix A, we have $A\mathbf{0} = \mathbf{0}$. This is a spot where perhaps we should be more careful with notation, and write $A\mathbf{0}_n = \mathbf{0}_m$.

The $n \times n$ *identity matrix* is the diagonal matrix I_n that has all 1's on its diagonal. So

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \text{etc.}$$

The identity matrix is special because it satisfies

$$AI_n = A = I_m A$$

for any $m \times n$ matrix A. In particular, $I_n\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. If the context is clear, we shall just write I instead of I_n.

2.4. **Transpose.** For reasons that are not immediately apparent, the following operation is important. Given an $m \times n$ matrix A, its *transpose* is the $n \times m$ matrix A^T whose rows are the columns of A (equivalently, whose columns are the rows of A). For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

then

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}, \qquad \mathbf{v}^T = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}.$$

One important property of the transpose operation is that if A is $m \times n$ and B is $n \times p$, then

$$(AB)^T = B^T A^T.$$

A similar property holds for larger products. For example,

$$(ABCD)^T = D^T C^T B^T A^T,$$

assuming the product $ABCD$ is defined.

2.5. **The linear system $A\mathbf{x} = \mathbf{b}$.** Every system of linear equations can be repackaged into the form $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x}$ is a column vector whose entries are the variables $x_1, x_2, \dots, x_n$. To see how this is done, consider the example

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 2 \\ 3x_1 + 5x_2 \qquad\quad = 1 \\ 2x_1 - \; x_2 + 2x_3 = 5 \end{cases}$$

These three equalities hold if and only if we have equality

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

of 3×1 column vectors. But the left hand side can be recognized as a matrix vector product:

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

So the original system is equivalent to the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 5 & 0 \\ 2 & -1 & 2 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}.$$

Here, our perspective is shifted slightly. Instead of thinking we must solve for three unknown scalars x_1, x_2, x_3, we must solve for one unknown vector $\mathbf{x}$. The matrix A above is called the *coefficient matrix* of the linear system.

Conversely, any equation of the form $A\mathbf{x} = \mathbf{b}$ can be transformed into a linear system. For example, if

$$A = \begin{bmatrix} 3 & 1 & 2 & -4 \\ 6 & 0 & 1 & 2 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 3 \end{bmatrix},$$

then $A\mathbf{x} = \mathbf{b}$ is the linear system

$$\begin{cases} 3x_1 + x_2 + 2x_3 - 4x_4 = 4 \\ 6x_1 \phantom{{}+x_2} + x_3 + 2x_4 = 3 \end{cases}$$

Observe that the augmented matrix of this linear system is

$$\begin{bmatrix} 3 & 1 & 2 & -4 & 4 \\ 6 & 0 & 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} A & \mathbf{b} \end{bmatrix},$$

where the shorthand notation $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ means the matrix which begins with the block A and has $\mathbf{b}$ as its final column. The augmented matrix is formed by "augmenting" the column $\mathbf{b}$ to the coefficient matrix A.

Theorem 2.2. *If A is $m \times n$ and $\mathbf{b}$ is $m \times 1$, then the equation $A\mathbf{x} = \mathbf{b}$ represents the linear system whose augmented matrix is $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$.*

Since we have changed perspective so that we solve a linear system $A\mathbf{x} = \mathbf{b}$ for the vector $\mathbf{x}$, let us revisit an example and present its solution as a vector. We saw an example above in which there were two free variables, and the general solution was

$$\begin{cases} x_1 = 3 - 2x_2 - x_5 \\ x_3 = 4 - 6x_5 \\ x_4 = 5 + x_5 \\ x_2, x_5 \text{ are free} \end{cases}$$

The vector $\mathbf{x}$ of unknowns can be presented as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 - 2x_2 - x_5 \\ x_2 \\ 4 - 6x_5 \\ 5 + x_5 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 4 \\ 5 \\ 0 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ x_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -x_5 \\ 0 \\ -6x_5 \\ x_5 \\ x_5 \end{bmatrix}, \text{ or}$$

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 4 \\ 5 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ -6 \\ 1 \\ 1 \end{bmatrix}.$$

This final form is referred to as the *parametric vector form* of the solution.

A *homogeneous* linear system is one of the form $A\mathbf{x} = \mathbf{0}$. Note that such a linear system must be consistent. Indeed, it is impossible for the augmented matrix $\begin{bmatrix} A & \mathbf{0} \end{bmatrix}$ to have a pivot in its final column. A more direct and useful explanation is that it is easy to identify one solution of $A\mathbf{x} = \mathbf{0}$, namely the vector $\mathbf{x} = \mathbf{0}$. This solution is called the *trivial solution*, because it works for trivial reasons. A *nontrivial solution* of $A\mathbf{x} = \mathbf{0}$ is a solution $\mathbf{x}$ for which $\mathbf{x} \neq \mathbf{0}$. If a nontrivial solution exists, then $A\mathbf{x} = \mathbf{0}$ has at least two distinct solutions, and hence it has infinitely many solutions. This means there must have been a free variable, so a column of A lacked a pivot.

Theorem 2.3. *The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$ if and only if there is a pivot in every column of A.*

2.6. **Linear combinations and span.** Let n denote a positive integer and let $\mathbb{R}^n$ denote the set of all $n \times 1$ column vectors. We often identify

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \ : \ x_1, x_2 \in \mathbb{R} \right\}$$

with the Cartesian plane. Similarly, we identify

$$\mathbb{R}^3 = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \ : \ x_1, x_2, x_3 \in \mathbb{R} \right\}$$

with 3-dimensional space. Although difficult to visualize, $\mathbb{R}^4, \mathbb{R}^5, \mathbb{R}^6$, etc. and $\mathbb{R}^n$ for larger n are also important. For each integer n, $\mathbb{R}^n$ is an example of a *vector space*, which (very roughly) means that $\mathbb{R}^n$ is a set of vectors which can be combined together to form new vectors using the operations of vector addition and scalar multiplication. If we mix together these two operation, we can start forming vectors like $2\mathbf{v}_1 + 5\mathbf{v}_2$ or $\mathbf{w}_1 - \mathbf{w}_2 - 17\mathbf{w}_3 + 9\mathbf{w}_4$.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors in $\mathbb{R}^n$. A *linear combination* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is a vector of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_p,$$

where $c_1, c_2, \ldots, c_p$ are scalars. There are many choices for the various c_i's, so there are many linear combinations of a given set of vectors. The set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is called the *span* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, and is denoted $\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$.

Example 2.4. If we consider a single vector $\mathbf{v} \in \mathbb{R}^n$, then $\mathrm{Span}\{\mathbf{v}\} = \{c\mathbf{v} \ : \ c \in \mathbb{R}\}$ is simply the set of all scalar multiples of $\mathbf{v}$. In $\mathbb{R}^2$ or $\mathbb{R}^3$, the span of a single (nonzero) vector $\mathbf{v}$ can be visualized as a line through the origin pointing in the direction of $\mathbf{v}$.

Example 2.5. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We claim that $\mathrm{Span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$, which means that every vector in $\mathbb{R}^2$ can be written as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. As an example, $\begin{bmatrix} 7 \\ 11 \end{bmatrix}$ is a vector in $\mathbb{R}^2$. How can we write it as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$? The answer is simply

$$\begin{bmatrix} 7 \\ 11 \end{bmatrix} = 7\mathbf{e}_1 + 11\mathbf{e}_2.$$

More generally, any $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ from $\mathbb{R}^2$ can be written as the linear combination

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2.$$

In general, $\text{Span}\{\mathbf{v}_1, \ldots, v_p\}$ is a subset of $\mathbb{R}^n$. If $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \mathbb{R}^n$, then we say that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ *span* $\mathbb{R}^n$. For example, the vectors $\mathbf{e}_1, \mathbf{e}_2$ of the previous example span $\mathbb{R}^2$.

Suppose A is an $m \times n$ matrix whose columns are the n column vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. We signify this notationally by $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}$. There is a convenient interpretation of the matrix vector product $A\mathbf{x}$ in terms of linear combinations:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n.$$

In other words, the product $A\mathbf{x}$ is the linear combination of the columns of A obtained using the entries of $\mathbf{x}$ as the coefficients. In fact,

$$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \{A\mathbf{x} \; : \; \mathbf{x} \in \mathbb{R}^n\}.$$

That is, the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is precisely the set of all vectors which have the form $A\mathbf{x}$, where $\mathbf{x}$ could be any vector in $\mathbb{R}^n$.

In light of the notation above, there is another perspective one can take on matrix multiplication. Say A is $m \times n$ and B is $n \times p$ with columns given by $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_1 & \ldots & \mathbf{b}_p \end{bmatrix}$. Then the product AB is given by

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_1 & \ldots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_1 & \ldots & A\mathbf{b}_p \end{bmatrix}.$$

So every column of the matrix AB is also in the span of the columns of A. This also shows how the general definition of matrix multiplication can be given in terms of the matrix-vector product.

Example 2.6. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$. We'll determine if $\mathbf{b}$ is an element of $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. That is, we will determine whether $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. We seek scalars x_1, x_2 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{b}.$$

In light of the previous discussion, this equation is $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$. This becomes the linear system

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

Specifically, we are interested in knowing if this system is consistent. If it is, then $\mathbf{b}$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, and therefore is in their span. We form the augmented matrix and do row reduction:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 4 \\ 1 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

In row echelon form, we see there is no pivot in the final column and therefore the system is consistent. So $\mathbf{b}$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. To know what the actual linear combination is, we proceed to RREF:

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

The solution is $x_1 = -3$ and $x_2 = 2$. So $-3\mathbf{v}_1 + 2\mathbf{v}_2 = \mathbf{b}$, which we can confirm with a computation.

Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^m$, how can we tell if they span $\mathbb{R}^m$? Let $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}$. If they do span $\mathbb{R}^m$, then the linear system $A\mathbf{x} = \mathbf{b}$ is consistent for *every* possible choice of $\mathbf{b} \in \mathbb{R}^m$. So after row reducing the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$, we never get a pivot in the final column.

Example 2.7. Let's show that $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 8 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ -5 \\ 3 \end{bmatrix}$ span $\mathbb{R}^3$. We'll show that

$$\begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & -5 \\ -2 & 8 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

is consistent for every possible right hand side. We row reduce the augmented matrix

$$\begin{bmatrix} 1 & -3 & 0 & b_1 \\ 0 & 1 & -5 & b_2 \\ -2 & 8 & 3 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 0 & b_1 \\ 0 & 1 & -5 & b_2 \\ 0 & 2 & 3 & 2b_1 + b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 0 & b_1 \\ 0 & 1 & -5 & b_2 \\ 0 & 0 & 13 & 2b_1 - 2b_2 + b_3 \end{bmatrix}.$$

Since all pivots are in the first three columns, we will never get one in the fourth column. The system is consistent regardless of the values of b_1, b_2, b_3. Hence, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$.

In the previous example, we could have remarked that since

$$A = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & -5 \\ -2 & 8 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 13 \end{bmatrix}$$

already has a pivot in every row, then there is not way that the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ could have a pivot in its final column.

Theorem 2.8. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors in $\mathbb{R}^m$ and let $A = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}$. Then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span $\mathbb{R}^m$ if and only if A has a pivot in every row.*

Note that if A is $m \times n$, then its number of pivots is bounded by both m and by n. Since the above scenario requires exactly m pivots, we must have $m \leq n$.

Corollary 2.9. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors in $\mathbb{R}^m$. If $n < m$, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ do not span $\mathbb{R}^m$.*

So 2 vectors cannot span $\mathbb{R}^3$, and 5 vectors cannot span $\mathbb{R}^7$. However, if you had 11 vectors in $\mathbb{R}^7$, they may or may not span $\mathbb{R}^7$. It depends on the vectors.

2.7. **Linear independence.** Roughly speaking, vectors are linearly dependent when relations such as

$$\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2 + 6\mathbf{v}_4 \quad \text{or} \quad 2\mathbf{w}_1 + 6\mathbf{w}_5 = 3\mathbf{w}_2 - 4\mathbf{w}_3 + 7\mathbf{w}_4$$

exist. Note that these relations can be rewritten as

$$-2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 - 6\mathbf{v}_4 = \mathbf{0} \quad \text{and} \quad 2\mathbf{w}_1 - 3\mathbf{w}_2 + 4\mathbf{w}_3 - 7\mathbf{w}_4 + 6\mathbf{w}_5 = \mathbf{0}$$

respectively. In each case, the left hand side is a linear combination of the vectors and the equation says it equals the zero vector. For any collection $\mathbf{v}_1, \ldots, \mathbf{v}_p$ of vectors in $\mathbb{R}^n$, we can always trivially obtain the zero vector as a linear combination if we use all 0's as coefficients:

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \ldots + 0\mathbf{v}_p = \mathbf{0}.$$

But the examples above are interesting because we are using nonzero coefficients. This motivates the following definition.

Definition 2.10. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is *linearly independent* if the linear system

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution $x_1 = 0, x_2 = 0, \ldots, x_p = 0$.

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of vectors is *linearly dependent* if it is not linearly independent. This means that there exist scalars $x_1, \ldots, x_p$, not all zero, such that

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = \mathbf{0}.$$

Note that some of the coefficients can be zero, as long as some are nonzero. For example, if $\mathbf{v}_1 = 4\mathbf{v}_2$, then the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ must be linearly dependent because

$$(1)\mathbf{v}_1 + (-4)\mathbf{v}_2 + (0)\mathbf{v}_3 + (0)\mathbf{v}_4 = \mathbf{0}.$$

The following characterization of linear dependence is useful intuitively.

Theorem 2.11. *A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors is linearly dependent if and only if (at least) one of the vectors in the set can be written as a linear combination of the others.*

Note that if one of the vectors $\mathbf{v}_i$ is the zero vector $\mathbf{0}$, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is automatically linearly dependent. A set $\{\mathbf{v}_1\}$ with a single vector $\mathbf{v}_1$ is linearly independent if and only if $\mathbf{v}_1 \neq \mathbf{0}$. A set $\{\mathbf{v}_1, \mathbf{v}_2\}$ with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other (that is, they "point in different directions"). For three or more vectors, the issue is a bit more subtle than just being about scalar multiples.

Example 2.12. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$. Note that none of the three vectors is a scalar multiple of any of the others. However, $\mathbf{v}_3 = 3\mathbf{v}_1 + 4\mathbf{v}_2$, which shows that $\mathbf{v}_3$ is a *linear combination* of the others. This implies $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

How do we practically determine whether a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent? This amounts to investigating the solution set of a homogeneous linear system. Note that the equation

$$x_1\mathbf{v}_1 + \dots + x_p\mathbf{v}_p = \mathbf{0}$$

is equation to the linear system

$$A\mathbf{x} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \mathbf{0},$$

where A is the matrix whose columns are $\mathbf{v}_1, \dots, \mathbf{v}_p$. The homogeneous linear system will have nontrivial solutions if and only if we have a free variable.

Theorem 2.13. *Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix}$. Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent if and only if A has a pivot in every column.*

Since we are allowed at most one pivot per row, this requirement cannot be satisfied if we have too many columns compared to rows. This leads to the following useful fact.

Corollary 2.14. *Let $\{\mathbf{v}_1, \dots \mathbf{v}_p\}$ be a set of p vectors in $\mathbb{R}^n$. If $p > n$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent.*

So if you have four vectors in $\mathbb{R}^3$, it must be possible to write one of them as a linear combination of the other three.

3. Linear Transformations

A linear transformation is a certain type of function that transforms vectors into other vectors, possibly with different dimensions. Many important geometric transformations such as rotations, reflections, dilations (stretching), projections, and shears turn out to be linear transformations.

Definition 3.1. A *linear transformation* is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and
(2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

The most important example of a linear transformation is a *matrix transformation*. Given any $m \times n$ matrix A, we can define a function

$$T : \mathbb{R}^n \to \mathbb{R}^m, \qquad T(\mathbf{x}) = A\mathbf{x}.$$

Note that the input $\mathbf{x}$ comes from $\mathbb{R}^n$, and the output $A\mathbf{x}$ is in $\mathbb{R}^m$. This is a linear transformation by properties of matrix multiplication:

(1) $T(\mathbf{u} + \mathbf{v}) = A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$, and
(2) $T(c\mathbf{u}) = A(c\mathbf{u}) = c(A\mathbf{u}) = cT(\mathbf{u})$.

This says that multiplication by a given matrix A is a linear transformation. This gives us a new perspective on matrices: we can think of a matrix A as a linear transformation.

3.1. **Standard matrix of a linear transformation.** The reason why matrix transformations are the most important examples of linear transformations is because they are the *only* examples of linear transformations. Every single linear transformation (such as rotations, reflections, projections, etc.) can be described by a matrix.

To understand why this is the case and to know how to produce the matrix, it is useful to introduce the *standard basis vectors* in $\mathbb{R}^n$. The standard basis vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ in $\mathbb{R}^n$ are the columns of the $n \times n$ identity matrix I_n. In $\mathbb{R}^2$, we have

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

whereas in $\mathbb{R}^3$, we have

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The vector $\mathbf{e}_j$ always has a 1 in the j-th position and 0's in all other positions. The main point going forward is that a linear transformation T is completely determined by what it does to the standard basis vectors.

Example 3.2. Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation with the property that

$$T(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad T(\mathbf{e}_2) = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \qquad T(\mathbf{e}_3) = \begin{bmatrix} -1 \\ -4 \end{bmatrix}.$$

Let's determine the vector $T(\mathbf{u})$ where $\mathbf{u} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}$. The key observation is that we can write $\mathbf{u}$ as a linear combination of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ in an "obvious" way:

$$\mathbf{u} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} = 4\mathbf{e}_1 + 3\mathbf{e}_2 + (-2)\mathbf{e}_3.$$

Then we use the properties in the definition of linear transformation:

$$\begin{aligned} T(\mathbf{u}) &= T(4\mathbf{e}_1 + 3\mathbf{e}_2 + (-2)\mathbf{e}_3) \\ &= T(4\mathbf{e}_1) + T(3\mathbf{e}_2) + T(-2\mathbf{e}_3) \\ &= 4T(\mathbf{e}_1) + 3T(\mathbf{e}_2) + (-2)T(\mathbf{e}_3) \\ &= 4\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 3\begin{bmatrix} 6 \\ 0 \end{bmatrix} + (-2)\begin{bmatrix} -1 \\ -4 \end{bmatrix} \\ &= \begin{bmatrix} 24 \\ 20 \end{bmatrix}. \end{aligned}$$

Note we can repeat this identical argument with a generic $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in place of $\mathbf{u}$ to obtain

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) \\ &= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + T(x_3\mathbf{e}_3) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + x_3T(\mathbf{e}_3) \\ &= x_1\begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 6 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} -1 \\ -4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 6 & -1 \\ 3 & 0 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \end{aligned}$$

We've shown $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 6 & -1 \\ 3 & 0 & -4 \end{bmatrix}$. That is, we've found the explicit matrix which describes this linear transformation T as a matrix transformation. Observe also that its columns are simply the vectors $T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)$.

This example generalizes to give the following Theorem:

Theorem 3.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. Then there is a unique $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. The matrix A is given by $A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{bmatrix}$, where $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ are the standard basis vectors for $\mathbb{R}^n$.*

The matrix A above is called the *standard matrix* of the linear transformation T.

Example 3.4. Counterclockwise rotation about the origin by an angle of $\pi/2$ radians (90 degrees) defines a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$. Let's find its standard matrix. The columns of its standard matrix are $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$, so it suffices to determine what T does to the standard basis vectors. If we rotate $\mathbf{e}_1$ counterclockwise by 90 degrees, we obtain $\mathbf{e}_2$ (draw a picture), so $T(\mathbf{e}_1) = \mathbf{e}_2$. If we rotate $\mathbf{e}_2$ counterclockwise by 90 degrees, we obtain $-\mathbf{e}_1$ (again, draw a picture), so $T(\mathbf{e}_2) = -\mathbf{e}_1$. The standard matrix is

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_2 & -\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

So counterclockwise rotation about the origin by 90 degrees is given by

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\mathbf{x}.$$

We can use this to rotate any vector by 90 degrees. For example, if $\mathbf{x} = \begin{bmatrix} 9 \\ 13 \end{bmatrix}$, then

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 9 \\ 13 \end{bmatrix} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}.$$

The rotated vector is $\begin{bmatrix} -13 \\ 9 \end{bmatrix}$.

Example 3.5. Counterclockwise rotation about the origin by θ radians is a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$. Find its standard matrix. (Follow the method of the previous example. You will have to do a little trigonometry. The answer will involve $\cos\theta$ and $\sin\theta$.)

3.2. **Composition of linear transformations.** Suppose $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations. Say $T(\mathbf{x}) = A\mathbf{x}$ for some $m \times n$ matrix A and $S(\mathbf{x}) = B\mathbf{x}$ for some $n \times p$ matrix B. The composition transformation

$$T \circ S : \mathbb{R}^p \to \mathbb{R}^m, \qquad (T \circ S)(\mathbf{x}) = T(S(\mathbf{x}))$$

is also linear. Note the composition $T \circ S$ means "apply S, then apply T." Since $T \circ S$ is also linear, it has its own standard matrix, and in fact

$$(T \circ S)(\mathbf{x}) = T(S(\mathbf{x})) = A(B(\mathbf{x})) = (AB)\mathbf{x},$$

where the last equality follows from the associativity of matrix multiplication. This shows that the matrix product AB is the standard matrix of the composite transformation $T \circ S$.

The above discussion is a bit backwards, because the definition of matrix multiplication was designed so that the standard matrix of $T \circ S$ equals the matrix product AB. Matrix multiplication **is** function composition. This should be emphasized, so we will:

Matrix multiplication is function composition of linear transformations.

Just as $T \circ S$ means "apply S, then apply T", the matrix AB (as a matrix transformation) means "multiply by B, then multiply by A." You should read matrix multiplication right-to-left, as you do with function composition. This interpretation of matrix multiplication as composition also explains some of the quirks of matrix multiplication. Why are we only allowed to multiply an $m \times n$ matrix by an $n \times p$ matrix? It's because to compose transformations, we need the codomain of the first (right) one to match the domain of the second (left) one. Why is the resulting matrix $m \times p$? It's because the composite function goes from $\mathbb{R}^p$ to $\mathbb{R}^m$.

The final quirk of matrix multiplication that this illuminates is the noncommutativity. Why is matrix multiplication noncommutative? The answer is because function composition is noncommutative. The order of linear transformations in a composition matters. If you reflect and then rotate, you may get a different result than if you first rotate and then reflect.

Example 3.6. Counterclockwise rotation about the origin is given by $T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, whereas reflection about the line $x_1 = x_2$ is given by $S(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\mathbf{x}$. Let's find the standard matrix of the linear transformation that first reflects about $x_1 = x_2$ and then rotates 90 degrees. This is $T \circ S$, and its standard matrix is

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

On the other, the transformation that first rotates 90 degrees and then reflects about $x_1 = x_2$ is $S \circ T$, and its standard matrix is

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

Note that $T \circ S$ and $S \circ T$ are different transformations.

3.3. **One-to-one and onto.** Let's briefly step away from linear algebra and discuss general functions. Suppose X and Y are sets and $f : X \to Y$ is a function. This means that to each $x \in X$, f assigns a unique element $f(x) \in Y$. It is unique in the sense that if you evaluate $f(x)$ today, and then evaluate $f(x)$ again tomorrow, you will always get the same value $f(x)$. Every element $x \in X$ must be assigned to some $f(x) \in Y$. It is allowed to have two different $x_1, x_2 \in X$ be assigned to the same element of Y. That is, $f(x_1) = f(x_2)$ is allowed, even if $x_1 \neq x_2$. There is no rule that says that every $y \in Y$ must be f of some x. Some lonely y's may not have any x's assigned to them by f. The *range* (or *image*) of f is the set of all $y \in Y$ that have some $x \in X$ assigned to them:

$$\text{range}(f) = \{f(x) \ : \ x \in X\}.$$

Thinking of x as the input, and $f(x)$ as the output, the range is just the set of all outputs produced by a function.

A function $f : X \to Y$ is called *one-to-one* (or *injective*) if distinct elements of X are always assigned to distinct elements of Y. That is, $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. (Equivalently, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.)

A function $f : X \to Y$ is called *onto* (or *surjective*) if range$(f) = Y$. That is, for every $y \in Y$, there is some $x \in X$ for which $f(x) = y$.

A function $f : X \to Y$ is called a *one-to-one correspondence* (or a *bijection*) if f is both one-to-one and onto. These are the functions that have inverses. If $f : X \to Y$ is a one-to-one correspondence, then there is an inverse function $f^{-1} : Y \to X$ such that

$$f^{-1}(f(x)) = x, \qquad f(f^{-1}(y)) = y, \qquad \text{for all } x \in X,\ y \in Y.$$

Returning to a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$, we can ask if T is one-to-one or onto. Since T is determined completely by A, we should be able to determine the answer from A.

Let's start with onto. This condition says that for every $\mathbf{b} \in \mathbb{R}^m$, then is some $\mathbf{x} \in \mathbb{R}^n$ for which $T(\mathbf{x}) = \mathbf{b}$. But $T(\mathbf{x}) = A\mathbf{x}$, so this condition says that $A\mathbf{x} = \mathbf{b}$ is consistent for every possible $\mathbf{b} \in \mathbb{R}^m$. In other words, T being onto is equivalent to the columns of A spanning $\mathbb{R}^m$.

Theorem 3.7. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. Then T maps onto $\mathbb{R}^m$ if and only if there is a pivot in every row of A.*

Now let's consider whether T is one-to-one. If it isn't, then there are two distinct vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ such that $T(\mathbf{u}) = T(\mathbf{v})$. Using linearity,

$$\begin{aligned} T(\mathbf{u}) &= T(\mathbf{v}) \\ T(\mathbf{u}) - T(\mathbf{v}) &= \mathbf{0} \\ T(\mathbf{u} - \mathbf{v}) &= \mathbf{0} \\ A(\mathbf{u} - \mathbf{v}) &= \mathbf{0}. \end{aligned}$$

Since $\mathbf{u}, \mathbf{v}$ are distinct, the vector $\mathbf{x} = \mathbf{u} - \mathbf{v}$ is a nontrivial solution to $A\mathbf{x} = \mathbf{0}$. Conversely, if we knew that $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions, then we have infinitely many $\mathbf{x}$ such that $T(\mathbf{x}) = A\mathbf{x} = \mathbf{0}$, which is a clear failure of being one-to-one. So we've shown that T is one-to-one if and only if $A\mathbf{x} = \mathbf{0}$ has no nontrivial solutions.

Theorem 3.8. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. Then T is one-to-one if and only if there is a pivot in every column of A.*

Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by $T(\mathbf{x}) = A\mathbf{x}$. The previous theorem implies that if $n > m$, then T cannot be one-to-one. The theorem before it implies that if $n < m$, then T cannot be onto. So for T to be a one-to-one correspondence, it is necessary that $T : \mathbb{R}^n \to \mathbb{R}^n$, and consequently that A is a square $n \times n$ matrix. Further, an $n \times n$ matrix A has a pivot in every row if and only if it has n pivots if and only if it has a pivot in every column. So $T : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one if and only if it is onto. When either condition holds, T is a one-to-one correspondence and therefore has an inverse transformation $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. The inverse is also linear, so it is natural to inquire about its standard matrix, which will be denoted A^{-1}.

4. Invertible Matrices

An $n \times n$ matrix A is called *invertible* if there is another $n \times n$ matrix A^{-1} with the property that

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

The matrix A^{-1} is called the *inverse* of A. If A is invertible, then its inverse is actually unique, which justifies why we call it *the* inverse of A and also makes the notation A^{-1} well-defined.

Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T(\mathbf{x}) = A\mathbf{x}$ and suppose A is invertible. Then we can define a transformation $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ given by $T^{-1}(\mathbf{x}) = A^{-1}(\mathbf{x})$. We have

$$T^{-1}(T(\mathbf{x})) = A^{-1}(A\mathbf{x}) = I_n\mathbf{x} = \mathbf{x}$$

and

$$T(T^{-1}(\mathbf{x})) = A(A^{-1}\mathbf{x}) = I_n\mathbf{x} = \mathbf{x}$$

for all $\mathbf{x} \in \mathbb{R}^n$. This shows that T and T^{-1} are actually inverse functions, so the inverse matrix corresponds to the inverse linear transformation. Since T has an inverse, it is is a one-to-one correspondence. Consequently the invertible matrix A must have n pivots. The converse holds as well.

Theorem 4.1. *An $n \times n$ matrix A is invertible if and only if A has n pivots.*

Example 4.2. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Since $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$, we see A has only 1 pivot. So A is not invertible.

In general, it is difficult to tell if a matrix is invertible without doing some work (row operations).

Exercise 4.3. Let $A = \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$. Show that one of these matrices is inverible and the other is not invertible.

4.1. **Finding an inverse.** The following theorem gives an algorithm for finding the inverse of an $n \times n$ invertible matrix A.

Theorem 4.4. *If A is an invertible $n \times n$ matrix, then the reduced row echelon form of $\begin{bmatrix} A & I_n \end{bmatrix}$ is $\begin{bmatrix} I_n & A^{-1} \end{bmatrix}$.*

Example 4.5. Let's try it out by finding the inverse of $A = \begin{bmatrix} 3 & 2 \\ 7 & 4 \end{bmatrix}$. We do row reduction to $\begin{bmatrix} A & I_2 \end{bmatrix}$:

$$\begin{aligned}
\begin{bmatrix} 3 & 2 & 1 & 0 \\ 7 & 4 & 0 & 1 \end{bmatrix} &\sim \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & -2/3 & -7/3 & 1 \end{bmatrix} \\
&\sim \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 7/2 & -3/2 \end{bmatrix} \\
&\sim \begin{bmatrix} 3 & 0 & -6 & 3 \\ 0 & 1 & 7/2 & -3/2 \end{bmatrix} \\
&\sim \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 7/2 & -3/2 \end{bmatrix}
\end{aligned}$$

We conclude that $A^{-1} = \begin{bmatrix} -2 & 1 \\ 7/2 & -3/2 \end{bmatrix}$.

Exercise 4.6. Find the inverse of the invertible matrix of the previous exercise.

Exercise 4.7. Show that the inverse of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $A^{-1} = \dfrac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, provided $ad - bc \neq 0$.

Note that if we try this approach and the left half of the RREF is not I_n, then the matrix A does not have n pivots and so A is not invertible.

4.2. **Why does it work?** Let's explain why this algorithm produces the inverse. When looking for the inverse, we are seeking an $n \times n$ matrix X with the property that $AX = I_n$. Write $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$, which is ultimately the inverse of A, but we are thinking of it as one big unknown to solve for now. Note that the columns of I_n are the standard basis vectors, so that $I_n = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix}$. We want

$$\begin{aligned}
AX &= I_n \\
A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix} \\
\begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \dots & A\mathbf{x}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix}.
\end{aligned}$$

This shows that each column $\mathbf{x}_i$ of X must satisfy $A\mathbf{x}_i = \mathbf{e}_i$. So we obtain n linear systems, each of which has n variables:

$$\begin{aligned}
A\mathbf{x}_1 &= \mathbf{e}_1 \\
A\mathbf{x}_2 &= \mathbf{e}_2 \\
&\vdots \\
A\mathbf{x}_n &= \mathbf{e}_n.
\end{aligned}$$

Each $\mathbf{x}_i$ is solved for by forming an augmented matrix and going to RREF:

$$\begin{aligned}
\begin{bmatrix} A & \mathbf{e}_1 \end{bmatrix} &\overset{\text{RREF}}{\sim} \begin{bmatrix} I_n & \mathbf{x}_1 \end{bmatrix} \\
\begin{bmatrix} A & \mathbf{e}_2 \end{bmatrix} &\overset{\text{RREF}}{\sim} \begin{bmatrix} I_n & \mathbf{x}_2 \end{bmatrix} \\
&\vdots \\
\begin{bmatrix} A & \mathbf{e}_n \end{bmatrix} &\overset{\text{RREF}}{\sim} \begin{bmatrix} I_n & \mathbf{x}_n \end{bmatrix}.
\end{aligned}$$

This is not very economical, because we are repeating the work that goes into row reducing A to I_n in each system. So to be more economical, we augment all of the "right hand sides" to A at once to obtain

$$\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} I_n & \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}.$$

But this is exactly the same as writing

$$\begin{bmatrix} A & I_n \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} I_n & X \end{bmatrix}.$$

This justifies the method because the matrix X we are solving for is really the inverse of A (remember it satisfies $AX = I_n$.)

4.3. **Using the inverse to solve $A\mathbf{x} = \mathbf{b}$.** If the inverse of an $n \times n$ matrix A is known, then the linear system $A\mathbf{x} = \mathbf{b}$ can be solved quite efficiently.

Theorem 4.8. *If A is an invertible $n \times n$ matrix, then the system $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.*

It works because we can (left) multiply the original equation $A\mathbf{x} = \mathbf{b}$ to solve for $\mathbf{x}$:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^{-1}A\mathbf{x} &= A^{-1}\mathbf{b} \\ I_n\mathbf{x} &= A^{-1}\mathbf{b} \\ \mathbf{x} &= A^{-1}\mathbf{b}. \end{aligned}$$

Example 4.9. We can solve $\begin{bmatrix} 3 & 2 \\ 7 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 5 \\ -4 \end{bmatrix}$ quickly using the inverse:

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -2 & 1 \\ 7/2 & -3/2 \end{bmatrix} \begin{bmatrix} 5 \\ -4 \end{bmatrix} = \begin{bmatrix} -14 \\ 47/2 \end{bmatrix}$$

Besides being a handy formula for the solution of a linear system, this illustrates a nice property of an invertible matrix A: For any choice of $\mathbf{b}$, the system $A\mathbf{x} = \mathbf{b}$ has a unique solution. The system $A\mathbf{x} = \mathbf{b}$ cannot be inconsistent, and its solution set cannot have a free variable.

5. Determinants

The determinant of a matrix is an important number attached to a matrix. Determinants are defined recursively. For a 2×2 matrix,

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

For example,

$$\det \begin{bmatrix} 3 & 2 \\ -1 & 4 \end{bmatrix} = 3(4) - 2(-1) = 14.$$

For a 3×3 matrix,

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

This expression is called a cofactor expansion along the first row. Note the following features:

- There is one term for each entry in the first row.
- Each term involves a determinant of a smaller matrix. This matrix is obtained by deleting the row an column which contain the corresponding entry from the first row.
- The signs alternate between $+$ and $-$.

By following these rules, we can define determinants of larger matrices. For example,

$$\det \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = a \cdot \det \begin{bmatrix} f & g & h \\ j & k & l \\ n & o & p \end{bmatrix} - b \cdot \det \begin{bmatrix} e & g & h \\ i & k & l \\ m & o & p \end{bmatrix} + c \cdot \det \begin{bmatrix} e & f & h \\ i & j & l \\ m & n & p \end{bmatrix} - d \cdot \det \begin{bmatrix} e & f & g \\ i & j & k \\ m & n & o \end{bmatrix}.$$

The following is one of the most important properties of the determinant. It says that the determinant can be used as an invertibility detector.

Theorem 5.1. *An $n \times n$ matrix A is invertible if and only if* $\det A \neq 0$.

Example 5.2. Let's show that $A = \begin{bmatrix} 2 & 0 & 1 \\ 3 & 1 & 0 \\ -1 & 3 & -5 \end{bmatrix}$ is not invertible. We evaluate

$$\begin{aligned} \det \begin{bmatrix} 2 & 0 & 1 \\ 3 & 1 & 0 \\ -1 & 3 & -5 \end{bmatrix} &= 2 \cdot \det \begin{bmatrix} 1 & 0 \\ 3 & -5 \end{bmatrix} - 0 \cdot \det \begin{bmatrix} 3 & 0 \\ -1 & -5 \end{bmatrix} + 1 \cdot \det \begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix} \\ &= 2(-5-0) - 0 + (9+1) = 0 \end{aligned}$$

Since $\det A = 0$, A is not invertible.

Here are several other properties of determinants.

Theorem 5.3.

(1) $\det(AB) = (\det A)(\det B)$ *for all $n \times n$ matrices A and B.*
(2) $\det I_n = 1$.
(3) $\det(A^{-1}) = \dfrac{1}{\det A}$ *for any invertible A.*
(4) $\det(cA) = c^n A$ *for any scalar c and $n \times n$ matrix A.*
(5) $\det(A^T) = \det A$.

The determinant of an upper or lower triangular matrix is equal to the product of its diagonal entries. In particular, this applies to diagonal matrices. For example,

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} = 1 \cdot 1 \cdot 1 = 1,$$

$$\det \begin{bmatrix} 5 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ -1 & 1 & -2 & 0 \\ 4 & 2 & 9 & 8 \end{bmatrix} = 5 \cdot 0 \cdot (-2) \cdot 8 = 0,$$

$$\det \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix} = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \cdot \ldots \cdot \lambda_n.$$

5.1. **More with cofactors, Cramer's Rule.** For a general $n \times n$ matrix with entries denoted a_{ij}, let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained by deleting row i and column j from A. Then the determinant of A is defined by

$$\det A = a_{11} \cdot \det A_{11} - a_{12} \cdot \det A_{12} + a_{13} \cdot \det A_{13} - \ldots + (-1)^{n+1} a_{1n} \cdot \det A_{1n}.$$

Cofactor expansions can actually be done along any row of our choosing, or down any column of our choosing. To carefully state the result, first defined the (i,j)-cofactor of A to be

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

In this notation,

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n}.$$

Note that the alternating signs are accounted for in the definition of the cofactor.

Theorem 5.4.

(1) *(Cofactor expansion along row i) For any i,*

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}.$$

(2) *(Cofactor expansion down column j) For any j,*

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}.$$

The signs in these expansions are always alternating. One just needs to be careful as to whether it starts with $+$ or $-$.

If A is invertible, Cramer's Rule gives a formula for the solution of $A\mathbf{x} = \mathbf{b}$ in terms of determinants. This can be used to give a formula for the inverse of a matrix, which involves the cofactors. The *adjugate* of A is the transpose of the matrix of cofactors:

$$\operatorname{adj} A = \begin{bmatrix} C_{11} & C_{21} & C_{31} & \dots & C_{n1} \\ C_{12} & C_{22} & C_{32} & \dots & C_{n2} \\ C_{13} & C_{23} & C_{33} & \dots & C_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & C_{3n} & \dots & C_{nn} \end{bmatrix}.$$

Note that the indices on the cofactors is the opposite of our usual convention in which row number comes before column number. This is why we said adj A is the transpose of the matrix of cofactors.

Theorem 5.5. *If A is invertible, then*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

This is an explicit formula for the inverse of a matrix. For example, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$\operatorname{adj} A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(Note that a 1×1 matrix is a scalar, and the determinant of a 1×1 matrix is itself.) The formula recovers the usual formula for the inverse of a 2×2 matrix:

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

Note that the formula looks much messier for larger matrices. One interesting aspect of this formula is that it can be used to compute a single entry of A^{-1} without computing the entire matrix. The (i, j) entry of A^{-1} is $\frac{C_{ji}}{\det A}$.

Example 5.6. Let $A = \begin{bmatrix} 4 & 0 & 1 \\ -1 & 3 & 2 \\ 0 & 2 & 1 \end{bmatrix}$. Let's find the entry of A^{-1} in the middle of the top row, that is, the $(1, 2)$ entry. It is given by $\frac{C_{21}}{\det A}$. We compute

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

and

$$C_{21} = (-1)^{2+1} \det \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = (-1)^3(0 - 2) = 2.$$

So the $(1, 2)$ entry of A^{-1} is

$$\frac{C_{21}}{\det A} = \frac{2}{-6} = -\frac{1}{3}.$$

6. Subspaces of $\mathbb{R}^n$

We won't treat abstract vector spaces here. Informally, a vector space is a set of vectors which is closed under the operations of vector addition and scalar multiplication. That means that the sum of two vectors in the vector space is also in the vector space and also any scalar multiple of a vector in the vector space is also in the vector space. Important examples of vector spaces include $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$, and these are the only examples we shall consider.

Definition 6.1. A subset W of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ if

(1) $\mathbf{0} \in W$,
(2) for every $\mathbf{w}_1, \mathbf{w}_2 \in W$, we have $\mathbf{w}_1 + \mathbf{w}_2 \in W$, and
(3) for every $\mathbf{w} \in W$ and scalar c, we have $c\mathbf{w} \in W$.

Example 6.2. The set $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^3$. You should verify all three conditions.

Example 6.3. The set $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$ is NOT a subspace of $\mathbb{R}^3$. You should verify that all three conditions fail. (It is enough that one fails.)

Theorem 6.4. *Given any set* $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ *of vectors in* $\mathbb{R}^n$*, the set* $W = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ *is a subspace of* $\mathbb{R}^n$*.*

This amounts to verifying the following assertions:

(1) The zero vector $\mathbf{0}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$.
(2) If we add two linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_p$, the result is also a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$.
(3) If we multiply a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$ by a scalar c, the result is also a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$.

Example 6.5. Let's show that $W = \left\{ \begin{bmatrix} s+t \\ 2t \\ s-t \end{bmatrix} : s, t \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^3$. Observe that

$$\begin{bmatrix} s+t \\ 2t \\ s-t \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

This shows that every vector in W is a linear combination of $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$. Conversely, any such linear combination has the form of a vector in W. This shows that

$$W = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \right\}.$$

Since the span of anything is a subspace, we conclude that W is a subspace of $\mathbb{R}^3$.

The previous theorem has the following converse.

Theorem 6.6. *If* W *is a subspace of* $\mathbb{R}^n$*, then there exist vectors* $\mathbf{v}_1, \dots, \mathbf{v}_p$ *in* W *such that* $W = \operatorname{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$*. Further, these vectors can be chosen so that* $p \leq n$*.*

So every subspace of $\mathbb{R}^n$ looks like the span of something. Typical examples of subspaces include lines through the origin (the span of one nonzero vector), planes through the origin (the span of two linearly independent vectors), and their higher dimensional analogs.

6.1. Null space and column space.

To any $m \times n$ matrix A, we can associate certain subspace of $\mathbb{R}^n$ or $\mathbb{R}^m$.

Definition 6.7. The *null space* of an $m \times n$ matrix A is the set of all $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

That is, the null space of A is the solution set of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. The null space is denoted $\operatorname{Nul} A$. For any matrix A, $\operatorname{Nul} A$ is a subspace of $\mathbb{R}^n$. (You should verify the three conditions yourself.) If we want to describe the null space of some specific matrix, it simply amounts to solving the linear system $A\mathbf{x} = \mathbf{0}$.

The zero vector $\mathbf{0}$ is always in $\operatorname{Nul} A$, because it is the trivial solution of $A\mathbf{x} = \mathbf{0}$. A nonzero vector $\mathbf{x}$ is in $\operatorname{Nul} A$ if and only if $\mathbf{x}$ is a nontrivial solution to $A\mathbf{x} = \mathbf{0}$. Recall that nontrivial solutions exist if and only if the columns of A are linearly independent.

Definition 6.8. The *column space* of an $m \times n$ matrix A is the span of its columns.

The column space is denoted $\operatorname{Col} A$. Since $\operatorname{Col} A$ is the span of some collection of vectors in $\mathbb{R}^m$, it follows immediately that $\operatorname{Col} A$ is a subspace of $\mathbb{R}^m$. Since every linear combination of the columns of A has the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, we have

$$\operatorname{Col} A = \{A\mathbf{x} \ : \ \mathbf{x} \in \mathbb{R}^n\}.$$

It follows from this description, that $\operatorname{Col} A$ is the range of the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$. So T is onto if and only if $\operatorname{Col} A = \mathbb{R}^m$.

A vector $\mathbf{b} \in \mathbb{R}^m$ is an element of $\operatorname{Col} A$ if and only if $\mathbf{b}$ is a linear combination of the columns of A. This happens if and only if the linear system $A\mathbf{x} = \mathbf{b}$ is consistent.

One can also speak of $\operatorname{Row} A$, the *row space* of A, which is the subspace of $\mathbb{R}^n$ spanned by the rows of A. However, $\operatorname{Row} A$ naturally identifies with $\operatorname{Col}(A^T)$, so it fits into the discussion of column space. Similarly, one may want to consider $\operatorname{Nul}(A^T)$ (sometimes called the *left null space* of A).

6.2. **Basis and dimension.** A *basis* for a subspace W of $\mathbb{R}^n$ is a set $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ of vectors in W such that

(1) $\operatorname{Span}\{\mathbf{b}_1, \dots, \mathbf{b}_p\} = W$, that is, $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ spans W, and
(2) $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is linearly independent.

Recall that we already remarked that every subspace W of $\mathbb{R}^n$ is the span of some set of vectors in W. If that spanning set is linearly dependent, then one of the vectors is a linear combination of the others (that is, it is already in the span of the others,) and if we remove that vector from the set, the remaining vectors will still span W. In this way, we can keep removing vectors from a spanning set until we obtain a linearly independent spanning set, i.e. a basis. This argument shows that every subspace W has a basis.

A basis for W can be thought of from two perspectives. It is a spanning set for W that is as small as possible, or it is a linearly independent subset of W that is as large as possible. Here small/large refers to the number of vectors in the set.

Theorem 6.9. *Suppose $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for W. Then every $\mathbf{w} \in W$ can be written uniquely as a linear combination of $\mathbf{b}_1, \dots, \mathbf{b}_p$. That is, for every $\mathbf{w} \in W$, there exist unique scalars $c_1, \dots, c_p$ such that*

$$\mathbf{w} = c_1\mathbf{b}_1 + \ldots + c_p\mathbf{b}_p.$$

Note every $\mathbf{w} \in W$ is a linear combination of the basis because the basis spans W. The uniqueness of this linear combination is a result of the linear independence of the basis. So to any $\mathbf{w} \in W$, we can associate the unique coefficients $c_1, \dots, c_p$, which are called the *coordinates* of $\mathbf{w}$ (relative to the given basis). The mapping

$$\mathbf{w} \mapsto \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

is a linear transformation from W to $\mathbb{R}^p$ that is both one-to-one and onto. Such a transformation is called a *linear isomorphism.* A linear isomorphism allows us to think of W as being the "same" as $\mathbb{R}^p$.

Note the number p above refers to the number of vectors in the basis for W. This number is important, and is called the dimension of W.

Definition 6.10. The *dimension* of a subspace W of $\mathbb{R}^n$ is the number of vectors in any basis for W.

It is a fact that any two bases for W must have the same number of elements, so this definition makes sense.

Example 6.11. The standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$ that has exactly n vectors in it, so $\dim \mathbb{R}^n = n$.

Note that any subspace of W of $\mathbb{R}^n$ must have $\dim W \leq n$ (because you can't have more than n linearly independent vectors in $\mathbb{R}^n$).

We saw that a line L through the origin in $\mathbb{R}^3$ is a subspace. Such a line has $\dim L = 1$. For a basis, one can take a single vector that points in the direction of the line.

Let $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. A plane in $\mathbb{R}^3$ such as $W = \operatorname{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$ has $\dim W = 2$. Here, $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an example of a basis. They clearly span W (by definition of W) and they are linearly independent because there are only two of them, and they are not multiples of each other.

There are always infinitely many choices of a basis for a subspace W. Consider the important example of $W = \mathbb{R}^n$. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a collection of n vectors in $\mathbb{R}^n$, and let $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$. Then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ spans $\mathbb{R}^n$ if and only if A has a pivot in every row, and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent if and only if A has a pivot in every column. Since A is $n \times n$, either condition happens if and only if A has exactly n pivots. This happens if and only if A is invertible. Hence we conclude:

Theorem 6.12. *Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a set of n vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$. Then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$ if and only if A is an invertible matrix.*

6.3. **Basis and dimension for null space and column space.** To find a basis for $\operatorname{Nul} A$, the main step is to solve the linear system $A\mathbf{x} = \mathbf{0}$ and express the solutions in parametric vector form. Let's consider the example

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

When solving $A\mathbf{x} = \mathbf{0}$, we form the augmented matrix and reduce to RREF:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 0 \\ 2 & 4 & 1 & 3 & 0 \\ 2 & 4 & 5 & 7 & 0 \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We have $x_1 = -2x_2 - x_4$, $x_3 = -x_4$, and x_2, x_4 are free. The solution in parametric vector form is

$$\mathbf{x} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

This shows that

$$\operatorname{Nul} A = \operatorname{Span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

Since these two vectors are linearly independent, they form a basis for $\operatorname{Nul} A$ and consequently $\dim(\operatorname{Nul} A) = 2$.

In general, to find a basis for $\operatorname{Nul} A$, one solves $A\mathbf{x} = \mathbf{0}$ in parametric vector form. The solution will look like a linear combination of some vectors, and those vectors will always be a basis for $\operatorname{Nul} A$. The fact that the solution is the set of all linear combinations of them shows that they span $\operatorname{Nul} A$. The fact that they are linearly independent is a consequently of reduced row echelon form. Note that there will be one vector in this basis for each free variable in the solution. The number of free variables in A is the difference of the number of columns of A minus the number of pivots in A (the number of pivots is $\operatorname{rank} A$). So if A is $m \times n$, then

$$\dim(\operatorname{Nul} A) = n - \operatorname{rank} A.$$

Now let's consider the column space of A. Suppose R is a row echelon form of A, so that $A \sim R$. In general, $\operatorname{Col} A \neq \operatorname{Col} R$. However,

$$A\mathbf{x} = \mathbf{0} \qquad \Leftrightarrow R\mathbf{x} = \mathbf{0}.$$

(The whole point of row operations is so that the previous line holds.) This means that the columns of A have the exact same linear dependence relations as the columns of R. Here is an example:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 2 & 4 & 5 & 7 \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R.$$

For example, notice that the 4th column of R is the sum of its first column and its third column. It follows that the same statement holds for the original A (check it!) We see from inspection of the columns of R that the first and third columns serve as a basis for $\operatorname{Col} R$ (they are linearly independent and all other columns can be written as a linear combination of them.) It follows that the same holds for the columns of A: the first and third columns of A are a basis for $\operatorname{Col} A$:

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} \right\} \text{ is a basis for } \operatorname{Col} A.$$

Consequently $\dim(\operatorname{Col} A) = 2$.

In general, the columns of R which contain the pivots will form a basis for $\operatorname{Col} R$. It follows that in general, the columns of A which correspond to the pivot columns will form a basis for $\operatorname{Col} A$. The number of vectors in the basis will coincide with the number of pivots, so

$$\dim(\operatorname{Col} A) = \operatorname{rank} A.$$

Combining this with the above result about $\operatorname{Nul} A$ shows that

$$\dim(\operatorname{Nul} A) + \dim(\operatorname{Col} A) = n \qquad (= \#\text{ columns of } A).$$

7. Theorems that Unite the Concepts

You may have noticed several concepts in linear algebra seem to occur in pairs, such as:

existence of solutions	vs.	uniqueness of solutions
span	vs.	linear independence
onto	vs.	one-to-one
column space	vs.	null space

In general, the concepts on the left and independent of the concepts on the right. For example, onto does not imply one-to-one, or vice-versa. However the concepts on the left are all related to each other, and the concepts on the right are all related to each other. Here are theorems illustrating these relationships.

Theorem 7.1. *Let A be an $m \times n$ matrix with columns $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following statements are equivalent.*

(1) For every $\mathbf{b} \in \mathbb{R}^m$, the linear system $A\mathbf{x} = \mathbf{b}$ is consistent.
(2) The columns of A span $\mathbb{R}^m$.
(3) The linear transformation T maps onto $\mathbb{R}^m$.
(4) $\operatorname{Col} A = \mathbb{R}^m$.
(5) $\operatorname{rank} A = m$.
(6) There is a pivot in every row of A.

In the case where one of these statements is true (hence all of them are true), it must be the case that $m \le n$.

Theorem 7.2. *Let A be an $m \times n$ matrix with columns $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following statements are equivalent.*

(1) The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(2) The columns of A are linearly independent.
(3) The linear transformation T is one-to-one.
(4) $\operatorname{Nul} A = \{\mathbf{0}\}$.
(5) $\operatorname{rank} A = n$.
(6) There is a pivot in every column of A.

In the case where one of these statements is true (hence all of them are true), it must be the case that $m \ge n$.

Now if A is an $n \times n$ matrix, then A has a pivot in every row if and only if A has a pivot in every column. This happens if and only if A is invertible. So in the square matrix case, all of the conditions in the previous two theorems are equivalent to each other, and also equivalent to the condition that A is invertible. We state this as one gigantic theorem.

Theorem 7.3 (Invertible Matrix Theorem)**.** *Let A be an $n \times n$ matrix with columns $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = A\mathbf{x}$. The following statements are equivalent.*

(1) *A is invertible.*
(2) *A has n pivots (that is, $\operatorname{rank} A = n$).*
(3) *For every $\mathbf{b} \in \mathbb{R}^n$, the linear system $A\mathbf{x} = \mathbf{b}$ is consistent.*
(4) *The columns of A span $\mathbb{R}^n$.*
(5) *The linear transformation T maps onto $\mathbb{R}^n$.*
(6) $\operatorname{Col} A = \mathbb{R}^n$.
(7) $\dim(\operatorname{Col} A) = n$.
(8) *The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(9) *The columns of A are linearly independent.*
(10) *The linear transformation T is one-to-one.*
(11) $\operatorname{Nul} A = \{\mathbf{0}\}$.
(12) $\dim(\operatorname{Nul} A) = 0$.
(13) *For every $\mathbf{b} \in \mathbb{R}^n$, the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution.*
(14) *The columns of A form a basis for $\mathbb{R}^n$.*
(15) *The transformation T is a linear isomorphism (i.e. T is both one-to-one and onto.)*
(16) $\det A \neq 0$.

8. Inner Product and Orthogonality

The *dot product* (also called *inner product* or *scalar product*) of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is defined by $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w}$. In terms of the entries of the vectors:

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = v_1w_1 + v_2w_2 + \dots + v_nw_n = \sum_{i=1}^{n} v_iw_i.$$

The *length* (or *norm*) of a vector $\mathbf{v} \in \mathbb{R}^n$ is related to the dot product, and is given by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}.$$

When a vector is visualized geometrically as an arrow, the norm is exactly the length of the arrow.

A *unit vector* is a vector $\mathbf{u}$ such that $\|u\| = 1$. Given any nonzero vector $\mathbf{v}$, the vector $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is a unit vector that points in the same direction as $\mathbf{v}$.

Although it seems very arithmetic, the dot product has a lot to do with geometry. The following is a geometric interpretation for the dot product of two vectors.

Theorem 8.1. *Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Then*

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \cos\theta,$$

where θ is the angle between the vectors $\mathbf{v}$ and $\mathbf{w}$. In particular, we have

$$\theta = \cos^{-1}\left(\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \, \|\mathbf{w}\|}\right).$$

The sign of the dot product gives information about the type of angle between two nonzero vectors. Indeed, if $\mathbf{v}$ and $\mathbf{w}$ are nonzero vectors, then $\|\mathbf{v}\|$ and $\|\mathbf{w}\|$

are positive. It follows that the sign of $\mathbf{v} \cdot \mathbf{w}$ is the same as the sign of $\cos\theta$. We conclude the following:

- $\mathbf{v} \cdot \mathbf{w} > 0$ if and only if $0 \leq \theta < \pi/2$.
- $\mathbf{v} \cdot \mathbf{w} = 0$ if and only if $\theta = \pi/2$.
- $\mathbf{v} \cdot \mathbf{w} < 0$ if and only if $\pi/2 < \theta \leq \pi$.

The second statement is important because it isolates the special case in which there is a right angle between two vectors. Right angles can easily be detected with a dot product computation. We say $\mathbf{v}$ and $\mathbf{w}$ are *orthogonal* if $\mathbf{v} \cdot \mathbf{w} = 0$.

8.1. **Orthogonal/Orthonormal bases.** A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of vectors in $\mathbb{R}^n$ is called an *orthogonal set* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for all $i \neq j$. That is, any two distinct vectors chosen from the set are orthogonal to each other. A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ of vectors in $\mathbb{R}^n$ is called an *orthonormal set* if it is an orthogonal set, and additionally each $\mathbf{u}_i$ is a unit vector. An *orthogonal basis* (resp. *orthonormal basis*) for a subspace W of $\mathbb{R}^n$ simply means a basis that is an orthogonal set (resp. orthonormal set).

A nice feature of orthogonality is that it implies linear independence.

Proposition 8.2. *Any orthogonal set of nonzero vectors in $\mathbb{R}^n$ is linearly independent.*

The word nonzero needs to be included because the zero vector is orthogonal to every vector: $\mathbf{0} \cdot \mathbf{v} = 0$. So it is possible to have an orthogonal set that contains the zero vector (and that wouldn't be linearly independent).

So when considering whether a set is an orthogonal/orthonormal basis for a subspace, linear independence becomes a non-issue.

Corollary 8.3. *Suppose W is a subspace of $\mathbb{R}^n$ and $\dim W = p$.*

(1) Any orthogonal set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of p nonzero vectors in W is an orthogonal basis for W.
(2) Any orthonormal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ of p vectors in W is an orthonormal basis for W.

Note that in the second statement, we do not need to insist that the vectors are nonzero because they are assumed to be unit vectors (hence are nonzero).

Given a basis for a subspace W, every vector in W can be written uniquely as a linear combination of that basis. When the basis is orthogonal/orthonormal, it is much easier to find the coefficients in the linear combination.

Theorem 8.4. *Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w} \in W$. Then*

$$\mathbf{w} = c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p \quad \textit{where} \quad c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}.$$

So the coefficients can be easily computed via dot product computations. Note that the denominators are 1 is the special case of an orthonormal basis. So if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for W and $\mathbf{w} \in W$, then we have

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_1 + \ldots + (\mathbf{w} \cdot \mathbf{u}_p)\mathbf{u}_p.$$

Suppose now that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is any set of vectors in $\mathbb{R}^n$, and let $A = \begin{bmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_p\end{bmatrix}$. Consider the matrix product A^TA, which is a $p \times p$ matrix. The interesting observation is that each entry of A^TA can be interpreted as a dot product of vectors in our original set:

$$A^T A = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_p^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \dots & \mathbf{v}_1 \cdot \mathbf{v}_p \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \dots & \mathbf{v}_2 \cdot \mathbf{v}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_p \cdot \mathbf{v}_1 & \mathbf{v}_p \cdot \mathbf{v}_2 & \dots & \mathbf{v}_p \cdot \mathbf{v}_p \end{bmatrix}.$$

From this, the following is immediate.

Theorem 8.5. *Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix}$.*

(1) The columns of A are an orthogonal set if and only if $A^T A$ is a diagonal matrix.
(2) The columns of A are an orthonormal set if and only if $A^T A = I_p$.

Definition 8.6. An $n \times n$ matrix Q is called an *orthogonal matrix* if its columns are an orthonormal set.

We immediately note that the terminology is bad. (It would make more sense to call it an orthonormal matrix, but people have been using the terminology for many years.) Also take note that we are currently only discussing square matrices.

Theorem 8.7. *Let Q denote an $n \times n$ matrix. The following are equivalent.*

(1) Q is an orthogonal matrix.
(2) $Q^T Q = I_n$.
(3) $Q^{-1} = Q^T$.
(4) The columns of Q are an orthonormal basis for $\mathbb{R}^n$.
(5) The rows of Q are an orthonormal basis for $\mathbb{R}^n$.

Note that the statement about rows follows from the fact that Q^T is also an orthogonal matrix. One of the nice features of an orthogonal matrix Q is that it is easy to write down its inverse, since $Q^{-1} = Q^T$.

Exercise 8.8. Every orthogonal 2×2 matrix has one of the two following forms:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

Verify all conditions in the previous theorem for these matrices.

As linear transformations, orthogonal matrices correspond to rotations and reflections. In fact, every orthogonal Q satisfies $\det Q = \pm 1$. The orthogonal matrices with determinant 1 are the rotations and the orthogonal matrices with determinant -1 are the reflections. Note that the two families of 2×2 matrices in the previous exercise are precisely the rotations and reflections in $\mathbb{R}^2$.

8.2. **Orthogonal complements.** Given a subspace W of $\mathbb{R}^n$, we say that a vector $z \in \mathbb{R}^n$ is orthogonal to W if $\mathbf{z} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in W$. The *orthogonal complement* of W is the set $W^\perp$ of all vectors in $\mathbb{R}^n$ that are orthogonal to W.

Example 8.9. If W is a line through the origin in $\mathbb{R}^2$, then $W^\perp$ will be the line through the origin that is perpendicular to W.

Example 8.10. If W is a plane through the origin in $\mathbb{R}^3$, then $W^\perp$ is the line through the origin that makes a right angle with W (often called a normal line to the plane).

Example 8.11. Recall that $\{\mathbf{0}\}$ and $\mathbb{R}^n$ are examples of subspaces of $\mathbb{R}^n$. They are actually orthogonal complements:

$$\{\mathbf{0}\}^\perp = \mathbb{R}^n \qquad \text{and} \qquad (\mathbb{R}^n)^\perp = \{\mathbf{0}\}.$$

The first says that every vector in $\mathbb{R}^n$ is orthogonal to the zero vector. The second says that the *only* vector in $\mathbb{R}^n$ that is orthogonal to *every* vector in $\mathbb{R}^n$ is the zero vector.

Proposition 8.12. *Let W be a subspace of $\mathbb{R}^n$.*

(1) $W^\perp$ is a subspace of $\mathbb{R}^n$.
(2) $(W^\perp)^\perp = W$.
(3) $\dim W + \dim(W^\perp) = n$.
(4) A vector $\mathbf{z}$ is in $W^\perp$ if and only if $\mathbf{z}$ is orthogonal to each vector in a spanning set for W.

If A is an $m \times n$ matrix, it is natural to consider the orthogonal complements of $\operatorname{Nul} A$ and $\operatorname{Col} A$. Note that these two spaces will not be orthogonal complements to each other in general, because $\operatorname{Nul} A$ is a subspace of $\mathbb{R}^n$ and $\operatorname{Col} A$ is a subspace of $\mathbb{R}^m$.

Theorem 8.13. *Let A be an $m \times n$ matrix. Then*

$$(\operatorname{Col} A)^\perp = \operatorname{Nul}(A^T) \qquad \textit{and} \qquad (\operatorname{Row} A)^\perp = \operatorname{Nul} A.$$

Note that the first assertion is equivalent to the second (we just have A replaced with A^T). To understand the second, just note that the equation $A\mathbf{x} = \mathbf{0}$ can be reinterpreted as the assertion that $\mathbf{x}$ is orthogonal to each row of A. (Each entry in $A\mathbf{x}$ is a dot product of a row of A with $\mathbf{x}$.)

8.3. **Orthogonal projection.** Orthogonal projection is based on the following theorem.

Theorem 8.14 (Orthogonal Decomposition Theorem)**.** *Let W be a subspace of $\mathbb{R}^n$. Then every vector $\mathbf{v} \in \mathbb{R}^n$ can be written uniquely as*

$$\mathbf{v} = \mathbf{w} + \mathbf{z}, \qquad \textit{where } \mathbf{w} \in W \textit{ and } \mathbf{z} \in W^\perp.$$

The unique vector $\mathbf{w} \in W$ from this theorem that is associated to a given $\mathbf{v} \in \mathbb{R}^n$ is called the *orthogonal projection of* $\mathbf{v}$ *onto* W, and is denoted $\mathbf{w} = \operatorname{proj}_W \mathbf{v}$. Visually, the orthogonal projection of $\mathbf{v}$ onto W looks like a shadow of $\mathbf{v}$ in W.

Since $(W^\perp)^\perp = W$, it follows that the vector $\mathbf{z} \in W^\perp$ from the theorem is $\mathbf{z} = \operatorname{proj}_{W^\perp} \mathbf{v}$. Consequently, every $\mathbf{v} \in \mathbb{R}^n$ can be written as

$$\mathbf{v} = \operatorname{proj}_W \mathbf{v} + \operatorname{proj}_{W^\perp} \mathbf{v}.$$

To compute an orthogonal projection, one typically uses the following.

Theorem 8.15. *Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then for any $\mathbf{v} \in \mathbb{R}^n$,*

$$\operatorname{proj}_W \mathbf{v} = \left(\frac{\mathbf{v} \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\right)\mathbf{w}_1 + \ldots + \left(\frac{\mathbf{v} \cdot \mathbf{w}_p}{\mathbf{w}_p \cdot \mathbf{w}_p}\right)\mathbf{w}_p.$$

It is crucial that an *orthogonal* basis for W is used, but it doesn't matter which orthogonal basis is chosen. Recall that the orthogonal decomposition theorem tells us there is a unique way to decompose $\mathbf{v} = \mathbf{w} + \mathbf{z}$, where $\mathbf{w} \in W$ and $\mathbf{z} \in W^\perp$. The previous result explicitly gives a way to compute $\mathbf{w} = \operatorname{proj}_W \mathbf{v}$. Once $\mathbf{w}$ is found, $\mathbf{z} = \operatorname{proj}_{W^\perp} \mathbf{v}$ can be easily found as $\mathbf{z} = \mathbf{v} - \mathbf{w}$.

8.4. **Gram-Schmidt process.** To compute an orthogonal projection onto W, it is necessary to first find an orthogonal basis for W. This raises a natural question: how can we find an orthogonal basis for a given subspace? Often, we have methods for producing a basis for a subspace (for example, this was discussed earlier for $\operatorname{Nul} A$ and $\operatorname{Col} A$), but those methods may not result in an orthogonal basis. So we will pose the following: Given a basis $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$ for a subspace W of $\mathbb{R}^n$, how can we construct an orthogonal basis for W? The answer the is Gram-Schmidt process.

Theorem 8.16 (Gram-Schmidt Process)**.** *Let $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ be a basis for a subspace W of $\mathbb{R}^n$. Iteratively define the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ by the following formulas:*

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \left(\frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 - \dots - \left(\frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\right)\mathbf{v}_{p-1}
\end{aligned}$$

The $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W.

Let us briefly explain these formulas and why they work. Notice that the right hand said of each equation looks like $\mathbf{x}_k$ minus a projection. For example, $\mathbf{v}_2$ is $\mathbf{x}_2$ minus the projection of $\mathbf{x}_2$ onto $\mathbf{v}_1$, and $\mathbf{v}_3$ is $\mathbf{x}_3$ minus the projection of $\mathbf{x}_3$ onto the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. More precisely, let W_k denote the subspace spanned by the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. Then the k-th Gram-Schmidt formula is

$$\mathbf{v}_k = \mathbf{x}_k - \operatorname{proj}_{W_{k-1}} \mathbf{x}_k.$$

But based on the previous discussion, this is just

$$\mathbf{v}_k = \operatorname{proj}_{W_{k-1}^{\perp}} \mathbf{x}_k,$$

which means $\mathbf{v}_k \in W_{k-1}^{\perp} = \operatorname{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}\}^{\perp}$, and so we immediately see that $\mathbf{v}_k$ is orthogonal to each of the previously constructed $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$.

The Gram-Schmidt process takes a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a subspace W of $\mathbb{R}^n$ and produces an orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ for that same subspace. If we want an *orthonormal* basis for W, we just rescale each vector in the orthogonal basis that they become unit vectors. (Note that stretching vectors does not change the angle between vectors.) That is, if we define $\mathbf{u}_k = \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k$ for each k, then $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for W.

9. Eigenvalues and Eigenvectors

The discussion of eigenvectors and eigenvalues is specific to square matrices. Let A denote an $n \times n$ matrix. We shall be interested in situations in which the equation

$$A\mathbf{v} = \lambda \mathbf{v}$$

is satisfied. Here $\mathbf{v} \in \mathbb{R}^n$ and λ is a scalar. This says that when we multiply the vector $\mathbf{v}$ by A, the result is the same as if we just stretched $\mathbf{v}$ by a factor of λ.

Here is a little bit of motivation for why one would consider this. Imagine you (somewhat randomly) wrote down a 2×2 matrix A, and you want to view it as the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2, T(\mathbf{x}) = A\mathbf{x}$ of the plane. How could you precisely attempt to describe how your transformation moves vectors around the plane? Certainly the image $T(\mathbf{x})$ depends very much on which $\mathbf{x}$ was chosen. For some $\mathbf{x}$, it might look like $\mathbf{x}$ was rotated counterclockwise by 72 degrees and then stretched by a factor of 1.2. For other $\mathbf{x}$, it may look like $\mathbf{x}$ was rotated by 20 degrees clockwise and stretched by a factor of 1.8. A whole continuum of different possibilities occurs in between. It looks like total chaos! How can you possibly attempt to explain to someone else what this transformation does?

Then you notice that you found a vector $\mathbf{v}$ for which $T(\mathbf{v})$ points in the same direction as $\mathbf{v}$, but is twice as long. That is, $T(\mathbf{v}) = 2\mathbf{v}$. This can be concisely described to someone else. Say you also find a vector $\mathbf{w}$ for which $T(\mathbf{w}) = (1/3)\mathbf{w}$. So now you can describe this transformation by saying "T is the linear transformation which stretches $\mathbf{v}$ by 2 and shrinks $\mathbf{w}$ by a factor of 1/3. It turns out that this is a complete description of T, because $\{\mathbf{v}, \mathbf{w}\}$ is a basis for $\mathbb{R}^2$ and any linear transformation is completely determined by what it does to any given basis. More precisely any $\mathbf{x} \in \mathbb{R}^2$ could be written as

$$\mathbf{x} = c\mathbf{v} + d\mathbf{w}$$

for unique scalars c, d. Then by linearity, we could compute

$$T(\mathbf{x}) = T(c\mathbf{v} + d\mathbf{w}) = cT(\mathbf{v}) + dT(\mathbf{w}) = c(2\mathbf{v}) + d(1/3\mathbf{w}) = 2c\mathbf{v} + \frac{d}{3}\mathbf{w}.$$

Definition 9.1. Let A be an $n \times n$ matrix. An *eigenvector* for A is a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ. The scalar λ is called an *eigenvalue* for A.

Eigenvectors and eigenvalues occur in pairs. Each eigenvector has an associated eigenvalue. Conversely, when we say that λ is an eigenvalue for A, then we mean that there exists at least one eigenvector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. We insist that an eigenvector $\mathbf{v}$ is nonzero because the zero vector trivially satisfies this equation for all possible λ. It turns out that each A has very few eigenvalues, and so they are quite special.

9.1. **Finding eigenvalues and eigenvectors.** To find eigenvalues/eigenvectors, we need to be able to solve the equation $A\mathbf{x} = \lambda\mathbf{x}$ for both $\mathbf{x}$ and λ. In practice, we usually find the eigenvalues first, and then solve this equation for $\mathbf{x}$. The equation has two occurrences of $\mathbf{x}$, so we do some algebra to fix this:

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x} \\ A\mathbf{x} - \lambda\mathbf{x} &= \mathbf{0} \\ A\mathbf{x} - \lambda I_n\mathbf{x} &= \mathbf{0} \\ (A - \lambda I_n)\mathbf{x} &= \mathbf{0}. \end{aligned}$$

So we see that $A\mathbf{x} = \lambda\mathbf{x}$ if and only if $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$. The latter is a homogeneous linear system, which always has the trivial solution. But by definition, an eigenvector is not the zero vector, so an eigenvector is a nontrivial solution. So we see

$$\begin{aligned}\lambda \text{ is an eigenvalue for } A &\iff (A-\lambda I_n)\mathbf{x}=\mathbf{0} \text{ has nontrivial solutions}\\ &\iff (A-\lambda I_n) \text{ is not invertible}\\ &\iff \det(A-\lambda I_n)=0.\end{aligned}$$

The beauty of this is that we now have a way to find eigenvalues, and it doesn't involve the actual eigenvectors at all.

Theorem 9.2. *Let A be an $n \times n$ matrix. Then λ is an eigenvalue for A if and only if* $\det(A - \lambda I_n) = 0$.

The equation $\det(A - \lambda I_n) = 0$ is called the *characteristic equation* of A. It turns out that the characteristic equation is always a polynomial equation of degree n. The expression $\det(A - \lambda I_n)$ is a degree n polynomial in λ, which is referred to as the *characteristic polynomial* of A. To find the eigenvalues of A, we solve the characteristic equation. Each eigenvalue of A is a root (i.e. a zero) of the characteristic polynomial. Since a degree n polynomial has at most n roots, we conclude that an $n \times n$ matrix A has at most n eigenvalues.

Once we know a specific eigenvalue λ for A, we can now search for eigenvectors. Since $A\mathbf{x} = \lambda\mathbf{x}$ is equivalent to $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$, the eigenvectors are precisely the nontrivial solutions of the homogeneous linear system $(A-\lambda I_n)\mathbf{x} = \mathbf{0}$. As with any other linear system, we can solve via row reduction and present the solutions in parametric vector form (there must be at least one free variable if λ is an eigenvalue). The solution set of $(A-\lambda I_n)\mathbf{x} = \mathbf{0}$ is called an *eigenspace* of A. Each eigenvalue of A gets its own eigenspace. By definition, the eigenspace for λ is equal to $\mathrm{Nul}(A-\lambda I_n)$. This proves that the eigenspace is a subspace of $\mathbb{R}^n$. If we want a basis for an eigenspace, we just treat it as a null space and find a basis the same way we find a basis for a null space.

As a set, the eigenspace of A for λ consists of all eigenvectors for A whose eigenvalue is λ, and also the zero vector (remember we refuse the call $\mathbf{0}$ an eigenvector, but it is in the eigenspace). This can make it awkward to reap the benefits of the fact that the eigenspace is a subspace. Here are things that are true:

- If $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, and c is a nonzero scalar, then $c\mathbf{v}$ is an eigenvector of A with the same eigenvalue λ.
- If $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors of A with the same eigenvalue λ, and if $\mathbf{v} \neq -\mathbf{w}$, then the sum $\mathbf{v} + \mathbf{w}$ is also an eigenvector of A with the same eigenvalue λ.

The first point says we can stretch an eigenvector and the result is also an eigenvector. Thus we immediately see that any eigenvalue of A always has infinitely many eigenvectors. We typically measure the "amount" of eigenvectors by the dimension of the eigenspace.

When you mix eigenvectors from different eigenspaces, you will not get eigenvectors. Sometimes students mistakenly think that the sum of two eigenvectors with dinstinct eigenvalues will also be an eigenvector. This is not the case. Suppose $A\mathbf{v} = \lambda\mathbf{v}$ and $A\mathbf{w} = \mu\mathbf{w}$. Then

$$A(\mathbf{v}+\mathbf{w}) = A\mathbf{v} + A\mathbf{w} = \lambda\mathbf{v} + \mu\mathbf{w} \neq (\lambda+\mu)(\mathbf{v}+\mathbf{w}).$$

The vector $\lambda\mathbf{v} + \mu\mathbf{w}$ is not a scalar multiple of $\mathbf{v} + \mathbf{w}$. Here is something nice that is true about eigenvectors with different eigenvalues.

Proposition 9.3. *Let A be an $n \times n$ matrix and suppose $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are eigenvectors for A with* distinct *eigenvalues $\lambda_1, \ldots, \lambda_p$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set in $\mathbb{R}^n$.*

This gives a second explanation for why an $n \times n$ matrix has at most n eigenvalues. If it had more, then you would be able to find a linearly independent set of more than n eigenvectors in $\mathbb{R}^n$, which is contradictory.

Example 9.4. Consider the matrix $A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$. Let's first find the eigenvalues of A by solving the characteristic equation

$$\begin{aligned} \det(A - \lambda I_2) &= 0 \\ \det\left(\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) &= 0 \\ \det \begin{bmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{bmatrix} &= 0 \\ (1-\lambda)(4-\lambda) - 1(-2) &= 0 \\ \lambda^2 - 5\lambda + 6 &= 0 \\ (\lambda - 2)(\lambda - 3) &= 0 \\ \lambda &= 2, 3. \end{aligned}$$

So the eigenvalues of A are 2 and 3. For each eigenvalue, we will describe the eigenspace. Let's first consider $\lambda = 2$. We solve the system

$$\begin{aligned} (A - 2I_2)\mathbf{x} &= \mathbf{0} \\ \left(\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\mathbf{x} &= \mathbf{0} \\ \begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix}\mathbf{x} &= \mathbf{0}, \end{aligned}$$

which has augmented matrix

$$\begin{bmatrix} -1 & 1 & 0 \\ -2 & 2 & 0 \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So x_2 is free and $x_1 = x_2$. So every $\mathbf{x}$ in the eigenspace has the form

$$\mathbf{x} = \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

So $\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ is a basis for the $\lambda = 2$ eigenspace, which is 1-dimensional. So $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector, and so is every (nonzero) scalar multiple of it. Let's verify it is an eigenvector:

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

Now let's consider the eigenvalue $\lambda = 3$. We solve the system

$$\begin{aligned}(A - 3I_2)\mathbf{x} &= \mathbf{0}\\ \left(\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\mathbf{x} &= \mathbf{0}\\ \begin{bmatrix} -2 & 1 \\ -2 & 1 \end{bmatrix}\mathbf{x} &= \mathbf{0},\end{aligned}$$

which has augmented matrix

$$\begin{bmatrix} -2 & 1 & 0 \\ -2 & 1 & 0 \end{bmatrix} \overset{\text{RREF}}{\sim} \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So x_2 is free and $x_1 = (1/2)x_2$. So every $\mathbf{x}$ in the eigenspace has the form

$$\mathbf{x} = \begin{bmatrix} (1/2)x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}.$$

So $\left\{ \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \right\}$ is a basis for the $\lambda = 2$ eigenspace, which is also 1-dimensional. Let's verify that $\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$ is an eigenvector:

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

If you don't like fractions, you can consider $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ as a basis for the eigenspace (scalar multiple of an eigenvector is an eigenvector). Let's verify that this really is an eigenvector as well:

$$\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Proposition 9.5. *Suppose A is an upper (or lower) triangular $n \times n$ matrix. Then the eigenvalues of A are precisely the diagonal entries of A.*

This follows because the determinant of a triangular matrix is the product of its diagonal entries. For example, when we solve the characteristic equation for

$$A = \begin{bmatrix} 6 & 11 & -1 \\ 0 & 4 & 44 \\ 0 & 0 & -1 \end{bmatrix},$$

we get

$$\begin{aligned}\det(A - \lambda I_3) &= 0\\ \det\begin{bmatrix} 6-\lambda & 11 & -1 \\ 0 & 4-\lambda & 44 \\ 0 & 0 & -1-\lambda \end{bmatrix} &= 0\\ (6-\lambda)(4-\lambda)(-1-\lambda) &= 0\\ \lambda &= 6, 4, -1.\end{aligned}$$

In particular, the eigenvalues of a diagonal matrix are its diagonal entries.

9.2. **Multiplicity and dimension.** It is possible for a number to be a repeated root of a polynomial equation. For eigenvalues, this leads to the notion of *multiplicity*. When $(\lambda - a)^k$ appears in the factorization of the characteristic polynomial, we see that $\lambda = a$ is an eigenvalue with multiplicity k. For example, say we computed the characteristic equation of a 6×6 matrix, factored it up and got

$$\lambda^2(\lambda+4)(\lambda-7)^3 = 0.$$

The only solutions are $\lambda = 0, -4, 7$, but they appear with different multiplicities. We see $\lambda = 0$ has multiplicity 2, $\lambda = -4$ has multiplicity 1, and $\lambda = 7$ has multiplicity 3. Every $n \times n$ matrix will have n (possibly complex) eigenvalues, when they are counted with multiplicity. In this example, it is typical to say that the eigenvalues are

$$\lambda = 0, 0, -4, 7, 7, 7.$$

Note that the sum of the multiplicities is always n. In this case, $2 + 1 + 3 = 6$.

Theorem 9.6. *Suppose λ is an eigenvalue for A with multiplicity m, and let V_λ denote the corresponding eigenspace. Then*

$$1 \leq \dim V_\lambda \leq m.$$

The most typical thing that happens is that an eigenvalue has multiplicity 1. This result implies that in this case, the eigenspace *must* be 1-dimensional (a line). Higher dimensional eigenspaces can only happen in the presence of nontrivial multiplicity. However, the dimension of the eigenspace does not necessarily equal the multiplicity. It can be that $\dim V_\lambda < m$ when $m \geq 2$.

Exercise 9.7. For each of the following matrices, determine all eigenvalues and multiplicities, and determine the dimensions of the eigenspaces.

$$\begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix}, \qquad \begin{bmatrix} 7 & 1 & 0 \\ 0 & 7 & 1 \\ 0 & 0 & 7 \end{bmatrix}, \qquad \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 1 \\ 0 & 0 & 7 \end{bmatrix}.$$

9.3. **Diagonalization.**

Definition 9.8. An $n \times n$ matrix A is called *diagonalizable* if there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$.

The factorization $A = PDP^{-1}$ is called a diagonalization of A. To understand this better, we can (right-)multiply both sides of the equation $A = PDP^{-1}$ by P to obtain $AP = PD$. This equation shows that eigenvalues and eigenvectors immediately get involved.

Lemma 9.9. *Let A, P, D be $n \times n$ matrices, and suppose D is diagonal. Write*

$$P = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Then $AP = PD$ if and only if $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for each $i = 1, \dots, n$.

Proof. By definition of matrix multiplication, we have

$$AP = A\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \dots & A\mathbf{v}_n \end{bmatrix}$$

and

$$PD = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \dots & \lambda_n\mathbf{v}_n \end{bmatrix}.$$

So $AP = PD$ if and only if they are equal column-by-column, which happens if and only if $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for all i. □

So the columns of P must be eigenvectors for A and the diagonal entries of D are the corresponding eigenvalues. This lemma does not require that P be invertible, but the definition of diagonalizable does. The impact of this is that the columns of P must form a basis for $\mathbb{R}^n$. The next theorem follows.

Theorem 9.10. *An $n \times n$ matrix A is diagonalizable if and only if there is a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors for A. In this case, $A = PDP^{-1}$ where $P = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ has the eigenvectors as its columns and $D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ has the corresponding eigenvalues as its entries.*

Example 9.11. We saw that the matrix $A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$ has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 3$ with corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$. So we can diagonalize $A = PDP^{-1}$ with

$$P = \begin{bmatrix} 1 & 1/2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Note that $P^{-1} = \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}$. We can confirm directly that

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

Any basis of eigenvectors will work for the columns of P. In the previous example, we could replace the second column with the eigenvector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Note that this will change P^{-1} as well. The order of the columns of P does not matter, provided that it corresponds exactly to the order of the diagonal entries of D. For example, we could also diagonalize the matrix in the previous example using

$$P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

Diagonalization fails when we "don't have enough eigenvectors". We always have infinitely many eigenvectors, so what we mean by this is that we don't have enough dimensions of eigenspaces. We need n dimensions to make a basis with n eigenvectors. Recall that the dimension of each eigenspace is bounded by the multiplicity of the eigenvalue, and the sum of the multiplicities is n. This leads to the following important characterization of diagonalizability.

Theorem 9.12. *An $n \times n$ matrix A is diagonalizable if and only if the dimension of each eigenspace equals the multplicity of the corresponding eigenvalue.*

Recall that the dimension of each eigenspace is less than or equal to the multiplicity. We need them to be equal for diagonalizability. If the dimension of one eigenspace is strictly less than its multiplicity, then diagonalization cannot be done.

Corollary 9.13. *If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.*

The previous result is nice, because it is what usually happens for a matrix. Here, the multiplicities are all trivial (they are all 1). This tells us that diagonalization fails, then there must be some nontrivial multiplicity. However, diagonalization can still sometimes succeed in the presence of nontrivial multiplicity. One just needs to manually inspect the eigenspaces to see if the dimensions are as large as they need to be.

Exercise 9.14. Consider the matrices

$$A = \begin{bmatrix} 3 & -1 & -1 \\ -2 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & -4 \\ 4 & 3 & 8 \\ 0 & 0 & 3 \end{bmatrix}.$$

For both matrices, the eigenvalues are $\lambda = 1$ (multiplicity 1) and $\lambda = 3$ (multiplicity 2). Show that one of these matrices is diagonalizable and the other isn't. Diagonalize the one for which it is possible to do so.

10. Symmetric Matrices and the Spectral Theorem

Let's mix in some orthogonality with our eigenvectors. When an $n \times n$ matrix A is diagonalizable, there is a basis of eigenvectors for A. What if there is an *orthogonal* basis for $\mathbb{R}^n$ consisting of eigenvectors for A? In this case, we could rescale the eigenvectors to obtain an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of eigenvectors for A. So we could diagonalize $A = PDP^{-1}$ using $P = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{bmatrix}$. Here, the columns of P are orthonormal, which makes P an orthogonal matrix. A consequence of this is that $P^{-1} = P^T$. So our diagonalization takes the form

$$A = PDP^T.$$

This is called an *orthogonal diagonalization*, which just means that we diagonalized A using an orthogonal matrix P.

So it is natural to want to know what types of matrices are orthogonally diagonalizable. Suppose $A = PDP^T$ has been orthogonally diagonalized (we are also assuming all eigenvalues, the entries of D, are real). Note that if we took the transpose of this equation, we get

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A.$$

So the matrix A must satisfy the condition $A^T = A$.

Definition 10.1. An $n \times n$ matrix A is called *symmetric* if $A^T = A$.

This simply means that the entries of A have a reflectional symmetry about the diagonal. For example, matrices of the form

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

are symmetric. Every orthogonally diagonalizable matrix must be symmetric. It is truly remarkable that the converse statement is true as well. This result is called the spectral theorem.

Theorem 10.2 (Spectral Theorem)**.** *Let A be an $n \times n$ symmetric matrix. Then the following hold.*

(1) Every eigenvalue of A is real.
(2) Any pair of eigenvectors for A that have different eigenvalues are orthogonal.
(3) The dimension of each eigenspace for A equals the multiplicity of the corresponding eigenvalue.
(4) There is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.
(5) A is orthogonally diagonalizable: $A = PDP^T$ where P is an orthogonal matrix whose columns are an orthonormal basis of eigenvectors for A.

To orthogonally diagonalize a matrix, we start by diagonalizing it. Part of the spectral theorem is that any two eigenspaces for A will be orthogonal to each other. So when constructing an eigenvector basis, we will get a lot of orthogonality between eigenvectors for free. The exception to this is when an eigenspace has dimension greater than 1. When we do the usual process for finding a basis for such an eigenspace, we may realize that the basis we get is not orthgonal. In this case, we would apply the Gram-Schmidt process to this eigenspace. After producing an orthogonal basis of eigenvectors for A, we normalize them to be unit vectors, so that P will be an orthogonal matrix.

Exercise 10.3. Orthogonally diagonalize the matrices

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Note that the eigenvalues of B are $\lambda = -2, 7, 7$.