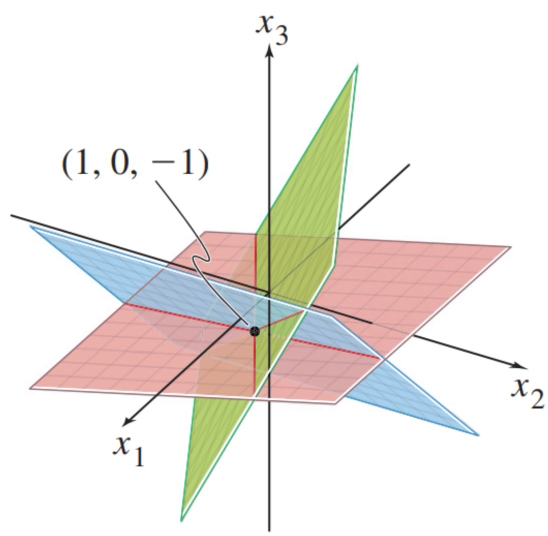
Section 0.1

What Are We Doing?



Objectives:

- Understand the premises of linear algebra

Problem: Solving Systems of Linear Equations

In school we are taught how to solve systems of the following form:

$$x_1 + 2x_2 = 3$$
$$4x_1 + 5x_2 = 6$$

There are two typical methods of solution. One, we can solve for one variable in one equation ($x_1 = 3 - 2x_2$), then substitute all instances of that variable in the other equation with what that variable is equal to. This is called **substitution**. Two, we can manipulate both sides of each equation so that the addition of the two equations **eliminates** one variable entirely:

$$-4x_1 - 8x_2 = -12 \qquad \text{(multiplied by } -4\text{)}$$
$$4x_1 + 5x_2 = 6 \qquad \text{(kept the same)}$$

This latter method is called **elimination**.

One way to simplify this method is to simply create a **matrix of the coefficients in this equation** (Section 1.1), like this:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right]$$

Now elimination amounts to adding a scalar multiple of one row to another row:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right] \xrightarrow{R_2 \to R_2 - 4R_1} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & -3 & -6 \end{array}\right]$$

This bottom row now reads "$-3x_2 = -6$", which is what we would get from applying elimination as above. Our next step in solving would involve multiplying both sides of this equation by $-\frac{1}{3}$. In this matrix form, this correlates to multiplying the entire row by a scalar:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & -3 & -6 \end{array}\right] \xrightarrow{R_2 \to -\frac{1}{3}R_2} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 2 \end{array}\right]$$

The bottom row now reads "$(0x_1+)x_2 = 2$". At this point, we have solved for a variable, and we can use back-substitution to solve for all the other variables in our system. This form of a matrix is called **row echelon form** (Section 1.2).

We can also *substitute* x_2 back into our equation in the first row by doing our row-adding operation again:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 2 \end{array}\right] \xrightarrow{R_1 \to R_1 - 2R_2} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right]$$

This form of the matrix is called **reduced row echelon form**, and the matrix simply holds the solution to our system of equations: $x_1 = -1$, and $x_2 = 2$. We have applied what are called **elementary row operations** (Section 1.5) to solve a system of linear equations.

Abstraction: Linear Transformations

In the 19th century mathematicians began applying this theory to other systems that acted like lines. Lines in $\mathbb{R}^2$ are characterized by two things: point and slope. In particular, lines through the origin are only differentiated by their slope. These lines $L(x) = rx$ have two properties that tell us how points interact with addition and multiplication:

$$L(x+y) = L(x) + L(y)$$
$$L(rx) = rL(x)$$

Note that these two properties prevent the line from breaking off (discontinuity), curling in one direction or the other (changing slope), or stopping at some point (domain restriction).

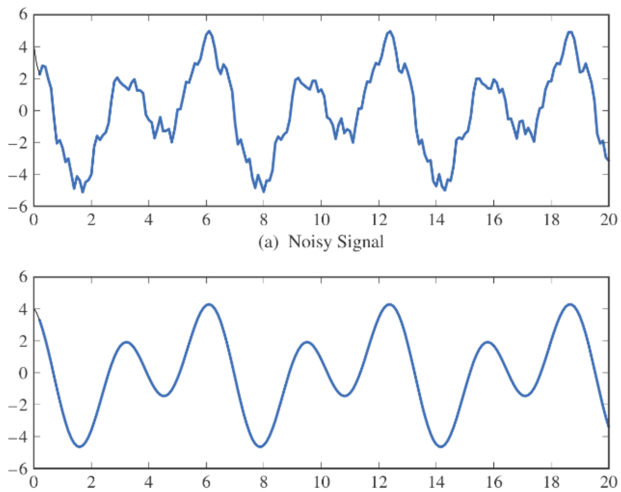
There are two other very important mathematical objects that share these two properties: the *derivative* and the *integral*.

$$(f+g)' = f' + g' \qquad \int (f+g)\,dx = \int f\,dx + \int g\,dx$$
$$(rf)' = rf' \qquad \int rf\,dx = r\int f\,dx$$

Objects that satisfy these two properties are called **linear transformations** (Section 4.1) - they "act like" lines through the origin. Note that we have replaced the numbers x and y with *functions* f and g - we generalize the notion of these two by calling them all **vectors**, and the collections in which they live **vector spaces** (Section 3.1). Just like the systems of equations we saw in the prior subsection, *linear transformations have their own corresponding matrices* based on the systems of equations they create. (There is a caveat here regarding **dimension** - Section 3.4.)

Application: Signal Processing

Because these functions can be written as matrices, we can apply matrix operations to manipulate the data they contain. The picture below in (a) is a depiction of a signal that a computer might receive from a peripheral, written as a function:



(a) Noisy Signal

(b) Filtered Signal

We can see that the *noise* in the signal is filtered out in the picture given by (b). What has been done is this: the function above has been approximated as a **linear combination** (Section 1.3) of trigonometric functions of the form $a_0 + a_k \cos(kx) + b_k \sin(kx)$ (here a_0, a_k, b_k are real numbers). Writing more and more terms in this linear combination gives us a better and better approximation of the original signal until we reach the desired error tolerance. These are useful since allowing k to be different integers gives us an **orthogonal set** of functions (Section 5.5). (Similar methods are used in data compression.)

Remarkably, this application only scratches the surface of where linear algebra can be found. The theory is perhaps most useful in the study of **linear differential equations** (Section 6.2), which can even involve partial derivatives (ours will not).

Our course will begin by discussing the theory of matrices, discuss this abstraction of linear equations to linear transformations, then give introductory theory (orthogonal sets, diagonalization of matrices) for where matrices and linear transformations can be applied.

Section 0.2
The Beauty of Mathematics

Objectives:

- Understand the differences in mathematical pedagogy in calculus versus linear algebra

It's typical for a mathematics professor to view the subject matter they teach as "beautiful". But what does this mean?

Take a look at the picture near the top of the page. Many would find this as a serene image, and several things can contribute to this.

Perhaps there is an air of familiarity to it: you've been at a log cabin and have had pleasant memories there, or you've seen it in other pictures.

Maybe it's because of how everything connects: the trees surrounding the cabin give off a feeling of security, while the colorful flowers portrayed give you a glimpse of how beautiful the view must be from the front porch.

There is also an absence of hustle, bustle, roads, construction, and crowds - this can evoke a reaction of peace and seclusion.

A mathematics professor thinks their subject is beautiful for the same reasons you might think this picture is beautiful. It is familiar, it is beautifully connected, and it is a welcome distraction from other duties.

Mathematicians often rate their happiness in their profession well above average, and it's not just about salary. They find a natural beauty to mathematics and are excited to work with it each day.

However, a test question on a math exam can sometimes feel like this:

Question: Color in the picture with the same colors used in the original.

There are a few colors that are obvious from context: the trees and bushes are green, the log cabin is brown. Just for fun, I re-colorized this photo by plugging it into an AI, and the AI was able to nail those colors. (It failed at most of the others.)

But what about the colors of the flowers? It's a lot easier for a mathematician who feels comfortable with this picture to remember these colors as pivotal to the experience of the photo, but as someone seeing the photo for the first time you'd have to truly study it to be able to recollect these details.

There are two bad approaches to remembering the details of this image:

- **The "I'll just wing it" approach:** While there are definitely some details you can gather from the given context, it's difficult to finish a question completely without having done some specific study to each type of question.
- **The "memorize every detail" approach:** While there are definitely some details you would have to memorize to recall, it's difficult to recollect everything within a timely manner to finish an exam!

The best approach to studying for a math exam is somewhere in the middle. Look for connections between different parts of this course - you'll be surprised how little you have to remember this way. **But there will still be some things you'll want to recall,** if for nothing else then to just save yourself the time of coming up with them.

Linear algebra is special because of its reliance on **theory** over **computation**.

- **Computation** is the use of a definition, algorithm, or technique applied to a specific problem. Examples: $2+2=4$, $\frac{d}{dx}x^2=2x$, $\int_0^1 x\,dx=\frac{1}{2}$.
- **Theory** is the set of ideas that spur computation forward. We cannot reliably compute unless theory tells us it's possible. Examples: the definition of the derivative, the Riemann sum.

Many engineering calculus courses focus nearly exclusively on building computational skills. In linear algebra we will add in a focus of building theory skills as well. Here's an example of what a theory-based question might look like. (You'll be doing some of this in the context of linear algebra yourself in this course!)

Question: Write 8^4-4^4 into a product of prime factors. (For example, $30=2\cdot3\cdot5$.)

This question seems crazy! I have to compute 8^4, 4^4, subtract the two, and then just hope I find the right factors? While this might seem wild on the outset, theory says this type of question can be solved much more easily:

Difference of squares formula: $a^2-b^2=(a-b)(a+b)$.

But how does this formula apply? Nothing's being squared, it's all being raised to the fourth power! For this we need to be a bit creative: in order to make this formula work with what's given, we have to make a decision: let $a=8^2$ and $b=4^2$.

Note that this works because $a^2=(8^2)^2=8^4$ and $b^2=(4^2)^2=4^4$. So $a^2-b^2=8^4-4^4$. Hence we can rewrite

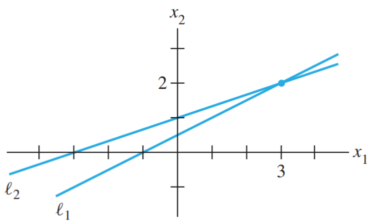
$$\begin{aligned}8^4-4^4=a^2-b^2=(a-b)(a+b)\\&=(8^2-4^2)(8^2+4^2)\\&=(8-4)(8+4)(8^2+4^2)=(4)(12)(80).\end{aligned}$$

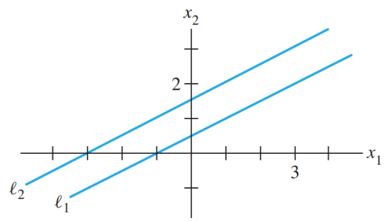
On the last line of that string of equalities you can see that we used the difference of squares formula again, this time on (8^2-4^2). While the product we have is not a product of prime factors, it is now much easier to factor - we only have to factor 4, 12, and 80.

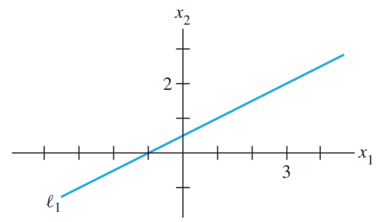
The goal of learning theory is to make computations as easy as possible. Just as we would have had a rough time factoring this big number without the difference of squares formula, we would have an incredibly difficult time understanding linear algebra without discussing quite a bit of theory.

There will be some sections in this course where you see a *lot* more definitions and theories than numbers and computations in this class. That's okay! We will warm up to this idea a bit in Section 1-5 and then continue to build upon it in the chapters to come.

Section 1.1
Matrices







Objectives:

- Define equivalent systems of linear equations
- Solve systems of linear equations using matrices

Recall that a **system of linear equations** is a collection of equalities of the form

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

where a_{ij} is a real number for $i \in [m]$, $j \in [n]$. (Here and elsewhere $\boldsymbol{[k]} := \{1, 2, \ldots, k\}$.)

Definition 1.1.1. The **coefficient matrix** of this system of equations compiles the coefficients of this system in an array where coefficients in the ith row are terms of the ith equality and coefficients in the jth column are coefficients of the same variable x_j.

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

Note that the matrix has m rows and n columns, and the indices of each entry are indexed by row, then column (instead of column, then row).

The **augmented matrix** of this system of equations adds one more column for the constant terms to the right of the equality. To show that this extra column is not associated with another variable, we add a vertical line to distinguish this column. Think of the vertical line as an "equals sign".

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

Example 1.1.1. Solve the below system of equations using substitution and elimination.

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

We say that two systems of equations are **equivalent** if they have the same solution set. In the above example, using elimination allowed us to get equivalent systems of equations to the original one given. If we restrict our solution methods to leaving constants on the right-hand side of the equality and variable terms on the left-hand side, there are three operations to get equivalent systems:

(I) Switch the order of any two equations.

(II) Multiply both sides of any one equation by a nonzero number.

(III) Add a multiple of one equation to another equation.

These three operations correspond directly to the following **elementary row operations** that can be applied to an augmented matrix:

(I) Switch any two rows.

(II) Multiply a row by a nonzero number.

(III) Add a multiple of one row to another row.

Example 1.1.2. Use elementary row operations to the augmented matrix $\left[A|\mathbf{b}\right]$ below so that $a_{21} = a_{31} = a_{32} = 0$. Then use **back-substitution** to solve the system of equations.

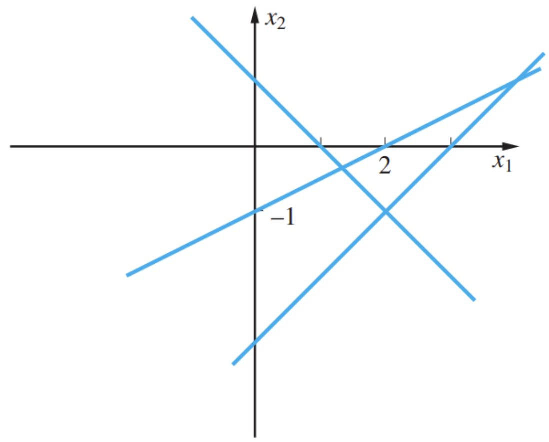
$$\left[\begin{array}{ccc|c} 3 & 2 & -1 & -2 \\ -3 & -1 & 1 & 5 \\ 3 & 2 & 1 & 2 \end{array}\right]$$

Note that our methods are very similar to those in Example 1 - our solution techniques are in fact identical. This method of solving a system of equations is called **Gauss-Jordan reduction**.

Example 1.1.3. Solve the following system of equations using Gauss-Jordan reduction.

$$\begin{aligned} 2x_1 + x_2 &= 8 \\ 4x_1 - 3x_2 &= 6 \end{aligned}$$

Section 1.2
Row Echelon Form



Objectives:

- Expand use of Gauss-Jordan reduction to overdetermined and underdetermined systems
- Recognize row echelon form and use reduction to find equivalent systems in this form
- Discover applications of matrices

We ended up with two types of matrices in the examples in the last section. In Example 1.1.2 we ended in **row echelon form**, which we could then use to find the solution by using back-substitution. In Example 1.1.3 we continued from this form to **reduced row echelon form**, which gave us the solution outright.= We will want a formal definition to ensure we end up in this form, because some systems of equations are **overdetermined** or **underdetermined**.

Definition 1.2.1. We say that a system of equations as given at the top of Section 1.1 is **overdetermined** if there are more equations than there are variables. We say that a system of equations is **underdetermined** if there are fewer equations than there are variables.

NOTE: Overdetermined systems are *usually* inconsistent (i.e., have no solutions), although they are not necessarily inconsistent. Underdetermined systems are *usually* dependent systems, meaning that they have infinitely many solutions, although they are not necessarily dependent.

Example 1.2.1. Solve the following system:

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

Example 1.2.2. Solve the following system:

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

The final matrix in Example 1.2.1 was in row echelon form:

Row Echelon Form

We say a matrix A is in **row echelon form** if

(i) All rows consisting entirely of zeros are at the bottom of the matrix.

(ii) The first nonzero entry in each nonzero row is 1.

(iii) If row k has nonzero entries, the number of leading zero entries in row $k+1$ is *greater* than the number of leading zero entries in row k.

Here is an example of a matrix in row echelon form. Let $*$ denote a generic number, which could be zero or nonzero:

$$\left[\begin{array}{ccccc|c} 1 & * & * & * & * & * \\ 0 & 1 & * & * & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Think that there must be a "staircase of zeroes" heading down and to the right. It is okay for the staircase to be uneven. But the place where you step down must be a 1 each time.

The final matrix in Example 1.2.2 was in **reduced row echelon form**:

Reduced Row Echelon Form

We say a matrix A is in **reduced row echelon form** if

(i) The matrix is in row echelon form.

(ii) The first nonzero entry in each row is the *only* nonzero entry in its column.

Here is an example of a matrix in *reduced* row echelon form:

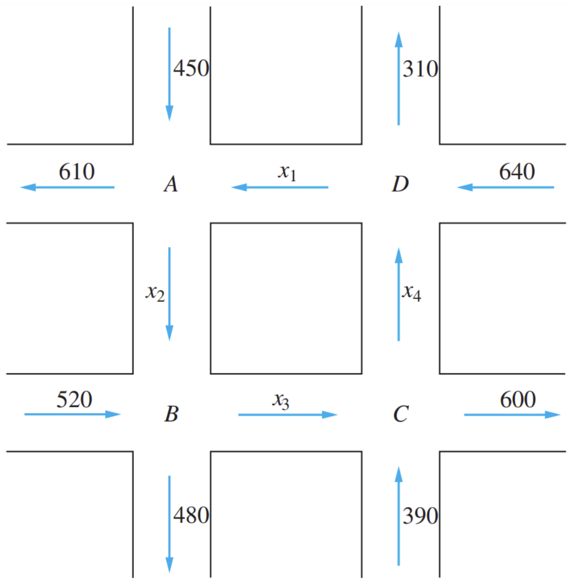
$$\left[\begin{array}{ccccc|c} 1 & 0 & * & 0 & 0 & * \\ 0 & 1 & * & 0 & 0 & * \\ 0 & 0 & 0 & 1 & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Once we isolate the "leading ones", by which we mean the first 1 on each row, we check to make sure that they are the *only* 1 on their respective columns.

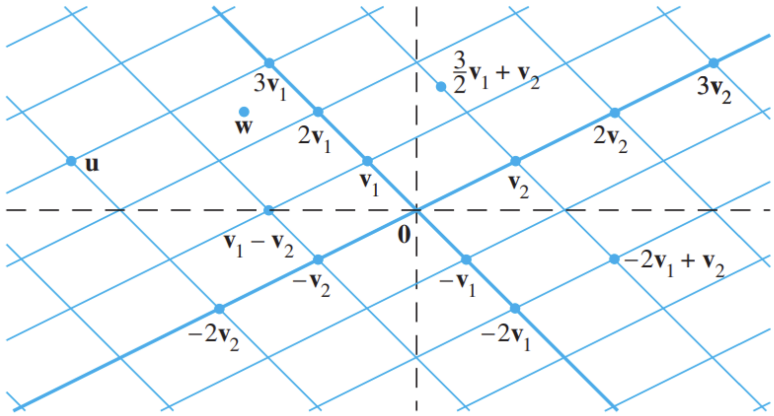
Example 1.2.3. Solve the system

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

Example 1.2.4. In the downtown section of a certain city, two sets of one-way streets intersect as shown below. The average hourly volume of traffic entering and leaving this section during rush hour is given in this diagram. Determine the amount of traffic between each of the four intersections.



Section 1.3
Matrix Arithmetic



Objectives:

- Visualize matrices geometrically in $\mathbb{R}^n$
- Define operations on matrices in ways consistent with geometric intuition

We are now ready to define a matrix:

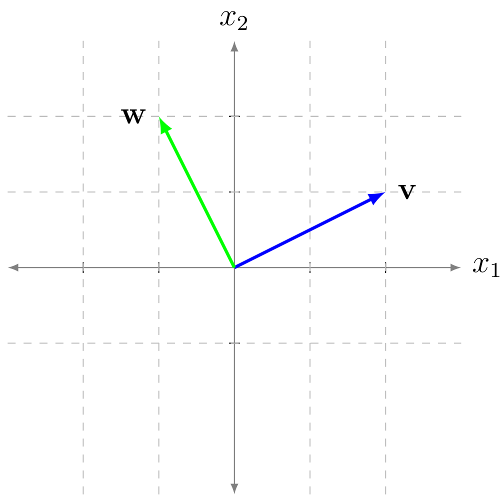
Definition 1.3.1. A **matrix** is an $m \times n$ array of real numbers. If the matrix is a $1 \times n$ matrix, we say it is a **row vector**. If the matrix is an $m \times 1$ matrix, we say it is a **column vector**.

The **vector** language here is very intentional: while matrices are best introduced as *algebraic* objects, they are best understood as *geometric* objects.

Vectors: A Brief Review

Recall that a **vector** $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is a list of numbers of length n. (Note that we refer to **vectors**, without any modifier, as if they are column vectors - for reasons we will describe shortly.) The **dot product** of two vectors $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ is given by multiplying corresponding entries and adding the resulting products: $\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$.

Here are the vectors $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$:



Note that these two vectors make a right angle with each other. This turns out to coincide directly with the fact that the *dot product of **v** and **w** equals zero*: $\mathbf{v} \cdot \mathbf{w} = 2(-1) + 1(2) = 0$.

How do matrices work with vectors? The answer can be found in the systems of equations we've been working with in the previous sections.

Example 1.3.1. Write the system of equations

$$\begin{aligned} 0x_1 - x_2 &= -1 \\ x_1 - 0x_2 &= 2 \end{aligned}$$

as a **matrix equation** of the form $A\mathbf{x} = \mathbf{b}$.

Notice that in this case, the vector $\mathbf{b}$ is equal to the vector $\mathbf{w}$ given above. Generally the vector $\mathbf{x}$ is unknown - it's something we have to solve for. Gratefully, thanks to what we've learned in our previous sections, we can solve for $\mathbf{x}$:

We have learned what the matrix A does to the vector $\mathbf{v}$ now: it *rotates the vector counter-clockwise by 90 degrees.*

So the matrix A does something to vectors! We can now make an even better definition:

Definition 1.3.2. A **matrix** A is a type of *function* that maps vectors to vectors. The **matrix product** $A\mathbf{x}$, for an appropriately-sized vector $\mathbf{x}$, is given by

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

The **matrix equation** $A\mathbf{x} = \mathbf{b}$, then, corresponds to the following equality of matrices:

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

This is the exact system of equations we gave at the beginning of Section 1.1.

Example 1.3.2. Let A be an $m \times n$ matrix and say that we wish to analyze the matrix equation $A\mathbf{x} = \mathbf{b}$. What must the size of $\mathbf{x}$ be? What must the size of $\mathbf{b}$ be?

Matrix Arithmetic

Now that the matrix has been defined as a function, we can ask a few questions we normally ask of functions. Can we *add* two matrices? Can we *multiply* a matrix by a constant? Can we *compose* two matrices? We want our definitions here to line up with how adding functions work. That is, for functions f, g and matrices A, B:

$$\begin{aligned} (f+g)(x) = f(x) + g(x) \quad &\Longleftrightarrow \quad (A+B)\mathbf{x} = A\mathbf{x} + B\mathbf{x} \\ (cf)(x) = cf(x) \quad &\Longleftrightarrow \quad (cA)\mathbf{x} = cA\mathbf{x} \\ (f \circ g)(x) = f(g(x)) \quad &\Longleftrightarrow \quad (AB)\mathbf{x} = A(B\mathbf{x}) \end{aligned}$$

Matrix Addition and Scalar Multiplication

Given two $m \times n$ matrices A, B and a scalar c, we define:

(i) $A + B = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$

NOTE: In order to add two matrices, the two matrices **must** be of the same size. Otherwise the addition is undefined.

(ii) $cA = \begin{bmatrix} ca_{11} & \cdots & ca_{1n} \\ \vdots & \ddots & \vdots \\ ca_{m1} & \cdots & ca_{mn} \end{bmatrix}$

(iii) $A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$

Briefly, these definitions come about since they mirror exactly how we *add* and *multiply by scalars* a pair of two vectors of the same size. One can think of a matrix A as if it were an array of column vectors stacked side-by-side:

$$A = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & \cdots & | \end{bmatrix}$$

So we can see that $A+B$ is formed by simply adding corresponding column vectors together, and that cA is formed by simply scaling each column vector.

Example 1.3.3. Write the system of equations

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

as a matrix equation $A\mathbf{x} = \mathbf{b}$. Then write $A\mathbf{x}$ as a **linear combination** of column vectors.

To be formal, we give the definition of linear combination:

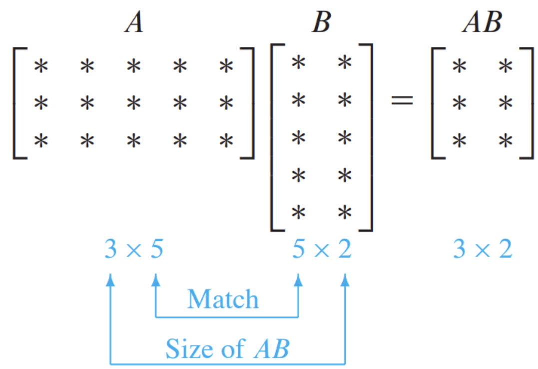
Definition 1.3.3. If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are vectors in $\mathbb{R}^m$ and $c_1, \ldots, c_n$ are scalars, than a vector $\mathbf{v}$ is equal to a *linear combination* of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ if

$$\mathbf{v} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n.$$

This leads us to our first theorem of this course:

Theorem 1.1. *A linear system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if $\boldsymbol{b}$ can be written as a linear combination of the column vectors of A.*

Section 1.4
Matrix Algebra



Objectives:

- Learn matrix multiplication and the properties of matrix arithmetic
- Recognize the identity matrix and check whether two matrices are inverse to each other

Matrix Multiplication We have not yet discussed how to *compose* two matrix functions. This process can seem strange until we see it in action:

Example 1.4.1. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$. Find $AB\mathbf{x}$, where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

In short, *if* the product AB is defined, then the (i, j)th entry of AB is equal to

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Example 1.4.2. Say that the product AB is defined. What must be true about the sizes of A and B? What is the resulting size of AB? Must BA also be defined?

Example 1.4.3. Show that, for matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix},$$

the matrix products AB and BA both exist but are unequal.

The typical *commutative property of multiplication* that we would expect, therefore, does not apply to matrices. However, lots of other rules do apply and can be taken for granted: for matrices A, B, C of the appropriate sizes and scalars α, β,

1. $A + B = B + A$
2. $(A+B)+C = A+(B+C)$
3. $(AB)C = A(BC)$
4. $A(B+C) = AB+AC$
5. $(A+B)C = AC+BC$
6. $(\alpha\beta)A = \alpha(\beta A)$
7. $\alpha(AB) = (\alpha A)B = A(\alpha B)$
8. $(\alpha + \beta)A = \alpha A + \beta A$
9. $\alpha(A + B) = \alpha A + \alpha B$.

Identity and Inverses

Example 1.4.4. Can we come up with a matrix O that acts like the number 0 for matrices? That is, $A+O=O+A=A$, and $AO=OA=O$? (Note that for this last equation to hold, O must be an $n \times n$ matrix.)

Example 1.4.5. Can we come up with a matrix I that acts like the number 1 for matrices? That is, $AI=IA=A$? (Note that for this last equation to hold, $I =: I_n$ must be an $n \times n$ matrix.)

The reason we bring up the **identity matrix** I_n for $n \times n$ matrices is this. We have defined addition, subtraction (since $A - B = A + (-B)$), and multiplication of matrices, but we have not yet defined matrix division. Recall that for *non-zero* numbers a we define $\frac{1}{a}$, or a^{-1}, to be the number such that

$$a\frac{1}{a} = \frac{1}{a}a = 1.$$

So given an $n \times n$ matrix A, we would like to see if there is a matrix A^{-1} such that

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

We say such a matrix A is **invertible** or **nonsingular**. An $n \times n$ matrix A is **singular** if it does *not* have a multiplicative inverse.

Example 1.4.6. Show that $\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$ and $\begin{bmatrix} -\frac{1}{10} & \frac{2}{5} \\ \frac{3}{10} & -\frac{1}{5} \end{bmatrix}$ are inverses of each other.

Example 1.4.7. Show that $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is singular.

The above example shows us that *some non-zero matrices can still be singular*. The process of finding the inverse of a matrix is very similar to the Gauss-Jordan elimination technique we saw earlier: we can solve for each entry of the inverse matrix by reducing the augmented matrix $\left[A|I_n\right]$ so that the left-hand side is in reduced row echelon form. If the augmented matrix now reads $\left[I_n|B\right]$ for some matrix B, then A is invertible and $B = A^{-1}$.

Example 1.4.8. Use Gauss-Jordan reduction to find $\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}^{-1}$.

Example 1.4.9. Use the properties of inverses to show that $(AB)^{-1} = B^{-1}A^{-1}$.

We close with a few properties of transpose matrices that we may take for granted:

1. $(A^T)^T = A$
2. $(\alpha A)^T = \alpha A^T$
3. $(A + B)^T = A^T + B^T$
4. $(AB)^T = B^T A^T$

Section 1.5

Elementary Matrices

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

Objectives:

- Recognize elementary matrices and use them to determine invertibility of a matrix
- Use elementary matrices to calculate the LU factorization of a matrix

Now that we have learned matrix multiplication, we can write each of our row operations (I)-(III) from Section 1.1 as *matrix* transformations. Let's do one together.

Example 1.5.1. Write the row operation (II) as a matrix transformation. Say we have a 3×3 matrix A and that we multiply the first row by some non-zero scalar α to get a new matrix B. Find the matrix E such that $EA = B$.

Example 1.5.2. Show that applying the matrix transformation $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ to the matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ switches the first two rows of A.

Example 1.5.3. Verify that $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is equal to $\begin{bmatrix} a & b & c \\ d & e & f \\ g-4a & h-4b & i-4c \end{bmatrix}$.

Each of the matrices E_1, E_2, and E_3 given in our examples above are **elementary matrices**.

Definition 1.5.1. An **elementary matrix** is a matrix that is obtained by performing a single elementary operation (I)-(III) on an identity matrix.

Example 1.5.4. Show that E_1, E_2, and E_3 are all invertible.

We have proven the following theorem:

Theorem 1.2. *All elementary matrices are invertible, and the inverse of an elementary matrix is also an elementary matrix.*

Recall in Section 1.1 that we said that a system of equations was equivalent to another if they have the same solution set. This is the same notion as being able to get from one system to another by applying a sequence of row operations. This inspires the following definition for matrices:

Definition 1.5.2. A matrix B is **row equivalent** to a matrix A if there exists a finite sequence $E_1, E_2, \ldots, E_k$ of elementary matrices such that

$$B = E_k E_{k-1} \cdots E_1 A.$$

Example 1.5.5. Prove the following theorem: if A is an $n \times n$ matrix, the following statements are the same:

This leads us to the following result: the system $A\mathbf{x} = \mathbf{b}$ of n linear equations in n unknowns has a unique solution if and only if A is nonsingular, in which case $\mathbf{x} = A^{-1}\mathbf{b}$.

Example 1.5.6. Solve the system

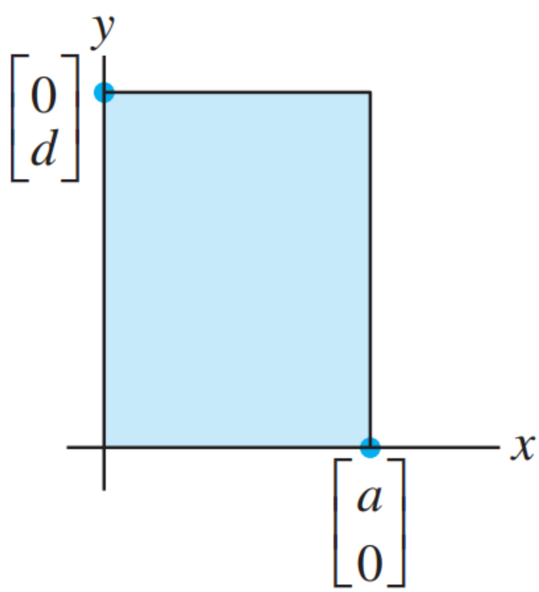
$$\begin{aligned} 2x_1 + 4x_2 &= 10 \\ 3x_1 + x_2 &= -20. \end{aligned}$$

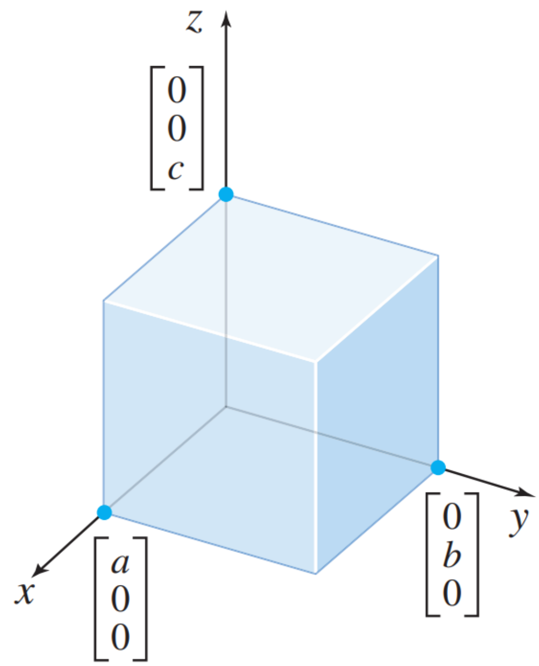
Example 1.5.7. Let $A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 5 & 2 \\ 4 & -1 & 9 \end{bmatrix}$. Using only row operation (III), find a matrix U in row echelon form that is equivalent to A. Then write $A = LU$, where L is the product of elementary matrices of row operation (III).

This is called a **decomposition** of matrices, where a matrix A is written as a product of less-complicated matrices. The matrix L is said to be **lower triangular**, since all entries above the main diagonal are zero. The matrix U is said to be **upper triangular**, since all entries below the main diagonal are zero. This decomposition where all diagonal entries of the matrix L are 1's is called the **LU factorization** of the matrix A. For large matrices, it is less computationally expensive (i.e., takes less time) to first factorize the matrix into its LU decomposition before analyzing its data further.

Section 2.1

Determinants





Objectives:

- Calculate the determinant of $n \times n$ matrices

In the previous section, we found some equivalent ways to state that a matrix A is invertible. We will be able to add one more statement to that theorem in Example 1.5.5 in this chapter: a matrix A is invertible iff $\det(A) \neq 0$.

A **determinant of order 2** is defined to be

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

A **determinant of order 3** can be defined in terms of order 2 determinants:

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix}.$$

This is only one of the ways to define this determinant. While our scalar values here come from the first row, we will show how we can proceed similarly down *any* row or column of this matrix in Example 2.1.2.

Example 2.1.1. Find the determinant of the matrix $A = \begin{bmatrix} 2 & 5 & 4 \\ 3 & 1 & 2 \\ 5 & 4 & 6 \end{bmatrix}$.

Example 2.1.2. Find the determinant of the matrix $A = \begin{bmatrix} 2 & 0 & 4 \\ 3 & 0 & 2 \\ 5 & 0 & 6 \end{bmatrix}$.

Example 2.1.3. Show that a 2×2 matrix A is row equivalent to I if and only if $\det(A) \neq 0$.

We continue defining the determinant of a matrix recursively, just as we did going from order 2 to order 3. For an order 4 matrix, pick any column or row of the matrix. Then for each entry in that column or row, multiply that entry times the determinant of order 3 formed by deleting the entire row and column that entry is in. The signs of each term will be determined by the **matrix of signs**:

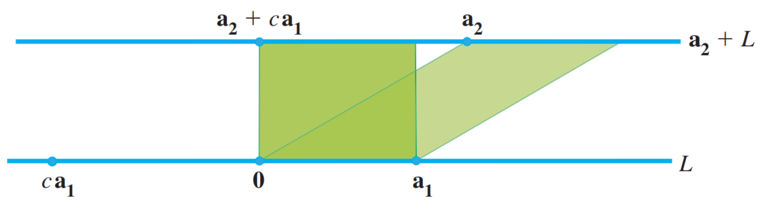
$$\begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix},$$

where the matrix is formed by starting with a + in the top-left corner, then alternating signs for every row and column traversed. The sum of these terms is the **determinant** of the larger matrix. We can do a similar process to find the determinant of larger matrices.

Example 2.1.4. Write the determinant of $A = \begin{bmatrix} 1 & 5 & 3 & -7 & 3 \\ 6 & 0 & -9 & 2 & 4 \\ 6 & 0 & 1 & -2 & -4 \\ 9 & 1 & 8 & -9 & 3 \\ 1 & 0 & -8 & 5 & 5 \end{bmatrix}$ as a *linear combination* of determinants of order 4.

We won't discuss this further in this course, but one other definition for the determinant is the following: given a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, form the column vectors $\begin{bmatrix} a \\ c \end{bmatrix}$ and $\begin{bmatrix} b \\ d \end{bmatrix}$ of the matrix. Then $|\det(A)|$ is equal to the *area of the parallelogram* formed by these column vectors starting at the origin in $\mathbb{R}^2$. If the angle from $\begin{bmatrix} a \\ c \end{bmatrix}$ to $\begin{bmatrix} b \\ d \end{bmatrix}$ is less than 180 degrees, the determinant is positive; otherwise the determinant is negative. A similar definition can be made in larger dimensions.

Section 2.2
Determinant Properties



Objectives:

- Determine how row operations affect the determinant of a matrix
- Establish that determinants are linked to invertibility of a matrix

Example 2.2.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then suppose $E_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is the elementary matrix associated to row operation (I) where we switch the two rows of this matrix. Find $\det(E_1 A)$.

Example 2.2.2. Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ be a general 3×3 matrix, and let $E_2 = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ be the elementary matrix associated to multiplying the first row of this matrix by a nonzero scalar α. Find $\det(E_2 A)$.

Example 2.2.3. Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ be a general 3×3 matrix. Then consider $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{bmatrix}$ to be the elementary matrix associated to adding α times the first row to the third row of the original matrix. Find $\det(E_3 A)$.

When combining our discoveries above with the fact that, for elementary matrices E,

$$\det(E) = \begin{cases} -1 & \text{if } E \text{ is of type I} \\ \alpha \neq 0 & \text{if } E \text{ is of type II} \\ 1 & \text{if } E \text{ is of type III} \end{cases}.$$

we find the following:

Theorem 2.1. *If E is an elementary matrix, then*

$$\det(EA) = \det(E)\det(A).$$

In particular,

(I) interchanging two rows of a matrix changes the sign of the determinant.

(II) multiplying a single row of a matrix by a scalar has the effect of multiplying the value of the determinant by that scalar.

(III) adding a multiple of one row to another does not change the value of the determinant.

Since $\det(AE) = \det((AE)^T) = \det(E^T A^T)$, we can repeat statements (I) through (III) about switching *columns* of the matrix A, since these are the *rows* of the matrix A^T.

Example 2.2.4. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 8 & -1 & 2 \end{bmatrix}$. Find $\det(A)$ using row reduction and elementary matrices.

We now arrive at the promised result:

Theorem 2.2. *An $n \times n$ matrix A is singular if and only if* $\det(A) = 0$.

Proof. Write A in its row-echelon form U, so that

$$U = E_k E_{k-1} \cdots E_1 A$$

for some elementary matrices E_i for $i \in [k]$. What must be true if A is singular?

What must be true if A is nonsingular?

□

In a similar way to how we showed that $\det(EA) = \det(E)\det(A)$ whenever E is an elementary matrix, we can show that $\det(BA) = \det(B)\det(A)$ for *arbitrary matrices* B and A.

Example 2.2.5. Recall that a 3×3 matrix A might have an LU factorization of the form

$$\begin{bmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}.$$

Here the values of ℓ_{ij} and u_{ij} (for i, j appropriately defined) are arbitrary real numbers. Calculate the determinant of LU, using the values given above.

Section 2.3

Uses of Determinants

$$x = \frac{\det\begin{bmatrix} -3 & 1 & -2 \\ 0 & -1 & -1 \\ 13 & 2 & 3 \end{bmatrix}}{\det\begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & -1 \\ 1 & 2 & 3 \end{bmatrix}} = \frac{-36}{-18} = 2, \quad y = \frac{\det\begin{bmatrix} 1 & -3 & -2 \\ 2 & 0 & -1 \\ 1 & 13 & 3 \end{bmatrix}}{\det\begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & -1 \\ 1 & 2 & 3 \end{bmatrix}} = \frac{-18}{-18} = 1, \quad z = \frac{\det\begin{bmatrix} 1 & 1 & -3 \\ 2 & -1 & 0 \\ 1 & 2 & 13 \end{bmatrix}}{\det\begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & -1 \\ 1 & 2 & 3 \end{bmatrix}} = \frac{-54}{-18} = 3.$$

Objectives:

- Utilize Cramer's Rule to calculate solutions to systems of equations

Our use of determinants has an application in what is known as *Cramer's Rule.*

Theorem 2.2. *Let A be a nonsingular $n \times n$ matrix, and let $\boldsymbol{b} \in \mathbb{R}^n$. Let A_i be a matrix obtained by replacing the ith column of A by $\boldsymbol{b}$. If $\boldsymbol{x}$ is the unique solution of $A\boldsymbol{x} = \boldsymbol{b}$, then the ith component of the solution $\boldsymbol{x}$, x_i, is equal to*

$$x_i = \frac{\det(A_i)}{\det(A)}.$$

Proof. We will only show this for an arbitrary 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. First note that, by Question 1 on our Homework 3,

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Hence by multiplying out the top row of this matrix by $\mathbf{b}$, we get

$$x_1 = \frac{b_1 \cdot d - b_2 \cdot b}{\det(A)} = \frac{\det\left(\begin{bmatrix} b_1 & b \\ b_2 & d \end{bmatrix}\right)}{\det(A)} = \frac{\det(A_1)}{\det(A)}.$$

Also

$$x_2 = \frac{-b_1 \cdot c + b_2 \cdot a}{\det(A)} = \frac{\det\left(\begin{bmatrix} a & b_1 \\ c & b_2 \end{bmatrix}\right)}{\det(A)} = \frac{\det(A_2)}{\det(A)}.$$

□

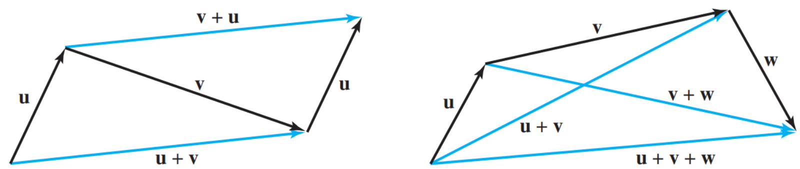
Example 2.3.1. Use Cramer's Rule to solve

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 5 \\ 2x_1 + 2x_2 + x_3 &= 6 \\ x_1 + 2x_2 + 3x_3 &= 9 \end{aligned}$$

Example 2.3.2. Why does Cramer's Rule only work for nonsingular matrices?

Section 3.1

Vector Spaces



Objectives:

- Identify vector spaces and non-vector spaces
- Explore functional vector spaces and their applications

MOTIVATION: P_n In Section 0.1 we saw that one application of linear algebra does not have to do with vectors in $\mathbb{R}^n$ at all. Instead, we replace our object of study with *functions* rather than coordinates. The properties we saw with matrices back in Chapters 1 and 2 still carry over into the world of functions, in a manner that we will be discussing more in Chapter 4: matrices are *linear transformations* over their respective vector space.

To see how we might be able to connect these two, let's start by defining the space P_n:

Definition 3.1.1. Given a natural number $n \in \{1, 2, 3, \dots\}$, we define P_n to be the set of all polynomials

$$a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x^1 + a_0$$

of degree less than n. Here the values of a_i (for $i \in [n-1] \cup \{0\}$) are real numbers.

Note that, for example, $x^{n-1} + x^{n-2} + \cdots + x^2 + x + 2$ is in P_n by this definition, but it is *not* in P_{n-1} since its degree of $n-1$ is *not* less than $n-1$. The polynomial x^2 is in P_n as long as $n \geq 3$, but $\frac{1}{x}$ is *not* in P_n for any n. The **coordinate** corresponding to a polynomial $p \in P_n$ is given by

$$P_n \ni a_0 + a_1x^1 + \cdots + a_{n-2}x^{n-2} + a_{n-1}x^{n-1} \quad \leftrightarrow \quad \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-2} \\ a_{n-1} \end{bmatrix} \in \mathbb{R}^n.$$

Note that this mapping uniquely defines a point in $\mathbb{R}^n$ for every polynomial $p \in P_n$. What's more, every point in $\mathbb{R}^n$ can uniquely yield a polynomial in P_n just by putting the numbers back as coefficients of the original polynomial. This starts helping P_n look a *lot* like $\mathbb{R}^n$.

We can even go farther: how do we add two polynomials in P_n? The way we are used to adding polynomials is combining like terms, like so:

$$\begin{array}{rl} & a_0 + a_1 x + \cdots + a_{n-2}x^{n-2} + a_{n-1}x^{n-1} \\ + & \underline{b_0 + b_1 x + \cdots + b_{n-2}x^{n-2} + b_{n-1}x^{n-1}} \\ & (a_0+b_0) + (a_1+b_1)x + \cdots + (a_{n-2}+b_{n-2})x^{n-2} + (a_{n-1}+b_{n-1})x^{n-1} \end{array}$$

This looks a lot like how we would add two vectors in $\mathbb{R}^n$:

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-2} \\ a_{n-1} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-2} \\ b_{n-1} \end{bmatrix} = \begin{bmatrix} a_0 + b_0 \\ a_1 + b_1 \\ \vdots \\ a_{n-2} + b_{n-2} \\ a_{n-1} + b_{n-1} \end{bmatrix}.$$

Comparing the summed polynomial and the summed vector once again,

$$(a_0+b_0) + (a_1+b_1)x + \cdots + (a_{n-2}+b_{n-2})x^{n-2} + (a_{n-1}+b_{n-1})x^{n-1} \quad \leftrightarrow \quad \begin{bmatrix} a_0 + b_0 \\ a_1 + b_1 \\ \vdots \\ a_{n-2} + b_{n-2} \\ a_{n-1} + b_{n-1} \end{bmatrix},$$

we see that this is exactly the same correspondence we started out with!

We can see how multiplying a polynomial by a scalar multiple does the same thing as multiplying an $\mathbb{R}^n$ vector by this same scalar under this correspondence. Any two spaces which are identical under a correspondence and which add and scalar multiply the same way are said to be **vector space isomorphic** to each other - they are functionally the same space.

We will call both these coordinates and these functions **vectors** in this chapter. This means that a **vector** is more than just a coordinate; it could also denote a function or an abstract object in what is known as a **vector space**. This notion of **vector space** carries with it all three things we saw above: the collection of objects (vectors), the addition operation (between vectors), and the multiplication operation (with a scalar). These three things together define a vector space.

Vector Space Definition

Definition 3.1.2. We will define a **vector space** $(V, +, \cdot)$ to be three things altogether: a collection of objects V, an addition operation between vectors $+$ (which looks like $\mathbf{x}+\mathbf{y}$), and a multiplication operation between a scalar and a vector $\cdot$ (which looks like $\alpha\cdot\mathbf{x}$, or $\alpha\mathbf{x}$). We ask for the following axioms to be satisfied:

A1. For any $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x}+\mathbf{y} \in V$, and

$$\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}.$$

A2. For any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

$$(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z}).$$

A3. There is a number $\mathbf{0} \in V$ so that, for all $\mathbf{x} \in V$,

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

A4. For any vector $\mathbf{x} \in V$ there is a vector denoted by $-\mathbf{x}$ with the property that

$$\mathbf{x}+(-\mathbf{x}) = \mathbf{0}.$$

A5. For any $\alpha \in \mathbb{R}$, $\mathbf{x}, \mathbf{y} \in V$, $\alpha\mathbf{x}$ and $\alpha\mathbf{y}$ are in V, and

$$\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x}+\alpha\mathbf{y}.$$

A6. For any $\alpha, \beta \in \mathbb{R}$, $\mathbf{x} \in V$,

$$(\alpha+\beta)\mathbf{x} = \alpha\mathbf{x}+\beta\mathbf{x}.$$

A7. For any $\alpha, \beta \in \mathbb{R}$, $\mathbf{x} \in V$,

$$(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x}).$$

A8. For any $\mathbf{x} \in V$,

$$1\mathbf{x} = \mathbf{x}.$$

All of these axioms are quite sensible. However, very few combinations of V, $+$, and $\cdot$ satisfy them. If one of these axioms fails, the combination $(V, +, \cdot)$ cannot be a vector space. Let us see some *non*-examples.

Example 3.1.1. Define $V = \{(a, 1) : a \in \mathbb{R}\}$ to be the line in $\mathbb{R}^2$ where the second coordinate is always 1. (This is the line $y = 1$.) With standard addition $+$ and multiplication $\cdot$, find all axioms that show $(V, +, \cdot)$ is *not* a vector space.

Example 3.1.2. Let $V = \mathbb{R}^n$ and let $+$ be the standard addition. However, define $\odot$ to be a multiplication between a scalar r and a vector $\mathbf{x}$ such that

$$r \odot \mathbf{x} = \mathbf{0}$$

rather than the typical multiplication. Find all axioms that show that $(V, +, \odot)$ is *not* a vector space.

We now continue to one of our most common examples of a vector space: the *functional* vector space.

Example 3.1.3. Let $a < b$ be real numbers, and define $C[a, b]$ to be the space of all *continuous* functions defined on the closed interval $[a, b]$. Let $+$ denote standard function addition and let $\cdot$ denote standard scalar multiplication of functions. Show that $(C[a, b], +, \cdot)$ is a vector space.

Remember, if something fails *one* of the axioms, it is not a vector space!

We close with a few properties that all vector spaces share which we can take for granted. (Can you use the axioms to prove these?)

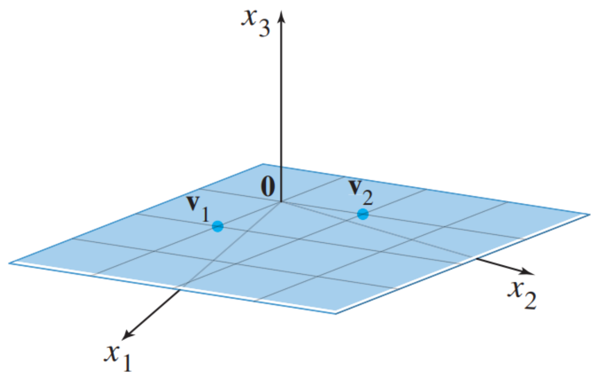
Theorem 3.1. *If V is a vector space and $\boldsymbol{x}$ is any element of V, then*

(i) $0\boldsymbol{x} = \boldsymbol{0}$.

(ii) if $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0}$, *then* $\boldsymbol{y} = -\boldsymbol{x}$ *as defined in A3.*

(iii) $(-1)\boldsymbol{x} = -\boldsymbol{x}$ *as defined in A3.*

Section 3.2
Subspaces



Objectives:

- Identify subspaces of vector spaces
- Discover nullspaces and spans of collections of vectors as subspaces
- Define spans of vectors and spanning sets of vector spaces

Example 3.2.1. Let's revisit Example 3.1.1. This time, let's define $S = \{(a, 0) : a \in \mathbb{R}\}$ to be the line in $\mathbb{R}^2$ where the second coordinate is always 0. (This is the line $y = 0$.) With standard addition $+$ and multiplication $\cdot$, show that $(S, +, \cdot)$ *is* a vector space.

The axioms that didn't hold last time now hold easily since our vector space S, which is a line, now also contains the origin. In fact, many of the axioms we checked hold automatically since S belongs to $V = \mathbb{R}^2$, which is already a vector space. We define a **subspace** to be a subset S of a larger space V which is by itself a vector space when equipped with the addition and multiplication of V. So for example, the line $y = 0$ is a subspace of the vector space $(\mathbb{R}^2, +, \cdot)$.

If we take out the axioms that automatically hold, we are left three small axioms to test, given below:

Subspace Test

Let $(V, +, \cdot)$ be a vector space. For a given subset $S \subset V$, S is a subspace of V if and only if all of the following are true:

(i) $\mathbf{0} \in S$.

(ii) $\alpha\mathbf{x} \in S$ whenever $\mathbf{x} \in S$ and for any scalar α

(iii) $\mathbf{x} + \mathbf{y} \in S$ whenever $\mathbf{x}, \mathbf{y} \in S$.

That is, if S is a nonempty set closed under vector addition and scalar multiplication, then S is a subspace of V.

NOTE: A subset $S \subset V$ only needs to fail *one* of these properties in order to not be a subspace.

Example 3.2.2. Let $S = \{(x_1, x_2, x_3)^T : x_1 = x_2\}$. Is S a subspace? Prove or disprove your answer.

Define the **space of n-times differentiable functions** $C^n[a, b]$ to be the space of continuous functions on $[a, b]$ whose nth derivatives are defined and also continuous.

Example 3.2.3. Show that $|x|$ is in $C[a, b]$ but not in $C^1[a, b]$. Then show that $x|x|$ is in $C^1[a, b]$ but not in $C^2[a, b]$. (Hint: $\frac{d}{dx}|x| = \text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$.)

The space $C^n[a, b]$ is non-empty since the zero function is infinitely differentiable. Since the sum and scalar multiplication of n-times differentiable functions is also n-times differentiable, $C^n[a, b]$ is a subspace of $C[a, b]$ by the Subspace Test.

Example 3.2.4. Let S be the set of all f in $C^2[a, b]$ such that

$$f''(x) + f(x) = 0.$$

Show that S is a subspace of $C^2[a, b]$.

Example 3.2.5. Let A be an arbitrary $m \times n$ matrix, and define the space $N(A)$ given below to be the **nullspace** of A:

$$N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

Show that $N(A)$ is a subspace of $\mathbb{R}^n$.

Example 3.2.6. Determine $N(A)$ if

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

To describe the nullspace in the above example, we could have also said

$$N(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We say that the **span** of this collection of vectors, $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, is the *set of all linear combinations* of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. That is,

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k : \alpha_i \in \mathbb{R}, i \in [k]\}.$$

Example 3.2.7. Find the span of the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$.

Example 3.2.8. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an arbitrary collection of vectors in $\mathbb{R}^m$. Show that $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a subspace of $\mathbb{R}^m$.

Example 3.2.9. For the vector space $\mathbb{R}^2$, we define $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to be the **standard basis vectors** for $\mathbb{R}^2$. Show that $\{\mathbf{e}_1, \mathbf{e}_2\}$ form a **spanning set** for $\mathbb{R}^2$ - that is, show that $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$.

Example 3.2.10. Can a single vector form a spanning set for $\mathbb{R}^2$? If so, give it.

Example 3.2.11. Is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix} \right\}$ a spanning set for $\mathbb{R}^3$?

Example 3.2.12. Is $\{1-x^2, x+2, x^2\}$ a spanning set for P_3?

We close with a few facts to help us remember the nullspace of an $m \times n$ matrix A.

- the nullspace $N(A)$ is the collection of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. Hence the vectors $\mathbf{x}$ are always in $\mathbb{R}^n$ (that is, their size is determined by the number of columns of A - otherwise $A\mathbf{x}$ would be undefined)
- if $\mathbf{x}_1$ and $\mathbf{x}_2$ *both* solve the equation $A\mathbf{x} = \mathbf{b}$, then

$$A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

  so $\mathbf{x}_1 - \mathbf{x}_2 \in N(A)$.
- conversely, if $\mathbf{z} \in N(A)$ and $\mathbf{x}$ is such that $A\mathbf{x} = \mathbf{b}$, then

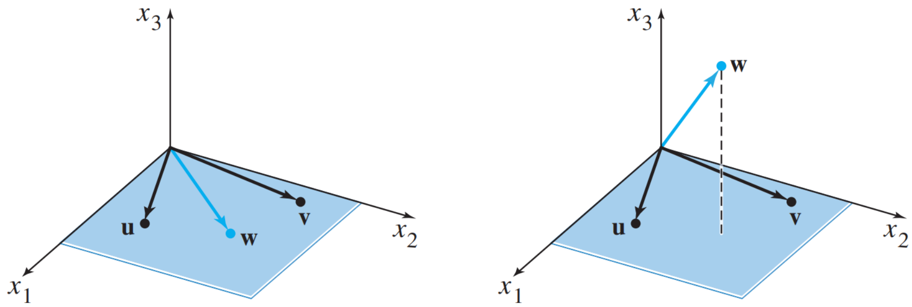
$$A(\mathbf{x} + \mathbf{z}) = A\mathbf{x} + A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

These last two facts yield the following theorem:

Theorem 3.2. *If the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}_0$, then $\boldsymbol{y}$ is also a solution if and only if $\boldsymbol{y} = \boldsymbol{x}_0 + \boldsymbol{z}$ for some vector $\boldsymbol{z} \in N(A)$.*

We call the vector $\mathbf{x}_0$ the *particular solution* of $A\mathbf{x} = \mathbf{b}$ and consider $N(A)$ to be the set of *general solutions*. This comes up in the study of linear differential equations.

Section 3.3
Linear Independence



Objectives:

- Discuss relationships of linearly independent vectors
- Relate linear independence of column vectors to invertibility of matrices
- Discover the Wronskian as a way to find linearly independent functions

Example 3.3.1. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ be the **standard basis vectors** for $\mathbb{R}^3$. Determine whether $\left\{ \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ forms a spanning set for $\mathbb{R}^3$.

While the set above certainly *does* form a spanning set, it seems a little larger than it needs to be. The vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ seem to be able to do all the work - what is the need for $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$?

Perhaps we are missing something - is there a vector in $\mathbb{R}^3$ that we need $\begin{bmatrix}1\\1\\0\end{bmatrix}$ for?

Example 3.3.2. Show that $\begin{bmatrix}1\\1\\0\end{bmatrix}$ is a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Then say $\mathbb{R}^3 \ni \mathbf{v} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \alpha_3\mathbf{e}_3 + \beta\begin{bmatrix}1\\1\\0\end{bmatrix}$. Show that we could write $\mathbf{v}$ as a linear combination of only $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

This motivates the following definition:

Definition 3.3.1. We say that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a vector space V are **linearly dependent** if one vector can be written as a linear combination of the others. More formally, there exist scalars $c_1, c_2, \ldots, c_n$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}.$$

Example 3.3.3. Show that $\{\mathbf{0}\}$ is linearly dependent.

Example 3.3.4. Show the set from Example 3.3.1 is linearly dependent.

Example 3.3.5. Show that the set $\{\mathbf{e}_1, \mathbf{e}_2\}$ is **linearly independent** in $\mathbb{R}^3$ - that is, show that if $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 = \mathbf{0}$, then *both* c_1 *and* c_2 *must* be zero.

While the example above is "small enough" to be linearly independent, it is now too small to be a spanning set - there is no way the span of these two vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ can reach $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

So $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a minimal spanning set of $\mathbb{R}^3$, which we call a **basis** of $\mathbb{R}^3$.

Example 3.3.6. Show that $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$.

In a geometric sense of $\mathbb{R}^2$ or $\mathbb{R}^3$, one can look at the figure at the beginning of this section for some intuition. Note that in the left graphic, $\mathbf{w}$ is linearly *dependent* on $\mathbf{u}$ and $\mathbf{v}$ since it is *contained within the same plane that contains* $\boldsymbol{u}$ *and* $\boldsymbol{v}$. In the right graphic, $\mathbf{w}$ can no longer be written as a linear combination of $\mathbf{u}$ and $\mathbf{v}$ since it is in a different plane, so it is linearly *independent*.

Note that, once these vectors above are made column vectors of a matrix, linear independence is equivalent to saying that

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{0}$$

has *only* the trivial solution. We have previously proven that this is equivalent to the determinant of A being nonzero. So the above example helps us see an easy trick to proving linear independence. We summarize our results here:

Theorem 3.2. *Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ be n vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix}$. Then vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ are linearly dependent if and only if A is singular.*

Example 3.3.7. Determine whether the vectors $\begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix}$ are linearly independent.

Example 3.3.8. Determine whether the vectors $p_1(x) = x^2 - 2x + 3$, $p_2(x) = 2x^2 + x + 8$, and $p_3(x) = x^2 + 8x + 7$ are linearly independent in P_3.

From this last example we have learned that, given n equations and n unknowns (or equivalently an $n \times n$ matrix), the matrix of coefficeints is invertible if and only if the column vectors span and are linearly independent. We formalize this in a theorem:

Theorem 3.3. *Let A be an $n \times n$ matrix. Then A is an invertible matrix if and only if its columns form a basis of $\mathbb{R}^n$.*

We end this section with a discussion of linear independence in the set $C^{(n-1)}[a, b]$. We begin by assuming we have n functions $f_1, \dots, f_n$ that are linearly dependent. That is,

$$c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) = 0$$

for *all* $x \in [a, b]$ and for constants $c_1, \dots, c_n$, not all zero. Taking derivatives of both sides we get the equation

$$c_1 f_1'(x) + c_2 f_2'(x) + \cdots + c_n f_n'(x) = 0.$$

If we continue taking derivatives of both sides as long as we are able to, we get the system

$$\begin{aligned} c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) &= 0 \\ c_1 f_1'(x) + c_2 f_2'(x) + \cdots + c_n f_n'(x) &= 0 \\ \vdots &= \vdots \\ c_1 f_1^{(n-1)}(x) + c_2 f_2^{(n-1)}(x) + \cdots + c_n f_n^{(n-1)}(x) &= 0. \end{aligned}$$

For each fixed x, then, we get the matrix equation

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

From just our beginning assumption, we have that the nontrivial solution $(c_1, c_2, \dots, c_n)^T$ solves this system. We say that the *determinant* of the matrix on the left is called the **Wronskian** of $f_1, \dots, f_n$.

In the case that this nontrivial solution exists, the determinant of this matrix (or the Wronskian) is equal to 0. If there is *no* nontrivial solution for some x, however, then this means our functions $f_1, \dots, f_n$ are linearly independent. This gives us the following theorem:

Theorem 3.4. *Let $f_1, f_2, \dots, f_n$ be elements of $C^{(n-1)}[a, b]$. If there exists a point x_0 in $[a, b]$ such that $W[f_1, f_2, \dots, f_n](x_0) \neq 0$, then $f_1, f_2, \dots, f_n$ are linearly independent.*

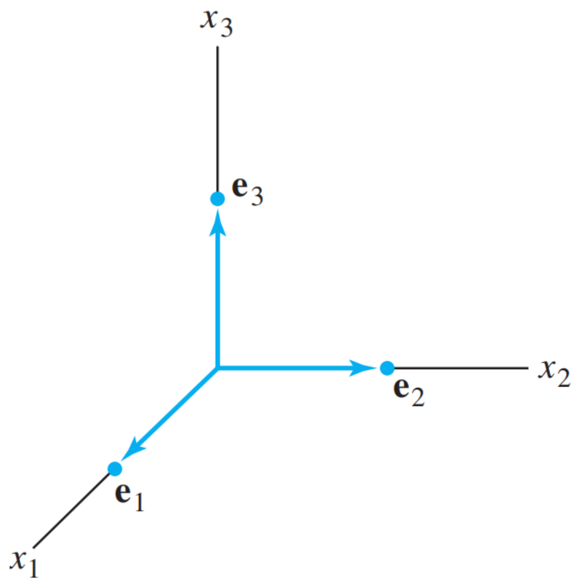
NOTE: functions that are differentiable *more* than $n-1$ times are still in the vector space $C^{(n-1)}[a,b]$. For example, $e^x \in C^{(n-1)}$ for *all* values $n \geq 2$.

Example 3.3.9. Show that e^x and e^{-x} are linearly independent in $C(-\infty,\infty)$.

Example 3.3.10. Show that the vectors $1, x, x^2, x^3$ are linearly independent in $C(-\infty,\infty)$.

Section 3.4

Basis and Dimension



Objectives:

- Compute various bases for vector spaces
- Discover theorems relating all bases of a vector space together

Example 3.4.1. Use your work in Example 3.2.6 to find a basis for the nullspace of A.

We reiterate a discovery from the previous section:

Definition 3.4.1. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form a **basis** for a vector space V if and only if the following are true:

(i) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent; and

(ii) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span $\mathbb{R}^n$.

We show that all bases are connected: they all have the same size.

Theorem 3.3. *If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is a basis for a vector space V, then ANY collection of m vectors in V, where $m > n$, is linearly dependent.*

Proof. Let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be m vectors in V, where $m > n$. Then since these are in V and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a spanning set for V, we have the equations

$$\mathbf{u}_i = a_{i1}\mathbf{v}_1 + a_{i2}\mathbf{v}_2 + \cdots + a_{in}\mathbf{v}_n \quad \text{for} \quad i \in [m].$$

Now consider a linear combination $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m$. If we show that there exist $c_1, \ldots, c_m$, not all zero, such that this combination equals zero, then by definition $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ are linearly dependent and we are done. We just mentioned that each of the vector $\mathbf{u}_1, \ldots, \mathbf{u}_m$ can be written as a linear combination of the $\mathbf{v}_i$ using the coefficients a_{ij} above. Plugging these into this linear combination, we get that

$$c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \left(\sum_{i=1}^{m} a_{i1}c_i\right)\mathbf{v}_1 + \cdots + \left(\sum_{i=1}^{m} a_{in}c_i\right)\mathbf{v}_n.$$

Now $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, so if this equals $\mathbf{0}$ each of these sums are equal to 0. But this yields n equations of the form $\sum_{i=1}^{m} a_{ij}c_i = 0$ - one for each $\mathbf{v}_i$ - with m unknowns, c_1 through c_m. Since $m > n$, this is an underdetermined homogeneous system and hence has a nontrivial solution $\hat{c}_1, \ldots, \hat{c}_m$, which gives us our nontrivial solution to the original linear combination

$$\hat{c}_1\mathbf{u}_1 + \cdots + \hat{c}_m\mathbf{u}_m = \mathbf{0}.$$

So these vectors are linearly dependent, and we are done. □

From this immediately follows the following theorem:

Theorem 3.4. *All bases for a vector space V have the same size - this is called the **dimension** of V.*

Proof. We have just shown any collection bigger than a basis is linearly dependent and hence cannot be a basis. Hence if we have two bases, one cannot be bigger than the other, so they must be the same size. □

Example 3.4.2. Show that the functions $1, x, x^2, \ldots, x^{n-1}$ are all linearly independent in P, the space of all polynomials. Conclude that P is **infinite dimensional**, meaning that no finite set of vectors can span P.

The following theorems give us shortcuts to determining when we have a basis for any **finite-dimensional** vector space (which we will be dealing with most of the time).

Theorem 3.5. *If a vector space V has dimension n, then*

(i) any set of n linearly independent vectors for V is a basis for V.

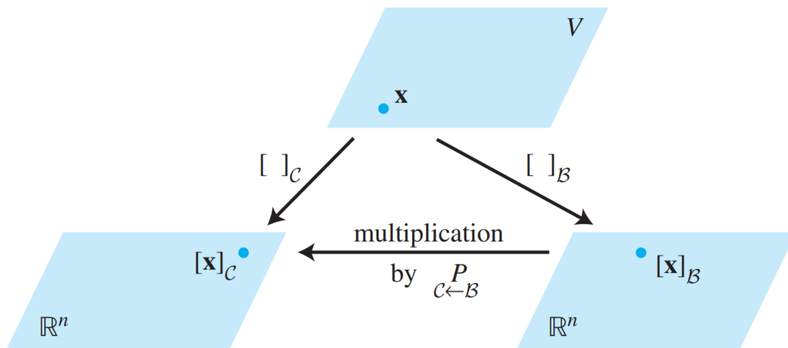
(ii) any n vectors that span V form a basis for V.

(iii) no fewer than n vectors can span V, but you can extend the set to form a basis for V by adding more vectors.

(iv) any spanning set of more than n vectors cannot be a basis for V, but it can be pared down by removing vectors to form a basis for V.

Example 3.4.3. The vectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 5 \\ 4 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$, $\mathbf{x}_4 = \begin{bmatrix} 2 \\ 7 \\ 4 \end{bmatrix}$, $\mathbf{x}_5 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ span $\mathbb{R}^3$. Pare down the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_5\}$ to form a basis for $\mathbb{R}^3$.

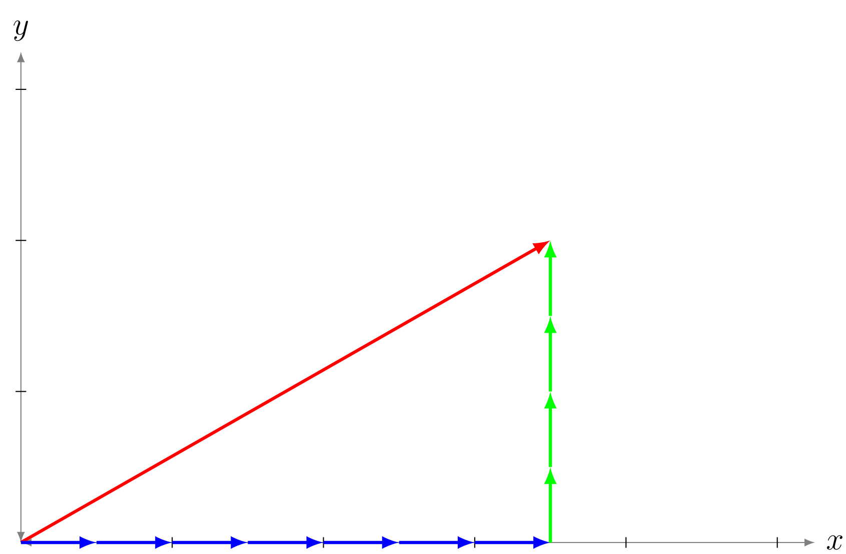
Section 3.5
Change of Basis



Objectives:

- Determine how to write vectors as linear combinations of different bases
- Define the transition matrix of a change of basis and use it to find new coordinates

The **standard coordinates** of a vector in $\mathbb{R}^2$ are written as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. For example, the vector $(7, 4)$ can also be written as $7\mathbf{e}_1 + 4\mathbf{e}_2$. Visually, this set of coordinates can be seen as the addition of several vectors end-to-end:



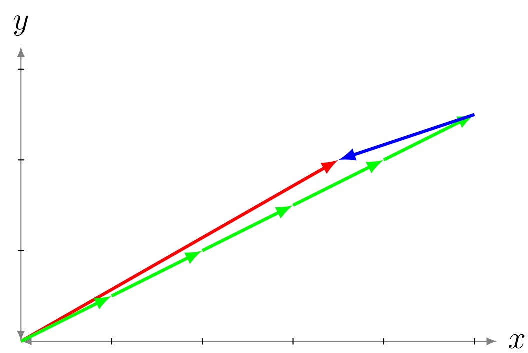
Given another basis $\mathbf{u}_1, \mathbf{u}_2$ of $\mathbb{R}^2$, we can similar write a vector in terms of these coordinates. The vector $(7, 4)$ is an element of $\mathbb{R}^2$, so it can be written as

$$\begin{bmatrix} 7 \\ 4 \end{bmatrix} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2,$$

and we say that the **coordinates** of $\begin{bmatrix} 7 \\ 4 \end{bmatrix}$ with respect to the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ is $(c_1, c_2)_{\{\mathbf{u}_1,\mathbf{u}_2\}}$. (If we ask for coordinates of a vector without specifying a basis, we mean the *standard* coordinates.)

Example 3.5.1. The vectors $\mathbf{u}_1 = (3, 1)$ and $\mathbf{u}_2 = (2, 1)$ form a basis for $\mathbb{R}^2$. Find the coordinates of the vector $\mathbf{v} = (7, 4)$ with respect to the basis $\mathbf{u}_1$, $\mathbf{u}_2$.

As an assistive visual, here are those vectors seated end-to-end ending at $(7, 4)$:



Example 3.5.2. Say that $(7,4)_{\mathbf{u}_1,\mathbf{u}_2}$ are the coordinates for a vector written with respect to the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$, where these vectors are the same as they were in the previous example. Find the coordinates of this vector with respect to the standard basis.

The matrix U whose columns are the basis vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ written in terms of the standard coordinates $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, is called the **transition matrix** from $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ to $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. This has two good uses:

(I) if we would like the coordinates of a vector $\mathbf{x} = (x_1, x_2)^T$ in terms of a basis $\{\mathbf{u}_1, \mathbf{u}_2\}$, we calculate $\mathbf{v}_{\{\mathbf{u}_1,\mathbf{u}_2\}} = U^{-1}\mathbf{x}$.

(II) if we are given the coordinates of a vector $\mathbf{v}_{\{\mathbf{u}_1,\mathbf{u}_2\}}$, we can write this vector in the standard basis by calculating $\mathbf{x} = U\mathbf{v}$.

Example 3.5.3. Let $\mathbf{u}_1 = (3,2)^T$, $\mathbf{u}_2 = (1,1)^T$, and $\mathbf{x} = (7,4)^T$. Find the coordinates of $\mathbf{x}$ with respect to $\mathbf{u}_1$ and $\mathbf{u}_2$.

We can also find the transition matrix between two non-standard bases. This takes two steps: first we transition one basis into the standard basis ($\{\mathbf{v}_1, \mathbf{v}_n\}$ into $\{\mathbf{e}_1, \mathbf{e}_2\}$), then transition the standard basis into the second basis ($\{\mathbf{e}_1, \mathbf{e}_2\}$ into $\{\mathbf{u}_1, \mathbf{u}_2\}$).

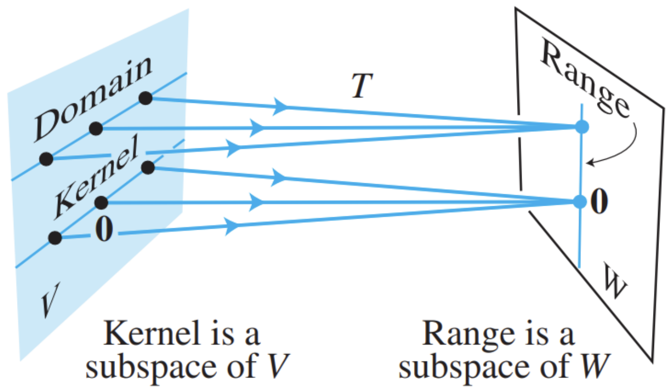
Example 3.5.4. Find the transition matrix corresponding to the change of basis from $\{\mathbf{v}_1, \mathbf{v}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \quad \text{and} \quad \mathbf{u}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Example 3.5.5. Suppose that in P_3 we want to change from the ordered basis $[1, x, x^2]$ to the ordered basis $[1, 2x, 4x^2 - 2]$. Find the transition matrix, then calculate the coordinates for x^2 with respect to this new basis.

Section 3.6

Row Space and Column Space



Objectives:

- Define the column and row spaces of a matrix
- Connect row spaces and nullspaces with the rank-nullity theorem

Recall that the nullspace of an $m \times n$ matrix A is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$. This tells us how much of the *domain* of the matrix A gets sent to zero. A related opposing set would be the *range* of the matrix A, which would tell us all of the vectors $\mathbf{y}$ such that, for some vector $\mathbf{x}$, $A\mathbf{x} = \mathbf{y}$. This would be a subset of $\mathbb{R}^m$ (rather than $\mathbb{R}^n$, since $A\mathbf{x}$ has entries equal to the number of rows of A) and is called the **column space** of A. It turns out that the nullspace and the column space (or equivalent, the nullspace and the row space) have sizes that are strictly related to each other - this will be the rank-nullity theorem that we will discuss later in this section.

We make the following definitions:

Definition 3.6.1. If A is an $m \times n$ matrix, the subspace of $\mathbb{R}^{1\times n}$ spanned by the row vectors of A is called the **row space** of A, row(A). The subspace of $\mathbb{R}^m$ spanned by the column vectors of A is called the **column space** of A, col(A).

Note that the row space of A is the column space of A^T, and vice versa.

Example 3.6.1. Describe the row and column spaces of $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Note that any of our elementary row operations on A only replace rows with linear combinations of the rows. Hence the *row space does not change under the elementary row operations.* We will use this to our advantage in this example.

Example 3.6.2. Define the **rank** of a matrix A to be the dimension of the row space of A. Find the rank of

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

We will similarly define the **nullity** of a matrix A to be the dimension of the null space of A. The matrix in Example 3.2.6 has a nullity of 2 - as we saw in Example 3.4.1, it has a basis of 2 vectors. This leads us to the major theorem of this section:

Theorem 3.6 (Rank-Nullity Theorem)**.** *If A is an $m \times n$ matrix, then the rank of A plus the nullity of A is equal to the number of columns of A. That is,* $\text{rank}(A) + \text{nullity}(A) = n$.

Proof. Let U be the reduced row echelon form of A. The system $A\mathbf{x} = \mathbf{0}$ is the equivalent to the system $U\mathbf{x} = \mathbf{0}$ by our first theorem from this section.

If A has rank r, then U will have r non-zero rows and hence r leading 1's. Consequently the system $U\mathbf{x} = \mathbf{0}$ will have $n - r$ free variables.

The dimension of $N(A)$ is equal to the dimension of the free variables, since these form the general solutions. So $\text{nullity}(A) = n - r$. Since A has rank r, we have $\text{rank}(A) + \text{nullity}(A) = r + (n - r) = n$ as desired. □

Example 3.6.3. Let $A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 4 & -3 & 0 \\ 1 & 2 & 1 & 5 \end{bmatrix}$. Find a basis for the row space of A and a basis for $N(A)$. Verify that the rank-nullity theorem applies for A.

Let's look back at the matrix U formed in the proof for the rank-nullity theorem, which was formed by writing A in reduced row-echelon form. Note that the dimension of the *column space* of U is also equal to r, the rank of A, since we can just consider the columns with leading 1's to be a basis. However, column space is *not* preserved in general by row equivalence. So this doesn't necessarily say (yet) that $\dim \text{col}(A) + \text{nullity}(A) = n$.

But let's play a trick. Let's delete the columns of U that *don't* have leading 1's. Call that U_L, for "matrix of leading 1's". Delete those same columns in A and call it A_L. Well, U_L is still the row reduction of A_L, so if $A_L\mathbf{x} = \mathbf{0}$ has a solution, then $U_L\mathbf{x} = \mathbf{0}$ has the same solution.

Now that we've deleted all the free-variables, the columns of U_L are linearly independent. So the only solution to $U_L\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. Hence this is the only solution to $A_L\mathbf{x} = \mathbf{0}$. This implies that the columns of A_L are linearly independent.

So returning to the matrix A, which has *more* columns, we know that the dimension of the column space of A is *at least* the dimension of the column space of A_L (which is the same dimension as U_L, which is r). So the dimension of the column space is *at least* the rank of A.

On the other hand, we can repeat this argument for A^T, where row space and column space switch roles (see our comment right after our first definition). So we have the following:

$$\begin{aligned} r = \operatorname{rank}(A) &= \dim \operatorname{col}(A^T) \\ &\geq \operatorname{rank}(A^T) \\ &= \dim \operatorname{col}(A). \end{aligned}$$

This leaves the following theorem:

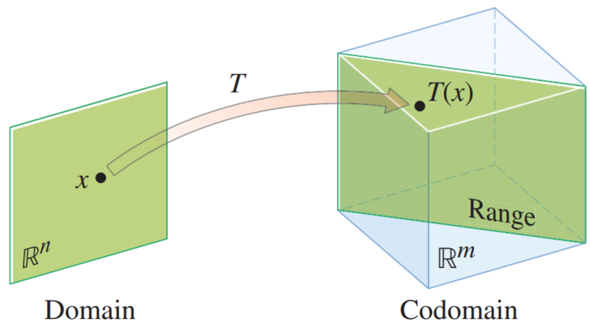
Theorem 3.7. *If A is an $m \times n$ matrix, the dimension of the row space is equal to the dimension of the column space of A.*

Example 3.6.4. Let

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

Find a basis for the column space of A. (Hint: first row reduce to a matrix U, then delete the columns in both A and U corresponding to the free variables in U.)

Section 4.1
Linear Transformations



Objectives:

- Define and see examples of linear transformations between vector spaces
- Calculate ranges and kernels of linear transformations

Recall that a vector space, as defined at the beginning of Chapter 3, is closed under *vector addition* and *scalar multiplication*. There are plenty of functions on these vector spaces: for example, the function $f(x) = (x^2, x^3)$ is *a function from* $\mathbb{R}$ *to* $\mathbb{R}^2$ since it takes a real number x and returns a coordinate (x^2, x^3) written in terms of the input x.

The function $f(x) = (x^2, x^3)$, however, does *not* satisfy the linearity conditions we discussed in Section 0.1. This is where we asked two questions. For x, y in the domain of f and a scalar multiple r,

Is $f(x + y) = f(x) + f(y)$**?** $\quad f(x+y) = ((x+y)^2, (x+y)^3) \neq (x^2+y^2, x^3+y^3) = f(x)+f(y).$

$$\textbf{Is } f(rx) = rf(x)\textbf{?} \quad f(rx) = ((rx)^2, (rx)^3) \neq (rx^2, rx^3) = rf(x).$$

This means that f is *not* a linear transformation from $\mathbb{R}$ to $\mathbb{R}^2$. Even if the function failed *one* of these two conditions, the mapping f could not be linear. We formalize that definition here:

Definition 4.1.1. A mapping $L : V \to W$ from a vector space V into a vector space W is said to be a **linear transformation** if

$$L(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \alpha L(\mathbf{v}_1) + \beta L(\mathbf{v}_2)$$

for all $\mathbf{v}_1, \mathbf{v}_2$ in V and for all scalars α, β. If $V = W$, we may also say L is a **linear operator** on V.

Example 4.1.1. Show that $L : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $L(\mathbf{x}) = 3\mathbf{x}$ is linear in *two* different ways.

Example 4.1.2. For a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, show that the mapping $L : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $L(\mathbf{x}) = x_1\mathbf{e}_1$ is linear.

Example 4.1.3. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $L(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x}$. Show that L is a linear operator.

Example 4.1.4. Show that $L : \mathbb{R}^2 \to \mathbb{R}^1$ defined by $L(\mathbf{x}) = x_1 + x_2$ is a linear transformation (*not* an operator).

Example 4.1.5. Show that the mapping $M : \mathbb{R}^2 \to \mathbb{R}^1$ defined by $M(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$ is *not* linear.

Example 4.1.6. Let $I : C[a, b] \to \mathbb{R}^1$ defined by

$$L(f) = \int_a^b f(x)\, dx.$$

Show that I is linear.

Example 4.1.7. Let $D : C^1[a, b] \to C[a, b]$ be given by

$$D(f) = f'.$$

Show that D is linear.

Given a linear transformation $L : V \to W$, we can use the properties of linear transformations to find where $L(\mathbf{0}_V)$ goes. (Here we are using $\mathbf{0}_V$ to denote the zero vector of V, It may be different than the zero vector of W, $\mathbf{0}_W$, but they share the same properties of zero vectors in their respective spaces.) For example, we know from Section 3.1 that $0 \cdot \mathbf{v} = \mathbf{0}_V$ for any vector $\mathbf{v} \in V$. So

$$L(\mathbf{0}_V) = L(0 \cdot \mathbf{v}) \overset{L \text{ linear}}{=} 0L(\mathbf{v}) = \mathbf{0}_W.$$

So *a linear transformation always maps $\mathbf{0}_V$ to $\mathbf{0}_W$*. This is different than most lines we know, which can have varying y-intercepts. *Linear transformations in $\mathbb{R}^n$ always go through the origin.* (However, linear *functions* of the form $y = ax + b$ do not have to go through the origin. In this class we will use "transformation" to make the distinction.)

We can use this fact to find out, using the properties of $-\mathbf{v}$, that

$$\mathbf{0}_W = L(\mathbf{0}_V) = L(\mathbf{v} - \mathbf{v}) \overset{L \text{ linear}}{=} L(\mathbf{v}) + L(-\mathbf{v}).$$

Once we subtract by $L(-\mathbf{v})$ on both sides, it follows that $L(-\mathbf{v}) = -L(\mathbf{v})$.

Example 4.1.8. The mapping $L : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $L(\mathbf{x}) = (x_2, x_1, x_1 + x_2)^T$ is linear. Find the **matrix of** L.

Now that we have established a connection between matrices and transformations in the previous example, let's define some corresponding subspaces. See if you recognize what these definitions would be if the transformation L were replaced by a matrix.

Definition 4.1.2. Given a linear transformation $L : V \to W$, we say the **kernel** of L is defined by

$$\ker(L) = \{\mathbf{v} \in V : L(\mathbf{v}) = \mathbf{0}_W\}.$$

The **range** of L is defined as

$$\operatorname{ran}(L) = \{\mathbf{w} \in W : \mathbf{w} = L(\mathbf{v}) \text{ for some } \mathbf{v} \in S\}.$$

One should think of the kernel of L like the nullspace of a matrix. The range of L, on the other hand, is like the column space of a matrix. We will be connecting these concepts in the next section. Much like the nullspace and column space of a matrix, we get the following theorem:

Theorem 4.1. *If $L : V \to W$ is a linear transformation, then*

(i) $\ker(L)$ *is a subspace of* V.

(ii) $\operatorname{ran}(L)$ *is a subspace of* W.

Example 4.1.9. Let L be a linear operator on $\mathbb{R}^2$ be defined by $L(\mathbf{x}) = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$. Find the kernel and range of L and state their dimensions.

Example 4.1.10. Let $D : P_3 \to P_3$ be the differentiation operator. Find the kernel and range of D and state their dimensions.

Section 4.2

Matrices of Transformations

Objectives:

- Represent linear transformations as matrix multiplications
- Use matrices to construct dilations, shifts, rotations, and other transformations in $\mathbb{R}^2$

Example 4.2.1. Let $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ be a vector and let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates a given vector by 90 degrees counterclockwise. Find $L(\mathbf{v})$.

In a similar way to how we found the matrix of the transformation L in the previous example, we can find a matrix corresponding to *any* linear transformation:

Theorem 4.2. *If L is a linear transformation mapping $\mathbb{R}^n$ into $\mathbb{R}^m$, there is an $m \times n$ matrix A such that*

$$L(\boldsymbol{x}) = A\boldsymbol{x}$$

for each $\boldsymbol{x} \in \mathbb{R}^n$. In fact, the jth column vector of A is given by $\boldsymbol{a}_j = L(\boldsymbol{e}_j)$.

The matrix A defined in this way is called the **matrix transformation** of L, or simply the **matrix** of L.

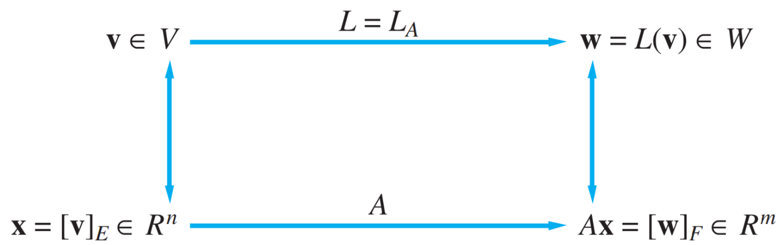
Example 4.2.2. Define the linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$L(\mathbf{x}) = (x_1 + x_2, x_2 + x_3)^T.$$

Using the theorem above, find the matrix of L.

We return to Section 3.5 and recall that a transition matrix U from a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ to $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ helps us transition from the coordinates in one basis to back to the standard coordinates. Say we wish to write a linear transformation $L : V \to W$ with respect to the basis $E = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in V and the basis $F = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ in W.

Recall that before we have to construct the transition matrix U_1 from E to the standard basis, as well as the transition matrix U_2 from F to the standard basis, then get the coordinate matrix $U_2^{-1}U_1$ to go from E to F. We are now inserting a linear transformation into the middle: once we have transitioned out of the basis of V into the standard basis, we map our vectors through the transformation L, then map them back into their desired final basis. This gives us the following:



Theorem 4.3. *Let $L : V \to W$ be a linear transformation. Let A be the matrix of L. Further suppose that $E = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is a basis of V with transition matrix U_1, and $F = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ is a basis of W with transition matrix U_2. Then the* ***matrix transformation*** *of L with respect to the bases E and F is given by*

$$U_2^{-1} A U_1.$$

We can follow this path using the diagram given above.

Example 4.2.3. Let $L : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation defined by

$$L(\mathbf{x}) = (x_2, x_1 + x_2, x_1 - x_2)^T.$$

Find the matrix representation of L with respect to the ordered bases $\{\mathbf{u}_1, \mathbf{u}_2\}$ and $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, where

$$\mathbf{u}_1 = (1, 2)^T, \quad \mathbf{u}_2 = (3, 1)^T, \quad \mathbf{b}_1 = (1, 0, 0)^T, \quad \mathbf{b}_2 = (1, 1, 0)^T, \quad \mathbf{b}_3 = (1, 1, 1)^T.$$

NOTE: If a linear transformation is *already given* in terms of the bases we need, then the matrix transformation can be calculated much more quickly quickly.

Example 4.2.4. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by

$$L(\alpha \mathbf{b}_1 + \beta \mathbf{b}_2) = (\alpha + \beta)\mathbf{b}_1 + 2\beta \mathbf{b}_2,$$

where $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Find the matrix representation of L with respect to $\{\mathbf{b}_1, \mathbf{b}_2\}$.

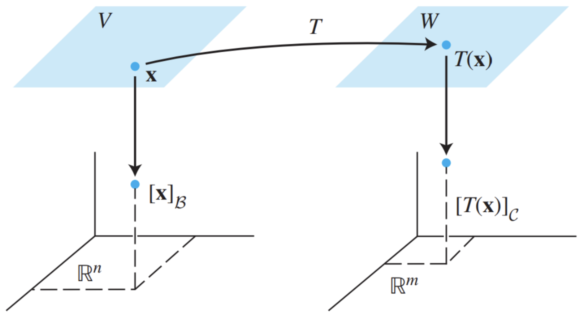
Example 4.2.5. Let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation defined by

$$L(\mathbf{x}) = x_1 \mathbf{b}_1 + (x_2 + x_3)\mathbf{b}_2,$$

where $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Find the matrix of L with respect to the ordered bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and $\{\mathbf{b}_1, \mathbf{b}_2\}$.

Here is a page to take notes on common $\mathbb{R}^2$ transformations.

Section 4.3
Similarity



Objectives:

- Discover a common formula used when analyzing linear operators in different bases

We begin this chapter by analyzing linear *operators* $L : V \to V$ and their matrix representations with respect to non-standard bases.

Example 4.3.1. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $L(\mathbf{x}) = (2x_1, x_1 + x_2)$. Write the matrix of L with respect to the ordered basis $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.

Note that our formula for finding the matrix transformation simplifies to

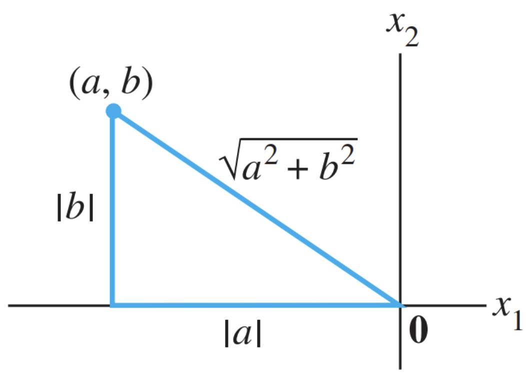
$$B = U^{-1}AU$$

when L is a *linear operator* and when we are using only *one* ordered basis rather than two. If two matrices A and B satisfy the equation above for *any* invertible matrix U, we say that A is **similar** to B (or, equivalently, that B is similar to A, or simply that A and B are similar).

Example 4.3.2. Let D be the differentiation operator on P_3. Find the matrix A representing D with respect to $\{1, x, x^2\}$. Then find the matrix B representing D with respect to $\{1, 2x, 4x^2 - 2\}$.

Example 4.3.3. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ by defined by $L(\mathbf{x}) = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix} \mathbf{x}$. Find the matrix representing L with respect to the ordered basis $\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Section 5.1
The Dot Product



Objectives:

- Review vector geometry: projections, dot product, distance, and orthogonality

Motivation: Orthogonal Basis

In Section 1.3 we discussed that **orthogonal** vectors - vectors which make a right-angle to each other when both are placed in the same plane - are characterized by having a dot product of 0. We say a collection $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ is an **orthogonal basis** if the following two things are true:

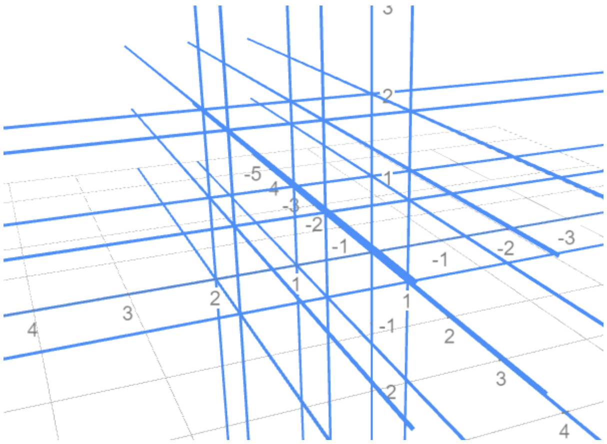
(i) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ forms a basis for $\mathbb{R}^n$

(ii) if $i \neq j \in [n]$, then $\mathbf{v}_i$ is orthogonal to $\mathbf{v}_j$.

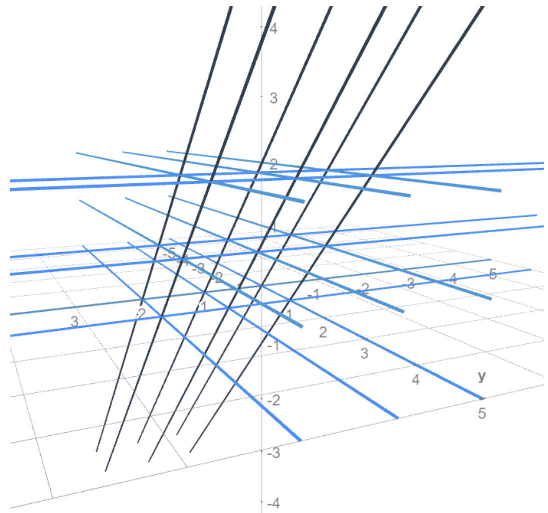
Example 5.1.1. Determine whether the following bases are orthogonal bases on $\mathbb{R}^3$.

(a) $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ (b) $\{\begin{bmatrix}1 & 0 & 0\end{bmatrix}^T, \begin{bmatrix}0 & 1 & 0\end{bmatrix}^T, \begin{bmatrix}0 & 1 & 1\end{bmatrix}^T\}$

Orthogonal bases are more powerful than general bases - not only do their linear combinations completely cover a space, but the geometry of these linear combinations shows that they are each at right angles with each other.



Lattice formed by $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$



Lattice formed by $\{\mathbf{e}_1, \mathbf{e}_2, \begin{bmatrix}0 & 1 & 1\end{bmatrix}^T\}$

However, *we have never seen how to make angles with functional vector spaces*. Is there a way to make an orthogonal basis in a functional vector space? **Yes!** It is also incredibly useful, as it gives us an easy way to **reconstruct complicated functions by looking at their simpler projections onto an orthogonal basis**. This is exactly the idea behind *Fourier series*.

In order to discuss angles in functional vector spaces, we need to generalize the *dot product* into something called an **inner product**. That is the thrust of this chapter. Before beginning, let's do a quick review of vectors.

Algebra of Vectors

Let $\mathbf{a} = \langle a_1, a_2, a_3 \rangle, \mathbf{b} = \langle b_1, b_2, b_3 \rangle$ be vectors in $\mathbb{R}^3$ and let c be a **scalar**.

(a) **Scalar Multiplication** $c\mathbf{a} = c\langle a_1, a_2, a_3 \rangle = \langle ca_1, ca_2, ca_3 \rangle$

(b) **Vector Magnitude** $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$

(c) **Vector Sum** $\mathbf{a} + \mathbf{b} = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle$

(d) **Vector Difference** $\mathbf{a} - \mathbf{b} = \langle a_1 - b_1, a_2 - b_2, a_3 - b_3 \rangle$

(e) **Unit Vector** The vector $\frac{\mathbf{a}}{\|\mathbf{a}\|}$ is a **unit vector** of length 1 in the direction of $\mathbf{a}$.

(f) **Standard Unit Vectors** The vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the directions of the positive x, y, and z-axes, respectively.

Definition 5.1.1. Let $\mathbf{x}, \mathbf{y}$ be vectors in $\mathbb{R}^n$. We define the **dot product** of two vectors $\mathbf{x}$ and $\mathbf{y}$ to be

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

The **magnitude** of a vector $\mathbf{x}$ is defined as $(\mathbf{x} \cdot \mathbf{x})^{1/2} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

The **distance** between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined to be $\|\mathbf{x} - \mathbf{y}\|$.

The **angle** between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\theta = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|x\|\|y\|}\right).$$

(If $\mathbf{x}, \mathbf{y}$ are unit vectors, meaning that $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$, then this formula can be simplified to merely $\cos^{-1}(\mathbf{x} \cdot \mathbf{y})$.)

Theorem 5.1 (Cauchy-Schwarz Inequality)**.** *If $\boldsymbol{x}, \boldsymbol{y}$ are nonzero vectors in $\mathbb{R}^n$, then*

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|.$$

The only time equality holds is if one vector is a multiple of the other.

Proof. Unraveling the angle equation in the above definition and solving for $\mathbf{x} \cdot \mathbf{y}$, we get that

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta.$$

Since $|\cos\theta| \leq 1$, we get that $|\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\||\cos\theta| \leq \|\mathbf{x}\|\|\mathbf{y}\|$. If these equality holds, this must mean that $|\cos\theta| = 1$, so solving for θ we get that θ is either 0 degrees or 180 degrees. Hence the two vectors are either in the same or opposite directions, meaning the two vectors are multiple of each other. □

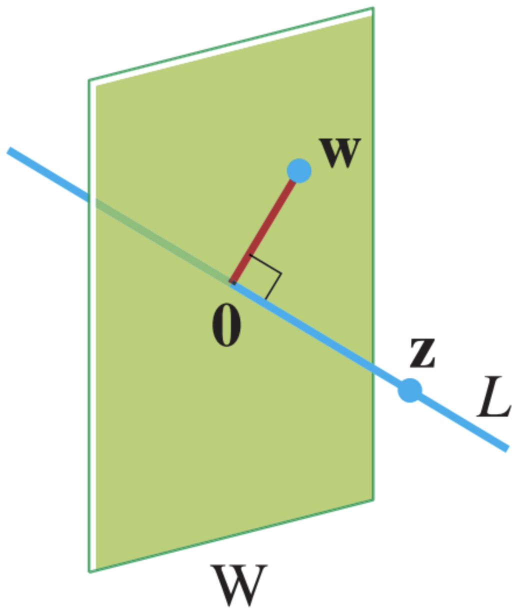
Definition 5.1.2. We define the **vector projection** of a vector $\mathbf{b}$ onto a vector $\mathbf{a}$ to be

$$\text{proj}_{\mathbf{a}}\, \mathbf{b} = \left(\frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}}\right)\mathbf{a}.$$

The length of this vector is called the **scalar projection** of $\mathbf{b}$ onto $\mathbf{a}$, which can be calculated as $\dfrac{\mathbf{b} \cdot \mathbf{a}}{\|\mathbf{a}\|}$.

Example 5.1.2. Find the projection of $\left[1, 4\right]^T$ onto the line $y = \frac{1}{3}x$.

Section 5.2
Orthogonal Subspaces



Objectives:

- Decompose $\mathbb{R}^n$ into subspaces and their orthogonal complements
- Revisit nullspaces and ranges of A, then connect these spaces with nullspaces and ranges of A^T

The concept of orthogonality can also apply to subspaces, as is pictured above. Note that, in the plane above, *any* vector in the plane would be considered orthogonal to *any* vector in the line. This is an example of the below definition:

Definition 5.2.1. Two subspaces X and Y of $\mathbb{R}^n$ are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$ for every $\mathbf{x} \in X$ and every $\mathbf{y} \in Y$. If X and Y are orthogonal, we write $X \perp Y$.

We define the **orthogonal complement** of X, $X^\perp$, to be the set of all vectors in V which are orthogonal to every vector in X.

Note that $Y^\perp$ contains *all* vectors orthogonal to Y.

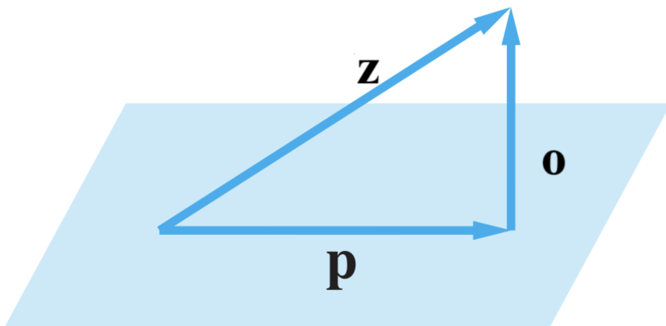
Example 5.2.1. Show that the line Y spanned by $\left[1,-1,1\right]^T$ is orthogonal to the plane $x-y+z=3$.

Example 5.2.2. Would $Y^\perp = \text{span}\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}\right\}$? Would $Y^\perp = \{(x,y,z)\in\mathbb{R}^3 : x-y+z=3\}$? Explain.

Here are some facts about orthogonal complements:

- If $\mathbf{x}\in X$, $\mathbf{x}\in Y$, and $X\perp Y$, then $\mathbf{x}$ is orthogonal to itself. So $\|x\|^2=\mathbf{x}\cdot\mathbf{x}=0$, so $\mathbf{x}=\mathbf{0}$.
- If S is a subspace of $\mathbb{R}^n$, then $S^\perp$ is also a subspace of $\mathbb{R}^n$.
- If S is a subspace of $\mathbb{R}^n$, then every vector $\mathbf{z}$ can be written uniquely as a sum of vectors $\mathbf{p}+\mathbf{o}$, where $\mathbf{p}\in S$ and $\mathbf{o}\in S^\perp$.
- If S is a subspace of $\mathbb{R}^n$, then $(S^\perp)^\perp = S$.

This third bullet point will come in handy in the next section, so let's get an intuitive reason as to why it might be true with a picture. Note that the vector $\mathbf{p}$ below is the *projection* of $\mathbf{z}$ onto the plane below, while $\mathbf{o} = \mathbf{z} - \mathbf{p}$.



Let's look at a place where we might be surprised to find orthogonal complements. Before we do this, we quickly define the **cross product** of two vectors:

Definition 5.2.2. Given two non-zero vectors $\mathbf{v} = \left[v_1, v_2, v_3\right]^T$ and $\mathbf{w} = \left[w_1, w_2, w_3\right]^T$ in $\mathbb{R}^3$, we define $\mathbf{v} \times \mathbf{w}$, the **cross product** of $\mathbf{v}$ and $\mathbf{w}$, to be

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

we recall that $\mathbf{v} \times \mathbf{w}$ has the property of being orthogonal to both $\mathbf{v}$ *and* $\mathbf{w}$.

Example 5.2.3. Find the orthogonal complement of the subspace of $\mathbb{R}^3$ spanned by $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

Example 5.2.4. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix}$. Find the bases for $N(A)$, $R(A^T)$, $N(A^T)$, and $R(A)$.

Example 5.2.5. Find the orthogonal complement of $R(A)$ and of $R(A^T)$. Then write two equations connecting pairs among $N(A)$, $R(A^T)^\perp$, $N(A^T)$, and $R(A)^\perp$.

This connection between these subspaces is not coincidental. The subspaces $N(A)$, $R(A)$, $N(A^T)$, and $R(A^T)$ are known as the **fundamental subspaces** of A. Let's try to get to the bottom of this fact.

Theorem 5.2. *If A is an $m \times n$ matrix, then $N(A) = R(A^T)^\perp$ and $N(A^T) = R(A)^\perp$.*

Proof. Let $\mathbf{x} \in N(A)$. Then by definition, $A\mathbf{x} = \mathbf{0}$, so by isolating each row of this multiplication we get

$$\begin{bmatrix} a_{i1}, a_{i2}, \dots, a_{in} \end{bmatrix}^T \cdot \begin{bmatrix} x_1, x_2, \dots, x_n \end{bmatrix}^T = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0$$

for $i \in [m]$ being any row of the matrix A.

This row of A would be a column of A^T. This means that $\begin{bmatrix} a_{i1}, a_{i2}, \dots, a_{in} \end{bmatrix}^T$ is in the column space of A^T, or in the range of A^T. So the above says that the dot product of any column of A^T and any element in $N(A)$ is zero, which is the definition of orthogonal.

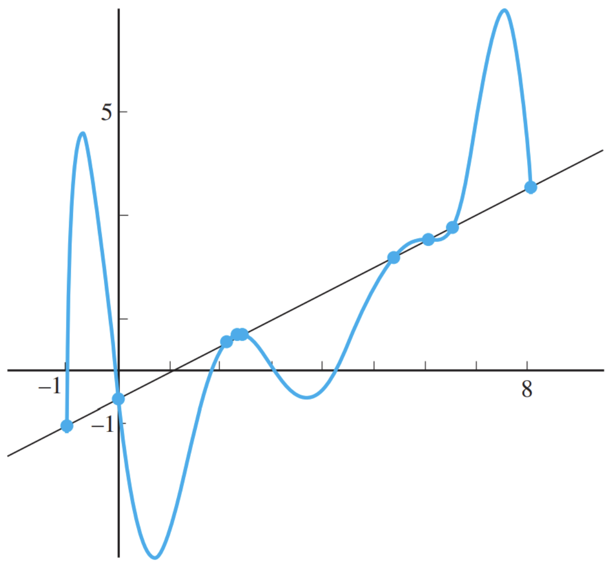
On the other hand, if $\mathbf{x} \in R(A^T)^\perp$, then $\mathbf{x}$ is orthogonal to each of the column vectors of A^T, or all the row vectors of A. By retracing our steps through the paragraphs above, this means that $A\mathbf{x} = \mathbf{0}$, which means $\mathbf{x} \in N(A)$. So $N(A) = R(A^T)^\perp$ as desired. The second equation comes from applying this result to $B = A^T$:

$$N(B) = R(B^T)^\perp \iff N(A^T) = R((A^T)^T)^\perp = R(A)^\perp.$$

□

Example 5.2.6. Find $R(A)$ if $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$. Then use the Fundamental Subspaces theorem to find $N(A^T)$.

Section 5.3
Application: Least Squares Problems



Objectives:

- Use linear regression to find least-squares solution lines to more complicated data lines

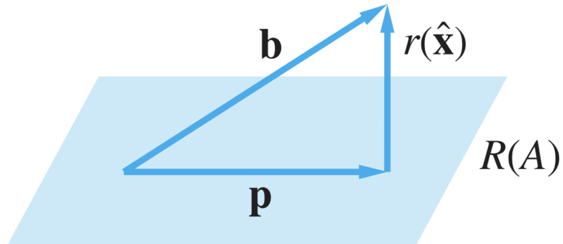
At this point we have gathered enough theory to develop a useful tool for data regression. When given an *overdetermined* system of equations of the form $A\mathbf{x} = \mathbf{b}$ for a fixed matrix A and vector $\mathbf{b}$, it is typical that we will not find a solution $\mathbf{x}$. This is because we have more equations than unknowns - there are so many constraints on the set of what $\mathbf{x}$ can be. However, we have now developed a notion of *distance* in $\mathbb{R}^m$, so we can set a new goal:

$$\text{Find } \mathbf{x} \in \mathbb{R}^n \text{ such that } \|\mathbf{b} - A\mathbf{x}\| \text{ is minimized.}$$

It is possible that the minimum value of $\|\mathbf{b} - A\mathbf{x}\|$ is 0. In that case, $\mathbf{b} - A\mathbf{x} = 0$, and the vector $\mathbf{x}$ satisfies the equation $A\mathbf{x} = \mathbf{b}$. Great! In any other case, the vector $\mathbf{x}$ will be the vector such that $A\mathbf{x}$ is the "closest" it can get to $\mathbf{b}$. Let's define the **residual** to be

$$r(\mathbf{x}) := \|\mathbf{b} - A\mathbf{x}\|.$$

Another context where we are finding the "closest" point to a given line or plane is with projections... so our picture from our previous section will be useful to bring back.



Note that the vector $\mathbf{b}$ is not in the subspace $R(A)$, the range of A. So there is no vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$, because $\mathbf{b}$ is not in the range of A. However, the vector $\mathbf{p}$ - the projection of $\mathbf{b}$ onto $R(A)$ - is very close to $\mathbf{b}$. Indeed, at this point $r(\mathbf{x})$ will be the smallest possible distance!

Let $\hat{\mathbf{x}}$ be the value of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{p}$. Then note that $r(\hat{\mathbf{x}}) \in R(A)^{\perp}$. From our previous section we know that $R(A)^{\perp} = N(A^T)$. So $\mathbf{b} - A\hat{\mathbf{x}} \in N(A^T)$, meaning that

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \quad \iff \quad A^T A\mathbf{x} = A^T\mathbf{b}.$$

Is this new equation useful? It is: $A^T A$ is now an $n \times n$ matrix, meaning that there are n equations with n unknowns. If A has linearly independent vectors, then so does $A^T A$, and we get the following theorem:

Theorem 5.3. *If A is an $m \times n$ matrix of rank n, the* ***normal equations***

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

have a unique solution

$$\hat{x} = (A^T A)^{-1} A^T \boldsymbol{b}.$$

The vector $\hat{x}$ is the unique ***least squares solution*** *of the system $A\boldsymbol{x} = \boldsymbol{b}$.*

Example 5.3.1. Find the least squares solution of the system

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

Example 5.3.2. Given the data $\begin{array}{c|c|c|c} x & 0 & 3 & 6 \\ \hline y & 1 & 4 & 5 \end{array}$, find the best least squares fit by a linear function.

Example 5.3.3. Find the best quadratic least squares fit to the data:

x	0	1	2	3
y	3	2	4	4

Section 5-4

Inner Product Spaces

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$

Objectives:

- Generalize the dot product for $\mathbb{R}^n$ to the concept of inner product for a special class of vector spaces
- Redefine vector geometry: projections, distance, and orthogonality

The definition in the graphic above is an example of an inner product on $C[a, b]$. Much like the dot product, it inputs two vectors (in this case, functions f and g), and outputs a numerical value. This mapping needs to satisfy a few more conditions in order to be called an inner product:

Definition 5.4.1. An **inner product** on a vector space V is an operation on V that assigns to each pair of vectors $\mathbf{x}, \mathbf{y} \in V$ a real number $\langle \mathbf{x}, \mathbf{y}\rangle$ satisfying the following:

(i) $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$ with equality iff $\mathbf{x} = \mathbf{0}$; we define $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ to be the **norm** of $\mathbf{x}$.

(ii) $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$ (that is, the order of $\mathbf{x}$ and $\mathbf{y}$ doesn't matter in the inner product)

(iii) $\langle \cdot, \cdot\rangle$ is *linear* in the first argument; that is,

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z}\rangle = \alpha\langle \mathbf{x}, \mathbf{z}\rangle + \beta\langle \mathbf{y}, \mathbf{z}\rangle.$$

A vector space $(V, +, \cdot, \langle \cdot, \cdot\rangle)$ with an inner product is called an **inner product space**.

Note that (ii) and (iii) together combine to show that the inner product is *also* linear in the second argument, because

$$\langle \mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z}\rangle = \langle \alpha\mathbf{y} + \beta\mathbf{z}, \mathbf{x}\rangle = \alpha\langle \mathbf{y}, \mathbf{x}\rangle + \beta\langle \mathbf{z}, \mathbf{x}\rangle = \alpha\langle \mathbf{x}, \mathbf{y}\rangle + \beta\langle \mathbf{x}, \mathbf{z}\rangle.$$

Example 5.4.1. Show that the dot product is an inner product on $\mathbb{R}^2$.

Example 5.4.2. Define $\langle f, g\rangle = \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)g(x)\,dx$ as the inner product on $C[-\pi, \pi]$. Find $\langle \sin(x), \cos(x)\rangle$.

Given an inner product space V, we say that two vectors $\mathbf{u}, \mathbf{v} \in V$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Now that we can define angles in inner product spaces, we can make similar theorems about inner product spaces that we can in $\mathbb{R}^n$:

Theorem 5.3 (The Pythagorean Law)**.** *If $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal vectors in an inner product space V, then*

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2.$$

Example 5.4.3. Show that $\sin(x) \perp \cos(x)$ in $C[-\pi, \pi]$. Then find $\|\sin(x) + \cos(x)\|$.

Theorem 5.4 (Cauchy-Schwarz Inequality). *If* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are any two vectors in an inner product space* V*, then*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|.$$

The proof of this theorem is slightly trickier, since we have not defined a bona fide angle θ yet. We will do so once we have proven this theorem.

Proof. If either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ this is clear, so let $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$. Let $\mathbf{p}$ be the vector projection of $\mathbf{u}$ onto $\mathbf{v}$. Then we have two facts:

(i) $\|\mathbf{u}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{u} - \mathbf{p}\|^2$. (Pythagorean Law)

(ii) $\dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|} = \|\mathbf{p}\|$. (Definition of scalar projection)

Combining these two facts we get the equation

$$\frac{(\langle \mathbf{u}, \mathbf{v} \rangle)^2}{\|\mathbf{v}\|^2} = \|\mathbf{p}\|^2 = \|\mathbf{u}\|^2 - \|\mathbf{u} - \mathbf{p}\|^2.$$

Multiplying both sides by $\|\mathbf{v}\|^2$ we get

$$(\langle \mathbf{u}, \mathbf{v} \rangle)^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{p}\|^2\|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2.$$

Therefore, taking square roots of both sides gives us that $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$, as desired. □

We define the **angle** θ between two non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ to be

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

Example 5.4.4. Let $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$ be the inner product on $C[0, 1]$. Find the angle between the vectors 1 and x on $C[0, 1]$.

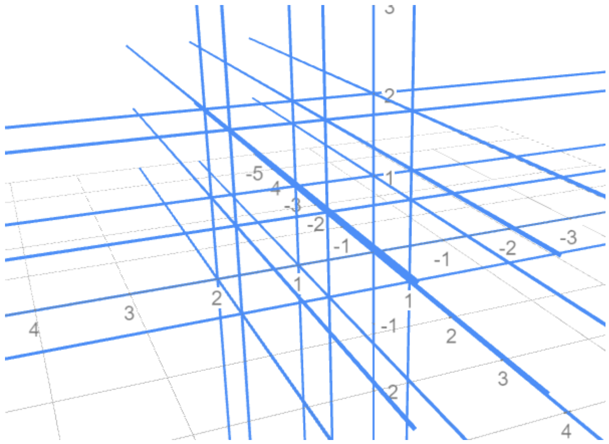
Example 5.4.5. Find the angle between the vectors 1 and x in the following vector spaces:

(a) $C[0,1]$ with the inner product $\langle f, g\rangle = \int_0^1 fg\,dx$

(b) $C[-1,1]$ with the inner product $\langle f, g\rangle = \int_{-1}^1 fg\,dx$

(c) P_2 with the inner product $\langle p, q\rangle = p(0)q(0) + p(1)q(1)$

Section 5.5
Orthonormal Sets



Objectives:

- Revisit orthogonal sets and find their applications to least squares problems
- Define orthogonal matrices and use them to compute vector projections

We begin by revisiting the definition of orthogonal basis that we introduced in Section 5-1. Recall that this definition had two components: it asked that our set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space while also having the property that $i \neq j \Rightarrow \mathbf{v}_i \perp \mathbf{v}_j$. This latter property is important enough to merit its own definition:

Definition 5.5.1. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be nonzero vectors in an inner product space V. If $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be an **orthogonal set** of vectors.

If we suppose further that $\|\mathbf{v}_i\| = 1$ for all i - that is, that each vector in the set is a unit vector - then we say that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an **orthonormal set** of vectors.

Here are a few facts about these sets of vectors:

- If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal set of vectors, then **normalizing** these vectors by dividing each one by its length does not change the angles between the vectors - it only scales them. So $\left\{ \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \ldots, \dfrac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \right\}$ is an *orthonormal* set of vectors.
- If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal set of vectors, then it is also a linearly independent set.

Let's give a brief explanation of this last item. Suppose we have this orthogonal set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ as part of the linear independence equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}.$$

Then taking the inner product with $\mathbf{v}_1$ on both sides, we get

$$\langle \mathbf{v}_1, c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \rangle = c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle + \cdots + c_n\langle \mathbf{v}_1, \mathbf{v}_n \rangle = \langle \mathbf{v}_1, \mathbf{0} \rangle = 0.$$

By definition of orthogonality, the left hand side simplifies down to just $c_1\langle \mathbf{v}_1, \mathbf{v}_1 \rangle$. Since $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle$ is not zero, we must have that c_1 is equal to 0. Repeat this process by taking the inner product with each $\mathbf{v}_i$, we get that *all* constants equal 0, as desired.

Example 5.5.1. Show that $\left\{ \left[1, 1, 1\right]^T, \left[2, 1, -3\right]^T, \left[4, -5, 1\right] \right\}$ is an orthogonal set in $\mathbb{R}^3$. Then form an orthonormal set of scalar multiples of these vectors.

Example 5.5.2. Show that $\{1, x\}$ is an orthogonal set in $C[-1, 1]$. Then form an orthonormal set of scalar multiples of these vectors.

Example 5.5.3. Show that $\{1, \cos(x), \cos(2x), \ldots, \cos(nx)\}$ is an orthogonal set of vectors in $C[-\pi, \pi]$ when equipped with the inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x)\, dx.$$

Then form an orthonormal set of scalar multiples of these vectors. (Hint: $\cos(kx)\cos(jx) = \frac{1}{2}[\cos((k-j)x) + \cos((k+j)x)]$.)

We say that an **orthonormal basis** is an orthonormal set that is also a basis. The set we found in Example 5.5.1 is an orthonormal basis, while the set we found in Example 5.5.2 is *not* an orthonormal basis. (We know this because $C[-\pi, \pi]$ contains all polynomials, which we know to be infinite dimensional from a previous section.)

Theorem 5.5. *Let $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ be an orthonormal basis for an inner product space V. Let $\boldsymbol{u}, \boldsymbol{v}$ be vectors in V. Then we have the following results:*

(i) $\boldsymbol{v} = \sum_{i=1}^{n} \langle \boldsymbol{v}, \boldsymbol{u}_i \rangle \boldsymbol{u}_i$.

(ii) If $\boldsymbol{u} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i$ *and* $\boldsymbol{v} = \sum_{i=1}^{n} b_i \boldsymbol{u}_i$*, then*

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{n} a_i b_i.$$

(iii) (Parseval's Identity) $\|\boldsymbol{v}\|^2 = \sum_{i=1}^{n} (\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle)^2$.

The first is true by orthonormality: since $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a basis, we have $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ for some constant c_i for $i \in [n]$. Then

$$\langle \mathbf{v}, \mathbf{u}_1 \rangle = \langle c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_1 \rangle = c_1\langle \mathbf{u}_1, \mathbf{u}_1 \rangle + \cdots + c_n\langle \mathbf{u}_n, \mathbf{u}_n \rangle.$$

Note that the only nonzero term in this right-hand side is the first one, which is equal to c_1 all by definition of orthonormal. Repeating this process for all other vectors $\mathbf{u}_i$ we get the result.

The second item follows directly from the first and the linearity of the inner product. The third item comes from applying the second item to the same vector and calculating $\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$, then substituting in the constants $\langle \mathbf{v}, \mathbf{u}_i \rangle$ from the first item.

Example 5.5.4. The set $\left\{ \frac{1}{\sqrt{2}}, \cos(2x) \right\}$ is orthonormal in $C[-\pi, \pi]$ with the inner product given in Example 5.5.3. The function $\sin^2(x)$ can be written as a linear combination of this orthonormal set in the following way:

$$\sin^2(x) = \left(\frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} + \left(-\frac{1}{2} \right) \cos(2x).$$

Using this information, **rewrite the integral $\int_{-\pi}^{\pi} \sin^4(x)\,dx$ as an inner product and calculate its value**.

We note that $\int_{-\pi}^{\pi} \sin^4(x)\,dx = \pi \cdot \frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2(x) \cdot \sin^2(x)\,dx = \pi \langle \sin^2(x), \sin^2(x) \rangle = \pi \|\sin^2(x)\|^2$.

Now we can use the Pythagorean Law!

$$\|\sin^2(x)\|^2 = \left\| \left(\frac{1}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} \right\|^2 + \left\| \left(-\frac{1}{2} \right) \cos(2x) \right\|^2 = \left(\frac{1}{\sqrt{2}} \right)^2 \left\| \frac{1}{\sqrt{2}} \right\|^2 + \left(-\frac{1}{2} \right)^2 \|\cos(2x)\|^2.$$

We notice that, since the set $\left\{\frac{1}{\sqrt{2}}, \cos(2x)\right\}$ **is orthonormal, both of these vectors have length 1. So** $\left\|\frac{1}{\sqrt{2}}\right\|^2 = \|\cos(2x)\| = 1$**. So**

$$\|\sin^2(x)\|^2 = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(-\frac{1}{2}\right)^2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Therefore,

$$\int_{-\pi}^{\pi} \sin^4(x)\,dx = \pi\|\sin^2(x)\|^2 = \frac{3\pi}{4}.$$

Example 5.5.5. Continuing from Example 5.5.4: find $\frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(x)\cos(2x)\,dx$.

We begin by rewriting this integral as an inner product.

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(x)\cos(2x)\,dx = \langle \sin^2(x), \cos(2x)\rangle.$$

We cannot use Pythagorean Law here because, in its current form, this expression cannot be written as a length. Instead, we use the decomposition of $\sin^2(x)$ **once more:**

$$\begin{aligned}\langle \sin^2(x), \cos(2x)\rangle &= \left\langle \left(\frac{1}{\sqrt{2}}\right)\frac{1}{\sqrt{2}} + \left(-\frac{1}{2}\right)\cos(2x), \cos(2x)\right\rangle \\ &= \frac{1}{\sqrt{2}}\left\langle \frac{1}{\sqrt{2}}, \cos(2x)\right\rangle + \left(-\frac{1}{2}\right)\langle \cos(2x), \cos(2x)\rangle.\end{aligned}$$

Since the set $\left\{\frac{1}{\sqrt{2}}, \cos(2x)\right\}$ **is orthonormal, we can evaluate these two inner products! The first one is 0, since the two vectors are orthonormal. The second one is 1, since this is the length of this vector squared which is 1! So**

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(x)\cos(2x)\,dx = \langle \sin^2(x), \cos(2x)\rangle = \frac{1}{\sqrt{2}}\cdot 0 - \frac{1}{2}\cdot 1 = -\frac{1}{2}.$$

We already have described an $n \times n$ *invertible* matrix as a matrix where its column vectors form a basis for $\mathbb{R}^n$. We have another definition in the case that a matrix's column vectors form an *orthonormal* basis for $\mathbb{R}^n$:

Definition 5.5.2. An $n \times n$ matrix Q is said to be an **orthogonal matrix** if the column vectors of Q form an orthonormal set in $\mathbb{R}^n$.

The following properties are true for all orthogonal matrices:

Orthogonal Matrices

If Q is an $n \times n$ orthogonal matrix, then

(a) the column vectors of Q form an orthonormal basis for $\mathbb{R}^n$

(b) $Q^TQ = I$

(c) $Q^T = Q^{-1}$

(d) $\langle Q\mathbf{x}, Q\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$

(e) $\|Q\mathbf{x}\| = \|\mathbf{x}\|$

The property "$Q^TQ = I$" holds even when the columns of the matrix form an orthonormal *set* rather than an orthonormal basis for our space. That is, whenever a matrix A has more rows than columns - like in the case of a least-squares problem - it is still true that $A^TA = I$. Recall that a least-squares solution to the equation $A\mathbf{x} = \mathbf{b}$ is equal to $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$. We can simplify using this property to get:

Theorem 5.6. *If the column vectors of an $m \times n$ matrix A form an orthonormal set of vectors in $\mathbb{R}^m$, then the solution to the least squares problem $A\boldsymbol{x} = \boldsymbol{b}$ is*

$$\hat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b} = A^T\boldsymbol{b}.$$

We can combine this theorem with the items from the first theorem in this section to find least squares approximations using the process outlined in the next example.

Example 5.5.6. Find the best least squares approximation to e^x on the interval $[0, 1]$ by a linear function.

We end this section by discussing **trigonometric polynomials**, which are polynomials of the form

$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos(kx) + b_k \sin(kx)).$$

As an extension of Example 5.5.2, we can see that the collection of functions

$$\left\{\frac{1}{\sqrt{2}}, \cos(x), \cos(2x), \ldots, \cos(nx)\right\}$$

is an orthonormal set in $C[-\pi, \pi]$ with respect to the inner product $\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx$. In fact, using more sum-to-product formulas from trigonometry, we can determine that the functions $\{\sin(x), \sin(2x), \ldots, \sin(nx)\}$ can *also* be added to this orthonormal set to give us

$$\left\{\frac{1}{\sqrt{2}}, \cos(x), \cos(2x), \ldots, \cos(nx), \sin(x), \sin(2x), \ldots, \sin(nx)\right\}$$

is an orthonormal set in $C[-\pi, \pi]$.

Even if a function f is *not* in the span of this set, we can use this set to find a least squares approximation. Just as in our previous example, the constants a_k, b_k for $k \in \{1, 2, \ldots\}$ above can be found in the following way (we will find a_0 below):

$$a_k = \langle f, \cos(kx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,dx$$
$$b_k = \langle f, \sin(kx)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\,dx$$

Note that $\langle f, \frac{1}{\sqrt{2}}\rangle\frac{1}{\sqrt{2}} = \langle f, 1\rangle\frac{1}{2}$ by linearity of the inner product, so we define

$$a_0 = \langle f, 1\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx$$

since we are already dividing our function by 2 in the trigonometric polynomial formula. The constants a_k and b_k are called the **Fourier coefficients** of f.

Example 5.5.7. Find the best least squares approximation to $f(x) = |x|$ on $[-\pi, \pi]$ by a trigonometric polynomial of degree less than or equal to 2.

A trigonometric polynomial of degree (at most) 2 is of the form

$$\frac{a_0}{2} + a_1 \cos(x) + b_1 \sin(x) + a_2 \cos(2x) + b_2 \sin(2x).$$

We use the formulas from the previous page to calculate:

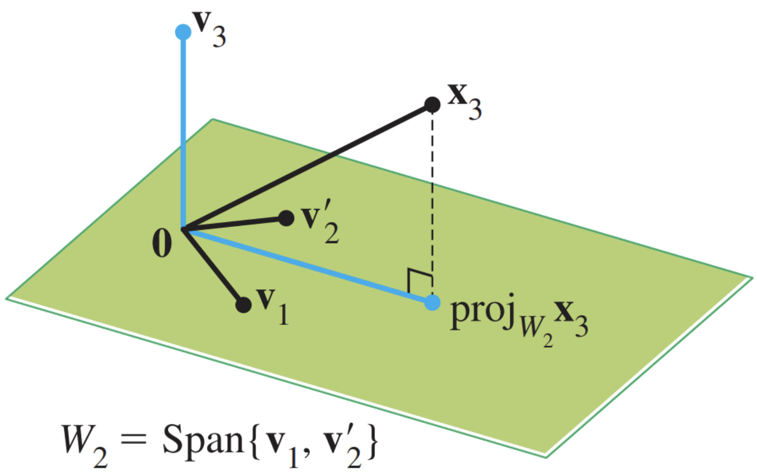
$$a_0 = \langle f(x), 1 \rangle = \langle |x|, 1 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \, dx = \frac{2}{\pi} \int_0^{\pi} x \, dx = \pi$$
$$a_1 = \langle f(x), \cos(x) \rangle = \langle |x|, \cos(x) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \cos(x) \, dx = \frac{2}{\pi} \int_0^{\pi} x \cos(x) = -\frac{4}{\pi}$$
$$b_1 = \langle f(x), \sin(x) \rangle = \langle |x|, \sin(x) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \sin(x) = 0$$
$$a_2 = \langle f(x), \cos(2x) \rangle = \langle |x|, \cos(2x) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \cos(2x) = \frac{2}{\pi} \int_0^{\pi} x \cos(2x) = -\frac{1}{\pi}$$
$$b_2 = \langle f(x), \sin(2x) \rangle = \langle |x|, \sin(2x) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} |x| \sin(2x) = 0$$

So plugging these in we get our trigonometric polynomial to be

$$\frac{\pi}{2} - \frac{4}{\pi} \cos(x) + 0 \sin(x) - \frac{1}{\pi} \cos(2x) + 0 \sin(2x).$$

Section 5.6

Gram-Schmidt Orthogonalization



Objectives:

- Construct an orthogonal basis from any basis using vector projections
- Formulate QR factorizations to solve least squares problems

Example 5.6.1. Let $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

Example 5.6.2. Let $\mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$. If $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \subset \mathbb{R}^4$, find an orthogonal basis of W.

This process of orthogonalizing is called **Gram-Schmidt orthogonalization**. We summarize this process below:

Gram-Schmidt Orthogonalization

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ be a basis for a nonzero subspace $W \subset \mathbb{R}^n$. Then define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1\\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1\\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2\\
&\vdots\\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}.
\end{aligned}$$

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

Example 5.6.3. Let

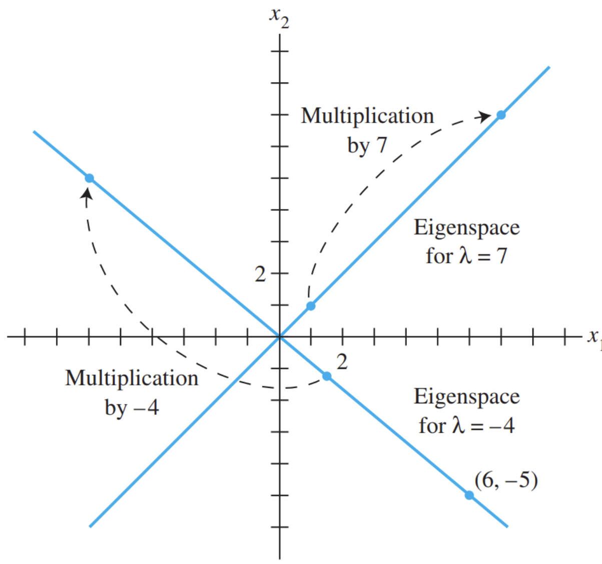
$$A = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{bmatrix}.$$

Find an orthonormal basis for the column space of A.

Example 5.6.4. Starting with the basis $\{1, x\}$ for P_2, use Gram-Schmidt orthogonalization to find an orthonormal basis for P_2.

Section 6.1

Eigenvalues and Eigenvectors



Objectives:

- Compute and recognize the eigenvalues and eigenvectors of a matrix
- Describe the geometry of a matrix transformation in $\mathbb{R}^2$

We return to our geometric visualization of matrices. In this chapter we will learn how to describe what $A\mathbf{v}$ is without calculating it exactly. Let's begin with a little exploration.

Example 6.1.1. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Are there any vectors $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{v}$? Are there any vectors $\mathbf{v}$ such that $A\mathbf{v} = -\mathbf{v}$?

Any vector that does not rotate but only scales when mapped through a linear operator - such as the vectors we found in the previous example - is called an **eigenvector**, and the factor it scales by is called its **eigenvalue**. More formally, we have the following definition:

Definition 6.1.1. Let A be an $n \times n$ matrix. A scalar λ is said to be an **eigenvalue** of A if there exists a nonzero vector $\mathbf{x}$ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The vector $\mathbf{x}$ is said to be an **eigenvector** belonging to λ.

In the previous example, the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ has an eigenvector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ with eigenvalue 1 and an eigenvector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ with eigenvalue -1.

Example 6.1.2. Show that $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$. Compute its eigenvalue.

There is a *geometric* way to interpret eigenvalues and eigenvectors. In the above, we can say that the matrix A stretches any vector near $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ by a factor of 3. Once we have enough eigenvalues and eigenvectors we can describe what A does to *any* vector.

The key to finding eigenvalues and eigenvectors lies in solving the equation $A\mathbf{x} = \lambda\mathbf{x}$. If we subtract $\lambda\mathbf{x}$ from both sides, we get $A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$. We could factor out $\mathbf{x}$ if both A and λ were matrices. So we multiply $\mathbf{x}$ by the identity matrix I on the left and get that $A\mathbf{x} - \lambda I\mathbf{x} = \mathbf{0}$, which implies

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

As a result we get the following statements to be equivalent:

(a) λ is an eigenvalue of A.

(b) $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

(c) $N(A - \lambda I) \neq \{\mathbf{0}\}$.

(d) $A - \lambda I$ is singular.

(e) $\det(A - \lambda I) = 0$.

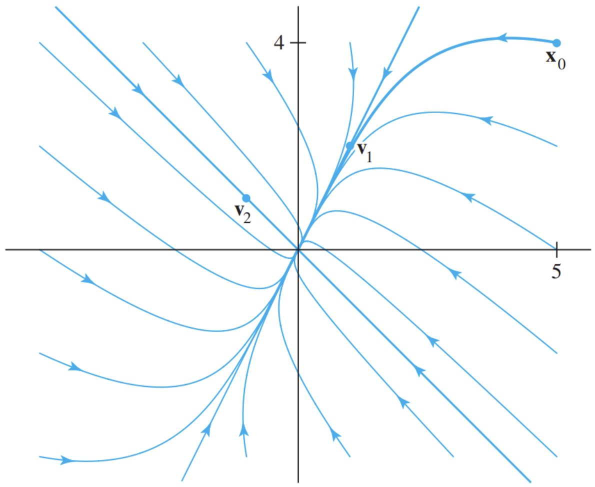
This last condition in particular is useful - it is called the **characteristic polynomial**. We will see the reason for this naming in the example below:

Example 6.1.3. Find the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$.

Sometimes an eigenvalue can belong to a set of more than one linearly independent vectors. We say that an **eigenspace** corresponding to an eigenvalue λ is the space of all eigenvectors of λ. We say that λ has **geometric multiplicity** equal to the dimension of its eigenspace.

Example 6.1.4. Find the eigenvalues and corresponding eigenvectors of $A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$.

Section 6.2
Linear Differential Equations



Objectives:

- Solve linear differential equations using matrices

Recall the differentiation operator $D : C^1[a,b] \to C[a,b]$ defined by $D(f) = f'$. This operator can be extended to apply to multiple functions at the same time. For example, for functions $y_1, y_2 \in C^1[a,b]$, the system of differential equations

$$y_1' = f(x), \qquad y_2' = g(x)$$

can be written as $D(\mathbf{Y}) = \begin{bmatrix} f(x) \\ g(x) \end{bmatrix}$, where $\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. We will simplify this notation by writing $\mathbf{Y}' = D(\mathbf{Y})$.

We define the equations below to be a system of **first-order linear differential equations**:

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &\vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n. \end{aligned}$$

Simplifying this using our notation above, we get that this system is

$$\mathbf{Y}' = A\mathbf{Y},$$

where A is the matrix of coefficients on the right-hand side of this system.

Example 6.2.1. Solve the first-order linear differential equation $y' = \lambda y$.

Example 6.2.2. Solve the first-order system of linear differential equations $\mathbf{Y}' = \lambda \mathbf{Y}$.

Example 6.2.3. Now let λ be an eigenvalue of A and let $\mathbf{x}$ be an eigenvector corresponding to λ. Show our solution from Example 6.2.2 solves $\mathbf{Y}' = A\mathbf{Y}$.

Example 6.2.4. Show that, if $\mathbf{Y}_1$ and $\mathbf{Y}_2$ solve the system $\mathbf{Y}' = A\mathbf{Y}$, then so does $\alpha\mathbf{Y}_1 + \beta\mathbf{Y}_2$.

Let's apply this theory to an example.

Example 6.2.5. Solve the system

$$\begin{aligned} y_1' &= 3y_1 + 4y_2 \\ y_2' &= 3y_1 + 2y_2. \end{aligned}$$

Using this information, solve the **initial value problem** when given that, when $t = 0$, $y_1 = 6$ and $y_2 = 1$:

$$\begin{aligned} y_1' &= 3y_1 + 4y_2 \\ y_2' &= 3y_1 + 2y_2. \end{aligned} \qquad \mathbf{Y}(0) = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

Example 6.2.6. Solve the system below:

$$y_1' = 4y_1 + 3y_2$$
$$y_2' = y_2$$

Further suppose that, when $t = 0$, $y_1 = 1$ and $y_2 = 1$. Find the solution to the initial value problem.

Section 6.3
Diagonalization

$$A = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1^{\,T} \\ \vdots \\ x_n^{\,T} \end{bmatrix}$$

Objectives:

- Discover matrices which are similar to diagonal matrices
- Use the matrix exponential to solve linear differential equations

A final use for eigenvalues and eigenvectors has us returning to the definition of *similar matrices* from before. Recall that we said that row equivalent matrices A and B are similar, in the sense that there exists some invertible matrix P such that

$$A = PBP^{-1}.$$

We say that an $n \times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix. Let's see how this could be useful:

Example 6.3.1. Say that a matrix A is similar to the matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Calculate A^{2024}.

This is particularly useful in applications of Markov chains and solving linear differential equations - we will explore this latter application soon. We can use eigenvalues and eigenvectors to discover a large class of matrices that are diagonalizable in this way.

We begin with a theorem that helps connect these two worlds:

Theorem 6.1. *If $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of an $n \times n$ matrix A with corresponding eigenvectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$, then $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ are linearly independent.*

Proof. Consider these eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ and let $r \leq k$ be the largest number such that $\mathbf{x}_1, \dots, \mathbf{x}_r$ are linearly independent. If $r = k$, we are done, so let's assume $r < k$. Then since $r + 1$ is larger than r, the first $r + 1$ vectors must be linearly dependent. So

$$c_1\mathbf{x}_1 + \cdots + c_r\mathbf{x}_r + c_{r+1}\mathbf{x}_{r+1} = \mathbf{0}$$

has a nontrivial solution. Applying A to both sides of these equations, we get

$$c_1A\mathbf{x}_1 + \cdots + c_rA\mathbf{x}_r + c_{r+1}A\mathbf{x}_{r+1} = A\mathbf{0} = \mathbf{0}.$$

These $A\mathbf{x}_i$ can be simplified using the eigenvalue equation:

$$c_1\lambda_1\mathbf{x}_1 + \cdots + c_r\lambda_r\mathbf{x}_r + c_{r+1}\lambda_{r+1}\mathbf{x}_{r+1} = \mathbf{0}.$$

We return briefly to the first equation in our proof and multiply the entire thing by λ_{r+1}:

$$c_1\lambda_{r+1}\mathbf{x}_1 + \cdots + c_r\lambda_{r+1}\mathbf{x}_r + c_{r+1}\lambda_{r+1}\mathbf{x}_{r+1} = \mathbf{0}.$$

Here is the trick: notice what would happen if we subtracted these last two equations. The last terms would cancel out, meaning that we would have no $\mathbf{x}_{r+1}$ term. But the other terms would not, because (for example) $\lambda_1 - \lambda_{r+1} \neq 0$ - those two eigenvalues are distinct! This means that we have shown that $\mathbf{x}_1, \dots, \mathbf{x}_r$ are *linearly dependent* after all, which contradicts our choice of r. So there's no way for $r < k$, and once again we are done! □

Theorem 6.2. *If an $n \times n$ matrix A has n linearly independent eigenvectors, then A is diagonalizable.*

Proof. For $i \in [n]$, let $\mathbf{x}_i$ be an eigenvector corresponding to λ_i. Let $X := \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n\end{bmatrix}$. Then

$$AX = \begin{bmatrix}A\mathbf{x}_1 & \cdots & A\mathbf{x}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{x}_1 & \cdots & \lambda_n\mathbf{x}_n\end{bmatrix} = \begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}.$$

This latter matrix is a diagonal matrix with eigenvalues on the diagonal and zeroes elsewhere. Let's call it D. So $AX = XD$.

Since X has linearly independent columns, it is invertible, so we can multiply both sides on the right by X^{-1}. So $A = XDX^{-1}$ and A is diagonalizable by definition. □

Note that combining these two theorems yields: if A has n linearly independent eigenvectors OR n distinct eigenvalues, then A is diagonalizable. Let's see an example.

Example 6.3.2. Let $A = \begin{bmatrix} 2 & -3 \\ 2 & -5 \end{bmatrix}$. Find matrices X and D such that $A = XDX^{-1}$.

Example 6.3.3. Let $A = \begin{bmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{bmatrix}$. Diagonalize A if possible, or explain why diagonalizing is not possible.

We close with our application of diagonalization to linear differential equations. Recall from Example 6.2.1 that we found the solution of the single-variable first-order linear differential equation to be $y(t) = ce^{ta}$. If we were given an initial value of $y(0) = y_0$, we would plug in y_0 for c to get the solution of the initial value problem to be

$$y(t) = y_0 e^{ta}.$$

In multiple variables, this solution also works. We replace the variables y and y_0 with their vector forms, and we replace the value a with the matrix A:

$$\mathbf{Y}(t) = \mathbf{Y}_0 e^{tA}.$$

We have not seen the value e^{tA} defined before, but we can define it using power series. Recall that the Maclaurin series for the function e^x is defined as

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

We define the **matrix exponential** in the same way, replacing x to plug in A:

$$e^A := 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

In general this can be a cumbersome formula - we now have to solve for an infinite series! But plugging a *diagonal* matrix in for this matrix exponential yields a neat formula. Let's define D to be the following $n \times n$ matrix:

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

The blank space here and elsewhere denotes that the entries off of the diagonal are zeroes. Then we have

$$\begin{aligned} e^D &= \lim_{m\to\infty} \left(I + D + \frac{1}{2!}D^2 + \cdots + \frac{1}{m!}D^m \right) \\ &= \lim_{m\to\infty} \begin{bmatrix} \sum_{k=0}^m \frac{1}{k!}\lambda_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^m \frac{1}{k!}\lambda_n^k \end{bmatrix} = \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n} \end{bmatrix}. \end{aligned}$$

Hence

$$e^{tXDX^{-1}} = Xt\left(1 + D + \frac{D^2}{2} + \cdots \right) X^{-1} = X \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n} \end{bmatrix} X^{-1}.$$

Let's see this used in an example.

Example 6.3.4. Use the matrix exponential to solve the initial value problem

$$\mathbf{Y}' = A\mathbf{Y}, \qquad \mathbf{Y}(0) = \mathbf{Y}_0$$

where

$$A = \begin{bmatrix} 3 & 4 \\ 3 & 2 \end{bmatrix}, \qquad \mathbf{Y}_0 = \begin{bmatrix} 6 \\ 1 \end{bmatrix}.$$

Example 6.3.5. Solve the initial value problem $\mathbf{Y}' = A\mathbf{Y}$, where $A = \begin{bmatrix} 1 & -2 \\ 0 & -1 \end{bmatrix}$, where $\mathbf{Y}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.