Chapter 1: Vectors

(Last Updated: September 28, 2021)

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed).

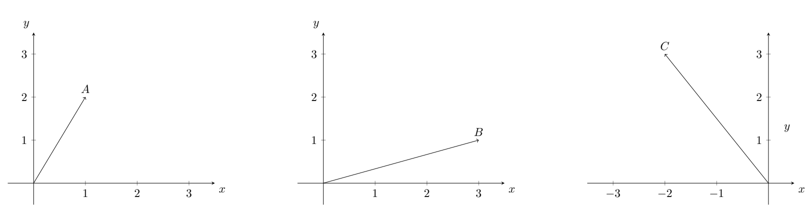
1. The geometry of algebra and vectors

Definition: Vector, initial point, terminal point

A *vector* is a directed line segment that corresponds to the displacement from a point A to another point B in the Cartesian plane (xy-plane, $\mathbb{R}^2$). We denote this vector by $\overrightarrow{AB}$. Here, A is the *initial point* (tail) and B is the *terminal point* (head).

But this is just one *representation* of the vector. Any two points giving the same direction and magnitude should produce the same vector. So, we will choose to represent our vectors with initial point at the origin O (such vectors are said to be in *standard position*). In this way, we need only indicate the terminal point in order to define the vector. If A is the point $(1, 2)$, then the corresponding vector is $\mathbf{a} = \overrightarrow{OA} = [1, 2]$. (Note we often use lowercase bold letters to denote vectors, though it is often common to use $\overrightarrow{a}$.) The square brackets indicate we mean the vector from the origin to the point $(1, 2)$, and not just the point itself. The entries of the vector $[1, 2]$ are called the *components* of $\mathbf{a}$. The representation $[1, 2]$ is called a *row vector*. We will more often use *column vectors*, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, to denote our vectors.

Example. The vectors $\mathbf{a} = \overrightarrow{OA} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{b} = \overrightarrow{OB} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{c} = \overrightarrow{OC} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ are visualized below.



The *zero vector* (in $\mathbb{R}^2$) is $\mathbf{0} = \overrightarrow{OO} = [0, 0]$. (When we move into $\mathbb{R}^n$, there will be a zero vector for each dimension.) Two vectors are *equal* (in $\mathbb{R}^2$) if they have the same length and the same direction. You can think of this as translating a vector so that its tail is on the origin. The next example shows how to do this algebraically.

Example. Consider the points $A = (1,1)$, $B = (5,2)$, $C = (-1,6)$, and $D = (3,7)$. Then

$$\overrightarrow{AB} = [5-1, 2-1] = [4,1] \quad \text{and} \quad \overrightarrow{CD} = [3-(-1), 7-6] = [4,1].$$

If $E = (4,1)$, then $\overrightarrow{AB} = \overrightarrow{OE} = \overrightarrow{CD}$.

There are two algebraic operations on vectors: vector addition and scalar multiplication.

Vector Addition

If $\mathbf{u} = [u_1, u_2]$ and $\mathbf{v} = [v_1, v_2]$ are vectors in $\mathbb{R}^2$, then the *total displacement* $\mathbf{u}+\mathbf{v}$ is the *sum*

$$\mathbf{u}+\mathbf{v} = [u_1+v_1, u_2+v_2].$$

There are two ways to view this sum geometrically.

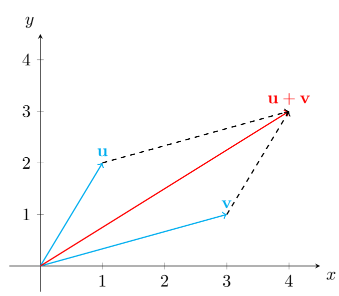
The Head-to-Tail Rule: Translate $\mathbf{v}$ so that its tail coincides with the head of $\mathbf{u}$ The sum $\mathbf{u}+\mathbf{v}$ is the vector from the tail of $\mathbf{u}$ to the head of $\mathbf{v}$.

The Parallelogram Rule: The sum $\mathbf{u}+\mathbf{v}$ is the vector in standard position along the diagonal of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$.

Example. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Algebraically, we find

$$\mathbf{u}+\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1+3 \\ 2+1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Geometrically, we find that $\mathbf{u}+\mathbf{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ as below:



Scalar Multiplication

Given a vector $\mathbf{v}$ in $\mathbb{R}^2$ and a real number c, the *scalar multiple* $c\mathbf{v}$ is the vector obtained by multiplying each component by c.

If $c > 0$, then $c\mathbf{v}$ points in the same direction as $\mathbf{v}$. If $c < 0$, then $c\mathbf{v}$ points in the opposite direction as $\mathbf{v}$. If $|c| > 1$, then $c\mathbf{v}$ is longer than $\mathbf{v}$. If $0 < |c| < 1$, then $c\mathbf{v}$ is shorter than $\mathbf{v}$. Of course, $1\mathbf{v} = \mathbf{v}$ and $0\mathbf{v} = \mathbf{0}$.

Example. Consider $\mathbf{v} = [8, -2]$. Then $2\mathbf{v} = [16, -4]$, $-2\mathbf{v} = [-16, 4]$, and $\frac{1}{2}\mathbf{v} = [4, -1]$.

Two vectors that are scalar multiples of one another are said to be *parallel*. To define *vector subtraction*, we first define the *negative* of a vector $\mathbf{v}$ as $-\mathbf{v} = (-1)\mathbf{v}$. Then the *difference* of vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ is defined as

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}).$$

Geometrically, $\mathbf{u} - \mathbf{v}$ corresponds to the other diagonal in the parallelogram created by $\mathbf{u}$ and $\mathbf{v}$.

These concepts all extend easily to $\mathbb{R}^3$. Instead of redoing everything above, we will skip to $\mathbb{R}^n$.

We define $\mathbb{R}^n$ to be the set of all ordered n-tuples of real numbers, written either as row or column vectors. So, a vector $\mathbf{v}$ in $\mathbb{R}^n$ can be written as

$$[v_1, v_2, \ldots, v_n] \quad \text{or} \quad \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

The entry v_i of $\mathbf{v}$ is called the *ith component* of $\mathbf{v}$.

We extend the standard operations on vectors in $\mathbb{R}^2$ to $\mathbb{R}^n$. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and let c be a real number:

Vector Addition and Scalar Multiplication

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \qquad c\mathbf{v} = c\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}.$$

Example. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}$. Compute $2\mathbf{u} - \mathbf{v}$.

$$2\mathbf{u} - \mathbf{v} = \begin{bmatrix} 4 \\ -2 \\ 10 \end{bmatrix} - \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 - 4 \\ -2 - 3 \\ 10 - (-2) \end{bmatrix} = \begin{bmatrix} 0 \\ -5 \\ 12 \end{bmatrix}$$

Theorem 1: Algebraic Properties of $\mathbb{R}^n$

Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c and d be scalars. Then

a. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$

b. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$

c. $\mathbf{u}+\mathbf{0}=\mathbf{u}$

d. $\mathbf{u}+(-\mathbf{u})=\mathbf{u}+(-1)\mathbf{u}=\mathbf{0}$

e. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$

f. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$

g. $c(d\mathbf{u})=(cd)\mathbf{u}$

h. $1\mathbf{u}=\mathbf{u}$.

Property (a) is commonly called *commutativity* while property (b) is called *associativity*. Properties (e) and (f) are called *distributivity*.

Often in this course, we will be interested in the *why* as much as the *what*. So, we will *prove* statements rather than just accepting them as fact. The properties above give us a good chance to practice with this.

Proof. (a) Write $\mathbf{u}=[u_1,u_2,\ldots,u_n]$ and $\mathbf{v}=[v_1,v_2,\ldots,v_n]$. By properties of real numbers, $u_i+v_i=v_i+u_i$ for all i. Hence,

$$\begin{aligned}\mathbf{u}+\mathbf{v}&=[u_1,u_2,\ldots,u_n]+[v_1,v_2,\ldots,v_n]\\&=[u_1+v_1,u_2+v_2,\ldots,u_n+v_n]\quad\text{(vector addition)}\\&=[v_1+u_1,v_2+u_2,\ldots,v_n+u_n]\quad\text{(commutivity of real numbers)}\\&=[v_1,v_2,\ldots,v_n]+[u_1,u_2,\ldots,u_n]\quad\text{(vector addition)}\\&=\mathbf{v}+\mathbf{u}.\end{aligned}$$

(b) Write $\mathbf{u}=[u_1,u_2,\ldots,u_n]$, $\mathbf{v}=[v_1,v_2,\ldots,v_n]$, and $\mathbf{w}=[w_1,w_2,\ldots,w_n]$. In this case we will use associativity of real numbers, so $(u_i+v_i)+w_i=u_i+(v_i+w_i)$ for all i. Hence,

$$\begin{aligned}(\mathbf{u}+\mathbf{v})+\mathbf{w}&=([u_1,u_2,\ldots,u_n]+[v_1,v_2,\ldots,v_n])+[w_1,w_2,\ldots,w_n]\\&=[u_1+v_1,u_2+v_2,\ldots,u_n+v_n]+[w_1,w_2,\ldots,w_n]\quad\text{(vector addition)}\\&=[(u_1+v_1)+w_1,(u_2+v_2)+w_2,\ldots,(u_n+v_n)+w_n]\quad\text{(vector addition)}\\&=[u_1+(v_1+w_1),u_2+(v_2+w_2),\ldots,u_n+(v_n+w_n)]\quad\text{(assoc of real numbers)}\\&=[u_1,u_2,\ldots,u_n]+[v_1+w_1,v_2+w_2,\ldots,v_n+w_n]\quad\text{(vector addition)}\\&=[u_1,u_2,\ldots,u_n]+([v_1,v_2,\ldots,v_n]+[w_1,w_2,\ldots,w_n])\quad\text{(vector addition)}\\&=\mathbf{u}+(\mathbf{v}+\mathbf{w}).\end{aligned}$$ $\square$

By property (b), we may now write $\mathbf{u}+\mathbf{v}+\mathbf{w}$ without confusion (since we can add the vectors in any way).

Example. Let $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ be vectors in $\mathbb{R}^n$. Simplify the expression $-3(\mathbf{a}-\mathbf{c})+2(\mathbf{a}+\mathbf{b})+3(\mathbf{c}-\mathbf{b})$ and indicate which property is used for each step.

We now come to one of the most fundamental definitions in this course.

Definition: Linear combination, coefficients

A vector $\mathbf{v}$ in $\mathbb{R}^n$ is a *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ (in $\mathbb{R}^n$) if there are scalars $c_1, c_2, \ldots, c_k$ such that $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$. The scalars c_i are called the *coefficients* of the linear combination.

Alternatively, we could say that the c_i are the *coordinates* of $\mathbf{v}$ with respect to the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

Example. Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$. Then $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ since $\mathbf{w} = (-1)\mathbf{u} + 2\mathbf{v}$. Finding methods for determining these coefficients will be a major goal of this course.

2. Length and Angle: The Dot Product

In this section we will study a certain product on vectors. This is but one product, and in fact it is a special case of an *inner product*, which we will study later in the course. Those who have taken Calc III will find this very familiar.

Definition: Dot Product

If $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, then the *dot product* $\mathbf{u} \cdot \mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Note that the dot product inputs two vectors but outputs a scalar.

Example. Let $\mathbf{a} = \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Then

$$\mathbf{a} \cdot \mathbf{b} = (-1)(2) + (3)(1) + (5)(-2) = -9.$$

You should check that $\mathbf{b} \cdot \mathbf{a} = -9$ as well.

Theorem 2: Algebraic properties of the dot product

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then

a. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
b. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \cdot \mathbf{v}) + (\mathbf{u} \cdot \mathbf{w})$
c. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$
d. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

These properties are easy to prove, and even easier if we are comfortable using summation notation.

Proof. (a) $\mathbf{u} \cdot \mathbf{v} = \displaystyle\sum_{i=1}^{n} (u_i \cdot v_i) = \sum_{i=1}^{n} (v_i \cdot u_i) = \mathbf{v} \cdot \mathbf{u}$.

(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \displaystyle\sum_{i=1}^{n} u_i(v_i + w_i) = \sum_{i=1}^{n} (u_i v_i) + (u_i w_i) = (\mathbf{u} \cdot \mathbf{v}) + (\mathbf{u} \cdot \mathbf{w})$.

(c) $(c\mathbf{u}) \cdot \mathbf{v} = \displaystyle\sum_{i=1}^{n} ((cu_i)v_i) = c \sum_{i=1}^{n} (u_i v_i) = c(\mathbf{u} \cdot \mathbf{v})$. □

The following definition is essentially a restatement of the Pythagorean Theorem.

Definition: Length/Norm

The *length* (or *norm*) of $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$ is the non-negative scalar $\|\mathbf{v}\|$ defined by

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

Example. Let $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Then

$$\|\mathbf{b}\| = \sqrt{\mathbf{b} \cdot \mathbf{b}} = \sqrt{(2)^2 + (1)^2 + (-2)^2} = \sqrt{9} = 3.$$

The next theorem is really just a specialization of properties of the dot product.

Theorem 3: Algebraic properties of length

Let $\mathbf{v}$ be a vector in $\mathbb{R}^n$ and let c be a scalar. Then

a. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{b} = \mathbf{0}$

b. $\|c\mathbf{v}\| = |c| \, \|\mathbf{v}\|$.

A *unit vector* is a vector of length 1. We can associate the set of unit vectors in $\mathbb{R}^2$ with the unit circle. If $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unit vector in the same direction as $\mathbf{v}$. This process is called *normalizing* a vector.

Example. Let $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$ as in the previous example, so $\|\mathbf{b}\| = 3$. The associated unit vector is

$$\frac{1}{\|\mathbf{b}\|}\mathbf{b} = \frac{1}{3}\mathbf{b} = \begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix}.$$

The next two theorems show how taking norms behave with respect to the dot product and vector addition, respectively. We won't prove the first, though interested students are encouraged to think about this. The second theorem follows easily from the first.

Theorem 4: The Cauchy-Schwarz Inequality

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|.$$

Theorem 5: The Triangle Inequality

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Proof. We have

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\| &= (\mathbf{u}+\mathbf{v}) \cdot (\mathbf{u}+\mathbf{v}) \\
&= \mathbf{u}\cdot\mathbf{u} + 2(\mathbf{u}\cdot\mathbf{v}) + \mathbf{v}\cdot\mathbf{v} \quad \text{by properties of dot product} \\
&\leq \|\mathbf{u}\|^2 + 2|\mathbf{u}\cdot\mathbf{v}| + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2\,\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\mathbf{u}+\mathbf{v})^2.
\end{aligned}$$

Since both sides of the original inequality are positive, taking square roots gives the result. □

Recall that the distance between two points (a_1, a_2) and (b_1, b_2) in $\mathbb{R}^2$ is determined by the well-known distance formula $\sqrt{(b_1-a_1)^2+(b_2-a_2)^2}$. We can similarly define distance between vectors.

Definition: Distance

The *distance* $d(\mathbf{u},\mathbf{v})$ between $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is defined by

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|.$$

Another way of sayin the above is that $d(\mathbf{u},\mathbf{v})$ is is the length of the vector $\mathbf{u}-\mathbf{v}$.

Example. Let $\mathbf{a} = \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. The distance between $\mathbf{a}$ and $\mathbf{b}$ is

$$\|\mathbf{a}-\mathbf{b}\| = \left\| \begin{bmatrix} -3 \\ 2 \\ 7 \end{bmatrix} \right\| = \sqrt{(-3)^2+(2)^2+(7^2)} = \sqrt{62}.$$

Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in $\mathbb{R}^2$. By the Law of Cosines,

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta.$$

Rearranging gives

$$
\begin{aligned}
\|\mathbf{u}\| \|\mathbf{v}\| \cos\theta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right] \\
&= \frac{1}{2}\left[(u_1^2+u_2^2) + (v_1^2+v_2^2) - \left((u_1-v_1)^2 + (u_2-v_2)^2\right)\right] \\
&= \frac{1}{2}\left[(u_1^2+u_2^2) + (v_1^2+v_2^2) - (u_1^2 - 2u_1v_1 + v_1^2) - (u_2^2 - 2u_2v_2 + v_2^2)\right] \\
&= u_1v_1 + u_2v_2 \\
&= \mathbf{u}\cdot\mathbf{v}.
\end{aligned}
$$

Hence,

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

This leads to the following definition.

Definition: Angle between vectors

For nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

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

Recall that $\cos 90 = 0$, and since $\mathbf{u}, \mathbf{v}$ are assumed to be nonzero, then this implies (in $\mathbb{R}^2$), that they are perpendicular. The next definition is a generalization of this notion to higher dimensions.

Definition: Orthogonal

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are *orthogonal* to each other if $\mathbf{u}\cdot\mathbf{v} = 0$.

Example. Let $\mathbf{u} = [1, -2, 4]$ and $\mathbf{v} = [-2, 1, 1]$. Then

$$\mathbf{u}\cdot\mathbf{v} = -2 - 2 + 4 = 0,$$

so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Theorem 6: The Pythagorean Theorem

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

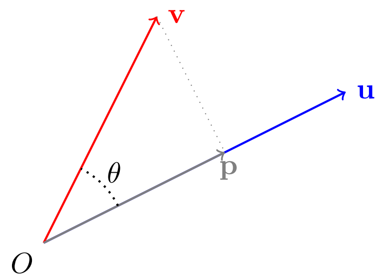
Proof. Recall that

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2.$$

Hence, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $2(\mathbf{u}\cdot\mathbf{v}) = 0$ (equivalently, $\mathbf{u}\cdot\mathbf{v} = 0$). $\square$

One application of the dot product, and the theory we've developed around it, is to the problem of finding the distance from a point to a line (Calc III students will have seen this already).

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors. Let $\mathbf{p}$ be the vector obtained by dropping a perpendcular from the head of $\mathbf{v}$ onto $\mathbf{u}$. Then $\mathbf{p}$ is parallel to $\mathbf{u}$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$.



The unit vector associated to $\mathbf{u}$ is $\hat{\mathbf{u}} = (1/\|\mathbf{u}\|)\mathbf{u}$, so $\mathbf{p} = \|\mathbf{p}\|\,\hat{\mathbf{u}}$. By trigonometry, $\cos\theta = \frac{\|\mathbf{p}\|}{\|\mathbf{v}\|}$ and by definition, $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$. Putting this all together we have

$$\mathbf{p} = \|\mathbf{v}\|\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right)\left(\frac{1}{\|\mathbf{u}\|}\right)\mathbf{u} = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|^2}\right)\mathbf{u} = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}.$$

Definition: Projection

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ and $\mathbf{u} \neq 0$, then the *projection of* $\mathbf{v}$ *onto* $\mathbf{u}$ is the vector $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ defined by

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}.$$

Example. Let $\mathbf{u} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$. Then

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{1}{2}\mathbf{u} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Consider the line ℓ in $\mathbb{R}^2$ with equation $2x + y = 0$. If we let $\mathbf{n} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, then we could alternatively express this line by the vector equation $\mathbf{n} \cdot \mathbf{x} = 0$. We call $\mathbf{n}$ the *normal vector* of the line ℓ. Note that $\mathbf{n}$ is orthogonal to any point on the line ℓ.

Alternatively, we might think of a particle moving (left-to-right) on the line ℓ at time t (in seconds). At time $t = 0$ it is at the origin. At time $t = 1$ it is at the point $(1, -2)$. That is, the x-coordinate changes 1 unit per second. Then we can express the equation of the line in *parametric form*:

$$\begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

We call the vector $\mathbf{d} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ the *direction vector* for the line. Then we can write the equation of the line as $\mathbf{x} = t\mathbf{d}$.

Now suppose we shift the line. Consider the line ℓ with equation $2x + y = 5$. The slope of this line is the same, so the vectors $\mathbf{d}$ and $\mathbf{n}$ are still direction and normal vectors, respectively, for this new line. Note that $\mathbf{n}$ is orthogonal to every vector parallel to ℓ. One point on the line is $P = (0, 5)$. If $X = (x, y)$ represents a general point on ℓ, then the vector $\overrightarrow{PX} = \mathbf{x} - \mathbf{p}$ is parallel to ℓ and so $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$. In other words, $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$. Observe that

$$\mathbf{n} \cdot \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 2x + y \quad \text{and} \quad \mathbf{n} \cdot \mathbf{p} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 5 \end{bmatrix} = 5.$$

The point P we chose on ℓ is not special. We could have chosen *any* point on ℓ for this.

Definition: Normal/general form of equation of a line

The *normal form of the equation of a line* ℓ in $\mathbb{R}^2$ is

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0 \quad \text{or} \quad \mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

where $\mathbf{p}$ corresponds to a specific point on ℓ and $\mathbf{n} \neq \mathbf{0}$ is a normal vector for ℓ.

The *general form of the equation of* ℓ is $ax + by = c$, where $\mathbf{n} = \begin{bmatrix} a \\ b \end{bmatrix}$ is a normal vector for ℓ.

Since every choice of $\mathbf{x}$, $\mathbf{x} - \mathbf{p}$ must be parallel to the direction vector $\mathbf{d}$, we obtain the parametric equation $\mathbf{x} - \mathbf{p} = t\mathbf{d}$ or $\mathbf{x} = \mathbf{p} + t\mathbf{d}$. (This should be reminiscent of the *point-slope form* of a line.) In our example, we have the *vector form* of ℓ:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and the corresponding *parametric form* of ℓ:

$$x = 1 + t$$
$$y = 3 - 2t.$$

These definitions generalize directly to $\mathbb{R}^3$.

Definition: Vector/parametric form of equation of a line

The *vector form of the equation of a line* ℓ in $\mathbb{R}^2$ and $\mathbb{R}^3$ is

$$\mathbf{x} = \mathbf{p} + t\mathbf{d}$$

where $\mathbf{p}$ corresponds to a specific point on ℓ and $\mathbf{d} \neq 0$ is a direction vector for ℓ.
The equations corresponding to the components of the vector form of the equation are called *parametric equations* of ℓ.

Example. Find a vector equation of the line through the points $P = (1, 2, 5)$ and $Q = (3, 1, 0)$.

One possible direction vector is $\mathbf{d} = \overrightarrow{PQ} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix}$. The vector equation is then

$$x = \mathbf{p} + t\mathbf{d} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} + t \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}.$$

Omitting much discussion, we will write the equation of a plane in $\mathbb{R}^3$ basically by analogy.

Definition: Normal/general form of equation of a plane

The *normal form of the equation of a plane* $\mathscr{P}$ in $\mathbb{R}^3$ is

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0 \quad \text{or} \quad \mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

where $\mathbf{p}$ corresponds to a specific point on $\mathscr{P}$ and $\mathbf{n} \neq \mathbf{0}$ is a normal vector for ℓ.

The *general form of the equation of* $\mathscr{P}$ is $ax+by+cz = d$, where $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is a normal vector for $\mathscr{P}$.

Whereas a line is determined by two points, a plane is determined by three non-collinear points.

Definition: Vector/parametric form of equation of a plane

The *vector form of the equation of a plane* $\mathscr{P}$ in $\mathbb{R}^3$ is

$$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$$

where $\mathbf{p}$ corresponds to a specific point on $\mathscr{P}$ and $\mathbf{u}, \mathbf{v}$ are direction vectors for $\mathscr{P}$ which are non-zero and parallel to $\mathscr{P}$, but not parallel to each other.
The equations corresponding to the components of the vector form of the equation are called *parametric equations* of $\mathscr{P}$.

Example. Write the vector equation of the line through the points $P = (1, -1, 3)$, $Q = (2, 5, -2)$, and $R = (0, -1, 4)$.

Two possible direction vectors are

$$\mathbf{u} = \overrightarrow{PQ} = \begin{bmatrix} 1 \\ 6 \\ -5 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \overrightarrow{PR} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

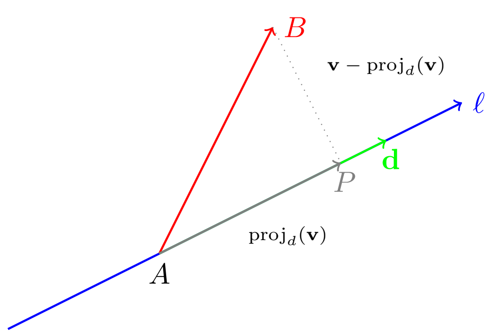
Thus, the vector equation is

$$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} + s \begin{bmatrix} 1 \\ 6 \\ -5 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

The distance from a point to a line is defined as the length of the perpendicular from the point to the line. We can use projections to compute this.

Let ℓ be a line with direction vector $\mathbf{d}$ and B a point not on ℓ (otherwise, the distance is just 0). Let P be the on ℓ at the foot of the perpendicular from B. Let A be any other point on the line and set $\mathbf{v} = \overrightarrow{AB}$. Now the distance is just

$$d(B, \ell) = \mathbf{v} - \text{proj}_{\mathbf{d}}(\mathbf{v}).$$



Example. Let $B = (1, 1, 3)$. Let ℓ be the line through the point $A = (0, 2, 1)$ with direction vector $\mathbf{d} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Find the distance from B to ℓ.

First we compute

$$\mathbf{v} = \overrightarrow{AB} = \mathbf{b} - \mathbf{a} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Now

$$\text{proj}_{\mathbf{d}}(\mathbf{v}) = \frac{\mathbf{d} \cdot \mathbf{v}}{\mathbf{d} \cdot \mathbf{d}} \mathbf{d} = \frac{2}{3} \mathbf{d} = \begin{bmatrix} 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}.$$

We want the length of the vector

$$\mathbf{v} - \text{proj}_{\mathbf{d}}(\mathbf{v}) = \begin{bmatrix} 1/3 \\ -5/3 \\ 4/3 \end{bmatrix}$$

which is

$$\|\mathbf{v} - \text{proj}_{\mathbf{d}}(\mathbf{v})\| = \left\| \begin{bmatrix} 1/3 \\ -5/3 \\ 4/3 \end{bmatrix} \right\| = \frac{1}{3} \left\| \begin{bmatrix} 1 \\ -5 \\ 4 \end{bmatrix} \right\| = \frac{1}{3}\sqrt{42}.$$

If the line has general form $ax + by = c$ and $B = (x_0, y_0)$, then the equation above becomes

$$d(B, \ell) = \frac{|ax_0 + by_0 - c|}{\sqrt{a^2 + b^2}}.$$

Let $\mathscr{P}$ be a plane with normal vector $\mathbf{n}$. Let B be a point not on P, let P be the foot of the perpendicular from B to $\mathscr{P}$, and let A be a distinct point on $\mathscr{P}$. Then the distance of a point B to the plane is defined as

$$d(B, \mathscr{P}) = \left\| \text{proj}_{\mathbf{n}}(\overrightarrow{AB}) \right\|.$$

Chapter 2: Systems of Linear Equations

(Last Updated: September 14, 2021)

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed).

1. Introduction to Systems of linear equations

Definition: Linear equations, coefficients, constant term

A *linear equation* in n variables $x_1, \ldots, x_n$ is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b.$$

The scalars $a_1, \ldots, a_n$ are called the *coefficients* and the scalar b is called the *constant term*

A linear equation is simply an equation of a "linear object" in $\mathbb{R}^n$ (line, plane, etc.). We will be concerned with systems of linear equations and their solutions (points of intersection).

Definition: Solution

A *solution* of a linear equations $a_1x_1+\cdots+a_nx_n = b$ is a vector $[s_1, \ldots, s_n]$ whose components satisfy the equation when we substitute s_i for x_i.

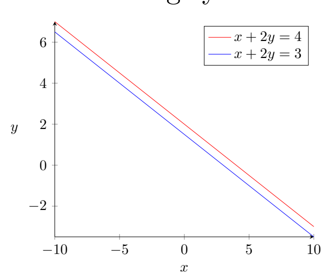
A linear equation will, by definition, have an infinite number of solutions. A greater challenge will be to determine the common solutions amongst a system.

Definition: System of linear equations, solution, solution set

A *system of linear equations* is a finite set of linear equations in the same variables. A *solution* of a system of linear equations is a vector that is a solution to each linear equation i the system. The *solution set* of a system of linear equations is the set of all solutions of the system.

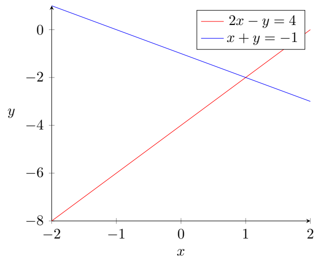
Example. Consider the following systems of equations:

$$x + 2y = 3$$
$$x + 2y = 4$$



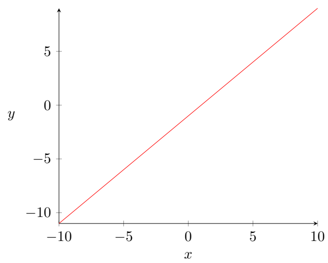
This system has no solutions because the lines are parallel and thus never intersect. One could also solve the system using methods from high school algebra (elimination or substitution).

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



This system has exactly one solution, the intersection point $(1, -2)$ of the two lines.

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



This system has infinitely many solutions as both equations correspond to the same line.

Definition: Consistent, inconsistent

A system of linear equations is *consistent* if it has at least one solution. A system with no solutions is called *inconsistent*.

Every system (with real coefficients) will either have a unique solution, infinitely many solutions, or no solutions.

When confronted with a system, we are most often interested in the following two questions:

- (Existence) Is the system consistent?
- (Uniqueness) If the system consistent, is there a *unique* solution?

Exercise. Solve the following system of equations using tools from high school algebra and interpret your solution geometrically.

$$
\begin{aligned}
x + y - z &= 4 \\
2x - y + 3z &= -13 \\
-x + 2y - z &= 8
\end{aligned}
$$

2. Direct methods for solving linear systems

A matrix is just a rectangular array. When we say a matrix is $m \times n$, we mean it has m rows and n columns. The following are examples of matrices:

$$\underset{2\times 3 \text{ matrix}}{\begin{bmatrix} 1 & 0 & -3 \\ -2 & 5 & 4 \end{bmatrix}} \quad \underset{3\times 2 \text{ matrix}}{\begin{bmatrix} 1 & -7 \\ 6 & 3 \\ 5 & 11 \end{bmatrix}} \quad \underset{2\times 2 \text{ matrix}}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$$

Definition: Coefficient matrix

Consider a system of m equations in n variables. The $m \times n$ matrix C formed by setting the (i, j) entry to be the coefficient of x_j in the ith equation is called the *coefficient matrix* of this system. The *augmented matrix* A is an $m \times (n + 1)$ matrix formed just as C but whose last column contains the constants of each equation.

Example.

A system of equations	coefficient matrix	augmented matrix
$\begin{aligned} x_1 + x_2 - x_3 &= 4 \\ 2x_1 - x_2 + 3x_3 &= -13 \\ -x_1 + 2x_2 - x_3 &= 8 \end{aligned}$	$\begin{bmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ -1 & 2 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & -1 & 3 & -13 \\ -1 & 2 & -1 & 8 \end{bmatrix}$

Our goal will be to perform operations on the augmented matrix corresponding to the system in a way that reveals the solution to the system.

Definition: (Reduced) row echelon form

A matrix is in *row echelon form* (REF) if satisfies the following:

(1) Any rows consisting entirely of zeros are at the botton.
(2) In each nonzero row, the first nonzero entry (called the *leading entry*) is in a column to the left of any leading entries below it.

A matrix is in *reduced row echelon form* (RREF) if it is in REF and is satisfies the following:

(3) The leading entry in each nonzero row is 1 (called a *leading* 1).
(4) Each leading 1 is the only nonzero entry in its column.

Example. The following matrices are in row echelon form. Only the second is in RREF.

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

The RREF of a matrix is unique, though there can be many ways to get to it. A matrix (corresponding to a system) in echelon form is easy to solve using *back substitution.*

Example. Consider the system

$$3x_1 - 5x_2 = 2$$
$$-x_2 = 1$$

From the second equation we clearly have $x_2 = -1$. Substituting this into the first equation gives $3x_1 + 5x_2 = 2$, so $3x_1 = -3$. Hence, the solution is $(-1, -1)$.

The following operations can be performed on any matrix without changing the solution set of the corresponding system. The process of putting a matrix in echelon form is called *row reduction.*

Elementary Row Operations

a. (Interchange) Interchange two rows.
b. (Scaling) Multiply a row by a nonzero constant.
c. (Replacement) Add a multiple of one row to another.

Initially we locate a (nonzero) leading entry as far to the left as possible, called a *pivot.* We use this to eliminate all the entries below it and then move to the right and identify our next pivot entry.

Example. Put the following matrix in RREF:

$$\begin{bmatrix} 1&2&4&5\\1&2&4&5\\2&4&5&4\\4&5&4&2 \end{bmatrix} \xrightarrow[\text{R2+(-1)R1}]{\text{Repl}} \begin{bmatrix} 1&2&4&5\\0&0&0&0\\2&4&5&4\\4&5&4&2 \end{bmatrix} \xrightarrow[\text{R3+(-2)R1}]{\text{Repl}} \begin{bmatrix} 1&2&4&5\\0&0&0&0\\0&0&-3&-6\\4&5&4&2 \end{bmatrix} \xrightarrow[\text{R4+(-4)R1}]{\text{Repl}} \begin{bmatrix} 1&2&4&5\\0&0&0&0\\0&0&-3&-6\\0&-3&-12&-18 \end{bmatrix}$$
$$\xrightarrow[\text{R2} \leftrightarrow \text{R4}]{\text{IC}} \begin{bmatrix} 1&2&4&5\\0&-3&-12&-18\\0&0&-3&-6\\0&0&0&0 \end{bmatrix} \xrightarrow[-\frac{1}{3}\cdot\,\text{R2}]{\text{Scale}} \begin{bmatrix} 1&2&4&5\\0&1&4&6\\0&0&-3&-6\\0&0&0&0 \end{bmatrix} \xrightarrow[-\frac{1}{3}\cdot\,\text{R3}]{\text{Scale}} \begin{bmatrix} 1&2&4&5\\0&1&4&6\\0&0&1&2\\0&0&0&0 \end{bmatrix}$$
$$\xrightarrow[\text{R2+(-4)R3}]{\text{Repl}} \begin{bmatrix} 1&2&4&5\\0&1&0&-2\\0&0&1&2\\0&0&0&0 \end{bmatrix} \xrightarrow[\text{R1+(-2)R2}]{\text{Repl}} \begin{bmatrix} 1&0&4&9\\0&1&0&-2\\0&0&1&2\\0&0&0&0 \end{bmatrix} \xrightarrow[\text{R1+(-4)R3}]{\text{Repl}} \begin{bmatrix} 1&0&0&1\\0&1&0&-2\\0&0&1&2\\0&0&0&0 \end{bmatrix}$$

This example corresponds to a system with a unique solution, $(x_1, x_2, x_3) = (1, -2, 2)$.

Example. The following matrix row reduces as

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

The reduced form of the augmented matrix has a pivot in each column. This corresponds to an inconsistent system. The reason is that, translating back to a system, we get the equations $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, and $0 = 1$. This last equation is impossible so there is no solution to the system.

A (linear) system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column. If a linear system is consistent, then the solution set contains either (i) a unique solution (no free variables) or (ii) infinitely many solutions. Condition (i) is equivalent to no row of the form $\begin{bmatrix} 0 & \cdots & 0 & b \end{bmatrix}$, $b \neq 0$ in the RREF form of the matrix.

Definition: Row equivalent

Two matrices are said to be *row equivalent* if one is obtainable from the other by a series of elementary row operations.

Theorem 1: Row equivalent matrices have same echelon form

Two matrices are row equivalent if and only if they can be reduced to the same row echelon form.

Here we lay out the algorithm to reducing matrices more explicitly. It is not strictly necessary to follow these steps but this should serve as a good guide in the beginning.

Row reduction algorithm (Gaussian Elimination)

a. Begin with the leftmost nonzero column (this is a pivot column).
b. Interchange rows as necessary so the top entry is nonzero.
c. Use row operations to create zeros in all positions below the pivot.
d. Ignoring the row containing the pivot, repeat 1-3 to the remaining submatrix. Repeat until there are no more rows to modify.
e. Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. Make each pivot 1 by scaling.

Example. The following matrix is in RREF and corresponds to the given system

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 - 5x_3 &= 1 \\ x_2 + x_3 &= 4 \\ 0 &= 0 \end{aligned}$$

The variables x_1 and x_2 corresponding to pivot columns are called *basic variables*. The variable x_3 does not correspond to a pivot column and is called a *free variable*. This is because there is a solution for *any* choice of x_3. If we solve each equation for its leading term, we have

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

Assigning the parameter t to x_3, that is $x_3 = t$, we can express in the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1+5t \\ 4-t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} + t \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}.$$

That is, the solution set corresponds to a line though the point $(1, 4, 0)$ with direction vector $\left[\begin{smallmatrix} 5 \\ -1 \\ 1 \end{smallmatrix}\right]$.

Definition: Rank

The *rank* of a matrix is the number of nonzero rows in its row echelon form.

Note that the rank is also the number of pivot entries in the matrix.

Theorem 2: The Rank Theorem

Let A be the coefficient matrix of a system of linear equations with n variables. If the system is consistent, then

$$\text{number of free variables} = n - \text{rank}(A).$$

The rank of the matrix in the previous example was 2.

Definition: Homogeneous system

A system of linear equations is *homogeneous* if each constant term is zero.

A homogeneous system always have at least one solution: the zero vector, which we call the *trivial solution*. The next theorem tells us a case where it is guaranteed tot have infinitely many solutions.

Theorem 3: Infinitely many solutions for homogeneous systems

If $[A|0]$ is a homogeneous system of m linear equations with n variables where $m < n$, then the system has infinitely many solutions.

Note however that even when $m \geq n$ there many be infinitely many solutions.

Example. Consider the following homogeneous 'system' with one equation: $3x_1 + 2x_2 - 5x_3 = 0$.

A general solution is $x_1 = -\frac{2}{3}x_2 + \frac{5}{3}x_3$ with x_2 and x_3 free. Write $x_2 = s$ and $x_3 = t$ so

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3}x_2 + \frac{5}{3}x_3 \\ s \\ t \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} t.$$

That is, the solution set represents a plane through the origin.

Example. Let $\mathbf{v}_1 = \begin{bmatrix} 5 \\ -5 \\ 25 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 3 \\ -6 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 10 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 5 \end{bmatrix}$.

(1) Find all solutions to the homogeneous system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$.

We form the augmented matrix of the system and row reduce,

$$\begin{bmatrix} 5 & 0 & 2 & 0 \\ -5 & 3 & -2 & 0 \\ 25 & -6 & 10 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & \frac{2}{5} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then $x_1 = (-2/5)x_3$, $x_2 = 0$ and x_3 is a free variable. Setting $x_3 = s$ we have

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (-2/5)s \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} -2/5 \\ 0 \\ 1 \end{bmatrix} s.$$

(2) Find all solutions to the system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$.

Again we form the augmented matrix of the system and row reduce,

$$\begin{bmatrix} 5 & 0 & 2 & 1 \\ -5 & 3 & -2 & -1 \\ 25 & -6 & 10 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & \frac{2}{5} & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The free variable is still x_3 so setting $x_3 = s$ gives the solution

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

This solution is similar to the one before. We call the portion of the solution containing the free variable x_3 the *homogeneous solution* of the system.

We now show how our method of solving systems may be applied to balancing chemical equations.

Example. When propane gas burns, propane C_3H_8 combines with oxygen O_2 to form carbon dioxide CO_2 and water H_2O according to an equation of the form

$$(x_1)C_3H_8 + (x_2)O_2 \longrightarrow (x_3)CO_2 + (x_4)H_2O.$$

To *balance the equation* means to find x_i such that the total number of atoms on the left equals the total on the right. We translate the chemicals into vectors $\left[\begin{smallmatrix} C \\ H \\ O \end{smallmatrix}\right]$:

$$C_3H_8 = \left[\begin{smallmatrix} 3 \\ 8 \\ 0 \end{smallmatrix}\right], \qquad O_2 = \left[\begin{smallmatrix} 0 \\ 0 \\ 2 \end{smallmatrix}\right], \qquad CO_2 = \left[\begin{smallmatrix} 1 \\ 0 \\ 2 \end{smallmatrix}\right], \qquad H_2O = \left[\begin{smallmatrix} 0 \\ 2 \\ 1 \end{smallmatrix}\right].$$

Balancing now becomes the linear system,

$$x_1\left[\begin{smallmatrix} 3 \\ 8 \\ 0 \end{smallmatrix}\right] + x_2\left[\begin{smallmatrix} 0 \\ 0 \\ 2 \end{smallmatrix}\right] = x_3\left[\begin{smallmatrix} 1 \\ 0 \\ 2 \end{smallmatrix}\right] + x_4\left[\begin{smallmatrix} 0 \\ 2 \\ 1 \end{smallmatrix}\right] \Rightarrow x_1\left[\begin{smallmatrix} 3 \\ 8 \\ 0 \end{smallmatrix}\right] + x_2\left[\begin{smallmatrix} 0 \\ 0 \\ 2 \end{smallmatrix}\right] - x_3\left[\begin{smallmatrix} 1 \\ 0 \\ 2 \end{smallmatrix}\right] - x_4\left[\begin{smallmatrix} 0 \\ 2 \\ 1 \end{smallmatrix}\right] = \mathbf{0}.$$

We form the augmented matrix and row reduce,

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

Thus, the solution is $x_1 = (1/4)x_4$, $x_2 = (5/4)x_4$, $x_3 = (3/4)x_4$ with x_4 free. Since only nonnegative solutions make sense in this context, any solution with $x_4 \geq 0$ is valid. For example, when $x_4 = 4$,

$$C_3H_8 + 5O_2 \longrightarrow 3CO_2 + 4H_2O.$$

As another example, we discuss how solving systems may be applied to balancing an economy.

Example. Suppose an economy has four sectors: Agriculture (A), Energy (E), Manufacturing (M), and Transportation (T).

- A sells 10% of its output to E, 25% to M, and retains the rest.
- E sells 30% of its output to A, 35% to M, 25% to T, and retains the rest.
- M sells 30% of its output to A, 15% to E, 40% to T, and retains the rest.
- T sells 20% of its output to A, 10% to E, 30% to M, and retains the rest.

Leontif proved that there exist *equilibrium prices* that can be assigned to the total outputs so that the income of each sector balances its expenses. First, we construct an exchange table:

Distribution of output from:	A	E	M	T	Purchased by
	.65	.3	.3	.2	A
	.1	.1	.15	.1	E
	.25	.35	.15	.3	M
	0	.25	.4	.4	T

Denote the prices of the total annual outputs of Agriculture, Energy, Manufacturing, and Transportation by p_A, p_E, p_M, and p_T, respectively. Agriculture pays for 65% of its own output, 30% of Energy's output, 30% of Manufacturing's output, and 20% of Transportation's output. To make Agriculture's income equal its expenses, we need

$$p_A = .65p_A + .3p_E + .3p_M + .2p_T \Rightarrow .35p_A - .3p_E - .3p_M - .2p_T = 0.$$

This leads to the following system of equations:

$$\begin{aligned} .35p_A - .3p_E - .3p_M - .2p_T &= 0 \\ -.1p_A + .9p_E - .15p_M - .1p_T &= 0 \\ -.25p_A - .35p_E + .85p_M - .3p_T &= 0 \\ -.25p_E - .4p_M + .6p_T &= 0. \end{aligned}$$

We form the corresponding augmented matrix and row reduce:

$$\begin{bmatrix} .35 & -.3 & -.3 & -.2 & 0 \\ -.1 & .9 & -.15 & -.1 & 0 \\ -.25 & -.35 & .85 & -.3 & 0 \\ 0 & -.25 & -.4 & .6 & 0. \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2.03 & 0 \\ 0 & 1 & 0 & -.53 & 0 \\ 0 & 0 & 1 & -1.17 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution in parametric vector form is

$$p = \begin{bmatrix} p_A \\ p_E \\ p_M \\ p_T \end{bmatrix} = \begin{bmatrix} 2.03 \\ .53 \\ 1.17 \\ 1 \end{bmatrix} p_T.$$

Setting $p_T = 100$ gives $p_A = 203$, $p_E = 53$, and $p_M = 117$.

Example. Determine whether $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}$ is a linear combination of $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$.

We are asking whether there exists x_1, x_2 such that $x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{w}$, which is equivalent to

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 3x_1 + 2x_2 \\ -x_1 + x_2 \\ x_1 - 3x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}.$$

This gives a system of three equations and two unknowns. We form the corresponding augmented matrix and row reduce to find the solution $x_1 = -1$ and $x_2 = 2$.

Theorem 4: Linear equations and linear combinations

A system of linear equations with augmented matrix $[A|\mathbf{b}]$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Definition: Span, spanning set

The *span* of a set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is the set of all linear combinations of the vectors in S, denoted $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ or $\text{span}(S)$. If $\text{span}(S) = \mathbb{R}^n$, then S is called a *spanning set* of $\mathbb{R}^n$.

Example. Show that the set $\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right)$ is a spanning set for $\mathbb{R}^2$.

Let $[a, b]$ be an arbitrary vector in $\mathbb{R}^2$. We must show that it is a linear combination of the given vectors. Equivalently, this asks whether there exists scalars x_1, x_2 such that

$$x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

We set up the corresponding augmented matrix and row reduce:

$$\begin{bmatrix} 1 & -2 & a \\ 2 & 3 & b \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -2 & a \\ 0 & 7 & b-2a \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -2 & a \\ 0 & 1 & (b-2a)/7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & (3a+2b)/7 \\ 0 & 1 & (b-2a)/7 \end{bmatrix}.$$

Hence, this system is consistent and so a solution exists (what that solution is here is irrelevant). Therefore, the given vectors span $\mathbb{R}^2$.

A *standard unit vector* in $\mathbb{R}^n$ is a vector $\mathbf{e}_k$ whose kth component is 1 and the other components are zero. It is clear that $\mathbb{R}^n = \text{span}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$.

Geometrically, we think of the span of one vector $\mathbf{v}$ as line through the origin since any vector in $\text{span}\{\mathbf{v}\}$ is of the form $x\mathbf{v}$ for some scalar (weight) x. Similarly, the span of two vectors $\mathbf{v}_1, \mathbf{v}_2$ which are not scalar multiples forms a plane through the three points $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2$.

The goal now will be to develop *efficient* spanning sets. That is, ones that use the fewest possible number of vectors.

Definition: Linearly dependent, linearly independent

An set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is *linearly dependent* if there exists scalars (called *weights*) $c_1, \ldots, c_k$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}. \tag{1}$$

If no such scalars exist, then the set is *linearly independent.*

Example. (a) Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix}$.

We set up the augmented matrix corresponding to the equation (1) and row reduce

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

Hence, the only solution is the trivial one and so the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

(b) Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

We set up the augmented matrix corresponding to the equation (1) and row reduce

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

There is a nontrivial solution (in particular, x_3 is free) so the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

(c) The set $\{\mathbf{0}\}$ is linearly *dependent* because $c\mathbf{0} = \mathbf{0}$ for all $c \in \mathbb{R}$. Moreover, any set containing the zero vector is linearly dependent. To see this, suppose $\mathbf{v}_1 = \mathbf{0}$, then $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$. A similar argument holds if any $\mathbf{v}_i$ is the zero vector.

(d) A set of one nonzero vector, $\{\mathbf{v}\}$ is linearly independent because $c\mathbf{v} = \mathbf{0}$ implies $c = 0$.

Theorem 5: Linearly dependence implies linear combination

A set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ of vectors in $\mathbb{R}^n$ is linearly dependent if at least one of the vectors can be expressed as a linear combination of the others.

Proof. Suppose S is linearly dependent. Then there exists weights $c_1, \dots, c_m$ such that

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}.$$

If $c_1 \neq 0$, then $\mathbf{v}_1 = \frac{c_2}{c_1}\mathbf{v}_2 + \cdots + \frac{c_p}{c_1}\mathbf{v}_m$, so $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_m$. Note that it must be true that at least one of $c_2, \dots, c_m$ is nonzero.

Conversely, if $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_m$, then there exists weights $c_2, \dots, c_m$ such that $\mathbf{v}_1 = c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$. Equivalently, $-\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$, so S is linearly dependent.

Both arguments hold with any vector in place of $\mathbf{v}_1$. □

Example 6. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{u}$, $\mathbf{v}$ linearly independent. Then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent if and only if one of the vectors is in the span of the other two.

Theorem 7: Linear dependence and homogeneous systems

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a set of (column) vectors in $\mathbb{R}^n$ and let A be the $m \times n$ matrix $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{bmatrix}$. Then S is linearly dependent if and only if the homogeneous system with augmented matrix $[A|\mathbf{0}]$ has only the trivial solution.

Proof. This is clear once we realize that any solution to the homogeneous system with augmented matrix $[A|\mathbf{0}]$ is also a solution to (1). □

Theorem 8: Linear dependence and rank

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a set of (row) vectors in $\mathbb{R}^n$ and let A be the matrix

$$A = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{bmatrix}.$$

Then S is linearly dependent if and only if $\operatorname{rank}(A) < m$.

Corollary 9: Linear dependence in $\mathbb{R}^n$

Any set of m vectors is linearly dependent if $m > n$.

Chapter 3: Matrices

(Last Updated: October 7, 2021)

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed). The section on Markov Chains includes some material written by my colleague and collaborator Robert Won.

1. Matrix Operations

Recall that a *matrix* is a rectangular array of numbers called *entries* (or *elements*) of the matrix. The *size* of a matrix is $m \times n$ where m is the number of rows and n is the number of columns. A *row vector* is a $1 \times m$ matrix, while a *column vector* is a $n \times 1$ matrix.

If A is a matrix, we denote the element is row i and column j by a_{ij}, so sometimes we write $A = [a_{ij}]_{m\times n}$ to indicate that A is the matrix

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

Definition: Square, digaonal entries, diagonal/scalar/identity/zero matrix

Let $A = [a_{ij}]_{m\times n}$.
- The matrix A is *square* if $m = n$.
- The *diagonal entries* of a matrix A are $a_{11}, a_{22}, a_{33}, \ldots$. Generally, those a_{ij} with $i = j$.
- A square matrix whose nondiagonal entries are zero is *diagonal* ($a_{ij} = 0$ for all $i \neq j$).
- A diagonal matrix whose diagonal entries are all equal is a *scalar matrix*.
- A scalar matrix whose diagonal entries are all 1 is an *identity matrix* (denoted I_n).
- A matrix A is a *zero matrix* if $a_{ij} = 0$ for all i, j. We often write O for the zero matrix when m and n are implied.

Two matrices $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{r\times s}$ are *equal* if $m = r$, $n = s$, and $a_{ij} = b_{ij}$ for all i, j.

We can now define two operations on matrices, which will resemble operations on vectors in $\mathbb{R}^n$. There is a very good reason for this. The set of $m \times n$ matrices (with entries in $\mathbb{R}$) is another example of a *vector space*.

Matrix addition and scalar multiplication

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be matrices of the *same size* $m \times n$ and $c \in \mathbb{R}$.

- Matrix addition: $A + B = [a_{ij} + b_{ij}]$ is the $m \times n$ matrix whose (i, j)-entry is $a_{ij} + b_{ij}$.
- Scalar multiplication: $cA = [ca_{ij}]$ is the $m \times n$ matrix whose (i, j)-entry is ca_{ij}.

Example. Let $A = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix}$. Then

$$A + 3B = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} + 3\begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} + \begin{bmatrix} 3 & 15 \\ -6 & 9 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 17 \\ -5 & 14 \\ 2 & 3 \end{bmatrix}.$$

The matrix $(-1)A$ is written $-A$ and so we can define substraction of matrices by $A - B = A + (-1)B$. If A is $m \times n$ and O is the $m \times n$ zero matrix, then clearly $A + O = A = O + A$ and $A - A = O$.

Here we define matrix multiplication algebraically. Later we will see matrix multiplication arise organically through the study of linear transformations.

Matrix multiplication

Let $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{n \times r}$. The *product* $C = AB$ is an $m \times r$ matrix with entries

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

This way of multiplying matrices is known as the *row-column rule*.

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

$$AB = \begin{bmatrix} 2(1) + 1(-3) & 2(2) + 1(1) & 2(4) + 1(6) \\ -1(1) + 5(-3) & -1(2) + 5(1) & -1(4) + 5(6) \end{bmatrix} = \begin{bmatrix} -1 & 5 & 14 \\ -16 & 3 & 26 \end{bmatrix}.$$

Note that the product BA is not defined in the previous example. In general, even if AB and BA are both defined, we would expect that $AB \neq BA$. Also, cancellation does not work. That is, it is possible for $AB = 0$ even if $A, B \neq 0$. For example,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{ but } \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

If A is an $n \times n$ matrix, then the product AA is defined. In general, for $k > 0$ we define the *kth matrix power* of A to be

$$A^k = A \cdot A \cdots A \ (k \text{ times}).$$

Thus, $A^1 = A$. We set, by convention, $A^0 = I_n$, the $n \times n$ identity matrix.

Properties of matrix powers

If A is a square matrix and r and s are nonnegative integers, then

a. $A^r A^s = A^{r+s}$

b. $(A^r)^s = A^{rs}$

Given a linear system, we can express as a matrix equation like so:

$$\begin{aligned} x_1 + x_2 - x_3 &= 4 \\ 2x_1 - x_2 + 3x_3 &= -13 \\ -x_1 + 2x_2 - x_3 &= 8 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -13 \\ 8 \end{bmatrix}$$

In fact, every linear system can be expressed in the form $A\mathbf{x} = \mathbf{b}$.

Theorem 1: Rows and columns of matrices

Let A be an $m \times n$ matrix.

(1) If $\mathbf{e}_i$ is a $1 \times m$ standard unit vector, then $\mathbf{e}_i A$ is the ith row of A.

(2) If $\mathbf{e}_j$ a $n \times 1$ standard unit vector, then $A\mathbf{e}_j$ is the jth column of A.

Definition: Partition

A *partition* of a matrix A is a decomposition of A into rectangular submatrices $A_1, \ldots, A_n$ such that each entry in A lies in some submatrix.

Example. The matrix

$$A = \begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 1 & 0 & 7 \\ 4 & 5 & 2 & 1 \end{bmatrix}$$

can be partitioned as $A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$ where

$$A_{11} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 0 \end{bmatrix} \quad A_{12} = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \quad A_{21} = \begin{bmatrix} 4 & 5 & 2 \end{bmatrix} \quad A_{22} = \begin{bmatrix} 1 \end{bmatrix}.$$

There are many other partitions of this matrix and in general a matrix will have several partitions. I leave it as an exercise to write another partition of this matrix.

Suppose A is $m \times n$ and B is $n \times r$ so that the product AB is defined. If we decompose B into its column vectors as $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_r \end{bmatrix}$. Then the product AB can be computed as

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_r \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_r \end{bmatrix}.$$

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

$$A\mathbf{b}_1 = \begin{bmatrix} -1 \\ -16 \end{bmatrix}, A\mathbf{b}_2 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}, A\mathbf{b}_3 = \begin{bmatrix} 14 \\ 26 \end{bmatrix}.$$

So we recover our previous result that

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} -1 & 5 & 14 \\ -16 & 3 & 26 \end{bmatrix}.$$

On the other hand, if we decompose A into its row vectors $A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$ then

$$AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} B = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}.$$

Exercise. Recompute the previous example using this method.

Definition: Transpose, symmetric

If A is an $m \times n$ matrix, the *transpose* of A, denoted A^T, is an $n \times m$ matrix whose columns are the rows of A. That is, $(A^T)_{ij} = A_{ji}$. A matrix A is symmetric if $A^T = A$.

It should be clear from the definition that a symmetric matrix is necessarily square.

Example. Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{bmatrix}$. Then $A^T = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}$, $B^T = \begin{bmatrix} 1 & -3 \\ 2 & 1 \\ 4 & 6 \end{bmatrix}$. Moreover,

$$B^T A^T = \begin{bmatrix} 1(2) + (-3)(1) & 1(-1) + (-3)(5) \\ 2(2) + 1(1) & 2(-1) + 1(5) \\ 4(2) + 6(1) & 4(-1) + 6(5) \end{bmatrix} = \begin{bmatrix} -1 & -16 \\ 5 & 3 \\ 14 & 26 \end{bmatrix} = (AB)^T.$$

The transpose can be used to give an alternate definition of the dot product. Let $\mathbf{u}$ and $\mathbf{v}$ be column vectors in $\mathbb{R}^n$ (so $n \times 1$ matrices). Then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

The first theorem should be compared to algebraic properties of vectors.

Theorem 2: Algebraic Properties of Matrix Addition and Scalar Multiplication

Let A, B, and C be matrices of the same size and let c and d be scalars. Then

a. $A + B = B + A$

b. $(A+B)+C = A+(B+C)$

c. $A + O = A$

d. $A + (-A) = O$

e. $c(A + B) = cA + cB$

f. $(c + d)A = cA + dA$

g. $c(dA) = (cd)A$

h. $1A = A$.

Let $A_1, A_2, \ldots A_k$ be matrices of the same size and let $c_1, c_2, \ldots, c_k$ be scalars. A *linear combination* of the A_i is an expression of the form

$$c_1A_1 + c_2A_2 + \cdots + c_kA_k.$$

The scalars c_i are called *coefficients*.

Example. Determine whether $B = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$ is a linear combination of

$$A_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \text{ and } A_3 = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}.$$

We need coefficients c_1, c_2, c_3 such that $c_1A_1 + c_2A_2 + c_3A_3 = B$. This gives

$$\begin{aligned} c_1 \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \\ \begin{bmatrix} c_1 + c_3 & c_1 - c_2 - c_3 \\ -c_1 + c_2 & c_1 \end{bmatrix} &= \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}. \end{aligned}$$

Hence, we have a linear system whose augmented matrix we row reduce:

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

Thus $2A_1 - 3A_2 + A_3 = B$.

There is a natural correspondence between matrices and column vectors which turns the above problem into the more familiar vector version. We will explore this correspondence more later.

A set of matrices $A_1, A_2, \ldots A_k$ of the same size are *linearly independent* if the only solution to the equation

$$c_1A_1 + c_2A_2 + \cdots + c_kA_k = O$$

is the trivial one ($c_1 = c_2 = \cdots = c_k = 0$). Otherwise the set is linearly dependent.

Example. The matrices from the previous example are linearly independent. If in the augmented column of the matrix we take all zeros (corresponding to the zero matrix), then our computations show that the unique solution to the homogeneous equation is the trivial one.

Theorem 3: Properties of Matrix multiplication

Let A, B, and C be matrices of the appropriate sizes (so the given operations are defined) and let k be a scalar. Then

a. $A(BC) = (AB)C$
b. $A(B + C) = AB + AC$
c. $(A + B)C = AC + BC$
d. $k(AB) = (kA)B = A(kB)$
e. $I_m A = A = AI_n$ if A is $m \times n$.

In general we have

$$(A + B)^2 = A^2 + AB + BA + B^2.$$

However, as noted previously, $AB \neq BA$ in general. Hence, this *does not* simplify to $A^2 + 2AB + B^2$ except in certain special cases.

Theorem 4: Properties of the Transpose

Let A and B be matrices of the appropriate sizes (so the given operations are defined) and let k be a scalar. Then

a. $(A^T)^T = A$
b. $(A + B)^T = A^T + B^T$
c. $(kA)^T = k(A^T)$
d. $(AB)^T = B^T A^T$
e. $(A^r)^T = (A^T)^r$ for all nonnegative integers r.

The next theorem is easy to check with examples, but also easy to verify in general.

Theorem 5: Sums and products with transpose

a. If A is a square matrix, then $A + A^T$ is a symmetric matrix.
b. For any matrix A, AA^T and $A^T A$ are symmetric matrices.

Proof. All of these follow from properties of the transpose. In the case of (a) we have

$$(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T.$$

For part (b), we have $(AA^T)^T = (A^T)^T A^T = AA^T$. The other one follows similarly. $\square$

3. Matrix Inverse

Ideally, we would like to be able to solve matrix equations as easily as algebraic ones. Matrix inverse allows us to do this, though it will be somewhat more computational.

Definition: Matrix inverse

If A is an $n \times n$ matrix, then an *inverse* of A is another matrix A' such that

$$AA' = I_n \quad \text{and} \quad A'A = I_n.$$

If such an A exists, then A is called invertible.

Note that the symmetry in the definition implies that if A is invertible with inverse A', then A' is also invertible with inverse A.

Example. Let $A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$ and $A' = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$. Then a simple check shows that $AA' = I_2$ and $A'A = I_2$. Hence, A is invertible.

Not every matrix is invertible. For example, it is fairly easy to see that the zero matrix is not invertible. Here is another example.

Example. Show that the matrix $B = \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}$ is not invertible.

Suppose that B has an inverse $B' = \begin{bmatrix} w & x \\ y & z \end{bmatrix}$. Then $BB' = I_2$ so

$$\begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} w & x \\ y & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which gives the system of equations

$$\begin{aligned} 2w + 6y &= 1 \\ 2x + 6z &= 0 \\ w + 3y &= 0 \\ x + 3z &= 1. \end{aligned}$$

Row reducing gives a pivot in the last column, so the system is inconsistent.

Theorem 6: Inverse of a matrix is unique

If A is an invertible matrix, then its inverse is unique.

Proof. Let A' and A'' be inverses of A. Then $A'A = I = AA'$ and $A''A = I = AA''$. Thus,

$$A'' = A''I = A''(AA') = (A''A)A' = IA' = A'.$$

$\square$

We denote the unique inverse of the (invertible) matrix A by A^{-1}. When a matrix is invertible and one knows the inverse, it is easy to solve matrix equations involving A.

Theorem 7: Invertible matrices imply unique solutions

If A is an invertible $n \times n$ matrix, then the system of linear equations given by $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^n$.

Proof. If A is invertible, then clearly,

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}.$$

Hence, $A^{-1}\mathbf{b}$ is a solution. We claim this solution is unique. Suppose $\mathbf{u}$ is another solution, then $A\mathbf{u} = \mathbf{b} = A(A^{-1}\mathbf{b})$. Multiplying both sides by A^{-1} gives,

$$\begin{aligned} A^{-1}(A\mathbf{u}) &= A^{-1}(A(A^{-1}\mathbf{b})) \\ \Rightarrow (A^{-1}A)\mathbf{u} &= (A^{-1}A)(A^{-1}\mathbf{b}) \\ \Rightarrow I_n\mathbf{u} &= I_n(A^{-1}\mathbf{b}) \\ \Rightarrow \mathbf{u} &= A^{-1}\mathbf{b}. \end{aligned}$$

Hence, the solution is unique. $\square$

In the 2×2 case there is a simple formula for the inverse. Be careful, however! This does not generalize in an obvious way to larger matrices.

Theorem 8: Inverse of a 2×2 invertible matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

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

If $ad - bc = 0$, then A is not invertible.

We call $ad - bc$ the *determinant* of the 2×2 matrix, denoted $\det(A)$.

Example. Consider the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

We set up and solve the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

By the previous theorem,

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

So,

$$\mathbf{x} = I_2\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

Theorem 9: Algebraic properties of the inverse

Let A and B be invertible $n \times n$ matrices, c a nonzero scalar, and n a nonnegative integer.

a. A^{-1} is invertible and $(A^{-1})^{-1} = A$.
b. cA is invertible and $(cA)^{-1}) = \frac{1}{c}A^{-1}$.
c. AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
d. A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.
e. A^n is invertible and $(A^n)^{-1} = (A^{-1})^n$.

In all of these cases, it suffices to show that we have *an inverse* of the given matrix. Uniqueness then implies that it is *the inverse*.

The third property is often referred to as *socks-on, shoes-on, shoes-off, socks-off*. The last property allows us to define for any positive integer n:

$$A^{-n} = (A^{-1})^n = (A^n)^{-1}.$$

Just like with real numbers, the inverse of a sum is *not* the sum of the inverses. In fact, the sum, of two invertible matrices may not be invertible.

Now we wish to demonstrate how to find the inverse of an $n \times n$ invertible matrix. In order to do this, we first need to introduce special kinds of matrices.

Definition: Elementary matrix

An *elementary matrix* is one obtained by performing a single row operation to an identity matrix.

Example. The following are elementary matrices.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} (R_3 + (-4)R_1), \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, (R_1 \leftrightarrow R_2), \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}, (5R_3)$$

The next result tells us how row operations correspond to multiplication by elementary matrices.

Theorem 10: Elementary matrices and row operations

Let E be the elementary matrix obtained by performing a single row operation on I_n. If A is an $n \times r$ matrix, then EA is the same matrix obtained by performing the same operation on A.

Proof. Let E be the elementary matrix obtained by multiplying row k of I_n by $m \neq 0$. That is,

$$E = \begin{bmatrix} \mathbf{e}_1 & \cdots & m\mathbf{e}_k & \cdots & \mathbf{e}_n \end{bmatrix}.$$

Let A be any $n \times n$ matrix. Then,

$$EA = E \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} E\mathbf{a}_1 & \cdots & E\mathbf{a}_n \end{bmatrix}.$$

It is clear that the result of the multiplication $E\mathbf{a}_i$ is the multiply the kth entry of $\mathbf{a}_i$ by m. Thus, the resulting matrix EA is the same as A, except each entry in the kth row is multiplied by m.

The proofs for the other two row operations are similar. □

It is not hard to find the inverse of an elementary matrix.

Theorem 11: Inverse of an elementary matrix

Each elementary matrix is invertible, and its inverse is an elementary matrix of the same type.

For example, if $a \neq 0$ and E is obtained by multiplying row i of I_n by a, then E^{-1} is obtained by multiplying row i of I_n by $1/a$.

Proof. Let E be an elementary matrix and F the elementary matrix that transforms E back into I. Clearly such an F exists because every row operation is reversible. Moreover, E must reverse F. Hence, $FE = I_n$ and $EF = I_n$, so F is the inverse of E. □

We can now state our main result. This theorem is very important in this class and as we go through we will expand upon it. It has many names, including simply the *Invertible Matrix Theorem*. But your text uses a name that aligns with other major theorems in math courses you have taken, including the Fundamental Theorem of Algebra and the Fundamental Theorem of Calculus.

Theorem 12: The Fundamental Theorem of Invertible Matrices

Let A be an $n \times n$ matrix. The following statements are equivalent.

a. A is invertible.
b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$.
c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
d. The reduced row echelon form of A is I_n.
e. A is a product of elementary matrices.

Proof. (a) $\Rightarrow$ (b) This was proved previously.

(b) $\Rightarrow$ (c) Since $\mathbf{0}$ is one solution, then (b) implies it is the unique solution.

(c) $\Rightarrow$ (d) We row reduce the augmented matrix $[A|\mathbf{0}]$. Since the system has only the trivial solution, then $\text{rank}(A) = n$. That is, A row reduces to the identity matrix.

(d) $\Rightarrow$ (e) Since A row reduces to I_n, there is a sequence of elementary matrices such that $E_k \cdots E_2 E_1 A = I_n$. But each E_i is invertible and so their product is invertible. Thus

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

That is, A is invertible. $\square$

There are several major consequences of this theorem. The first is that checking invertibility only requires checking on one side.

Theorem 13: One-sided inverse implies inverse

Let A be a square matrix. If B is a square matrix such that $AB = I$ or $BA = I$, then $B = A^{-1}$.

Proof. Suppose $BA = I$ and consider the equation $A\mathbf{x} = \mathbf{0}$. Left multiplying both sides of the equation by B gives $(BA)\mathbf{x} = B\mathbf{0} = \mathbf{0}$, so $\mathbf{x} = \mathbf{0}$. Hence, $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, so A is invertible by the Fundamental Theorem. Hence, A^{-1} exists. Since $BA = I$,

$$B = BI = B(AA^{-1}) = (BA)A^{-1} = IA^{-1} = A^{-1}.$$ $\square$

The next theorem tells us how to compute the inverse of an invertible matrix.

Theorem 14: Row operations and the inverse

Let A be a square matrix. If a sequence of elementary row operations reduces A to I, then the same sequence of elementary row operations transform I into A.

Proof. Let $E_1, E_2, \ldots, E_k$ be the sequence of elementary row operations that transform A to I. Hence, $E_k \cdots E_2 E_1 A = I$. Set $B = E_k \cdots E_2 E_1$, so $BA = I$ and by the previous theorem $A^{-1} = B = BI = E_k \cdots E_2 E_1 I$. Thus, the same row operations transform I into $B = A^{-1}$. $\square$

Suppose A is an invertible matrix. We now form a "super augmented" matrix $[A|I]$. The previous theorem shows that performing row reduction so that $A \to I$ will transform $I \to A^{-1}$. That is, $[A|I] \to [I|A^{-1}]$.

Example. Find the inverse of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$.

We form the augmented matrix $[A|I_3]$ and row reduce A to I_3.

$$[A|I_3] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right]$$

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

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right].$$

Thus, $A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$.

Here is another perspective which aligns with our understanding of systems of equations. Let $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$ be the inverse of a square matrix $AX = I$. So

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} = I = AX = A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_n \end{bmatrix}.$$

That is, $\mathbf{x}_1$ is a solution to the equation $A\mathbf{x}_1 = \mathbf{e}_1$. To solve this, we form the augmented matrix $[A|\mathbf{e}_1]$ and row reduce. But A is invertible and so it row reduces to the identity. Hence, this will give us the unique solution to $A\mathbf{x}_1 = \mathbf{e}_1$. So, using the method above allows us to simultaneously solve for each $\mathbf{x}_i$ at once.

A plane in $\mathbb{R}^3$ through the origin is essentially a copy of $\mathbb{R}^2$. For example, the xy-plane or yz-plane are basically indistinguishable from $\mathbb{R}^2$. We will make this idea more precise by introducing the concept of *subspces*.

Definition: Subspace

A *subspace* of $\mathbb{R}^n$ is any collection S of vectors in $\mathbb{R}^n$ such that:

a. The zero vector $\mathbf{0}$ is in S.
b. If $\mathbf{u}$ and $\mathbf{v}$ are in S, then $\mathbf{u}+\mathbf{v}$ is in S (*closure under addition*).
c. If $\mathbf{u}$ is in S and c is a scalar, then $c\mathbf{u}$ is in S (*closure under scalar multiplication*).

The last two properties could be combined to say simply that S is *closed under linear combinations*.

Example. (1) Consider the set $\{\mathbf{0}\}$ (the set containing only the zero vector for some n). Obviously, $\mathbf{0} \in \{\mathbf{0}\}$. Since $\mathbf{0}+\mathbf{0}=\mathbf{0}$ and $c\mathbf{0}=\mathbf{0}$ for all $c \in \mathbb{R}$, then $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ (for every n.) We call this the *trivial subspace* of $\mathbb{R}^n$.

(2) $\mathbb{R}^n$ is a subspace of itself. We call any subspace of $\mathbb{R}^n$ that is not $\mathbb{R}^n$ itself a *proper subspace*.

(3) Let S be the set of vectors $\begin{bmatrix} x & y & z \end{bmatrix}$ in $\mathbb{R}^3$ such that $x = 3y$ and $z = -2y$. We claim that S is a subspace.

Substituting the conditions, we see that S is simply the set of vectors $\begin{bmatrix} 3 & 1 & -2 \end{bmatrix} y$ with $y \in \mathbb{R}$. Letting $y = 0$ we see that $\mathbf{0}$ is in S. If y and y' are real numbers, then

$$\begin{bmatrix} 3 & 1 & -2 \end{bmatrix} y + \begin{bmatrix} 3 & 1 & -2 \end{bmatrix} y' = \begin{bmatrix} 3 & 1 & -2 \end{bmatrix} (y+y').$$

So S is closed under addition. Similarly, if c is a real number, then

$$c\left(\begin{bmatrix} 3 & 1 & -2 \end{bmatrix} y\right) = \begin{bmatrix} 3 & 1 & -2 \end{bmatrix} (cy)$$

is in S. Thus, S is closed under scalar multiplication.

Theorem 19: Span of a set is a subspace

Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be vectors in $\mathbb{R}^n$. Then $\operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ is a subspace of $\mathbb{R}^n$.

Proof. Let $S = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. Note that $\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k$, so $\mathbf{0}$ is in S. Now let $\mathbf{u}$ and $\mathbf{v}$ be vectors in S. Then

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$
$$\mathbf{v} = d_1\mathbf{v}_2 + \cdots + d_k\mathbf{v}_k,$$

for some scalars c_i and d_i. Then

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) + (d_1\mathbf{v}_2 + \cdots + d_k\mathbf{v}_k) \\ &= (c_1+d_1)\mathbf{v}_1 + \cdots + (c_k+d_k)\mathbf{v}_k,\end{aligned}$$

which is an element of S. Thus, S is closed under addition. Now let c be a scalar, then

$$c\mathbf{u} = c(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = (cc_1)\mathbf{v}_1 + \cdots + (cc_k)\mathbf{v}_k,$$

which is an element of S. Thus, S is closed under scalar multiplication. It follows that S is a subspace. □

Example. (1) Consider (3) from the previous example. In this case, $S = \text{span}(\begin{bmatrix}3 & 1 & -2\end{bmatrix})$, so it follows from the theorem that S is a subspace.

(2) Let S be the set of vectors $\begin{bmatrix}x & y & z\end{bmatrix}$ in $\mathbb{R}^3$ such that $x = 3y+1$ and $z = -2y$. This is the set of vectors of the form

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

This set does not contain the zero vector, and so S is not a subspace.

(3) Define a set H in $\mathbb{R}^2$ by the following property: $\mathbf{v} \in H$ if $\mathbf{v}$ has exactly one nonzero entry. Then H is not a subspace because it does not contain $\mathbf{0}$.

(4) Consider the set S of vectors $\begin{bmatrix}x \\ y\end{bmatrix}$ where $y = x^2$. This set contains $\mathbf{0}$, but it is not a subspace. Note that $\begin{bmatrix}1 \\ 1\end{bmatrix}$ is an element of S, but

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

is not an element of S.

Next we'll define two subspaces related to matrices.

Definition: Row space, column space

Let A be an $m \times n$ matrix.

a. The *row space* of a matrix A is the subspace $\text{row}(A)$ of $\mathbb{R}^n$ spanned by the rows of A.
b. The *column space* of a matrix A is the subspace $\text{col}(A)$ of $\mathbb{R}^m$ spanned by the columns of A.

Recall that $\text{col}(A)$ is the set of all vectors of the form $A\mathbf{x}$ where $\mathbf{x}$ is in $\mathbb{R}^n$.

Example. Let $A = \begin{bmatrix} 0 & 3 \\ -2 & 5 \\ 1 & 2 \end{bmatrix}$. (1) Is $\mathbf{b} = \begin{bmatrix} 1 \\ 5 \\ -1 \end{bmatrix}$ in $\mathrm{col}(A)$? (2) Is $\mathbf{c} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ in $\mathrm{row}(A)$?

(1) This is the equivalent to asking whether $\mathbf{b}$ is in the span of the columns of A. So we form the corresponding augmented matrix and row reduce:

$$\begin{bmatrix} 0 & 3 & 1 \\ -2 & 5 & 5 \\ 1 & 2 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -5/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, the system is consistent so $\mathbf{b}$ is in $\mathrm{col}(A)$.

(2) Whether or not $\mathbf{c}$ is in the row space of A is equivalent to asking whether $\mathbf{c}^T$ is in the column space of A^T. So we form this augmented matrix and row reduce:

$$\begin{bmatrix} 0 & -2 & 1 & 3 \\ 3 & 5 & 2 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 9/8 & 17/8 \\ 0 & 1 & -1/2 & -3/2 \end{bmatrix}.$$

Thus, the system is consistent so $\mathbf{c}$ is in $\mathrm{row}(A)$.

Theorem 20: Row equivalent matrices have the same row space

If A and B are row equivalent matrices, then $\mathrm{row}(A) = \mathrm{row}(B)$.

Theorem 21: Null space is a subspace

Let A be an $m \times n$ matrix and let N be the set of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. Then N is a subspace of $\mathbb{R}^n$.

Proof. By definition, N is a set in $\mathbb{R}^n$. Clearly $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in \mathrm{null}(A)$. Let $\mathbf{x}, \mathbf{y}$ be in N so $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. Then $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus, $\mathbf{x} + \mathbf{y}$ is in N. Finally, suppose $\mathbf{x}$ is in N and $c \in \mathbb{R}$. Then $A(c\mathbf{x}) = c(A\mathbf{x}) = c\mathbf{0} = \mathbf{0}$, so $c\mathbf{x}$ is in N. Thus, N is a subspace of $\mathbb{R}^n$. $\square$

Definition: Null space

Let A be an $m \times n$. The null space of A is the subspace of $\mathbb{R}^n$ consisting of solutions to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. It is denoted $\mathrm{null}(A)$.

Example. Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$. To compute null$(A)$, we consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. This row reduces as

$$\begin{bmatrix} 1 & -3 & -4 & 0 \\ -4 & 6 & -2 & 0 \\ -3 & 7 & 6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The solution in vector form (with $x_3 = s$) is $\mathbf{x} = \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix} s$. Hence, $\text{null}(A) = \text{span}\left(\begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix}\right)$.

Theorem 22: Trichotomy in solutions

Let A be a matrix whose entries are real numbers. For any system of linear equations $A\mathbf{x} = \mathbf{b}$, exactly one of the following is true:

a. There is no solution.
b. There is a unique solution.
c. There are infinitely many solutions.

Proof. Assume there are at least two solutions, $\mathbf{x}_1 \neq \mathbf{x}_2$. Then $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2 \neq \mathbf{0}$ and

$$A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

That is, $\mathbf{x}_0$ is a nontrivial element in null(A). Hence, $\text{span}(x_0) \subset \text{null}(A)$. Now for any scalar c,

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

Thus, the infinitely many elements $\mathbf{x}_1 + c\mathbf{x}_0$ are solutions to $A\mathbf{x} = \mathbf{b}$. □

Every subspace is the span of the vectors it contains. Preferably, we would like to express a subspace as a span of as few elements as possible. This is the idea behind a basis.

Definition: Basis

A *basis* for a subspace S of $\mathbb{R}^n$ is a linearly independent set in S that spans S.

Example. (1) The set $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis for $\mathbb{R}^n$. This is why they are sometimescalled the standard basis vectors for $\mathbb{R}^n$

(2) In a previous example, we showed that $\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix}\right)$ is a spanning set for $\mathbb{R}^2$. These vectors are also linearly independent and so this set is a basis for $\mathbb{R}^2$.

(3) Let $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$. Then $\mathbf{w} = \mathbf{v} - \mathbf{u}$. Since $\mathbf{u}$ and $\mathbf{v}$ are linearly independent. Then a basis for $\text{span}(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is $\{\mathbf{u}, \mathbf{v}\}$.

Example. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & 4 \end{bmatrix}$. Find a basis for $\text{row}(A)$, $\text{col}(A)$, and (A).

The matrix A row reduces to

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

A basis for R is

$$\left\{ \begin{bmatrix} 1 & -2 & 0 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right\}.$$

Since R is row equivalent to A, then $\text{row}(A) = \text{row}(R)$ and so this is also a basis for $\text{row}(A)$.

To find a basis for $\text{col}(A)$, we could compute a basis for $\text{row}(A^T)$. Or, we could note that non-pivot columns correspond to free variables and thus dependence relations amongst the columns of A. It follows that the pivot columns (corresponding to basic variables) are linearly independent and thus form a basis of $\text{col}(A)$. *However*, in this case we must use original columns of A.

From the above, we see that $\text{col}(A) = \text{span}(\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5)$ because those are pivot columns. Thus, a basis for $\text{col}(A)$ is

$$\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ 4 \end{bmatrix} \right\}.$$

Finally, we compute a basis for (A). We give the free variables parameters $x_2 = s$ and $x_4 = t$. Then the solution space of the matrix equation $A\mathbf{x} = \mathbf{0}$ is

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

These are linearly independent, and so a basis for (A) is

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

We now come to our main result regarding bases.

Theorem 23: The Basis Theorem

Let S be a subspace of $\mathbb{R}^n$. Then any two subspaces for S have the same number of vectors.

Proof. Let $\mathscr{B} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ be a basis of S and let $\mathscr{C} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$ be any collection of vectors in S with $s > r$. We will show that this forces a linear dependence relation amongst the $\mathbf{v}_i$.

Suppose there are scalars such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_s\mathbf{v}_2 = \mathbf{0}. \tag{1}$$

Since $\mathscr{B}$ is a basis of S, then each $\mathbf{v}_i$ may be written as a linear combination of the $\mathbf{u}_j$:

$$\begin{aligned} \mathbf{v}_1 &= a_{11}\mathbf{u}_1 + a_{12}\mathbf{u}_2 + \cdots + a_{1r}\mathbf{u}_r \\ \mathbf{v}_2 &= a_{21}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + \cdots + a_{2r}\mathbf{u}_r \\ &\vdots \\ \mathbf{v}_s &= a_{s1}\mathbf{u}_1 + a_{s2}\mathbf{u}_2 + \cdots + a_{sr}\mathbf{u}_r. \end{aligned}$$

Substituting these into (1) and regrouping we have

$$(c_1a_{11} + c_2a_{12} + \cdots + c_sa_{s1})\mathbf{u}_1 + (c_1a_{12} + c_2a_{22} + \cdots + c_sa_{s2})\mathbf{u}_2 + \cdots + (c_1a_{1r} + c_2a_{1r} + \cdots + c_sa_{sr})\mathbf{u}_r.$$

Since $\mathscr{B}$ is a basis, the $\mathbf{u}_j$ are linearly independent and hence every coefficient is zero. This gives a homogeneous linear system:

$$\begin{aligned} c_1a_{11} + c_2a_{12} + \cdots + c_sa_{s1} &= 0 \\ c_1a_{12} + c_2a_{22} + \cdots + c_sa_{s2} &= 0 \\ &\vdots \\ c_1a_{1r} + c_2a_{1r} + \cdots + c_sa_{sr} &= 0. \end{aligned}$$

This is a homogeneous linear system in the c_i. Since $r < s$, then there are infinitely many solutions. That is, there is a nontrivial solution, so $\mathscr{C}$ is not linearly independent.

Note that the same argument shows that there cannot be a basis with fewer elements that $\mathscr{B}$ as this would contradict that the $\mathbf{u}_j$ are linearly independent. Thus, every basis has r elements. $\square$

The Basis Theorem shows that the next definition is well-defined.

Definition: Dimension

The *dimension* of a nonzero subspace S of $\mathbb{R}^n$, denoted $\dim S$, is the number of vectors in any basis of S.

The trivial subspace cannot have a basis, and so we set $\dim(\{0\}) = 0$.

Example. (1) Since the standard basis for $\mathbb{R}^n$ has n vectors, then $\dim \mathbb{R}^n = n$.

(2) In the previous example, we have $\dim(\text{row}(A)) = 3$, $\dim(\text{col}(A)) = 3$, and $\dim((A)) = 2$.

It turns out that the fact that the row space and column space have the same dimension in the last example is not a coincidence, as we prove in the next theorem.

Theorem 24: Row and column space are equidimensional

The row and column space of a matrix A have the same dimension.

Proof. We need only observe that the dimension of $\text{row}(A)$ is the number of nonzero rows, which is equal to the number of leading ones (pivots) in the RREF of A. But this is exactly the number of linearly independent columns of A, which is the dimension of $\text{col}(A)$. $\square$

Definition: Rank, nullity

The *rank* of a matrix is the dimension of its row space and is denoted $\text{rank}(A)$. The nullity of a matrix A is the dimension of its null space and is denoted $\text{nullity}(A)$.

Theorem 25: Rank is invariant under transpose

For any matrix A, $\text{rank}(A^T) = \text{rank}(A)$.

Proof. We have $\text{rank}(A^T) = \dim(\text{col}(A^T)) = \dim(\text{row}(A)) = \text{rank}(A)$. $\square$

The next theorem is very important, but it is also easy to prove by observing that rank is the number of basic variables (pivot columns) and nullity is the number of free variables (non-pivot columns). It is sometimes called the *Rank-Nullity Theorem.*

Theorem 26: The Rank Theorem

If A is an $m \times n$ matrix, then $\text{rank}(A) + \text{nullity}(A) = n$.

Example. (1) Let $M = \begin{bmatrix} 7 & 2 \\ 4 & 8 \\ 6 & 6 \\ 6 & 5 \end{bmatrix}$. Then the columns of M are linearly independent and so $\text{rank}(M) = 2$. Thus, $\text{nullity}(M) = 2 - 2 = 0$.

(2) Let $N = \begin{bmatrix} 1 & -5 & 4 & -4 \\ 0 & 5 & -4 & 3 \\ 3 & 0 & 0 & -3 \end{bmatrix}$. Through row reduction we can show that $\text{rank}(N) = 2$. Thus, $\text{nullity}(M) = 4 - 2 = 2$.

We can now extend our fundamental theorem.

Theorem 27: The Fundamental Theorem of Invertible Matrices

Let A be an $n \times n$ matrix. The following statements are equivalent.

a. A is invertible.
b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$.
c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
d. The reduced row echelon form of A is I_n.
e. A is a product of elementary matrices.
f. $\text{rank}(A) = n$
g. $\text{nullity}(A) = 0$
h. The column vectors of A are linearly independent.
i. The column vectors of A span $\mathbb{R}^n$.
j. The column vectors of A form a basis for $\mathbb{R}^n$.
k. The row vectors of A are linearly independent.
l. The row vectors of A span $\mathbb{R}^n$.
m. The row vectors of A form a basis for $\mathbb{R}^n$.

Proof. We have already proved (a) through (e) are equivalent. Note that (d) $\Leftrightarrow$ (f) since row reduction does not change the number of pivots ($= \text{rank}(A)$). Now (f) $\Leftrightarrow$ (g) follows from the Rank Theorem.

(f) $\Rightarrow$ (h) Since $\text{rank}(A) = n$, then A has n pivots. Thus, in $\text{RREF}(A)$ we have

(h) $\Rightarrow$ (i) If the columns are linearly independent, then $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Since (c) $\Rightarrow$ (b), $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$. This implies (i).

(i) $\Rightarrow$ (j) Since the columns span $\mathbb{R}^n$, then they must also be linearly independent. If they weren't, then we could remove (at least) one vector without changing the span. Continuing in this way, we arrive at a linearly independent set that spans $\mathbb{R}^n$ but has fewer than n elements, a contradiction.

(j) $\Rightarrow$ (f) The columns form a basis and are therefore linearly independent. Thus, A has n pivot positions and so $\text{rank}(A) = n$. Thus, we have that (a) through (j) are equivalent.

Note that $\text{row}(A) = \text{col}(A^T)$. By the previous theorem, $\text{rank}(A) = \text{rank}(A^T)$. Hence, (k), (l), and (m) are equivalent to (f). $\square$

The fundamental theorem provides an easy way to check if a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ forms a basis for $\mathbb{R}^n$. We could check whether the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ row reduces to the identity. But in fact it is sufficient to verify that the columns are linearly independent, which we can determine from the row echelon form of A.

Example. Consider the vectors $\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix}$. The matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 4 & 4 & 4 \\ 1 & 2 & 5 \end{bmatrix} \quad \text{reduces to} \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, $\text{rank}(A) = 3$ so the vectors form a basis of $\mathbb{R}^3$.

The next theorem will be useful later. Note that if A is $m \times n$, then $A^T A$ is $n \times n$.

Theorem 28: The matrix $A^T A$

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

a. $\text{rank}(A^T A) = \text{rank}(A)$.

b. The $n \times n$ matrix $A^T A$ is invertible if and only if $\text{rank}(A) = n$.

Proof. Statement (b) follows easily from statement (a) and the fundamental theorem.

We first show that $(A) = (A^T)$. Let $\mathbf{x}$ be in (A). Then

$$(A^T A)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}.$$

On the other hand, suppose $\mathbf{y} \in (A^T A)$. Then $(A^T A)\mathbf{y} = \mathbf{0}$, so

$$(A\mathbf{y}) \cdot (A\mathbf{y}) = (A\mathbf{y})^T(A\mathbf{y}) = (\mathbf{y}^T A^T)(A\mathbf{y}) = \mathbf{y}^T(A^T A\mathbf{y}) = \mathbf{0}.$$

It follows that $\text{nullity}(A) = \text{nullity}(A^T A)$. But A and A^T have the same number of columns, so

$$\text{rank}(A) + \text{nullity}(A) = n = \text{rank}(A^T A) + \text{nullity}(A^T A).$$

The result follows. $\square$

We now see how we can explicitly view a plane in $\mathbb{R}^3$ as a copy of $\mathbb{R}^2$.

Theorem 29: Unique linear combination

Let S be a subspace of $\mathbb{R}^n$ and $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ a basis for S. If $\mathbf{v}$ is a vector in S, then there is exactly one way to write $\mathbf{v}$ as a linear combination of the vectors in $\mathscr{B}$.

Proof. Since $\mathscr{B}$ is a basis, then it spans S. Hence, there is *at least* one way to write $\mathbf{v}$ as a linear combination of the vectors in $\mathscr{B}$. Suppose there are two ways:

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$
$$\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k.$$

Then

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k$$

and so

$$(c_1 - d_1)\mathbf{v}_1 + \cdots + (c_k - d_k)\mathbf{v}_k = \mathbf{0}.$$

Since the $\mathbf{v}_i$ are linearly independent, then $c_1 - d_1 = \cdots = c_k - d_k = 0$. Thus, $c_i = d_i$ for all i. □

Definition: Coordinates, coordinate vector

Let S be a subspace of $\mathbb{R}^n$ and $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ a basis for S. Let $\mathbf{v}$ be a vector S, and write $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. The scalars $c_1, c_2, \ldots, c_k$ are called the *coordinates of* $\mathbf{v}$ *with respect to* $\mathscr{B}$ and the vector

$$[\mathbf{v}]_{\mathscr{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

is called the *coordinate vector of* $\mathbf{v}$ *with respect to* $\mathscr{B}$.

Example. Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$, and let S. Since $\mathbf{u}$ and $\mathbf{v}$ are linearly independent, then $\mathscr{B} = \{\mathbf{u}, \mathbf{v}\}$ is a basis for S.

Set $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}$, then $\mathbf{w} = -\mathbf{u} + 2\mathbf{v}$. Thus,

$$[\mathbf{w}]_{\mathscr{B}} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

6. Introduction to Linear Transformations

Linear transformations are ways to map vectors $\mathbb{R}^n \to \mathbb{R}^m$ in a way that preserves the properties of the vector space. We will see that this heavily connected to geometric ideas in $\mathbb{R}^n$ as well as to algebraic properties of matrices.

Definition: Transformation, domain, codomain, image

A *transformation* (or *mapping* or *function*) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{v}$ in $\mathbb{R}^n$ (the *domain*) a unique vector $T(\mathbf{v})$ in $\mathbb{R}^m$ (the *codomain*). Our notation for this is $T : \mathbb{R}^n \to \mathbb{R}^m$. The vector $T(\mathbf{v})$ is called the *image* of $\mathbf{v}$ under T. The set of all possible images $T(\mathbf{v})$ is called the *range* (or *image*) of T.

You have seen several such functions. The functions you study in calculus are of the form $f : \mathbb{R} \to \mathbb{R}$. The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is a transformation, but it does not preserve linearity: that is, $f(x + y) \neq f(x) + f(y)$. Here is a simple example that encapsulates what we are after.

Example. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}.$$

(1) Find $T(\mathbf{u})$.

We compute $T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = (2) \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} + (-1) \begin{bmatrix} -3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$.

(2) Find some $\mathbf{x} \in \mathbb{R}^2$ whose image under T is $\mathbf{b}$.

We need to solve the matrix equation $A\mathbf{x} = \mathbf{b}$, so we row reduce the augmented matrix $[A \mid \mathbf{b}]$:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 3/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \mathbf{x} = \begin{bmatrix} 3/2 \\ -1/2 \end{bmatrix}.$$

(3) Is $\mathbf{c}$ in the range of T?

This is basically the same question as before. We need to solve the matrix equation $A\mathbf{x} = \mathbf{c}$, so we form the augmented matrix $[A \mid \mathbf{c}]$ and row reduce.

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

The system is inconsistent, so $\mathbf{c}$ is not in the range of T.

In the previous example, we could also look at the action on an arbitrary vector $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$. We have

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x - 3y \\ 3x + 5y \\ -x + 7y \end{bmatrix}.$$

We could represent the transformation this way by explicitly saying what it does to each vector.

Definition: Linear transformation

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if

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

Another way of stating this definition is to write that

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$$

for all vectors $\mathbf{v}_1, \mathbf{v}_2$ in $\mathbb{R}^n$ and all scalars c_1, c_2.

Example. Consider the map $T : \mathbb{R} \to \mathbb{R}^2$ given by $T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$.

This map T is not linear (because $(a + b)^2 \neq a^2 + b^2$ in general). For example,

$$\begin{bmatrix} 3 \\ 9 \end{bmatrix} = T(3) = T(1 + 2) = T(1) + T(2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix},$$

which is absurd.

Example. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + y \\ x - y \end{bmatrix}$.

It is clear that T is a function. We will verify that T is linear.

Let $\mathbf{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ be in $\mathbb{R}^2$. Then

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} (x_1 + x_2) + (y_1 + y_2) \\ (x_1 + x_2) - (y_1 + y_2) \end{bmatrix} = \begin{bmatrix} (x_1 + y_1) + (x_2 + y_2) \\ (x_1 - y_1) + (x_2 - y_2) \end{bmatrix} \\ &= \begin{bmatrix} x_1 + y_1 \\ x_1 - y_1 \end{bmatrix} + \begin{bmatrix} x_2 + y_2 \\ x_2 - y_2 \end{bmatrix} = T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) + T\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = T(\mathbf{u}) + T(\mathbf{v}). \end{aligned}$$

Similarly, let $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ and let c be a scalar. Then

$$T(c\mathbf{v}) = T\left(c\begin{bmatrix} x \\ y \end{bmatrix}\right) = T\left(\begin{bmatrix} cx \\ cy \end{bmatrix}\right) = \begin{bmatrix} cx + cy \\ cx - cy \end{bmatrix} = \begin{bmatrix} c(x+y) \\ c(x-y) \end{bmatrix} = c\begin{bmatrix} x+y \\ x-y \end{bmatrix} = cT(\mathbf{v}).$$

You will be pleased to find out that these checks are not normally needed. This is because, it turns out, all linear transformation maps are determined by multiplication by some matrix, and maps determined by multiplication by some matrix are linear.

Theorem 30: Matrix transformations are linear

Let A be an $m \times n$ matrix. The matrix transformation $T_A(\mathbf{x}) = A\mathbf{x}$ (for $\mathbf{x}$ in $\mathbb{R}^n$) is a linear transformation.

Proof. We will employ our algebraic properties of matrices. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and c a scalar. Then

$$T_A(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T_A(\mathbf{u}) + T_A(\mathbf{v}),$$

and

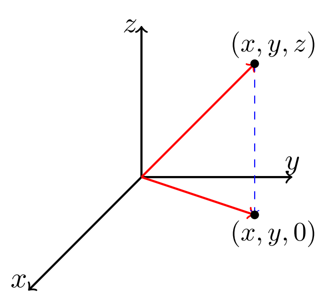
$$T_A(c\mathbf{v}) = A(c\mathbf{v}) = c(A\mathbf{v}) = cT_A(\mathbf{v}).$$

Hence, T_A is a linear transformation. $\square$

Example. In the previous example, if we set $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, then $T(\mathbf{x}) = A\mathbf{x}$. Hence, T is linear.

Example. Let $T_A : \mathbb{R}^3 \to \mathbb{R}^3$ be the matrix transformation with $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

The range of T is the xy-plane, which is a copy of $\mathbb{R}^2$. We say T is a *projection* of $\mathbb{R}^3$ onto $\mathbb{R}^2$.



We now prove the converse to the previous theorem: that every linear transformation is determined by a matrix. This proof also tells us how to find the matrix for T.

Theorem 31: Standard matrix of a linear transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is a matrix transformation. Specifically, $T = T_A$ where A is the $m \times n$ matrix

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}.$$

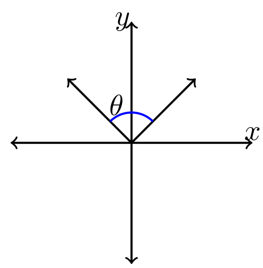
Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis vectors of $\mathbb{R}^n$. For $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$. By linearity,

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n) \\ &= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \mathbf{x} = A\mathbf{x}. \end{aligned}$$

$\square$

We call A the *standard matrix* of T, sometimes denoted simply as $[T]$.

Example. Let $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ be the matrix transformation with $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Then T_A rotates each vector in $\mathbb{R}^2$ by θ° counterclockwise.



If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ are functions, then the *composition* $S \circ T$ is defined by

$$(S \circ T)(\mathbf{x}) = S(T(\mathbf{x})) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, bend left, "T"] \arrow[rr, bend right, "S \circ T"'] & \mathbb{R}^m \arrow[r, bend left, "S"] & \mathbb{R}^p \end{tikzcd}$$

Theorem 32: Composition of linear transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations with standard matrices A and B, respectively. Then $S \circ T$ is a linear transformation with standard matrix BA.

Proof. Since A and B are the standard matrices for T and S, respectively, then for any $\mathbf{x} \in \mathbb{R}^n$,

$$(S \circ T)(\mathbf{x}) = S(T(\mathbf{x})) = S(A\mathbf{x}) = B(A\mathbf{x}) = BA(\mathbf{x}).$$

Hence, $S \circ T$ is a matrix transformation (with standard matrix BA), and so it is linear. (Note that BA will be a $p \times n$ matrix.) $\square$

Composition of linear transformations is *associative*. One can prove this in general for functions, but now it follows because matrix multiplication is associative.

Example. Let T be the reflection across the y-axis in $\mathbb{R}^2$, and let S be a rotation by $90°$ counterclockwise in $\mathbb{R}^2$. What are the standard matrices for $S \circ T$ and $T \circ S$?

The standard matrix of T is $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ and the standard matrix for S is $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Hence,

$$[S \circ T] = BA = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad [T \circ S] = AB = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Geometrically, $S \circ T$ is a reflection across $y = -x$, while $T \circ S$ is a reflection across $y = x$.

An *identity transformation* is a map $I : \mathbb{R}^n \to \mathbb{R}^n$ given by $I(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. It should be clear that this map is linear and its standard matrix is the $n \times n$ identity matrix.

Definition: Invertible linear transformation

A transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is *invertible* if there exists a transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S \circ T = I$ and $T \circ S = I$.

Just as with matrices, we say S is an *inverse* of T. Again, as with matrices, the inverse is unique and so we typically denote it as T^{-1}.

Theorem 33: Standard matrix of the inverse

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation with standard matrix A. Then $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with standard matrix A^{-1}.

Proof. Let $S : \mathbb{R}^n \to \mathbb{R}^n$ be given by $S(\mathbf{x}) = A\mathbf{x}$. Then clearly $S \circ T = I$ and $T \circ S = I$. By uniqueness, $S = T^{-1}$. $\square$

Example. Let S be a rotation by $90°$ counterclockwise in $\mathbb{R}^2$, so it has standard matrix $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Then $B^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, which corresponds to a rotation by $270°$.

7. Applications - Markov Chains

Markov Chains are used in a host of applications, including Google's PageRank algorithm.

Example. Suppose we have the following data from a research study on two toothpastes: Brand A and Brand B. In any given month, of those using Brand A, 70% continue using it the next month and 30% switch to Brand B. Of those using Brand B, 80% continue using it the following month and 20% switch to Brand A.

Suppose initially, 120 people are using Brand A and 80 people are using Brand B. In the next month $(.70)(120) + (.20)(80) = 100$ will be using Brand A, while $(.30)(120) + (.80)(80) = 100$ will be using Brand B. We could represent this using the matrix equation:

$$\begin{bmatrix} .70 & .20 \\ .30 & .80 \end{bmatrix} \begin{bmatrix} 120 \\ 80 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}.$$

Call the matrix on the left P, $\mathbf{x}_0 = \begin{bmatrix} 120 \\ 80 \end{bmatrix}$, and $\mathbf{x}_1 = \begin{bmatrix} 100 \\ 100 \end{bmatrix}$. Now we can compute the numbers for the second month by

$$\mathbf{x}_2 = P\mathbf{x}_1 = \begin{bmatrix} .70 & .20 \\ .30 & .80 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix} = \begin{bmatrix} 90 \\ 110 \end{bmatrix}.$$

The vectors $\mathbf{x}_k$ are called the *state vectors* of the Markov chain and P is the *transition matrix*. The Markov chain satisfies:

$$\mathbf{x}_{k+1} = P\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \ldots.$$

Note that this implies that for any $k = 0, 1, 2, \ldots$, we have

$$\mathbf{x}_k = P^k \mathbf{x}_0.$$

In essence, a *Markov chain* is a probabilistic process where you can reliably make predictions of the future based *only* on the present state (without knowing the full history).

A Markov chain has n states, and for each pair (i, j), there is a probability of moving from state i to state j, p_{ij}. Let $P = [p_{ij}]$. (This might be a bit weird. What it means is that we put all the probabilities of moving out of state i in the ith column. That's what we did in the example above.)

A *probability vector* is a vector $\mathbf{x}$ in $\mathbb{R}^n$ with nonnegative entries whose entries add up to 1. A *stochastic matrix* is a square matrix whose columns are probability vectors. The probability vector keeps track of the state of the Markov chain (the ith entry of $\mathbf{x}$ is the probability of being in state i). The matrix in our Miami example, the transition matrix is a stochastic matrix.

There are many, many examples of Markov chains. Here are a few. (Take STA 427 to learn more about them!).

1. Speech analysis: Voice recognition, keyboard word prediction.

2. Brownian motion: You know the probability of the particle moving to a certain position given its current position.
3. Create randomly generated words or sentences that look meaningful. (The website `http://projects.haykranen.nl/markov/demo/` is quite fun. You can generate random phrases by making the next word be chosen randomly based on the previous 4 words, according to the distribution of words in the source. As a source, you can use the Wikipedia article about Calvin and Hobbes, Alice and Wonderland, or Kant.)
4. Board games: You could consider the Markov chain whose states, say, are the 40 spaces on a Monopoly board. If you know what space you are currently on, then because you know the probabilities of dice rolls and the probabilities of chance cards, you can compute the probability that you move from your current square to any other square. This is why you have heard that in Monopoly you should aim to obtain the orange and light blue properties. In the long run, these squares give a good return on investment, because players are more likely to be on those squares.

How can we determine the long-term behavior in a Markov chain? A *steady-state* vector $\mathbf{q}$ is a probability vector for which $P\mathbf{q} = \mathbf{q}$. (For those that have heard the term, this is the same thing as an eigenvector with eigenvalue 1!) It turns out that any stochastic matrix that is *regular* (that is, there is some $k > 0$ for which all the entries in P^k are positive) has a unique steady-state vector.

Suppose $\mathbf{x}$ is a steady-state vector. Then $P\mathbf{x} = \mathbf{x} = I\mathbf{x}$. Written another way, $(I - P)\mathbf{x} = \mathbf{0}$. Thus, we can find $\mathbf{x}$ by solving the corresponding homogeneous system.

Example. Consider the toothpaste example with transition matrix

$$P = \begin{bmatrix} .70 & .20 \\ .30 & .80 \end{bmatrix}.$$

Then

$$I - P = \begin{bmatrix} .30 & -.20 \\ -.30 & .20 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -2/3 \\ 0 & 0 \end{bmatrix}.$$

Setting our free variable $x_2 = t$, then we have the solution set

$$\mathbf{x} = \begin{bmatrix} 2/3 \\ 1 \end{bmatrix} t.$$

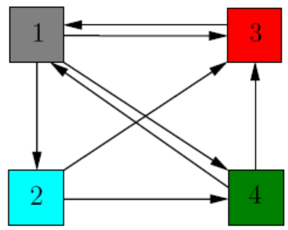
Since we require $\mathbf{x}$ to be a probability vector. Then we have

$$1 = x_1 + x_2 = \frac{2}{3}t + t = \frac{5}{3}t.$$

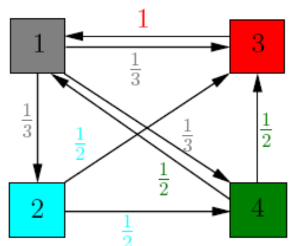
So, $t = \frac{3}{5}$. This gives the vector $\begin{bmatrix} .4 \\ .6 \end{bmatrix}$. This means that, in the long run, 40% will be using Brand A and 60% will be using Brand B.

The vector $\mathbf{x}$ is called an *eigenvector* (corresponding to the *eigenvalue* 1). We will learn about eigenvectors and eigenvalues soon.

Example (PageRank)**.** Suppose we are given by the following directed graph:



We think of the nodes as web pages all related to the phrase "Miami University". We put an arrow from i to j if i links to j. In our model, each page should transfer its importance evenly to the pages that it links to. Node 1 has 3 outgoing edges, so it will pass on 1/3 of its importance to each of the other 3 nodes. Node 3 has only one outgoing edge, so it will pass on all of its importance to node 1. In general, if a node has k outgoing edges, it will pass on $1/k$ of its importance to each of the nodes that it links to. Let us better visualize the process by assigning weights to each edge.



Let us denote by A the transition matrix of the graph,

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

Notice that this is a stochastic matrix! What does it represent? We can think of a *random surfer*, who starts at some web page, and then clicks some link on the web page at random. So if you start at page 1, the surfer has a 1/3 probability of landing at pages 2, 3, or 4. For our random surfer, this vector gives the probability that he will be at each web page (after surfing randomly for a long time). This vector is our PageRank vector[1]

To find the steady-state vector, we solve $(I - A)\mathbf{x} = \mathbf{0}$. The solution space is

$$\text{span}\left\{ \begin{bmatrix} 12 \\ 4 \\ 9 \\ 6 \end{bmatrix} \right\}$$

[1]Almost, there are some problems that arise which we do not address here.

Of course this isn't the steady-state vector yet because it is not a probability vector. We have to normalize so that we choose the eigenvector whose entries sum to 1. The PageRank vector is

$$\begin{bmatrix} 12/31 \\ 4/31 \\ 9/31 \\ 6/31 \end{bmatrix} \approx \begin{bmatrix} .38 \\ .12 \\ .29 \\ .19 \end{bmatrix}$$

Chapter 4: Eigenstuff

(Last Updated: November 29, 2021)

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed).

1. Introduction to eigenvalues and eigenvectors

We have already seen eigenvalues and eigenvectors in the context of Markov chains. In the language of that section, an eigenvector is a steady-state vector (for the eigenvalue 1). In this chapter we study this phenomenon in more detail and see some of its consequences.

Definition: Eigenvector, eigenvalue

An *eigenvector* of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an *eigenvalue* of A if there is a nontrivial solution of $A\mathbf{x} = \lambda\mathbf{x}$.

We say $\mathbf{x}$ is the eigenvector corresponding to λ.

Example. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Observe that $A\mathbf{u} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\mathbf{u}$, so $\mathbf{u}$ is an eigenvector corresponding to the eigenvalue -4 for the matrix A. Geometrically, we can interpret this as A stretching the vector $\mathbf{u}$.

We have already seen the process of finding eigenvectors for an eigenvalue of 1, so the next example should seem familar.

Example. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$. We will show that 7 is an eigenvalue of A.

If 7 is an eigenvalue, then there is some eigenvector $\mathbf{x}$ such that $A\mathbf{x} = 7\mathbf{x}$. This is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or $(A - 7I)\mathbf{x} = \mathbf{0}$. That is, $\mathbf{x}$ is a solution to the homogeneous matrix equation for the matrix $A - 7I$. Hence, we row reduce:

$$A - 7I = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}.$$

The general solution is of the form $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_2$. Hence, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector for A.

We can check this:

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Note that any vector in the null space of $A - 7I$ will also be an eigenvector.

In general, λ is an eigenvalue of A if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. The solution set is the null space the matrix $A - \lambda I$.

Definition: Eigenspace

For an eigenvalue λ of an $n \times n$ matrix A, the *eigenspace* E_λ of A corresponding to λ is the null space of the matrix $A - \lambda I$.

Any eigenspace of a matrix A is automatically a subspace of $\mathbb{R}^n$ because null spaces are subspaces.

Example. One eigenvalue of $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$ is $\lambda = 2$. We will find a basis for E_2.

We need a basis for the null space of $A - 2I$ so we set up the corresponding matrix and row reduce:

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

Hence, a basis of E_2 is

$$\left\{ \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

In this case, E_2 is 2-dimensional.

Suppose λ is an eigenvalue of an $n \times n$ matrix A, then $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero $\mathbf{x}$. Said another way, the matrix equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. That is, $(A - \lambda I)$ is **not** invertible. In the 2×2 case, we know how to check this using the determinant. That is, we simply need to find which values λ give $\det(A - \lambda I) = 0$.

Example. Let $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$. Then

$$\det(A - \lambda I) = \det \begin{bmatrix} 2 - \lambda & 3 \\ 3 & -6 - \lambda \end{bmatrix} = (2 - \lambda)(-6 - \lambda) - 9 = \lambda^2 + 4\lambda - 21 = (\lambda + 7)(\lambda - 3).$$

Thus, $\det(A - \lambda I) = 0$ if $\lambda = 3$ or $\lambda = -7$. So, the eigenvalues of A are $3, -7$.

Now we compute a basis of each eigenspace:

$$A - 3I = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix} \qquad \text{basis of } E_3 : \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$$

$$A + 7I = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1/3 \\ 0 & 0 \end{bmatrix} \qquad \text{basis of } E_{-7} : \left\{ \begin{bmatrix} -1/3 \\ 1 \end{bmatrix} \right\}$$

Hence, each eigenspace is 1-dimensional.

2. Determinants

In the last section we used determinants to find the eigenvalues of a 2×2 matrix. To extend this to $n\times n$ matrices, we need to first study determinants of larger (square) matrices. Another motivation comes from invertibility. A 2×2 matrix is invertible if and only if its determinant is nonzero. In this chapter we'll develop a similar criteria for $n\times n$ matrices.

For an $n\times n$ matrix $A=(a_{ij})$, we denote by A_{ij} the submatrix obtained by deleting the ith row and jth column. The value $\det A_{ij}$ is called the (i,j)-*minor* of A.

Example. Let $A=\begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then $A_{23}=\begin{bmatrix} 2 & -4 \\ 1 & 4 \end{bmatrix}$ and $A_{33}=\begin{bmatrix} 2 & -4 \\ 3 & 1 \end{bmatrix}$.
The $(2,3)$-minor of A is $\det A_{23}=2(4)-(-4)(1)=20$.

The definition of the determinant (that we use) is *recursive*. That means, in order to compute the determinant of an $n\times n$ matrix we need to know how to compute the determinant of an $(n-1)\times(n-1)$ matrix. This is ok because we already know how to compute the determinant of a 2×2 matrix.

For a 3×3 matrix A, the determinant has the following form:

$$\det(A)=a_{11}(a_{22}a_{33}-a_{23}a_{32})-a_{12}(a_{21}a_{12}-a_{23}a_{31})+a_{13}(a_{21}a_{32}-a_{22}a_{31}).$$

This is difficult to remember, but not so much when we recognize that it can be written in terms of submatrices:

$$\det(A)=a_{11}\det(A_{11})-a_{12}\det(A_{12})+a_{13}\det(A_{13}).$$

Example. Let $A=\begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then

$$A_{11}=\begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix}, \det(A_{11})=-9, \quad A_{12}=\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}, \det(A_{12})=-5 \quad A_{13}=\begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}, \det(A_{13})=11.$$

Hence,

$$\begin{aligned} \det(A) &= a_{11}\det(A_{11})-a_{12}\det(A_{12})+a_{13}\det(A_{13}) \\ &= 2(-9)-(-4)(-5)+(3)(11)=-5. \end{aligned}$$

Another method for computing this is known as the *butterfly method*[1].

[1]Not to be confused with the *The Butterfly Effect*, the grand masterpiece of Ashton Kutcher's acting career.

Definition 1. For $n \geq 2$, the *determinant* of an $n \times n$ matrix $A = (a_{ij})$ is the sum of n terms of the form $\pm a_{1j} \det(A_{1j})$ with alternating $\pm$ signs:

$$\begin{aligned}\det(A) = |A| &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{n+1}a_{1n}\det(A_{1n}) \\ &= \sum_{j=1}^{n}(-1)^{j+1}a_{1j}\det(A_{1j}).\end{aligned}$$

The definition of the determinant we gave is one of many (equivalent) definitions. In particular, the choice of using the first row is completely arbitrary.

The (i,j)-cofactor is $C_{ij} = (-1)^{i+j}\det(A_{ij})$. Our earlier formula is then

$$\det(A) = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

Theorem 1: Laplace (cofactor) Expansion Theorem

The determinant of any $n \times n$ matrix can be determined by cofactor expansion along any row or column. In particular,

(ith row) $$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}$$

(jth col) $$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}C_{ij}.$$

Example. Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Expanding on the second column gives

$$\begin{aligned}\det(A) &= (-1)^{1+2}a_{12}\det(A_{12}) + (-1)^{2+2}a_{22}\det(A_{22}) + (-1)^{3+2}a_{32}\det(A_{32}) \\ &= -(-4)(-5) + (1)(-5) - 4(-5) = -5\end{aligned}$$

Certain types of matrices have easy-to-compute determinants.

Example. Let $A = \begin{bmatrix} 1 & -3 & 7 & 15 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$. Expanding on the first column gives

$$\det(A) = (-1)^{1+1}(1)\det(A_{11}) = \det\begin{bmatrix} -2 & 5 & 3 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix}.$$

Now we expand again on the first column to get

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

Using the formula for 2×2 matrices we have

$$\det A = (1)(-2)(3)(4) = -24.$$

That is, the determinant is just the product of the diagonal entries of the original matrix.

The next result is proved using the notion of *mathematical induction*, which you will study more in MTH331

Theorem 2: Determinants of triangular matrices

If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.

For larger matrices, we will want to use row reduction to more efficiently compute determinants. The next theorem tells us how row reduction changes the determinant of a matrix.

Theorem 3: Properties of determinants

Let A be a square matrix.

a. If B is obtained by interchanging two rows of A, then $\det B = -\det A$.
b. If B is obtained by multiplying a row of A by k, then $\det B = k \det A$.
c. If B is obtained by adding a multiple of one row of A to another, then $\det B = \det A$.

Note that an immediate consequence of property (b) is that for an $n \times n$ matrix A:

$$\det(kA) = k^n A.$$

These rules can be proved using elementary matrices or going through the recursive definition. Before we go through this, we will look at an example.

Example. Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. We will compute the determinant of A using row reduction.

$$\begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 4 & -1 \\ 3 & 1 & 2 \\ 2 & -4 & 3 \end{bmatrix} \xrightarrow[R_2+(-3)R_1]{R_3+(-2)R_1} \begin{bmatrix} 1 & 4 & -1 \\ 0 & -11 & 5 \\ 0 & -12 & 5 \end{bmatrix} \xrightarrow{R_3+(-\frac{12}{11})R_2} \begin{bmatrix} 1 & 4 & -1 \\ 0 & -11 & 5 \\ 0 & 0 & -5/11 \end{bmatrix}$$

This last matrix is upper triangular, so its determinant is just the product of the diagonal entries. Hence, the determinant of the last matrix is 5. From our original matrix, we use row addition (which doesn't change the determinant) and we interchanged once (which scales the determinant by -1). Hence, $\det(A) = (-1)(5) = -5$.

A special case of the previous theorem replaces B with an elementary matrix E and A with the identity matrix I_n. Using this, we can show that

$$\det(EB) = \det(E)\det(B).$$

This leads to one of our big theorems regarding determinants.

Theorem 4: Properties of determinants

Let A and B be $n \times n$ matrices.

a. $\det(AB) = \det(A)\det(B)$
b. $\det(A) = \det(A^T)$
c. The matrix A is invertible if and only if $\det A \neq 0$.
d. If A is invertible, then $\det(A^{-1}) = \frac{1}{\det A}$.

Proof. (a) If one of A or B is not invertible, then AB is not invertible. Hence, both sides of the equation yield 0. Now suppose both A and B are invertible. Then simply expand A and B according to their elementary matrix decomposition and apply the above rule.

(b) The follows by noting that expansion along the first row of A is the same as expanding on the first column of A^T.

(c) Let $R = \text{RREF}(A)$. Then there is a sequence of elementary matrices $E_1, \dots, E_r$ such that $E_r \cdots E_1 A = R$. Hence, $\det(A) \neq 0$ if and only if $\det(R) \neq 0$. If A is invertible, then $R = I_n$ so $\det(A) \neq 0$. On the other hand, if A is not invertible, then R contains a zero row so $\det(A) = 0$.

(d) Applying part (a): $1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. □

Now we study Cramer's rule, which has many applications but is not suitable for large computations. For a $n \times n$ matrix A and any vector $\mathbf{b}$ in $\mathbb{R}^n$, we obtain $A_i(\mathbf{b})$ by replacing column i in A with $\mathbf{b}$.

Theorem 5: Cramer's Rule

Let A be an invertible $n \times n$ matrix and let $\mathbf{b}$ be a vector in $\mathbb{R}^n$. Then the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \quad \text{for } i = 1, \dots, n.$$

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. If $A\mathbf{x} = \mathbf{b}$, then

$$\begin{aligned} A \cdot I_i(\mathbf{x}) = A\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}). \end{aligned}$$

Now $\det(A \cdot I_i(\mathbf{x})) = \det A_i(\mathbf{b})$ so

$$\det(A)\det I_i(\mathbf{x}) = \det A_i(\mathbf{b}).$$

But $\det(I_i(\mathbf{x})) = x_i$, so the above equation reduces to $\det(A)x_i = \det A_i(\mathbf{b})$. The result now follows. $\square$

Example. We will use Cramer's Rule to solve the following system:

$$3x_1 + 4x_2 = 3$$
$$5x_1 + 6x_2 = 7.$$

We have $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$ so $\det(A) = -2$. Since $\mathbf{b} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$, then

$$A_1(\mathbf{b}) = \begin{bmatrix} 3 & 4 \\ 7 & 6 \end{bmatrix} \qquad \det(A_1(\mathbf{b})) = -10 \qquad x_1 = \frac{\det(A_1(\mathbf{b}))}{\det(A)} = \frac{-10}{-2} = 5$$

$$A_2(\mathbf{b}) = \begin{bmatrix} 3 & 3 \\ 5 & 7 \end{bmatrix} \qquad \det(A_2(\mathbf{b})) = 6 \qquad x_2 = \frac{\det(A_2(\mathbf{b}))}{\det(A)} = \frac{6}{-2} = -3.$$

Let A be an $n \times n$ invertible matrix. Cramer's Rule also leads to a formula for the inverse of A.

The jth column of A^{-1} is the unique solution to $A\mathbf{x} = \mathbf{e}_j$. By Cramer's Rule, the ith entry of $\mathbf{x}$ is $x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A}$. Cofactor expansion down column i of $A_i(e_j)$ shows that

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji}.$$

Definition: Adjoint matrix

The *adjoint* (or *adjugate*) of A, denoted $\operatorname{adj} A$, is the $n \times n$ matrix with (i, j)-entry C_{ji}.

Theorem 6: Cramer's Inverse Formula

Let A be an invertible $n \times n$ matrix. Then

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

The next example illustrates why this isn't a good method for computing the inverse matrix.

Example. We will use Cramer's Inverse Formula to find the inverse of $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$.

$$C_{11} = (-1)^{1+1} \det(A_{11}) = -9 \qquad C_{12} = (-1)^{1+2} \det(A_{12}) = 5 \qquad C_{13} = (-1)^{1+3} \det(A_{13}) = 11$$
$$C_{21} = (-1)^{2+1} \det(A_{21}) = 8 \qquad C_{22} = (-1)^{2+2} \det(A_{22}) = -5 \qquad C_{23} = (-1)^{2+3} \det(A_{23}) = -12$$
$$C_{31} = (-1)^{3+1} \det(A_{11}) = -11 \qquad C_{32} = (-1)^{3+2} \det(A_{12}) = 5 \qquad C_{33} = (-1)^{3+3} \det(A_{13}) = 14$$

Hence,

$$\operatorname{adj} A = \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} -9 & 8 & -11 \\ 5 & -5 & 5 \\ 11 & -12 & 14 \end{bmatrix}.$$

Therefore,

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A = -\frac{1}{5} \begin{bmatrix} -9 & 8 & -11 \\ 5 & -5 & 5 \\ 11 & -12 & 14 \end{bmatrix}$$

3. Eigenvalues and Eigenvectors for nxn matrices

We now have the tools to study eigenstuff for larger matrices.

Definition: Characteristic polynomial

For an $n \times n$ matrix A, the *characteristic polynomial* of A is $\det(A - \lambda I)$. The *characteristic equation* is $\det(A - \lambda I) = 0$.

As we saw in the first section, the roots of the characteristic equation are the eigenvalues of A.

Example. Find the eigenvalues and corresponding eigenspaces for the following matrix:

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

First we find the characteristic polynomial:

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} 1-\lambda & 0 & 2 \\ 3 & -1-\lambda & 3 \\ 2 & 0 & 1-\lambda \end{bmatrix} \\ &= (1-\lambda)\det \begin{bmatrix} -1-\lambda & 3 \\ 0 & 1-\lambda \end{bmatrix} + (2)\det \begin{bmatrix} 3 & -1-\lambda \\ 2 & 0 \end{bmatrix} \\ &= (1-\lambda)((-1-\lambda)(1-\lambda) - 0) + 2(0 - (-1-\lambda)(2)) \\ &= (1-\lambda)(-1-\lambda)(1-\lambda) - 4(-1-\lambda) \\ &= (-1-\lambda)(\lambda^2 - 2\lambda - 3) = -(1+\lambda)(\lambda-3)(\lambda+1). \end{aligned}$$

Hence, the eigenvalues are $\lambda = -1, 3$. Now we find a basis of each eigenspace:

$$E_{-1} : A + I = \begin{bmatrix} 2 & 0 & 2 \\ 3 & 0 & 3 \\ 2 & 0 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{basis of } E_{-1}\text{: } \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$E_3 : A - 3I = \begin{bmatrix} -2 & 0 & 2 \\ 3 & -3 & 3 \\ 2 & 0 & -2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -3/2 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{basis of } E_3\text{: } \left\{ \begin{bmatrix} 1 \\ 3/2 \\ 1 \end{bmatrix} \right\}.$$

The next definitions are important in the study of diagonalization in the next section.

Definition: Algebraic multiplicity, geometric multiplicity

Let A be a square matrix with eigenvalue λ. The *algebraic multiplicity* of λ is the multiplicity of λ as a root of the characteristic polynomial. The *geometric multiplicity* is the dimension of the eigenspace E_λ.

Example. In the previous example, -1 has algebraic and geometric multiplicity 2, while 3 has algebraic and geometric multiplicity 1.

The following result comes directly from our theorem on determinants of triangular matrices.

Theorem 7: Eigenvalues of triangular matrices

The eigenvalues of a triangular matrix are the entries along the main diagonal.

Proof. If A is triangular, then so is $A - \lambda I$ and hence

$$\det(A - tI) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

Thus, the roots of the characteristic equation are the diagonal entries of A. $\square$

Example. Let $A = \begin{bmatrix} 2 & 1 & 7 & 0 \\ 0 & 5 & 2 & 4 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$. Then

$$\det(A - \lambda I) = (2 - \lambda)(5 - \lambda)(-1 - \lambda)(2 - \lambda).$$

Thus, the eigenvalues are $2, 5, -1, 2$, which are exactly the diagonal entries of the matrix.

When a matrix has an eigenvalue of 0, then the null space of $A - 0I = A$ is nontrivial. Hence, A is not invertible. On the other hand, if A is not invertible, then (A) is nontrivial. This proves the following:

Theorem 8: Invertibility and eigenvalue 0

A matrix A is invertible if and only if 0 is not an eigenvalue of A.

If we know an eigenvalue of a matrix A, then we can determine eigenvalues of several related matrices.

Theorem 9: Eigenvalues of related matrices

Let A be a square matrix with eigenvalue λ.

a. For any positive integer n, λ^n is an eigenvector of A^n.
b. If A is invertible, then $1/\lambda$ is an eigenvalue of A^{-1}.
c. If A is invertible, then for any integer n, λ^n is an eigenvector of A^n.

Let A be a square matrix with eigenvectors $\mathbf{x}$ and $\mathbf{y}$ for distinct nonzero eigenvalues λ and μ, respectively. Then $\mathbf{x}$ and $\mathbf{y}$ are linearly independent. To see this, suppose $a\mathbf{x} + b\mathbf{y} = \mathbf{0}$. Multiplying both by A gives $a\lambda\mathbf{x} + b\mu\mathbf{y} = \mathbf{0}$. If $a \neq 0$, then

$$\mathbf{x} = -\frac{b\mu}{a\lambda}\mathbf{y}.$$

But this implies that

$$\lambda \mathbf{x} = A\mathbf{x} = A(-\frac{b\mu}{a\lambda}\mathbf{y}) = \mu(-\frac{b\mu}{a\lambda}\mathbf{y}) = \mu\mathbf{x},$$

so $\lambda = \mu$. A similar argument applies if $b \neq 0$. We conclude that $a = b = 0$. This argument extends to larger sets of vectors.

Theorem 10: Eigenvectors are linearly independent

Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly independent.

Proof. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent. Because $\mathbf{v}_1 \neq \mathbf{0}$, then there exists $p \in \{2, \ldots, m\}$ such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent and $\mathbf{v}_{p+1}$ is a linear combination of those vectors. That is, there exist $c_1, \ldots, c_p$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1}.$$

Multiplying both sides by A gives

$$\begin{aligned}
c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\
c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}\mathbf{v}_{p+1} \\
c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \\
c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p &= \mathbf{0}.
\end{aligned}$$

This contradicts the linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ because none of the $\lambda_i - \lambda_{p+1}$ are zero by hypothesis. $\square$

4. Similarity and diagonalization

Definition: Similar matrices

Let A and B be $n \times n$ matrices. Then A is said to be *similar* to B if there exists an invertible matrix P such that $B = P^{-1}AP$.

When A is similar to B we write $A \sim B$.

Similarity is an example of an *equivalence relation*. That means it is

- Reflexive: A is similar to itself
$$A = I^{-1}AI$$
- Symmetric: If A is similar to B, then B is similar to A
$$B = P^{-1}AP \Rightarrow A = PBP^{-1}$$
- Transitive: If A is similar to B and B is similar to C, then A is similar to C
$$A = P^{-1}BP \text{ and } B = Q^{-1}CQ \Rightarrow A = P^{-1}(Q^{-1}CQ)P = (QP)^{-1}CQP$$

Example. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}$. Take $P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, then one can check that $P^{-1}AP = B$ or, more simply $AP = PA$.

Where the matrix P comes from in the previous example is not at all obvious and is not the primary goal of this section. We will be interested in determining when a matrix is similar to a diagonal matrix, and determining that matrix.

Similar matrices have many of the same properties.

Theorem 11: Properties of similar matrices

Let A and B be $n \times n$ matrices with $A \sim B$. Then

a. $\det A = \det B$
b. A is invertible if and only if B is invertible.
c. A and B have the same rank.
d. A and B have the same characteristic polynomial.
e. A and B have the same eigenvalues.
f. $A^m \sim B^m$ for all integers $m \geq 0$.
g. If A is invertible, then $A^m \sim B^m$ for all integers m.

Proof. Since $A \sim B$, then there is some invertible matrix P such that $B = P^{-1}AP$.

(a) We have
$$\det(B) = \det(P^{-1}AP) = \det(P)^{-1}\det(A)\det(P) = \det(A).$$

(d) Since

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(A - \lambda I)P,$$

then by part (a), $\det(B - \lambda I) = \det(A - \lambda I)$.

The remaining properties are easy to check. $\square$

Definition: Diagonalizable

An $n \times n$ matrix A is *diagonalizable* if there is a diagonal matrix D such that $A \sim D$.

To *diagonalize* a matrix A is to find an invertible matrix P and a diagonal matrix D such that $D = P^{-1}AP$ (equivalently, $A = PDP^{-1}$).

We will see how to use eigenvalues and eigenvectors to diagonalize matrices. The following example shows one advantage of this process:

Example. Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Then $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$. Then

$$A^3 = (PDP^{-1})(PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)D(P^{-1}P)DP^{-1} = PD^3P^{-1}.$$

Note that $D^3 = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$. So

$$A^3 = PD^3P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 223 & 98 \\ -196 & -71 \end{bmatrix}.$$

Theorem 12: Criteria for diagonalizability

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the (linearly independent) eigenvectors of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Set $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and D to be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$. By IMT, P is invertible and

$$AP = A\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} = PD.$$

Hence, $D = P^{-1}AD$ so A is diagonalizable.

Conversely, suppose A is diagonalizable. Then there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. Write $P = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$ and let $c_1, \ldots, c_n$ be the diagonal entries of D. As above, $AP = PD$ implies $A\mathbf{x}_i = c_i\mathbf{x}_i$ for each i. But then each $\mathbf{x}_i$ is an eigenvector for A. Because P is invertible, the vectors $\mathbf{x}_i$ are linearly independent. $\square$

Under the hypotheses of the theorem, the eigenvectors of A form a basis of $\mathbb{R}^n$ called an *eigenbasis* of $\mathbb{R}^n$ with respect to A.

Using the theorem, we have the following steps diagonalize an $n \times n$ matrix A.

a. Find the eigenvalues of A using the characteristic polynomial.
b. Find an eigenbasis for each eigenvalue.
c. Construct P with columns from eigenvectors (assuming there are n linearly independent eigenvalues).
d. Construct D from eigenvalues in order corresponding to P.
e. Check!

If A has n distinct eigenvalues, then each eigenspace has dimension one and so we have an eigenbasis. However, this is not the only situation.

Example. Let $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$. Then

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix} = (1-\lambda)(2+\lambda)^2.$$

Hence, the eigenvalues for this matrix are 1 and -2. We compute a basis for each eigenspace:

$$A - (1)I = \begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\},$$

$$A - (-2)I = \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thus, A is diagonalizable by the matrices

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

Example. Consider the matrix $A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$.

The characteristic polynomial of A is $(\lambda+2)^2(\lambda-1)$, so the eigenvalues are 1 and -2. However, the two eigenspaces E_{-2} and E_1 both have dimension one, so A does not have three linearly independent eigenvectors. Hence, A is not diagonalizable.

We can summarize our examples in the following theorem.

Theorem 13: The Diagonalization Theorem

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

a. A is diagonalizable.

b. The union of the bases of the eigenspaces of A contain n vectors.

c. The algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

We will discuss a few applications of eigenstuff and determinants.

The cross product and area. Those that have taken some Calc III have undoutably seen another product operation on vectors: the cross product. This has an important geometric interpretation.

Definition: Cross product

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$. The *cross product* of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

Note that the output of the cross product is a *vector*, as opposed to the dot product whose output is a *scalar*. The cross product formula is a little clunky and hard to remember. But the terms should look familar. In fact, it's much easier to compute by viewing this as a determinant-like object:

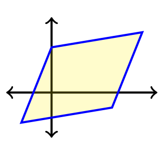
$$\mathbf{u} \times \mathbf{v} = \det \begin{bmatrix} \mathbf{e}_1 & u_1 & v_1 \\ \mathbf{e}_2 & u_2 & v_2 \\ \mathbf{e}_3 & u_3 & v_3 \end{bmatrix} = (u_2v_3 - u_3v_2)\mathbf{e}_1 - (u_1v_3 - u_3v_1)\mathbf{e}_2 + (u_1v_2 - u_2v_1)\mathbf{e}_3.$$

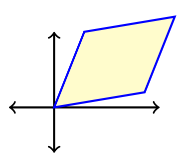
Area of a parallelogram

If A is a 2×2 matrix, then $|\det A|$ is the area of the parallelogram determined by the columns of A.

The key point to the above theorem is that if $\mathbf{a}_1, \mathbf{a}_2 \neq \mathbf{0}$, then for any scalar c the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ is equal to that of $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

Example. Consider the parallelogram determined by the points $(-2,-2), (0,3), (4,-1), (6,4)$. We first translate the parallelogram by adding $(2,2)$ to each vertex.





Then the points are $(0,0), (2,5), (6,1), (8,6)$. This is the parallelogram determined by the vectors through the points $(2,5)$ and $(6,1)$. Thus,

$$\text{Area } = \begin{vmatrix} 2 & 6 \\ 5 & 1 \end{vmatrix} = |-28| = 28.$$

Volume of a parallelepiped

If A is a 3×3 matrix, then $|\det A|$ is the volume of the parallelepiped determined by the columns of A.

Differential equations.

Example. Let $x(t)$ be a differentiable function of t. We can use calculus to find all solutions to the differential equation

$$x'(t) = 2x(t).$$

Write $x = x(t)$. By separation of variables, we have

$$\begin{aligned} \frac{dx}{dt} &= 2x \\ \frac{dx}{x} &= 2dt \\ \ln x &= 2t + C. \end{aligned}$$

Hence, the general solution is $x = ce^{2t}$ for any constant c.

A *system of linear first-order differential equations* (DEs) is a set of equations

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

where the x_i are all differentiable functions in t and the a_{ij} are all constants.

We are using shorthand here with x_i in place of $x_i(t)$ and x_i' for the derivative of x_i (with respect to t). We can represent the entire system in matrix form as

$$\mathbf{x}'(t) = A\mathbf{x}(t),$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \qquad A = (a_{ij}).$$

A *solution* to the system is a vector-valued function $\mathbf{x}(t)$ that satisfies $\mathbf{x}'(t) = A\mathbf{x}(t)$. An *initial value problem* is a system of DEs along with an initial condition $\mathbf{x}_0 = \mathbf{x}(0)$.

Note that $\mathbf{x}'(t) = A\mathbf{x}(t)$ is linear. If $c, d \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v}$ are solutions, then

$$(c\mathbf{u} + d\mathbf{v})' = c\mathbf{u}' + d\mathbf{v}' = c(A\mathbf{u}) + d(A\mathbf{v}) = A(c\mathbf{u} + d\mathbf{v}).$$

Moreover, $\mathbf{0}$ is (trivially) a solution and so the set of solutions is a subspace of $\mathbb{R}^n$.

Example. Let $\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ and $A = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$.

It follows that our solutions are $x_1 = c_1 e^{3t}$ and $x_2 = c_2 e^{-5t}$. Thus, the solution space is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-5t} \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{-5t}.$$

This example suggests that solutions to a system of linear equations might have a simple presentation as a linear combination of functions of the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$.

Suppose a solution to $\mathbf{x}'(t) = A\mathbf{x}(t)$ has the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$. By calculus, $\mathbf{x}'(t) = \lambda \mathbf{v} e^{\lambda t}$. But then $A\mathbf{x} = A\mathbf{v}e^{\lambda t}$ (this uses the fact that $e^{\lambda t} \neq 0$ for all t). Hence, $A\mathbf{v} = \lambda \mathbf{v}$, that is, $\mathbf{v}$ is an eigenvector with eigenvalue λ. Solutions of this form are called *eigenfunctions*.

Example. We will solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ given

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \qquad A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}.$$

The eigenvalues of A are $-.5, -2$ with corresponding eigenvectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Thus, the eigenfunctions are

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} \quad \text{and} \quad \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t},$$

and any solution is a linear combination of these two. That is,

$$\mathbf{x} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}.$$

We can use the initial conditions to solve for c_1, c_2. So,

$$\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \mathbf{x}_0 = \mathbf{x}(0) = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

To solve for c_1, c_2 we solve the vector equation by row reducing the corresponding augmented matrix,

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

Hence, the solution to the given IVP is,

$$\mathbf{x} = 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} - 2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

Discrete dynamical systems.

Example. Wood rats are the primary source of food for spotted owls in the Redwood forest. Denote the population of owls and wood rats at time k by

$$\mathbf{x}_k = \begin{bmatrix} O_k \\ R_k \end{bmatrix},$$

where k is the time in months, O_k is the number of owls in the region, and R_k is the number of wood rats in the region (in thousands).

The populations at time $k+1$ are related to the populations at time k by the *dynamical system*,

$$\begin{aligned} O_{k+1} &= (.5)O_k + (.4)R_k \\ R_{k+1} &= -p \cdot O_k + (1.1)R_k, \end{aligned}$$

where $1000p$ denotes the average number of rats eaten by one owl in a month. Note that we can denote this system by

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{where} \quad A = \begin{bmatrix} .5 & .4 \\ -p & 1.1 \end{bmatrix}.$$

We will study the long-term behavior (or *evolution*) of this system.

Let $p = .104$. Then the eigenvalues of A are $\lambda_1 = 1.02$ and $\lambda_2 = .58$ with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 10 \\ 13 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathbb{R}^2$, then any initial vector $\mathbf{x}_0$ can be written as $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ for scalars c_1, c_2. Then,

$$\mathbf{x}_1 = A\mathbf{x}_0 = A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1(A\mathbf{v}_1) + c_2(A\mathbf{v}_2) = c_1(1.02)\mathbf{v}_1 + c_2(.58)\mathbf{v}_2.$$

In general,

$$\mathbf{x}_k = c_1(1.02)^k\mathbf{v}_1 + c_2(.58)^k\mathbf{v}_2.$$

Assume $c_1 > 0$. As $k \to \infty$, $(.58)^k \to 0$. Hence, for sufficiently large k we have

$$\mathbf{x}_k \approx c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix}.$$

As k increases, the approximation gets better. So for large k,

$$\mathbf{x}_{k+1} \approx c_1(1.02)^{k+1} \begin{bmatrix} 10 \\ 13 \end{bmatrix} = (1.02)c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix} = (1.02)\mathbf{x}_k.$$

This says both populations grow by a factor of about 1.02 every month, but the ratio of critters (10 owls to 13 thousand rats) stays the same.

In general, assume a dynamical system is described by the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ and that A is diagonalizable with n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues and arrange them so that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$, any initial vector $\mathbf{x}_0$ can be written (uniquely) as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

As in our example, we have

$$\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n.$$

In general we will study the behavior as $k \to \infty$.

Chapter 5: Orthogonality

(Last Updated: November 9, 2021)

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed).

1. Orthogonality in $\mathbb{R}^n$

We previously saw the idea of orthogonality in the study of vector geometry. Here we will discuss the idea more thoroughly. Recall that the inner product of $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n.$$

Alternatively, $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$ using regular matrix multiplication.

The following properties are put here for your convenience. At this point, you should be able to justify these facts.

Theorem 1: Algebraic properties of inner product

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then we have the following:

a. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
b. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$.
c. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$.
d. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Recall that two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are said to be *orthogonal* (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$. We now extend this definition to sets.

Definition: Orthogonal set, orthonormal set

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called an *orthogonal set* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$, $i, j = 1 \ldots, k$. If, in addition, each $\mathbf{u}_i$ is a unit vector, then the set is said to be *orthonormal*.

Example. Consider the vectors

$$\mathbf{u} = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}.$$

Then $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$, so the set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is orthogonal. The set is *not* orthonormal because the vectors are not unit vectors. However, we could replace each one by its associated unit vector to obtain an orthonormal set with the same span.

Orthogonal sets (of nonzero vectors) can easily be shown to be linearly independent.

Theorem 2: Orthogonal sets are linearly independent

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then the set is linearly independent.

Proof. Write $\mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. Then

$$0 = \mathbf{0} \cdot \mathbf{u}_1 = (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_1 = c_1(\mathbf{v}_1 \cdot \mathbf{v}_1) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_1) = c_1(\mathbf{v}_1 \cdot \mathbf{v}_1).$$

Since $\mathbf{v}_1 \cdot \mathbf{v}_1 \neq 0$ (because $\mathbf{v}_1 \neq 0$), then $c_1 = 0$. Repeating this argument with $\mathbf{v}_2, \ldots, \mathbf{v}_k$ gives $c_2 = \cdots = c_k = 0$. Hence, the set is linearly independent. ☐

Definition: Orthogonal basis, orthonormal basis

An *orthogonal basis* for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set. An orthogonal basis is an *orthonormal basis* if the basis is an orthonormal set.

The standard basis is an orthonormal basis of $\mathbb{R}^n$.

Example. We will find an orthogonal basis for the subspace

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - y + 2z = 0 \right\}.$$

This is a plane through the origin and so W is 2-dimensional. One vector in W is $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. It suffices to find another vector $\mathbf{v}$ in W that is orthogonal to $\mathbf{u}$.

Write $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. If $\mathbf{u} \cdot \mathbf{v}$, then we have $a + b = 0$. But $\mathbf{v}$ is in W so $a - b + 2c = 0$. We solve this system through row reduction:

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

This gives $a+c=0$ and $b-c=0$. So one such vector is $\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$. If we take this to be $\mathbf{v}$ then it is easy to check that $\mathbf{v}$ is in W and $\mathbf{u} \cdot \mathbf{v} = 0$. Hence, an orthogonal basis for W is $\{\mathbf{u}, \mathbf{v}\}$.

An orthonormal basis could be obtained by replacing $\mathbf{u}$ and $\mathbf{v}$ by their associated unit vectors. So an orthonormal basis for W is $\{\mathbf{q}_1, \mathbf{q}_2\}$ where

$$\mathbf{q}_1 = \frac{1}{\|u\|}\mathbf{u} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \frac{1}{\|v\|}\mathbf{u} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

The following strategy is suggested by the previous theorem.

Theorem 3: Coordinates for orthogonal basis

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Let $\mathbf{y} \in W$ and write

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$$

where the scalars are given by

$$c_i = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \qquad i = 1, \ldots, k.$$

Note that this theorem simplifies in the case of an orthonormal basis since in this case, $\mathbf{u}_i \cdot \mathbf{u}_i = 1$.

Proof. In a similar way to the previous theorem, we have

$$\mathbf{y} \cdot \mathbf{u}_i = (c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k) \cdot \mathbf{u}_i = c_i(\mathbf{u}_i \cdot \mathbf{u}_i).$$

Since $\mathbf{u}_i \cdot \mathbf{u}_i \neq 0$, then the result follows. □

Example. Define the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}.$$

The vectors in S are orthogonal and hence the set S is linearly independent. Since S consists of 3 linearly independent vectors in $\mathbb{R}^3$, it is a basis for $\mathbb{R}^3$.

Let $\mathbf{x} = \begin{bmatrix} 8 \\ -4 \\ -3 \end{bmatrix}$. Then we can find the coordinates of $\mathbf{x}$ with respect to S using the method above. Denote the vectors in S by $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, respectively. Then

$$c_1 = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} = \frac{5}{2}, \quad c_2 = \frac{\mathbf{x} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} = -\frac{3}{2}, \quad c_3 = \frac{\mathbf{x} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} = 2.$$

Using our previous notation, we have

$$[\mathbf{x}]_S = \begin{bmatrix} 5/2 \\ -3/2 \\ 2 \end{bmatrix}.$$

We now show that matrices with orthonormal columns have very special properties. Such matrices frequently appear in applications.

Theorem 4: Characterization of orthonormal matrices

An $m \times n$ matrix Q has orthonormal columns if and only if $Q^TQ = I$.

Proof. Write $Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$. Then

$$Q^TQ = \begin{bmatrix} \mathbf{q}_1^T \\ \cdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T\mathbf{q}_1 & \mathbf{q}_1^T\mathbf{q}_2 & \cdots & \mathbf{q}_1^T\mathbf{q}_n \\ \mathbf{q}_2^T\mathbf{q}_1 & \mathbf{q}_2^T\mathbf{q}_2 & \cdots & \mathbf{q}_2^T\mathbf{q}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{q}_n^T\mathbf{q}_1 & \mathbf{q}_n^T\mathbf{q}_2 & \cdots & \mathbf{q}_n^T\mathbf{q}_n \end{bmatrix}.$$

Hence, $Q^TQ = I$ if and only if $\mathbf{q}_i \cdot \mathbf{q}_i = 1$ for all i and $\mathbf{q}_i \cdot \mathbf{q}_j = 0$ for all $i \neq j$. $\square$

Square matrices with orthonormal columns are especially important. Not the apparent incongruence in the definition here.

Definition: Orthogonal matrix

An $n \times n$ matrix whose columns form an orthonormal set is called an *orthogonal matrix.*

From the previous theorem it is clear that an orthogonal matrix Q is invertible and

$$Q^{-1} = Q^T.$$

Example. The following matrices are easily seen to be orthogonal:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Hence,

$$A^{-1} = A^T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B^{-1} = B^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Both matrices in the previous example are examples of an *isometry*, a length-preserving transformation in $\mathbb{R}^3$. In fact, this length-preserving property characterizes orthogonal matrices.

Theorem 5: Alternate characterizations of orthogonal matrices

Let Q be an $n \times n$. The following statements are equivalent:

a. Q is orthogonal.

b. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

c. $(Q\mathbf{x}) \cdot (Q\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Proof. (a) $\Rightarrow$ (b) Suppose Q is orthogonal. Write $Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$. Then

$$\begin{aligned} \|Q\mathbf{x}\|^2 = Q\mathbf{x} \cdot Q\mathbf{x} &= (x_1\mathbf{q}_1 + \cdots x_n\mathbf{q}_n) \cdot (x_1\mathbf{q}_1 + \cdots x_n\mathbf{q}_n) \\ &= \sum_{i,j} (x_i\mathbf{q}_i) \cdot (x_j\mathbf{q}_j) = \sum_{i,j} x_i x_j (\mathbf{q}_i \cdot \mathbf{q}_j) = \sum_i x_i^2 (\mathbf{q}_i \cdot \mathbf{q}_i) = \sum_i x_i^2 = \|\mathbf{x}\|^2 . \end{aligned}$$

(b) $\Rightarrow$ (c) Assume property (b). Then

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right) \\ &= \frac{1}{4}\left(\|Q(\mathbf{x}+\mathbf{y})\|^2 - \|Q(\mathbf{x}-\mathbf{y})\|^2\right) \\ &= \frac{1}{4}\left(\|Q\mathbf{x}+Q\mathbf{y}\|^2 - \|Q\mathbf{x}-Q\mathbf{y}\|^2\right) \\ &= Q\mathbf{x} \cdot Q\mathbf{y}. \end{aligned}$$

(c) $\Rightarrow$ (a) Taking standard vectors $\mathbf{e}_i$ and $\mathbf{e}_j$, we have by property (c),

$$\mathbf{q}_i \cdot \mathbf{q}_j = Q\mathbf{e}_i \cdot Q\mathbf{e}_j = \mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Hence, the columns of Q are orthonormal, so Q is an orthogonal matrix. $\square$

We conclude this section with some properties of orthogonal matrices.

Theorem 6: Properties of orthogonal matrices

Let Q be an orthogonal matrix.

a. The rows of Q form an orthonormal set.

b. Q^{-1} is orthogonal.

c. $\det Q = \pm 1$.

d. If λ is an eigenvalue of Q, then $|\lambda| = 1$.

e. If Q_1 and Q_2 are orthogonal $n \times n$ matrices, then so is $Q_1 Q_2$.

Proof. (a) Since $Q^{-1} = Q^T$, then Q^T is invertible and so

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

Hence Q^T is an orthogonal matrix. Thus, the columns of Q^T (which are the rows of Q) form an orthonormal set.

(b) This is similar. By (a) we have $(Q^{-1})^T = Q = (Q^{-1})^{-1}$.

(c) Since $\det(Q) = \det(Q^T)$ for any $n \times n$ matrix, then

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

(d) Let λ be an eigenvalue of Q with eigenvector $\mathbf{v}$. Then

$$\|\mathbf{v}\| = \|Q\mathbf{v}\| = \|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\|\,.$$

Since $\mathbf{v} \neq \mathbf{0}$, then $|\lambda| = 1$.

(e) If both Q_1 and Q_2 are orthogonal, then

$$(Q_1Q_2)^T(Q_1Q_2) = (Q_2^TQ_1^T)(Q_1Q_2) = Q_2^T(Q_1^TQ_1)Q_2 = Q_2^TQ_2 = I_n.$$

Hence, Q_1Q_2 is orthogonal. $\square$

2. Orthogonal complements

Are discussion in this section generalizes the notion of a normal vector. Recall that the normal vector $\mathbf{n}$ to a line ℓ has the property that it is orthogonal to every vector on the line.

Definition: Complement of a subspace, orthogonal to a set

Let W be a subspace of $\mathbb{R}^n$. The *complement of* W, is the set

$$W^{\perp} = \{\mathbf{v} \text{ in } \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \text{ in } W\}.$$

An element $\mathbf{v}$ in $W^{\perp}$ is said to be *orthogonal* to W.

Note that an element is in $W^{\perp}$ if it is orthogonal to every element in W.

Example. Let W be a plane through the origin in $\mathbb{R}^3$ and let L be a line through $\mathbf{0}$ perpendicular to W. If $\mathbf{z} \in L$ and $\mathbf{w} \in W$ are nonzero, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$. In fact, $L = W^{\perp}$ and $W = L^{\perp}$.

Theorem 7: Properties of complements

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

a. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

b. $W \cap W^{\perp} = \{\mathbf{0}\}$.

c. If $W = \operatorname{span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, then $\mathbf{v}$ is in $W^{\perp}$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for $i = 1, \ldots, k$.

Proof. (a) Since $\mathbf{0} \cdot \mathbf{w} = \mathbf{0}$ for all $\mathbf{w}$ in W, then $\mathbf{0}$ is in $W^{\perp}$. Let $\mathbf{u}$ and $\mathbf{v}$ be in $W^{\perp}$, let $\mathbf{w}$ be in W, and let λ be a scalar, then

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w}) = 0 + 0 = 0$$
$$(\lambda \mathbf{u}) \cdot \mathbf{w} = \lambda(\mathbf{u} \cdot \mathbf{w}) = \lambda \cdot 0 = 0.$$

Hence, $\mathbf{u} + \mathbf{v}$ and $(\lambda \mathbf{u})$ are in $W^{\perp}$. Thus, $W^{\perp}$ is a subspace.

(b) Let $\mathbf{v}$ be in W and $W^{\perp}$. Then $\mathbf{v} \cdot \mathbf{v} = 0$. But this implies that $\mathbf{v} = \mathbf{0}$.

(c) On one hand, if $\mathbf{v}$ is in $W^{\perp}$, then it is orthogonal to each $\mathbf{w}_i$ (since they are elements of W). On the other hand, suppose $\mathbf{v}$ is orthogonal to each $\mathbf{w}_i$. Let $\mathbf{w}$ be in W and write

$$\mathbf{w} = c_1 \mathbf{w}_1 + \cdots + c_k \mathbf{w}_k.$$

Then

$$0 = \mathbf{v} \cdot \mathbf{w} = c_1(\mathbf{v} \cdot \mathbf{w}_1) + \cdots + c_k(\mathbf{v} \cdot \mathbf{w}_k) = 0.$$

Hence, $\mathbf{v}$ is in $W^{\perp}$. $\square$

Orthogonality gives important relationships between column and row spaces, and null spaces.

Theorem 8: Complements of row and column spaces are null spaces

Let A be an $m \times n$ matrix. Then $(\operatorname{row} A)^\perp = \operatorname{null} A$ and $(\operatorname{col} A)^\perp = \operatorname{null} A^T$.

Proof. If $\mathbf{x} \in \operatorname{null} A$, then $A\mathbf{x} = \mathbf{0}$ by definition. Hence, $\mathbf{x}$ is orthogonal to each row of A. Since the rows of A span row A, then $\mathbf{x} \in (\operatorname{row} A)^\perp$.

Conversely, if $\mathbf{x} \in (\operatorname{row} A)^\perp$, then $\mathbf{x}$ is orthogonal to each row of A and $A\mathbf{x} = 0$, so $\mathbf{x} \in \operatorname{null} A$.

For the second statement, we can apply the first part to get

$$(\operatorname{col} A)^\perp = (\operatorname{row} A^T)^\perp = \operatorname{null} A^T \qquad \square.$$

Example. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & 4 \end{bmatrix}$.

We previously showed that a basis for row(A) is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ where

$$\mathbf{u}_1 = \begin{bmatrix} 1 & -2 & 0 & -1 & 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 & 0 & 1 & 2 & 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We also showed that a basis for null(A) is $\{\mathbf{x}_1, \mathbf{x}_2\}$ where

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}.$$

One can easily check that $\mathbf{u}_i \cdot \mathbf{x}_j$ for all i, j.

Recall that, in $\mathbb{R}^2$, the projection of a vector $\mathbf{v}$ onto a nonzero vector $\mathbf{u}$ is given by

$$\operatorname{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}.$$

The vector $\operatorname{perp}_{\mathbf{u}}(\mathbf{v}) = \mathbf{v} - \operatorname{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to $\operatorname{proj}_{\mathbf{u}}(\mathbf{v})$ and we can decompose $\mathbf{v}$ as

$$\mathbf{v} = \operatorname{proj}_{\mathbf{u}}(\mathbf{v}) + \operatorname{perp}_{\mathbf{u}}(\mathbf{v}).$$

Definition: Orthogonal projection onto a subspace, orthogonal component

Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. For any vector $\mathbf{v} \in \mathbb{R}^n$, the *orthogonal projection of* $\mathbf{v}$ *onto* W is given by

$$\operatorname{proj}_W(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{v} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \mathbf{u}_k.$$

The *component of* $\mathbf{v}$ *orthogonal to* W is

$$\operatorname{perp}_W(\mathbf{v}) = \mathbf{v} - \operatorname{proj}_W(\mathbf{v}).$$

This definition matches our previous one when W is 1-dimensional. Note that $\text{proj}_W(\mathbf{v})$ is an element of W because it is a linear combination of basis elements. In particular,

$$\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \text{proj}_{\mathbf{u}_k}(\mathbf{v}).$$

Also note that the definition simplifies when the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is orthonormal. In this case, if we let $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{bmatrix}$, then $\text{proj}_W(\mathbf{v}) = UU^T\mathbf{v}$ for all $\mathbf{v}$ in $\mathbb{R}^n$.

Example. Let W be the plane in $\mathbb{R}^3$ with equation $x - y + 2z = 0$. An orthogonal basis for W is $\{\mathbf{u}_1, \mathbf{u}_2\}$ where

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

Let $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$. Then

$$\text{proj}_W(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{v} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \frac{2}{2}\mathbf{u}_1 + \frac{-2}{3}\mathbf{u}_2 = \frac{1}{3}\begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}.$$

We further have that

$$\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(\mathbf{v}) = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 4 \\ -4 \\ 8 \end{bmatrix}.$$

The next result shows that the decomposition of a vector into a sum of its projection and its orthogonal component is unique in a certain sense. This result relies on a fact, justified in the next section, that every subspace has an *orthogonal basis*.

Theorem 9: Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$.

Proof. We must show two things: (1) that a decomposition exists and (2) that it is unique.

(Existence) Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W (note that if $W = \{\mathbf{0}\}$ then this theorem is trivial). Let $\mathbf{w} = \text{proj}_W(\mathbf{v})$, which is in W, and $\mathbf{w}^\perp = \text{perp}_W(\mathbf{v})$. Then

$$\mathbf{w} + \mathbf{w}^\perp = \text{proj}_W(\mathbf{v}) + \text{perp}_W(\mathbf{v}) = \text{proj}_W(\mathbf{v}) + (\mathbf{v} - \text{proj}_W(\mathbf{v})) = \mathbf{v}.$$

It remains only to show that $\mathbf{w}^{\perp}$ is in $W^{\perp}$. It is enough to show that $\mathbf{w}^{\perp} \cdot \mathbf{u}_i$ for each basis vector $\mathbf{u}_i$. We have,

$$\begin{aligned}
\mathbf{w}^{\perp} \cdot \mathbf{u}_i &= \operatorname{perp}_W(\mathbf{v}) \cdot \mathbf{u}_i \\
&= (\mathbf{v} - \operatorname{proj}_W(\mathbf{v})) \cdot \mathbf{u}_i \\
&= \mathbf{v} \cdot \mathbf{u}_i - \left(\frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{v} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \mathbf{u}_k \right) \cdot \mathbf{u}_i \\
&= \mathbf{v} \cdot \mathbf{u}_i - \frac{\mathbf{v} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} (\mathbf{u}_i \cdot \mathbf{u}_i) \\
&= \mathbf{v} \cdot \mathbf{u}_i - \mathbf{v} \cdot \mathbf{u}_i = 0.
\end{aligned}$$

Hence, $\mathbf{w}^{\perp}$ is in $W^{\perp}$ as desired.

To prove uniqueness, let $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^{\perp}$ be another decomposition with $\mathbf{w}_1$ in W and $\mathbf{w}_1^{\perp}$ in $W^{\perp}$. Then $\mathbf{w} + \mathbf{w}^{\perp} = \mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^{\perp}$, so $(\mathbf{w} - \mathbf{w}_1) = (\mathbf{w}_1^{\perp} - \mathbf{w}^{\perp})$, which is simulateneously a vector in W and $W^{\perp}$. Since $W \cap W^{\perp} = \{\mathbf{0}\}$, then $\mathbf{w} - \mathbf{w}_1 = \mathbf{0}$, so $\mathbf{w} - \mathbf{w}_1$. Similarly, $\mathbf{w}^{\perp} = \mathbf{w}_1^{\perp}$. ☐

We can now finish establishing our properties for subspace complements.

Proposition 10: Further properties of complements

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

a. $(W^{\perp})^{\perp} = W$.
b. $\dim W + \dim W^{\perp} = n$.

Proof. (a) Let $\mathbf{w}$ be in W and $\mathbf{x}$ in $W^{\perp}$. Then $\mathbf{w} \cdot \mathbf{x}$, so $\mathbf{w}$ is in $(W^{\perp})^{\perp}$. That is $W \subseteq (W^{\perp})^{\perp}$.

Now let $\mathbf{v}$ be in $(W^{\perp})^{\perp}$. By the Orthogonal Decomposition Theorem, $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$ with $\mathbf{w}$ in W and $\mathbf{w}^{\perp}$ in $W^{\perp}$. Then

$$0 = \mathbf{v} \cdot \mathbf{w}^{\perp} = (\mathbf{w} + \mathbf{w}^{\perp}) \cdot \mathbf{w}^{\perp} = (\mathbf{w} \cdot \mathbf{w}^{\perp}) + (\mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp}) = 0 + (\mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp}) = (\mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp}).$$

Hence, $\mathbf{w}^{\perp} = 0$, so $\mathbf{v} = \mathbf{w}$ in W. Hence, $(W^{\perp})^{\perp} \subseteq W$. By double inclusion, $W = (W^{\perp})^{\perp}$.

(b) Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis of W and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ be an orthogonal basis of $W^{\perp}$. Note that $\mathbf{u}_i \cdot \mathbf{v}_j = 0$ for any i, j. It follows that $\mathscr{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ is a linearly independent set in $\mathbb{R}^n$. On the other hand, by the Orthogonal Decomposition Theorem, $\mathscr{B}$ spans $\mathbb{R}^n$. It follows that $\mathscr{B}$ is a basis of $\mathbb{R}^n$. Hence, $k + \ell = n$. ☐

3. The Gram-Schmidt Process

A lingering question from the previous section was whether it was possible to construct an orthogonal (or orthonormal) basis for a subspace W of $\mathbb{R}^n$. In this section we discuss an algorithm for doing this.

Example. Let $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$.

We will construct an orthogonal basis for W. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{span}\{\mathbf{v}_1\}$. It suffices to find a vector $\mathbf{v}_2 \in W$ orthogonal to W_1. Let $\mathbf{p} = \text{proj}_{W_1} \mathbf{x}_2 \in W_1$. Then $\mathbf{x}_2 = \mathbf{p} + (\mathbf{x}_2 - \mathbf{p})$ where $\mathbf{x}_2 - \mathbf{p} \in W_1^{\perp}$. Let

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \begin{bmatrix} 9 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}.$$

Now $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and $\mathbf{v}_1, \mathbf{v}_2 \in W$. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W.

To obtain an orthonormal basis for W, take the unit vectors associated to $\mathbf{v}_1$ and $\mathbf{v}_2$.

This process could continue. Say W was three-dimensional. We could then let $W_2 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and find the projection of $\mathbf{x}_3$ onto W_2. We'll prove the next theorem using this idea.

Theorem 11: The Gram-Schmidt Process

iven a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace $W \subset \mathbb{R}^n$, 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. In addition,

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$$

for all $1 \le k \le p$.

Proof. For $1 \le k \le p$, set $W_k = \text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and $V_k = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Since $\mathbf{v}_1 = \mathbf{x}_1$, then it holds (trivially) that $W_1 = V_1$ and $\{\mathbf{v}_1\}$ is orthogonal.

Suppose for some k, $1 \le k < p$, that $W_k = V_k$ and that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \operatorname{proj}_{W_k} \mathbf{x}_{k+1} \in W_k^{\perp} \subset W_{k+1}.$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Since $\mathbf{x}_{k+1} \in W_{k+1}$, then $\mathbf{v}_{k+1} \in W_{k+1}$. Hence, $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set of $k+1$ nonzero vectors in W_{k+1} and hence a basis of W_{k+1}. Hence, $W_{k+1} = V_{k+1}$. The result now follows by induction. $\square$

Example. Let $W = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

We will construct an orthogonal basis for W.

Set $\mathbf{v}_1 = \mathbf{x}_1$. Then

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix}.$$

Now,

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} - \frac{5}{2}\begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}.$$

Hence, an orthogonal basis for W is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

4. Orthogonal diagonalization of symmetric matrices

We have seen already that it is quite time intensive to determine whether a matrix is diagonalizable. We'll see that there are certain cases when a matrix is always diagonalizable.

Recall that a matrix is *symmetric* if $A^T = A$. It is (reasonably) easy to show that the product of symmetric matrices is symmetric, and the inverse of a symmetric

Example. Let $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

The eigenvalues of A are $\lambda_1 = -2$ and $\lambda_2 = 4$ and corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Hence, A is diagonalizable. Set $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$, then $P^{-1}AP = D$ where $D = \begin{bmatrix} -2 & 0 \\ 0 & 4 \end{bmatrix}$.

But note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Normalizing these vectors we have

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Set $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$. Since Q is an orthogonal matrix, then $Q^{-1} = Q^T$, so we have $Q^T AQ = D$. This is adventageous since computing the transpose is much easier than computing the inverse.

Definition: Orthogonally diagonlizable

An square matrix A is *orthogonally diagonalizable* if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$.

The next theorem is easy to prove. Its converse is also true, but this will take much more work to prove.

Theorem 12: Orthogonally diagonalizable implies symmetric

If A is orthogonally diagonalizable, then A is symmetric.

Proof. Since A is orthogonally diagonalizable, then $Q^T AQ = D$ for some orthogonal matrix Q and diagonal matrix D. But then $A = (Q^T)^{-1}DQ^T = QDQ^T$. Hence,

$$A^T = (QDQ^T) = (Q^T)^T D^T Q^T = QDQ^T = A. \qquad \square$$

Recall that matrices like $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, though all the entries are real, have complex eigenvalues. It turns out that this never happens with symmetric matrices.

We write complex numbers in $\mathbb{C}$ as $a+bi$ where a and b are real numbers and $i=\sqrt{-1}$. The *complex conjugate* of $z=a+bi$ is $\overline{z}=\overline{a+bi}=a-bi$. We can extend this notion to complex matrices. The conjugate of a matrix $A=[a_{ij}]$ with complex entries is $\overline{A}=[\overline{a_{ij}}]$.

Theorem 13: Eigenvalues of symmetric matrices are real

If A is a real symmetric matrix, then the eigenvalues of A are real.

Proof. Suppose λ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}$. Then $A\mathbf{v}=\lambda\mathbf{v}$ and so

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

Now we take transposes. Using the fact that A is symmetric we have

$$\overline{\mathbf{v}}^TA=\overline{\mathbf{v}}^TA^T=(A\overline{\mathbf{v}})^T=(\overline{\lambda}\overline{\mathbf{v}})^T=\overline{\lambda}\overline{\mathbf{v}}^T.$$

Therefore,

$$\lambda(\overline{\mathbf{v}}^T\mathbf{v})=\overline{\mathbf{v}}^T(\lambda\mathbf{v})=\overline{\mathbf{v}}^T(A\mathbf{v})=(\overline{\mathbf{v}}^TA)\mathbf{v}=(\overline{\lambda}\overline{\mathbf{v}}^T)\mathbf{v}=\overline{\lambda}(\overline{\mathbf{v}}^T\mathbf{v}).$$

This shows that $(\lambda-\overline{\lambda})(\overline{\mathbf{v}}^T\mathbf{v})=0$. Write

$$\mathbf{v}=\begin{bmatrix}a_1+b_1i\\ \vdots\\ a_n+b_ni\end{bmatrix}\quad\text{so}\quad\overline{\mathbf{v}}=\begin{bmatrix}a_1-b_1i\\ \vdots\\ a_n-b_ni\end{bmatrix}.$$

Then

$$\overline{\mathbf{v}}^T\mathbf{v}=\overline{\mathbf{v}}\cdot\overline{\mathbf{v}}=(a_1^2+b_1)^2+\cdots+(a_n^2+b_n^2)\neq 0$$

because $\mathbf{v}\neq 0$. We conclude that $\lambda-\overline{\lambda}=0$. That is $\lambda=\overline{\lambda}$ so λ is real. $\square$

The next theorem is stronger than a previous result for general matrices.

Theorem 14: Eigenvectors of symmetric matrices

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\mathbf{v}_1,\mathbf{v}_2$ be eigenvectors for A with corresponding eigenvalues λ_1,λ_2, $\lambda_1\neq\lambda_2$. Then

$$\lambda_1(\mathbf{v}_1\cdot\mathbf{v}_2)=(\lambda_1\mathbf{v}_1)^T\mathbf{v}_2=(A\mathbf{v}_1)^T\mathbf{v}_2=\mathbf{v}_1^TA^T\mathbf{v}_2=\mathbf{v}_1^TA\mathbf{v}_2=\mathbf{v}_1^T(\lambda_2\mathbf{v}_2)=\lambda_2(\mathbf{v}_1\cdot\mathbf{v}_2).$$

Hence, $(\lambda_1-\lambda_2)(\mathbf{v}_1\cdot\mathbf{v}_2)=0$. Since $\lambda_1\neq\lambda_2$, then we must have $\mathbf{v}_1\cdot\mathbf{v}_2=0$. $\square$

Based on the previous theorem, we say that the eigenspaces of A are *mutually orthogonal*.

Example. Let $A=\begin{bmatrix}3&-2&4\\-2&6&2\\4&2&3\end{bmatrix}$.

The eigenvalues of A are -2 and 7. The eigenspaces have bases,

$$E_7 : \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} \right\}, \qquad E_{-2} : \left\{ \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix} \right\}.$$

It is easy now to verify that the basis vector for E_{-2} is orthogonal to those of E_7.

However, the two basis vectors for E_7 *are not* orthogonal. In order to orthgonally diagonalize A, we need an orthogonal basis for E_7. To do this, we use Gram-Schmidt:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Note for $\mathbf{v}_3$ we only scaled the vector. Finally, we normalize each vector,

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Now the matrix $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is orthogonal and so $Q^T Q = I$.

The set of eigenvalues of a matrix A is called the *spectrum of* A and is denoted σ_A.

Theorem 15: The (real) Spectral Theorem

Let A be an $n \times n$ real matrix. Then A is symmetric if and only if it is orthogonally diagonalizable.

Proof. We already proved that orthogonally diagonalizable implies symmetric. Now we assume A is symmetric and prove that it is orthgonally diagonalizable. Clearly the result holds when A is 1×1. Assume $(n-1) \times (n-1)$ symmetric matrices are orthogonally diagonalizable.

Let A be $n \times n$ and let λ_1 be an eigenvalue of A and $\mathbf{u}_1$ a (unit) eigenvector for λ_1. Set $W = \text{span}\{\mathbf{u}_1\}$. By the Gram-Schmidt process we may extend $\mathbf{u}_1$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ where $\{\mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for $W^\perp$. Set $Q_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$Q_1^T A Q_1 = \begin{bmatrix} \mathbf{u}_1^T A \mathbf{u}_1 & \cdots & \mathbf{u}_1^T A \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n^T A \mathbf{u}_1 & \cdots & \mathbf{u}_n^T A \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}.$$

The first column is as indicated because $\mathbf{u}_i^T A \mathbf{u}_1 = \mathbf{u}_i^T(\lambda \mathbf{u}_1) = \lambda(\mathbf{u}_i \cdot \mathbf{u}_1) = \lambda \delta_{ij}$. As $Q_1^T A Q_1$ is symmetric, $* = 0$ and B is a symmetric $(n-1) \times (n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \ldots, \lambda_n$ (by the inductive hypothesis). Because A and $Q_1^T A Q_1$ are similar, then the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q_2 such that $Q_2^T B Q_2 = D$, where the diagonal entries of D are $\lambda_2, \ldots, \lambda_n$. Now

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}^T \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q_2^T B Q_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D \end{bmatrix}.$$

Note that $\begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$ is orthogonal. Set $Q = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$. As the product of orthogonal matrices is orthogonal, Q is itself orthogonal and $Q^T A Q$ is diagonal. $\square$

Let A be orthogonally diagonalizable. Then $A = QDQ^T$ where

$$Q = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$$

and D is the diagonal matrix whose diagonal entries are the eigenvalues of A: $\lambda_1, \ldots, \lambda_n$. Then

$$A = QDQ^T = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T.$$

This is known as the *spectral decomposition of* A, or *projection form of the Spectral Theorem.* Each $\mathbf{u}_i \mathbf{u}_i^T$ is called a *projection matrix* because $(\mathbf{u}_i \mathbf{u}_i^T)\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\operatorname{span}\{\mathbf{u}_i\}$.

Example 16. Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$. An orthonormal basis of the column space is

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Setting $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ gives $Q^T A Q = D = \operatorname{diag}(7, 7, -2)$. The projection matrices are

$$\mathbf{u}_1 \mathbf{u}_1^T = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \mathbf{u}_2 \mathbf{u}_2^T = \frac{1}{18} \begin{bmatrix} 1 & -4 & -1 \\ -4 & 16 & 4 \\ -1 & 4 & 1 \end{bmatrix}, \mathbf{u}_3 \mathbf{u}_3^T = \frac{1}{9} \begin{bmatrix} 4 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 4 \end{bmatrix}.$$

The spectral decomposition is

$$7\mathbf{u}_1 \mathbf{u}_1^T + 7\mathbf{u}_2 \mathbf{u}_2^T - 2\mathbf{u}_3 \mathbf{u}_3^T = A.$$

Best Approximation/Linear Regression

In data science, one often wants to be able to approximate a set of data by a curve. One might hope to construct the line that best fits the data. This is known (by one name) as *linear regression*. We will study a linear algebraic approach to this problem.

Suppose the system $A\mathbf{x} = \mathbf{b}$ is inconsistent. Previously, we gave up all hope then of "solving" this system because no solution existed. However, if we give up the idea that we must find an *exact* solution and instead focus on finding an *approximate* solution, then we may have hope of solving.

Another way of saying $A\mathbf{x} = \mathbf{b}$ is inconsistent is to say that $\mathbf{b}$ does not lie in the column space of A. To find the point in the column space *closest* to $\mathbf{b}$, we use an orthogonal projection: $\mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})$. In fact, the *Best Approximation Theorem* formalizes this idea. The projection is the closest point in $\mathrm{col}\,A$ to $\mathbf{b}$ in the sense that for any other point $\mathbf{v}$ in $\mathrm{col}\,A$ with $\mathbf{v} \neq \mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})$, we have

$$\|\mathbf{b} - \mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})\| < \|\mathbf{b} - \mathbf{v}\|.$$

(Note if $A\mathbf{x} = \mathbf{b}$ is consistent, then $\mathbf{b}$ is in $\mathrm{col}(A)$ and so $\mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b}) = \mathbf{b}$.)

Hence, we can find some vector $\overline{\mathbf{x}}$ such that $A\overline{\mathbf{x}} = \mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})$. We call $\overline{\mathbf{x}}$ a *least squares solution* to the inconsistent system $A\mathbf{x} = \mathbf{b}$.

Recall that $\mathbf{b} - \mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})$ is in $\mathrm{col}(A)^{\perp}$, and $\mathrm{col}(A)^{\perp} = \mathrm{null}(A^T)$, then

$$0 = A^T(\mathbf{b} - \mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})) = A^T\mathbf{b} - A^T\,\mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b}).$$

That is,

$$(A^TA)\overline{\mathbf{x}} = A^T(A\overline{\mathbf{x}}) = A^T(\mathrm{proj}_{\mathrm{col}\,A}(\mathbf{b})) = A^T\mathbf{b}.$$

The least-squares solutions are now the solutions to the system $(A^TA)\mathbf{x} = A^T\mathbf{b}$.

Example. Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 5 & 1 \\ 4 & 1 \\ 3 & 0 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

We will use normal equations. First we compute

$$A^TA = \begin{bmatrix} 50 & 9 \\ 9 & 2 \end{bmatrix}, \qquad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

To solve the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ we invert A^TA.

$$\overline{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{19}\begin{bmatrix} 2 & -9 \\ -9 & 50 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/19 \\ 5/19 \end{bmatrix}.$$

Note in general, A^TA need not be invertible. Hence, when A^TA is invertible then the least-squares solution $\hat{\mathbf{x}}$ is unique and

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

The *error vector* is defined as

$$\mathbf{e} = \mathbf{b} - A\overline{\mathbf{x}}$$

The *least squares error* is then the length $\|\mathbf{e}\|$.

Example. In the previous example, we have

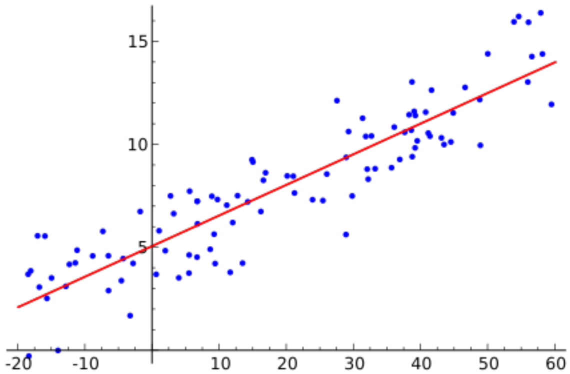
$$\mathbf{e} = \mathbf{b} - A\overline{\mathbf{x}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 & 1 \\ 4 & 1 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1/19 \\ 5/19 \end{bmatrix} = \frac{1}{19}\begin{bmatrix} 9 \\ -9 \\ -3 \end{bmatrix}.$$

Hence, the least squares error is

$$\|\mathbf{e}\| = \left\| \frac{1}{19}\begin{bmatrix} 9 \\ -9 \\ -3 \end{bmatrix} \right\| = \frac{3}{\sqrt{19}}.$$

We will see how to fit data to a line using least-squares. We denote the equation $A\mathbf{x} = \mathbf{b}$ by $X\boldsymbol{\beta} = \mathbf{y}$. The matrix X is referred to as the *design matrix*, $\boldsymbol{\beta}$ as the *parameter vector*, and $\mathbf{y}$ as the *observation vector*.

We will model a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ by a line. Denote this line by $y = \beta_0 + \beta_1 x$. The *residual* of a point (x_i, y_i) is the distance from that point to the line. The *least-squares line* minimizes the sum of the squares of the residuals.



Suppose the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ were all on the line. Then they would satisfy,

$$\beta_0 + \beta_1 x_1 = y_1$$

$$\vdots$$

$$\beta_0 + \beta_1 x_n = y_n.$$

We could write this system as $X\boldsymbol{\beta} = \mathbf{y}$ where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

If the data does not lie on the line, then we want the vector $\boldsymbol{\beta}$ to be the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ that minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$.

Example. Given the data points $(4,1)$, $(1,2)$, $(3,3)$, $(5,5)$. We will find the equation $y = \beta_0 + \beta_1 x$. We build the matrix X and vector $\mathbf{y}$ from the data,

$$X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}.$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, we have the normal equation $X^T X\boldsymbol{\beta} = X^T\mathbf{y}$ where

$$X^T X = \begin{bmatrix} 4 & 13 \\ 13 & 51 \end{bmatrix}, \quad X^T\mathbf{y} = \begin{bmatrix} 11 \\ 40 \end{bmatrix}.$$

Hence, the least squares solution is

$$\overline{\mathbf{x}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X^T X)^{-1} X^T \mathbf{y} = \begin{bmatrix} 41/35 \\ 17/35 \end{bmatrix}.$$

The least squares error is

$$\|\mathbf{y} - X\overline{\mathbf{x}}\| = \frac{3\sqrt{26}}{\sqrt{35}} \approx 2.58567$$

Say our data was better approximated by a parabola. In this case, we want a degree two polynomial

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

that minimizes the least squares error. A similar analysis as before shows that we end up with a system $X\boldsymbol{\beta} = \mathbf{y}$ where

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

(Note that X is a type of matrix called a *Vandermonde matrix*.)

Example. Consider the data from the previous example. We will adjust our method to find the parabola that best fits the data. From the data, we have

$$X = \begin{bmatrix} 1 & 4 & 16 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}.$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, we have the normal equation $X^T X\boldsymbol{\beta} = X^T\mathbf{y}$ where

$$X^T X = \begin{bmatrix} 4 & 13 & 51 \\ 13 & 51 & 217 \\ 51 & 217 & 963 \end{bmatrix}, \quad X^T\mathbf{y} = \begin{bmatrix} 11 \\ 40 \\ 170 \end{bmatrix}.$$

Hence, the least squares solution is

$$\overline{\mathbf{x}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} 18/5 \\ -391/220 \\ 17/44 \end{bmatrix}.$$

The least squares error is

$$\|\mathbf{y} - X\overline{\mathbf{x}}\| = \frac{23}{\sqrt{110}} \approx 2.19296$$

which is less than in the previous exercise.

Chapter 6: Vector Spaces

Though we are studying "abstract" vector spaces, we will quickly see that there are many concrete examples of spaces satisfying the same linearity properties as $\mathbb{R}^n$. We will generalize our notation of a *real* vector space to indicate spaces whose scalars comes from $\mathbb{R}$. However, much of what we do works when $\mathbb{R}$ is replaced by $\mathbb{C}$. We will also see that many of these vector spaces are in some sense indistinguishable from $\mathbb{R}^n$, and this will lead to the notion of an *isomorphism*.

1. Vector Spaces and Subspaces

Definition: Vector Space

A *vector space* (over $\mathbb{R}$) is a nonempty set V of objects (called *vectors*) along with two operations: addition and multiplication by scalars in $\mathbb{R}$, subject to the following axioms:

(1) If $\mathbf{u}, \mathbf{v}$ are in V, then $\mathbf{u} + \mathbf{v}$ is in V;
(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v}$ in V;
(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V;
(4) There exists an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v}$ in V;
(5) For all $\mathbf{v}$ in V, there exists an element $-\mathbf{v}$ in V such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
(6) For all $\mathbf{v}$ in V and all scalars c, $c\mathbf{v}$ is in V;
(7) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for all $\mathbf{u}, \mathbf{v}$ in V and all scalars c;
(8) $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$ for all $\mathbf{v}$ in V and all scalars c, d;
(9) $c(d\mathbf{v}) = (cd)\mathbf{v}$ for all scalars c, d and all $\mathbf{v}$ in V;
(10) $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$ in V.

Example. We have already seen, extensively, that $\mathbb{R}^n$ is a vector space. You should still be able to verify any of the above properties for $\mathbb{R}^n$.

Example. Let M denote the set of 2×3 matrices. We have shown previously that M is a vector space (we just didn't use that terminology). Here we review what that means.

First note that M is closed under matrix addition (1) and scalar multiplication (6) by definition. Moreover, we checked previously that matrix addition is commutative (2) and associative (3). The $\mathbf{0}$ element is the 2×3 zero matrix. It is easy to check that for all 2×3 matrices A, $A + \mathbf{0} = A$ (4) and $A + (-1)A = \mathbf{0}$ (5). We further checked that matrix addition and scalar multiplication distribute over one another (7,8,9). Finally, (10) is clear.

This argument works for any size $m \times n$ matrices, and we denote the vector space of such matrices by $\mathcal{M}_{mn}$ (or just $\mathcal{M}_n$ for square $n \times n$ matrices).

The material for these notes is derived primarily from *Linear Algebra: A Modern Introduction* by David Poole (4ed) and *Linear Algebra and its applications* by David Lay (4ed). Last Updated: December 4, 2021

Example. Let $\mathscr{P}_2$ denote polynomials (with real coefficients) of degree at most 2. Given two such polynomials $p(x)$ and $q(x)$, write

$$p(x) = a_0 + a_1 x + a_2 x^2 \quad \text{and} \quad q(x) = b_0 + b_1 x + b_2 x^2$$

with a_i, b_i real numbers. Then addition is defined by

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2$$

which is another element of $\mathcal{P}_2$. Similarly, if λ is a scalar then

$$\lambda p(x) = (\lambda a_0) + (\lambda a_1)x + (\lambda a_2)x^2$$

which again an element of $\mathscr{P}_2$. Thus, $\mathscr{P}_2$ is closed under the operations.

The zero element is the zero polynomial (in which $a_0 = a_1 = a_2 = 0$). The additive inverse of $p(x)$ is $(-1)p(x) = (-a_0) + (-a_1)x + (-a_2)x^2$. The axioms are now easy to check. We will check associative of addition. Let $p(x)$ and $q(x)$ be as above and write $r(x) = c_0 + c_1 x + c_2 x^2$. Then

$$\begin{aligned}
(p(x) + q(x)) + r(x) &= \big((a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2\big) + (c_0 + c_1 x + c_2 x^2) \\
&= ((a_0 + b_0) + c_0) + ((a_1 + b_1) + c_1)x + ((a_2 + b_2) + c_2)x^2 \\
&= (a_0 + (b_0 + c_0)) + (a_1 + (b_1 + c_1))x + (a_2 + (b_2 + c_2))x^2 \\
&= (a_0 + a_1 x + a_2 x^2) + \big((b_0 + c_0)) + (b_1 + c_1)x + (b_2 + c_2)x^2\big) \\
&= p(x) + (q(x) + r(x)).
\end{aligned}$$

You should work through the remaining properties. There is nothing special about degree two here. We could similarly define $\mathscr{P}_n$ to be the set of all polynomials of degree at most n.

Example. Let $\mathscr{F}$ denote the set of all real-valued functions with domain $\mathbb{R}$. If f and g are in $\mathscr{F}$ and c is a scalar, then $f + g$ and cf are defined by

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (cf)(x) = cf(x).$$

The zero vector is the function z defined by $z(x) = 0$ for all x in $\mathbb{R}$. You are encouraged to work through the axioms for these operations.

Theorem 1: Properties of vector spaces

Let V be a vector space, $\mathbf{u}$ a vector in V, and c a scalar. Then we have

a. $0\mathbf{u} = \mathbf{0}$,

b. $c\mathbf{0} = \mathbf{0}$,

c. $-\mathbf{u} = (-1)\mathbf{u}$,

d. If $c\mathbf{u} = \mathbf{0}$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$.

Proof. (a) By axiom (8),

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}.$$

Subtracting $0\mathbf{u}$ on both sides gives $0\mathbf{u} = \mathbf{0}$.

(b) Similarly, by axiom (7),

$$c\mathbf{0} = c(\mathbf{0} + \mathbf{0}) = c\mathbf{0} + c\mathbf{0}.$$

Subtracting $c\mathbf{0}$ on both sides gives $c\mathbf{0} = \mathbf{0}$.

(c) By part (a),

$$\mathbf{0} = 0\mathbf{u} = (1 + (-1))\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}.$$

Subtracting $\mathbf{u}$ on both sides gives $-\mathbf{u} = (-1)\mathbf{u}$.

(d) If $c = 0$, then we are done. Suppose c is not zero. Then we can multiply both sides by $1/c$ and use axiom (10) to obtain

$$\mathbf{u} = 1\mathbf{u} = (1/c)c\mathbf{u} = (1/c)\mathbf{0} = \mathbf{0}. \qquad \square$$

We now generalize our notion of subspaces to general vector spaces.

Definition: Subspace

A *subspace* of a vector space V is a subset W of V that has three properties:

a. The zero vector of V is in W.
b. (closure under addition) If $\mathbf{u}$ and $\mathbf{v}$ are in W, then $\mathbf{u} + \mathbf{v}$ is in W.
c. (closure under scalar multiplication) If $\mathbf{u}$ is in W and c is a scalar, then $c\mathbf{u}$ is in W.

The axioms imply that W is itself a vector space. That is, one could use the definition along with the properties of V to verify that W satisfies the axioms of a vector space. Many of these come "for free" because they are properties of the operation (such as associativity).

Example. Let W be the subset of $\mathcal{M}_n$ consisting of *symmetric* $n \times n$ matrices. (Recall a square matrix A is symmetric if $A^T = A$.) We claim W is a subspace of $\mathcal{M}_n$.

Clearly, the zero matrix is symmetric, so W contains the zero vector of $\mathcal{M}_n$. Let A and B be symmetric matrices. By properties of transpose,

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

Hence, $A + B$ is symmetric (and hence an element of W). So W is closed under addition.

Now let A be a symmetric matrix and let c be a scalar. Then again by properties of transpose,

$$(cA)^T = cA^T = cA,$$

so cA is symmetric (and hence an element of W). So W is closed under scalar multiplication.

Example. Let $\mathscr{C}$ denote the set of all continuous real-valued functions defined on $\mathbb{R}$. This is a subspace of $\mathscr{F}$ and hence a vector space, which we will take as given since the tools to check this rigorously are not available to us right now. Let $\mathscr{D}$ denote the subset of $\mathscr{C}$ consisting of differentiable functions on $\mathbb{R}$. Clearly the zero function is differentiable. If f and g are in $\mathscr{D}$ and c is a scalar, then by rules you learned in calculus,

$$(f+g)' = f' + g' \quad \text{and} \quad (cf)' = cf'.$$

Hence, $\mathscr{D}$ is closed under addition and scalar multiplication.

Example. The subset of a vector space V containing only the identity, $\{\mathbf{0}\}$ is a subspace of V called the *trivial subspace*. Any vector space V is a subspace of itself and any subspce H of V such that $H \neq V$ is called a *proper subspace* of V.

Example. Let H denote the subset of $\mathcal{M}_n$ consisting of $n \times n$ matrices of determinant zero. Clearly, H contains the zero matrix. Moreover, by our determinant rules, if $M \in H$ and $c \in \mathbb{R}$, then

$$\det(cM) = c^n \det(M) = 0.$$

Thus, $cM \in H$. However, if $M, N \in H$, then $M + N$ need not be an element of H. Consider the following example in $n = 2$.

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then $M + N = I_2 \notin H$. Hence, H is not a subspace of $\mathcal{M}_n$.

We continue to use much of our terminology for $\mathbb{R}^n$.

Definition: Linear combination, span

A *linear combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a vector space V is an expression of the form

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

with $c_1, \ldots, c_n$ scalars. The *span* of a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in V is the set of all linear combinations of those vectors, denoted $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

Example. (1) Any polynomial $p(x)$ in $\mathscr{P}_2$ can be written as $a(1) + b(x) + c(x^2)$ with a, b, c real numbers. Hence, $\mathcal{P}_2 = \text{span}(1, x, x^2)$. More generally, $\mathcal{P}_n = \text{span}(1, x, \ldots, x^n)$.

(2) The set of matrices below

$$E_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad E_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_{21} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad E_{22} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad E_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a spanning set for $\mathcal{M}_{2,3}$.

2. Linear Independence, Basis, and Dimension

Now we recall several additional definitions, stated in our more general setting.

Definition: Linear independence, linear dependence

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in a vector space V is *linearly independent* if whenever there are scalars c_i such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0},$$

then $c_i = 0$ for all i. The set is *linearly dependent* if there are scalars c_i not all zero such that $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$.

As we saw before, an alternative definition for linear dependence is that one vector in the set may be expressed as a linear combination of the other vectors.

Example. (1)The matrices E_{ij} in the previous example form a linearly independent set in $\mathcal{M}_{2,3}$. To see this, consider the equation

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \lambda_1 E_{11} + \lambda_2 E_{12} + \lambda_3 E_{13} + \lambda_4 E_{21} + \lambda_5 E_{22} + \lambda_6 E_{23} = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_4 & \lambda_5 & \lambda_6 \end{bmatrix}.$$

So clearly, $\lambda_i = 0$ for all i.

(2) In $\mathcal{M}_2$ the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix},$$

are linearly dependent because $A + B = C$.

(3) The set $\{1, x, \ldots, x^n\}$ in $\mathcal{P}_n$ is linearly independent. To see this, suppose there are scalars c_i such that

$$c_0 + c_1 x + \cdots + c_n x^n = 0.$$

If we set $x = 0$, then we obtain $c_0 = 0$. Now taking the first derivative we obtain

$$c_1 + 2c_2 x + \cdots + nc_n x^{n-1} = 0.$$

Again setting $x = 0$ gives $c_1 = 0$. Continuing in this way we obtain $c_i = 0$ for all i.

Definition: Basis

A subset $\mathscr{B}$ of a vector space V is a *basis* for V if $\mathscr{B}$ is linearly independent and spans $\mathscr{B}$.

Example. (1) We have seen already that $\{E_{ij}\}$ is a basis for $\mathcal{M}_{2,3}$.

(2) We have already seen that $\{1, x, \ldots, x^n\}$ is a basis for $\mathscr{P}_n$.

(3) The set $\{1 + x, x + x^2, 1 + x^2\}$ is also a basis for $\mathscr{P}_2$. We could work this out formally using the methods above, but we will soon see a better way to check this.

The following result we have seen before for $\mathbb{R}^n$. It is of critical importance for defining *coordinates*.

Theorem 2: Uniqueness of expression

Let V be a vector space with basis $\mathscr{B}$. For every vector $\mathbf{v}$ in V, there is exactly one way to write $\mathbf{v}$ as a linear combination of elements of $\mathscr{B}$.

Proof. Suppose $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and let $\mathbf{v}$ be in V. Since $\mathscr{B}$ spans V, then there is at least one way to write $\mathbf{v}$ in terms of the $\mathbf{v}_i$. Suppose there are two ways:

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \quad \text{and} \quad \mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n,$$

with scalars c_i, d_i. The setting them equal and combining like terms gives,

$$(c_1 - d_1)\mathbf{v}_1 + \cdots + (c_n - d_n)\mathbf{v}_n = \mathbf{0}.$$

But now linear independence implies that $c_i = d_i$ for each i. $\square$

Definition: Coordinates

Let $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. For $\mathbf{v}$ in V, there is a unique expression of $\mathbf{v}$ as a linear combination of the $\mathbf{v}_i$: $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. We call $c_1, \ldots, c_n$ the *coordinates of* $\mathbf{v}$ *with respect to* $\mathscr{B}$. The *coordinate vector of* $\mathbf{v}$ *with respect to* $\mathscr{B}$ is defined as:

$$[\mathbf{v}]_{\mathscr{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \text{ in } \mathbb{R}^n$$

Example 3. Consider $\mathscr{P}_2$ with basis $\mathscr{B} = \{1, x, x^2\}$. Let $p(x) = 2 - 3x + 5x^2$. Then

$$[p(x)]_{\mathscr{B}} = \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}.$$

On the other hand, consider the basis $\mathscr{B}' = \{1 + x, x + x^2, 1 + x^2\}$, then

$$p(x) = -3(1 + x) + 0(x + x^2) + 5(1 + x^2).$$

(You should think about how I came up with this.) Then

$$[p(x)]_{\mathscr{B}'} = \begin{bmatrix} -3 \\ 0 \\ 5 \end{bmatrix}.$$

It is worth observing that the coordinate vector depends not only on the basis, but on the *order* of elements in the basis.

Theorem 4: Linearity of coordinates

Let $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V and let λ be a scalar. Then

a. $[\mathbf{u} + \mathbf{v}]_{\mathscr{B}} = [\mathbf{u}]_{\mathscr{B}} + [\mathbf{v}]_{\mathscr{B}}$

b. $[\lambda \mathbf{u}]_{\mathscr{B}} = \lambda [\mathbf{u}]_{\mathscr{B}}$

Proof. Write

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$
$$\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n,$$

for scalars c_i, d_i. Then

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_n + d_n)\mathbf{v}_n.$$

Taking coordinates gives the result. Similarly,

$$\lambda\mathbf{u} = (\lambda c_1)\mathbf{v}_1 + \cdots + (\lambda c_n)\mathbf{v}_n.$$

Again, taking coordinates gives the result. □

This result says that the map $V \to \mathbb{R}^n$ is a linear transformation. In fact, it is more. It is an *isomorphism*, which means in part that it takes a basis to a basis (and vice-versa).

Theorem 5: Coordinates preserve linear independence

Let $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be a set of vectors in V. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent in V if and only if $\{[\mathbf{u}_1]_{\mathscr{B}}, \ldots, [\mathbf{u}_k]_{\mathscr{B}}\}$ is linearly independent in $\mathbb{R}^n$.

Proof. Suppose $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent in V. We claim $\{[\mathbf{u}_1]_{\mathscr{B}}, \ldots, [\mathbf{u}_k]_{\mathscr{B}}\}$ is linearly independent in $\mathbb{R}^n$. Let c_i be scalars such that

$$c_1[\mathbf{u}_1]_{\mathscr{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathscr{B}} = \mathbf{0}$$

Using the linearity theorem,

$$[c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathscr{B}} = \mathbf{0}.$$

But the only vector whose coordinates are all zero is the zero vector itself in V. Hence,

$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}.$$

By linear independence of the $\mathbf{u}_i$, $c_i = 0$ for all i.

The converse is similar. Suppose $\{[\mathbf{u}_1]_{\mathscr{B}}, \ldots, [\mathbf{u}_k]_{\mathscr{B}}\}$ is linearly independent in $\mathbb{R}^n$. We claim $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent in V. Let d_i be scalars such that

$$d_1\mathbf{u}_1 + \cdots + d_k\mathbf{u}_k = \mathbf{0}.$$

Now taking coordinates on both sides gives

$$[d_1\mathbf{u}_1 + \cdots + d_k\mathbf{u}_k]_{\mathscr{B}} = [\mathbf{0}]_{\mathscr{B}} = \mathbf{0}.$$

Now again using linearity,

$$d_1[\mathbf{u}_1]_{\mathscr{B}} + \cdots + d_k[\mathbf{u}_k]_{\mathscr{B}} = \mathbf{0}.$$

So by linear independence of then $[\mathbf{u}_i]_{\mathscr{B}}$, we have $d_i = 0$ for all i. $\square$

Using coordinates, we can prove the following theorem for general (finite dimensional) vector spaces. It is easy to prove using the above result, along with a corresponding result for spanning.

Theorem 6: The Basis Theorem

If a vector space V has a basis with n vectors, then every basis for V has n vectors.

Definition: Dimension

Let V be a nonzero vector space. The *dimension* of V is the number of elements in any basis of V. The dimension of the zero vector space is defined as zero.

Example. (1) The dimension of $\mathbb{R}^n$ is n.

(2) The dimension of $\mathscr{P}_n$ is $n+1$.

(3) The dimension of $\mathcal{M}_{m,n}$ is mn.

Another application is that we can use coordinates to check whether a given subset of a vector space is a basis by checking whether the coordinates of those elements (relative to some standard basis) is a basis of $\mathbb{R}^n$.

Example. Consider $\mathscr{P}_2$ with basis $\mathscr{B} = \{1, x, x^2\}$. Consider the set $S = \{1+x, x+x^2, 1+x^2\}$. Note we have already verified that this S is a basis, but this gives another method. Taking coordinates of the elements in S, we obtain the following set in $\mathbb{R}^3$:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Let A be the matrix whose columns are these three elements. Then $\det A \neq 0$, so A is invertible. Hence, the set is a basis for $\mathbb{R}^3$. It follows that S is a basis for $\mathscr{P}_2$.

3. Change of Basis

Say we have two bases, $\mathscr{B}$ and $\mathscr{C}$, for a vector space V. Given a vector $\mathbf{v}$ in V, written in terms of the basis $\mathscr{B}$, how do we write that vector in terms of $\mathscr{C}$?

Definition: Change-of-basis matrix

Let $\mathscr{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathscr{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be bases for a vector space V. The *change-of-basis matrix* is the $n \times n$ matrix

$$P_{\mathscr{C}\leftarrow\mathscr{B}} = \begin{bmatrix} [\mathbf{u}_1]_{\mathscr{C}} & [\mathbf{u}_2]_{\mathscr{C}} & \cdots & [\mathbf{u}_n]_{\mathscr{C}} \end{bmatrix}$$

That is, the columns of $P_{\mathscr{C}\leftarrow\mathscr{B}}$ are the basis vectors of $\mathscr{B}$ written in terms of $\mathscr{C}$. The reason for the backwards arrow is clear in that for for any vector $\mathbf{x}$ in V, we have

$$P_{\mathscr{C}\leftarrow\mathscr{B}}[\mathbf{x}]_{\mathscr{B}} = [\mathbf{x}]_{\mathscr{C}}.$$

To see this, write $\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ for (unique) scalars c_i. Then

$$[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

so we have by linearity,

$$\begin{aligned} [\mathbf{x}]_{\mathscr{C}} &= [c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n]_{\mathscr{C}} \\ &= c_1[\mathbf{u}_1]_{\mathscr{C}} + \cdots + c_n[\mathbf{u}_n]_{\mathscr{C}} \\ &= \begin{bmatrix} [\mathbf{u}_1]_{\mathscr{C}} & [\mathbf{u}_2]_{\mathscr{C}} & \cdots & [\mathbf{u}_n]_{\mathscr{C}} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\ &= P_{\mathscr{C}\leftarrow\mathscr{B}}[\mathbf{x}]_{\mathscr{B}}. \end{aligned}$$

The columns of the matrix $P_{\mathscr{C}\leftarrow\mathscr{B}}$ are linearly independent, and so it is invertible. One can show without much difficulty that

$$(P_{\mathscr{C}\leftarrow\mathscr{B}})^{-1} = P_{\mathscr{B}\leftarrow\mathscr{C}}.$$

Example. Let $\mathscr{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard basis for $\mathbb{R}^2$ and let $\mathscr{C} = \left\{ \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}$. Then $\mathscr{C}$ is also a basis for $\mathscr{B}$ (check!).

We want to compute $P_{\mathscr{C}\leftarrow\mathscr{B}}$ and $P_{\mathscr{B}\leftarrow\mathscr{C}}$. But note that it is *much* easier to compute the latter: the vectors in $\mathscr{C}$ written in terms of the standard basis are just the vectors themselves! So,

$$P_{\mathscr{B}\leftarrow\mathscr{C}} = \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix}.$$

Then

$$P_{\mathscr{C}\leftarrow\mathscr{B}} = (P_{\mathscr{B}\leftarrow\mathscr{C}})^{-1} = \frac{1}{(-1)}\begin{bmatrix} 4 & -1 \\ -5 & 1 \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}.$$

Now let $\mathbf{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ (so that $\mathbf{x} = 3\mathbf{e}_1 + (-1)\mathbf{e}_2$). We want to write $\mathbf{x}$ in terms of the basis $\mathscr{C}$. We know how to do this directly (set the vectors of $\mathscr{C}$ in a matrix with augmented column $\mathbf{x}$ and row reduce). But now we have another way to do it. By the above,

$$[\mathbf{x}]_{\mathscr{C}} = P_{\mathscr{C}\leftarrow\mathscr{B}}[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -13 \\ 16 \end{bmatrix}.$$

We can check this:

$$(-13)\begin{bmatrix} 1 \\ 5 \end{bmatrix} + (16)\begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \mathbf{x}.$$

Let's see this in an example with polynomials. The problem is similar with other vector spaces.

Example. Consider the bases $\mathscr{B} = \{1, x, x^2\}$ and $\mathscr{C} = \{1+x, x+x^2, 1+x^2\}$ for $\mathscr{P}_2$. Here $\mathscr{B}$ is like the *standard basis* for $\mathscr{P}_2$. So we again first compute $P_{\mathscr{B}\leftarrow\mathscr{C}}$:

$$[1+x]_{\mathscr{B}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad [x+x^2]_{\mathscr{B}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad [1+x^2]_{\mathscr{B}} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Now,

$$P_{\mathscr{B}\leftarrow\mathscr{C}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Then

$$P_{\mathscr{C}\leftarrow\mathscr{B}} = (P_{\mathscr{B}\leftarrow\mathscr{C}})^{-1} = \begin{bmatrix} 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 \end{bmatrix}$$

Now let $p(x) = 1 + 2x - x^2$. So

$$[p(x)]_{\mathscr{C}} = P_{\mathscr{C}\leftarrow\mathscr{B}}[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}.$$

That is,

$$2(1+x) + 0(x+x^2) + (-1)(1+x^2) = 1 + 2x - x^2 = p(x).$$

The last situation we want to consider is when we have two non-standard bases.

Let $\mathscr{E}$ denote the standard basis of $\mathbb{R}^n$. Let $\mathscr{B}$ and $\mathscr{C}$ be any two bases of $\mathbb{R}^n$ (one of them, or both, could also be the standard basis but that just produces some trivialities). Suppose we want to compute $P_{\mathscr{C}\leftarrow\mathscr{B}}$. Using the definition, we could show the following.

Write $\mathscr{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathscr{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ as before. Let B be the matrix whose columns are the $\mathbf{u}_i$, and let C be the matrix whose columns are the $\mathbf{v}_i$. Then

$$[C \mid B] \longrightarrow [I \mid P_{\mathscr{C}\leftarrow\mathscr{B}}]$$

As an alternative, consider the following diagram: Consider the following diagram.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathscr{B}} \arrow[dr, "P_{\mathscr{E}\leftarrow\mathscr{B}}"'] & & \mathbb{R}^n_{\mathscr{C}} \arrow[dl, "P_{\mathscr{E}\leftarrow\mathscr{C}}"] \\
& \mathbb{R}^n_{\mathscr{E}} &
\end{tikzcd}$$

Note that both of these matrices are easy to compute. Hence, to go from $\mathscr{B}$ to $\mathscr{C}$, we need to first go from $\mathscr{B}$ to $\mathscr{E}$ and then from $\mathscr{E}$ to $\mathscr{C}$:

$$\begin{tikzcd}
\mathbb{R}^n_{\mathscr{B}} \arrow[dr, "P_{\mathscr{E}\leftarrow\mathscr{B}}"'] & & \mathbb{R}^n_{\mathscr{C}} \arrow[dl, "P_{\mathscr{E}\leftarrow\mathscr{C}}"] \\
& \mathbb{R}^n_{\mathscr{E}} \arrow[ur, dashed, bend right, "P_{\mathscr{C}\leftarrow\mathscr{E}}"'] &
\end{tikzcd}$$

That is,

$$P_{\mathscr{C}\leftarrow\mathscr{B}} = P_{\mathscr{C}\leftarrow\mathscr{E}} P_{\mathscr{E}\leftarrow\mathscr{B}} = (P_{\mathscr{E}\leftarrow\mathscr{C}})^{-1} P_{\mathscr{E}\leftarrow\mathscr{B}}.$$

Example. Consider the following bases of $\mathbb{R}^2$:

$$\mathscr{B} = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \end{bmatrix} \right\} \quad \text{and} \quad \mathscr{C} = \left\{ \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}.$$

We will compute the change of basis matrix $P_{\mathscr{C}\leftarrow\mathscr{B}}$ using the strategy above.

First note that

$$P_{\mathscr{E}\leftarrow\mathscr{B}} = \begin{bmatrix} -1 & -2 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad P_{\mathscr{C}\leftarrow\mathscr{E}} = \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix}$$

Then as we computed before:

$$(P_{\mathscr{C}\leftarrow\mathscr{E}})^{-1} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}.$$

Hence,

$$P_{\mathscr{C}\leftarrow\mathscr{B}} = (P_{\mathscr{E}\leftarrow\mathscr{C}})^{-1}P_{\mathscr{E}\leftarrow\mathscr{B}} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 7 & 13 \\ -8 & -15 \end{bmatrix}.$$

Now suppose we have a vector $\mathbf{x}$ in $\mathbb{R}^2$ such that $[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Then

$$[\mathbf{x}]_{\mathscr{C}} = P_{\mathscr{C}\leftarrow\mathscr{B}}[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} 7 & 13 \\ -8 & -15 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -19 \\ -22 \end{bmatrix}.$$

We can check:

$$[\mathbf{x}]_{\mathscr{B}} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{so} \quad \mathbf{x} = (1)\begin{bmatrix} -1 \\ 3 \end{bmatrix} + (-2)\begin{bmatrix} -2 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -7 \end{bmatrix}$$

$$[\mathbf{x}]_{\mathscr{C}} = \begin{bmatrix} -19 \\ 22 \end{bmatrix} \quad \text{so} \quad \mathbf{x} = (-19)\begin{bmatrix} 1 \\ 5 \end{bmatrix} + (22)\begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -7 \end{bmatrix}.$$