Multivariable Calculus and Linear Algebra

Arnav Nayak

June 2020

Contents

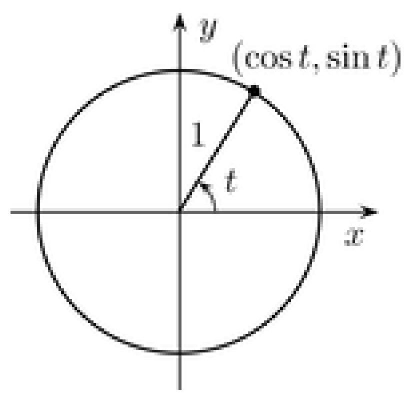
Part I

Multivariable Calculus

1 Chapter 11: Parametric Equations and Polar Coordinates

Oftentimes, we find situations where we can't describe curves in the form $y = f(x)$. Curves like these fail the vertical line test that describe a function. Instead, we can describe both the x and y coordinates of this curve as a function of a third variable t, $x = f(t)$ and $y = g(t)$. These are called parametric equations and the variable t is known as the **parameter**. We can call the curve they trace out a **parametric curve**. Certain situations where we can use parametric curves are when we trace the path of a particle with respect to time or the position of an object in space with respect to time, though the parameter does not always need to denote time.

For example, if we describe $x = cos(t)$ and $y = sin(t)$ on the interval $0 \leq t \leq 2\pi$, then we get the curve shown below:



Some curves can be expressed in the form $y = F(x)$ by eliminating the parameter. If we substitute the functions for our parametric equations in, we get the expression $g(t) = F(f(t))$. We can differentiate this expression using the chain rule, giving us the expression

$$g^{'}(t) = F^{'}(f(t))f^{'}(t) = F^{'}(x)f^{'}(t)$$
$$F^{'}(x) = \frac{g^{'}(t)}{f^{'}(t)}$$

We can now find the tangent of a parametric curve easily with

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}$$

We can derive again to find the second derivative

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}$$

We know that the infinitesimally small change along a curve, denoted by ds, is really just an infinitesimally small movement along the x and y direction. Thus,

$$\begin{aligned} ds &= idx + jdy \\ ds^2 &= ds \cdot ds \\ &= (idx + jdy) \cdot (idx + jdy) = dx^2 + dy^2 \\ ds &= \sqrt{dx^2 + dy^2} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \\ L &= \int_{\alpha}^{\beta} ds = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \end{aligned}$$

And just like that, we have the equation of **Arc Length** along a parametric curve.

We can describe another coordinate system called the **polar coordinate system**. The coordinates take the form (r, θ), where r is the length that the point is away from the coordinate and θ is the angle above the polar axis. We can convert between cartesian and polar by seeing that we can form a triangle with the polar axis and the line connecting the origin and the point. Thus,

$$\begin{aligned} x &= rcos\theta \\ y &= rsin\theta \end{aligned}$$

We can also see that

$$\begin{aligned} r^2 &= x^2 + y^2 \\ tan\theta &= \frac{y}{x} \end{aligned}$$

We know the tangent of a parametric curve. We can use that to find the equation of a polar equation as well, with the help of the product rule.

$$\frac{dy}{dx} = \frac{\frac{dy}{d\theta}}{\frac{dx}{d\theta}} = \frac{\frac{dr}{d\theta} sin\theta + rcos\theta}{\frac{dr}{d\theta} cos\theta - rsin\theta}$$

We also know that the infinitesimally small change in area of a polar curve is simply the change in r times the average arc length: $dA = dr \times \frac{(rd\theta + (r+dr)d\theta)}{2}$. Since $drd\theta$ is so small it is negligible, dA becomes $rdrd\theta$. We can integrate this to find

$$A = \int_{\alpha}^{\beta} \frac{1}{2} r^2 d\theta$$

which is the area of a section under a polar equation.

We can define some conic sections as well. An **ellipse** is the set of points in a plane the sum of whose distances from two fixed points is a constant. Its equation takes the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

which has foci $(c, 0)$ and $c^2 = a^2 - b^2$

We can also define the **hyperbola**, which is the set of points in a plane the difference of whose distances from two fixed points is a constant. Its equation is given by

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

which has focI $(c, 0)$ and $c^2 = a^2 + b^2$.

2 Chapter 12: Infinite Sequences and Series

A **sequence** is a list of numbers written in a definite order. Sequences can be described with a formula such as $\left[\frac{n}{n+1}\right]$, where the n-th term is given by substituting in n in the formula. If the **limit** $\lim_{n\to\infty} a_n = L$ exists, then we say the sequence **converges**, otherwise it **diverges**. For example, if we have the sequence $a_n = \frac{n}{n+1}$, then we can find its limit as n approaches infinity by

$$\lim_{n\to\infty} \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = \frac{\lim_{n\to\infty} 1}{\lim_{n\to\infty} 1 + \lim_{n\to\infty} \frac{1}{n}}$$
$$= \frac{1}{1+0} = 1$$

If $\lim_{x\to\infty} f(x) = L$ and $f(n) = a_n$, then $\lim_{n\to\infty} a_n = L$. This way we can relate sequences to functions.

For example, if we have the sequence $a_n = \frac{ln(n)}{n}$, then we can find its limit as it approaches infinity by recognize the related function $f(x) = \frac{ln(n)}{n}$. Thus, we can l'Hospital rule the function to find the limit of the function, which is equal to the limit of the sequence.

$$\lim_{x\to\infty} \frac{lnx}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0$$

There is also a theorem that states that if the $\lim -n \to \infty |a_n| = 0$, then $\lim -n \to \infty a_n = 0$.

For example, if we have the sequence $a_n = \frac{(-1)^n}{n}$, then we can find

$$\lim_{n\to\infty} \left|\frac{(-1)^n}{n}\right| = \lim_{n\to\infty} \frac{1}{n} = 0$$

A sequence is **increasing** if each consequent term is greater than the last. It is **decreasing** if it is smaller. It is **monotonic** if it is either increasing or decreasing. If there is a number M such that $a_n \leq M$ for any value of a_n, then the sequence is **bounded above**. If there is a number m such that $m \leq a_n$ for any value of a_n, then the sequence is **bounded below**. If a sequence is both bounded above and below, then it is a **bounded sequence**. The **monotonic sequence theorem** states that every bounded and monotonic sequence is convergent.

We can sum the terms in a sequence to obtain a **series**. A **partial sum** is essentially a sum of certain numbers of a series. For example, $s_1 = a_1$, $s_2 = a_1 + a_2$ etc. If the sequence s_n is convergent and $\lim_{n\to\infty}$ exists as a real number s, then the series s_n or Σa_n is **convergent** and s is the sum of the series. Otherwise it is **divergent**.

A **geometric series** is an infinite series that takes the form

$$a + ar + ar^2 + ar^3 + ... + ar^{n-1} = \Sigma ar^{n-1}$$

We can find the sum of a geometric series as shown below

$$\begin{aligned} s_n &= a + ar + ar^2 + ...ar^{n-1} \\ rs_n &= ar + ar^2 + ... + ar^{n-1} + ar^n \\ s_n - rs_n &= a - ar^n \\ s_n &= \frac{a(1-r^n)}{1-r} \end{aligned}$$

The sequence of this series is convergent if $|r| < 1$, because the r^n term approaches 0 as n approaches infinity only when $|r|$ is less than 1. The sum when $|r|$ is less than 1 is then $\frac{a}{1-r}$. If $|r| \geq 1$, then the series is divergent.

A specific type of series is known as the **harmonic series**. It is described by $\Sigma\frac{1}{n}$. The harmonic series diverges as shown below

$$\begin{aligned} &\Sigma\frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \\ &s_1 = 1 \\ &s_2 = 1 + \frac{1}{2} \\ &s_4 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} > 1 + \frac{1}{2} + \frac{1}{2} \\ &s_8 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + \frac{3}{2} \end{aligned}$$

We can always group the fractions to be greater than a half, which will always result in the summation of 1s, which is divergent.

If the series Σa_n is convergent, then $\lim_{n\to\infty} a_n = 0$. However, the converse is not necessarily true. The **divergence test** states that if $\lim_{n\to\infty} a_n \neq 0$, then the series must be divergent.

The **integral test** is a test where if f is a continuous, positive and decreasing function and series $a_n = f(n)$, then the series is convergent if the improper integral $\int_1^\infty f(x)dx$ is convergent. If the integral is divergent, then the series is divergent.

For example, if we have the series $a_n = \frac{2}{3x+5}$, then

$$\begin{aligned} f(x) &= \frac{2}{3x+5} \\ &= \int_1^\infty \frac{2}{3x+5}dx = \lim_{b\to\infty} \frac{2}{3}ln(3x+5)|_b^\infty \\ &= \infty \end{aligned}$$

Thus, both the integral of the function and the series diverge.

If we have the series $a_n = \frac{1}{n^p}$, otherwise known as the **p-series**, can be shown to converge if $p > 1$ and diverge if $p \leq 1$. If p¿1:

$$\int_1^\infty \frac{1}{x^p} dx = \frac{n^{1-p}}{1-p}$$

which only converges if $p > 1$.

If we suppose that a_n and b_n are series with positive terms, then if b_n is convergent and $a_n \leq b_n$ for all values of n, then a_n is also convergent. If b_n is divergent and $a_n \geq b_n$ for all values of n, then a_n is also divergent. If we have a sequence, we can compare it to a related sequence to prove convergence or divergence. This is known as the **comparison test**.

For example, if we have the series $a_n = \frac{3^n}{2+5^n}$, we can consider the series $b_n = \frac{3^n}{5^n}$, which will always be greater than a_n. The series b_n converges as it is simply a geometric series with a ration of $\frac{3}{5}$, which is less than 1. Thus, since b_n converges, a_n must converge as well.

We can extend this and use it with limits. If a_n and b_n are both series with positive terms, and if $\lim_{n\to\infty} \frac{a_n}{b_n} = c$, where $c > 0$, then both series either converge or diverge. This is the **limit comparison test**.

For example, if we have the series $a_n = \frac{\sqrt{n^3+1}}{3n^3+4n^2+2}$, then we can compare it to the function $b_n = \frac{\sqrt{n^3}}{3n^3}$

$$b_n = \frac{\sqrt{n^3}}{3n^3} = \frac{1}{3n^{3/2}}$$

$$\lim_{n\to\infty} \frac{\sqrt{n^3+1}}{3n^3+4n^2+2} \cdot 3n^{3/2} = 1$$

Thus, both series either converge or diverge. We know b_n converges as it is a p-series whose $p > 1$. Thus, a_n must also converge.

An **alternating series** is a series whose terms alternate between positive and negative. If an alternating series is decreasing and the limit of the absolute values of the series is 0, then the series is convergent. This is the **alternating series test**.

For example, if we have the series $a_n = (-1)^n \frac{1}{n!}$, then we know that the terms are decreasing. We also know that $\lim_{n\to\infty} a_n = 0$ because the terms in the denominator become significantly larger than the 1 in the numerator. Thus, the series is convergent.

If a series a_n is convergent and the series of its absolute values are also convergent, then the series is **absolutely convergent**. If the series is convergent but not absolutely convergent, then it is **conditionally convergent**. If $\lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1$, then the series is absolutely convergent. If L is greater than 1 or infinity, then the series is divergent. If it is equal to 1, then it is inconclusive. This is the **Ratio Test**.

For example, if series $a_n = (-1)^n \frac{n^3}{3^n}$. then

$$\begin{aligned}
\left|\frac{a_{n+1}}{a_n}\right| &= \left|\frac{(-1)^{n+1}(n+1)^3}{3^{n+1}} \cdot \frac{3^n}{(-1)^n n^3}\right| \\
&= \frac{(n+1)^3}{3^{n+1}} \cdot \frac{3^n}{n^3} \\
&= \frac{(n+1)^3}{3n^3} \\
\lim_{n\to\infty} \frac{(n+1)^3}{3n^3} &= \frac{1}{3}
\end{aligned}$$

Thus, the limit is less than 1 and the series is absolutely convergent.

A **Power Series** is a series of the form

$$\Sigma c_n x^n = c_0 + c_1 x + c_2 x^2 + ...$$

which can be thought of as a function of x.
A series of the form

$$\Sigma c_n (x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + ...$$

is a **power series in (x-a)** or a **power series centered about a**. For a given power series, the only three possibilities are that it converges when x=a, for all x, or a positive number R for which it converges if $|x-a| < R$ and diverges if $|x-a| > R$. This R is known as the **Radius of Convergence**. The **interval of convergence** is the interval that consists of all values of x for which the series converges. It is important to always test the endpoints of the interval for convergence as well.

We can represent functions as power series as well. If we have the function $ln(1-x)$, then

$$\begin{aligned}
-ln(1-x) &= \int \frac{1}{1-x} dx = \int (1 + x + x^2 + x^3 + ...)dx \\
&= x + \frac{x^2}{2} + \frac{x^3}{3} + ... + C = \Sigma \frac{x^n}{n} + C \\
&- ln(1) = C, C = 0 \\
ln(1-x) &= -\Sigma \frac{x^n}{n} + C
\end{aligned}$$

Thus, we can represent the function as an integral of the sum of a geometric series, which we can expand and integrate term by term to find the power series representation of a function.

If f has a power series representation about a, then its **coefficients** are given by the formula $c_n = \frac{f^{(n)}(a)}{n!}$. Thus, the power series expansion of a function is given by the form

$$
\begin{aligned}
f(x) &= \Sigma \frac{f^{(n)}(a)}{n!}(x-a)^n \\
&= f(a) + \frac{f^{'}(a)}{1!}(x-a) + \frac{f^{''}(a)}{2!}(x-a)^2 + ...
\end{aligned}
$$

This is known as the **Taylor series of the function f at a**. If a=0, then the expansion is known as a **Maclaurin Series**.

3 Chapter 13: Vectors and the Geometry of Space

The beginning of this chapter concerns the definition of a **vector** and the operations that can be done on them, such as scalar multiplication, addition, subtraction, the dot product and the cross product. These concepts are developed in the Linear Algebra section of this review packet on chapter 1. However, this section will explore some concepts and proofs.

The **dot product** is a the scalar product of two vectors, denoted by $a \cdot b$, and given by adding the product of each respective component,

$$a \cdot b = a_1 b_1 + a_2 b_2 + ...$$

It can also be found with the expression

$$|a||b| \cos \theta$$

We can find the angles that a vector makes with the positive x, y and z axes. These are known as **direction angles** and are given by finding the dot product of the vector with each unit vector. Thus,

$$\cos \alpha = \frac{\mathbf{a} \cdot \mathbf{i}}{|\mathbf{a}||\mathbf{i}|}$$
$$\cos \beta = \frac{\mathbf{a} \cdot \mathbf{j}}{|\mathbf{a}||\mathbf{j}|}$$
$$\cos \gamma = \frac{\mathbf{a} \cdot \mathbf{k}}{|\mathbf{a}||\mathbf{k}|}$$

As discussed in Chapter 1, we also can find a projection of one vector on another. We can describe this in scalar form or vector form. The **scalar projection** of vector **b** onto vector **a** is simply the component of b that falls on a, which is simply $\mathbf{b} \cos \theta$. This can be given by

$$\frac{a \cdot b}{|\mathbf{a}|}$$

To find the **vector projection of b onto a**, we take the scalar projection and multiply it by the directional unit vector along a, which is simply $\frac{\mathbf{a}}{|\mathbf{a}|}$. Thus, the vector projection is given by

$$\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|} \left(\frac{\mathbf{a}}{|\mathbf{a}|} \right)$$

The **cross product** is the vector multiplication of two vectors. This operation turns out to be very useful because the resulting vector is orthogonal to both initial vectors. If vectors **a** and **b** each have three components, then the cross product is given by

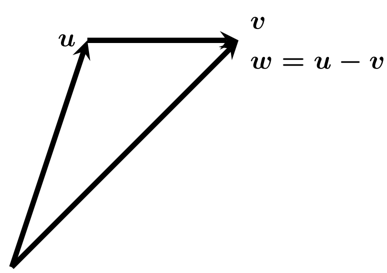
$$a \times b = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

The magnitude of the cross product is also given by the expression

$$|a||b|\sin\theta$$

The length of the cross product $\mathbf{a} \times \mathbf{b}$ is equal to the area of a parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$. Thus, the area of the triangle created by two vectors is simply half of the magnitude of the cross product.

Proof of law of cosines:
Let $\mathbf{w} = \mathbf{u} - \mathbf{v}$ and the angle between $\mathbf{u}$ and $\mathbf{v}$ to be θ.



$$\begin{aligned} w^2 &= (\mathbf{u} - \mathbf{v})^2 \\ &= u^2 + v^2 - 2\mathbf{u} \cdot \mathbf{v} \\ &= u^2 + v^2 - 2uv\cos\theta \end{aligned}$$

Proof of law of sines:

$$\begin{aligned} \mathbf{w} \times \mathbf{v} &= \mathbf{u} - \mathbf{v} \times \mathbf{v} = \mathbf{u} \times \mathbf{v} \\ \mathbf{w} \times \mathbf{u} &= \mathbf{u} - \mathbf{v} \times \mathbf{u} = \mathbf{u} \times \mathbf{v} \\ |\mathbf{w} \times \mathbf{v}| &= |\mathbf{w} \times \mathbf{u}| = |\mathbf{u} \times \mathbf{v}| \\ wu\sin\alpha &= uu\sin\beta = uv\sin\theta \\ \frac{\sin\beta}{v} &= \frac{\sin\alpha}{u} = \frac{\sin\theta}{w} \end{aligned}$$

Vectors can be utilized to describe equations of both lines and planes in different forms. For lines, we can define a specific vector known as the **normal vector**. The normal vector is the vector that is perpendicular to any vector $\mathbf{x}$ parallel to the line. Thus, $\mathbf{n} \cdot \mathbf{x} = \mathbf{0}$ as they are orthogonal. We can also define

$\mathbf{d}$ as the **direction vector** being a vector parallel to the line. Thus, $\mathbf{x}$ is really just a scalar multiple of $\mathbf{d}$. Thus, $\mathbf{x} =$p$\mathbf{d}$. For situations where the line does not pass through the origin, we must put the direction vector into standard position first by subtracting vector $\mathbf{p}$, a point on the line, from vector $\mathbf{x}$. Thus, $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = \mathbf{0}$ or $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$ This describes the **normal form** of the equation of a line.

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

The normal vector can be found from the **general form** of the equation of a line.

$$a\mathbf{x} + b\mathbf{y} = c$$
$$\mathbf{n} = \begin{bmatrix} a \\ b \end{bmatrix}$$

The **vector form** of the equation of a line simply stems from the definition of $\mathbf{x}$. The equations corresponding to the components of the vector form are called **parametric equations**.

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

For example, let l be a line in $\mathbb{R}^3$ passing through the point P = (1,2,-1) and parallel to the vector $\mathbf{d} = \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}$. Then,

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

(Vector Form)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + t \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}$$

(Parametric Form)

$$x = 1 + 5t$$
$$y = 2 - t$$
$$z = -1 + 3t$$

Another example. If we let $7\mathbf{x} + 3\mathbf{y} = 19$, then we can take any arbitrary point P = (1, 4)

$$\mathbf{n} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

(Normal Form)

$$\begin{bmatrix} 7 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

The same derivations can be done with the equation of a plane as well. If we let $ax + by + cz = d$ describe the general form of a plane and $\mathbf{p}$ be a specific point on the plane, then the normal form of the equation of a plane is given by

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

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

$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

Because of the second dimension, two direction vectors that are not parallel to each other are required to describe the vector form of a plane. If we let $\mathbf{u}$ and $\mathbf{v}$ be those direction vectors, then the vector form of a plane is given by the following. Again, the parametric equations are simply the equations of the corresponding components.

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

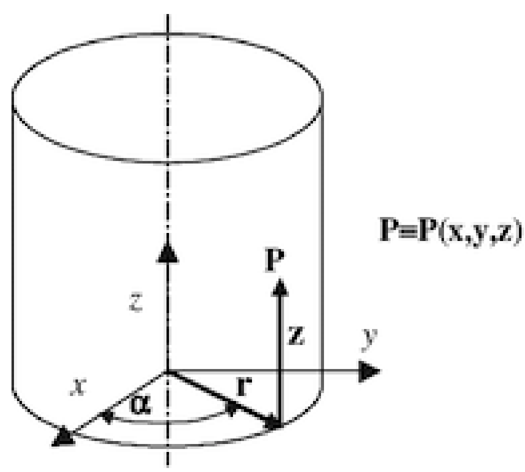
For example, if we have a plane that contains the point P = (5,7,3) and normal vector $\mathbf{n} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, then

$$\mathbf{n}\cdot\mathbf{p} = 1(5) + 2(7) + 3(3) = 5 + 14 + 9 = 28$$
$$\mathbf{n}\cdot\mathbf{x} = x + 2y + 3z = 28$$

We can find two other points on the plane to get two direction vectors.

$$Q = (3,2,7) \text{ and } R = (2,1,8)$$
$$\mathbf{u} = \begin{bmatrix} 3-5 \\ 2-7 \\ 7-3 \end{bmatrix} = \begin{bmatrix} -2 \\ -5 \\ 4 \end{bmatrix}$$
$$\mathbf{v} = \begin{bmatrix} 2-5 \\ 1-7 \\ 8-3 \end{bmatrix} = \begin{bmatrix} -3 \\ -8 \\ 5 \end{bmatrix}$$
$$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$$
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 3 \end{bmatrix} + s\begin{bmatrix} -2 \\ -5 \\ 4 \end{bmatrix} + t\begin{bmatrix} -3 \\ -8 \\ 5 \end{bmatrix}$$

We can define the **cylindrical coordinate system** as a three dimensional space where points are represented by (r, θ, z), where r is the magnitude of the length from the origin to the point, θ is the angle above the positive x-axis and z is the distance from the point to the xy-plane. We can use the following to convert between the cylindrical coordinate system and cartesian coordinate system.

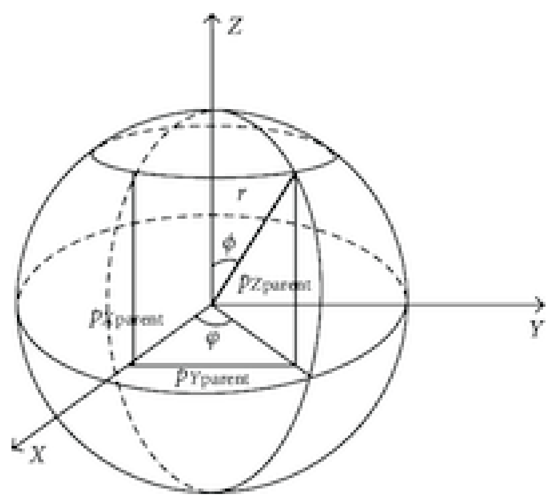


$$x = r\cos\theta$$
$$y = r\sin\theta$$
$$z = z$$

and to convert back, we can use

$$r^2 = x^2 + y^2$$
$$\tan\theta = \frac{y}{x}$$
$$z = z$$

The **spherical coordinate system** is another coordinate system that is useful for when we describe an object that is symmetrical about a point. Points are described by (ρ, θ, ϕ), where ρ is the length between origin and the point, θ is the angle above the positive x axis, and ϕ is the angle between the z axis and the point.



We can use the following to convert from spherical to cartesian

$$x = \rho\sin\phi\cos\theta$$
$$y = \rho\sin\phi\sin\theta$$
$$z = \rho\cos\phi$$

The distance formula shows us that

$$\rho^2 = x^2 + y^2 + z^2$$

4 Chapter 14: Vector Functions

A **vector function** is a function is a set of real numbers and whose range is a set of vectors. The components are determined with **component functions** and the vector function can be represented as such

$$\mathbf{r}(t) =< f(t), g(t), h(t) >$$

We can find the derivative of a vector function r' by taking the derivative of each of the component functions. By definition, it is the same as taking the derivative of a regular curve, where you find the secant line between two points and take the limit as the interval approaches 0. Doing this with a vector function gives you a vector that lies on the tangent line. Thus, the derivative of a vector function is the **tangent vector**. The **unit tangent vector** is given by:

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}$$

For example, if $\mathbf{r}(t) = (1 + t^3)\mathbf{i} + te^{-t}\mathbf{j} + \sin 2t\mathbf{k}$, then its derivative is

$$\mathbf{r}'(t) = 3t^2\mathbf{i} + (1 - t)e^{-t}\mathbf{j} + 2\cos 2t\mathbf{k}$$

and the unit tangent vector at t=0 is

$$\begin{aligned} \mathbf{r}(0) &= \mathbf{i} \\ \mathbf{r}'(0) &= \mathbf{j} + 2\mathbf{k} \\ \mathbf{T}(0) &= \frac{\mathbf{j} + 2\mathbf{k}}{\sqrt{1+4}} = \frac{\mathbf{j}}{\sqrt{5}} + \frac{2\mathbf{k}}{\sqrt{5}} \end{aligned}$$

The definite integral of a vector function works in a similar manner, by taking the definite integral of each of the component function. The result is still a vector.

When the unit tangent vector of a vector function changes slowly, the curve is fairly straight, but when the tangent vector changes direction quickly, the curve bends more sharply. Thus, we can define the **curvature** of a curve to be the rate of change of the unit tangent vector with respect to arc length.

$$\kappa = \left| \frac{d\mathbf{T}}{ds} \right|$$

We can find curvature with respect to the parameter t instead of arc length by using chain rule and dividing

$$\frac{d\mathbf{T}}{dt} = \frac{d\mathbf{T}}{ds}\frac{ds}{dt}$$
$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{T}/dt}{ds/dt}$$
$$\kappa = \left|\frac{d\mathbf{T}}{ds}\right| = \left|\frac{d\mathbf{T}/dt}{ds/dt}\right|$$
$$= \left|\frac{\mathbf{T}'(t)}{\mathbf{r}'(t)}\right|$$

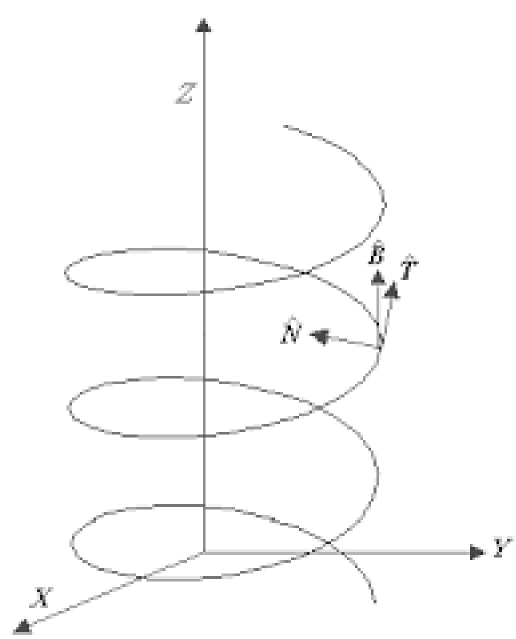
We know that $\mathbf{T}'(t)$ must be orthogonal to $\mathbf{T}(t)$, thus their dot product is 0. We can define the **unit Normal vector** as

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}$$

Geometrically, this is a unit vector pointing to the "center" of a curve, perpendicular to the tangent vector. The **Binormal vector** is the cross product of both the tangent and normal vector, resulting in a vector that is perpendicular to both vectors, usually pointing in or out of the page on a space curve. It is given by

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t)$$

Visually, all these vectors look like this on a space curve



We can derive another expression for curvature as well.

$$\begin{aligned}
\mathbf{T}(t) &= \frac{\mathbf{r}^{'}(t)}{|r^{'}(t)|} \\
\mathbf{r}^{'}(t) &= \mathbf{T}(t)|r^{'}(t)| \\
\mathbf{v} &= \mathbf{T}v \\
\frac{d\mathbf{v}}{dt} &= \frac{d(\mathbf{T}v)}{dt} = \mathbf{T}^{'}v + \mathbf{T}v^{'} \\
\mathbf{r}^{'} \times \mathbf{r}^{''} &= \mathbf{v} \times \mathbf{v}^{'} = \mathbf{v} \times (\mathbf{T}^{'}v + \mathbf{T}v^{'}) \\
\mathbf{v} \times \mathbf{T} &= 0 \\
\mathbf{r}^{'} \times \mathbf{r}^{''} &= v\mathbf{v} \times \mathbf{T}^{'} \\
\frac{|\mathbf{r}^{'} \times \mathbf{r}^{''}|}{|\mathbf{r}^{3}|} &= \frac{|\mathbf{v} \times \mathbf{T}^{'}|}{v^{2}} = \frac{|v\mathbf{T} \times \mathbf{N}|\mathbf{T}^{'}||}{v^{2}} \\
\mathbf{B}\mathbf{T} \times \mathbf{N}, &|B| = 1 \\
&= \frac{v|\mathbf{T}^{'}|}{v^{2}} = \frac{|\mathbf{T}^{'}|}{v} = \kappa
\end{aligned}$$

Thus,

$$\kappa = \frac{|\mathbf{r}^{'} \times \mathbf{r}^{''}|}{|\mathbf{r}^{3}|}$$

If we represent $\mathbf{r}(t)$ as the vector function for the path of the particle, the its derivative is the velocity vector and its second derivative is the acceleration vector function, similar to regular functions. We can resolve acceleration into two components, one in the direction of the tangent and the other in the direction of the normal (centripetal).

$$\begin{aligned}
\mathbf{T} &= \frac{\mathbf{r}^{'}}{r^{'}} = \frac{\mathbf{v}}{v} \\
\mathbf{v} &= v\mathbf{T} \\
\mathbf{a} &= \mathbf{v}^{'} = v^{'}\mathbf{T} + v\mathbf{T}^{'} \\
\kappa &= \frac{|\mathbf{T}^{'}|}{|\mathbf{r}^{'}|} = \frac{|\mathbf{T}^{'}|}{v} \\
|\mathbf{T}^{'}| &= \kappa v \\
\mathbf{N} &= \frac{\mathbf{T}^{'}}{|\mathbf{T}^{'}|} \\
\mathbf{T}^{'} &= |\mathbf{T}^{'}|\mathbf{N} = \kappa v\mathbf{N} \\
\mathbf{a} &= \mathbf{v}^{'} = v^{'}\mathbf{T} + \kappa v^{2}\mathbf{N}
\end{aligned}$$

We can find the unit vectors of the coordinate systems defined in the last chapter in terms of the unit vectors of the cartesian coordinate system. We can start with the conversions between cartesian and cylindrical systems.

$$
\begin{aligned}
x = r\cos\theta, y = r\sin\theta, z = z \\
dx &= \cos\theta dr - r\sin\theta d\theta \\
dy &= \sin\theta dr + \cos\theta d\theta \\
dz &= dz \\
ds &= dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k} \\
&= (\cos\theta dr - r\sin\theta d\theta)\mathbf{i} + (\sin\theta dr + \cos\theta d\theta)\mathbf{j} + dz\mathbf{k} \\
&= (\mathbf{i}\cos\theta + \mathbf{j}\sin\theta)dr + (-\mathbf{i}\sin\theta + \mathbf{j}\cos\theta)rd\theta + (\mathbf{k})dz \\
&= \mathbf{e_r}dr + \mathbf{e}_\theta rd\theta + \mathbf{e_k}dz \\
\mathbf{e_r} &= \mathbf{i}\cos\theta + \mathbf{j}\sin\theta \\
\mathbf{e}_\theta &= -\mathbf{i}\sin\theta + \mathbf{j}\cos\theta \\
\mathbf{e_k} &= \mathbf{k}
\end{aligned}
$$

We can do the same exercise with spherical coordinates

$$
\begin{aligned}
x &= \rho\sin\phi\cos\theta, y = \rho\sin\phi\sin\theta, z = \rho\cos\phi \\
dx &= \sin\phi\cos\theta d\rho - \rho\sin\phi\sin\theta d\theta + \rho\cos\phi\cos\theta d\phi \\
dy &= \sin\phi\sin\theta d\rho - \rho\sin\phi\cos\theta d\theta + \rho\cos\phi\sin\theta d\phi \\
dz &= \cos\phi d\rho - \rho\sin\phi d\phi \\
ds &= dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k} \\
ds &= (sin\phi\cos\theta d\rho - \rho\sin\phi\sin\theta d\theta + \rho\cos\phi\cos\theta d\phi)\mathbf{i} \\
&\quad + (\sin\phi\sin\theta d\rho - \rho\sin\phi\cos\theta d\theta + \rho\cos\phi\sin\theta d\phi)\mathbf{j} \\
&\quad + (\cos\phi d\rho - \rho\sin\phi d\phi)\mathbf{k} \\
ds &= (\mathbf{i}\sin\phi\cos\theta + \mathbf{j}\sin\phi\sin\theta + \mathbf{k}\cos\phi)d\rho \\
&\quad + (\mathbf{i}\cos\phi\cos\theta + \mathbf{j}\cos\phi\sin\theta - \mathbf{k}\sin\phi)\rho d\phi \\
&\quad + (-\mathbf{i}\sin\theta + \mathbf{j}\cos\theta)\rho\sin\phi d\theta \\
\mathbf{e}_\rho &= \mathbf{i}\sin\phi\cos\theta + \mathbf{j}\sin\phi\sin\theta + \mathbf{k}\cos\phi \\
\mathbf{e}_\phi &= \mathbf{i}\cos\phi\cos\theta + \mathbf{j}\cos\phi\sin\theta - \mathbf{k}\sin\phi \\
\mathbf{e}_\theta &= -\mathbf{i}\sin\theta + \mathbf{j}\cos\theta
\end{aligned}
$$

5 Chapter 15: Partial Derivatives

There are many scenarios where a function would depend on not only one variable, but two variables. We can think of the temperature as a function of two variables, the x and y position of the space we are considering. Or the volume as a function of both the radius and the height of a cylinder. Thus, the **function of two variables** is simply a rule that assigns a real ordered pair (x,y) in a set D a unique real number denoted by $f(x, y)$. Both x and y are the independent variables and $f(x, y)$, or z, is the dependent variable.

For example, if we wanted to find the domain and range of the multivariable function

$$f(x, y) = \frac{\sqrt{x + y + 1}}{x - 1}$$

we can see that there is no expression for f if the denominator is 0 or the quantity under the square root is nonnegative. Thus,

$$D = (x, y)|x + y + 1 \geq 0, x \neq 1$$

The **level curves** of a function of two variables are the curves with equations $f(x, y) = k$, where k i s a constant.

We can translate the notion of the limit to a multivariable function. We say that the **limit of f(x, y) as (x, y) approaches (a, b)** is denoted by

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

If we find the limit along one path to be a different value than the limit along another path, then the limit does not exist. For example, if we have the function

$$f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$$

then the limit as we approach (0, 0) from the x-axis is simply when y=0, then

$$\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2} = \frac{x^2}{x^2} = 1$$

whereas if we take the limit from the y-axis when x=0, then

$$\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2} = \frac{-y^2}{y^2} = -1$$

Since the values are not equal, the limit as the function approaches (0, 0) does not exist.

We can use the squeeze theorem to our advantage when finding limits of functions. If we define a function as

$$\frac{xy}{\sqrt{x^2 + y^2}}$$

then the limit of this function as it approaches zero can be found with the squeeze theorem. We know that

$$0 \leq \left| \frac{xy}{\sqrt{x^2 + y^2}} \right| \leq |x|, |y|$$
$$\lim_{x \to 0} |x| = 0$$
$$\lim_{y \to 0} |y| = 0$$

Thus, the limit as the function approaches (0, 0) is also 0.

The notion of the derivative carries to multivariable functions as well. However, the derivative at a certain point can be one of many tangents. For example, if there is a function of both x and y, then we can take the derivative of that function with respect to x or with respect to y. This is known as the **partial derivative**. It can be defined by

$$f_x(x, y) = \frac{f(x + h, y) - f(x, y)}{h}$$
$$f_y(x, y) = \frac{f(x, y + h) - f(x, y)}{h}$$

The partial derivative is also denoted by f_x, $\frac{\partial f}{\partial x}$, or $\frac{\partial}{\partial x} f(x, y)$

To find f_x, we hold y as a constant and differentiate with respect to x. To find f_y, we hold y as a constant and differentiate with respect to y.

For example, if $f(x, y) = x^3 + x^2y^3 - 2y^2$, then

$$f_x = 3x^2 + 2xy^3$$
$$f_y = 3y^2x^2 - 4y$$

We can see that implicit differentiation works in a similar manner if z is defined as a function of x and y. If $x^3 + y^3 + z^3 + 6xyz = 1$, and we differentiate with respect to x, then

$$3x^2 + 3z^2 \frac{\partial z}{\partial x} + 6yz + 6xy \frac{\partial z}{\partial x} = 0$$
$$\frac{\partial z}{\partial x} = -\frac{x^2 + 2yz}{z^2 + 2xy}$$

We can form **partial differential equations** to express physical laws. For example, the partial differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

is called the **Laplace's Equation**. Its solutions are called the harmonic functions. Another differential equation is

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$$

is called the **wave equation**

We can use what we know to find the **tangent plane** of a point on a multivariable curve. The tangent plane on a point $P(x_o, y_o, z_o)$ is

$$z - z_o = f_x(x_o, y_o)(x - x_o) + f_y(x_o, y_o)(y - y_o)$$

We can find differentials in a similar manner. In a regular function, the differential dx is the independent variable representing an infinitesimally small increment or decrement in the x direction. The differential dy is given by $dy = f^{'}(x)dx$. The differential dz in a multivariable function is given by

$$dz = f_x(x, y)dx + f_y(x, y)dy = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

We can prove the tangent plane expression with differentials

$$\begin{aligned} dx &= x - x_o \\ dy &= y - y_o \\ dz &= z - z_o \\ dz &= \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \\ z - z_o &= f_x(x_o, y_o)(x - x_o) + f_y(x_o, y_o)(y - y_o) \end{aligned}$$

We can translate the notion of the **chain rule** to multivariable functions as well. Suppose that $z = f(x, y)$ is a differentiable function of x and y and $x = g(t)$ and $y = h(t)$. Thus, z is also a differentiable function of t and its derivative is given by

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

However, this only holds true if x and y are differentiable functions of 1 variable. If we assume that u is a differentiable function of n variables and each of the n variables is a differentiable function of m variables, then the general form of the chain rule is

$$\frac{\partial u}{\partial t_i} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_i} + ... + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

For example, if $w = f(x, y, z, t)$ and $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, $t = t(u, v)$, then

$$\begin{aligned} \frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial u} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial u} \\ \frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial v} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial v} \end{aligned}$$

We know that the partial derivative with respect to x gives us the rate of change in the x- direction, along the unit vector **i**. The same is true with the partial derivative with respect to the y direction, along unit vector **j**. If we want to find the rate of change along any unit vector $\mathbf{u} =< a, b >$, then we use the **directional derivative** of the function

$$D_u f(x_o, y_o) = \lim_{h \to 0} \frac{f(x_o + ha, y_o + hb) - f(x_o, y_o)}{h}$$

The directional derivative is also given by

$$D_u f(x, y) = f_x(x, y)a + f_y(x, y)b$$
$$D_u f(x, y) =< f_x(x, y), f_y(x, y) > \cdot \mathbf{u}$$

The first vector of the dot product is known as the **gradient** of f and is denoted by ∇f, which is read as del f. More formally, del f is given by

$$\nabla f(x, y) =< f_x(x, y), f_y(x, y) >= \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$
$$D_u f(x, y) = \nabla f(x, y) \cdot \mathbf{u}$$

The maximum value of the directional derivative is given by $|\nabla f(\mathbf{x})|$.

We can utilize the **Second Derivatives Test** to test if a point is a **local maximum** or a **local minimum** on a multivariable function. Suppose $f_x(a, b) = 0 and f_y(a, b) = 0$, denoting that they are **critical points**. Let

$$D = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$$

If $D > 0$ and $f_{xx} > 0$, then $f(a, b)$ is a local minimum. If $D > 0$ and $f_{xx} < 0$, then $f(a, b)$ is a local maximum. If $D < 0$, then $f(a, b)$ is neither a local maximum or minimum. This is known as a **saddle point** and it is where the graph of f crosses its tangent plane.

We can use the idea of a **lagrange multiplier** to maximize or minimize a general function if it's subject to a constraining function. If the function is $f(x, y, z)$ and it is subject to a constraint $g(x, y, z) = k$, we find all the values of x, y, z and λ such that

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$$

We solve for the solutions of $f_x = \lambda g_x$, $f_y = \lambda g_y$, $f_z = \lambda g_z$. The maximum solution is the maximum solution and the minimum solution is the minimum solution.

For example, if a rectangular box of square area $12m^2$ is to be made without a square lid and we need to find the maximum value of such a box, then we can use lagrange multipliers to find the maximum solution. We let

$$V = xyz$$

subject to the constraint

$$g(x, y, z) = 2xz + 2yz + xy = 12$$

We look for values of x,y,z and λ such that $\nabla V = \lambda g$. This gives the equations $V_x = \lambda g_x$, $V_y = \lambda g_y$, $V_z = \lambda g_z$ and $2xz + 2yz + xy = 12$, which become

$$\begin{aligned} yz &= \lambda(2z + y) \\ xz &= \lambda(2z + x) \\ xy &= \lambda(2x + 2y) \\ 2xz + 2yz + xy &= 12 \end{aligned}$$

We solve the system of equations to get

$$\begin{aligned} 2xz + xy &= 2yz + xy \\ 2yz + xy &= 2xz + 2yz \\ x =& y = 2z \\ 4z^2 +& 4z^2 + 4z^2 = 12 \\ z = 1&, x = 2, x = 2 \end{aligned}$$

6 Chapter 16: Multiple Integrals

We can think of the area under a multivariable function in the same way we think of the area under a curve. We can use Riemann Sums in a similar manner, except we think of the "volume" under the curve as a summation of rectangular prisms of equal area. The domain of the multivariable function is simply the total area of these "Riemann Prisms". Because of the varying heights of the multivariable function, we divide the domain into subrectangles of equal area. Finding the height of each respective subrectangle and multiplying it by the area will result in the volume of that subprism. We add the volumes of each subprism to get a volume under a curve. The height in question is simply the function of the two points anywhere within that subrectangle. Thus, the **double integral** of f over the rectangle R is

$$\int\int_R f(x,y)dA = \lim_{m,n\to\infty} \Sigma_{i=1}^m \Sigma_{j=1}^n f(x_{ij}, y_{ij})\Delta A$$

Thus, the volume of the solid that lies above the rectangle R and below the surface is

$$V = \int\int_R f(x,y)dA$$

As for the sample point used in the Riemann sum, we can use any of the single variable counterparts - midpoint, trapezoidal etc. For example, if we want to estimate the value of

$$\int\int sin(x+y)dA$$

using a riemann sum with m=n=2 and a rectangle of dimension $R = [0,\pi] \times [0,\pi]$, then we simply split up our rectangle into subrectangles, each side composed of 2 smaller rectangles. We find the area of each rectangle to be $\pi^2/4$. Our sample points using the midpoint rule are

$$(\frac{\pi}{4}, \frac{\pi}{4}), (\frac{\pi}{4}, \frac{3\pi}{4}), (\frac{3\pi}{4}, \frac{\pi}{4}), (\frac{3\pi}{4}, \frac{3\pi}{4})$$

We can then estimate the volume of the solid under the curve to be the sum of the function of our sample points times the area of each subrectangle:

$$[sin(\frac{\pi}{4}+\frac{\pi}{4}) + sin(\frac{\pi}{4}+\frac{3\pi}{4}) + sin(\frac{3\pi}{4}+\frac{\pi}{4}) + sin(\frac{3\pi}{4}+\frac{3\pi}{4})] \times \frac{\pi^2}{4}$$
$$= 0$$

Note that the area below the xy plane is considered negative volume and can cancel out the volume above the xy plane.

To compute an **iterated integral**, we can work from the inside out. If we are computing an expression

$$\int_c^d \int_a^b f(x,y)dxdy$$

We can compute

$$\int_c^d \left[\int_a^b f(x,y)dx\right] dy$$

holding y constant. Then, we compute the outer integral holding x constant. For example if we have the expression

$$\begin{aligned}\int_0^3 \int_1^2 (x^2y)dydx \\ = \int_0^3 \left[x^2\frac{y^2}{2}\right] dx \\ = \int_0^3 \left[\frac{3x^2}{2}\right] dx \\ = \left[\frac{x^3}{2}\right] \\ = \frac{27}{2}\end{aligned}$$

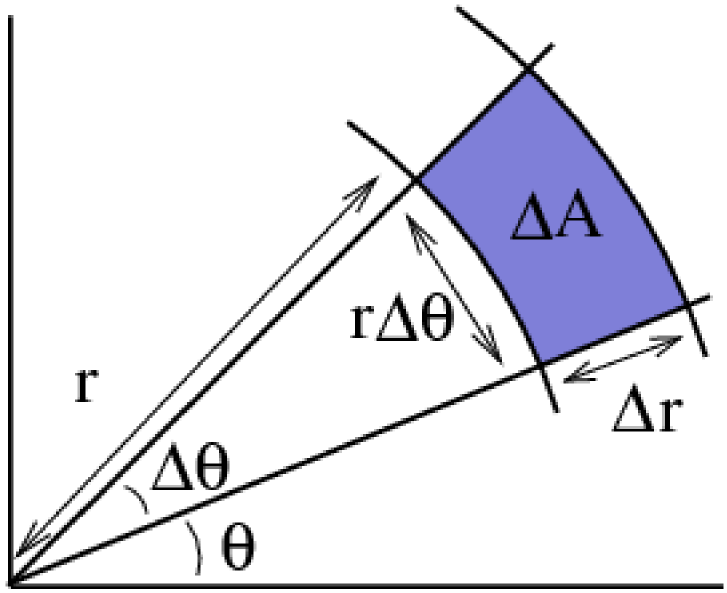
The order in which you integrate the iterated integral does not matter as long as f is bounded on R and the iterated integrals exist. This is **Fubini's theorem**. The region we find our volume over does not always have to be rectangles. If the region we're finding the volume above is a region D subtended by the two continuous functions of x, then the region is known as a **type I** region and its volume is found as follows

$$\int\int_D f(x,y)dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)dydx$$
$$D = [(x,y)|a \leq x \leq b, g_1(x) \leq y \leq g_2(x)]$$

If the region D is subtended by two continuous functions of y instead, then the region is a **type II** region and its volume is given by

$$\int\int_D f(x,y)dA = \int_c^d \int_{h_1(x)}^{h_2(x)} f(x,y)dxdy$$
$$D = [(x,y)|c \leq y \leq d, h_1(x) \leq x \leq h_2(x)]$$

We can find the volume over circular regions easier if we use a change in area in polar coordinates instead of cartesian. To change the function from cartesian to polar, we simply substitute $r\cos\theta$ for x and $r\sin\theta$ for y. The change in area must also be converted in terms of polar coordinates since $dxdy$ no longer represents the change in area in polar coordinates.



We see that the change in area can be thought of the area of a trapezoid with height dr, top side $rd\theta$ and bottom side $(r+dr)d\theta = rd\theta$. Note that $drd\theta$ is so small it's insignificant. The change in area is then $rdrd\theta$. Thus,

$$\int\int_R f(x,y)dA = \int_\alpha^\beta \int_a^b f(rcos(\theta), rsin(\theta))rdrd\theta$$

Previously, we could only find the center of mass or moment of inertias of a surface with a constant density with a single integral. Now, with double integrals, we can find the same with laminas of variable density in a 3D plane. If the density is a function of two variables, x and y, then $\rho(x,y)$ is the mass over unit area. To find this, we can divide the lamina into subrectangles and take the limit of the change in mass over the change in area as the dimensions of that subrectangle approach 0. Thus, the change in mass of one subrectangle is the density at a point within that subrectangle times the change in area. If we add up all of these change in masses, we get the **mass** of the entire lamina.

$$m = \int\int \rho(x,y)dA$$

The **moment about the x axis** is simply the mass times the distance to the x axis, which is the y value. The **moment about the y axis** is the same times x.

$$M_x = \int\int y\rho(x,y)dA$$
$$M_y = \int\int x\rho(x,y)dA$$

The center of mass, having coordinates (x_{center}, y_{center}), is defined so that $mx_{center} = M_y$ and $my_{center} = M_x$

$$x_{center} = \frac{M_y}{m} = \frac{1}{m}\int\int x\rho(x,y)dA$$
$$y_{center} = \frac{M_x}{m} = \frac{1}{m}\int\int y\rho(x,y)dA$$

The **moment of inertia** is defined as mr^2. In the case with a lamina of a variable density the moment of inertia about the x axis is

$$I_x = \int\int y^2\rho(x,y)dA$$

and the moment of inertia about the y axis is

$$I_y = \int\int x^2\rho(x,y)dA$$

The moment of inertia about the origin I_o is simple the sum of I_x and I_y.

We know that the infinitesimally small change in length in cartesian coordinates is

$$ds = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$$

To find the surface area of a curve, we can divide the region under the curve into subrectangles. If we take a sample point in any one of these subrectangles and find the point on the curve that corresponds to this sample point, then we can find the tangent plane at this point on the curve. We approximate the area of this tangent plane to the surface area of the small section on the curve. If we add all of these areas, we can get the equation of the surface area of a multivariable curve.

To find the area of the extremely small tangent plane, we can find the magnitude of the cross product between the two vectors that compose the tangent plane. The vectors that compose the tangent plane at that point is simply the partial derivative with respect to x times the change in x and the partial derivative with respect to y times the change in y.

$$\begin{aligned} dS &= \left(\frac{\partial s}{\partial x}dx\right) \times \left(\frac{\partial s}{\partial y}dy\right) \\ &= \left(\mathbf{i}dx + \mathbf{k}\frac{\partial z}{\partial x}dx\right) \times \left(\mathbf{j}dy + \mathbf{k}\frac{\partial z}{\partial y}dy\right) \\ &= \left(\mathbf{i} + \mathbf{k}\frac{\partial z}{\partial x}\right) \times \left(\mathbf{j} + \mathbf{k}\frac{\partial z}{\partial y}\right) dxdy \\ &= \left(\mathbf{k} - \mathbf{j}\frac{\partial z}{\partial y} - \mathbf{i}\frac{\partial z}{\partial x}\right) dxdy \\ \sqrt{dS \cdot dS} &= \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \end{aligned}$$

Just like there are single integrals for a function of one variable and double integrals for functions of two variables, there exists **triple integrals** for functions of three variables, which becomes harder to visualize. However, the same concepts are used. We can think the shape as a 3D plotted over a region D on the xy plane. To find the volume of this shape, we can divide the shape into subprisms and utilize a triple Riemann Sum, adding up all of the little subprisms, in order to find the volume. Thus, using the same concept,

$$\int\int\int f(x,y,z)dV$$

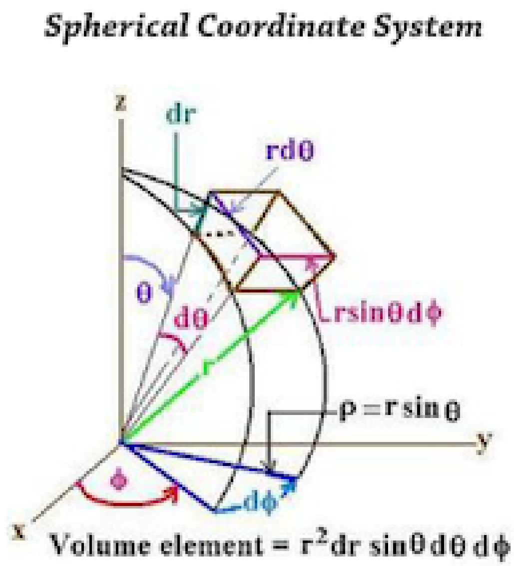
where dV is the infinitesimally small change in volume that would depend on the coordinate system being used.
In the cartesian system, if the solid region is subtended between two functions in the x direction and two functions in the y direction, then the volume of the solid becomes

$$\int\int\int f(x,y,z)dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)dzdydx$$

To find the same formula in the cylindrical coordinate plane, we figure out dV in that plane in the same way we found dA previously. dV is simply the same as dA multiplied by the change in the z direction. Thus, $dV = rdzdrd\theta$. Thus, in cylindrical coordinates the formula becomes

$$\int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(rcos\theta,rsin\theta)}^{u_2(rcos\theta,rsin\theta)} f(rcos\theta, rsin\theta, z)rdzdrd\theta$$

To convert the formula into spherical coordinates, we must find the dV in that system, which looks a little bit harder. We can visualize the change in volume as below



The little spherical wedge at the end is composed of three dimensions, dr, $rd\theta$, and $rsin\theta d\phi$. The third element can be thought of a leg of magnitude $rsin\theta$ moving through an angle ϕ, creating an arc. Thus, our dv is $r^2 \sin\theta dr d\theta d\phi$. Thus, the formula becomes

$$\int_c^d \int_\alpha^\beta \int_a^b f(\rho sin\phi cos\theta, \rho sin\phi sin\theta, \rho cos\phi)\rho^2 sin\phi d\rho d\phi d\theta$$

For any general change of variable transformation, we can find the extraneous terms in addition to the change in each direction using the **jacobian**. The jacobian of the transformation given by $x = g(u, v)$ and $y = h(u, v)$ is

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

Thus, the change in variables from cartesian to cylindrical can be thought of

$$\int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(rcos\theta, rsin\theta)}^{u_2(rcos\theta, rsin\theta)} f(rcos\theta, rsin\theta, z) J dz dr d\theta$$

where J is the jacobian of the transformation, which in this case turns out to be r.

The change in variables from cartesian to spherical can be thought of

$$\int_c^d \int_\alpha^\beta \int_a^b f(\rho sin\phi cos\theta, \rho sin\phi sin\theta, \rho cos\phi) J d\rho d\phi d\theta$$

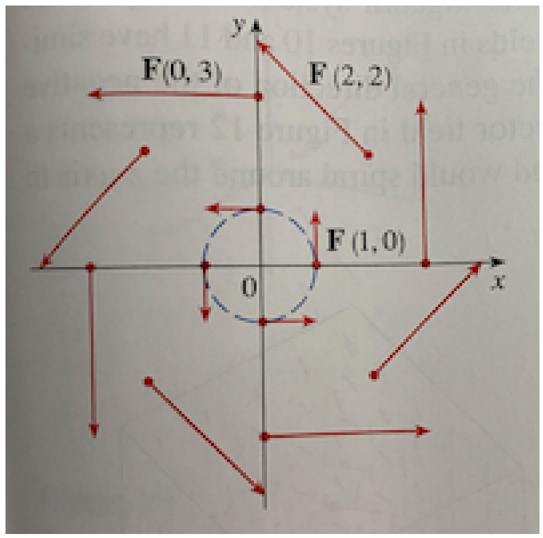
where J is the jacobian of the transformation, which in this case turns out to be $\rho^2 sin\phi$.

7 Chapter 17: Vector Calculus

A **vector field** is a function that assigns to each point (x, y) in D to a two dimensional vector F(x,y). For example, a vector field can take the form

$$F(x, y) = -y\mathbf{i} + x\mathbf{j}$$

and graphically, this looks like this



Vector fields like these can be used to represent any phenomena, ranging from an electric field from a charge to wind speeds in a 3d plane. The del operator we defined earlier is actually a vector field known as the **gradient vector field** because it takes the form

$$\nabla f(x, y) = f_x(x, y)\mathbf{i} + f_u(x, y)\mathbf{j}$$

We can define another type of integral. Instead of integrating over an interval $[a, b]$, we can integrate over a curve C. This is known as a **line integral**. We can figure out the line integral by dividing the "curtain" under the curve into subrectanges. This looks like small sub arcs on the xy plane. If we take a sample point, representing the height, and multiply it by the change in length, we get the area of one small "subcurtain". We can add up all of these to get the area of the "curtain" that a curve traverses through. It is given by

$$\int_C f(x, y)ds$$

If the curve is parametrized, then the change in length can be thought of as

$$ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

and the line integral becomes

$$\int_C f(x(t),y(t))\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}dt$$

Instead of finding the line integral with respect to the arc length, we can also find the line integral with respect to x and y. They are simply found with the same process, except we multiply the sample point times the change in the x or y direction

$$\int_C f(x,y)dx = \int_b^a f(x(t),y(t))x^{'}(t)dt$$
$$\int_C f(x,y)dy = \int_b^a f(x(t),y(t))y^{'}(t)dt$$

If there exists a continuous vector field defined on a smooth curve C given by the vector function **r**(t), then the **line integral of F along C** is given by

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t))\cdot \mathbf{r}^{'}(t)dt = \int_C \mathbf{F}\cdot T ds$$

The **fundamental theorem for line integrals** states that

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a))$$

We know this to be true because

$$\int_C \nabla f \cdot d\mathbf{r}$$
$$= \int \left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz\right)$$
$$= \int (df) = f(b) - f(a)$$

If we assume that a vector field is continuous on an open connected region exists, and if $\int_C F\cdot dr$ is independent of path, then **F** is a conservative vector, and there exists a function f such that $\nabla f = \mathbf{F}$

To determine if a vector field $F(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$ is conservativem then

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

Green's Theorem gives the relationship between a line integral around a simple closed curve C and a double integral over the plane region D bounded by C.

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \int_C Pdx + Qdy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dA$$

We can prove it in its vector form

$$
\begin{aligned}
\oint F \cdot nds &= \int F \cdot n \frac{ds}{dt} dt \\
&= \int (F_x n_x + F_y n_y) \frac{ds}{dt} dt \\
&= \int \frac{(F_x \frac{dy}{dt} - F_y \frac{dx}{dt})}{ds/dt} \frac{ds}{dt} dt \\
&= \int (F_x dy - F_y dx) = \iint (\frac{\partial F_x}{\partial x} dxdy - \frac{\partial F_y}{\partial y} dydx) \\
&= \iint (\frac{\partial F_x}{\partial x} dxdy + \frac{\partial F_y}{\partial y} dxdy) \\
&= \iint \nabla \cdot F dA
\end{aligned}
$$

We can define two operations that can be performed on vector fields that are applicable in many different concepts. If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is a vector field, then the **curl** of **F** is defined by

$$
\begin{aligned}
curl F &= \nabla \times \mathbf{F} \\
&= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}
\end{aligned}
$$

If f is a function of three variables that has continuous second order partial derivatives, then

$$
\begin{aligned}
curl(\nabla f) &= \nabla \times (\nabla f) \\
&= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{vmatrix} \\
&= \left(\frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y} \right) \mathbf{i} + \left(\frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z} \right) \mathbf{j} + \left(\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x} \right) \mathbf{k} \\
&= 0
\end{aligned}
$$

The **divergence of F** is given by

$$
\begin{aligned}
div\mathbf{F} &= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \\
&= \nabla \cdot \mathbf{F}
\end{aligned}
$$

Stokes' Theorem is the higher dimension version of Green's Theorem. Instead of considering a plane region D, we can think of a space curve S. The relation between the surface integral over a surface S to a line integral around the boundary curve of S is shown below

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_S curl\mathbf{F} \cdot d\mathbf{S}$$

Proof:

$$\int_C \mathbf{F} \cdot d\mathbf{r}$$

$$\begin{aligned}
&= \int (F_x dx + F_y dy + F_z dz) \\
&= \int (\int dF_x dx + \int dF_y dy + \int dF_z dz) \\
&= \int (\int (\frac{\partial F_x}{\partial y} dy + \frac{\partial F_x}{\partial z} dz) dx \\
&+ \int (\frac{\partial F_y}{\partial x} dx + \frac{\partial F_y}{\partial z} dz) dy + \int (\frac{\partial F_z}{\partial x} dx + \frac{\partial F_z}{\partial y} dy) dz) \\
&= \int (\int (-\frac{\partial F_x}{\partial y} dxdy + \frac{\partial F_x}{\partial z} dzdx) \\
&+ \int (\frac{\partial F_y}{\partial x} dxdy - \frac{\partial F_y}{\partial z} dydz) + \int (-\frac{\partial F_z}{\partial x} dzdx + \frac{\partial F_z}{\partial y} dydz)) \\
&= \int (\int (\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}) dydz \\
&+ \int (\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}) dzdx + \int (\frac{\partial F_y}{\partial x} + \frac{\partial F_x}{\partial y}) dxdy) \\
&= \int (\int (\nabla \times F)_x dA_x + \int (\nabla \times F)_y dA_y + \int (\nabla \times F)_z dA_z) \\
&= \iint (\nabla \times F) \cdot dA
\end{aligned}$$

If we let E be a simple solid region and S the boundary surface of E and F a vector field whose component functions have partial derivatives on a region that contains E, then

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_E div\mathbf{F} dV$$

This is **Divergence Theorem**.

Proof:

$$
\begin{aligned}
&\iint_S \mathbf{F} \cdot d\mathbf{S} \\
&\qquad = \iint (F_x dA_x + F_y dA_y + F_z dA_z) \\
&\qquad = \iint (F_x dydz + F_y dxdz + F_z dxdy) \\
&\qquad = \iiint (\frac{\partial F_x}{\partial x} dxdydz + \frac{\partial F_y}{\partial y} dydzdx + \frac{\partial F_z}{\partial z} dzdxdy) \\
&\qquad = \iiint (\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}) dxdydz \\
&\qquad = \iiint \nabla \cdot F dA
\end{aligned}
$$

8 Chapter 18: Second Order Differential Equations

A **Second order linear differential equation** has the form

$$P(x)\frac{d^2y}{dx^2} + Q(x)\frac{dy}{dx} + R(x)y = G(x)$$

The system is termed **homogenous** if G(x)=0.
If we know that two solutions $y_1(x)$ and $y_2(x)$ exist, then the linear combination $y = c_1y_1(x)+c_2y_2(x)$ is also a solution to the homogenous linear equation Proof:

$$\begin{aligned}
&Py_1^{''} + Qy_1^{'} + Ry_1 = 0\\
&Py_2^{''} + Qy_2^{'} + Ry_2 = 0\\
&Py^{''} + Qy^{'} + R = 0\\
&= P(c_1y_1 + c_2y_2)^{''} + Q(c_1y_1 + c_2y_2)^{'} + R(c_1y_1 + c_2y_2)\\
&= P(c_1y_1^{''} + c_2y_2^{''}) + Q(c_1y_1^{'} + c_2y_2^{'}) + R(c_1y_1 + c_2y_2)\\
&= c_1[Py_1^{''} + Qy_1^{'} + Ry_1] + c_2[Py_2^{''} + Qy_2^{'} + Ry_2]\\
&= c_1(0) + c_2(0) = 0
\end{aligned}$$

Thus, if we know two linearly independent solution, we know every solution. Thus we need to find a solution that satisfies the auxiliary equation

$$ay^{''} + by^{'} + cy = 0$$

The function $y = e^{rx}$ has a special property because its derivative is just a constant multiple of itself. Thus,

$$\begin{aligned}
ar^2e^{rx} + bre^{rx} + ce^{rx} &= 0\\
(ar^2 + br + c)e^{rx} &= 0
\end{aligned}$$

Thus, $y = e^{rx}$ is a solution if r is a root of the **characteristic equation**

$$ar^2 + br + c = 0$$

In the case that the discriminant of this auxiliary equation is positive, or $b^2 - 4ac > 0$, then the roots r_1 and r_2 are real and distinct and the general solution of the differential equation takes the form

$$y = c_1e^{r_1x} + c_2e^{r_1x}$$

For example, if we have the equation

$$3\frac{d^2y}{dx^2} + \frac{dy}{dx} - y = 0$$

then the auxiliary equation becomes

$$3r^2 + r - 1 = 0$$
$$r = \frac{-1 \pm \sqrt{13}}{6}$$

If the discriminant is 0, then the roots are real and equal. The auxiliary becomes

$$r = \frac{-b}{2a}$$
$$2ar + b = 0$$

We know that $y_1 = e^{rx}$. $y_2 = xe^{rx}$ is also a solution. Since they are both linearly independent, the general solution of a differential equation with one real root is

$$y = c_1 e^{rx} + c_2 x e^{rx}$$

If the discriminant of the auxiliary equation is less than 0, then the roots are complex numbers. They take the form

$$r_1 = \alpha + i\beta$$
$$r_2 = \alpha - i\beta$$

We can use Euler's equation to find a solution

$$e^{i\theta} = cos\theta + isin\theta$$

We then have

$$\begin{aligned} y &= C_1 e^{r_1 x} + C_2 e^{r_2 x} = C_1 e^{(\alpha + i\beta)x} + C_2 e^{(\alpha - i\beta)x} \\ &= C_1 e^{\alpha x}(cos\beta x + isin\beta x) + C_2 e^{\alpha x}(cos\beta x - isin\beta x) \\ &= e^{\alpha x}(c_1 cos\beta x + c_2 sin\beta x) \end{aligned}$$

Thus, the general form of the solution is

$$e^{\alpha x}(c_1 cos\beta x + c_2 sin\beta x)$$

We can now find a solution to a **nonhomogenous** equation. If we have a nonhomogenous equation, then the related homogenous equation is known as the **complementary equation**. Thus, the general solution of the nonhomogeneous equation is

$$y(x) = y_p(x) + y_c(x)$$

where y_p is a particular solution to the nonhomogenous equation and y_c is the general solution of the complementary.
For example, if we have the equation $y^{''} + y^{'} - 2y = x^2$, then we start solving by first finding the complementary equation to be

$$\begin{aligned}
y^{''} + y^{'} - 2y &= 0 \\
r^2 + r - 2 = (r-1)(r+2) &= 0 \\
r &= 1, -2 \\
y_c &= c_1 e^x + c_2 e^{-2x}
\end{aligned}$$

To find the particular equation, we know that

$$\begin{aligned}
y_p &= Ax^2 + Bx + C \\
y_p^{'} &= 2Ax + B \\
y_p^{''} &= 2A \\
-2Ax^2 + (2A - 2B)x + (2A + B - 2C) &= x^2 \\
A &= -\frac{1}{2} \\
B &= -\frac{1}{2} \\
C &= -\frac{3}{4} \\
y_p &= -\frac{1}{2}x^2 - \frac{1}{2}x - \frac{3}{4}
\end{aligned}$$

Therefore, the general solution is

$$y = y_c + y_p = c_1 e^x + c_2 e^{-2x} - \frac{1}{2}x^2 - \frac{1}{2}x - \frac{3}{4}$$

Part II
Linear Algebra

9 Chapter 1: Vectors

Vectors are quantities with both a magnitude and a direction, whereas scalars are quantities with only a magnitude. For example, 10 m/s West is a vector, while 10 m/s is a scalar. Geometrically, vectors can be denoted with an arrow, with the length of the arrow being proportional to its magnitude. The vector from point A to point B is denoted by $\overline{AB}$ where point A is the **initial** point and point B is the **terminal** point. Column vectors are often used to represent vectors conveniently. For example,

$$\mathbf{a} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

In this case, -2 and 3 are known as the **components** of the vector **a**. Graphically, an arrow would be drawn from the origin to point (-2,3).

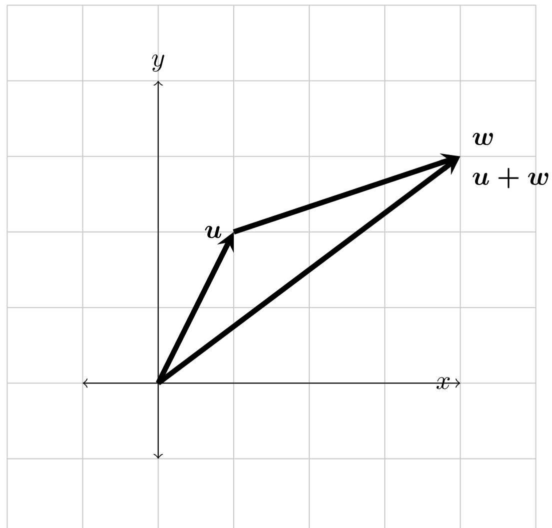
Two vectors are equal if they have the same length and direction, despite differences in their initial and terminal points. Any vector can be redrawn with its initial point in the origin, making it in standard position. For example, redrawing vector AB in **standard position** would look like

$$\begin{aligned} A &= (-1, 2) \\ B &= (3, 4) \\ \overline{AB} &= [3 - (-1), 4 - 2] = [4, 2] \end{aligned}$$

Vectors have different operators. The notion of vector addition is as follows. If $\mathbf{u} = [u_1, u_2]$ and $\mathbf{w} = [w_1, w_2]$, then

$$\mathbf{u} + \mathbf{w} = [u_1 + w_1, u_2 + w_2]$$

Geometrically, vector addition looks like this



$$\mathbf{u} = [1, 2]$$
$$\mathbf{w} = [3, 1]$$
$$\mathbf{u} + \mathbf{w} = [1 + 3, 2 + 1] = [4, 3]$$

The second basic operation is **scalar multiplication**. It is as follows

$$c\mathbf{v} = c[v_1, v_2] = [cv_1, cv_2]$$

Vector subtraction is adding by the negative of a vector. The negative of a vector is simply scalar multiplication done by -1.

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

The vectors above were in the set of all ordered doubles of real numbers and is denoted by $\mathbb{R}^2$. We can define $\mathbb{R}^n$ as the set of all ordered n-tuples of real numbers. Thus, a vector in $\mathbb{R}^n$ can be written as

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

The following are algebraic properties of all vectors in $\mathbb{R}^n$

$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$$
$$(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$$
$$\mathbf{u}+\mathbf{0}=\mathbf{u}$$
$$\mathbf{u}+(-\mathbf{u})=\mathbf{0}$$
$$c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$$
$$(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{d}$$
$$c(d\mathbf{u})=(cd)\mathbf{u}$$
$$1\mathbf{u}=\mathbf{u}$$

A vector $\mathbf{v}$ is a **linear combination** of vectors $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ if there are scalars $c_1, c_2, ..., c_3$ such that $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ ...\ c_k\mathbf{v_k}$. The scalars are called the **coefficients** of the linear combination. For example,

The vector $\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix}$

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

Another vector operation is known as the dot product. It is essentially the sum of the products of the corresponding components of two vectors. Note that the answer is a scalar. The **dot product** of two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + ... + u_nv_n$$

For example, when $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix}$ then,

$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot (-3) + 2 \cdot 5 + (-3) \cdot 2 = 1$$

The dot product can also be attained by

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

where θ is the angle between the two vectors

In general, these dot product properties are true for vectors in $\mathbb{R}^n$

$$\begin{gathered}
\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \\
\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \\
(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) \\
\mathbf{u} \cdot \mathbf{u} \geq 0 \text{ and } \mathbf{u} \cdot \mathbf{u} = 0 \text{ if and only if } \mathbf{u} = 0
\end{gathered}$$

The length of a vector $\mathbf{v}$ is defined by

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

For example,

$$||[2, 3]|| = \sqrt{2^2 + 3^2} = \sqrt{13}$$

A vector of length 1 is known as a **unit vector**. Given any nonzero vector $\mathbf{v}$, we can always find a unit vector in the same direction by dividing $\mathbf{v}$ by its own length. This process is known as **normalizing** a vector.

$$\mathbf{u} = \frac{\mathbf{v}}{||v||}$$

In $\mathbb{R}^n$, any unit vector which has 1 has its i-th component and 0 in all its other components are known as **standard unit vectors**.

The **Cauchy-Schwarz Inequality** is as follows

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

Proof (without angle):

$$\begin{aligned}
||\mathbf{u} - t\mathbf{v}||^2 &= (\mathbf{u} - t\mathbf{v}) \cdot (\mathbf{u} - t\mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{u} - \mathbf{u} \cdot t\mathbf{v} - t\mathbf{u} \cdot \mathbf{v} + t^2\mathbf{v} \cdot \mathbf{v} \\
&= ||\mathbf{v}||^2 t^2 - 2\mathbf{u} \cdot \mathbf{v} + ||\mathbf{u}||^2
\end{aligned}$$

This is an equation of a non-negative parabola and thus,

the determinant must be non negative as well

$$\begin{aligned}
D = (2\mathbf{u} \cdot \mathbf{v})^2 - 4||\mathbf{u}||^2||\mathbf{v}||^2 &\leq 0 \\
(\mathbf{u} \cdot \mathbf{v})^2 &\leq ||\mathbf{u}||^2||\mathbf{v}||^2 \\
|\mathbf{u} \cdot \mathbf{v}| &\leq ||\mathbf{u}||||\mathbf{v}||
\end{aligned}$$

Proof (with angle):

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v} &= ||\mathbf{u}||||\mathbf{v}|| \cos\theta \\
\cos\theta &\leq 1 \\
|\mathbf{u} \cdot \mathbf{v}| &\leq ||\mathbf{u}||||\mathbf{v}||
\end{aligned}$$

The **Triangle Inequality** is as follows

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

Proof:

$$\begin{aligned}
||\mathbf{u} \cdot \mathbf{v}||^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{u} + 2(\mathbf{u} \cdot \mathbf{v}) + \mathbf{v} \cdot \mathbf{v} \\
&\leq ||\mathbf{u}||^2 + 2|\mathbf{u} \cdot \mathbf{v}| + ||\mathbf{v}||^2 \\
&\leq ||\mathbf{u}||^2 + 2||\mathbf{u}||||\mathbf{v}|| + ||\mathbf{v}||^2 \\
&= (||\mathbf{u}|| + ||\mathbf{v}||)^2
\end{aligned}$$

The **distance** between two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined by

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

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** to each other if $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$. The Pythagoras' Theorem can now be seen as $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

We can now define the **projection** of a vector onto another vector. If θ is the angle between two vectors $\mathbf{u}$ and $\mathbf{v}$, then the projection of $\mathbf{v}$ onto $\mathbf{u}$ is

$$\begin{aligned}
\mathbf{p} &= ||p||\hat{u} \\
\mathbf{p} &= ||v|| \cos\theta \hat{u} \\
\mathbf{p} &= ||v|| \left(\frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||||\mathbf{v}||} \right) \left(\frac{1}{||\mathbf{u}||} \right) \mathbf{u} \\
\mathbf{p} &= \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}} \right) \mathbf{u}
\end{aligned}$$

For example, let $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix}$, then the projection of $\mathbf{v}$ onto $\mathbf{u}$ is

$$\mathbf{u} \cdot \mathbf{v} = 4(-2) + (2)(-1) + 1(3) = -8 - 2 + 3 = -7$$
$$\mathbf{u} \cdot \mathbf{u} = 4(4) + (2)(2) + 1(1) = 16 + 4 + 1 = 21$$
$$\mathbf{p} = \frac{-1}{3} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

Vectors can be utilized to describe equations of both lines and planes in different forms. For lines, we can define a specific vector known as the **normal vector**. The normal vector is the vector that is perpendicular to any vector $\mathbf{x}$ parallel to the line. Thus, $\mathbf{n} \cdot \mathbf{x} = \mathbf{0}$ as they are orthogonal. We can also define $\mathbf{d}$ as the **direction vector** being a vector parallel to the line. Thus, $\mathbf{x}$ is really just a scalar multiple of $\mathbf{d}$. Thus, $\mathbf{x} = \mathrm{p}\mathbf{d}$. For situations where the line does not pass through the origin, we must put the direction vector into standard position first by subtracting vector $\mathbf{p}$, a point on the line, from vector $\mathbf{x}$. Thus, $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = \mathbf{0}$ or $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$ This describes the **normal form** of the equation of a line.

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

The normal vector can be found from the **general form** of the equation of a line.

$$a\mathbf{x} + b\mathbf{y} = c$$
$$\mathbf{n} = \begin{bmatrix} a \\ b \end{bmatrix}$$

The **vector form** of the equation of a line simply stems from the definition of $\mathbf{x}$. The equations corresponding to the components of the vector form are called **parametric equations**.

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

For example, let l be a line in $\mathbb{R}^3$ passing through the point P = (1,2,-1) and parallel to the vector $\mathbf{d} = \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}$. Then,

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

(Vector Form)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + t \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}$$

(Parametric Form)

$$x = 1 + 5t$$
$$y = 2 - t$$
$$z = -1 + 3t$$

Another example. If we let $7\mathbf{x} + 3\mathbf{y} = 19$, then we can take any arbitrary point P = (1, 4)

$$\mathbf{n} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$
$$\mathbf{p} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$
$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

(Normal Form)

$$\begin{bmatrix} 7 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

The same derivations can be done with the equation of a plane as well. If we let $ax + by + cz = d$ describe the general form of a plane and **p** be a specific point on the plane, then the normal form of the equation of a plane is given by

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$
$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$$
$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$$

Because of the second dimension, two direction vectors that are not parallel to each other are required to describe the vector form of a plane. If we let **u** and **v** be those direction vectors, then the vector form of a plane is given by the following. Again, the parametric equations are simply the equations of the corresponding components.

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

For example, if we have a plane that contains the point P = (5,7,3) and normal vector $\mathbf{n} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, then

$$\mathbf{n} \cdot \mathbf{p} = 1(5) + 2(7) + 3(3) = 5 + 14 + 9 = 28$$
$$\mathbf{n} \cdot \mathbf{x} = x + 2y + 3z = 28$$

We can find two other points on the plane to get two direction vectors.

$$Q = (3, 2, 7) \text{ and } R = (2, 1, 8)$$
$$\mathbf{u} = \begin{bmatrix} 3-5 \\ 2-7 \\ 7-3 \end{bmatrix} = \begin{bmatrix} -2 \\ -5 \\ 4 \end{bmatrix}$$
$$\mathbf{v} = \begin{bmatrix} 2-5 \\ 1-7 \\ 8-3 \end{bmatrix} = \begin{bmatrix} -3 \\ -8 \\ 5 \end{bmatrix}$$
$$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$$
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 3 \end{bmatrix} + s\begin{bmatrix} -2 \\ -5 \\ 4 \end{bmatrix} + t\begin{bmatrix} -3 \\ -8 \\ 5 \end{bmatrix}$$

There is a simple way to find a normal vector given two nonparallel vectors. We can define a vector operation called the **cross product** to do so. Taking the cross product of two vectors will result in a third vector that is orthogonal to both original vectors. In chapter 3, we can use the determinant to find the cross product of two vectors. For now, the magnitude of the cross product can be found by

$$|\mathbf{u} \times \mathbf{v}| = uv\sin\theta$$

Finally, we can use these concepts to obtain the distance from a point to a line and the distance from a point to a plane. The distance between a point B and a line l is simply the perpendicular component of the triangle that the vector makes with the plane. In other words, it is the vector minus the component along the line, which is simple the projection of the vector onto the line. Thus,

$$
\begin{aligned}
d(B,l) &= ||\mathbf{v} - proj(\mathbf{v})_d|| \\
\text{if } \mathbf{v} &= \begin{bmatrix} x - x_o \\ y - y_o \end{bmatrix} \text{ then,} \\
d(B,l) &= \frac{|\mathbf{v} \cdot \mathbf{n}|}{||\mathbf{n}||} \\
&= \frac{a(x - x_o) + b(y - y_o)}{\sqrt{a^2 + b^2}} \\
&= \frac{ax_o + by_o - c}{\sqrt{a^2 + b^2}}
\end{aligned}
$$

The same exercise can be done to find the distance between a point and a plane.

$$
\begin{aligned}
d(B,P) &= ||\mathbf{v} - proj(\mathbf{v})_p|| \\
\text{if } \mathbf{v} &= \begin{bmatrix} x - x_o \\ y - y_o \\ z - z_o \end{bmatrix} \text{ then,} \\
d(B,l) &= \frac{|\mathbf{v} \cdot \mathbf{n}|}{||\mathbf{n}||} \\
&= \frac{a(x - x_o) + b(y - y_o) + c(z - z_o)}{\sqrt{a^2 + b^2 + c^2}} \\
&= \frac{ax_o + by_o + cz_o - d}{\sqrt{a^2 + b^2 + c^2}}
\end{aligned}
$$

10 Chapter 2: Systems of Linear Equations

A **linear equation** is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + ...a_nx_n = b$$

where $a_1, a_1, ..., a_n$ are **coeffecients** and the b is a constant term.

A **system of linear equations** is simply a finite set of linear equations, each with the same variables. A **solution** would be a vector that is a solution of each equation in the system. A system is known as **consistent** if there exists at least on solution and **inconsistent** if it does not. There are three possibilities with a system of equations with real coefficients. It either has one unique solution, infinitely many solutions or no solutions. If we have the following system

$$\begin{aligned} x - y - z &= 2 \\ y + 3z &= 5 \\ 5z &= 10 \end{aligned}$$

then we can solve the system by solving for the last equation and working backwards. This process is known as **back substitution**.

We can solve systems of equations through **matrixes**. We can take a system and turn it into an augmented matrix like so.

$$\begin{aligned} x - y - z &= 2 \\ 3x - 3y + 2z &= 16 \\ 2x - y + z &= 9 \end{aligned}$$

$$\begin{bmatrix} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{bmatrix}$$

We can solve this system by performing **elementary row operations** on the augmented matrix and turn it into **row echelon form**. Row Echelon form is simply a form of matrix where any zero rows are completely at the bottom and any nonzero rows are ordered in such a way that the first nonzero entry is in a column of its own, creating a sort of staircase pattern with the zeros. Elementary row operations are simply operations that you can do on the rows of the matrixes.

Elementary Row Operations:

Interchange two rows

Multiply a row by a nonzero constant

Add a multiple of a row to another row

Taking the previous augmented matrix, we can transform it into row echelon form.

$$\begin{bmatrix} 1 & -1 & -1 & 2 \\ 3 & -3 & 2 & 16 \\ 2 & -1 & 1 & 9 \end{bmatrix} \overset{r_2-3r_1}{\longrightarrow} \begin{bmatrix} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 2 & -1 & 1 & 9 \end{bmatrix}$$
$$\overset{r_3-2r_1}{\longrightarrow} \begin{bmatrix} 1 & -1 & -1 & 2 \\ 0 & 0 & 5 & 10 \\ 0 & 1 & 3 & 5 \end{bmatrix} \overset{r_2<->r_3}{\longrightarrow} \begin{bmatrix} 1 & -1 & -1 & 2 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 5 & 10 \end{bmatrix}$$

We can now do back substitution to solve the system. This method is known as **Gaussian Elimination**. If we find that the row echelon form gives us a system that has infinitely many solutions, then we will have to assign parameters to the **free variables** and write the **leading variables** in terms of those variables. For example,

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

gives us the equations

$$x - y - w + 2z = 1$$
$$y - z = 1$$
$$\begin{bmatrix} x \\ y \\ w \\ z \end{bmatrix} = \begin{bmatrix} 2+s-t \\ s \\ 1+t \\ t \end{bmatrix}$$

We number of free variables is simply the number of variables minus the leading variables. We can also see that the number of leading variables is also the number of nonzero rows in a row echelon matrix. We can define this number as the **rank** of a matrix. Thus, the **rank theorem** states

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

There is another form of matrix called the **reduced row echelon form**, where a matrix is in row echelon, but every leading entry is 1 and every column with a leading 1 has 0s everywhere else. We can solve a system by also placing it in reduced row echelon form and solving for the leading variables in terms of the free variables. This is known as **Gauss-Jordan Elimination**. Taking the previous system,

$$\begin{bmatrix} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{r_1+r_2} \begin{bmatrix} 1 & -1 & 0 & 1 & 2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

which gives us the equations

$$\begin{aligned} x - y + z &= 2 \\ w - z &= 1 \\ x &= 2 + y - z \\ w &= 1 + z \\ \begin{bmatrix} x \\ y \\ w \\ z \end{bmatrix} &= \begin{bmatrix} 2 + s - t \\ s \\ 1 + t \\ t \end{bmatrix} \end{aligned}$$

One unique type of system that can always be solved is a system where the constant term in each equation is 0. This type of system is known as **homogeneous**. It takes the form $[A|0]$. We can prove the following theorem

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

Proof:

$$\begin{gathered} rank(A) \leq m \\ \text{number of free variables} = n - rank(A) \geq n - m > 0 \end{gathered}$$

A general fact about system of linear equations. A system of linear equations with augmented matrix $[A|b]$ is consistent if and only if b is a linear combination of the columns of A. If $S = \mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ is a set of vectors in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ is known as the **span** of the set. If $span(S) = \mathbb{R}^n$, then S is called the **spanning set** for $\mathbb{R}^n$.

For example, if we want to show that $span\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right) = \mathbb{R}^n$, then we have to show that any arbitrary vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is a linear combination of the two vectors.

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

The augmented matrix is

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

Thus, the system is consistent for any choice of a and b.

If we consider the standard unit vectors $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$, then for any vector $\begin{bmatrix}x\\y\\z\end{bmatrix}$ we have

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=x\begin{bmatrix}1\\0\\0\end{bmatrix}+y\begin{bmatrix}0\\1\\0\end{bmatrix}+z\begin{bmatrix}0\\0\\1\end{bmatrix}$$

Thus, $span(\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}) = \mathbb{R}^3$

A set of vectors is **linearly dependent** if there are scalars $c_1, c_2, ...c_k$ at least one of which is not zero, such that

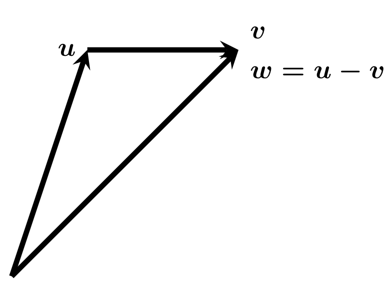
$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k} = 0$$

Let's assume that we have a linearly independent set $S = [u_1, u_2, ...u_m]$. We take a subset of this set $A = [u_1, u_2, ...u_k]$, where $k < m$. Assume that subset A has a vector that is linearly dependent. This would mean that this vector $\mathbf{u_n} = c_1\mathbf{u_1} + c_2\mathbf{u_2} + ...$ However since vector $\mathbf{u_n}$ is in the set S, then this would contradict the original claim that S is a linearly independent set. Thus, any subset of a linearly independent set is also a linearly independent set.

We can use what we have learned so far to prove both the law of sines and the law of cosines.

Proof of law of cosines:

Let $\mathbf{w} = \mathbf{u} - \mathbf{v}$ and the angle between $\mathbf{u}$ and $\mathbf{v}$ to be θ.



$$
\begin{aligned}
w^2 &= (\mathbf{u} - \mathbf{v})^2 \\
&= u^2 + v^2 - 2\mathbf{u} \cdot \mathbf{v} \\
&= u^2 + v^2 - 2uv \cos\theta
\end{aligned}
$$

Proof of law of sines:

$$
\begin{aligned}
\mathbf{w} \times \mathbf{v} = \mathbf{u} - \mathbf{v} \times \mathbf{v} = \mathbf{u} \times \mathbf{v} \\
\mathbf{w} \times \mathbf{u} = \mathbf{u} - \mathbf{v} \times \mathbf{u} = \mathbf{u} \times \mathbf{v} \\
|\mathbf{w} \times \mathbf{v}| = |\mathbf{w} \times \mathbf{u}| = |\mathbf{u} \times \mathbf{v}| \\
wu \sin\alpha = uu \sin\beta = uv \sin\theta \\
\frac{\sin\beta}{v} = \frac{\sin\alpha}{u} = \frac{\sin\theta}{w}
\end{aligned}
$$

11 Chapter 3: Matrices

A **matrix** is a rectangular array of numbers called the **entries**, or **elements** of the matrix. A matrix whose number of rows equal the number of columns is known as a **square** matrix and the square matrix whose nondiagonal entries are all 0 is known as a **diagonal matrix**.

A + B is simply the sum of both their entries; however, they both must have the same dimensions. Scalar multiplication of a matrix, cA, acts in the same way as scalar multiplication of a vector. The negative of a matrix is simply that matrix scalar multiplied by -1.

If A is an $m \times n$ matrix and B is an $n \times r$ matrix, then the two matrices can be multiplied with each other. Notice that the number of columns of A must equal the number of rows of B. The result is a matrix C with dimension $m \times r$. Matrix multiplication is as follows;

$$c_i j = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{in}b_{nj}$$

For example, if $A = \begin{bmatrix} 1 & 3 & -1 \\ -2 & -2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 4 \\ 2 & 3 \\ 3 & 5 \end{bmatrix}$, then

$$\begin{aligned}
c_{11} &= 1(1) + 3(2) + (-1)3 = 4 \\
c_{12} &= 1(4) + 3(3) + (-1)5 = 8 \\
c_{21} &= -2(1) + -2(2) + (1)3 = -3 \\
c_{21} &= -2(4) + -2(3) + (1)5 = -9 \\
AB &= \begin{bmatrix} 4 & 8 \\ -3 & -9 \end{bmatrix}
\end{aligned}$$

Matrix can have exponentials as well. $A^k = AA...A$, k times. $A^rA^s = A^{r+s}$ and $(A^r)^s = A^{rs}$

The **transpose** of an $m \times n$ matrix is simply an $n \times m$ matrix A^T where the rows and columns are interchanged. For example, the transpose of $\begin{bmatrix} 1 & 4 \\ 2 & 3 \\ 3 & 5 \end{bmatrix}$ is simply $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 3 & 5 \end{bmatrix}$. A square matrix A is **symmetric** is $A^T = A$.

In the same way linear combination is defined with vectors, a **linear combination** of matrices can be formed as well with $c_1A_1 + c_2A_2 + c_3A_3 + ...$ If $A_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, and matrix $B = \begin{bmatrix} 1 & 4 \\ 2 & 1 \end{bmatrix}$, then we can determine if B is a linear combination of the three matrices like so:

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

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

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

Thus, $c_1 = 1$, $c_2 = -2$ and $c_3 = 3$ and B is a linear combination of the matrices.

The notion of **linear independence** and **linear dependence** is the same with matrices as with vectors. We can define the **span** of a set of matrices to be the set of all linear combination of the matrices. Taking the previously defined vectors, we can find its span like so

$$c_1 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} c_2 + c_3 & c_1 + c_3 \\ -c_1 + c_3 & c_2 + c_3 \end{bmatrix}$$

$$\begin{bmatrix} c_2 + c_3 & c_1 + c_3 \\ -c_1 + c_3 & c_2 + c_3 \end{bmatrix} = \begin{bmatrix} w & x \\ -y & z \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 1 & w \\ 1 & 0 & 1 & x \\ -1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & x/2 - y/2 \\ 0 & 1 & 0 & -x/2 - y/2 + w \\ 0 & 0 & 1 & x/2 + y/2 \\ 0 & 0 & 0 & w - z \end{bmatrix}$$

Thus, $span(A_1, A_2, A_3) = \begin{bmatrix} w & x \\ y & w \end{bmatrix}$ since $w = z$

We also have some properties of Matrix Multiplication

$$\begin{aligned} A(BC) &= (AB)C \\ A(B + C) &= AB + BC \\ (A + B)C &= AC + BC \\ k(AB) &= (kA)B = A(kB) \\ I_m A &= A = AI_m \end{aligned}$$

And some properties of the transpose

$$
\begin{aligned}
(A^T)^T &= A \\
(A+B)^T &= A^T + B^T \\
(kA)^T &= k(A^T) \\
(AB)^T &= B^T A^T \\
(A^r)^T &= (A^T)^r
\end{aligned}
$$

If A is an $n \times n$ matrix, an **inverse** of A is an $n \times n$ matrix $A^{'}$ with the property that $AA^{'} = I$ and $A^{'}A = I$. A is known as **invertible** if such a matrix exists.

For any system of linear equations given by $A\mathbf{x} = \mathbf{b}$, the solution is given by $\mathbf{x} = A^{'}\mathbf{b}$

Proof:

$$
\begin{aligned}
A\mathbf{x} &= \mathbf{b} \\
A^{'}A\mathbf{x} &= A^{'}\mathbf{b} \\
I\mathbf{x} &= A^{'}\mathbf{b} \\
\mathbf{x} &= A^{'}\mathbf{b}
\end{aligned}
$$

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then A is invertible if $ad - bc \neq 0$, in which case

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

The expression $ad - bc$ is known as the **determinant**, denoted by $detA$.

For example, if we have the system of equations $x+2y = 3$ and $3x+4y = -2$, then

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
\mathbf{b} &= \begin{bmatrix} 3 \\ -2 \end{bmatrix} \\
A^{-1} &= \frac{1}{detA} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \\
&= \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \\
\mathbf{x} &= \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} \\
&= \begin{bmatrix} -8 \\ 11/2 \end{bmatrix}
\end{aligned}
$$

Here are some properties of Invertible Matrices:

$$
\begin{aligned}
(A^{-1})^{-1} &= A \\
(cA)^{-1} &= \frac{1}{c}A^{-1} \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(A^T)^{-1} &= (A^{-1})^T \\
(A^n)^{-1} &= (A^{-1})^n
\end{aligned}
$$

We can use matrices to perform row operations on other matrices. We do this by multiplying an **elementary matrix** with the matrix. To obtain this elementary matrix, we simple perform the intended row operation on an identity matrix. For example, the matrix $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ multiplied by any matrix is equivalent to performing $3R_2$ on that matrix.

We can use the Gauss-Jordan method to compute the inverse. If A is invertible, then computing

$$[A|I] \rightarrow [I|A^{-1}]$$

gives us the inverse.

We can factor matrices in an assortment of ways. One of these ways is **LU Factorization**. We do this by row reducing A to get matrix U. The row operations we use will have a multiplier. We place the multipliers required to get the first column of row echelon form under the first 1 of an identity matrix, and continue on with the second, third, etc. columns, until we get a matrix L.

For example, if we have matrix $A = \begin{bmatrix} 3 & 1 & 3 & -4 \\ 6 & 4 & 8 & -10 \\ 3 & 2 & 5 & -1 \\ -9 & 5 & -2 & -4 \end{bmatrix}$, then

$$\begin{bmatrix} 3 & 1 & 3 & -4 \\ 6 & 4 & 8 & -10 \\ 3 & 2 & 5 & -1 \\ -9 & 5 & -2 & -4 \end{bmatrix} \xrightarrow{r_2 - 2r_1} \xrightarrow{r_3 - r_1} \xrightarrow{r_4 - (-3r_1)} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 8 & 7 & -16 \end{bmatrix}$$

$$\xrightarrow{r_3 - \frac{1}{2}r_1} \xrightarrow{r_4 - 4r_2} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -1 & -8 \end{bmatrix}$$

$$\xrightarrow{r_4 - (-1)r_3} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & -4 \end{bmatrix} = U$$

Our L matrix is our multipliers under the diagonals of the identity matrix

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & \frac{1}{2} & 1 & 0 \\ -3 & 4 & -1 & 1 \end{bmatrix}$$

If our row reduction requires switching two rows, then we have to use $P^T LU$ Factorization instead. In $P^T LU$ factorization we find row echelon form of A, and we find our **permutation matrix**, P, by performing our row operations on an identity matrix - elementary matrices. We then find the LU factorization of PA instead, and factor with the transpose of P - P^T.

We can define a **subspace** of $\mathbb{R}^n$ as any collection S of vectors in $\mathbb{R}^n$ such that

The zero vector 0 is in S

If **u** and **v** are in S, then $\mathbf{u} + \mathbf{v}$ is in S

if **u** is in S, then c**u** is in S

The **row space** of a matrix A is the subspace row(A) of $\mathbb{R}^n$ spanned by the rows of A. The **column space** of a matrix A is the subspace col(A) of $\mathbb{R}^n$ spanned by the column of A. For example, if we have matrix $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 3 & -3 \end{bmatrix}$, and we want to determine if matrix $\mathbf{b} = \begin{bmatrix}1\\2\\3\end{bmatrix}$ is in the column space of A, we can say that **b** is a column space of A if $A\mathbf{x} = \mathbf{b}$ is consistent. Thus

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

The system is consistent and **b** is in col(A).

We can determine whether vector $\mathbf{w} = [4, 5]$ is in the row space of A through the same process. We can row reduce $\begin{bmatrix}A\\w\end{bmatrix}$ to $\begin{bmatrix}A'\\0\end{bmatrix}$ by

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

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

Thus, **w** is in the subspace row(A).

The **null space** is the subspace consisting of solutions to the system $A\mathbf{x} = 0$. A**basis** for a subspace S is a set of vectors in S that both spans S and is linearly independent. Any two bases for the same subspace have the same number of vectors, known as the **dimension** of S. The **rank** of a matrix A is the dimension of its row and column spaces, denoted by $rank(A)$. The **nullity** of a matrix A is the dimension of its null space and its denoted by $nullity(A)$. Thus, the **rank theorem** states that $rank(A) + nullity(A) = n$ if A is an $m \times n$ matrix. This allows us to easily find the nullity of a matrix without solving for $A\mathbf{x} = 0$.

Proof of Rank Theorem: If A is an $m \times n$ matrix, and R is the row echelon form of the matrix, then R has n rows and r leading variables, indicating n-r free variables in the solution $A\mathbf{x} = 0$. $rank(A) = r$. $dim(null(A)) = n - r$. Thus, $rank(A) + nullity(A) = r + n - r = n$

If we let $\beta = [v_1, v_2, ...]$ be a basis for subspace S, then there is one way to write **v** as a linear combination of the basis vectors: $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ...$ The scalars $c_1, c_2, ...$ are the **coordinates of v with respect to** β and $[v]_\beta = \begin{bmatrix}c_1\\c_2\\c_3\\...\end{bmatrix}$ is the **coordinate vectors of v with respect to** β.

A transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is called a **linear transformation** if $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ and $T(c\mathbf{v}) = cT(\mathbf{v})$

For example, the transformation defined by $T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 2x - y \\ 3x + 4y \end{bmatrix}$ is a linear transformation because

$$T(\mathbf{u} + \mathbf{v}) = T(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}) = \begin{bmatrix} x_1 + x_2 \\ 2(x_1 + x_2) - (y_1 + y_2) \\ 3(x_1 + x_2) + 4(y_1 + y_2) \end{bmatrix}$$

$$= \begin{bmatrix} x_1 \\ 2x_1 - y_1 \\ 3x_1 + 4y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ 2x_2 - y_2 \\ 3x_2 + 4y_2 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v})$$

$$T(c\mathbf{v}) = T(\begin{bmatrix} cx \\ cy \end{bmatrix}) = \begin{bmatrix} cx \\ c(2x - y) \\ c(3x + 4y) \end{bmatrix} = cT(\mathbf{v})$$

We can take the linear transformation T and place it in matrix form, where any matrix multiplied by this **standard matrix of transformation T** will result in the transformed form of the original matrix. We can also tie two transformations together with the **composition** of two transformations where T is a transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ and S is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$. To transform from $\mathbb{R}^m$ to $\mathbb{R}^p$, we can use $S(T(\mathbf{v}))$. The standard matrix of transformation between these two matrices is simply [S][T]. S and T are **inverse transformations** if $S \cdot T$ and $T \cdot S$ are both equal to the identity matrix.

12 Chapter 4: Eigenvalues and Eigenvectors

A scalar λ is called an **eigenvalue** of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. The vector $\mathbf{x}$ is called an **eigenvector**. The collection of all eigenvectors associated with the eigenvalue λ is called the **eigenspace** of λ and denoted by E_λ. We see that $A\mathbf{x} = \lambda\mathbf{x}$ can be rearranged to $(A - \lambda\mathbf{x}) = 0$. The set of all eigenvectors corresponding to an eigenvalue is simply the null space of $A - \lambda I$. Thus, matrix A has an eigenvalue if the null space of $A - \lambda I$ has a nontrivial solution. A matrix A also has a nontrivial null space if its determinant is 0. Thus, we can find the eigenvalues of a matrix by setting $det(A - \lambda I) = 0$. For example, if we want to find the eigenvalues of $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, then

$$\begin{aligned} det(A - \lambda I) &= 0 \\ det(\begin{bmatrix} 3-\lambda & 1 \\ 1-\lambda & 3 \end{bmatrix}) &= 0 \\ &= (3-\lambda)(3-\lambda) - 1 \\ &= \lambda^2 - 6\lambda + 8 \\ \lambda &= 2, 4 \end{aligned}$$

We also can find the equation of a determinant for any square matrix.

$$\begin{aligned} detA &= a_{11}detA_{11} - a_{12}detA_{12} + a_{13}detA_{13} \\ &= \Sigma_1^n (-1)^{i+j} a_{ij} detA_{ij} \end{aligned}$$

where A_{ij} is the sub matrix of A where the i-th row and the j-th column are removed.

For a 3×3 matrix, the determinant is

$$detA = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

The **Laplace Expansion Theorem** says that the determinant of an $n \times n$ matrix can be found with

$$detA = a_{i1}C_{i1} + a_{i2}c_{i2} + ... = \Sigma_1^n a_{ij}C_{ij}$$

or

$$detA = a_{1j}C_{1j} + a_{2j}c_{2j} + ... = \Sigma_1^n a_{nj}C_{nj}$$

where the **cofactor** is $C_{ij} = (-1)^{i+j} detA_{ij}$

For example, if $A = \begin{bmatrix} 2 & -3 & 1 & 0 & 4 \\ 0 & 3 & 2 & 5 & 7 \\ 0 & 0 & 1 & 6 & 0 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$, then its determinant is equal to

$$\begin{aligned} detA &= 2\begin{vmatrix} 3 & 2 & 5 & 7 \\ 0 & 1 & 6 & 0 \\ 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & -1 \end{vmatrix} \\ &= 2*3\begin{vmatrix} 1 & 6 & 0 \\ 0 & 5 & 2 \\ 0 & 0 & -1 \end{vmatrix} \\ &= 2*3*1\begin{vmatrix} 5 & 2 \\ 0 & -1 \end{vmatrix} \\ &= 2*3*1*(5(-1) - 2*0) \\ &= -30 \end{aligned}$$

$det(AB) = det(A)det(B)$, $det(kA) = k^n det(A)$, $det(A^{-1}) = (\frac{1}{det(A)})$, and $det(A) = det(A^T)$ are some properties of determinants. **Cramer's Rule** states that if A is an invertible matrix and **b** be a vector, then the solution **x** of $A\mathbf{x} = \mathbf{b}$ is given by

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

For example if we have the system

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

where $A_i(\mathbf{b})$ is the matrix A with the i-th column replaced with **b**. Then,

$$\begin{aligned} detA &= 6 \\ det(A_1(\mathbf{b})) &= \begin{vmatrix} 2 & 2 \\ 1 & 4 \end{vmatrix} = 6 \\ det(A_2(\mathbf{b})) &= \begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = 3 \\ x_1 &= \frac{6}{6} = 1 \\ x_2 &= \frac{3}{6} = \frac{1}{2} \end{aligned}$$

If matrices A and B are related to each other with an invertible matrix P such that $P^{-1}AP = B$, then we say that A is **similar to** B. This can be shown as $AP = PB$ as well. If A and B are similar then the following hold

$$detA = detB$$

A is invertible only if B is invertible

$$rankA = rankB$$

A and B have the same characteristic polynomial

A and B have the same eigenvalues

A square matrix A is **diagonalizable** if there is a diagonal matrix D such that A is similar to D and $P^{-1}AP = D$. A must have n linearly independent eigenvectors to be diagonalizable. We can find P by putting the linearly independent eigenvectors in a matrix. Using P and A, we can also solve for D.

13 Chapter 5: Orthogonality

A set of vectors $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ is an **orthogonal set** if all pairs of vectors in the set are orthogonal. For example, the set $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ where $\mathbf{v_1} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{v_3} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ is an orthogonal set because

$$\mathbf{v1} \cdot \mathbf{v2} = (2)(0) + (1)(1) + (-1)(1) = 0$$
$$\mathbf{v2} \cdot \mathbf{v3} = (0)(1) + (1)(-1) + (1)(1) = 0$$
$$\mathbf{v1} \cdot \mathbf{v3} = (2)(1) + (1)(-1) + (-1)(1) = 0$$

If $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ is an orthogonal set of nonzero vectors, then these vectors must be linearly independent.

Proof:
if $c_1, c_2, ..., c_k$ are scalars such that $c_1\mathbf{v_1} + ... + c_k\mathbf{v_k} = 0$ (linearly dependent), then

$$(c_1\mathbf{v_1} + ... + c_k\mathbf{v_k}) \cdot \mathbf{v_i} = 0 \cdot \mathbf{v_i} = 0$$
$$c_1(\mathbf{v_1} \cdot \mathbf{v_i}) + ...c_i(\mathbf{v_i} \cdot \mathbf{v_i}) + ... + c_k(\mathbf{v_k} \cdot \mathbf{v_i}) = 0$$
$$c_i(\mathbf{v_i} \cdot \mathbf{v_i}) = 0$$

Since $\mathbf{v_i} \neq 0$, c_i must be 0. This is true for all values of i. Thus, the set must be linearly independent since all coefficients are 0.

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set. If we let $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and $\mathbf{w}$ be any vector in W, then

$$\mathbf{w} = c_1\mathbf{v_1} + ... + c_k\mathbf{v_k}$$

and its scalars are given by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v_i}}{\mathbf{v_i} \cdot \mathbf{v_i}}$$

Proof:
Since the set is a basis, vector $\mathbf{w}$ can be defined by any combination of scalars $\mathbf{w} = c_1\mathbf{v_1} + ... + c_k\mathbf{v_k}$. We can take the dot product with $\mathbf{v_i}$ to obtain

$$
\begin{aligned}
\mathbf{w} \cdot \mathbf{v_i} &= (c_1\mathbf{v_1} + ... + c_k\mathbf{v_k}) \cdot \mathbf{v_i} \\
&= c_i(\mathbf{v_1} \cdot \mathbf{v_i}) + ... + c_i(\mathbf{v_i} \cdot \mathbf{v_i}) + ... + c_k(\mathbf{v_k} \cdot \mathbf{v_i}) \\
&= c_i(\mathbf{v_i} \cdot \mathbf{v_i}) \\
c_i &= \frac{\mathbf{w} \cdot \mathbf{v_i}}{\mathbf{v_i} \cdot \mathbf{v_i}}
\end{aligned}
$$

For example, if we let $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and the orthogonal basis $\beta = [\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}]$ (the vectors on the previous page), then

$$
\begin{aligned}
c_1 &= \frac{2+2-3}{4+1+1} = \frac{1}{6} \\
c_2 &= \frac{0+2+3}{0+1+1} = \frac{5}{2} \\
c_3 &= \frac{1-2+3}{1+1+1} = \frac{2}{3} \\
\mathbf{w} &= \frac{1}{6}\mathbf{v_1} + \frac{5}{2}\mathbf{v_1} + \frac{2}{3}\mathbf{v_1}
\end{aligned}
$$

We can use the notation below to show the coefficients of **w** in the orthogonal basis β.

$$
[\mathbf{w}]_\beta = \begin{bmatrix} 1/6 \\ 5/2 \\ 2/3 \end{bmatrix}
$$

A set of vectors is an **orthonormal set** if it is an orthogonal set of unit vectors. If the set is also a basis of subspace W of $\mathbb{R}^n$, then it is an **orthonormal set**. A square matrix Q whose columns form an orthonormal set is called an **orthogonal matrix**. The columns of a matrix Q form an orthonormal set if and only if $\mathbf{Q^TQ} = \mathbf{I}$. An important fact about orthogonal matrixes is that it implies that $\mathbf{Q^{-1}} = \mathbf{Q^T}$. Thus, to prove a matrix is orthogonal, we can just show that this theorem is true. We can also prove the following.

Q^{-1} is Orthonormal

Proof:

$$
\begin{aligned}
Q^{-1} &= Q^T \\
(Q^{-1})^{-1} &= Q \\
(Q^T)^T &= Q \\
(Q^{-1})^{-1} &= (Q^T)^T \\
(Q^{-1})^{-1} &= (Q^{-1})^T
\end{aligned}
$$

Thus, Q^{-1} is also orthonormal.

det Q = ±1
Proof:

$$\begin{aligned} det(Q^TQ) &= det(I) = 1 \\ det(Q^T)det(Q) &= 1 \\ (det(Q))^2 &= 1 \\ det(Q) &= \pm 1 \end{aligned}$$

Thus, det Q = ±1.

If λ is an eigenvalue of Q, then $|\lambda| = 1$. Orthogonal matrixes imply that $||Q\mathbf{x}|| = ||\mathbf{x}||$
Proof:

$$\begin{aligned} &Q\mathbf{v} = \lambda\mathbf{v} \\ &||\mathbf{v}|| = ||Q\mathbf{v}|| = ||\lambda\mathbf{v}|| = |\lambda|||\mathbf{v}|| \\ &|\lambda| = 1 \end{aligned}$$

Thus, $|\lambda| = 1$.

If Q_1 and Q_2 are both orthogonal, then so is Q_1Q_2.
Proof:

$$\begin{aligned} Q_1^TQ_1 &= I = Q_2^TQ_2 \\ (Q_1Q_2)^T(Q_1Q_2) &= Q_1^TQ_2^TQ_1Q_2 \\ &= Q_1^TQ_1Q_2^TQ_2 \\ &= I \end{aligned}$$

Thus, Q_1Q_2 is orthogonal.

Previously, one way we described **v** was as the sum of its perpendicular and parallel components onto another vector **u**. We called the parallel component of **v** onto **u** as its projection.

$$\mathbf{v} = proj(\mathbf{v})_u + perp(\mathbf{v})_u$$

We can do the same with vectors onto a subspace of $\mathbb{R}^n$. If we let W be a subspace of $\mathbb{R}^n$ and $[\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_k}]$ be an orthogonal basis for W, then for any vector **v** in $\mathbb{R}^n$, the **orthogonal projection of v onto W** is defined as

$$proj(\mathbf{v})_W = \left(\frac{\mathbf{u_1} \cdot \mathbf{v}}{\mathbf{u_1} \cdot \mathbf{u_1}}\right)\mathbf{u_1} + ... + \left(\frac{\mathbf{u_k} \cdot \mathbf{v}}{\mathbf{u_k} \cdot \mathbf{u_k}}\right)\mathbf{u_k}$$

Using the concepts above, we can now construct an orthonormal basis for a subspace W, given its span. Essentially, we just need to take the first vector in the span and find a second vector that is orthogonal to it by taking just its perpendicular component. Thus, if the span for W is $span(\mathbf{x_1}, \mathbf{x_2})$, then we can set the first vector in the orthogonal basis $\mathbf{v_1} = \mathbf{x_1}$ and find $perp(\mathbf{x_2})_{\mathbf{x_1}}$. This will be the second vector $\mathbf{v_2}$ in our orthogonal basis. We can continue for the rest of span. This process is known as the **Gram-Schmidt Process**.

If $[\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}]$ is a basis for a subspace W of $\mathbb{R}^n$, then

$$\begin{aligned}
&\mathbf{v_1} = \mathbf{x_1} \\
&\mathbf{v_2} = \mathbf{x_2} - \left(\frac{\mathbf{v_1} \cdot \mathbf{x_2}}{\mathbf{v_1} \cdot \mathbf{v_1}}\right)\mathbf{v_1} \\
&\mathbf{v_3} = \mathbf{x_3} - \left(\frac{\mathbf{v_1} \cdot \mathbf{x_3}}{\mathbf{v_1} \cdot \mathbf{v_1}}\right)\mathbf{v_1} - \left(\frac{\mathbf{v_2} \cdot \mathbf{x_3}}{\mathbf{v_2} \cdot \mathbf{v_2}}\right)\mathbf{v_2} \\
&\ldots
\end{aligned}$$

For example, if we have the subspace W $= span\,(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3})$, where $\mathbf{x_1} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{x_2} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{x_3} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix}$, then

$$\begin{aligned}
&\mathbf{v_1} = \mathbf{x_1} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 2 \end{bmatrix} \\
&\mathbf{v_2} = \mathbf{x_2} - \left(\frac{\mathbf{v_1} \cdot \mathbf{x_2}}{\mathbf{v_1} \cdot \mathbf{v_1}}\right)\mathbf{v_1} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \\ 1/2 \\ 1/2 \end{bmatrix}
\end{aligned}$$

We can scale $\mathbf{v_2}$ to make our calculations more convenient. $\mathbf{v_2}' = 2\mathbf{v_2}$

$$\mathbf{v_3} = \mathbf{x_3} - \left(\frac{\mathbf{v_1} \cdot \mathbf{x_3}}{\mathbf{v_1} \cdot \mathbf{v_1}}\right)\mathbf{v_1} - \left(\frac{\mathbf{v_2}' \cdot \mathbf{x_3}}{\mathbf{v_2}' \cdot \mathbf{v_2}'}\right)\mathbf{v_2}' = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \\ 1 \end{bmatrix}$$

Our orthogonal basis for subspace W is $[\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}]$. We can normalize each vector to then create an orthonormal basis.

We can utilize the Gram Schmidt Process to factorize a matrix A. If we assume that A is an $m \times n$ matrix with linearly independent columns, then A can be factorized into A = QR, where Q is another matrix with orthonormal columns and R is an upper triangular matrix. This process is known as **QR Factorization**.

For example, if we wanted to find the QR Factorization of A = $\begin{bmatrix} 1 & 2 & 2 \\ -1 & 1 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ (The vectors from the previous problem), we would first need an orthonormal basis for col(A), which we found before through the Gram-Schmidt Process.

$$\mathbf{q_1} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \mathbf{q_2} = \begin{bmatrix} 3\sqrt{5}/10 \\ 3\sqrt{5}/10 \\ \sqrt{5}/10 \\ \sqrt{5}/10 \end{bmatrix} \mathbf{q_3} = \begin{bmatrix} -\sqrt{6}/6 \\ 0 \\ \sqrt{6}/6 \\ \sqrt{6}/3 \end{bmatrix}$$

$$Q = \begin{bmatrix} 1/2 & 3\sqrt{5}/10 & -\sqrt{6}/6 \\ -1/2 & 3\sqrt{5}/10 & 0 \\ -1/2 & \sqrt{5}/10 & \sqrt{6}/6 \\ 1/2 & \sqrt{5}/10 & \sqrt{6}/3 \end{bmatrix}$$

$$Q^T A = Q^T QR = IR = R$$

$$R = Q^T A = \begin{bmatrix} 1/2 & -1/2 & -1/2 & 1/2 \\ 3\sqrt{5}/10 & 3\sqrt{5}/10 & \sqrt{5}/10 & \sqrt{5}/10 \\ -\sqrt{6}/6 & 0 & \sqrt{6}/6 & \sqrt{6}/3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ -1 & 1 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 1/2 \\ 0 & \sqrt{5} & 3\sqrt{5}/2 \\ 0 & 0 & \sqrt{6}/2 \end{bmatrix}$$

Thus,

$$A = \begin{bmatrix} 1/2 & 3\sqrt{5}/10 & -\sqrt{6}/6 \\ -1/2 & 3\sqrt{5}/10 & 0 \\ -1/2 & \sqrt{5}/10 & \sqrt{6}/6 \\ 1/2 & \sqrt{5}/10 & \sqrt{6}/3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1/2 \\ 0 & \sqrt{5} & 3\sqrt{5}/2 \\ 0 & 0 & \sqrt{6}/2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & 1 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Orthogonal vectors makes diagonalization of matrices easier. A square matrix is **orthogonally diagonalizable** if there exists and orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$. Note that since $Q^{-1} = Q^T$, the diagonalization $Q^{-1}AQ = D$ is still valid. For example, if we have a matrix A = $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, then

We begin by finding its eigenvalues

$$det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}\right) = 0$$

$$(2-\lambda)^2 - 1 = 0$$

$$\lambda^2 - 4\lambda + 3 = 0$$

$$(\lambda-3)(\lambda-1) = 0$$

$$\lambda = 3, 1$$

Now we find its eigenvectors

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{bmatrix} = 0$$

$$\begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{bmatrix} = 0$$

$$\begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Q must be columns of eigenvectors that are orthonormal to each other, so we must normalize the eigenvectors to get

$$\mathbf{q_1} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \mathbf{q_2} = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$$

$$Q = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus,

$$D = Q^T A Q = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

The **Spectral Theorem** states that a square matrix A is symmetric if and only if it is orthogonally diagonalizable. We can write A in a different form called the **spectral decomposition**, which is given by

$$A = \lambda \mathbf{q_1}\mathbf{q_1}^T + \lambda \mathbf{q_2}\mathbf{q_2}^T + ... + \lambda \mathbf{q_n}\mathbf{q_n}^T$$

If we take the normalized eigenvectors from before, we can find its spectral decomposition like so.

$$\mathbf{q_1} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \mathbf{q_2} = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$$

$$A = 3 \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} + 1 \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

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

If we, for example, want to find a symmetric matrix whose eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = -2$ and has eigenvectors $\mathbf{v_1} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{v_2} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

We have to normalize the eigenvectors before finding the spectral decomposition.

$$\mathbf{q_1} = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix} \mathbf{q_2} = \begin{bmatrix} -4/5 \\ 3/5 \end{bmatrix}$$

$$A = 3 \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix} \begin{bmatrix} 3/5 & 4/5 \end{bmatrix} - 2 \begin{bmatrix} -4/5 \\ 3/5 \end{bmatrix} \begin{bmatrix} -4/5 & 3/5 \end{bmatrix}$$

$$A = 3 \begin{bmatrix} 9/25 & 12/25 \\ 12/25 & 16/25 \end{bmatrix} - 2 \begin{bmatrix} -16/25 & -12/25 \\ -12/25 & 9/25 \end{bmatrix}$$

$$A = \begin{bmatrix} -1/5 & 12/5 \\ 12/5 & 6/5 \end{bmatrix}$$

An expression of the form $ax^2+by^2+cz^2+dxy+exz+fyz$ is called **quadratic form**. We can express a function with the **matrix associated with f**, in the form

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

where A = $\begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}$ if the quadratic expression is in form $ax^2 + by^2 + cxy$ and A = $\begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix}$ if the quadratic expression is in the form $ax^2 + by^2 + cz^2 + dxy + exz + fyz$.

For example, the matrix A= $\begin{bmatrix} 2 & 3 & 3/2 \\ 3 & -1 & 0 \\ 3/2 & 0 & 5 \end{bmatrix}$ represents the quadratic form $3x^2 - y^2 + 5z^2 + 6xy + 3xz$.

14 Chapter 6: Vector Spaces

Previously, we saw that we can perform similar operations on both vectors and matrices. We can generalize the term "vector" to refer to any abstract set of examples. Let V be the set on which two operations **addition and scalar multiplication** have been defined. If **u** and **v** are in V, then the sum is denoted by **u+v** and the scalar multiple of **u** is denoted by c**u**. If the following axioms hold for all vectors in V, then V is known as a **vector space** and its elements are known as **vectors**.

$$\mathbf{u} + \mathbf{v} \text{ is in } V$$
$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$
$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

There exists an element **0** V in called a **zero vector** such that **u + 0 = u**

For each **u** in V, there is an element**-u** in V such that **u + (-u) = 0**

$$c\mathbf{u} \text{ is in } V$$
$$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + d\mathbf{v}$$
$$(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$$
$$c(d\mathbf{u})) = (cd)\mathbf{u}$$
$$1\mathbf{u} = \mathbf{u}$$

For example, let P_2 denote the set of all polynomials of degree 2 or less with real coefficients. if $p(x) = a_0 + a_1x + a_2x^2$ and $q(x) = b_0 + b_1x + b_2x^2$, then

$$\begin{aligned}
p(x) + q(x) &= a_0 + b_0 + a_1x + b_1x + a_2x^2 + b_2x^2 \in V \\
p(x) + q(x) &= a_0 + a_1x + a_2x^2 + b_0 + b_1x + b_2x^2 \\
&= b_0 + b_1x + b_2x^2 + a_0 + a_1x + a_2x^2 = q(x) + p(x) \\
&... \\
cp(x) &= ca_0 + ca_1x + ca_2x^2 \in V \\
&...
\end{aligned}$$

Axioms 1, 2 and 6 are shown above. The other axioms are easy to see as well. Thus P_2 is a vector space. In general, the set P_n of all polynomials of degree less than or equal to n is a vector space.

We can also redefine subspaces with this new abstraction. A subset W of a vector space V is called a **subspace** if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.

Spanning sets also carry over to the notion of a general vector space. If S = $[\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}]$ is a set of vectors in a vector space, then the set of all linear combinations is called the **span** of the set and is denoted by span(S).

For example, if $p(x) = 1 - x + x^2$ and $q(x) = 2 + x - 3x^2$ and we wanted to determine whether $r(x) = 1 - 4x + 6x^2$ was in $span(p(x), q(x))$, then we simply need to find scalars c and d such that

$$\begin{aligned} c(1 - x + x^2) + d(2 + x - 3x^2) &= 1 - 4x + 6x^2 \\ (c + 2d) + (-c + d)x + (c - 3d)x^2 &= 1 - 4x + 6x^2 \\ c = 3d &= -1 \\ r(x) &\in span(p(x), q(x)) \end{aligned}$$

The notion of linear dependence also carries over. If a set of vectors in vector space V is **linearly independent**, at least one of the vectors can be expressed a linear combination of the others. If we want to show that the set $[1, x, x^2, ..., x^n]$ is linearly independent in P_n, we can assume linear dependence first, like so

$$\begin{aligned} p(x) &= c_0 + c_1 x + c_2 x^2 + ... + c_n x^n = 0 \\ x &= 0 \; c_0 = 0 \\ p(x) &= c_1 x + c_2 x^2 + ... + c_n x^n = 0 \\ p^{'}(x) &= c_1 + 2c_2 x + ... + nc_n x^{n-1} = 0 \\ x &= 0 \; c_1 = 0 \end{aligned}$$

If we repeat the process, we see that all coefficients are 0, proving linear independence.

A subset β of a vector space V is a **basis** for V if β spans V and β is linearly independent. S $= [1, x, x^2, ..., x^n]$ is called the **standard basis** for P_n.

For example, if we want to show that $\beta = [1 + x, x + x^2, 1 + x^2]$ is a basis for P_2, we must show that the vectors are linearly independent and spans P_2.

Linear Independence:

$$\begin{aligned} c_1(1 + x) + c_2(x + x^2) + c_3(1 + x^2) &= 0 \\ (c_1 + c_3) + (c_1 + c_2)x + (c_2 + c_3)x^2 &= 0 \\ c_1 = 0 \; c_2 = 0 \; c_3 &= 0 \end{aligned}$$

Span:

$$\begin{aligned} c_1(1 + x) + c_2(x + x^2) + c_3(1 + x^2) &= a + bx + cx^2 \\ (c_1 + c_3) + (c_1 + c_2)x + (c_2 + c_3)x^2 &= a + bx + cx^2 \\ c_1 + c_3 &= a \\ c_1 + c_2 &= b \\ c_2 + c_3 &= c \end{aligned}$$

The coefficient matrix has rank 3 and, thus, a solution exists, implying β is a basis for P_n

If $\beta = [\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}]$ is a basis for vector space V, and $\mathbf{v}$ is a vector such that $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k}$, then $c_1, c_2, ..., c_k$ are the **coordinates of v with respect to** β. The following vector is known as the **coordinate vector of v with respect to** β.

$$[\mathbf{v}]_\beta = \begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_k \end{bmatrix}$$

The number of vectors in a basis of vector space V is known as its **dimension**. A vector space with a finite number of vectors in its basis is **finite dimensional**. If it is not finite dimensional, then it is **infinite dimensional**. Dimensions are denoted by **dim V**.

Oftentimes, we may find a problem difficult to solve in a current coordinate system and we may find it easier to switch to a different coordinate system to make things more convenient. A **change of basis** allows us to do this. If we have basis $\beta = [\mathbf{u_1}, \mathbf{u_2}]$ and another basis $C = [\mathbf{v_1}, \mathbf{v_2}]$ where $\mathbf{u_1} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathbf{u_2} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{v_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $[\mathbf{x}]_\beta = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, then

We first have to find the old basis in terms of the new basis

$$\mathbf{u_1} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -3\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = -3\mathbf{v_1} + 2\mathbf{v_2}$$
$$\mathbf{u_2} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\mathbf{v_1} - \mathbf{v_2}$$

We can put the coefficients of these vectors in a matrix, known as the **change of basis** matrix.

$$P_{C-\beta} = \begin{bmatrix} -3 & 3 \\ 2 & -1 \end{bmatrix}$$

Now if we multiply our old coefficient vector with the change of basis matrix, we will get our new vector in its new basis.

$$[\mathbf{x}]_C = P_{C-\beta}[\mathbf{x}]_\beta$$
$$[\mathbf{x}]_C = \begin{bmatrix} -3 & 3 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ -1 \end{bmatrix}$$

A special fact about change of basis matrixes is that $(P_{C-\beta})^{-1} = P_{\beta-C}$

We can use the **Gauss Jordan Method** to calculate the change of basis matrix. If we have a basis β and C, then B is a matrix whose columns are the vectors of β in terms of any basis for V ϵ and C is a matrix whose columns are the vectors of C in terms of any basis for V ϵ. We can also see that $[C|B] - [I|P_{C-\beta}]$

A **linear transformation** from one vector space to another is a mapping T such that for all **u** and **v**

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$
$$T(c\mathbf{u}) = cT(\mathbf{u})$$

For example, if T is defined by $T(A) = A^T$

$$T(A+B) = (A+B)^T = A^T + B^T = T(A) + T(B)$$
$$T(cA) = (cA)^T = cA^T = cT(A)$$

Thus, T is a linear transformation.

Another example. If we let D be the **differential operator** and be defined by $D(f) = f^{'}$, then

$$D(f+g) = (f+g)^{'} = f^{'} + g^{'} = D(f) + D(g)$$
$$D(cf) = (cf)^{'} = cf^{'} = cD(f)$$

Thus, D is a linear transformation as well.

Another example. If we let S be defined by $S(f) = \int_a^b f(x)dx$

$$\begin{aligned} S(f+g) &= \int_a^b (f+g)(x)dx \\ &= \int_a^b (f(x)+g(x))dx \\ &= \int_a^b f(x)dx + \int_a^b g(x)dx \\ &= S(f) + S(g) \\ S(cf) &= \int_a^b (cf)(x)dx \\ &= \int_a^b c(f(x))dx \\ &= c\int_a^b (f(x))dx \\ &= cS(f) \end{aligned}$$

Thus, S is a linear transformation as well.

Transformations can be composed together as well. If T: U - V and S: V - W are linear transformations, then the **composition of S with T** is the mapping $S \circ T$ defined by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$$

We can also extend the notions of null space and column space to the kernel and range of a linear transformation. If we let T:V - W be a linear transformation, then the **kernel** of T denoted by $ker(T)$ is the set of all vectors in V that are mapped by T to 0 in W. The **range** of T, denoted by $range(T)$ is the set of all vectors in E that are the images of vectors vectors in V under T.

$$\begin{aligned} ker(T) &= [\mathbf{v} \text{ in } V : T(\mathbf{v} = 0)] \\ range(T) &= [T(v) : v \text{ in } V] \\ &= [\mathbf{w} \text{ in } W : \mathbf{w} = T(\mathbf{v})] \end{aligned}$$

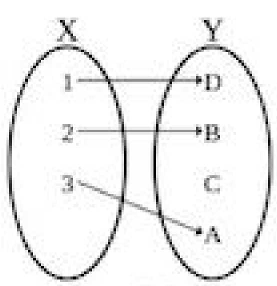
For example, if we have the differential operator D: $P_3 - P_2$ defined by $D(p(x)) = p^{'}(x)$ and we want to find the kernel and range, then

$$\begin{aligned} D(a + bx + cx^2 + dx^3) &= b + 2cx + 3dx^2 \\ ker(D) &= [a + bx + cx^2 + dx^3 : D(a + bx + cx^2 + dx^3) = 0] \\ ker(D) &= [a + bx + cx^2 + dx^3 : b + 2cx + 3dx^2 = 0] \\ b + 2cx + 3dx^2 &= 0 \ b = 2c = 3d = 0 \\ ker(D) &= [a : a \text{ in } \mathbb{R}^n] \end{aligned}$$

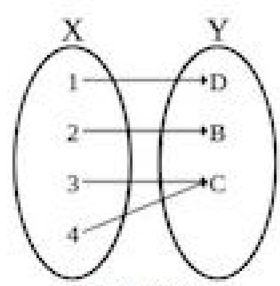
Thus, the kernel of D is the set of all constant polynomials.

The range of D is simple the subspace P_2 because every polynomial in P_2 is the image under D (derivative in this case) of some polynomial in P_3.

A linear transformation is said to be **one to one** if T maps distinct vectors in V to distinct vectors in W. If range(T) = W, then T is called **onto**.



one-to-one (injective), but not onto



onto (surjective), but not one-to-one

There is a rather faster way to check if a linear transformation is one to one. A linear transformation is one to one if and only if $ker(T) = 0$

From the above definitions of one to one and onto, we can say that a linear transformation T is **invertible** if and only if it is one to one and onto. If a linear transformation is one to one and onto, it is called an **isomorphism** and we say that the subspace V is isomorphic to W.

We can show that any linear transformation between finite dimensional vector spaces can be represented by a matrix transformation. If we let T be a linear transformation defined by $T : V - W$ and β and C be bases for V and W respectively, then $R(\mathbf{v}) = [\mathbf{v}]_\beta$ defines an isomorphism R and $S(\mathbf{w}) = [\mathbf{w}]_C$ Now that both matrixes are mapped to the set of real numbers, they can be mapped to each other through a matrix.

$$[T(\mathbf{v})]_C = [T]_{C-\beta}[\mathbf{v}]_C$$

The matrix is known as the **matrix of T with respect to the bases of** β **and** C. For example, let D: $P_3 - P_2$ be the differential operator $D(p(x)) = p^{'}(x)$. Let $\beta = [1, x, x^2, x^3]$ and $C = [1, x, x^2]$ be bases for P_3 and P_2 respectively.

First, find the images of the basis β under D

$$[D(1)] = 0 \; [D(x)] = 1 \; [D(x^2)] = 2x \; [D(x^3)] = 3x^2$$

Now find their coordinate vectors with respect to C.

$$[D(1)]_C = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \; [D(x)]_C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \; [D(x^2)]_C = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \; [D(x^3)]_C = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

Thus,

$$A = [D]_{C-\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

If we want to compute $D(5 - x + 2x^2)$, we can do so directly

$$D(5 - x + 2x^2) = -1 + 6x^2 = \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix}$$

Or we can use our newfound matrix transformation

$$[5 - x + 2x^3]_\beta = \begin{bmatrix} 5 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$

$$\begin{aligned} [D(5 - x + 2x^3)]_C &= [D]_{C-\beta}[5 - x + 2x^3]_\beta \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix} \end{aligned}$$

We can determine that $[S \circ T]_{D-\beta} = [S]_{D-C}[T]_{C-\beta}$

If we let T be a linear transformation from vector space V to vector space W, then $([T]_{c-\beta})^{-1} = [T^{-1}]_{\beta-C}$.

We can define the solution set of a **second order differential equation** as a subspace of F. If $y^{''} + ay^{'} + by = 0$ is a second order differential equation and λ_1 and λ_2 be the roots of the characteristic equation $\lambda^2 + a\lambda + b = 0$, then $[e^{\lambda_1 t}, te^{\lambda_2 t}]$ is a basis for the solution set. Thus, the solutions are of form $y = c_1 e^{\lambda_1 t} + c_2 te^{\lambda_2 t}$ For example, if we want to find all solutions of $y^{''} - 2y^{'} + 3y = 0$, then

$$\begin{aligned} \lambda^2 - 2\lambda + 3 = 0 \\ (\lambda - 3)(\lambda + 1) = 0 \\ \lambda = 3, -1 \end{aligned}$$

The solution takes the form

$$y = c_1 e^{3t} + c_2 e^{-t}$$

15 Chapter 7: Distance and Approximation

Previously, we defined the dot product of two vectors and used the operation often throughout the chapters. We can extend that operation to vector spaces other than $\mathbb{R}^2$. An **inner product** is an operation that assigns every pair of vectors $\mathbf{u}$ and $\mathbf{v}$ a real number $\langle \mathbf{u},\ \mathbf{v}\rangle$, such that the following properties are true.

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v}\rangle &= \langle \mathbf{v}, \mathbf{u}\rangle \\
\langle \mathbf{u}, \mathbf{v}+\mathbf{w}\rangle &= \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle \\
\langle c\mathbf{u}, \mathbf{v}\rangle &= c\langle \mathbf{u}, \mathbf{v}\rangle \\
\langle \mathbf{u}, \mathbf{u}\rangle &\geq 0
\end{aligned}$$

For example, if we define an operation $\langle \mathbf{u}, \mathbf{v}\rangle$ such that $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$, then this operation is an inner product space because

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v}\rangle &= 2u_1v_1 + 3u_2v_2 = 2v_1u_1 + 3v_2u_2 = \langle \mathbf{u}, \mathbf{v}\rangle \\
\langle \mathbf{u}, \mathbf{v}+\mathbf{w}\rangle &= 2u_1(v_1+w_1) + 3u_2(v_2+w_2) \\
&= 2u_1v_1 + 2u_1w_1 + 3u_2v_2 + 3u_2w_2 \\
&= 2u_1v_1 + 3u_2w_2 + 2u_1w_1 + 3u_2w_2 \\
&= \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle \\
\langle c\mathbf{u}, \mathbf{v}\rangle &= 2(cu_1)v_1 + 3(cu_2)v_2 \\
&= c(2u_1v_1 + 3u_2v_2) \\
&= c\langle \mathbf{u}, \mathbf{v}\rangle \\
\langle \mathbf{u}, \mathbf{u}\rangle &= 2u_1u_2 + 3u_2u_2 \\
&= 2u_1^2 + 3u_2^2 \geq 0 \\
u_1 = 0, u_2 &= 0, \langle \mathbf{u}, \mathbf{u}\rangle = 0
\end{aligned}$$

Thus, $\langle \mathbf{u}, \mathbf{v}\rangle$ defines an inner product space.

The **length** or **norm** of $\mathbf{v}$ is $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$. The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$. Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. A vector of length 1 is called a **unit vector**.

For example, if $\langle f, g\rangle = \int_a^b f(x)g(x)dx$, the inner product is on $C[0, 1]$, and $f(x) = x$ and $g(x) = 3x - 2$, then

$$
\begin{aligned}
||f|| &= \sqrt{\langle f, f \rangle} \\
&= \sqrt{\int_0^1 f^2(x)dx} \\
&= \sqrt{\int_0^1 x^2 dx} = \sqrt{\left.\frac{x^3}{3}\right|_0^1} \\
&= \sqrt{\frac{1}{3}} \\
d(f,g) &= ||f - g|| = \sqrt{\langle f-g, f-g \rangle} \\
&= \sqrt{\int_0^1 (2-2x)^2 dx} \\
&= \sqrt{\left. 4x - 4x^2 + \frac{4x^3}{3} \right|_0^1} \\
&= \sqrt{\frac{4}{3}} \\
\langle f, g \rangle &= \int_0^1 (3x^2 - 2)dx \\
&= x^3 - x^2 \Big|_0^1 = 0
\end{aligned}
$$

Thus, f and g are orthogonal.

Pythagoras' Theorem states that if **u** and **v** are vectors in an inner product space V, then they are orthogonal if and only if $||\mathbf{u} + \mathbf{v}|| = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$

The Cauchy Schwarz Inequality now states that $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq |\mathbf{u}||\mathbf{v}|$.

The Triangle Inequality stats that $||\mathbf{u} + \mathbf{v}|| \leq |\mathbf{u}| + |\mathbf{v}|$.

A **norm** on a vector space is a mapping that associates with each vector **v** a real number $||\mathbf{v}||$, called the **norm** of **v**, such that the following properties are satisfied.

$$
\begin{gathered}
||\mathbf{v}|| \geq 0 \text{ and } ||\mathbf{v}|| = 0 \text{ if and only if } \mathbf{v} = 0 \\
||c\mathbf{v}|| = |c|||\mathbf{v}|| \\
||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||
\end{gathered}
$$

A vector space with a norm is called a **normed linear space**.

The **sum norm**, or **1-norm**, of a vector $||\mathbf{v}||$ is the sum of the absolute values of its components. The **max norm**, or **∞-norm**, of a vector $||\mathbf{v}||$ is the largest number among the absolute values of its components.

In general, the definition of a norm

$$||\mathbf{v}||_p = (|v_1|^p + ... + |v_n|^p)^{1/p}$$

is known as the **p-norm**. A p-norm with p=2 is known as the **Euclidean norm**, or **2-norm**.

For any norm, we can define a **distance function**

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

The following properties hold for the distance function

$$d(\mathbf{u}, \mathbf{v}) \leq 0$$
$$d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$$
$$d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$$

A **matrix norm** is a mapping that associates a nxn matrix A a real number $||A||$, called the norm of A such that the properties are satisfied.

$$||A|| \geq 0$$
$$||cA|| = |c|||A||$$
$$||A + B|| \leq ||A|| + ||B||$$
$$||AB|| \leq ||A||||B||$$

The **Frobenius norm** of a matrix A is the square root of the sum of the squares of the entries of A. $||A||_1$ is the largest absolute column sum. $||A||_\infty$ is the largest absolution row sum.

A matrix A is **ill conditioned** if small changes in its entries can produced large changes in the solutions to $A\mathbf{x} = \mathbf{b}$. If small changes to A produces small changes in the solutions to $A\mathbf{x} = \mathbf{b}$ the A is **well conditioned**. We can determine the **condition number** of a matrix as

$$cond(A) = ||A^{-1}||||A||$$

If the condition number is large compared to one matrix norm, then it is large relative to any matrix norm. With the condition number, the following inequality holds:

$$\frac{||\Delta\mathbf{x}||}{||\mathbf{x}||} \leq cond(A)\frac{||\Delta A||}{||A||}$$

The **Best Approximation Theorem** states that if W is a subspace of an inner product space and if **v** is in V, then $proj_w(\mathbf{v})$ is the best approximation to **v** in W.

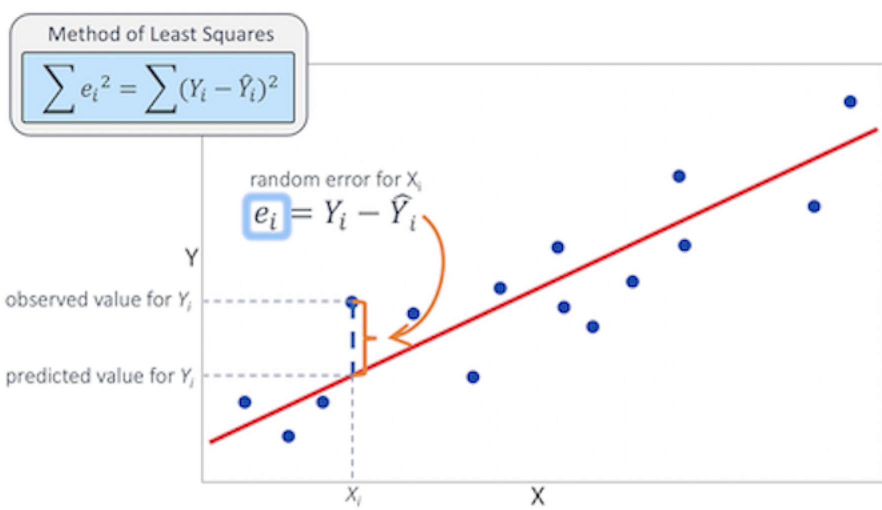
For example, if $\mathbf{u_1} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{u_2} = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, then the best approximation to **v** in $W = span(\mathbf{u_1}, \mathbf{u_2})$ is

$$\begin{aligned} proj_W(\mathbf{v}) &= \left(\frac{\mathbf{u_1} \cdot \mathbf{v}}{\mathbf{u_1} \cdot \mathbf{u_1}}\right)\mathbf{u_1} + \left(\frac{\mathbf{u_2} \cdot \mathbf{v}}{\mathbf{u_2} \cdot \mathbf{u_2}}\right)\mathbf{u_2} \\ &= \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \frac{8}{15}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ -2/5 \\ 1/5 \end{bmatrix} \end{aligned}$$

The distance from **v** to W is just

$$||\mathbf{v} - proj_W(\mathbf{v})|| = \sqrt{0^2 + (\frac{12}{5})^2 + (\frac{24}{5})^2} = \frac{12\sqrt{5}}{5}$$

We can now use what we've learned to find a curve that best fits a set of data points. The curve that best fits is simply a curve that minimizes the error between the line and the data points.



The **error vector** is then $\mathbf{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$.

To find the line of best fit, we need to make sure $||\mathbf{e}||$ as small as possible. We can use the Euclidean norm. The number $||\mathbf{e}||$ is called the **least squares error**.

For example, if $y = 1 + x$ and we have data points $(1, 2), (2, 2), (3, 4)$, then

$$\begin{aligned} \varepsilon &= y - (1 + x) \\ \varepsilon_1 &= 2 - (1 + 1) = 0 \\ \varepsilon_2 &= 2 - (1 + 2) = -1 \\ \varepsilon_3 &= 4 - (1 + 3) = 0 \\ \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 &= 0^2 + (-1)^2 + 0^2 = 1 \\ ||\mathbf{e}|| &= 1 \end{aligned}$$

The line $y = a + bx$ that minimizes the least squares error, where $\varepsilon_i = y_i - (a + bx_i)$, is known as the **least squares approximating line**.

If we let A be an $m \times n$ matrix and $\mathbf{b}$ be in $\mathbb{R}^m$, then $A\mathbf{x} = \mathbf{b}$ always as at least one least squares solution given by

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

For example, if we take the data points from before, then

$$\begin{aligned} y &= a + bx \\ A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} b &= \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\ A^T A &= \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} A^T \mathbf{b} = \begin{bmatrix} 8 \\ 18 \end{bmatrix} \\ \mathbf{x} &= \begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix} \\ y &= \frac{2}{3} + x \end{aligned}$$

We can use the same logic to find a parabolic curve of best fit. If we use the points $(-1,1),(0,-1),(1,0),(2,2)$.

$$y = a + bx + cx^2$$
$$a - b + c = 1$$
$$a = -1$$
$$a + b + c = 0$$
$$a + 2b + 4c = 2$$
$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} b = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$
$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$
$$\mathbf{x} = \begin{bmatrix} -\frac{7}{10} \\ -\frac{3}{5} \\ 1 \end{bmatrix}$$
$$y = -\frac{7}{10} - \frac{3}{5}x + x^2$$

We can also achieve a solution using QR Factorization. If A is an $m \times n$ matrix and A=QR is a QR factorization of A, then the unique least squares solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x} = R^{-1}Q^T\mathbf{b}$$

We can assign a value to $(A^TA)^{-1}A^T$ as the **pseudoinverse** of A, denoted by A^+.

We can now factor every matrix, symmetric or not, square or not, in the form $A = PDQ^T$. We defined SVD factorization as the **singular value decomposition** of a matrix. We must begin by. first finding the **singular values** of a matrix, which are simply just the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \sigma_2, ...$ They are conventionally ordered by increasing magnitude.

For example, matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

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

$$\lambda_1 = 3$$
$$\lambda_2 = 1$$
$$\sigma_1 = \sqrt{3}$$
$$\sigma_2 = 1$$

We can now factor an $m \times n$ matrix A as $A = U\Sigma V^T$, where U is an $m \times m$ matrix, V is an $n \times n$ matrix, and Σ is an $m \times n$ matrix. The diagonal elements of Σ will hold the singular values of A. The matrix V is simple the normalized eigenvectors of A^TA. We can compute U with the expression:

$$\mathbf{u_n} = \frac{1}{\sigma_1} A\mathbf{v_n}$$

If the vectors of u do not form an orthonormal basis of $\mathbb{R}^n$, then we must use the Gram-Schmidt Process to orthonormalize them.

For example, if we have the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, then

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

Its eigenvalues are $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0$ and its eigenvectors are

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

we normalize them to find

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

Thus, our V and Σ matrices are given by

$$V = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix}$$
$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

We can compute U like so

$$
\begin{aligned}
\mathbf{u_1} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1\sqrt{2} \\ 1\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
\mathbf{u_2} &= \frac{1}{1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
U &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}
$$

Thus,

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

Similar to spectral decomposition, we also have the **outer product form of the SVD**, given by:

$$
A = \sigma_1 \mathbf{u_1} \mathbf{v_1^T} + ... + \sigma_r \mathbf{u_r} \mathbf{v_r^T}
$$

We can also use the pseudoinverse of A to be to be $A^+ = V\Sigma^+U^T$, where Σ^+ is simple the matrix $\begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}$