1.6 Revision of Vectors

EXAMPLE. Find a unit vector in the direction of the following vectors:

(i) $\vec{v}_1 = (2, -3)$.

(ii) $\vec{v}_2 = (1, -2, 5)$.

(iii) $\vec{v}_3 = (1, 1, 1, 1)$.

SOLUTION: To get a unit vector in the direction of a given vector, all we need to do is scale the size of the vector by the inverse of its length.

(i) $\vec{v}_1 = (2, -3)$.
The length of $\vec{v}_1$ is $|\vec{v}_1| = \sqrt{2^2 + (-3)^2} = \sqrt{13}$.
A unit vector u_1 in the direction of v_1 may then be obtained as follows:

$$u_1 = \frac{1}{\sqrt{13}}(2, -3) = (\frac{2}{\sqrt{13}}, \frac{-3}{\sqrt{13}}).$$

(ii) $\vec{v}_2 = (1, -2, 5)$.
The length of $\vec{v}_2$ is $|\vec{v}_2| = \sqrt{1^2 + (-2)^2 + 5^2} = \sqrt{30}$.
A unit vector u_2 in the direction of v_2 may then be obtained as follows:

$$u_2 = \frac{1}{\sqrt{30}}(1, -2, 5) = (\frac{1}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{5}{\sqrt{30}}).$$

(iii) $\vec{v}_3 = (1, 1, 1, 1)$.
The length of $\vec{v}_3$ is $|\vec{v}_3| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = \sqrt{4} = 2$.
A unit vector u_3 in the direction of v_3 may then be obtained as follows:

$$u_3 = \frac{1}{2}(1, 1, 1, 1) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}).$$

EXAMPLE. Consider the two vectors $\vec{a} = (2, 1, -4)$ and $\vec{b} = (3, -2, 5)$ in 3-dimensional space. Find the cross-product of the two vectors.

SOLUTION:

$$\begin{aligned}
\vec{a} \times \vec{b} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & 1 & -4 \\ 3 & -2 & 5 \end{vmatrix} \\
&= \begin{vmatrix} 1 & -4 \\ -2 & 5 \end{vmatrix} \vec{i} - \begin{vmatrix} 2 & -4 \\ 3 & 5 \end{vmatrix} \vec{j} + \begin{vmatrix} 2 & 1 \\ 3 & -2 \end{vmatrix} \vec{k} \\
&= -3\vec{i} - 22\vec{j} - 7\vec{k} \\
&= (-3, -22, -7)
\end{aligned}$$

Quiz

1. Find $\frac{d}{dx}(\tan(x^2)\cos(e^x)$.

   SOLUTION:

$$\begin{aligned} \frac{d}{dx}(\tan(x^2)\cos(e^x) &= \tan(x^2)\frac{d}{dx}(\cos(e^x) + \cos(e^x)\frac{d}{dx}(\tan(x^2)) \\ &= \tan(x^2)(-\sin(e^x))e^x + \cos(e^x)(\sec^2(x^2))2x \end{aligned}$$

2. Find the stationary points of $y = x^3 - 3x$.

   SOLUTION: Let $f(x) = x^3 - 3x$. Since $f(x)$ is continuous, the stationary points of $f(x)$ are the solutions of the equation $f'(x) = 0$.

$$\begin{aligned} f(x) &= x^3 - 3x \\ f'(x) &= 3x^2 - 3 \\ 3x^2 - 3 &= 0 \\ 3(x-1)(x+1) &= 0 \end{aligned}$$

   The stationary points are: $\{1, -1\}$.

3. Find $\int x \ln x dx$.

   SOLUTION:

$$\begin{aligned} u &= \ln(x); \qquad v = \frac{x^2}{2} \\ du &= \frac{dx}{x}; \qquad dv = xdx \\ \int x\ln(x)dx &= \int udv = uv - \int vdu \\ &= \frac{x^2\ln(x)}{2} - \int \frac{x^2}{2}\frac{dx}{x} \\ &= \frac{x^2\ln(x)}{2} - \int \frac{x}{2}dx \\ &= \frac{x^2\ln(x)}{2} - \frac{x^2}{4} + C. \end{aligned}$$

4. Find $\int (1-x^2)^{-1/2}dx$.

SOLUTION:
$$\int \frac{1}{\sqrt{a^2-u^2}}du = \sin^{-1}(\frac{u}{a}) + c.$$
Therefore,
$$\int (1-x^2)^{-1/2}dx = \sin^{-1}(x) + c.$$

5. Find the determinant of $\begin{pmatrix} 2 & -1 \\ -6 & 2 \end{pmatrix}$.

SOLUTION:
$$\begin{vmatrix} 2 & -1 \\ -6 & 2 \end{vmatrix} = (2)(2) - (-1)(-6) = -2$$

6. Let $\vec{u} = (2,3,5)$, $\vec{v} = (1,1,-1)$. Find

(a) The unit vector in the direction of $\vec{v} = (1,1,-1)$ is,
$$\frac{1}{\sqrt{1^2+1^2+(-1)^2}}(1,1,-1) = \frac{1}{\sqrt{3}}(1,1,-1) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}).$$

(b)
$$3\vec{u} - 5\vec{v} = 3(2,3,5) - 5(1,1,-1) = (6,9,15) - (5,5,-5) = (1,4,20).$$

(c)
$$\vec{u}\cdot\vec{v} = (2,3,5).(1,1,-1) = 2+3-5 = 0.$$

(d)
$$\begin{aligned} \vec{u}\times\vec{v} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & 3 & 5 \\ 1 & 1 & -1 \end{vmatrix} \\ &= \begin{vmatrix} 3 & 5 \\ 1 & -1 \end{vmatrix}\vec{i} - \begin{vmatrix} 2 & 5 \\ 1 & -1 \end{vmatrix}\vec{j} + \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix}\vec{k} \\ &= (3(-1)-(5)(1))\vec{i} - (2(-1)-(5)(1))\vec{j} + ((2)(1)-(3)(1))\vec{k} \\ &= -8\vec{i} + 7\vec{j} - \vec{k} \\ &= (-8,7,-1) \end{aligned}$$

7. Determine if $\vec{u} = (-2, 2, 1, -1)$ is perpendicular to $\vec{v} = (-2, -3, 1, 1)$.

SOLUTION:

$$\vec{u}.\vec{v} = (-2)(-2) + (2)(-3) + (1)(1) + (-1)(1) = -2 \neq 0.$$

Hence the two vectors are not perpendicular.

2 Topics in 3D Geometry

In two dimensional space, we can graph curves and lines. In three dimensional space, there is so much extra space that we can graph planes and surfaces in addition to lines and curves. Here we will have a very brief introduction to Geometry in three dimensions.

2.1 Planes

Just as it is easy to write the equation of a line in 2D space, it is easy to write the equation of a plane in 3D space.

The point-normal equation of a plane

A vector perpendicular to a plane is said to be *normal to the plane* and is called a *normal vector*, or simply a *normal*.

To write the equation of a plane we need a point $P(x_0, y_0, z_0)$ on the plane and a normal vector $\vec{n} = (a, b, c)$ to the plane.

Let $P = (x_0, y_0, z_0)$ be a point on the plane and $\vec{n}$ be a vector perpendicular to the plane. Then a point $Q(x, y, z)$ lies on the plane,

$\Leftrightarrow$ the vector $\vec{PQ}$ lies on the plane,

$\Leftrightarrow$ $\vec{PQ}$ and $\vec{n}$ are perpendicular,

$\Leftrightarrow$ $\vec{n} \cdot \vec{PQ} = 0$,

$\Leftrightarrow$ $(a, b, c) \cdot (x - x_0, y - y_0, z - z_0) = 0$,

$\Leftrightarrow$ $a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$.

Definition. *The point-normal equation of a plane that contains the point $P(x_0, y_0, z_0)$ and has normal vector $\vec{n} = (a, b, c)$ is*

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Example. Let P be a plane determined by the points $A = (1, 2, 3)$, $B = (2, 3, 4)$, and $C = (-2, 0.3)$. Find a vector which is normal to the plane. Find an equation of the plane.

Solution: We need a point on the plane and a normal to the plane. The vector $\vec{AB} \times \vec{AC} = (2, -3, 1)$ is a normal to the plane and we take $A = (1, 2, 3)$ as a point on the plane (you can choose B or C instead of A if you want). The equation on the plane in point-normal form is:

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

or equivalently,

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

Observe that the coefficients of x, y and z are $(2, -3, 1)$ which is the normal to the plane.

2.2 Lines

Vector equation of a line

To write the vector equation of a line, we need a point $P(x_0, y_0, z_0)$ on the line and a vector $\vec{v} = (a, b, c)$ that is parallel to the line.

Definition. *The vector equation of a line that contains the point $P(x_0, y_0, z_0)$ and is parallel to the vector $\vec{v} = (a, b, c)$ is:*

$$P + t\vec{v} = \vec{r}, \quad \textit{where } t \textit{ is scalar.}$$

or,

$$\begin{aligned} (x_0, y_0, z_0) + t(a, b, c) &= (x, y, z) \\ (x_0 + ta, y_0 + tb, z_0 + tc) &= (x, y, z) \end{aligned}$$

Parametric equation of a line

The parametric equation of a line is derived from the vector equation of a line.

Definition. *The parametric equation of a line that contains the point $P(x_0, y_0, z_0)$ and is parallel to the vector $\vec{v} = (a, b, c)$ is:*

$$\begin{aligned} x &= x_0 + ta \\ y &= y_0 + tb \\ z &= z_0 + tc \end{aligned}$$

EXAMPLE. Let L which passes through the points $P(1,1,1)$ and $Q(3,2,1)$. Find a vector which is parallel to the line. Find the vector-equation and parametric equation of the line.

SOLUTION: The vector $\vec{PQ} = (2,1,0)$ is parallel to the line and we take the point $P(1,1,1)$ on the line.
The vector equation of the line:

$$(1,1,1) + t(2,1,0) = (x,y,z)$$

The parametric equations of the line:

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

EXAMPLE. Find the equation of the plane which contains the point $(0,1,2)$ and is perpendicular to the line $(1,1,1) + t(2,1,0) = (x,y,z)$.

2.3 Surfaces

The graph in 3D space of an equation in x, y and z is a surface. Often the graph is too difficut to draw, but here we sketch the graph of a few special types of equations whose graphs are easy to visualize.

Cylindrical surfaces

The graph in 3D space of an equation containing only one or two of the three variables x, y, z is called a *cylindrical surface.*

EXAMPLE. Plot $y = x^2$.

Plot $x^2 + y^2 = 5$.

Quadric Surfaces

The graph in 2D space of a second degree equation in x and y is an ellipse, parabola or hyperbola. In 3D space, the graph of a second degree equation in x, y and z is one of six quadric surfaces.

1. Ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$
2. Elliptic Cone $z^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$
3. Elliptic Paraboloid $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$
4. Hyperbolic Parabolid $z = \frac{y^2}{b^2} - \frac{x^2}{a^2}$
5. Elliptic Hyperboloid of one sheet $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$
6. Elliptic Hyperboloid of two sheets $-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

Cross-sections of some quadric surfaces

2.4 Functions of several variables

So far you have studied about functions of one variable, e.g.

$$f(x) = x + x^2.$$

You have learned how to graph these functions, perhaps how to determine their domain and range. You have gone much further in your quest to understand functions; you have learned how to differentiate them, then to calculate the maximum and minimum, then to integrate them. At the end of this term (in March) we will learn how to write a function as a sum of simpler functions, i.e. a Fourier Series Expansion of a function.

However now we will learn something new. We will learn about functions of several variables, e.g.

$$f(x,y) = x^2 + y^2, \quad g(x,y) = 2xy + 7, \quad h(x,y) = e^x + 2y.$$

You will very quickly see that although the concepts are new, the techniques are old and familiar.

Graphs and Level Sets

To draw the graph of $f(x) = x^2$, we drew the graph of the equation

$$y = x^2.$$

Similarly, to draw the graph of the equation $f(x,y) = x^2 + y^2$, we draw the graph of the equation

$$z = x^2 + y^2.$$

We now use the methods developed in the last lecture to draw graphs. Note however that it is difficult to graph general surfaces.

Remark. The graph of a function $f(x)$ of one variable is the graph of the equation $y = f(x)$, a curve in 2D space. The graph of a function $f(x,y)$ of two variables is the graph of the equation $z = f(x,y)$, a surface in 3D space. The graph of a function $f(x,y,z)$ is a set of points in 4D space and we cannot draw the graph.

One way to understand functions of two or more variables is by using *level sets*.

EXAMPLE. An example of level sets is a topographic map, which maps hills and valleys in a region by drawing curves indicating height or elevation. If $h(x, y)$ is the height function over a region, say the Himalayas, then if we mark all the points (x, y) on the ground at which the height of the mountain is $h(x, y) = 3000m$, we get the *level set* for $L = 3000$. If we draw the level sets for different heights, e.g 2000m, 3000m, 4000m, 5000m, 6000m, 7000m, 8000m, we get a rough topographic map for the Himalayas.

Note that we draw the level sets on the ground, i.e. in the domain.

DEFINITION. *We fix a number C. The level set of a function of two variables $f(x, y)$ is the set of points (x, y) in the domain which satisfy the equation*

$$f(x, y) = C.$$

For every real number C, we get a level set.

In general, for a fixed number C and a function of several variables f, we define the level set to be the collection of points in the domain which satisfy the equation $f = C$.

EXAMPLE. Let $f(x) = x^2$ and say the constant $c = 1$. The level set is the set of points such that

$$\begin{aligned} x^2 &= 1 \\ x &= 1, -1 \end{aligned}$$

Hence the level set for $c = 1$ is {1, -1}.

EXAMPLE. Find the level sets of the function $f(x, y) = x^2 + y^2$ for $C = 1, 4, 9$.

SOLUTION: The level sets are

$$\begin{array}{lcll} C = 1 & : & x^2 + y^2 = 1; & \text{a circle of radius 1} \\ C = 4 & : & x^2 + y^2 = 4; & \text{a circle of radius 2} \\ C = 9 & : & x^2 + y^2 = 9; & \text{a circle of radius 3} \end{array}$$

Example. Let $f(x, y, z) = x^2 + y^2 + z^2$ and say the constant $c = 1$. The level set is the set of points such that

$$x^2 + y^2 + z^2 \;=\; 1$$

Hence the level set for $c = 1$ is the set of points lying on the unit sphere.

2.5 Change of Coordinates

Before we describe cylindrical and spherical coordinate systems, we will recall the polar coordinate system in 2D space.

Polar Coordinates

The polar coordinate system is equivalent to the rectangular coordinate system. It locates points using two coordinates r and θ. The coordinate r is the distance from a point to the origin, and θ is the angle used in trignometry which measures the counterclockwise rotation from the positive X-axis.

Conversion from rectangular to polar coordinates:
Let (x, y) be a point in the rectangular coordinate system.

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

Conversion from polar to rectangular coordinates:
Let (r, θ) be a point in the polar coordinate system.

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned}$$

Example. Express in polar coordiantes the portion of the unit disc that lies in the first quadrant.

Solution: The region may be expressed in polar coordinates as

$$0 \le r \le 1; \qquad 0 \le \theta \le \pi/2$$

Example. Express in polar coordinates the function

$$f(x, y) = x^2 + y^2 + 2yx.$$

Solution: We substitute $x = r\cos\theta$ and $y = r\sin\theta$ in f, to get

$$f(r, \theta) = r^2\cos^2(\theta) + r^2\sin^2(\theta) + 2r^2\sin(\theta)\cos(\theta) = r^2(1 + \sin(2\theta)).$$

Cylindrical Coordinates

This is a three dimensional extension of plane polar coordinates. Cylindrical coordinates are given by the 3-tuple (r, θ, z), the polar coordinates of the X, Y plane and the rectangular coordinate z. Given a point (x, y, z) in 3-dimensional space, to calculate r and θ we project the point to the XY-plane $(x,\ y,\ z) \mapsto (x,\ y)$. Then calculate $(r,\ \theta)$ as in the previous section.

Cylindrical to Rectangular Conversion Formulas:

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z \end{aligned}$$

Rectangular to Cylindrical Conversion Formulas:

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}(\frac{y}{x}) \\ z &= z \end{aligned}$$

Example. 1. Express in cylindrical coordinates the function

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

2. Express in rectangular coordinates the equation

$$r = \sin\theta$$

Spherical coordinates

Spherical coordinates consist of the 3-tuple (ρ, θ, ϕ). These are determined as follows:

1. $\rho =$ the distance from the origin to the point.

2. $\theta =$ the same angle that we saw in polar/cylindrical coordinates.

3. $\phi =$ the angle between the positive z-axis and the line from the origin to the point.

Spherical to Rectangular Conversion Formulas:

$$\begin{aligned} x &= \rho\sin(\phi)\cos(\theta) \\ y &= \rho\sin(\phi)\sin(\theta) \\ z &= \rho\cos\phi \end{aligned}$$

Rectangular to Spherical Conversion Formulas:

$$\begin{aligned} \rho &= \sqrt{x^2+y^2+z^2} \\ \theta &= \tan^{-1}(\frac{y}{x}) \\ \phi &= \cos^{-1}(\frac{z}{\rho}) \end{aligned}$$

Example. 1. Express in spherical coordinates the function

$$f(x, y, z) = x^2 + y^2 + z^2$$

2. Express in rectangular coordinates the equation

$$\rho = 5$$

References

1. A complete set of notes on Pre-Calculus, Single Variable Calculus, Multivariable Calculus and Linear Algebra. Here is a link to the chapter on Lines, Planes and Quadric Surfaces. Also read the section on cylindrical and spherical coordiantes.
http://tutorial.math.lamar.edu/Classes/CalcIII/3DSpace.aspx.

2. A collection of examples, animations and notes quadric Surfaces.
Quadric Surfaces.

3. Another gallery of animated and graphical demonstrations of calculus and related topics, from the University of Minnesota.
http://www.math.umn.edu/%7Erogness/quadrics/.

4. Links to various resources on Calculus.
http://www.calculus.org/.

3 Partial Derivatives

3.1 First Order Partial Derivatives

A function $f(x)$ of one variable has a first order derivative denoted by $f'(x)$ or

$$\frac{df}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

It calculates the slope of the tangent line of the function f at x.

A function $f(x,y)$ of two variables has two first order partials $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$. Just like in the one variable case, the partial derivatives are also related to the *tangent plane* of the function at a point (x,y)

DEFINITION. $\frac{\partial f}{\partial x}$ *is defined as the derivative of f with respect to x with y treated as a constant.*

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

$\frac{\partial f}{\partial y}$ *is defined as the derivative of f with respect to y with x treated as a constant.*

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

Similarly for $f(x,y,z)$ we can define three first order partial derivatives: $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$.

EXAMPLE. 1.

$$\begin{aligned} f(x,y) &= x^2y^3 \\ \frac{\partial f}{\partial x} &= \frac{\partial x^2y^3}{\partial x} = y^3\frac{\partial x^2}{\partial x} = 2xy^3 \\ \frac{\partial f}{\partial y} &= \frac{\partial x^2y^3}{\partial y} = 3x^2y^2 \end{aligned}$$

2. $f(x,y,z) = ze^{2x+3y+4z}$

3. $f(x,y) = x^2 + y^2$

EXAMPLE. Consider the function $f(x,y) = x^2/y$. Calculate the first order partial derivatives and evaluate them at the point $P(2,1)$.

SOLUTION:

$$\begin{aligned}
f(x,y) &= \frac{x^2}{y} \\
\frac{\partial f}{\partial x} &= \frac{2x}{y} \\
\frac{\partial f}{\partial x}|_{\{x=2,y=1\}} &= 4 \\
\frac{\partial f}{\partial y} &= -\frac{x^2}{y^2} \\
\frac{\partial f}{\partial y}|_{\{x=2,y=1\}} &= -\frac{1}{9}
\end{aligned}$$

REMARK. Partial derivatives are used in the same manner as the derivative of a function of one variable. The partial of $f(x,y)$ with respect to x is the rate of change (or the slope) of f with respect to x as y stays constant. Similarly the partial of $f(x,y)$ with respect to y is the rate of change (or the slope) of f with respect to y as x stays constant. For instance in the above example, the slope of the function at the point $P(2,1)$ in the x direction is 4, and the slope of the function at P in the y direction is $-1/9$.

THEOREM *Let $f(x,y)$ be a function of two variables and let $P=(x_0,y_0,z_0)$ be a point on the graph. The equation of the tangent plane to the graph of $f(x,y)$ at the point P is given by,*

$$z - z_0 = \frac{\partial f}{\partial x}|_{\{x_0,y_0\}}(x-x_0) + \frac{\partial f}{\partial y}|_{\{x_0,y_0\}}(y-y_0).$$

We will see why this is true after we study about Directional Derivative and the Gradient. For now we accept it as a result and use it in some examples.

EXAMPLE. We consider the function $f(x,y) = x^3 + 2xy - y + 3$. Compute the tangent plane at the point $P_0 = (1,2,6)$ on the graph of the function.

SOLUTION: (SOLUTION:) We first check that the point $P_0(x_0,y_0,z_0) = (1,2,6)$ is on the graph of the function:

$$\begin{aligned}
f(1,2) &= 1^3 + 2.1.2 - 2 + 3 \\
&= 1 + 4 - 2 + 3 \\
&= 6
\end{aligned}$$

So the point P_0 does indeed lie on the graph of f.
Then the equation of the tangent plane is given by;

$$z - z_0 = \frac{\partial f}{\partial x}_{\{x_0,y_0\}}|(x - x_0) + \frac{\partial f}{\partial y}|_{\{x_0,y_0\}}(y - y_0).$$

$$\begin{aligned}
\frac{\partial f}{\partial x}|_{\{1,2\}} &= 3x^2 + 2y|_{\{1,2\}} = 7 \\
\frac{\partial f}{\partial y}|_{\{1,2\}} &= 2x - 1|_{\{1,2\}} = 1 \\
z - 6 &= 7(x - 1) + 1(y - 2) \\
&= 7x - 7 + y - 2 \\
z &= 7x + y - 3
\end{aligned}$$

Therefore the tangent plane to the graph of f at $(1, 2, 6)$ is

$$7x + y - z - 3 = 0.$$

References

1. Engineering Mathematics, by K. A. Stroud.
2. A complete set of notes on Pre-Calculus, Single Variable Calculus, Multivariable Calculus and Linear Algebra. Here is a link to the chapter on Partial Differentiation.
   http://tutorial.math.lamar.edu/Classes/CalcIII/PartialDerivatives.aspx.
3. A collection of examples, animations and notes on Multivariable Calculus.
   http://people.usd.edu/ jflores/MultiCalc02/WebBook/Chapter_15/
4. Links to various resources on Calculus.
   http://www.calculus.org/.

3.2 Higher Order Partial Derivatives

If f is a function of several variables, then we can find higher order partials in the following manner.

DEFINITION. *If $f(x,y)$ is a function of two variables, then $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are also functions of two variables and their partials can be taken. Hence we can differentiate them with respect to x and y again and find,*

$\frac{\partial^2 f}{\partial x^2}$ *, the derivative of f taken twice with respect to x,*

$\frac{\partial^2 f}{\partial x \partial y}$ *, the derivative of f with respect to y and then with respect to x,*

$\frac{\partial^2 f}{\partial y \partial x}$ *, the derivative of f with respect to x and then with respect to y,*

$\frac{\partial^2 f}{\partial y^2}$ *, the derivative of f taken twice with respect to y.*

We can carry on and find $\frac{\partial^3 f}{\partial x \partial y^2}$, which is taking the derivative of f first with respect to y twice, and then differentiating with respect to x, etc. In this manner we can find nth-order partial derivatives of a function.

THEOREM $\frac{\partial^2 f}{\partial x \partial y}$ *and* $\frac{\partial^2 f}{\partial y \partial x}$ *are called mixed partial derivatives. They are equal when* $\frac{\partial^2 f}{\partial x \partial y}$ *and* $\frac{\partial^2 f}{\partial y \partial x}$ *are continuous.*

NOTE. In this course all the fuunctions we will encounter will have equal mixed partial derivatives.

EXAMPLE. 1. Find all partials up to the second order of the function

$$f(x,y) = x^4 y^2 - x^2 y^6.$$

Solution:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 4x^3y^2 - 2xy^6 \\
\frac{\partial f}{\partial y} &= 2x^4y - 6x^2y^5 \\
\frac{\partial^2 f}{\partial x^2} &= 12x^2y^2 - 2y^6 \\
\frac{\partial^2 f}{\partial y \partial x} &= 8x^3y - 12xy^5 \\
\frac{\partial^2 f}{\partial y^2} &= 2x^4 - 30x^2y^4 \\
\frac{\partial^2 f}{\partial x \partial y} &= 8x^3 - 12xy^5
\end{aligned}$$

Notations:

$$\begin{aligned}
f_x &= \frac{\partial f}{\partial x} \\
f_y &= \frac{\partial f}{\partial y} \\
f_{xx} &= \frac{\partial^2 f}{\partial x^2} \\
f_{yy} &= \frac{\partial^2 f}{\partial y^2} \\
f_{xy} &= \frac{\partial^2 f}{\partial x \partial y}
\end{aligned}$$

3.3 Chain Rule

You are familiar with the chain rule for functions of one variable: if f is a function of u, denoted by $f = f(u)$, and u is a function of x, denoted $u = u(x)$. Then

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx}.$$

Chain Rules for First-Order Partial Derivatives

For a two-dimensional version, suppose z is a function of u and v, denoted

$$z = z(u, v)$$

and u and v are functions of x and y,

$$u = u(x, y) \text{ and } v = v(x, y)$$

then

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}$$
$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}$$

$$\begin{tikzcd}
 & & & z \arrow[dll, "\frac{\partial z}{\partial u}"'] \arrow[drr, "\frac{\partial z}{\partial v}"] & & & \\
 & u \arrow[dl, "\frac{\partial u}{\partial x}"'] \arrow[dr, "\frac{\partial u}{\partial y}"] & & & & v \arrow[dl, "\frac{\partial v}{\partial x}"'] \arrow[dr, "\frac{\partial v}{\partial y}"] & \\
x & & y & & x & & y
\end{tikzcd}$$

Example. 1. Find the first partial derivatives using chain rule.

$$\begin{aligned} z &= z(u, v) \\ u &= xy \\ v &= 2x + 3y \end{aligned}$$

SOLUTION:

$$\begin{aligned}
\frac{\partial z}{\partial x} &= \frac{\partial z}{u}\frac{\partial u}{\partial x} + \frac{\partial z}{v}\frac{\partial v}{\partial x} \\
&= y\frac{\partial z}{u} + 2\frac{\partial z}{v} \\
\frac{\partial z}{\partial y} &= \frac{\partial z}{u}\frac{\partial u}{\partial y} + \frac{\partial z}{v}\frac{\partial v}{\partial y} \\
&= x\frac{\partial z}{u} + 3\frac{\partial z}{v}
\end{aligned}$$

Chain Rule for Second Order Partial Derivatives

To find second order partials, we can use the same techniques as first order partials, but with more care and patience!

EXAMPLE. Let

$$\begin{aligned}
z &= z(u, v) \\
u &= x^2y \\
v &= 3x + 2y
\end{aligned}$$

1. Find $\frac{\partial^2 z}{\partial y^2}$.

   SOLUTION: We will first find $\frac{\partial^2 z}{\partial y^2}$.

   $$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = x^2\frac{\partial z}{\partial u} + 2\frac{\partial z}{\partial v}.$$

   Now differentiate again with respect to y to obtain

   $$\begin{aligned}
   \frac{\partial^2 z}{\partial y^2} &= \frac{\partial}{\partial y}(x^2\frac{\partial z}{\partial u}) + \frac{\partial}{\partial y}(2\frac{\partial z}{\partial v}) \\
   &= x^2\frac{\partial}{\partial y}(\frac{\partial z}{\partial u}) + 2\frac{\partial}{\partial y}(\frac{\partial z}{\partial v})
   \end{aligned}$$

   Note that z is a function of u and v, and u and v are functions of x and y. Then the partial derivatives $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$ are also initially functions of u and v and eventually functions of x and y. In other words we have the same dependence diagram as z.

NOTE. $\frac{\partial}{\partial y}(\frac{\partial z}{\partial u}) \neq \frac{\partial^2 z}{\partial y \partial u})$. Never write $\frac{\partial^2 z}{\partial y \partial u}$ as it is mathematically meaningless.

$$\begin{tikzcd}
 & & & \frac{\partial z}{\partial u} \arrow[dll, "\frac{\partial^2 z}{\partial u^2}"'] \arrow[drr, "\frac{\partial^2 z}{\partial v \partial u}"] & & & \\
 & u \arrow[dl, "\frac{\partial u}{\partial x}"'] \arrow[dr, "\frac{\partial u}{\partial y}"] & & & & v \arrow[dl, "\frac{\partial v}{\partial x}"'] \arrow[dr, "\frac{\partial v}{\partial y}"] & \\
x & & y & & x & & y
\end{tikzcd}$$

$$\begin{tikzcd}
 & & & \frac{\partial z}{\partial v} \arrow[dll, "\frac{\partial^2 z}{\partial u \partial v}"'] \arrow[drr, "\frac{\partial^2 z}{\partial v^2}"] & & & \\
 & u \arrow[dl, "\frac{\partial u}{\partial x}"'] \arrow[dr, "\frac{\partial u}{\partial y}"] & & & & v \arrow[dl, "\frac{\partial v}{\partial x}"'] \arrow[dr, "\frac{\partial v}{\partial y}"] & \\
x & & y & & x & & y
\end{tikzcd}$$

Therefore

$$\begin{aligned}
\frac{\partial^2 z}{\partial y^2} &= x^2(\frac{\partial^2 z}{\partial u^2}\frac{\partial u}{\partial y} + \frac{\partial^2 z}{\partial v \partial u}\frac{\partial v}{\partial y}) + 2(\frac{\partial^2 z}{\partial u \partial v}\frac{\partial u}{\partial y} + \frac{\partial^2 z}{\partial v^2}\frac{\partial v}{\partial y}) \\
&= x^2(x^2\frac{\partial^2 z}{\partial u^2} + 2\frac{\partial^2 z}{\partial v \partial u}) + 2(x^2\frac{\partial^2 z}{\partial u \partial v} + 2\frac{\partial^2 z}{\partial v^2})
\end{aligned}$$

The mixed partials are equal so the answer simplifies to

$$\frac{\partial^2 z}{\partial y^2} = x^4\frac{\partial^2 z}{\partial u^2} + 4x^2\frac{\partial^2 z}{\partial u \partial v} + 4\frac{\partial^2 z}{\partial v^2}.$$

2. Find $\frac{\partial^2 z}{\partial x \partial y}$

SOLUTION: We have

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y} = x^2\frac{\partial z}{\partial u} + 2\frac{\partial z}{\partial v}.$$

Now differentiate again with respect to x to obtain

$$\begin{aligned}
\frac{\partial^2 z}{\partial x \partial y} &= \frac{\partial}{\partial x}(x^2 \frac{\partial z}{\partial u}) + \frac{\partial}{\partial x}(2\frac{\partial z}{\partial v}) \\
&= x^2 \frac{\partial}{\partial x}(\frac{\partial z}{\partial u}) + 2x\frac{\partial z}{\partial u} + 2\frac{\partial}{\partial x}(\frac{\partial z}{\partial v}) \\
&= x^2(\frac{\partial^2 z}{\partial u^2}\frac{\partial u}{\partial x} + \frac{\partial^2 z}{\partial v \partial u}\frac{\partial v}{\partial x}) + 2x\frac{\partial z}{\partial u} + 2(\frac{\partial^2 z}{\partial u \partial v}\frac{\partial u}{\partial x} + \frac{\partial^2 z}{\partial v^2}\frac{\partial v}{\partial x}) \\
&= x^2(2xy\frac{\partial^2 z}{\partial u^2} + 3\frac{\partial^2 z}{\partial v \partial u}) + 2x\frac{\partial z}{\partial u} + 2(2xy\frac{\partial^2 z}{\partial u \partial v} + 3\frac{\partial^2 z}{\partial v^2}) \\
&= 2x^3 y\frac{\partial^2 z}{\partial u^2} + (3x^2 + 4xy)\frac{\partial^2 z}{\partial v \partial u}) + 2x\frac{\partial z}{\partial u} + 6\frac{\partial^2 z}{\partial v^2}
\end{aligned}$$

3.4 Maxima and Minima

Recall from 1-dimensional calculus, to find the points of maxima and minima of a function, we first find the critical points i.e where the tangent line is horizontal $f'(x) = 0$. Then

(i) If $f''(x) > 0$ the gradient is increasing and we have a local minimum.

(ii) If $f''(x) < 0$ the gradient is decreasing and we have a local maximum.

(iii) If $f''(x) = 0$ it is inconclusive.

Critical Points of a function of 2 variables

We are now interested in finding points of local maxima and minima for a function of two variables.

Definition. *A function $f(x, y)$ has a relative minimum (resp. maximum) at the point (a, b) if $f(x, y) \geq f(a, b)$ (resp. $f(x, y) \leq f(a, b)$) for all points (x, y) in some region around (a, b)*

Definition. *A point (a, b) is a critical point of a function $f(x, y)$ if one of the following is true*

(i) $f_x(a, b) = 0$ and $f_y(a, b) = 0$

(ii) $f_x(a, b)$ and/or $f_y(a, b)$ does not exist.

Classification of Critical Points

We will need two quantities to classify the critical points of $f(x, y)$:

1. f_{xx}, the second partial derivative of f with respect to x.

2. $H = f_{xx}f_{yy} - f_{xy}^2$ the Hessian

If the Hessian is zero, then the critical point is degenerate. If the Hessian is non-zero, then the critical point is non-degenerate and we can classify the points in the following manner:

case(i) If $H > 0$ and $f_{xx} < 0$ then the critical point is a relative maximum.

case(ii) If $H > 0$ and $f_{xx} > 0$ then the critical point is a relative minimum.

case(iii) If $H < 0$ then the critical point is a saddle point.

EXAMPLE. Find and classify the critical points of

$$12x^3 + y^3 + 12x^2y - 75y.$$

SOLUTION: We first find the critical points of the function.

$$\begin{aligned} f_x &= 36x^2 + 24xy = 12x(3x + 2y) \\ f_y &= 3y^2 + 12x^2 - 75 = 3(4x^2 + y^2 - 25). \end{aligned}$$

The critical points are the points where $f_x = 0$ and $f_y = 0$

$$\begin{aligned} f_x &= 0 \\ 12x(3x + 2y) &= 0 \end{aligned}$$

Therefore either $x = 0$ or $3x + 2y = 0$. We handle the two cases separately;

case(i) $x = 0$.
Then substituting this in f_y we get $f_y = 3(y^2 - 25) = 0$ implies $y = \pm 5$.

case(ii) $3x + 2y = 0$.

Then $y = -3x/2$ and substituting this in f_y we get,

$$\begin{aligned}
f_y &= 3(4x^2 + \frac{9x^2}{4} - 25) \\
&= \frac{3}{4}(16x^2 + 9x^2 - 100) \\
&= \frac{3}{4}(25x^2 - 100) \\
&= \frac{75}{4}(x^2 - 4) \\
f_y &= 0 \\
\frac{75}{4}(x^2 - 4) &= 0 \\
x^2 - 4 &= 0 \\
x &= \pm 2
\end{aligned}$$

Thus we have found four critical points: $(0, 5)$, $(0, -5)$, $(2, -3)$, $(-2, 3)$.

We must now classify these points.

$$\begin{aligned}
f_{xx} &= 72x + 24y = 24(3x + y) \\
f_{xy} &= 24x \\
f_{yy} &= 6y \\
H &= f_{xx}f_{yy} - f_{xy}^2 \\
&= (24)(3x + y)(6y) - (24x)^2
\end{aligned}$$

Points	f_{xx}	H	Type
$(0, 5)$	120	3600	Minimum
$(0, -5)$	-120	3600	Maximum
$(2, -3)$	72	-3600	Saddle
$(-2, 3)$	-72	-3600	Saddle

References

1. Engineering Mathematics, by K. A. Stroud.

2. A complete set of notes on Pre-Calculus, Single Variable Calculus, Multivariable Calculus and Linear Algebra.
   Here is a link to the chapter on Higher Order Partial Differentiation.
   http://tutorial.math.lamar.edu/Classes/CalcIII/HighOrderPartialDerivs.aspx.

   Here is a link to the chapter on Chain Rules.
   http://tutorial.math.lamar.edu/Classes/CalcIII/ChainRule.aspx.

   Here is a link to the chapter on Maxima and Minima.
   http://tutorial.math.lamar.edu/Classes/CalcIII/RelativeExtrema.aspx

3. A collection of examples, animations and notes on Multivariable Calculus.
   http://people.usd.edu/ jflores/MultiCalc02/WebBook/Chapter_15/

4. Links to various resources on Calculus.
   http://www.calculus.org/.

4 A little Vector Calculus

4.1 Gradient

Vector Function/ Vector Fields

The functions of several variables we have so far studied would take a point (x, y, z) and give a real number $f(x, y, z)$. We call these types of functions *scalar-valued* functions i.e. for example

$$f(x, y, z) = x^2 + 2xyz.$$

We are now going to talk about *vector-valued* functions, where we take a point (x, y, z) and the value of $f(x, y, z)$ is a vector. i.e. for example

$$f(x, y, z) = (y, x, z^2) = y\vec{i} + x\vec{j} + z^2\vec{k}.$$

Definition. *A vector function is a function that takes one or more variables and returns a vector.*

Example. 1. A vector function of a single variable:

$$r(t) = (2 + t, 3 + 2t, 1 - 3t).$$

Let us look at a few values.

$$\begin{aligned} f(2) &= (4, 7, -5), & f(-2) &= (0, -1, 7), \\ f(1) &= (3, 5, -2), & f(-1) &= (1, 1, 4), \\ f(0) &= (2, 3, 1) \end{aligned}$$

2. A vector function of three variables:

$$f(x, y, z) = (y, x, z^2).$$

Let us look at a few values.

$$\begin{aligned} f(0, 0, 0) &= (0, 0, 0), & f(1, 0, 0) &= (0, 1, 0), \\ f(0, 1, 0) &= (1, 0, 0), & f(0, 0, 1) &= (0, 0, 1), \\ f(1, 2, 1) &= (2, 1, 1), & f(2, 1, 2) &= (1, 2, 4) \end{aligned}$$

DEFINITION. *A vector field is an assignment of a vector to every point in space.*

EXAMPLE. 1. In a magnetic field, we can assign a vector which describes the force and direction to every point in space.

2. Normal to a surface. For a surface, at every point on the surface, we can get a tangent plane and hence a normal vector to every point on the surface.

EXAMPLE. Let $f(x, y, z) = x^2 + 2xyz$ be a scalar-valued function. We then define a vector-valued function by taking its partial derivatives.

$$\begin{aligned} \nabla f &= (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \\ &= (2x + 2yz, 2xz, 2xy) \end{aligned}$$

This kind of vector function has a special name, *the gradient.*

DEFINITION. *Suppose that $f(x, y)$ is a scalar-valued function of two variables. Then the gradient of f is the vector function defined as,*

$$\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j}.$$

Similarly if $f(x, y, z)$ is a scalar-valued function of three variables. Then the gradient of f is the vector function defined as,

$$\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j} + \frac{\partial f}{\partial z}\vec{k}.$$

REMARK. The gradient of a scalar function is a vector field which points in the direction of the greatest rate of increase of the scalar function, and whose magnitude is the greatest rate of change.

EXAMPLE. Consider the scalar-valued function defined by $f(x, y) = x^2 + y^2$. Find the gradient of f at the point $x = 2, y = 5$.

SOLUTION: Then gradient of f, is a vector function given by,

$$\begin{aligned} \nabla f &= (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \\ &= (2x, 2y) \\ \nabla f(2, 5) &= (4, 10) \end{aligned}$$

4.2 Directional Derivative

For a function of 2 variables $f(x, y)$, we have seen that the function can be used to represent the surface

$$z = f(x, y)$$

and recall the geometric interpretation of the partials:

(i) $f_x(a, b)$-represents the rate of change of the function $f(x, y)$ as we vary x and hold $y = b$ fixed.

(ii) $f_y(a, b)$-represents the rate of change of the function $f(x, y)$ as we vary y and hold $x = a$ fixed.

We now ask, at a point P can we calculate the slope of f in an arbitrary direction?

Recall the definition of the vector function ∇f,

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right).$$

We observe that,

$$\begin{aligned} \nabla f \cdot \hat{i} &= f_x \\ \nabla f \cdot \hat{j} &= f_y \end{aligned}$$

This enables us to calculate the directional derivative in an arbitrary direction, by taking the dot product of ∇f with a unit vector, $\vec{u}$, in the desired direction.

DEFINITION. *The directional derivative of the function f in the direction $\vec{u}$ denoted by $D_{\vec{u}}f$, is defined to be,*

$$D_{\vec{u}}f = \frac{\nabla f \cdot \vec{u}}{|\vec{u}|}$$

EXAMPLE. What is the directional derivative of $f(x, y) = x^2 + xy$, in the direction $\vec{i} + 2\vec{j}$ at the point $(1, 1)$?

SOLUTION: We first find ∇f.

$$\begin{aligned}
\nabla f &= (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \\
&= (2x + y, x) \\
\nabla f(1,1) &= (3,1)
\end{aligned}$$

Let $u = \vec{i} + 2\vec{j}$.

$$|\vec{u}| = \sqrt{1^2 + 2^2} = \sqrt{1+4} = \sqrt{5}.$$

$$\begin{aligned}
D_{\vec{u}} f(1,1) &= \frac{\nabla f \cdot \vec{u}}{|\vec{u}|} \\
&= \frac{(3,1).(1,2)}{\sqrt{5}} \\
&= \frac{(3)(1) + (1)(2)}{\sqrt{5}} \\
&= \frac{5}{\sqrt{5}} \\
&= \sqrt{5}
\end{aligned}$$

Properties of the Gradient deduced from the formula of Directional Derivatives

$$\begin{aligned}
D_{\vec{u}} f &= \frac{\nabla f \cdot \vec{u}}{|\vec{u}|} \\
&= \frac{|\nabla f||\vec{u}| \cos\theta}{|\vec{u}|} \\
&= |\nabla f| \cos\theta
\end{aligned}$$

1. If $\theta = 0$, i.e. i.e. $\vec{u}$ points in the same direction as ∇f, then $D_{\vec{u}} f$ is maximum. Therefore we may conclude that

   (i) ∇f **points in the steepest direction**.

   (ii) The **magnitude of ∇f gives the slope in the steepest direction**.

2. At any point P, $\nabla f(P)$ **is perpendicular to the level set** through that point.

EXAMPLE. 1. Let $f(x, y) = x^2 + y^2$ and let $P = (1, 2, 5)$. Then P lies on the graph of f since $f(1, 2) = 5$. Find the slope and the direction of the steepest ascent at P on the graph of f

SOLUTION:

- We use the first property of the Gradient vector. The direction of the steepest ascent at P on the graph of f is the direction of the gradient vector at the point $(1, 2)$.

$$\begin{aligned} \nabla f &= (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \\ &= (2x, 2y) \\ \nabla f(1,2) &= (2, 4). \end{aligned}$$

- The slope of the steepest ascent at P on the graph of f is the magnitude of the gradient vector at the point $(1, 2)$.

$$|\nabla f(1, 2)| = \sqrt{2^2 + 4^2} = \sqrt{20}.$$

2. Find a normal vector to the graph of the equation $f(x, y) = x^2 + y^2$ at the point $(1, 2, 5)$. Hence write an equation for the tangent plane at the point $(1, 2, 5)$.

SOLUTION: We use the second property of the gradient vector. For a function g, $\nabla g(P)$ **is perpendicular to the level set**. So we want our surface $z = x^2 + y^2$ to be the level set of a function.
Therefore we define a new function,

$$g(x, y, z) = x^2 + y^2 - z.$$

Then our surface is the level set

$$\begin{aligned} g(x, y, z) &= 0 \\ x^2 + y^2 - z &= 0 \\ z &= x^2 + y^2 \end{aligned}$$

$$\begin{aligned}\nabla g &= (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}) \\ &= (2x, 2y, -1) \\ \nabla g(1,2,5) &= (2,4,-1)\end{aligned}$$

By the above property, $\nabla g(P)$ is perpendicular to the level set $g(x,y,z) = 0$. Therefore $\nabla g(P)$ is the required normal vector.
Finally an equation for the tangent plane at the point $(1,2,5)$ on the surface is given by

$$2(x-1) + 4(y-2) - 1(z-5) = 0.$$

4.3 Curl and Divergence

We denoted the gradient of a scalar function $f(x,y,z)$ as

$$\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})$$

Let us separate or isolate the operator $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$. We can then define various physical quantities such as div, curl by specifying the action of the operator ∇.

Divergence

DEFINITION. *Given a vector field $\vec{v}(x,y,z) = (v_1(x,y,z), v_2(x,y,z), v_3(x,y,z))$, the divergence of $\vec{v}$ is a scalar function defined as the dot product of the vector operator ∇ and $\vec{v}$,*

$$\begin{aligned}\text{Div}\,\vec{v} &= \nabla \cdot \vec{v} \\ &= (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}) \cdot (v_1, v_2, v_3) \\ &= \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}\end{aligned}$$

EXAMPLE. Compute the divergence of $(x-y)\vec{i} + (x+y)\vec{j} + z\vec{k}$.

SOLUTION:

$$\begin{aligned}
\vec{v} &= ((x-y),(x+y),z) \\
\nabla &= (\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}) \\
\operatorname{Div}\vec{v} &= \nabla\cdot\vec{v} \\
&= (\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z})\cdot((x-y),(x+y),z) \\
&= \frac{\partial(x-y)}{\partial x}+\frac{\partial(x+y)}{\partial y}+\frac{\partial z}{\partial z} \\
&= 1+1+1 \\
&= 3
\end{aligned}$$

Curl

DEFINITION. *The curl of a vector field is a vector function defined as the cross product of the vector operator ∇ and $\vec{v}$,*

$$\begin{aligned}
\operatorname{Curl}\vec{v}=\nabla\times\vec{v} &= \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix} \\
&= (\frac{\partial v_3}{\partial y}-\frac{\partial v_2}{\partial z})i-(\frac{\partial v_3}{\partial x}-\frac{\partial v_1}{\partial z})j+(\frac{\partial v_2}{\partial x}-\frac{\partial v_1}{\partial y})k
\end{aligned}$$

EXAMPLE. Compute the curl of the vector function $(x-y)\vec{i}+(x+y)\vec{j}+z\vec{k}$.

SOLUTION:

$$\begin{aligned}
\operatorname{Curl}\vec{v}=\nabla\times\vec{v} &= \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ (x-y) & (x+y) & z \end{vmatrix} \\
&= (\frac{\partial z}{\partial y}-\frac{\partial(x+y)}{\partial z})i-(\frac{\partial z}{\partial x}-\frac{\partial(x-y)}{\partial z})j+(\frac{\partial(x+y)}{\partial x}-\frac{\partial(x-y)}{\partial y})k \\
&= (0-0)\vec{i}-(0-0)\vec{j}+(1-(-1))\vec{k} \\
&= 2\vec{k}
\end{aligned}$$

4.4 Laplacian

We have seen above that given a vector function, we can calculate the divergence and curl of that function. A scalar function f has a vector function ∇f associated to it. We now look at $\mathrm{Curl}(\nabla f)$ and $\mathrm{Div}(\nabla f)$.

$$\begin{aligned}
\mathrm{Curl}(\nabla f) &= \nabla \times \nabla f \\
&= (\frac{\partial f_z}{\partial y} - \frac{\partial f_y}{\partial z})i + (\frac{\partial f_x}{\partial z} - \frac{\partial f_z}{\partial x})j + (\frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y})k \\
&= (f_{yz} - f_{zy})i + (f_{zx} - f_{xz})j + (f_{xy} - f_{yx})k \\
&= 0 \\
\mathrm{Div}(\nabla f) &= \nabla \cdot \nabla f \\
&= (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}) \cdot (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \\
&= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}
\end{aligned}$$

DEFINITION. *The Laplacian of a scalar function $f(x, y)$ of two variables is defined to be* $\mathrm{Div}(\nabla f)$ *and is denoted by* $\nabla^2 f$*,*

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

The Laplacian of a scalar function $f(x, y, z)$ of three variables is defined to be $\mathrm{Div}(\nabla f)$ *and is denoted by* $\nabla^2 f$*,*

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

EXAMPLE. Compute the Laplacian of $f(x, y, z) = x^2 + y^2 + z^2$.

SOLUTION:

$$\begin{aligned}
\nabla^2 f &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \\
&= \frac{\partial 2x}{\partial x} + \frac{\partial 2y}{\partial y} + \frac{\partial 2z}{\partial z} \\
&= 2 + 2 + 2 \\
&= 6.
\end{aligned}$$

We have the following identities for the Laplacian in different coordinate systems:

$$
\begin{aligned}
\textit{Rectangular} &: \quad \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \\
\textit{Polar} &: \quad \nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\big(r\frac{\partial f}{\partial r}\big) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} \\
\textit{Cylindrical} &: \quad \nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\big(r\frac{\partial f}{\partial r}\big) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2} \\
\textit{Spherical} &: \quad \nabla^2 f = \frac{1}{\rho^2}\frac{\partial}{\partial \rho}\big(\rho^2\frac{\partial f}{\partial \rho}\big) + \frac{1}{\rho^2 \sin\theta}\frac{\partial}{\partial \theta}\big(\sin\theta\frac{\partial f}{\partial \theta}\big) + \frac{1}{\rho^2 \sin^2\theta}\frac{\partial^2 f}{\partial \phi^2}
\end{aligned}
$$

Example. Consider the same function $f(x, y, z) = x^2 + y^2 + z^2$. We have seen that in rectangular coordinates we get

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 6.$$

We now calculate this in cylindrical and spherical coordinate systems, using the formulas given above.

1. Cylindrical Coordinates.
   We have $x = r\cos\theta$ and $y = r\sin\theta$ so

   $$f(r, \theta, z) = r^2\cos^2\theta + r^2\sin^2\theta + z^2 = r^2 + z^2.$$

   Using the above formula:

   $$
   \begin{aligned}
   \nabla^2 f &= \frac{1}{r}\frac{\partial}{\partial r}\big(r\frac{\partial f}{\partial r}\big) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2} \\
   &= \frac{1}{r}\frac{\partial}{\partial r}(r2r) + 0 + \frac{\partial(2z)}{\partial z} \\
   &= \frac{1}{r}(4r) + 2 \\
   &= 4 + 2 \\
   &= 6
   \end{aligned}
   $$

2. Spherical Coordinates.
   We have $x = \rho\sin\phi\cos\theta$, $y = \rho\sin\phi\sin\theta$, $z = \rho\cos\phi$ and $\rho = \sqrt{x^2 + y^2 + z^2}$, so

   $$f(r, \theta, z) = \rho^2.$$

Using the above formula:

$$\begin{aligned}
\nabla^2 f &= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}\big(\rho^2\frac{\partial f}{\partial\rho}\big) + \frac{1}{\rho^2\sin\theta}\frac{\partial}{\partial\theta}\big(\sin\theta\frac{\partial f}{\partial\theta}\big) + \frac{1}{\rho^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2} \\
&= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}(\rho^2 2\rho) + 0 + 0 \\
&= \frac{1}{\rho^2}\frac{\partial}{\partial\rho}(2\rho^3) \\
&= \frac{1}{\rho^2}(6\rho^2) \\
&= 6.
\end{aligned}$$

These three different calculations all produce the same result because ∇^2 is a derivative with a real physical meaning, and does not depend on the coordinate system being used.

References

1. A briliant animated example, showing that the maximum slope at a point occurs in the direction of the gradient vector. The animation shows:

   - a surface
   - a unit vector rotating about the point (1, 1, 0), (shown as a rotating black arrow at the base of the figure)
   - a rotating plane parallel to the unit vector, (shown as a grey grid)
   - the traces of the planes in the surface, (shown as a black curve on the surface)
   - the tangent lines to the traces at (1, 1, f (1, 1)), (shown as a blue line)
   - the gradient vector (shown in green at the base of the figure)

   http://archives.math.utk.edu/ICTCM/VOL10/C009/dd.gif

2. A complete set of notes on Pre-Calculus, Single Variable Calculus, Multivariable Calculus and Linear Algebra.
   Here is a link to the chapter on Directional Derivatives.
   http://tutorial.math.lamar.edu/Classes/CalcIII/DirectionalDeriv.aspx.

   Here is a link to the chapter on Curl and Divergence.
   http://tutorial.math.lamar.edu/Classes/CalcIII/CurlDivergence.aspx

5 Linear Equations

5.1 Introduction

A linear equation is one in which all the unknown variables occur with a power of one. i.e. $2x + y = 5$ is a linear equation, but $2x^2 + y = 5$ is not a linear equation, since x has power 2. Systems of linear equations are common in all branches of science. In school you have learned how to simultaeously solve a system of two linear equations.

EXAMPLE. Solve the system of two linear equations:

$$\begin{aligned} 3x_1 + 2x_2 &= 7 \\ -x_1 + x_2 &= 6 \end{aligned}$$

Then $x_1 = -1$ and $x_2 = 5$ is a solution. Geometrically the solution, $(-1, 5)$ is the point of intersection of the two lines given by the two equations.

Here is an example of larger system of equations from Chemistry.

EXAMPLE. Under certain controled conditions, mix toulene C_7H_8 and nitric acid HNO_3 to produce trinitrotoluene (TNT) $C_7H_5O_6N_3$ along with water. In what proportion should those components be mixed, i.e. solve for x, y, z, w?

SOLUTION: The number of atoms of each element before the reaction must equal the number present afterward!

$$xC_7H_8 + yHNO_3 \rightarrow zC_7H_5O_6N_3 + wH_2O$$

So we get the system:

$$\begin{aligned} 7x &= 7z \\ 8x + y &= 5z + 2w \\ y &= 3z \\ 3y &= 6z + w \end{aligned}$$

To find the answer we need to solve the above system of linear equations. In this chapter we will learn an easy method to solve such a system.

5.2 Row Echelon Form

In this section we learn Gauss' Method to solve a system of linear equations. We will use an example to understand the method.

EXAMPLE. Find the solution set of the linear system

$$\begin{aligned} x - \ \ y \qquad\qquad\quad &= 0 \\ 2x - 2y + \ \ z + 2w &= 4 \\ y \qquad + \ \ w &= 0 \\ 2z + \ \ w &= 5 \end{aligned}$$

We can abbreviate this linear system with the linear array of numbers, whose entries are the coefficients of the equations:

$$\left(\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 2 & -2 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right).$$

The vertical bar just reminds us that the coefficients of the system are on the left hand side of the bar and the constants are on the right. We call this an *augmented matrix*. We can now proceed with Gauss' method.

$$\left(\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 2 & -2 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right) \overset{R_2=-2R_1+R2}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right)$$

$$\overset{R_2\leftrightarrow R_3}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right)$$

$$\overset{R_4=-2R_3+R4}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & -3 & -3 \end{array}\right)$$

The resulting equations are:

$$\begin{aligned} x - y \quad\quad\quad &= 0 \\ y \;+\; w &= 0 \\ z + 2w &= 4 \\ w &= 1 \end{aligned}$$

Back substitution now gives $w = 1, z = 2, y = -1, x = -1$.

REMARK. We transformed the original set of equations into a new simpler set of equations, using row operations to put the array into *Row Echelon Form*. You probably have two questions;

1. Why do both the original system and the new simpler system has the same solution set?
   Answer: There is a theorem which says that, if a linear system is changed to another by one of these operations

   - an equation is swapped with another
   - an equation has both sides multiplied by a nonzero constant
   - an equation is replaced by the sum of itself and a multiple of another

   then the two systems have the same set of solutions.

2. What is Row Echelon Form?
   Answer: A matrix is in Row Echelon Form if it has the following form:

   - any all zero rows are at the bottom of the reduced matrix
   - in a non-zero row, the first-from-left nonzero value is 1
   - the first 1 in each row is to the right of the first 1 in the row above

Algorithm to put a matrix into Row Echelon Form

(i) Make the first entry of the first row non-zero by doing a swap if necessary.

(ii) Make the first entry of the first row 1 by multiplying by the reciprocal. The first row is now in the correct form.

(iii) Make the first entry in all the rows below the first row 0, by using the 1 in the first row above to subtract. The first column should be $\begin{pmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \end{pmatrix}$

(iv) Now we can disregard the first row and first column because they are in the correct form, and only consider the remaining sub-matrix. Repeat steps (i)-(iv) for the smaller sub-matrix. Continue till you run out of rows.

5.3 Soving Systems of Linear Equations

The solution set of a system of linear equations can be;

(i) Empty

(ii) A one point set, i.e. a unique solution

(iii) Infinite

An example where the solution set is empty.

Example. Consider the system

$$\begin{aligned} x + 2y &= 4 \\ x + 2y &= 8 \end{aligned}$$

The augmented matrix:

$$\left(\begin{array}{cc|c} 1 & 2 & 4 \\ 1 & 2 & 8 \end{array}\right) \overset{R_2 = -R_1 + R2}{\longrightarrow} \left(\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 0 & 4 \end{array}\right)$$

We get the absurd relation $0 = 4$. This implies that the system has no solutions. Geometrically the two equations are two parallel lines and hence there is no point of inersection.

An example where the solution set is infinite.

Example. Consider the linear system of equations

$$\begin{aligned} x_1 + 2x_2 \phantom{{}- x_3} &= 4 \\ x_2 - x_3 &= 0 \\ x_1 + 2x_3 &= 4. \end{aligned}$$

The associated augmented matrix:

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 4 \end{array}\right)$$

We now reduce the matrix to row echelon form:

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 4 \end{array}\right) \overset{R_3=-R_1+R3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & -2 & 2 & 0 \end{array}\right)$$

$$\overset{R_3=2R_2+R_3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

We have a bottom row of zeros. The second row gives $x_2 = x_3$ and the first row gives $x_1 + 2x_2 = 4$. We can express x_1 in terms of x_3 to get $x_1 = 4 - 2x_3$. Thus we can write the solution set for the system in the following manner:

$$\begin{aligned} S &= \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_2 = x_3 \text{ and } x_1 = 4 - 2x_3 \} \\ &= \{ \begin{pmatrix} 4 & -2x_3 \\ & x_3 \\ & x_3 \end{pmatrix} \} \\ &= \{ \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} x_3 | x_3 \in \mathbb{R} \} \end{aligned}$$

Example. Consider the linear system

$$\begin{aligned} x + y + \ z - \ w &= 1 \\ y - \ z + \ w &= -1 \\ 3x \qquad + 6z - 6w &= 6 \\ - y + \ z - \ w &= 1 \end{aligned}$$

The associated augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 3 & 0 & 6 & -6 & 6 \\ 0 & -1 & 1 & -1 & 1 \end{array}\right)$$

Gauss' Method:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 3 & 0 & 6 & -6 & 6 \\ 0 & -1 & 1 & -1 & 1 \end{array}\right) \xrightarrow{R_3=-3R_1+R3} \left(\begin{array}{cccc|c} 1 & 1 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & -3 & 3 & -3 & 3 \\ 0 & -1 & 1 & -1 & 1 \end{array}\right)$$

$$\xrightarrow{R_3=3R_2+R_3, R_4=R_2+R_4} \left(\begin{array}{cccc|c} 1 & 1 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

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

From the second equation we get, $y = -1 + z - w$ and substituting this in equation one we get, $x + (-1 + z - w) + z - w = 1$. Solving for x we get $x = 2 - 2z + 2w$. The soution set:

$$\begin{aligned} S &= \{\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} | y = -1 + z - w \text{ and } x = 2 - 2z + 2w\} \\ &= \{\begin{pmatrix} 2 & -2z & +2w \\ -1 & +z & -w \\ & z & \\ & & w \end{pmatrix}\} \\ &= \{\begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} z + \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix} w | z, w \in \mathbb{R}\} \end{aligned}$$

6 Linear Independence and Gram Schmidt

6.1 Linear Combination of Vectors

A linear combination of vectors in $\mathbb{R}^2$ is a sum of the form

$$\alpha \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \beta \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

e.g.

$$2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + (-4) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

A linear combination of vectors in $\mathbb{R}^3$ is a sum of the form

$$\alpha \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \beta \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

e.g.

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

EXAMPLE. Write $\begin{pmatrix} 9 \\ 8 \\ 3 \end{pmatrix}$ as a linear combination of the vectors $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$.

SOLUTION: We need to find scalars like x, y, z such that

$$\begin{aligned} x \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + y \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix} &= \begin{pmatrix} 9 \\ 8 \\ 3 \end{pmatrix} \\ \begin{pmatrix} x & +2y & +3z \\ 2x & -y & +z \\ 3x & & -z \end{pmatrix} &= \begin{pmatrix} 9 \\ 8 \\ 3 \end{pmatrix} \end{aligned}$$

We get a system of three linear equations:

$$\begin{aligned} x + 2y + 3z &= 9 \\ 2x - y + z &= 8 \\ 3x - z &= 3 \end{aligned}$$

We can use the methods of the last chapter to find the solution set of this system of linear equations. The associated augmented matrix:

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 3 & 0 & -1 & 3 \end{array}\right)$$

We now reduce the matrix to row echelon form:

$$\begin{aligned}
\left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 3 & 0 & -1 & 3 \end{array}\right) &\overset{R_2=-2R_1+R2, R_3=-3R_1+R_3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & -5 & -5 & -10 \\ 0 & -6 & -10 & -24 \end{array}\right) \\
&\overset{R_2=\frac{-1}{5}R_2}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 1 & 2 \\ 0 & -6 & -10 & -24 \end{array}\right) \\
&\overset{R_3=6R_2+R_3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -4 & -12 \end{array}\right) \\
&\overset{R_3=\frac{-1}{4}R_3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)
\end{aligned}$$

Therefore we get the relations:

$$\begin{aligned}
z &= 3 \\
y &= 2 - z = 2 - 3 = -1 \\
x &= 9 - 2y - 3z = 9 - 2(-1) - 3(3) = 2
\end{aligned}$$

Therefore we write,

$$(2)\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + (-1)\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + (3)\begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \\ 3 \end{pmatrix}$$

6.2 Linear Independence of Vectors

DEFINITION. *A set of vectors S is linearly independent if there is no non-trivial linear combination of vectors that sums to zero.*

REMARK. If a set is not linearly independent, it is linearly dependent..

EXAMPLE. 1. Consider the linear combination which sums to zero,

$$\begin{aligned}
\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \beta \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} \alpha \\ \beta \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\alpha &= 0 \\
\beta &= 0
\end{aligned}$$

is a unique solution. Therefore the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are linearly independent.

2. Consider the linear combination which sums to zero

$$\begin{aligned}
x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} + z \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y \end{pmatrix} + + \begin{pmatrix} z \\ z \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} x+z \\ y+z \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
x + z &= 0; \quad x = -z \\
y + z &= 0; \quad y = -z
\end{aligned}$$

So for any value of z, say $z = 1$, we can choose $x = y = -z = -1$ and get

$$(-1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (-1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

a non-trivial linear combination which sums to zero.
So the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are **not** linearly independent.

3. Determine whether the set of vectors $\{[1,1,0],[1,0,1],[2,1,1]\}$ is linearly independent.

SOLUTION: Take a linear combination of the vectors which sums to zero.

$$x\begin{pmatrix}1\\1\\2\end{pmatrix}+y\begin{pmatrix}1\\0\\1\end{pmatrix}+z\begin{pmatrix}0\\1\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

$$\begin{pmatrix}x & +y & +2z\\x & & +z\\ & y & +z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

We get a system of three linear equations:

$$\begin{aligned} x+y+2z&=0\\ x\qquad+\ z&=0\\ y+\ z&=0 \end{aligned}$$

We can use the methods of the last chapter to find the solution set of this system of linear equations. If $x=y=z=0$ is a unique solution, then the vectors are linearly independent, other wise we can find a dependency. The associated augmented matrix:

$$\left(\begin{array}{ccc|c}1&1&2&0\\1&0&1&0\\0&1&1&0\end{array}\right)$$

We now reduce the matrix to row echelon form:

$$\left(\begin{array}{ccc|c}1&1&2&0\\1&0&1&0\\0&1&1&0\end{array}\right)\xrightarrow{R_2=-R_1+R2}\left(\begin{array}{ccc|c}1&1&2&0\\0&-1&-1&0\\0&1&1&0\end{array}\right)$$

$$\xrightarrow{R_3=R_2+R_3}\left(\begin{array}{ccc|c}1&1&2&0\\0&1&1&0\\0&0&0&0\end{array}\right)$$

We have a bottom row of zeros and hence cannot have a unique solution. The second row gives $y+z=0$ and the first row gives $x+y+2z=0$. So

$$\begin{aligned} y &= -z\\ x &= -y-2z=-(-z)-2z=-z \end{aligned}$$

If we choose $z = 1$, then $x = -1$ and $y = -1$we get a non-trivial linear combination which sums to zero:

$$(-1)\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + (-1)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + (1)\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Therefore the set of vectors is **not** linearly independent.

THEOREM *Let $S = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be a set of n, n-dimensional vectors, then the set S is linearly independent if and only if the determinant of the matrix having the n vectors as columns is non-zero.*

EXAMPLE. 1. We can apply this result to the example above. Let $\{[1,1,0],[1,0,1],[2,1,1]\}$ be three, 3-dimensional vectors.

$$\begin{vmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 1(0-1) - 1((1-0) + 2(1-0) = -1 - 1 + 2 = 0.$$

Since the determinant is zero, the set of vectors are not linearly independent.

2. Consider the set $\{[1,-1,1],[1,0,1],[1,1,2]\}$.

$$\begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 1(0-1) - 1((-1)(2) - (1)(1)) + 1((-1)(1) - 0) = 1 \neq 0.$$

Since the determinant is non-zero, the set of vectors are linearly independent.

6.3 Orthogonal and orthonormal vectors

DEFINITION. *We say that 2 vectors are orthogonal if they are perpendicular to each other. i.e. the dot product of the two vectors is zero.*

DEFINITION. *We say that a set of vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ are mutually orthogonal if every pair of vectors is orthogonal. i.e.*

$$\vec{v}_i.\vec{v}_j = 0, \text{ for all } i \neq j.$$

EXAMPLE. The set of vectors $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix}$ is mutually orthogonal.

$$\begin{aligned} (1,0,-1).(1,\sqrt{2},1) &= 0 \\ (1,0,-1).(1,-\sqrt{2},1) &= 0 \\ (1,\sqrt{2},1).(1,-\sqrt{2},1) &= 0 \end{aligned}$$

DEFINITION. *A set of vectors S is orthonormal if every vector in S has magnitude 1 and the set of vectors are mutually orthogonal.*

EXAMPLE. We just checked that the vectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \vec{v}_3 = \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix}$$

are mutually orthogonal. The vectors however are not normalized (this term is sometimes used to say that the vectors are not of magnitude 1). Let

$$\begin{aligned} \vec{u}_1 &= \frac{\vec{v}_1}{|\vec{v}_1|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix} \\ \vec{u}_2 &= \frac{\vec{v}_2}{|\vec{v}_2|} = \frac{1}{2} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ \sqrt{2}/2 \\ 1/2 \end{pmatrix} \\ \vec{u}_3 &= \frac{\vec{v}_3}{|\vec{v}_3|} = \frac{1}{2} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -\sqrt{2}/2 \\ 1/2 \end{pmatrix} \end{aligned}$$

The set of vectors $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is orthonormal.

PROPOSITION *An orthogonal set of non-zero vectors is linearly independent.*

6.4 Gram-Schmidt Process

Given a set of linearly independent vectors, it is often useful to convert them into an orthonormal set of vectors. We first define the projection operator.

DEFINITION. *Let $\vec{u}$ and $\vec{v}$ be two vectors. The projection of the vector $\vec{v}$ on $\vec{u}$ is defined as folows:*

$$Proj_{\vec{u}}\vec{v} = \frac{(\vec{v}.\vec{u})}{|\vec{u}|^2}\vec{u}.$$

EXAMPLE. Consider the two vectors $\vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\vec{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
These two vectors are linearly independent.
However they are not orthogonal to each other. We create an orthogonal vector in the following manner:

$$\begin{aligned}
\vec{v}_1 &= \vec{v} - (\mathrm{Proj}_{\vec{u}}\vec{v}) \\
\mathrm{Proj}_{\vec{u}}\vec{v} &= \frac{(1)(1)+(1)(0)}{(\sqrt{1^2+0^2})^2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = (1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\vec{v}_1 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - (1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\end{aligned}$$

$\vec{v}_1$ thus constructed is orthogonal to $\vec{u}$.

The Gram-Schmidt Algorithm:

Let $v_1, v_2, ..., v_n$ be a set of n linearly independent vectors in $\mathcal{R}^n$. Then we can construct an orthonormal set of vectors as follows:

Step 1. Let $\vec{u}_1 = \vec{v}_1$. $\quad \vec{e}_1 = \frac{\vec{u}_1}{|\vec{u}_1|}$.

Step 2. Let $\vec{u}_2 = \vec{v}_2 - \mathrm{Proj}_{\vec{u}_1}\vec{v}_2$. $\quad \vec{e}_2 = \frac{\vec{u}_2}{|\vec{u}_2|}$.

Step 3. Let $\vec{u}_3 = \vec{v}_3 - \mathrm{Proj}_{\vec{u}_1}\vec{v}_3 - \mathrm{Proj}_{\vec{u}_2}\vec{v}_3$. $\quad \vec{e}_3 = \frac{\vec{u}_3}{|\vec{u}_3|}$.

Step 4. Let $\vec{u}_4 = \vec{v}_4 - \mathrm{Proj}_{\vec{u}_1}\vec{v}_4 - \mathrm{Proj}_{\vec{u}_2}\vec{v}_4 - \mathrm{Proj}_{\vec{u}_3}\vec{v}_4$. $\quad \vec{e}_4 = \frac{\vec{u}_4}{|\vec{u}_4|}$.

-
-

EXAMPLE. We will apply the Gram-Schmidt algorithm to orthonormalize the set of vectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

To apply the Gram-Schmidt, we first need to check that the set of vectors are linearly independent.

$$\begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 1(0-1) - 1((-1)(2) - (1)(1)) + 1((-1)(1) - 0) = 1 \neq 0.$$

Therefore the vectors are linearly independent.
Gram-Schmidt algorithm:

Step 1. Let

$$\begin{aligned} \vec{u}_1 &= \vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \\ \vec{e}_1 &= \frac{\vec{u}_1}{|\vec{u}_1|} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}. \end{aligned}$$

Step 2. Let

$$\begin{aligned} \vec{u}_2 &= \vec{v}_2 - \mathrm{Proj}_{\vec{u}_1} \vec{v}_2 \\ \mathrm{Proj}_{\vec{u}_1} \vec{v}_2 &= \frac{(1,0,1).(1,-1,1)}{1^2 + (-1)^2 + 1^2} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \frac{2}{3} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \\ \vec{u}_2 &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \end{pmatrix} \\ \vec{e}_2 &= \frac{\vec{u}_2}{|\vec{u}_2|} = \frac{3}{\sqrt{6}} \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \end{pmatrix}. \end{aligned}$$

Step 3. Let

$$
\begin{aligned}
\vec{u}_3 &= \vec{v}_3 - \mathrm{Proj}_{\vec{u}_1}\vec{v}_3 - \mathrm{Proj}_{\vec{u}_2}\vec{v}_3 \\
\mathrm{Proj}_{\vec{u}_1}\vec{v}_3 &= \frac{(1,1,2).(1,-1,1)}{1^2+(-1)^2+1^1}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \\
\mathrm{Proj}_{\vec{u}_2}\vec{v}_3 &= \frac{(1,1,2).(1/3,2/3,1/3)}{(1/3)^2+(2/3)^2+(1/3)^2}\begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \end{pmatrix} = \frac{5}{2}\begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \end{pmatrix} \\
\vec{u}_3 &= \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} - \frac{5}{2}\begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \end{pmatrix} \\
&= \begin{pmatrix} -1/2 \\ 0 \\ 1/2 \end{pmatrix} \\
\vec{e}_3 &= \frac{\vec{u}_3}{|\vec{u}_3|} = \sqrt{2}\begin{pmatrix} -1/2 \\ 0 \\ 1/2 \end{pmatrix}.
\end{aligned}
$$

EXAMPLE. Consider the vectors $\{[3,0,4],[-1,0,7],[2,9,11]\}$ Check that the vectors are linearly independent and use the Gram-Schmidt process to find orthogonal vectors.

Ans. $\{[3,0,4],[-4,0,3],[0,9,0]\}$ Check that the vectors are mutually orthogonal.

7 Matrices

7.1 Introduction

Definition. *A $m \times n$ matrix is a rectangular array of numbers having m rows and n columns.*

$$A = \begin{pmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & . & a_{2n} \\ a_{31} & a_{32} & . & . & . & a_{3n} \\ . & . & & & . & . \\ . & . & & & . & . \\ a_{m1} & a_{m2} & . & . & . & a_{mn} \end{pmatrix}$$

The element a_{ij} represents the entry in the ith row and jth column. We sometimes denote A by $(a_{ij})_{m\times n}$.

7.2 Matrix Operations

Addition

We can only add two matrices of the same dimension i.e. same number of rows and columns. We then add element-wise.

Example.

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} + \begin{pmatrix} 4 & 3 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ 6 & 9 \end{pmatrix}.$$

Scalar Multiplication

If c is a real number and $A = (a_{ij})_{m\times n}$ is a matrix then $cA = (ca_{ij})_{m\times n}$.

Example.

$$5\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 10 & 5 \\ 5 & 15 \end{pmatrix}.$$

Matrix Multiplication

Given a matrix $A = (a_{ij})_{m\times n}$ and a matrix $B = (b_{ij})_{r\times s}$, we can only multiply them if $n = r$. In such a case the multiplication is defined to be the matrix

$C = (c_{ij})_{m \times s}$ as follows:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

We may view the c_{ij}th element as the dot product of the ith row of the matrix A and jth column of the matrix B.

EXAMPLE. 1.

$$\begin{aligned} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 5 & 6 \end{pmatrix} &= \begin{pmatrix} (2)(4) + (1)(5) & (2)(3) + (1)(6) \\ (1)(4) + (3)(5) & (1)(3) + (3)(6) \end{pmatrix} \\ &= \begin{pmatrix} 13 & 12 \\ 19 & 21 \end{pmatrix}. \end{aligned}$$

2. $$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 1 & 3 & 2 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 12 & 4 \\ 5 & 3 & 2 \\ 3 & 11 & 6 \end{pmatrix}$$

7.3 Special matrices

(a) A matrix is called a square matrix if the number of rows is equal to the number of columns.

(b) The transpose of a square matrix $A = (a_{ij})$ is the matrix $A^T = (a_{ji})$. The rows of A become the columns of A^T.

EXAMPLE. The transpose of $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}$ is

$$A^T = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 3 & 0 & 1 \end{pmatrix}.$$

(c) A square matrix is said to be *symmetric* if $A = A^T$, i.e. $a_{ij} = a_{ji}, \forall i, j$.

(d) A square matrix is said to be a diagonal matrix, if all the non-diagonal elements in the matrix are zero.

$$e.g. \begin{pmatrix} 3 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(e) A square matrix is said to be the *identity matrix*, if it is a diagonal matrix and all the non-zero elements are 1.

$$e.g. \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The identity marix has the property that

$$AI = A = IA$$

for all square matrices A.

7.4 Determinants

DEFINITION. *The determinant of a* 2×2 *matrix is defined as follows:*

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

DEFINITION. *The determinant of a* 3×3 *matrix is defined as follows:*

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

DEFINITION. *The definition of determinants can be generalized for a square matrix of any size. Let* $A = (a_{ij})$ *be a square* $n \times n$ *matrix. Fix a row* i*. Then the determinant is defined to be,*

$$det\ (A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}\, det\ (A_{ij}),$$

where A_{ij} *is the minor matrix obtained by deleting row* i *and column* j*. It is called the* $ij \quad Cofactor$ *of* A*.* T

EXAMPLE. Find the determinant of

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

Choose row $i = 1$.

$$A_{11} = \begin{pmatrix} -1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\det\ (A_{11}) = -2$$

$$A_{12} = \begin{pmatrix} 0 & 1 & 4 \\ 0 & 2 & 2 \\ 1 & 1 & 2 \end{pmatrix}$$

$$\det\ (A_{11}) = -6$$

$$A_{13} = \begin{pmatrix} 0 & -1 & 4 \\ 0 & 0 & 2 \\ 1 & 0 & 2 \end{pmatrix}$$

$$\det\ (A_{11}) = -2$$

$$A_{14} = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\det\ (A_{11}) = -\ 2$$

$$\begin{aligned} \det\ (A) &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det\ (A_{1j}) \\ &= (-1)^2(1)\det\ (A_{11}) + (-1)^3(1)\det\ (A_{12}) + (-1)^4(0)\det\ (A_{13}) + (-1)^5(-1)\det\ (A_{14}) \\ &= (1)(1)(-2) + (-1)(1)(-6) + (1)(0)(-2) + (-1)(-1)(-2) \\ &= 2 \end{aligned}$$

7.5 Inverses

DEFINITION. *A matrix I which has 1's on the diagonal and 0's everywhere else is called the identity matrix. This matrix has the property that $AI = A$.*

EXAMPLE. 1. A 2×2 identity matrix has the form $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

2. A 3×3 identity matrix has the form $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

DEFINITION. *Let $A = (a_{ij})$ be a square matrix. A matrix $B = (b_{ij})$ is called the inverse of A if*

$$AB = BA = I.$$

REMARK. A matrix A has an inverse if and only if $\det(A) \neq 0$.

Inverse of a 2×2 matrix

DEFINITION. *The inverse of a 2×2 matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is given by,*

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

EXAMPLE. Find the inverse of the matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

SOLUTION: We first check that the inverse of A exists!

$$\det(A) = (1)(2) - (1)(1) = 1 \neq 0.$$

Hence the inverse of A must exist and is given by

$$A^{-1} = \frac{1}{1} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

We check that this is correct by multiplying $A^{-1}A$ to see if we get the identity matrix.

$$\begin{aligned} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} &= \begin{pmatrix} (2)(1) + (-1)(1) & (1)(1) + (-1)(2) \\ (-1)(1) + (1)(2) & (-1)(1) + (1)(2) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

We can find the inverse in another way: Write as $(A|I_2)$

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right).$$

Then use matrix row operations to get it into the form $(I_2|B)$.

$$\left(\begin{array}{cc|cc} 1 & 0 & a & b \\ 0 & 1 & c & d \end{array}\right).$$

You will find that $B = A^{-1}$.

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right) \overset{R_2=R_2-R_1}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array}\right) \overset{R_1=R_1-R_2}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right)$$

Observe that the matrix obtained

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

is the inverse of A.

Inverse of a 3×3 matrix

The method is the same as in the 2×2 case:

- Check determinant A is non-zero.
- Rewrite as $(A|I_3)$.
- Use row operations to put in the form $(I_3|A^{-1})$.

Example. Let $A = \begin{pmatrix} 1 & -1 & 1 \\ 0 & -2 & 1 \\ -2 & -3 & 0 \end{pmatrix}$. Find A^{-1}.

1. Check that $\det(A) \neq 0$.

$$\begin{aligned} \det(A) &= 1(0-(-3)(1)) - (-1)(0-(-2)(1)) + 1(0-(-2)(-2)) \\ &= 3+2-4 \\ &= 1 \neq 0 \end{aligned}$$

Therefore the inverse exists.

2. Rewrite as $(A|I_3)$.

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 1 & 0 \\ -2 & -3 & 0 & 0 & 0 & 1 \end{array}\right).$$

3. Use row operations to put in the form $(I_3|A^{-1})$.

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 1 & 0 \\ -2 & -3 & 0 & 0 & 0 & 1 \end{array}\right) \overset{R_3=R_3+2R_1}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & 1 & 0 \\ 0 & -5 & 2 & 2 & 0 & 1 \end{array}\right)$$

$$\overset{R_2=\frac{-1}{2}R_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1/2 & 0 & -1/2 & 0 \\ 0 & -5 & 2 & 2 & 0 & 1 \end{array}\right)$$

$$\overset{R_3=R_3+5R_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1/2 & 0 & -1/2 & 0 \\ 0 & 0 & -1/2 & 2 & -5/2 & 1 \end{array}\right)$$

$$\overset{R_3=-2R_3}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 1 & -4 & 5 & -2 \end{array}\right)$$

$$\overset{R_2=R_2+\frac{1}{2}R_3}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 2 & -1 \\ 0 & 0 & 1 & -4 & 5 & -2 \end{array}\right)$$

$$\overset{R_1=R_1-R_3}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 5 & -5 & 2 \\ 0 & 1 & 0 & -2 & 2 & -1 \\ 0 & 0 & 1 & -4 & 5 & -2 \end{array}\right)$$

$$\overset{R_1=R_1+R_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & -2 & 2 & -1 \\ 0 & 0 & 1 & -4 & 5 & -2 \end{array}\right)$$

4.

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

8 Eigenvalues and Eigenvectors and Diagonalization

8.1 Introduction

DEFINITION. *Suppose that A is an $n \times n$ square matrix. Suppose also that $\vec{x}$ is a non-zero vector in $\mathbb{R}^n$ and that λ is a scalar so that,*

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

We then call $\vec{x}$ an eigenvector of A and λ an eigenvalue of A.

EXAMPLE. Suppose $A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$.
Then $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is an eigenvector associated to the eigenvalue 5 because

$$A\vec{x} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 5 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 5\vec{x}.$$

Also $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector associated to the eigenvalue 2 because

$$A\vec{x} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 2\vec{x}.$$

In this chapter we will learn how to find these eigenvalues and eigenvectors.

8.2 Method to find eigenvalues and eigenvectors

We start with $A\vec{x} = \lambda\vec{x}$ and rewrite it as follows,

$$\begin{aligned} A\vec{x} &= \lambda I\vec{x} \\ \lambda I\vec{x} - A\vec{x} &= 0 \\ (\lambda I - A)\vec{x} &= 0 \end{aligned}$$

THEOREM *λ is an eigenvalue of A if and only if $\lambda I - A$ is not invertible if and only if $\det(\lambda I - A) = 0$.*

Steps to find eigenvalues and eigenvectors:

1. Form the characteristic equation
$$\det(\lambda I - A) = 0.$$
2. To find all the eigenvalues of A, solve the characteristic equation.
3. For each eigenvalue λ, to find the corresponding set of eigenvectors, solve the linear system of equations
$$(\lambda I - A)\vec{x} = 0$$

EXAMPLE. 1. $A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$. Find the eigenvalues and eigenvectors of A.

SOLUTION:

- We will find the characteristic equation of A. We will first find the matrix $\lambda I - A$.
$$\begin{aligned} \lambda I - A &= \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \\ &= \begin{pmatrix} \lambda - 4 & 2 \\ 1 & \lambda - 3 \end{pmatrix} \end{aligned}$$
The determinant of this matrix gives us the characteristic polynomial.
$$\begin{aligned} \det(\lambda I - A) &= 0 \\ \begin{vmatrix} \lambda - 4 & 2 \\ 1 & \lambda - 3 \end{vmatrix} &= 0 \\ (\lambda - 4)(\lambda - 3) - 2 &= 0 \\ \lambda^2 - 7\lambda + 12 - 2 &= 0 \\ \lambda^2 - 7\lambda + 10 &= 0 \end{aligned}$$
- To find all the eigenvalues of A, solve the characteristic equation.
$$\begin{aligned} \lambda^2 - 7\lambda + 10 &= 0 \\ (\lambda - 5)(\lambda - 2) &= 0 \end{aligned}$$
So we have two eigenvalues $\lambda_1 = 2$, $\lambda_2 = 5$.

- For each eigenvalue λ, to find the corresponding set of eigenvectors, we simply solve the linear system of equations given by,

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

case(i) $\lambda_1 = 2$.

$$\begin{aligned}
(2I - A)\vec{x} &= 0 \\
\begin{pmatrix} 2-4 & -2 \\ -1 & 2-3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} -2 & -2 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} -2x_1 - 2x_2 \\ -x_1 - x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

We get two equations:

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

Both equations give the relation $x_1 = -x_2$. Therefore the set of eigenvectors corresponding to $\lambda = 2$ is given by:

$$\begin{aligned}
&\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} | x_1 = -x_2\} \\
&\{ \begin{pmatrix} -x_2 \\ x_2 \end{pmatrix} \} \\
&\{ \ x_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} | x_2 \text{ is a real number.}\}
\end{aligned}$$

An eigenvector corresponding to $\lambda_1 = 2$ is $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

case(i) $\lambda_2 = 5$.

$$\begin{aligned}
(5I - A)\vec{x} &= 0 \\
\begin{pmatrix} 5-4 & -2 \\ -1 & 5-3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} x_1 - 2x_2 \\ -x_1 + 2x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

We get two equations:

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

Both equations give the relation $x_1 = 2x_2$. Therefore a general eigenvector corresponding to $\lambda_2 = 2$ is,

$$\begin{aligned} &\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} | x_1 = 2x_2 \} \\ &\{ \begin{pmatrix} 2x_2 \\ x_2 \end{pmatrix} \} \\ &\{ x_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} | x_2 \text{ is a real number.} \} \end{aligned}$$

An eigenvector corresponding to $\lambda_2 = 5$ is $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

In this lecture we will find the eigenvalues and eigenvectors of 3×3 matrices.

EXAMPLE. An example of three distinct eigenvalues.

$$A = \begin{pmatrix} 4 & 0 & 1 \\ -1 & -6 & -2 \\ 5 & 0 & 0 \end{pmatrix}.$$

SOLUTION: Recall,

Steps to find eigenvalues and eigenvectors:

1. Form the characteristic equation
$$\det(\lambda I - A) = 0.$$
2. To find all the eigenvalues of A, solve the characteristic equation.
3. For each eigenvalue λ, to find the corresponding set of eigenvectors, solve the linear system of equations
$$(\lambda I - A)\vec{x} = 0$$

Step 1. **Form the Characteristic Equation.**
The characteristic equation is:

$$\begin{aligned} \det\,(\lambda I - A) &= 0 \\ \begin{vmatrix} \lambda - 4 & 0 & -1 \\ 1 & \lambda + 6 & 2 \\ -5 & 0 & \lambda \end{vmatrix} &= 0 \\ (\lambda - 4)((\lambda + 6)(\lambda) - 0) - 1(0 - (-5)(\lambda + 6)) &= 0 \\ \lambda^3 + 2\lambda^2 - 29\lambda - 30 &= 0 \end{aligned}$$

Step 2. **Find the eigenvalues.**
We need to solve the characteristic equation. i.e. we need to factorize the characteristic polynomial. We can factorize it by either using long division or by directly trying to spot a common factor.

Method 1: Long Division.
We want to factorize this cubic polynomial. In general it is quite difficult to guess what the factors may be. We try $\lambda = \pm 1, \pm 2, \pm 3, etc.$ and hope to quickly find one factor. Let us try $\lambda = -1$. We divide the polynomial $\lambda^3 + 2\lambda^2 - 29\lambda - 30$ by $\lambda + 1$, to get,

$$
\begin{array}{rrrrr}
 & & \lambda^2 & +\lambda & -30 \\ \cline{2-5}
\lambda + 1 & |\lambda^3 & +2\lambda^2 & -29\lambda & -30; \\
 & -\lambda^3 & +\lambda^2 & & \\ \cline{2-3}
 & & \lambda^2 & -29\lambda & -30 \\
 & - & \lambda^2 & +\lambda & \\ \cline{3-5}
 & & & -30\lambda & -30 \\
 & & - & -30\lambda & -30 \\ \cline{4-5}
 & & & & 0
\end{array}
$$

The quotient is $\lambda^2 + \lambda - 30$.
The remainder is 0.
Therefore $\lambda^3 + 2\lambda^2 - 29\lambda - 30 = (\lambda + 1)(\lambda^2 + \lambda - 30) + 0$.

$$
\begin{aligned}
\lambda^3 + 2\lambda^2 - 29\lambda - 30 &= 0 \\
(\lambda + 1)(\lambda^2 + \lambda - 30) &= 0 \\
(\lambda + 1)(\lambda + 6)(\lambda - 5) &= 0
\end{aligned}
$$

Therefore the eigenvalues are: $\{-1, -6, 5\}$.
Method 2: Direct factorization by spotting common factor.

$$
\begin{aligned}
\begin{vmatrix} \lambda - 4 & 0 & 0 \\ 1 & \lambda + 6 & 2 \\ -5 & 0 & \lambda \end{vmatrix} &= 0 \\
(\lambda - 4)((\lambda + 6)(\lambda) - 0) - 1(0 - (-5)(\lambda + 6)) &= 0 \\
(\lambda - 4)((\lambda + 6)(\lambda)) - 5(\lambda + 6) &= 0 \\
(\lambda + 6)(\lambda(\lambda - 4) - 5) &= 0 \\
(\lambda + 6)(\lambda^2 - 4\lambda - 5) &= 0 \\
(\lambda + 6)(\lambda - 5)(\lambda + 1) &= 0
\end{aligned}
$$

Therefore the eigenvalues of A are: $\lambda_1 = -1$, $\lambda_2 = -6$ and $\lambda_3 = 5$.

Step 3. **Find Eigenvectors corresponding to each Eigenvalue:**
We now need to find eigenvectors corresponding to each eigenvalue.

case(i) $\lambda_1 = -1$.
The eigenvectors are the solution space of the following system:

$$\begin{pmatrix} -5 & 0 & -1 \\ 1 & 5 & 2 \\ -5 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned} -5x_1 - x_3 &= 0; \quad x_1 = \frac{-1}{5}x_3 \\ x_1 + 5x_2 + 2x_3 &= 0; \quad x_2 = \frac{-9}{25}x_3 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda_1 = -1$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} |x_1 = \frac{-1}{5}x_3, x_2 = \frac{-9}{25}x_3\}$$

$$\{ \begin{pmatrix} \frac{-1}{5}x_3 \\ \frac{-9}{25}x_3 \\ x_3 \end{pmatrix} |x_3 \text{ is a non-zero real number}\}$$

$$\{ x_3 \begin{pmatrix} \frac{-1}{5} \\ \frac{-9}{25} \\ 1 \end{pmatrix} |x_3 \text{ is a nn-zero real number}\}$$

An eigenvector corresponding to $\lambda_1 = -1$ is $\begin{pmatrix} \frac{-1}{5} \\ \frac{-9}{25} \\ 1 \end{pmatrix}$.

case(ii) $\lambda_2 = -6$.
The eigenvectors are the solution space of the following system:

$$\begin{pmatrix} -10 & 0 & -1 \\ 1 & 0 & 2 \\ -5 & 0 & -6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned} -10x_1 - x_3 &= 0; \quad x_3 = -10x_1 \\ x_1 + 2x_3 &= 0; \quad x_1 = -2x_3 = -2(-10)x_1 \\ x_1 &= 0 = x_3 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda_2 = -6$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = x_3 = 0\}$$

$$\{ \begin{pmatrix} 0 \\ x_2 \\ 0 \end{pmatrix} | x_2 \text{ is a non-zero real number}\}$$

$$\{ x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} | x_2 \text{ is a non-zero real number}\}$$

An eigenvector corresponding to $\lambda_2 = -6$ is $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

case(iii) $\lambda_3 = 5$.

The eigenvectors are the solution space of the following system:

$$\begin{pmatrix} 1 & 0 & -1 \\ 1 & 11 & 2 \\ -5 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned} x_1 &= x_3 \\ x_2 &= \frac{-3}{11} x_3 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda_3 = 5$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = x_3, x_2 = \frac{-3}{11} x_3\}$$

$$\{ \begin{pmatrix} x_3 \\ \frac{-3}{11} x_3 \\ x_3 \end{pmatrix} | x_3 \text{ is a non-zero real number}\}$$

$$\{ x_3 \begin{pmatrix} 1 \\ \frac{-3}{11} \\ 1 \end{pmatrix} | x_3 \text{ is a non-zero real number}\}$$

An eigenvector corresponding to $\lambda_3 = 5$ is $\begin{pmatrix} 1 \\ \frac{-3}{11} \\ 1 \end{pmatrix}$.

EXAMPLE. An example of repeated eigenvalue having only two eigenvectors.

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

SOLUTION: Recall,

Steps to find eigenvalues and eigenvectors:

1. Form the characteristic equation

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

2. To find all the eigenvalues of A, solve the characteristic equation.

3. For each eigenvalue λ, to find the corresponding set of eigenvectors, solve the linear system of equations

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

Step 1. **Form the Characteristic Equation.**
The characteristic equation is:

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

Step 2. **Find the eigenvalues.**
We need to solve the characteristic equation. i.e. we need to factorize the characteristic polynomial. We can factorize it by either using long division or by directly trying to spot a common factor.
Method 1: Long Division.
We want to factorize this cubic polynomial. In general it is quite difficult to guess what the factors may be. We try $\lambda = \pm 1, \pm 2, \pm 3, etc.$

and hope to quickly find one factor. Let us try $\lambda = -1$. We divide the polynomial $\lambda^3 + 2\lambda^2 - 29\lambda - 30$ by $\lambda + 1$, to get,

$$\begin{array}{rrrrr}
 & & \lambda^2 & -\lambda & -2 \\
\cline{2-5}
\lambda + 1 & \multicolumn{1}{|r}{\lambda^3} & & -3\lambda & -2; \\
 & -\lambda^3 & +\lambda^2 & & \\
\cline{2-3}
 & & -\lambda^2 & -3\lambda & -2 \\
 & - & -\lambda^2 & -\lambda & \\
\cline{3-4}
 & & & -2\lambda & -2 \\
 & & - & -2\lambda & -2 \\
\cline{4-5}
 & & & & 0
\end{array}$$

The quotient is $\lambda^2 - \lambda - 2$.
The remainder is 0.
Therefore $\lambda^3 - 3\lambda - 2 = (\lambda + 1)(\lambda^2 - \lambda - 2) + 0$.

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

Therefore the eigenvalues are: $\{-1, 2\}$.
Method 2: Direct factorization by spotting common factor.

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

Therefore the eigenvalues of A are: $\{-1, 2\}$.

Continuing from the last lecture...

EXAMPLE. An example of repeated eigenvalue having only two eigenvectors.

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

SOLUTION: Recall,

Steps to find eigenvalues and eigenvectors:

1. Form the characteristic equation
$$\det(\lambda I - A) = 0.$$
2. To find all the eigenvalues of A, solve the characteristic equation.
3. For each eigenvalue λ, to find the corresponding set of eigenvectors, solve the linear system of equations
$$(\lambda I - A)\vec{x} = 0$$

Step 1. **Form the Characteristic Equation.**
The characteristic equation is:

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

Step 2. **Find the eigenvalues.**
We need to solve the characteristic equation. i.e. we need to factorize the characteristic polynomial. We can factorize it by either using long division or by directly trying to spot a common factor.

Method 1: Long Division.
We want to factorize this cubic polynomial. In general it is quite difficult to guess what the factors may be. We try $\lambda = \pm 1, \pm 2, \pm 3, etc.$ and hope to quickly find one factor. Let us try $\lambda = -1$. We divide the polynomial $\lambda^3 + 2\lambda^2 - 29\lambda - 30$ by $\lambda + 1$, to get,

$$
\begin{array}{rrrrr}
 & & \lambda^2 & -\lambda & -2 \\ \cline{2-5}
\lambda + 1 & \big| \lambda^3 & & -3\lambda & -2; \\
 & -\lambda^3 & +\lambda^2 & & \\ \cline{2-3}
 & & -\lambda^2 & -3\lambda & -2 \\
 & - & -\lambda^2 & -\lambda & \\ \cline{3-4}
 & & & -2\lambda & -2 \\
 & & - & -2\lambda & -2 \\ \cline{4-5}
 & & & & 0
\end{array}
$$

The quotient is $\lambda^2 - \lambda - 2$.
The remainder is 0.
Therefore $\lambda^3 - 3\lambda - 2 = (\lambda + 1)(\lambda^2 - \lambda - 2) + 0$.

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

Therefore the eigenvalues are: $\{-1, 2\}$.
Method 2: Direct factorization by spotting common factor.

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

Therefore the eigenvalues of A are: $\{-1, 2\}$.

Step 3. **Find Eigenvectors corresponding to each Eigenvalue:**

We now need to find eigenvectors corresponding to each eigenvalue.

case(i) $\lambda_1 = -1$.

The eigenvectors are the solution space of the following system:

$$\begin{aligned} \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ -x_1 - x_2 - x_3 &= 0; \quad x_1 = -x_2 - x_3 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda_1 = -1$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = -x_2 - x_3 \}$$

$$\{ \begin{pmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{pmatrix} | \text{ at least one of } x_2 \text{ and } x_3 \text{ is a non-zero real number} \}$$

$$\{ x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} | \text{ at least one of } x_2 \text{ and } x_3 \text{ is a non-zero real number} \}$$

We therefore can get two linearly independent eigenvectors corresponding to $\lambda_1 = -1$:

$$\vec{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

case(ii) $\lambda_2 = 2$.

The eigenvectors are the solution space of the following system:

$$\begin{aligned} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ 2x_1 - x_2 - x_3 &= 0 \\ -x_1 + 2x_2 - x_3 &= 0 \\ -x_1 - x_2 + 2x_3 &= 0 \end{aligned}$$

Since this system of equations looks fairly complicated, it may be a good idea to use row reduction to simplify the system.

$$\left(\begin{array}{ccc|c} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \end{array}\right) \xrightarrow{R_3 \leftrightarrow R_1} \left(\begin{array}{ccc|c} -1 & -1 & 2 & 0 \\ -1 & 2 & -1 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right)$$

$$\xrightarrow{R_1=(-1)R_1} \left(\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ -1 & 2 & -1 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right)$$

$$\xrightarrow[R_2=R_1+R_2]{R_3=R_2-2R_1} \left(\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ 0 & 3 & -3 & 0 \\ 0 & -3 & 3 & 0 \end{array}\right)$$

$$\xrightarrow{R_2=1/3R_2} \left(\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -3 & 3 & 0 \end{array}\right)$$

$$\xrightarrow{R_3=R_3+3R_2} \left(\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The resulting set of equations is:

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

The set of eigenvectors corresponding to $\lambda_2 = 2$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = x_2 = x_3 \}$$

$$\{ \begin{pmatrix} x_3 \\ x_3 \\ x_3 \end{pmatrix} | x_3 \text{ is a non-zero real number} \}$$

$$\{ x_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} | x_3 \text{ is a non-zero real number} \}$$

An eigenvector corresponding to $\lambda_2 = 2$ is $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

EXAMPLE. An example of a repeated eigenvalue having only one linearly independent eigenvector.

$$A = \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{pmatrix}.$$

SOLUTION: The characteristic polynomial for A:

$$\begin{aligned} \begin{vmatrix} \lambda - 6 & -3 & 8 \\ 0 & \lambda + 2 & 0 \\ -1 & 0 & \lambda + 3 \end{vmatrix} &= 0 \\ (\lambda - 6)((\lambda + 2)(\lambda + 3) - 0) - 3(0 - 0) + 8(0 - (-1)(\lambda + 2)) &= 0 \\ \lambda^3 - \lambda^2 - 16\lambda - 20 &= 0 \\ (\lambda + 2)(\lambda^2 - 3\lambda - 10) &= 0 \\ (\lambda + 2)^2(\lambda - 5) &= 0 \end{aligned}$$

Therefore the eigenvalues of A are: $\lambda_1 = -2$, $\lambda_2 = 5$.

case(i) $\lambda_1 = -2$.

The eigenvectors are the solution space of the following system:

$$\begin{aligned} \begin{pmatrix} -8 & -3 & 8 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ -x_1 + x_3 &= 0 \implies x_1 = x_3 \\ -8x_1 - 3x_2 + 8x_3 &= 0 \implies x_2 = 0 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda_1 = -2$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = x_3, x_2 = 0 \}$$

$$\{ \begin{pmatrix} x_3 \\ 0 \\ x_3 \end{pmatrix} | x_3 \text{ is a non-zero real number} \}$$

$$\{ x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} | x_3 \text{ is a non-zero real number} \}$$

An eigenvector corresponding to $\lambda_1 = -2$ is $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

case(ii) $\lambda_2 = 5$.
The eigenvectors are the solution space of the following system:

$$\begin{aligned}
\begin{pmatrix} -1 & -3 & 8 \\ 0 & 7 & 0 \\ -1 & 0 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
-x_1 + 8x_3 &= 0 \Longrightarrow x_1 = 8x_3 \\
7x_2 &= 0 \Longrightarrow x_2 = 0
\end{aligned}$$

The set of eigenvectors corresponding to $\lambda_2 = 5$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = 8x_3, x_2 = 0\}$$

$$\{ \begin{pmatrix} 8x_3 \\ 0 \\ x_3 \end{pmatrix} | x_3 \text{ is a non-zero real number}\}$$

$$\{ x_3 \begin{pmatrix} 8 \\ 0 \\ 1 \end{pmatrix} | x_3 \text{ is a non-zero real number}\}$$

An eigenvector corresponding to $\lambda_2 = 5$ is $\begin{pmatrix} 8 \\ 0 \\ 1 \end{pmatrix}$.

EXAMPLE. An example of an eigenvalue repeated three times having only two linearly independent eigenvectors.

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

SOLUTION: The characteristic equation for A:

$$\begin{aligned} \det\,(\lambda I - A) &= 0 \\ \begin{vmatrix} \lambda-4 & 0 & 1 \\ 0 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{vmatrix} &= 0 \\ \lambda^3 - 9\lambda^2 + 27\lambda - 27 &= 0 \\ (\lambda-3)^3 &= 0 \end{aligned}$$

Therefore A has only one eigenvalue: $\lambda = 3$.

The eigenvectors corresponding to $\lambda = 3$ are the solution space of the following system:

$$\begin{aligned} \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ -x_1 + x_3 &= 0 \Longrightarrow x_1 = x_3 \end{aligned}$$

The set of eigenvectors corresponding to $\lambda = 3$ is,

$$\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} | x_1 = x_3 \}$$

$$\{ \begin{pmatrix} x_3 \\ x_2 \\ x_3 \end{pmatrix} | \text{ at least one of } x_2 \text{ and } x_3 \text{ is a non-zero real number} \}$$

$$\{ \ x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} | \text{ at least one on } x_2 \text{ and } x_3 \text{ is a non-zero real number} \}$$

Therefore we can find two linearly independent eigenvectors corresponding to the eigenvalue $\lambda = 3$:

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

8.3 Dimension of Eigenspace

Definition. *The rank of a matrix is defined to be the number of linearly independent rows or columns.*

Remark. We use the following results to determine the rank of a matrix.

- Let A be a square matrix and let $A_{echelon}$ be the row echelon form of the matrix. Then both A and $A_{echelon}$ have the same number of linearly independent rows.
- Consequently, $\text{Rank}(A) = \text{Rank}(A_{echelon})$.
- The number of linearly independent rows of a matrix in row echelon form is just the number of non-zero rows.
- We may conclude that

$$\text{Rank}(A) = \text{Rank}(A_{echelon}) = \text{number of non-zero rows in } A_{echelon}.$$

Example. 1. $\text{Rank}\begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} = \text{Rank}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 1.$

2. $\text{Rank}\begin{pmatrix} -8 & -3 & 8 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \text{Rank}\begin{pmatrix} 1 & 0 & -1 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 2.$

3. $\text{Rank}\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \text{Rank}\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 1.$

Definition. *Let A be a square matrix of size $n \times n$ having an eigenvalue λ. The set of eigenvectors corresponding to the eigenvalue λ, alongwith the zero vector, is called the eigenspace corresponding to eigenvalue λ.*
The dimension of the eigenspace is defined to be the maximum number of linearly independent vectors in the eigenspace.

Theorem *Let A be a matrix of size $n \times n$ and let λ be an eigenvalue of A repeating k times. Then the dimension of the eigenspace of λ or equivalently the maximum number of linearly independent eigenvectors corresponding to λ is:*

$$n - Rank(\lambda I - A).$$

Example. Let us use this result to confirm our calculations in the previous examples.

1. Recall $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

   - has eigenvalue $\lambda = -1$ occurring with multiplicity 2.
   - We found two linearly independent eigenvectors corresponding to this eigenvalue.
     $\vec{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

   We will use the above theorem to confirm that we should indeed expect to find two linearly independent eigenvectors. By the above theorem, the maximum number of linearly independent eigenvectors corresponding to $\lambda = -1$ is:

$$\begin{aligned} &= 3 - \text{Rank}(\lambda I - A) \\ &= 3 - \text{Rank} \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \\ &= 3 - 1 \\ &= 2 \end{aligned}$$

2. Recall $A = \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ -1 & 0 & 1 \end{pmatrix}$

   - has eigenvalue $\lambda = -2$ occurring with multiplicity 2.

- We found only one linearly independent eigenvector corresponding to this eigenvalue.
$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

We will use the above theorem to confirm that we should indeed expect to find only one linearly independent eigenvector. By the above theorem, the maximum number of linearly independent eigenvectors corresponding to $\lambda = -2$ is:

$$\begin{aligned} &= 3 - \text{Rank}(\lambda I - A) \\ &= 3 - \text{Rank} \begin{pmatrix} -8 & -3 & 8 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \\ &= 3 - 2 \\ &= 1 \end{aligned}$$

3. Recall $A = \begin{pmatrix} 4 & 0 & -1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$

- has eigenvalue $\lambda = 3$ occurring with multiplicity 3.
- We found only two linearly independent eigenvectors corresponding to this eigenvalue.
$\vec{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

We will use the above theorem to confirm that we should indeed expect to find only two linearly independent eigenvectors. By the above theorem, the maximum number of linearly independent eigenvectors corresponding to $\lambda = 3$ is:

$$
\begin{aligned}
&= 3 - \mathrm{Rank}(\lambda I - A) \\
&= 3 - \mathrm{Rank}\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \\
&= 3 - 1 \\
&= 2
\end{aligned}
$$

8.4 Diagonalization

Definition. *A square matrix of size n is called a diagonal matrix if it is of the form*

$$\begin{pmatrix} d_1 & 0 & . & . & . & 0 \\ 0 & d_2 & 0 & . & . & 0 \\ . & . & & & & . \\ . & . & & & & . \\ 0 & . & . & & d_{n-1} & 0 \\ 0 & . & . & . & 0 & d_n \end{pmatrix}$$

Definition. *Two square matrices A and B of size n are said to be similar, if there exists an invertible matrix P such that*

$$B = P^{-1}AP$$

Definition. *A matrix A is diagonalizable if A is similar to a diagonal matrix. i.e. there exists an invertible matrix P such that $P^{-1}AP = D$ where D is a diagonal matrix. In this case, P is called a diagonalizing matrix.*

Example. Recall the example we did in the last chapter.

$$A = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}.$$

Solution: In previous lectures we have found,
the eigenvalues of A are: $\lambda_1 = 2$, $\lambda_2 = 5$.
An eigenvector corresponding to $\lambda_1 = 2$ is $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.
An eigenvector corresponding to $\lambda_2 = 5$ is $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.
Define a new matrix P such that its columns are the eigenvectors of A.

$$P = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix}.$$

Note that since eigenvectors corresponding to distinct eigenvalues are linearly independent, the columns of P are linearly independent, hence P is an invertible matrix. (We can also confirm this by verifying that the determinant of P is non-zero).

$$P^{-1} = \frac{-1}{3} \begin{pmatrix} 1 & -2 \\ -1 & -1 \end{pmatrix}$$

$$
\begin{aligned}
P^{-1}AP &= \frac{-1}{3}\begin{pmatrix} 1 & -2 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \\
&= \frac{-1}{3}\begin{pmatrix} 2 & -4 \\ -5 & -5 \end{pmatrix}\begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \\
&= \frac{-1}{3}\begin{pmatrix} -6 & 0 \\ 0 & -15 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix},
\end{aligned}
$$

which is a diagonal matrix. Therefore A is diagonalizable and a diagonalizing matrix for A is

$$P = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix},$$

and the corresponding diagonal matrix D is

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}.$$

THEOREM *Let A be a square matrix of size $n \times n$. Then A is diagonalizable if and only if there exists n linearly independent eigenvectors. In this case a diagonalizing matrix P is formed by taking as its columns the eigenvectors of A,*

$$P^{-1}AP = D.$$

where D is a diagonal matrix whose diagonal entries are eigenvalues corresponding to the eigenvectors of A. The ith diagonal entry is the eigenvalue corresponding to the eigenvector in the ith column of P.

EXAMPLE. An example where the matrix is not diagonalizable. Recall

$$A = \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

(i) Find the eigenvalues and eigenvectors of A.

(ii) Is A diagonalizable?

(iii) If it is diagonalizable, find a diagonalizing matrix P such that $P^{-1}AP = D$, where D is a diagonal matrix.

SOLUTION: In previous lectures we have found,
A had two eigenvalues:
$\lambda_1 = -1$ occurring with multiplicity 2, and
$\lambda_2 = 5$ occurring with multiplicity 1.
We could find only one linearly independent eigenvector corresponding to $\lambda_1 = -1$,
and one linearly independent eigenvector corresponding to $\lambda_2 = 5$.
Therefore we could find only two linearly independent eigenvectors. By the above theorem, A is not diagonalizable.

PROPOSITION *Let A be a square matrix of size $n \times n$. If λ_1 and λ_2 are two distinct eigenvalues of A, and $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors corresponding to λ_1 and λ_2 respectively, then $\vec{v}_1$ and $\vec{v}_2$ are linearly independent.*

COROLLARY *Let A be a square matrix of size n. If A has n distinct eigenvalues then A is diagonalizable.*

EXAMPLE. Recall

$$A = \begin{pmatrix} 4 & 0 & 1 \\ -1 & -6 & -2 \\ 5 & 0 & 0 \end{pmatrix}.$$

(i) Find the eigenvalues and eigenvectors of A.

(ii) Is A diagonalizable?

(iii) Find a diagonalizing matrix P such that $P^{-1}AP = D$, where D is a diagonal matrix.

SOLUTION: (i) In previous lectures we have found,
The eigenvalues of A are: $\lambda_1 = -1$, $\lambda_2 = -6$ and $\lambda_3 = 5$.
An eigenvector corresponding to $\lambda_1 = -1$ is $\begin{pmatrix} \frac{-1}{5} \\ \frac{-9}{25} \\ 1 \end{pmatrix}$.

An eigenvector corresponding to $\lambda_2 = -6$ is $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

An eigenvector corresponding to $\lambda_3 = 5$ is $\begin{pmatrix} 1 \\ \frac{-3}{11} \\ 1 \end{pmatrix}$.

(ii) The matrix A is diagonalizable, since A is a square matrix of size 3×3 and we have found three linearly independent eigenvectors for A.

(iii) A diagonalizing matrix P has the eigenvectors as its columns. Let

$$P = \begin{pmatrix} \frac{-1}{5} & 0 & 1 \\ \frac{-9}{25} & 1 & \frac{-3}{11} \\ 1 & 0 & 1 \end{pmatrix}.$$

The corresponding diagonal matrix D will have the eigenvalues as diagonal entries corresponding to the columns of P. i.e. the ith diagonal entry will be the eigenvalue corresponding to the eigenvector in the ith column.

$$D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

We will now verify that $P^{-1}AP = D$.

$$\begin{aligned}
AP &= \begin{pmatrix} 4 & 0 & 1 \\ -1 & -6 & -2 \\ 5 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{-1}{5} & 0 & 1 \\ \frac{-9}{25} & 1 & \frac{-3}{11} \\ 1 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1/5 & 0 & 5 \\ 9/25 & -6 & -15/11 \\ -1 & 0 & 5 \end{pmatrix} \\
PD &= \begin{pmatrix} \frac{-1}{5} & 0 & 1 \\ \frac{-9}{25} & 1 & \frac{-3}{11} \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\
&= \begin{pmatrix} 1/5 & 0 & 5 \\ 9/25 & -6 & -15/11 \\ -1 & 0 & 5 \end{pmatrix} \\
AP &= PD \qquad (1)
\end{aligned}$$

Since the columns of P are linearly independent, P has non-zero determinant and is therefore invertible. We multiply (1) by P^{-1} on both sides.

$$\begin{aligned} P^{-1}AP &= P^{-1}PD, \\ &= ID \\ &= D \end{aligned}$$

8.5 Diagonalization of symmetric matrices

DEFINITION. *Let A be a square matrix of size n. A is a symmetric matrix if $A^T = A$*

DEFINITION. *A matrix P is said to be orthogonal if its columns are mutually orthogonal.*

DEFINITION. *A matrix P is said to be orthonormal if its columns are unit vectors and P is orthogonal.*

PROPOSITION *An orthonormal matrix P has the property that*

$$P^{-1} = P^T.$$

THEOREM *If A is a real symmetric matrix then there exists an orthonormal matrix P such that*

(i) *$P^{-1}AP = D$, where D a diagonal matrix.*

(ii) *The diagonal entries of D are the eigenvalues of A.*

(iii) *If $\lambda_i \neq \lambda_j$ then the eigenvectors are orthogonal.*

(iv) *The column vectors of P are linearly independent eigenvectors of A, that are mutually orthogonal.*

EXAMPLE. Recall

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

(i) Find the eigenvalues and eigenvectors of A.

(ii) Is A diagonalizable?

(iii) Find an orthonormal matrix P such that $P^T AP = D$, where D is a diagonal matrix.

SOLUTION: We have found the eigenvalues and eigenvectors of this matrix in a previous lecture.

(i), (ii) Observe that A is a real symmetric matrix. By the above theorem, we know that A is diagonalizable. i.e. we will be able to find a sufficient number of linearly independent eigenvectors.

The eigenvalues of A were; $-1, 2$. We found two linearly independent eigenvectors corresponding to $\lambda_1 = -1$: $\vec{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

And one eigenvector corresponding to $\lambda_2 = 2$: $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

(iii) We now want to find an orthonormal diagonalizing matrix P.

Since A is a real symmetric matrix, eigenvectors corresponding to distinct eigenvalues are orthogonal.

$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

However the eigenvectors corresponding to eigenvalue $\lambda_1 = -1$, $\vec{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ are not orthogonal to each other, since we chose them from the eigenspace by making arbitrary choices*. We will have to use Gram Schmidt to make the two vectors orthogonal.

$$\begin{aligned}
\vec{u}_1 &= \vec{v}_1 \\
\text{Proj}_{\vec{u}_1}\vec{v}_2 &= \frac{((-1,1,0).(-1,0,1))}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \\
\vec{u}_2 &= \vec{v}_2 - \text{Proj}_{\vec{u}_1}\vec{v}_2 \\
&= \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} -1/2 \\ 1/2 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}
\end{aligned}$$

We now have a set of orthogonal vectors:

$$\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}\}.$$

We normalize the vectors to get a set of orthonormal vectors:

$$\{\begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix}\}.$$

We are now finally ready to write the orthonormal diagonalizing matrix:

$$P = \begin{pmatrix} 1\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix}$$

and the corresponding diagonal matrix D

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We will now verify that $P^TAP = D$.

$$AP = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix}$$
$$= \begin{pmatrix} 2\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 2\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 2\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix}$$
$$PD = \begin{pmatrix} 1\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
$$= \begin{pmatrix} 2\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 2\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 2\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix}$$
$$AP = PD \tag{1}$$

Since the columns of P are linearly independent, P has non-zero determinant and is therefore invertible. We multiply (1) by P^{-1} on both sides.

$$\begin{aligned} P^{-1}AP &= P^{-1}PD, \\ &= ID \\ &= D \end{aligned} \tag{2}$$

Also since P is orthonormal, we have

$$P^{-1} = P^T$$

i.e. $PP^T = I = P^TP$.

$$\begin{pmatrix} 1\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1\sqrt{2} & 1/\sqrt{2} & 0 \\ -1\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore from (2) and since $P^{-1} = P^T$ we finally get the relation

$$P^TAP = D.$$

NOTE. *:Look back at how we selected the eigenvecors $\vec{v}_1$ and $\vec{v}_2$; we chose $x_2 = 1, x_3 = 0$ to get $\vec{v}_1$ and $x_2 = 0, x_3 = 1$ to get $\vec{v}_2$. If we had chosen $x_2 = 1, x_3 = 0$ to get $\vec{v}_1$ and $x_2 = -1/2, x_3 = 1$ to get $\vec{v}_2$, then $\vec{v}_1$ and $\vec{v}_2$ would be orthogonal. However it is much easier to make arbitrary choices for x_1 and x_2 and then use the Gram Schmidt Process to orthogonalize the vectors as we have done in this example.

9 Operators and Commutators

9.1 Operators

Operators are commonly used to perform a specific mathematical operation on another function. The operation can be to take the derivative or integrate with respect to a particular term, or to multiply, divide, add or subtract a number or term with regards to the initial function. Operators are commonly used in physics, mathematics and chemistry.

EXAMPLE. 1. A regular function can be thought of as an operator.

(i)

$$f : x \mapsto ax,$$

where a is a real number.
Operator: the function f
Operates on: real numbers
Action: multiply by a.

(ii) Squaring

$$f : x \mapsto x^2$$

Operator: the function f
Operates on: real numbers
Action: squaring

2. Operators which acts on functions:

(i)

$$\hat{a} : f(x) \mapsto af(x),$$

where a is a real number.
Operator: $\hat{a}$
Operates on: scalar functions
Action: multiply by a.

(ii)

$$\hat{x} : f(x) \mapsto xf(x),$$

Operator: $\hat{x}$
Operates on: scalar functions
Action: multiply by x.

(iii) Differentiation Operator

$$D : f(x) \mapsto \frac{d}{dx} f(x)$$

Operator: D
Operates on: scalar functions
Action: Differentiate with respect to x.
We can similarly define partial differentiation operators, grad operator, integral operator, etc.

(iv) The Momentum Operator

$$\hat{P} = \frac{\hbar}{i} \frac{\partial}{\partial x}$$

where:

- $\hbar$ is Planck's constant,
- i is the imaginary unit.

$$\hat{P} = \frac{\hbar}{i} \frac{\partial}{\partial x} : \psi(x,t) \mapsto \frac{\hbar}{i} \frac{\partial \psi(x,t)}{\partial x}$$

Operator: $\hat{P}$
Operates on: wave function, $\psi(x,t)$.

(iv) The Energy Operator

$$\hat{E} = \frac{\hbar}{i} \frac{\partial}{\partial t}$$

where:

- $\hbar$ is Planck's constant,
- i is the imaginary unit.

$$\hat{E} = \frac{\hbar}{i} \frac{\partial}{\partial t} : \psi(x,t) \mapsto \frac{\hbar}{i} \frac{\partial \psi(x,t)}{\partial t}$$

Operator: $\hat{E}$
Operates on: wave function, $\psi(x,t)$.

(iv) The Hamiltonian Operator

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$$

where:

– $\hbar$ is Planck's constant,
– m is the mass of the particle (regard as constant).
– i is the imaginary unit.

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} : \psi(x,t) \mapsto -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial \psi(x,t) x^2}$$

Operator: $\hat{H}$
Operates on: wave function, $\psi(x,t)$.

3. Operators which act on vectors.

   (i) Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ be a square matrix and $\vec{v} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ be an arbitrary vector. Define an operator in the following manner:

$$\begin{array}{rcl} A : \vec{v} & \mapsto & A\vec{v} \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \mapsto & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix}. \end{array}$$

Operator: A
Operates on: two dimensional vectors
Action: maps a vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ to $\begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix}$.

9.2 Linear Operators

An operator $\mathcal{O}$ is a linear operator if it satisfies the following two conditions:

(i) $\mathcal{O}(f+g) = \mathcal{O}(f) + \mathcal{O}(g)$.

(ii) $\mathcal{O}(\lambda f) = \lambda\mathcal{O}(f)$, where λ is a scalar.

EXAMPLE. Determine if the following operators are linear:

1. $\mathcal{O} = \hat{a} : f(x) \mapsto af(x)$.

SOLUTION:

$$\begin{aligned}
\mathcal{O}(f+g) &= a(f+g) \\
&= af+ag \\
&= \mathcal{O}(f)+\mathcal{O}(g) \\
\mathcal{O}(\lambda f) &= a(\lambda f) \\
&= \lambda a f \\
&= \lambda \mathcal{O}(f)
\end{aligned}$$

Therefore $\mathcal{O}$ is linear.

2. $\mathcal{O} = squaring : x \mapsto x^2$.

SOLUTION:

$$\begin{aligned}
\mathcal{O}(x+y) &= (x+y)^2 \\
&= x^2+y^2+2xy \\
&= \mathcal{O}(x)+\mathcal{O}(y)+2xy \\
\mathcal{O}(1+2) &= \mathcal{O}(3)=9 \\
\mathcal{O}(1)+\mathcal{O}(2) &= 1+4=5
\end{aligned}$$

Therefore $\mathcal{O}$ is not linear.

3. $D : f(x) \mapsto \frac{d}{dx}f(x)$.

SOLUTION:

$$\begin{aligned}
D(f+g) &= \frac{d}{dx}(f+g) \\
&= \frac{d}{dx}f+\frac{d}{dx}g \\
&= D(f)+D(g) \\
D(\lambda f) &= \frac{d}{dx}(\lambda f) \\
&= \lambda \frac{d}{dx} f \\
&= \lambda D(f)
\end{aligned}$$

Therefore D is linear.

4. $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : \vec{v} \mapsto A\vec{v}$.

SOLUTION:

$$\begin{aligned}
A(\vec{v}+\vec{w}) &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}) \\
&= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}) \\
&= \begin{pmatrix} (x_1 + y_1) + (x_2 + y_2) \\ x_2 + y_2 \end{pmatrix} \\
&= \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 + y_2 \\ y_2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
&= A(\vec{v}) + A(\vec{w}) \\
A(\lambda\vec{v}) &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (\lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) \\
&= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix} \\
&= \begin{pmatrix} \lambda x_1 + \lambda x_2 \\ \lambda x_2 \end{pmatrix} \\
&= \lambda \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} \\
&= \lambda \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \lambda A\vec{v}
\end{aligned}$$

Therefore A is linear.

9.3 Composing Operators

When we have two operators and we want to apply them to a function in sequence, we use the notation $\mathcal{O}_1 \circ \mathcal{O}_2$.

$$\mathcal{O}_1 \circ \mathcal{O}_2 : f \mapsto \mathcal{O}_1(\mathcal{O}_2(f)).$$

First apply $\mathcal{O}_2$ and then $\mathcal{O}_1$. The order is important.

Example. 1. Consider the following two operators.

$$\begin{aligned} \mathcal{O}_1 &: f(x) \mapsto xf(x) \\ \mathcal{O}_2 &: g(x) \mapsto (g(x))^2 \end{aligned}$$

Find $\mathcal{O}_1 \circ \mathcal{O}_2$ and $\mathcal{O}_2 \circ \mathcal{O}_1$.

Solution: We first find $\mathcal{O}_1 \circ \mathcal{O}_2$.

$$\begin{aligned} \mathcal{O}_1 \circ \mathcal{O}_2(h(x)) &= \mathcal{O}_1(\mathcal{O}_2(h(x)) \\ &= \mathcal{O}_1((h(x)^2) \\ &= x(h(x)^2) \\ \mathcal{O}_1 \circ \mathcal{O}_2 : h(x) &\mapsto x(h(x)^2 \end{aligned}$$

We now find $\mathcal{O}_2 \circ \mathcal{O}_1$.

$$\begin{aligned} \mathcal{O}_2 \circ \mathcal{O}_1(h(x)) &= \mathcal{O}_2(\mathcal{O}_1(h(x)) \\ &= \mathcal{O}_2(xh(x)) \\ &= x^2(h(x)^2) \\ \mathcal{O}_2 \circ \mathcal{O}_1 : h(x) &\mapsto x^2(h(x)^2) \end{aligned}$$

Observe that $\mathcal{O}_1 \circ \mathcal{O}_2 \neq \mathcal{O}_2 \circ \mathcal{O}_1$. Hence the order in which you apply the operators is important.

9.4 Commutators

Let $\mathcal{O}_A$ and $\mathcal{O}_B$ be two operators. The commutator of $\mathcal{O}_A$ and $\mathcal{O}_B$ is the operator defined as

$$[\mathcal{O}_A, \mathcal{O}_B] = \mathcal{O}_A \circ \mathcal{O}_B - \mathcal{O}_B \circ \mathcal{O}_A.$$

EXAMPLE. 1. Consider the following two operators.

$$\begin{aligned} \mathcal{O}_A &: \vec{v} \mapsto A\vec{v}, \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ \mathcal{O}_B &: \vec{v} \mapsto B\vec{v}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

Find $[\mathcal{O}_A, \mathcal{O}_B]$.

SOLUTION: We first find $\mathcal{O}_A \circ \mathcal{O}_B$.

$$\begin{aligned} \mathcal{O}_A \circ \mathcal{O}_B(\vec{v}) &= \mathcal{O}_A(\mathcal{O}_B(\vec{v})) \\ &= \mathcal{O}_A(B\vec{v}) \\ &= AB\vec{v} \\ \mathcal{O}_A \circ \mathcal{O}_B : \vec{v} &\mapsto AB\vec{v}, \quad AB = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ 0 \end{pmatrix} \end{aligned}$$

We now find $\mathcal{O}_2 \circ \mathcal{O}_1$.

$$\begin{aligned} \mathcal{O}_B \circ \mathcal{O}_A(\vec{v}) &= \mathcal{O}_B(\mathcal{O}_A(\vec{v})) \\ &= \mathcal{O}_B(A\vec{v}) \\ &= BA\vec{v} \\ \mathcal{O}_B \circ \mathcal{O}_A : \vec{v} &\mapsto BA\vec{v}, \quad BA = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 \end{pmatrix} \end{aligned}$$

Therefore the commutator of $\mathcal{O}_A$ and $\mathcal{O}_B$ is

$$\begin{aligned} [\mathcal{O}_A, \mathcal{O}_B] &= \mathcal{O}_A \circ \mathcal{O}_B - \mathcal{O}_B \circ \mathcal{O}_A \\ [\mathcal{O}_A, \mathcal{O}_B](\vec{v}) &= (\mathcal{O}_A \circ \mathcal{O}_B - \mathcal{O}_B \circ \mathcal{O}_A)(\vec{v}) \\ &= (AB - BA)(\vec{v}) \end{aligned}$$

2. Consider the following two operators.

$$\begin{aligned} D &: f(x) \mapsto (D)f(x) = \frac{d}{dx}f(x) \\ xD &: f(x) \mapsto (xD)f(x) = x\frac{d}{dx}f(x) \end{aligned}$$

Find $[D, xD]$.

SOLUTION:

$$\begin{aligned} (D \circ xD)(f(x)) &= D(xD(f(x))) \\ &= D(x\frac{d}{dx}f(x)) \\ &= \frac{d}{dx}(x\frac{d}{dx}f(x)) \\ &= x\frac{d^2f}{dx^2} + \frac{df}{dx} \\ &= (xD^2 + D)(f(x)) \\ (xD \circ D)(f(x)) &= xD(D(f(x))) \\ &= xD(\frac{d}{dx}f(x)) \\ &= x\frac{d}{dx}(\frac{d}{dx}f(x)) \\ &= x\frac{d^2f}{dx^2} \\ &= (xD^2)(f(x)) \end{aligned}$$

Therefore the commutator of D and xD is

$$\begin{aligned} [D, xD] &= D \circ xD - xD \circ D \\ &= xD^2 + D - xD^2 \\ &= D \end{aligned}$$

3. Find the commutator $[x\frac{\partial}{\partial y}, y\frac{\partial}{\partial x}]$.

Solution:

$$[x\frac{\partial}{\partial y}, y\frac{\partial}{\partial x}] = x\frac{\partial}{\partial y} \circ y\frac{\partial}{\partial x} - y\frac{\partial}{\partial x} \circ x\frac{\partial}{\partial y}$$

$$\begin{aligned}
(x\frac{\partial}{\partial y} \circ y\frac{\partial}{\partial x})(f) &= x\frac{\partial}{\partial y}(y\frac{\partial}{\partial x}(f) \\
&= x(y\frac{\partial^2 f}{\partial y\partial x} + \frac{\partial f}{\partial x}) \\
&= (xy\frac{\partial^2}{\partial y\partial x} + x\frac{\partial}{\partial x})(f) \\
(y\frac{\partial}{\partial x} \circ x\frac{\partial}{\partial y})(f) &= y\frac{\partial}{\partial x}(x\frac{\partial}{\partial y}(f) \\
&= y(x\frac{\partial^2 f}{\partial x\partial y} + \frac{\partial f}{\partial y}) \\
&= (xy\frac{\partial^2}{\partial x\partial y} + y\frac{\partial}{\partial x})(f)
\end{aligned}$$

Therefore the commutator is:

$$\begin{aligned}
[x\frac{\partial}{\partial y}, y\frac{\partial}{\partial x}] &= x\frac{\partial}{\partial y} \circ y\frac{\partial}{\partial x} - y\frac{\partial}{\partial x} \circ x\frac{\partial}{\partial y} \\
&= (xy\frac{\partial^2}{\partial y\partial x} + x\frac{\partial}{\partial x}) - (xy\frac{\partial^2}{\partial x\partial y} + y\frac{\partial}{\partial x}) \\
&= x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}
\end{aligned}$$

Recall we defined the Commutators of two operators in the following manner: Let $\mathcal{O}_A$ and $\mathcal{O}_B$ be two operators. The commutator of $\mathcal{O}_A$ and $\mathcal{O}_B$ is the operator defined as

$$[\mathcal{O}_A, \mathcal{O}_B] = \mathcal{O}_A \circ \mathcal{O}_B - \mathcal{O}_B \circ \mathcal{O}_A.$$

Example. 1. Find the commutator of the following two operators.

$$\begin{aligned} \hat{E} &: \psi(x,t) \mapsto i\hbar\frac{\partial}{\partial t}\psi(x,t) \\ \hat{t} &: \psi(x,t) \mapsto t\psi(x,t) \end{aligned}$$

Solution:

$$\begin{aligned} [\hat{E},\hat{t}]\psi(x,t) &= (i\hbar\frac{\partial}{\partial t} \circ \hat{t} - \hat{t} \circ i\hbar\frac{\partial}{\partial t})(\psi(x,t)) \\ &= (i\hbar\frac{\partial}{\partial t} \circ \hat{t})(\psi(x,t)) - (\hat{t} \circ i\hbar\frac{\partial}{\partial t})(\psi(x,t)) \\ &= i\hbar\frac{\partial}{\partial t}(t\psi(x,t)) - ti\hbar\frac{\partial(\psi(x,t))}{\partial t} \\ &= i\hbar t\frac{\partial(\psi(x,t))}{\partial t} + i\hbar(\psi(x,t))\frac{\partial t}{\partial t} - ti\hbar\frac{\partial(\psi(x,t))}{\partial t} \\ &= i\hbar(\psi(x,t)) \end{aligned}$$

2. Find the commutator of the following two operators.

$$\begin{aligned} \hat{H} &: \psi(x,t) \mapsto -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) \\ \hat{x} &: \psi(x,t) \mapsto x\psi(x,t) \end{aligned}$$

Solution:

$$
\begin{aligned}
[\hat{H}, \hat{x}] &= \hat{H} \circ \hat{x} - \hat{x} \circ \hat{H} \\
(\hat{H} \circ \hat{x})(\psi(x,t)) &= -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}(\hat{x}(\psi)) \\
&= -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}(x(\psi)) \\
&= -\frac{\hbar^2}{2m}\frac{\partial}{\partial x}(\frac{\partial}{\partial x}(x\psi)) \\
&= -\frac{\hbar^2}{2m}\frac{\partial}{\partial x}(x\frac{\partial\psi}{\partial x} + \psi) \\
&= -\frac{\hbar^2}{2m}(\frac{\partial}{\partial x}(x\frac{\partial\psi}{\partial x}) + \frac{\partial\psi}{\partial x}) \\
&= -\frac{\hbar^2}{2m}(x\frac{\partial^2\psi}{\partial x^2} + 2\frac{\partial\psi}{\partial x}) \\
\hat{H} \circ \hat{x} &= -\frac{\hbar^2 x}{2m}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{m}\frac{\partial}{\partial x} \\
(\hat{x} \circ \hat{H})(\psi(x,t)) &= (\hat{x} \circ -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2})(\psi) \\
&= -\frac{\hbar^2 x}{2m}\frac{\partial^2\psi}{\partial x^2} \\
\hat{x} \circ \hat{H} &= -\frac{\hbar^2 x}{2m}\frac{\partial^2}{\partial x^2} \\
[\hat{H}, \hat{x}] &= -\frac{\hbar^2}{2m}x\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{m}\frac{\partial}{\partial x} + \frac{\hbar^2 x}{2m}\frac{\partial^2}{\partial x^2} \\
&= -\frac{\hbar^2}{m}\frac{\partial}{\partial x}
\end{aligned}
$$

3. Prove the following identity for three operators A,B, C:

$$[A \circ B, C] = A \circ [B, C] + [A, C] \circ B.$$

Solution:

$$
\begin{aligned}
A \circ [B, C] + [A, C] \circ B &= A \circ (B \circ C - C \circ B) + (A \circ C - C \circ A) \circ B \\
&= A \circ B \circ C - A \circ C \circ B + A \circ C \circ B - C \circ A \circ B \\
&= (A \circ B) \circ C - C \circ (A \circ B) \\
&= [A \circ B, C]
\end{aligned}
$$

4. Find the commutator $[D^2, e^x]$.

SOLUTION: We can find the commutator directly

$$[D^2, e^x] = D^2 \circ e^x - e^x \circ D^2$$

or by using the above identity we get,

$$[D^2, e^x] = D \circ [D, e^x] + [D, e^x] \circ D.$$

Try the direct method at home. In class we will use the identity to find the commutator.

We first find the commutator $[D, e^x]$.

$$\begin{aligned}
(D \circ e^x)(f) &= \frac{d}{dx}(e^x f) \\
&= e^x \frac{df}{dx} + f \frac{de^x}{dx} \\
&= e^x \frac{df}{dx} + f e^x \\
&= (e^x D + e^x)(f) \\
(e^x \circ D)(f) &= (e^x D)(f) \\
[D, e^x] &= (e^x D + e^x) - (e^x D) \\
&= e^x \\
[D^2, e^x] &= D \circ [D, e^x] + [D, e^x] \circ D \\
&= D \circ e^x + e^x \circ D \\
&= (e^x D + e^x) + e^x D \\
&= e^x (2D + 1)
\end{aligned}$$