Calculus 3 Full Notes

Isabelle

2023-2024 School Year

Contents

1 Vectors and the Geometry of Space

Chapter 12: Page 794

1.1 3D Coordinate Systems

Points in 3D space are represented as triples (a, b, c) of real numbers. a is the x position, b is the y position, and c is the z position.

We organize 3D space by choosing an origin O, and then drawing three axes for x, y, and z through it. x and y are typically drawn horizontally, and z is typically drawn vertically. The direction of the z axis is given by the right-hand rule. If you curl your fingers from the $+x$ direction to the $+y$ direction, your thumb will point in the direction of the $+z$ direction.

There are three **coordinate planes** that are defined by 3D space.

1. The xy plane is the set of all vectors with a zero z-coordinate.
2. The xz plane is the set of all vectors with a zero y-coordinate.
3. The yz plane is the set of all vectors with a zero x-coordinate.

These planes divide space into eight **octants**, with each octant being a unique combination of signs for the x, y and z coordinates. The first octant is the set of all vectors with $x, y, z > 0$.

If $P = (a, b, c)$ is a point in space, a represents the distance from the yz plane to P. b represents the distance from the xz plane to P, and c represents the distance from the xy plane to P. We can find the **projections** of P onto any of these planes by setting the respective coordinate equal to zero:

1. The projection of P onto the yz plane is given by $(0, b, c)$.
2. The projection of P onto the xz plane is given by $(a, 0, c)$.
3. The projection of P onto the xy plane is given by $(a, b, 0)$.
4. The projection of P onto the x axis is given by $(a, 0, 0)$.

The cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) | x, y, z \in \mathbb{R}\}$ is the set of all ordered triples. This is also denoted by $\mathbb{R}^3$. This is called the **three-dimensional rectangular coordinate system**, and has basis vectors

$$\begin{aligned} \hat{\boldsymbol{\imath}} &= \langle 1, 0, 0 \rangle \\ \hat{\boldsymbol{\jmath}} &= \langle 0, 1, 0 \rangle \\ \hat{\boldsymbol{k}} &= \langle 0, 0, 1 \rangle. \end{aligned}$$

An equation relating x, y, and z in $\mathbb{R}^3$ is called a *surface* in $\mathbb{R}^3$. Some examples of surfaces are:

1. $y = 3$: a plane given by the set $\{(a, 3, b) | a, b \in \mathbb{R}\}$.
2. $y = x$: a plane given by the set $\{(a, a, b) | a, b, \in \mathbb{R}\}$. This will look like the line $y = x$ in $\mathbb{R}^2$, except stretched upwards infinitely in the z direction.

The distance between two points in $\mathbb{R}^n$ given by $P_1 = (a_1, a_2, \ldots a_n)$ and $P_2 = (b_1, b_2, \ldots, b_n)$ is given by:

$$|P_1 P_2| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}.$$

This can be shown with $n - 1$ repeated applications of the Pythagorean theorem.

Example 1.1.1. Find the distance between $P(2, -1, 7)$ and $Q(1, -3, 5)$.

Solution:

$$\begin{aligned} |PQ| &= \sqrt{(1-2)^2 + (-3+1)^2 + (5-7)^2} \\ &= \sqrt{1 + 4 + 4} = 3 \end{aligned}$$

△

Example 1.1.2. Find an equation for a sphere with radius r and center $C(h, k, l)$.

Solution: A sphere is defined as a shape where all points that lie on it are the same distance r from its center. In other words, for any point $P(x, y, z)$ on the surface of the sphere,

$$\sqrt{(x-h)^2 + (y-k)^2 + (z-l)^2} = r$$

or,

$$(x-h)^2 + (y-k)^2 + (z-l)^2 = r^2$$

This result will come in handy often, and is worth remembering. $\triangle$

Example 1.1.3. Show that $x^2 + y^2 + z^2 + 4x - 6y + 2z + 6 = 0$ is the equation of a sphere. Find its center and radius.

Solution: First, let's rewrite our equation in a form that is more convenient to us:

$$(x^2 + 4x) + (y^2 - 6y) + (z^2 + 2z) = -6$$

We can complete the square to rewrite our equation:

$$[(x+2)^2 - 4] + [(y-3)^2 - 9] + [(z+1)^2 - 1] = -6.$$

Or,

$$(x+2)^2 + (y-3)^2 + (z+1)^2 = 8.$$

From this, we can see that the circle is centered at $C(-2, 3, 1)$ with radius $\sqrt{8}$. $\triangle$

Example 1.1.4. What region in $\mathbb{R}^3$ is represented by the following inequalities?

$$1 \leq x^2 + y^2 + z^2 \leq 4 \qquad z \leq 0$$

Soln: The first inequality necessitates that the points are within the sphere centered at O of radius $\sqrt{4} = 2$, but outside the sphere centered at O of radius $\sqrt{1} = 1$. The second inequality necessitates that the spheres be below or on the xy plane. Together, these describe a half-shell, with an inner radius of 1 and an outer radius of 2, that is entirely beneath or on the plane $z = 0$. $\triangle$

1.2 Vectors

uh something something magnitude direction. go read about it yourself

1.3 The Dot Product

Definition 1.3.1. If $\boldsymbol{a} = \langle a_1, a_2, \ldots, a_n \rangle$ and $\boldsymbol{b} = \langle b_1, b_2, \ldots, b_n \rangle$, then the dot product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

This gives us a useful way to "multiply" two vectors, leaving us with a scalar quantity.

Example 1.3.2. Find

$$\langle 2, 4 \rangle \cdot \langle 3, -1 \rangle,$$
$$\langle -1, 7, 4 \rangle \cdot \left\langle 6, 2, -\frac{1}{2}, \right\rangle$$
$$\text{and} \quad (\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} - 3\hat{\boldsymbol{k}}) \cdot (2\hat{\boldsymbol{\jmath}} - \hat{\boldsymbol{k}}).$$

Solution:

$$\begin{aligned}
\langle 2,4\rangle \cdot \langle 3,-1\rangle &= 2\cdot 3 + 4\cdot(-1) = 2 \\
\langle -1,7,4\rangle \cdot \left\langle 6,2,-\frac{1}{2}, \right\rangle &= -1\cdot 6 + 7\cdot 2 + 4\cdot -\frac{1}{2} \\
&= -6 + 14 - 2 = 6 \\
(\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} - 3\hat{\boldsymbol{k}}) \cdot (2\hat{\boldsymbol{\jmath}} - \hat{\boldsymbol{k}}) &= 1\cdot 0 + 2\cdot 2 + (-3)\cdot(-1) \\
&= 4 + 3 = 7
\end{aligned}$$

△

Theorem 1.3.3 (Properties of the Dot Product)**.** If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$, where V is a vector space, and $k \in \mathbb{R}$:

1. $\boldsymbol{a} \cdot \boldsymbol{a} = |\boldsymbol{a}|^2$
2. $\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}$
3. $\boldsymbol{a} \cdot \boldsymbol{0} = 0$
4. $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$
5. $(k\boldsymbol{a}) \cdot b = k(\boldsymbol{a} \cdot \boldsymbol{b}) = a \cdot (k\boldsymbol{b})$

These can all be proven quite easily:

Proof. Let $\boldsymbol{a} = \langle a_1, a_2, \ldots, a_n\rangle$, $\boldsymbol{b} = \langle b_1, b_2, \ldots, b_n\rangle$, $\boldsymbol{c} = \langle c_1, c_2, \ldots, c_n\rangle$, and let $k \in \mathbb{R}$. Then,

$$\begin{aligned}
\boldsymbol{a} \cdot \boldsymbol{a} &= a_1a_1 + a_2a_2 + \cdots + a_na_n = a_1^2 + a_2^2 + \cdots + a_n^2 \\
&= |\boldsymbol{a}|^2 \text{ (Property 1).} \\
\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) &= a_1(b_1 + c_1) + a_2(b_2 + c_2) + \cdots + a_n(b_n + c_n) \\
&= a_1b_1 + a_1c_1 + a_2b_2 + a_2c_2 + \cdots + a_nb_n + a_nc_n \\
&= \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c} \text{ (Property 2).} \\
\boldsymbol{a} \cdot \boldsymbol{0} &= 0a_1 + 0a_2 + \cdots + 0a_n \\
&= 0 \text{ (Property 3).} \\
\boldsymbol{a} \cdot \boldsymbol{b} &= a_1b_1 + a_2b_2 + \cdots + a_nb_n \\
&= b_1a_1 + b_2a_2 + \cdots + b_na_n \\
&= \boldsymbol{b} \cdot \boldsymbol{a} \text{ (Property 4).} \\
(k\boldsymbol{a}) \cdot \boldsymbol{b} &= (ka_1)b_1 + (ka_2)b_2 + \cdots + (ka_n)b_n \\
&= k(a_1b_1 + a_2b_2 + \cdots + a_nb_n) \\
&= k(\boldsymbol{a} \cdot \boldsymbol{b}) \\
(k\boldsymbol{a}) \cdot \boldsymbol{b} &= (ka_1)b_1 + (ka_2)b_2 + \cdots + (ka_n)b_n \\
&= a_1(kb_1) + a_2(kb_2) + \cdots + a_n(kb_n) \\
&= \boldsymbol{a} \cdot (k\boldsymbol{b}) \text{ (Property 5).}
\end{aligned}$$

□

There is also a geometric interpretation of the dot product,

Theorem 1.3.4. If $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ and θ is the angle between them, then

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}| \cos\theta$$

Proof. Create a triangle with vertices OAB. Define $\boldsymbol{a} = \overrightarrow{OA}$ and $\boldsymbol{b} = \overrightarrow{OB}$. Then, by the law of cosines,

$$\begin{aligned}|\overrightarrow{AB}|^2 &= |\overrightarrow{OA}|^2 + |\overrightarrow{OB}|^2 - 2|\overrightarrow{OA}|\,|\overrightarrow{OB}|\cos\theta \\ &= |\boldsymbol{a}|^2 + |\boldsymbol{b}|^2 - 2|\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta\end{aligned}$$

We can also note that $\overrightarrow{AB} = \boldsymbol{a} - \boldsymbol{b}$. Then,

$$\begin{aligned}|\boldsymbol{a} - \boldsymbol{b}|^2 &= |\boldsymbol{a}|^2 + |\boldsymbol{b}|^2 - 2|\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta \\ |\boldsymbol{a}|^2 - 2(\boldsymbol{a}\cdot\boldsymbol{b}) + |\boldsymbol{b}|^2 &= |\boldsymbol{a}|^2 + |\boldsymbol{b}|^2 - 2|\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta \\ -2(\boldsymbol{a}\cdot\boldsymbol{b}) &= -2|\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta \\ \boldsymbol{a}\cdot\boldsymbol{b} &= |\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta\end{aligned}$$

□

Using this, we can come to a quite significant conclusion:

Corollary 1.3.5. If θ is the angle between two nonzero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, then

$$\cos\theta = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|\,|\boldsymbol{b}|}$$

This can be used to find the angle between two vectors, given their components (or their magnitudes and dot products).

Example 1.3.6. Find the angle between $\boldsymbol{a} = \langle 2, 2, -1\rangle$ and $\boldsymbol{b} = \langle 5, -3, 2\rangle$.

Solution: First, determine that

$$|\boldsymbol{a}| = \sqrt{2^2 + 2^2 + (-1)^2} = 3$$

and

$$|\boldsymbol{b}| = \sqrt{5^2 + (-3^2) + 2^2} = \sqrt{38}.$$

Then, find that

$$\boldsymbol{a}\cdot\boldsymbol{b} = 2\cdot 5 - 3\cdot 2 - 2\cdot 1 = 2.$$

Then, calculate that

$$\theta = \cos^{-1}\left(\frac{2}{3\sqrt{38}}\right) \approx 1.46 \approx 0.46\pi$$

△

Theorem 1.3.4 also leads to an interesting conclusion about orthogonal vectors.

Theorem 1.3.7. Two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ are orthogonal if and only if

$$\boldsymbol{a}\cdot\boldsymbol{b} = 0.$$

Proof. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, and let θ be the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

By the definition of orthogonal, the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ must be $\theta = \pi/2$. Then, by the definition of the dot product,

$$\boldsymbol{a}\cdot\boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}|\cos\theta = |\boldsymbol{a}|\,|\boldsymbol{b}|\cos\frac{\pi}{2} = 0.$$

□

Further, since $\cos\theta > 0$ for $\theta \in (0, \pi/2)$, $\boldsymbol{a}\cdot\boldsymbol{b} > 0$ implies that the angle between them is acute. Similarly, since $\cos\theta < 0$ for $\theta \in (\pi/2, \pi)$, $\boldsymbol{a}\cdot\boldsymbol{b} < 0$ implies that the angle between them is obtuse.

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are pointing in the same direction, the angle between them is 0 and $\boldsymbol{a}\cdot\boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}|$.

Likewise, if $\boldsymbol{a}$ and $\boldsymbol{b}$ are pointing in opposite directions, the angle between them is π radians and $\boldsymbol{a}\cdot\boldsymbol{b} = -|\boldsymbol{a}|\,|\boldsymbol{b}|$.

1.3.1 Direction Angles and Direction Cosines

The *direction angles* of a nonzero vector $\boldsymbol{a} \in \mathbb{R}^3$ are the angles α, β, and γ that $\boldsymbol{a}$ makes with the positive x-, y-, and z- axes.

The cosines of these direction angles, $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are called the *direction cosines* of $\boldsymbol{a}$. By Corollary 1.3.5, we can obtain formulas for each of these:

$$\cos\alpha = \frac{\boldsymbol{a}\cdot\hat{\boldsymbol{\imath}}}{|\boldsymbol{a}|} = \frac{a_1}{|\boldsymbol{a}|}$$
$$\cos\beta = \frac{\boldsymbol{a}\cdot\hat{\boldsymbol{\jmath}}}{|\boldsymbol{a}|} = \frac{a_2}{|\boldsymbol{a}|}$$
$$\cos\gamma = \frac{\boldsymbol{a}\cdot\hat{\boldsymbol{k}}}{|\boldsymbol{a}|} = \frac{a_3}{|\boldsymbol{a}|}$$

Another consequence of this is that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = \frac{1}{|\boldsymbol{a}|^2}(a_1^2 + a_2^2 + a_3^2) = 1$$

Example 1.3.8. Find the direction angles of $\boldsymbol{a} = \langle 1, 2, 3\rangle$.

Solution: First, find

$$|\boldsymbol{a}| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

then,

$$\alpha = \cos^{-1}\frac{1}{\sqrt{14}} \quad \beta = \cos^{-1}\frac{2}{\sqrt{14}} \quad \gamma = \cos^{-1}\frac{3}{\sqrt{14}}$$

△

1.4 Projections

With dot products, we can *project* two vectors onto each other. Consider two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. We can denote the *vector projection* of $\boldsymbol{b}$ onto $\boldsymbol{a}$ with $\operatorname{proj}_{\boldsymbol{a}} \boldsymbol{b}$.

Similarly, we can define the *scalar projection* of $\boldsymbol{b}$ onto $\boldsymbol{a}$ (also called the *component of* $\boldsymbol{b}$ *along* $\boldsymbol{a}$) as the magnitude of the vector projection, or:

$$\operatorname{comp}_{\boldsymbol{a}} \boldsymbol{b} = |\operatorname{proj}_{\boldsymbol{a}} \boldsymbol{b}|$$

This can also be found with:

$$\operatorname{comp}_{\boldsymbol{a}} \boldsymbol{b} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|} = |\boldsymbol{b}|\cos\theta$$

Alternatively,

$$\operatorname{comp}_{\boldsymbol{a}} \boldsymbol{b} = \frac{\boldsymbol{a}}{|\boldsymbol{a}|}\cdot\boldsymbol{b}$$

Then, the vector projection is:

$$\begin{aligned}\operatorname{proj}_{\boldsymbol{a}} \boldsymbol{b} &= \operatorname{comp}_{\boldsymbol{a}}(\boldsymbol{b})\frac{\boldsymbol{a}}{|\boldsymbol{a}|}\\ &= \left(\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|}\right)\frac{\boldsymbol{a}}{|\boldsymbol{a}|}\\ &= \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|^2}\boldsymbol{a}\end{aligned}$$

Example 1.4.1. Find the scalar and vector projection of $\boldsymbol{b} = \langle 1, 1, 2\rangle$ onto $\boldsymbol{a} = \langle -2, 3, 1\rangle$.

Solution: First,

$$\text{comp}_{\boldsymbol{a}}\,\boldsymbol{b} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|} = \frac{1\cdot(-2)+1\cdot 3+2\cdot 1}{\sqrt{(-2)^2+3^2+1^2}} = \frac{3}{\sqrt{14}}$$

△

Then,

$$\text{proj}_{\boldsymbol{a}}\,\boldsymbol{b} = \hat{a}\,\text{comp}_{\boldsymbol{a}}\,\boldsymbol{b} = \frac{3}{\sqrt{14}}\left\langle \frac{-2}{\sqrt{14}}, \frac{3}{\sqrt{14}}, \frac{1}{\sqrt{14}}\right\rangle = \left\langle -\frac{3}{7}, \frac{9}{14}, \frac{3}{14}\right\rangle$$

1.5 The Cross Product

The cross product is a special operation defined only for three-dimensional vectors. Specifically,

Definition 1.5.1. If $\boldsymbol{a} = \langle a_1, a_2, a_3\rangle$ and $\boldsymbol{b} = \langle b_1, b_2, b_3\rangle$, then the cross product between $\boldsymbol{a}$ and $\boldsymbol{b}$ is:

$$\boldsymbol{a}\times\boldsymbol{b} = \langle a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1\rangle$$

This is easily remembered as the determinant of the matrix:

$$\boldsymbol{a}\times\boldsymbol{b} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Example 1.5.2. Show that $\boldsymbol{a}\times\boldsymbol{a} = \boldsymbol{0}$ for any $\boldsymbol{a}\in\mathbb{R}^3$.

Solution: Let $\boldsymbol{a} = \langle a_1, a_2, a_3\rangle$. Then,

$$\boldsymbol{a}\times\boldsymbol{a} = \langle a_2a_3 - a_3a_2, a_3a_1 - a_1a_3, a_1a_2 - a_2a_1\rangle = \langle 0, 0, 0\rangle = \boldsymbol{0}$$

△

Another critical property of the cross product,

Theorem 1.5.3. For any $\boldsymbol{a}, \boldsymbol{b}\in\mathbb{R}^3$, $\boldsymbol{a}\times\boldsymbol{b}$ is orthogonal to both $\boldsymbol{a}$ and $\boldsymbol{b}$.

Proof. First, $\boldsymbol{a}\times\boldsymbol{b}$ is orthogonal to $\boldsymbol{a}$:

$$\begin{aligned}(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{a} &= a_1(a_2b_3 - a_3b_2) - a_2(a_1b_3 - a_3b_1) + a_3(a_1b_2 - a_2b_1)\\ &= a_1a_2b_3 - a_1a_3b_2 - a_1a_2b_3 + a_2a_3b_1 + a_1a_3b_2 - a_2a_3b_1\\ &= (a_1a_2b_3 - a_1a_2b_3) + (a_1a_3b_2 - a_1a_3b_2) + (a_2a_3b_1 - a_2a_3b_1)\\ &= 0\end{aligned}$$

Then, $\boldsymbol{a}\times\boldsymbol{b}$ is orthogonal to $\boldsymbol{b}$:

$$\begin{aligned}(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{b} &= b_1(a_2b_3 - a_3b_2) - b_2(a_1b_3 - a_3b_1) + b_3(a_1b_2 - a_2b_1)\\ &= a_2b_1b_3 - a_3a_1b_2 - a_1b_2b_3 + a_3b_1b_2 + a_1b_2b_3 - a_2b_1b_3\\ &= (a_2b_1b_3 - a_2b_1b_3) + (a_3a_1b_2 - a_3a_1b_2) + (a_1b_2b_3 - a_1b_2b_3)\\ &= 0\end{aligned}$$

□

This means that if $\boldsymbol{a}$ and $\boldsymbol{b}$ are represented as directed line segments coming from the same initial point, then $\boldsymbol{a}\times\boldsymbol{b}$ point in a direction perpendicular to the plane through $\boldsymbol{a}$ and $\boldsymbol{b}$. The direction of this is determined by the *right-hand rule*: If the fingers of your right hand curl in the direction of rotation from $\boldsymbol{a}$ to $\boldsymbol{b}$, then your thumb points in the direction of $\boldsymbol{a}\times\boldsymbol{b}$.

Another important property of the cross product:

Theorem 1.5.4. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ be nonzero vectors, and let θ be the angle between them. Then,

$$|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}|\,|\boldsymbol{b}|\,\sin\theta$$

Proof.

$$\begin{aligned}
|\boldsymbol{a} \times \boldsymbol{b}|^2 &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\
&= a_2^2b_3^2 - 2a_2b_3a_3b_2 + a_3^2b_2^2 + a_3^2b_1^2 - 2a_3b_1a_1b_3 + a_1^2b_3^2 \\
&+ a_1^2b_2^2 - 2a_1b_2a_2b_1 + a_2^2b_1^2 \\
&= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\
&= |\boldsymbol{a}|^2|\boldsymbol{b}|^2 - (\boldsymbol{a} \cdot \boldsymbol{b})^2 \\
&= |\boldsymbol{a}|^2|\boldsymbol{b}|^2 - |\boldsymbol{a}|^2|\boldsymbol{b}|^2\cos^2\theta \\
&= |\boldsymbol{a}|^2|\boldsymbol{b}|^2(1 - \cos^2\theta) \\
&= |\boldsymbol{a}|^2|\boldsymbol{b}|^2\sin^2\theta \\
|\boldsymbol{a} \times \boldsymbol{b}| &= |\boldsymbol{a}|\,|\boldsymbol{b}|\sin\theta
\end{aligned}$$

□

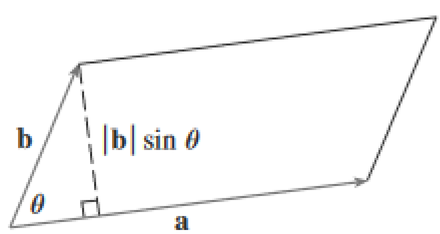
Additionally,

Corollary 1.5.5. Two nonzero vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ are parallel or antiparallel if and only if

$$\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0}$$

This easily follows from the magnitude definition of the cross product, where $\sin\theta = 0$ if $\theta = 0$ or $\theta = \pi$.

There is also a geometric interpretation of the cross product: if $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$, then $|\boldsymbol{a} \times \boldsymbol{b}|$ is the area of the parallelogram determined by $\boldsymbol{a}$ and $\boldsymbol{b}$ (see the below image).



Example 1.5.6. Find a vector perpendicular to the plane that passes through the points $P(1, 4, 6)$, $Q(-2, 5, -1)$, and $R(1, -1, 1)$.

Solution: First, consider the vectors $\overrightarrow{PQ} = \langle -3, 1, -7 \rangle$ and $\overrightarrow{PR} = \langle 0, -5, -5 \rangle$. Then, if $\hat{N}$ is the unit normal vector to the plane,

$$\hat{N} = \frac{\overrightarrow{PQ} \times \overrightarrow{PR}}{|\overrightarrow{PQ} \times \overrightarrow{PR}|} = \frac{\langle -40, -15, 15 \rangle}{\sqrt{40^2 + 15^2 + 15^2}} = \frac{1}{\sqrt{82}} \langle -8, -3, 3 \rangle$$

Then, for any $a \in \mathbb{R}$, $a\hat{N}$ is perpendicular to the plane. △

Example 1.5.7. Find the area of the triangle with vertices P, Q, R (from the previous example).

Solution: We've already determined that

$$|\overrightarrow{PQ} \times \overrightarrow{PR}| = 5\sqrt{82}$$

in the previous part. We can interpret this as the area of the parallelogram formed by $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. The triangle will have half the area of the parallelogram, or $A = \frac{5}{2}\sqrt{82}$. △

Some other important properties of the cross product,

Theorem 1.5.8. If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$ and $x \in \mathbb{R}$,

1. $\boldsymbol{a} \times \boldsymbol{b} = -(\boldsymbol{b} \times \boldsymbol{a})$.
2. $(x\boldsymbol{a}) \times \boldsymbol{b} = x(\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{a} \times (x\boldsymbol{b})$.
3. $\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c}$.
4. $(\boldsymbol{a} + \boldsymbol{b}) \times \boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{c} + \boldsymbol{b} \times \boldsymbol{c}$.
5. $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$.
6. $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{c})\boldsymbol{b} - (\boldsymbol{a} \cdot \boldsymbol{b})\boldsymbol{c}$.

These properties can be shown relatively easily (although it is a bit of annoying algebra).

One of these properties, Property 5, is especially important. It is known as the **scalar triple product** of $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. If $\boldsymbol{a} = \langle a_1, a_2, a_3 \rangle$, and $\boldsymbol{b} \times \boldsymbol{c} = \langle d_1, d_2, d_3 \rangle$, we can rewrite:

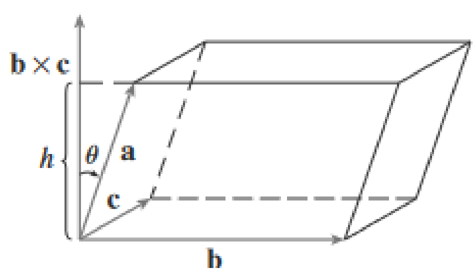
$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = a_1(d_1) + a_2(d_2) + a_3(d_3)$$

Which is also equal to the determinant

$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix},$$

since $d_1 = \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix}$, $d_2 = \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix}$, and $d_3 = \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}$.

There is also a geometric interpretation of the scalar triple product. Consider the parallelepiped formed by $\boldsymbol{a}, \boldsymbol{b}$, and $\boldsymbol{c}$ (below). The area of the base of the shape is given by $A = |\boldsymbol{b} \times \boldsymbol{c}|$. If θ is the angle between $\boldsymbol{a}$



and $\boldsymbol{b} \times \boldsymbol{c}$, then the height is given by $h = |\boldsymbol{a}|\,|\cos\theta|$ (note that we must use $|\cos\theta|$ instead of $\cos\theta$ in case $\theta > \pi/2$). Then, the total volume,

$$V = Ah = |\boldsymbol{a}|\,|\cos\theta|\,|\boldsymbol{b} \times \boldsymbol{c}| = |\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})|$$

which we can recognize as the scalar triple product between $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.

Another application of this formula is when the volume of the paralellepiped is 0, which we can recognize happens when $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ exist in the same plane. This is referred to as them being **coplanar**.

Example 1.5.9. Use the scalar triple product to show that the vectors $\boldsymbol{a} = \langle 1, 4, -7 \rangle$, $\boldsymbol{b} = \langle 2, -1, 4 \rangle$, and $\boldsymbol{c} = \langle 0, -9, 18 \rangle$ are coplanar.

Solution:

$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \begin{vmatrix} 1 & 4 & -7 \\ 2 & -1 & 4 \\ 0 & -9 & 18 \end{vmatrix}$$

which we will compute with a cofactor expansion along the third row.

$$\begin{vmatrix} 1 & 4 & -7 \\ 2 & -1 & 4 \\ 0 & -9 & 18 \end{vmatrix} = 0 \begin{vmatrix} 4 & -7 \\ -1 & 4 \end{vmatrix} - (-9) \begin{vmatrix} 1 & -7 \\ 2 & 4 \end{vmatrix} + 18 \begin{vmatrix} 1 & 4 \\ 2 & -1 \end{vmatrix}$$
$$= 9(4+14) + 18(-1-8) = 9(18) - 9(18) = 0$$

△

Cross products come up often in physics. One simple example is torque. Consider a force $\boldsymbol{F}$ acting on a rigid body that is free to rotate. The force is applied at a displacement of $\boldsymbol{r}$ relative to the axis of rotation of the object. Then, the torque $\boldsymbol{\tau}$ of the object relative to the axis of rotation is

$$\boldsymbol{\tau} = \boldsymbol{r} \times \boldsymbol{F}$$

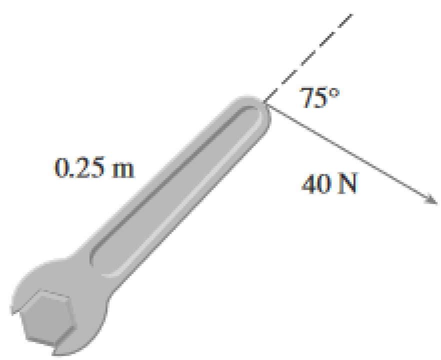
The magnitude of the torque represents how much "rotational force" is applied. The direction of the torque is the axis of rotation, and determines which way the object will rotate.

The magnitude of the torque is given by

$$|\boldsymbol{\tau}| = |\boldsymbol{r} \times \boldsymbol{F}| = |\boldsymbol{r}|\,|\boldsymbol{F}|\,\sin\theta$$

if θ is the angle between $\boldsymbol{r}$ and $\boldsymbol{F}$. This formula can be understood to say that only the component of $\boldsymbol{F}$ *perpendicular* to $\boldsymbol{r}$ can cause rotation, which makes sense if you consider the physical implications.

Example 1.5.10. A bolt is tightened by applying a 40-N force to a 0.25-m wrench at an angle of 75° as shown below. Find the magnitude of the torque about the center of the bolt.



Solution: This is as simple as plugging in:

$$|\boldsymbol{\tau}| = |\boldsymbol{r}|\,|\boldsymbol{F}|\,\sin\theta = (0.25)(40)\sin 75°$$
$$\approx 9.66\,\mathrm{Nm}$$

△

1.6 Equations of Lines

A line in the xy-plane is determined by a point and a direction (slope). Likewise, a line in $\mathbb{R}^3$ is determined by a point and a direction.

Consider a line L. Given that $P_0(x_0, y_0, z_0)$ is a known point on the line and that $\boldsymbol{v}$ is a vector in the direction of L, let $P(x, y, z)$ be an arbitrary point on L and let $\boldsymbol{r}$ and $\boldsymbol{r}_0$ be the position vectors of P and P_0 respectively. If $\boldsymbol{a} = \overrightarrow{P_0P}$, then we know that $\boldsymbol{r} = \boldsymbol{r}_0 + \boldsymbol{a}$ by the triangle law. Because we know that $\boldsymbol{r}$ and $\boldsymbol{a}$

are parallel (due to the fact that $\boldsymbol{r}$ is offset from $\boldsymbol{a}$ by a constant vector), then there must exist some scalar t such that $\boldsymbol{a} = t\boldsymbol{v}$. Therefore, we can rewrite our equation:

$$\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{v}$$

Now, we have parameterized the equation for our line in terms of an independent variable t and a starting point $\boldsymbol{r}_0$. This is the n-dimensional equivalent to the point slope form. If $\boldsymbol{v} = \langle a, b, c \rangle$, then we can rewrite this again:

$$\langle x, y, z \rangle = \langle x_0 + ta, y_0 + tb, z_0 + tc \rangle$$

The set of equations $x = x_0 + ta$, $y = y_0 + tb$, and $z = z_0 + tc$ are called the **parametric equations** of L through P_0 and parallel to $\boldsymbol{v}$. Every value of $t \in \mathbb{R}$ gives a point on L, and every point on L has a corresponding t-value.

An interesting sidenote (this can be safely ignored if you don't understand it), this equation forms a bijection between L and $\mathbb{R}$, so the cardinality of a line is the same as the cardinality of the real numbers. Similar arguments can be used to show that any nonzero subspace $V \subseteq \mathbb{R}^n$ has cardinality

$$|V| = |\mathbb{R}^m| = 2^{\aleph_0}$$

for arbitrary $m \in \mathbb{N}$.

Example 1.6.1. Find a vector equation and parametric equations for the line that passes through $P_0(5, 1, 3)$ and is parallel to the vector $\boldsymbol{v} = \hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}} - 2\hat{\boldsymbol{k}}$. If we double the value of t, what happens?

Solution: This is pretty simple:

$$\langle x, y, z \rangle = \overrightarrow{OP_0} + t\boldsymbol{a} = \langle 5 + t, 1 + 4t, 3 - 2t \rangle .$$

For part B, doubling t has no effect on the line. If $\boldsymbol{f}(t)$ gives the point on the line given by t and $\boldsymbol{g}(t)$ gives the point on the line given by $2t$, $\boldsymbol{f}$ and $\boldsymbol{g}$ are easily observed to be bijective with the correspondence $\boldsymbol{f}(2t) = \boldsymbol{g}(t)$. △

In general, if $\boldsymbol{v} = \langle a, b, c \rangle$ is used to describe the direction of a line L, then a, b, and c are called **direction numbers** of L. Since scalar multiples of $\boldsymbol{v}$ are along the same direction as $\boldsymbol{v}$, we can use any $c\boldsymbol{v}$, $c \in \mathbb{R}$ to describe the direction of L.

Another way to describe a line L is to solve for t, giving

$$t = \frac{x - x_0}{a}$$
$$t = \frac{y - y_0}{b}$$
$$t = \frac{z - z_0}{c}$$

These can be equated to each other to get an expression describing L that is independent of t:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{z}$$

Equations of this form are called **symmetric equations** for L. Notice that these are only valid if $a, b, c \neq 0$. If one or more direction numbers are equal to zero we can instead say

$$x = x_0, \qquad \frac{y - y_0}{b} = \frac{z - z_0}{z}$$

or similar if b or c is zero.

Example 1.6.2. Find symmetric equations to the line passing through $A(2, 4, -3)$ and $B(3, -1, 1)$. At what point does this line intersect the xy-plane?

Solution: First, the direction of the line can be described with the vector

$$\overrightarrow{AB} = \langle 1, -5, -2 \rangle .$$

Then, a symmetric equation can be found,

$$\frac{x-2}{1} = \frac{y-4}{-5} = \frac{z+3}{4}$$

To find an intersection point with the xy-plane, set $z = 0$ and solve

$$x - 2 = \frac{4-y}{5} = \frac{3}{4}$$

to find that

$$\langle x, y, 0 \rangle = \left\langle \frac{11}{4}, \frac{1}{4}, 0 \right\rangle$$

△

In general, the direction numbers of the line L through the points $P_0(x_0, y_0, z_0)$ and $P_1(x_1, y_1, z_1)$ are given by

$$\begin{aligned} a &= x_1 - x_0 \\ b &= y_1 - y_0 \\ c &= z_1 - z_0 \end{aligned}$$

Often, we will want a description of a line *segment* instead of the whole line. For instance, how could we describe the line segment AB in Example 1.6.2? If we put $t = 0$ into the parametric equations for L, we get $\overrightarrow{OA}$. Similarly, if we put $t = 1$ into these equations, we get $\overrightarrow{OB}$. Thus, we can describe the segment AB with the parametric equations

$$x = 2 + t, \quad y = 4 - 5t, \quad z = -3 + 4t, \quad 0 \le t \le 1$$

or the corresponding vector equation

$$\boldsymbol{r}(t) = \langle 2 + t, 4 - 5t, -3 + 4t \rangle \quad 0 \le t \le 1$$

In general, consider a line L passing through two arbitrary nonequal points P_0 and P_1. If $\boldsymbol{r}_0 = \overrightarrow{OP_0}$ and $\boldsymbol{r}_1 = \overrightarrow{OP_1}$, The vector $\overrightarrow{P_1P_0} = \boldsymbol{r}_1 - \boldsymbol{r}_0$ encodes the direction numbers for L. However, these direction numbers are special, since they also include the exact values required to go from P_0 to P_1. Therefore, the line equation

$$\boldsymbol{r} = \boldsymbol{r}_0 + (\boldsymbol{r}_1 - \boldsymbol{r}_0)t, \quad 0 \le t \le 1$$

describes the segment P_0P_1 of L. We can rewrite this as

$$\boldsymbol{r} = \boldsymbol{r}_0(1 - t) + \boldsymbol{r}_1 t, \quad 0 \le t \le 1.$$

Example 1.6.3. Show that the lines L_1 and L_2 represented by

$$\begin{aligned} \boldsymbol{r}_1 &= \langle 1 + t, -2 + 3t, 4 - t \rangle \\ \boldsymbol{r}_2 &= \langle 2s, 3 + s, -3 + 4s \rangle \end{aligned}$$

Are **skew lines**–that is, they do not intersect and are not parallel (therefore, they do not lie in the same plane).

Solution: First, let $\boldsymbol{v}_1$ be parallel to L_1 and $\boldsymbol{v}_2$ be parallel to L_2. Then,

$$\boldsymbol{v}_1 = \langle 1, 3, -1 \rangle , \quad \boldsymbol{v}_2 = \langle 2, 1, 4 \rangle .$$

Because $\boldsymbol{v}_1$ is not a scalar multiple of $\boldsymbol{v}_2$, they (and their corresponding lines) are not parallel.

If L_1 intersects with L_2, then there must exist some t and s such that

$$1 + t = 2s \quad -2 + 3t = 3 + s \quad 4 - t = -3 + 4s$$

We can rearrange this to a matrix equation:

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

Augmenting the coefficient matrix with the output vector and row reducing,

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

So the system is inconsistent and there is no solution. △

1.7 Equations of Planes

A plane in space, similar to a line, can be defined with just one point and one vector.

Consider some plane in space containing the point $P_0(x_0, y_0, z_0)$ with a normal vector $\boldsymbol{n} \in \mathbb{R}^3$. Then, if we define $\overrightarrow{OP_0} = \boldsymbol{r}_0$, and define $\boldsymbol{r}$ as an arbitrary point on the plane, the vector $\boldsymbol{r}_0 - \boldsymbol{r}$ will be perpendicular to the normal vector. That is,

$$\boldsymbol{n} \cdot (\boldsymbol{r} - \boldsymbol{r}_0) = 0$$

Which can be rewritten as

$$\boldsymbol{n} \cdot \boldsymbol{r} = \boldsymbol{n} \cdot \boldsymbol{r}_0$$

If $\boldsymbol{r} = \langle x, y, z \rangle$ and $\boldsymbol{n} = \langle a, b, c \rangle$, we can rewrite this equation:

$$\langle a, b, c \rangle \cdot \langle x - x_0, y - y_0, z - z_0 \rangle = 0$$

or

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

This is called the **scalar equation of the plane** through P_0 with normal vector $\boldsymbol{n}$.

Example 1.7.1. Find the equation of the plane through the point $(2, 4, -1)$ with normal vector $\boldsymbol{n} = \langle 2, 3, 4 \rangle$. Find the intercepts.

Solution: Plugging into the scalar equation of the plane formula,

$$2(x - 2) + 3(y - 4) + 4(z + 1) = 0$$

Then, simplifying,

$$2x + 3y + 4z = 12$$

We can find the x-intercept by setting $y = z = 0$. Similar processes work for the y and z intercepts. Thus, $x^* = 6$, $y^* = 4$, and $z^* = 3$. △

As we saw in the previous example, we can rewrite the equation of the plane as

$$ax + by + cz + d = 0$$

where $d = -(ax_0 + by_0 + cz_0)$. This is called a **linear equation** in x, y, and z. If a, b, c are all nonzero, then this represents the plane with normal vector $\langle a, b, c \rangle$.

Example 1.7.2. Find an equation of the plane that passes through the points $P(1,3,2)$, $Q(3,-1,6)$, and $R(5,2,0)$.

Solution: First, compute $\boldsymbol{r}_1 = \overrightarrow{PQ}$ and $\boldsymbol{r}_2 = \overrightarrow{PR}$.

$$\begin{aligned} \boldsymbol{r}_1 &= \langle 3-1, -1-3, 6-2 \rangle = \langle 2, -4, 4 \rangle \\ \boldsymbol{r}_2 &= \langle 5-1, 2-3, 0-2 \rangle = \langle 4, -1, -2 \rangle \end{aligned}$$

Then, we know that the normal vector $\boldsymbol{n}$ will be perpendicular to both $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. That is,

$$\begin{aligned} \boldsymbol{n} &= \boldsymbol{r}_1 \times \boldsymbol{r}_2 \\ &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 2 & -4 & 4 \\ 4 & -1 & -2 \end{vmatrix} = 12\hat{\boldsymbol{\imath}} + 20\hat{\boldsymbol{\jmath}} + 14\hat{\boldsymbol{k}} \end{aligned}$$

Since we can use any scalar multiple, we will multiply this by a factor of 0.5 to simplify calculations. Thus, $\boldsymbol{n} = \langle 6, 10, 7 \rangle$. Now, our equation can be written.

$$6(x-1) + 10(y-3) + 7(z-2) = 0$$

or

$$6x + 10y + 7z = 50$$

△

Example 1.7.3. Find the point at which the line with parametric equations $x = 2 + 3t$, $y = -4t$, $z = 5 + t$ intersects the plane $4x + 5y - 2z = 18$.

Solution: Plugging each parametric equation into the plane equation, we get

$$\begin{aligned} 18 &= 4(2+3t) + 5(-4t) - 2(5+t) \\ &= 8 + 12t - 20t - 10 - 2t \\ 20 &= -10t \\ t &= -2 \end{aligned}$$

Plugging this back into our parametric equations,

$$\langle x, y, z \rangle = \langle -4, 8, 3 \rangle$$

△

Example 1.7.4. Two parts:

1. Find the angle between the planes $x + y + z = 1$ and $x - 2y + 3z = 1$.
2. Find symmetric equations for the line of intersection L between these two planes.

Solution: The angle between the planes is equal to the angle between their normal vectors. The normal vector for the first plane is $\boldsymbol{n}_1 = \langle 1, 1, 1 \rangle$ and the second is $\boldsymbol{n}_2 = \langle 1, -2, 3 \rangle$. Then,

$$\begin{aligned} \cos\theta &= \frac{\boldsymbol{n}_1 \cdot \boldsymbol{n}_2}{|\boldsymbol{n}_1|\,|\boldsymbol{n}_2|} = \frac{2}{\sqrt{3} \cdot \sqrt{14}} = \frac{2}{\sqrt{42}} \\ \theta &= \arccos\frac{2}{\sqrt{42}} \approx 72^\circ \end{aligned}$$

Now, the line of intersection is perpendicular to both normal vectors, so its direction $\boldsymbol{r}$ is given by

$$\boldsymbol{n}_1 \times \boldsymbol{n}_2 = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 1 & 1 & 1 \\ 1 & -2 & 3 \end{vmatrix} = \langle 5, -2, -3 \rangle$$

Then, a point of intersection between the two lines can be found. For instance, take $sz = 0$. So the point that satisfies both equations can be found:

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

so $x = 1$ and $y = 0$. Then, the symmetric equation for the plane can be written:

$$\frac{x-1}{5} + \frac{y}{-2} + \frac{z}{-3} = 0$$

△

Example 1.7.5. Find a formula for the distance D from a point $P_1(x_1, y_1, z_1)$ to the plane $ax+by+cz+d = 0$.

Solution: Let $P_0(x_0, y_0, z_0)$ be an arbitrary point in the plane. Then, let $\boldsymbol{b} = \overrightarrow{P_0P_1} = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0\rangle$. We can note that the distance between P_1 and the plane is the scalar projection of $\boldsymbol{b}$ onto $\boldsymbol{n} = \langle a, b, c\rangle$. Therefore,

$$\begin{aligned} D = |\text{comp}_{\boldsymbol{n}}\, \boldsymbol{b}| &= \left|\frac{\boldsymbol{n} \cdot \boldsymbol{b}}{|\boldsymbol{n}|}\right| \\ &= \left|\frac{a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)}{\sqrt{a^2 + b^2 + c^2}}\right| \\ &= \frac{|ax_1 + by_1 + cz_1 - (ax_0 + by_0 + cz_0)}{\sqrt{a^2 + b^2 + c^2}} \end{aligned}$$

Note that from the plane equation, $ax_0 + by_0 + cz_0 = -d$. Then,

$$D = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

△

Example 1.7.6. Find the distance D between the parallel planes $a_0x + b_0y + c_0z + d_0 = 0$ and $a_1x + b_1y - c_1z + d_1 = 0$.

Solution: Choose some point $P_0(x_0, y_0, z_0)$ on the first plane and some other point $P_1(x_1, y_1, z_1)$ on the second plane. Then, the distance between the planes is the projection of the vector $\boldsymbol{d} = \overrightarrow{P_0P_1}$ onto the normal vector of either plane. If we choose $\boldsymbol{n} = \langle a_1, b_1, c_1\rangle$ to be normal to both planes, then

$$\begin{aligned} D = |\,\text{comp}_{\boldsymbol{n}}\, \boldsymbol{d}| &= \frac{|\boldsymbol{n} \cdot \boldsymbol{d}|}{|\boldsymbol{n}|} \\ &= \frac{|a_1(x_1 - x_0) + b_1(y_1 - y_0) + c_1(z_1 - z_0)}{\sqrt{a_1^2 + b_1^2 + c_1^2}} \\ &= \frac{|a_1x_1 + b_1y_1 + c_1z_1 - (a_1x_0 + b_1y_0 + c_1z_0)|}{\sqrt{a_1^2 + b_1^2 + c_1^2}} \end{aligned}$$

because these planes are parallel, then the vector $\boldsymbol{n}_2 = \langle a_0, b_0, c_0\rangle$ is a scalar multiple of $\boldsymbol{n}_1$. We can thus rewrite our first equation as

$$a_1x + b_1y + c_1z + \frac{a_1}{a_0}d_0 = 0$$

Then,

$$\begin{aligned} D &= \frac{|d_1 - \frac{a_1d_0}{a_0}|}{\sqrt{a_1^2 + b_1^2 + c_1^2}} \\ &= \frac{|a_0d_1 - a_1d_0|}{|a_0|\sqrt{a_1^2 + b_1^2 + c_1^2}} \end{aligned}$$

△

1.8 Cylinders and Quadric Surfaces

Two other types of surfaces in $\mathbb{R}^3$ are cylinders and quadric surfaces. These allow us to plot more complex structures in three dimensions.

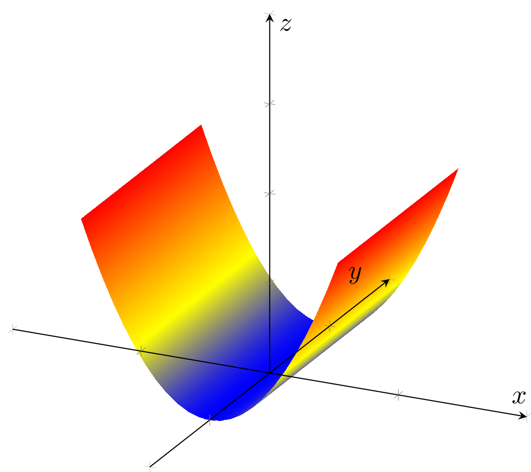
The most important problem-solving method for understanding the graphs of complex functions in $\mathbb{R}^3$ is the idea of **traces**. Traces are cross-sections taken by fixing one variable, so we can take a look at the structure through several functions in $\mathbb{R}^2$ projected onto a given plane.

1.8.1 Cylinders

A **cylinder** is a surface that consists of all lines (called **rulings**) that are parallel to a given line and pass through a given plane curve.

Example 1.8.1. Sketch the graph of the surface $z = x^2$.

Solution: Notice that the equation for this surface is independent of y. That means that any vertical plane given by $y = k$ (parallel to the xz-plane) intersects the surface with the equation $z = x^2$. So the vertical traces of this surface are parabolas. This type of surface is known as a **parabolic cylinder**. △



Example 1.8.2. Identify the surfaces $x^2 + y^2 = 1$ and $y^2 + z^2 = 1$.

Solution: For the first surface, there is no z dependence. So we can look at z traces and recognize them as the equations of circles with radius one in the xy plane. The graph will be a cylinder with radius 1 running parallel to the z axis.

For the second surface, there is no x dependence. So we can look at x traces and recognize them as equations of circles with radius one in the yz plane. The graph will be a cylinder with radius one running parallel to the x axis. △

1.8.2 Quadric Surfaces

A **quadric surface** is the graph of a second degree equation in terms of x, y, and z. The general form of a quadric surface is

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0$$

where $A, B, C, \ldots, J$ are constants. These surfaces can be somewhat difficult to analyze, but it is doable using traces.

Example 1.8.3. Use traces to describe the quadric equation

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

Solution: First, setting $x = k$, we get $\frac{y^2}{b} + \frac{z^2}{c} = 1 - \frac{k^2}{a}$. From this, we can see that the equation is only valid when $x \in [-\sqrt{a}, \sqrt{a}]$, and it will sketch ovals in the yz plane.

Setting $y = k$, we get $\frac{x^2}{a} + \frac{z^2}{c} = 1 - \frac{k^2}{b}$. This is another oval, and we can see that $y \in [-\sqrt{b}, \sqrt{b}]$.

Finally, setting $z = k$, we get $\frac{x^2}{a} + \frac{y^2}{b} = 1 - \frac{k^2}{c}$. This is another oval, with $z \in [-\sqrt{c}, \sqrt{c}]$.

Combining all of these, we can see that the overall curve will be an egg shape centered at the origin with a radius of $\sqrt{b}$ in the y direction, a radius of $\sqrt{a}$ in the x direction, and a radius of $\sqrt{c}$ in the z direction. △

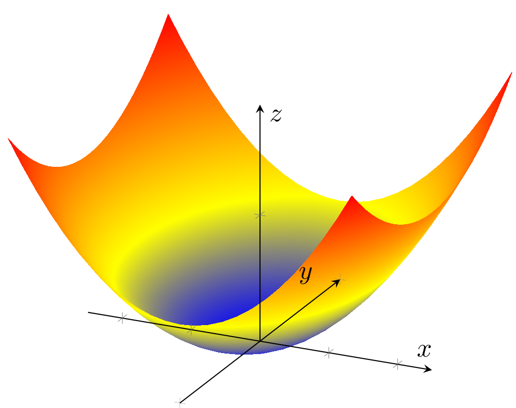
Example 1.8.4. Use traces to describe the surface $ax^2 + by^2 = cz$.

Solution: First, setting $z = k$, we get $ax^2 + by^2 = ck$, which is an ellipse in the xy plane of x-radius $\sqrt{cka^{-1}}$ and y-radius $\sqrt{ckb^{-1}}$.

Setting $x = k$, we get $by^2 - cz = ak^2$, which describes a parabola in the yz-plane, opening in the z direction.

Setting $y = k$, we get $ax^2 - cz = bk^2$, which describes a parabola in the xz-plane, opening in the z direction.

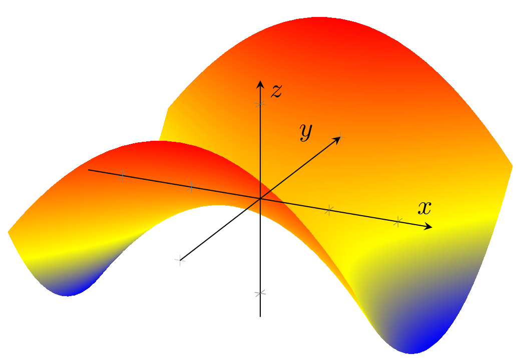
Overall, these three traces describe a 3D parabolic shape opening in the z direction. This type of surface is known as an **elliptic paraboloid**. △



Example 1.8.5. Sketch the surface $z = y^2 - x^2$.

Solution: First, let's get the z traces by setting $z = k$. So, $y^2 - x^2 = k$, which describes a *hyperbola* in the xy-plane. Then, the x traces tell us that $z = y^2 - k^2$, which describes a parabola opening in the $+z$ direction in the yz plane. The y traces tell use that $z = -x^2 + k^2$ which describes a parabola opening in the $-z$ direction in the xz plane.

Together, these traces sketch a graph that looks like the one below:



This is known as a **hyperbolic paraboloid**. △

Some other graphs of common quadric surfaces can be seen below: quadrics.png

Example 1.8.6. Identify and sketch the surface $4x^2 - y^2 + 2z^2 + 4 = 0$.

Solution: First, let's put the equation into standard form:

$$-x^2 + \frac{1}{4}y^2 - \frac{1}{2}z^2 = 1.$$

Now, we analyze the z traces to find $\frac{1}{4}y^2 - \frac{1}{2}z^2 = 1 + k^2$, which describes a hyperbola in the $x = k$ planes. The y traces tell us that $-x^2 - \frac{1}{2}z^2 = 1 - \frac{1}{4}k^2$, which can be rearranged to say $x^2 + \frac{1}{2}z^2 = \frac{1}{4}k^2 - 1$, which describe ovals in the $y = k$ planes, but with a limited domain of $|k| > 2$. The x traces tell us that $\frac{1}{4}y^2 - \frac{1}{2}z^2 = 1 + k^2$, which describes a hyperbola in the $x = k$ planes.

Together, these traces describe a **hyperboloid of two sheets**, opening in the y direction. △

Example 1.8.7. Classify the quadric surface $x^2 + 2z^2 - 6x - y + 10 = 0$.

Solution: Let's first complete the square on the x terms,

$$[(x-3)^2 - 9] + 2z^2 - y + 10 = 0$$

And rearrange into standard form:

$$y - (x-3)^2 - 2z^2 = 1$$

Looking at the traces, we can see:

1. $z = k$ gives the equation $y - (x-3)^2 = 1 + 2k^2$, describing a parabola with a positive y-intercept and facing in the $+y$ direction.
2. $y = k$ gives the equation $(x-3)^2 + 2z^2 = k - 1$, describing a circle with center $(3, 0, k)$ and radius k (for $k > 1$).
3. $x = k$ gives the equation $y - 2z^2 = 1 + (k-3)^2$, describing a parabola with a positive y-intercept that opens in the $+y$ direction.

Together, these describe a hyperbola with a vertex at the coordinate $(3, 1, 0)$. △

1.9 Cylindrical and Spherical Coordinates

Remember polar coordinates from single-variable calculus? It's back, with a vengaence. And now there are two of them.

1.9.1 Cylindrical Coordinates

Cylindrical coordinates describe points in the form (r, θ, z), where r is the distance from the origin, θ is the angle of rotation about the z axis, and z is the height on the z axis. These are basically the same as polar coordinates, but with an added z dimension. The conversion equations are similar to polar:

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

and

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

Example 1.9.1. Convert $P_c(2, 2\pi/3, 1)$ from cylindrical coordinates to rectangular coordinates.

Solution: From this, we can see that $r = 2$, $\theta = 2\pi/3$, and $z = 1$. From our conversion equations, we can determine that $x = r\cos\theta = 2\cos(2\pi/3) = -1$ and $y = r\sin\theta = 2\sin(2\pi/3) = \sqrt{3}$. Therefore, the point is equivalent to $P_r(-1, \sqrt{3}, 1)$. △

Example 1.9.2. Find cylindrical coordinates for the point with rectangular coordinates $(3, -3, -7)$.

Solution: We can see that $x = 3$, $y = -3$, and $z = -7$. With our conversion formulas,

$$r = \sqrt{x^2 + y^2} = \sqrt{18} = 3\sqrt{2}$$

and

$$\theta = \tan^{-1} yx^{-1} = \tan^{-1}(-1) = 3\pi/4 + n\pi.$$

Because x is positive and y is negative, θ must be in the fourth quadrant, and we pick the solutions $\theta = 7\pi/4 + 2\pi n$ for any $n \in \mathbb{Z}$. So our coordinate in cylindrical is $(3\sqrt{2}, 7\pi/4 + 2\pi n, -7)$. △

Cylindrical coordinates come in handy for problems that have symmetry about an axis, which we choose to be the z-axis. For example, instead of the equation $x^2 + y^2 = c^2$ to describe a cylinder, we can simply use $r = c$ to describe the same surface.

Example 1.9.3. Describe the surface whose equation in cylindrical coodinates is $z = r$.

Solution: For this, any coordinate (r, θ, r) is valid. In the $z = k$ planes, the surface inscribes circles of radius $|k|$. This suggests that our surface is a cone, which coincides with the rectangular equation $x^2 + y^2 = z^2$, which we can recognize in this case as equivalent to the identity $x^2 + y^2 = r^2$. △

Example 1.9.4. Find an equation in cylindrical coordinates for the ellipsoid $4x^2 + 4y^2 + z^2 = 1$

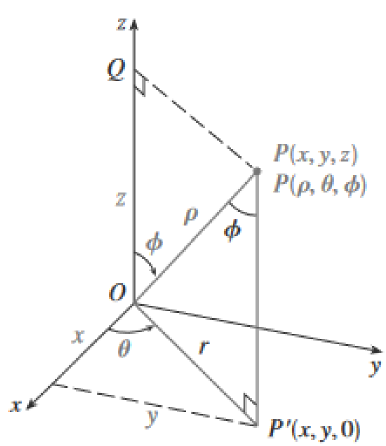
Solution: First, let's rearrange the equation to the form $4(x^2 + y^2) + z^2 = 1$. Then, with the substitution $r^2 = x^2 + y^2$, we can say that $4r^2 + z^2 = 1$, or $z^2 = 1 - 4r^2$. △

1.9.2 Spherical Coordinates

Spherical coordinates come in the form $P(\rho, \theta, \phi)$. $\rho = ||\overrightarrow{OP}||$ is the distance from the origin to P, θ is the same angle as in spherical–the angle about the z axis, and ϕ is the angle between the $+z$ axis and P. Note that

$$\rho \geq 0, \quad 0 \leq \rho \leq \pi.$$

The spherical coordinate system is handy in problems with symmetry about a point, such as cones, spheres, etc. For example, the equation $\rho = c$ describes a sphere of radius C. $\theta = c$ describes a vertical half-plane running along the z axis. $\phi = c$ describes a cone, stemming from the origin at an angle of c from the $+z$ axis. The relationships between spherical and rectangular coordinates are shown in the picture below. Namely,



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

And that

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

Example 1.9.5. Find the rectangular coordinates corresponding to the point $P_s(2, \pi/4, \pi/3)$.

Solution: First, note that $\rho = 2$, $\theta = \pi/4$, and $\phi = \pi/3$. Therefore,

$$\begin{aligned} x &= \rho\cos\theta\sin\phi = 2\cdot\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} = \frac{\sqrt{6}}{2} \\ y &= \rho\sin\theta\sin\phi = 2\cdot\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} = \frac{\sqrt{6}}{2} \\ z &= \rho\cos\phi = 2\cdot\frac{1}{2} = 1 \end{aligned}$$

So the point in rectangular is $P_r(\sqrt{6}/2, \sqrt{6}/2, 1)$. △

Example 1.9.6. Find the spherical coordinates corresponding to the point $P_r(0, 2\sqrt{3}, -2)$.

Solution: First, we can see that

$$\rho = \sqrt{x^2+y^2+z^2} = \sqrt{0^2 + (2\sqrt{3})^2 + 2^2} = 4.$$

Then, we can see that

$$\theta = \tan^{-1}(y/x) = \pi/4 + 2\pi n$$

(because $\lim_{r\to\infty}\tan^{-1} r = \pi/4 + 2\pi n$). Finally, we can see that

$$\phi = \cos^{-1}(z/\rho) = \cos^{-1}\frac{-2}{3\sqrt{2}} = \cos^{-1}\left(-\frac{1}{2}\right) = \frac{2\pi}{3}$$

So the point in spherical coordinates is $P_s(4, \pi/4 + 2\pi n, 2\pi/3)$. △

Example 1.9.7. Find an equation in spherical coordinates for the hyperboloid of two sheets with equation $x^2 - y^2 - z^2 = 1$.

Solution: We can plug in our spherical coordinate version of x, y, and z:

$$\begin{aligned} 1 &= (\rho\cos\theta\sin\phi)^2 - (\rho\sin\theta\sin\phi)^2 - (\rho\cos\phi)^2 \\ &= \rho^2\cos^2\theta\sin^2\phi - \rho^2\sin^2\theta\sin^2\phi - \rho^2\cos^2\phi \\ &= \rho^2\left[\cos^2\theta\sin^2\phi - \sin^2\theta\sin^2\phi - \cos^2\phi\right] \\ \rho &= \left[\sin^2\phi\left(\cos^2\theta - \sin^2\theta\right) - \cos^2\phi\right]^{-1/2} \end{aligned}$$

△

Example 1.9.8. Find a rectangular equation for the surface whose spherical equation is $\rho = \sin\theta\sin\phi$.

Solution: From our conversion equations, $y = \rho\sin\theta\sin\phi$, so $\sin\theta\sin\phi = y/\rho$. Then,

$$\begin{aligned} \rho &= \sin\theta\sin\phi \\ &= \frac{y}{\rho} \\ y &= \rho^2 \\ &= x^2 + y^2 + z^2 \end{aligned}$$

Therefore,

$$x^2 + (y - 1/2)^2 + z^2 = 1/4$$

Which is the equation for a sphere with center $(0, 1/2, 0)$ and radius $1/2$. △

2 Vector Functions

Chapter 13: Page 849

2.1 Vector Functions and Space Curves

In general, a function is just a mapping between an input set (domain) and an output set (range). We are most interested in functions that output three-dimensional vectors–that is, their range is $\mathbb{R}^3$.

Example 2.1.1. If $\boldsymbol{r}(t) = \left\langle t^3, \ln(3-t), \sqrt{t}\right\rangle$, then the component functions are

$$f(t) = t^3 \quad g(t) = \ln(3-t) \quad h(t) = \sqrt{t}$$

By convention, the domain of $\boldsymbol{r}$ consists of all t values for which *all* component functions are defined. Because $f(t)$ is defined for all $t \in \mathbb{R}$, $g(t)$ is defined when $t \in [0,3)$ and $h(t)$ is defined when $t \in [0,\infty)$, $\boldsymbol{r}(t)$ is defined for $t \in [0,3)$. △

The **limit** of a vector function $\boldsymbol{r}$ is defined by taking the limits of the component functions.

Definition 2.1.2. If $\boldsymbol{r}(t) = \langle f(t), g(t), h(t)\rangle$, then

$$\lim_{t\to a} \boldsymbol{r}(t) = \left\langle \lim_{t\to a} f(t), \lim_{t\to a} g(t), \lim_{t\to a} h(t)\right\rangle$$

provided the limits of the component functions exist.

Example 2.1.3. Find $\lim_{t\to 0} \boldsymbol{r}(t)$ if

$$\boldsymbol{r}(t) = (1+t^3)\hat{\boldsymbol{\imath}} + te^{-t}\hat{\boldsymbol{\jmath}} + \frac{\sin t}{t}\hat{\boldsymbol{k}}$$

Solution:

$$\boldsymbol{L} = \lim_{t\to 0} \boldsymbol{r}(t) = \left[\lim_{t\to 0} 1 + t^3\right]\hat{\boldsymbol{\imath}} + \left[\lim_{t\to 0} \frac{t}{e^t}\right]\hat{\boldsymbol{\jmath}} + \left[\lim_{t\to 0} \frac{\sin t}{t}\right]\hat{\boldsymbol{k}}$$

Applying L'Hopital's rule,

$$\begin{aligned} \boldsymbol{L} &= [1]\hat{\boldsymbol{\imath}} + [0]\hat{\boldsymbol{\jmath}} + \left[\lim_{t\to 0} \frac{\cos t}{1}\right]\hat{\boldsymbol{k}} \\ &= \langle 1, 0, 1\rangle . \end{aligned}$$

△

Definition 2.1.4. A vector function $\boldsymbol{r}(t)$ is continuous at a if

$$\lim_{t\to a} \boldsymbol{r}(t) = \boldsymbol{r}(a)$$

There is a close connection between continuous vector functions and space curves. Suppose that f, g, and h are all continuous real-valued functions on an interval I. Then, the set of all points (x,y,z) in space, where

$$x = f(t) \quad y = g(t) \quad z = h(t)$$

for $t \in I$, is called a **space curve**. The above equations are called the *parametric equations* of C and t is called a *parameter*. We can think of C as being traced out by a moving particle whose position at time t_0 is $(f(t_0), g(t_0), h(t_0))$. If we have a vector-valued function

$$\boldsymbol{r}(t) = \langle f(t), g(t), h(t)\rangle ,$$

then $\boldsymbol{r}(t_0)$ describes the position of the particle at $t = t_0$.

Example 2.1.5. Describe the curve defined by the vector-values function

$$\boldsymbol{r}(t) = \langle 1+t, 2+5t, -1+6t \rangle$$

Solution: The corresponding parametric equations are

$$x = 1+t \quad y = 2+5t \quad -1+6t$$

which we can recognize as the parametric equations of a line passing through $P_0(1,2,-1)$ parallel to the vector $\boldsymbol{v} = \langle 1,5,6 \rangle$. This line has equation

$$\boldsymbol{r}(t) = \langle 1,2,-1 \rangle + t \langle 1,5,6 \rangle .$$

△

We can also represent curves in $\mathbb{R}^2$ with vector notation. For instance, the surface described by equations $x = t^2 - 2t$ and $y = t+1$ can be represented with the vector equation

$$\boldsymbol{r}(t) = \left\langle t^2 - 2t, t+1 \right\rangle .$$

Example 2.1.6. Describe the curve with vector equation

$$\boldsymbol{r}(t) = \langle \cos t, \sin t, t \rangle$$

Solution: Since $x^2 + y^2 = \cos^2 t + \sin^2 t = 1$, the curve must lie on the circle of radius one centered about the z axis. Because z is linearly correlated with t, the z coordinate is increasing at a constant rate as the x and y coordinates trace a circle. This describes a helix shape. △

Example 2.1.7. Find a vector equation and parametric equations for the line segment that connects the points $P(1,3,-2)$ and $Q(2,-1,3)$.

Solution: First, define $\boldsymbol{v} = \overrightarrow{QP} = \langle 1,-4,5 \rangle$. Then, the vector equation for the line is

$$\langle 1,3,-2 \rangle + t \langle 1,-4,5 \rangle$$

The corresponding parametric equations are

$$x = 1+t \quad y = 3-4t \quad z = -2+5t$$

If we limit the line segment to only include the points between (and including) P and Q, then we limit t to the domain $[0,1]$. △

Example 2.1.8. Find a vector function that represents the curve of intersection of the cylinder $x^2 + y^2 = 1$ and the plane $y + z = 2$.

Solution: The curve of intersection satisfies both equations. The first equation tells us that x and y can be described by the parametric equations $x = \cos t$ and $y = \sin t$ for $t \in [0, 2\pi]$. The second equation tells us that $z = 2 - y$, which we can rewrite as $2 - \sin t$. Therefore, the vector function is,

$$\boldsymbol{r}(t) = \langle \cos t, \sin t, 2 - \sin t \rangle , \quad t \in [0, 2\pi]$$

This is known as a *paramaterization* of the curve. △

2.2 Derivatives and Integrals of Vector Functions

omg we're actually doing calculus now lesgo

2.2.1 Derivatives

The derivative of a vector function is defined similarly as for real-valued functions:

$$\left.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right|_{t=t_0} = \lim_{h\to 0} \frac{\boldsymbol{r}(t_0+h)-\boldsymbol{r}(t_0)}{h}$$

We call $\boldsymbol{r}'(t_0)$ the **tangent vector** to $\boldsymbol{r}$ at $t=t_0$, and the line

$$\boldsymbol{L} = \boldsymbol{r}(t_0) + \boldsymbol{r}'(t_0)(t-t_0)$$

the **tangent line** to $\boldsymbol{r}$ at $t=t_0$.

Theorem 2.2.1. If $\boldsymbol{r}(t) = \langle f_1(t), f_2(t), f_3(t)\rangle$, where f_1, f_2, and f_3 are differentiable real-valued functions, then $\boldsymbol{r}'(t) = \langle f_1'(t), f_2'(t), f_3'(t)\rangle$.

Proof.

$$\begin{aligned}
\boldsymbol{r}'(t) &= \lim_{h\to 0} \frac{\boldsymbol{r}(t+h) - r(t)}{h} \\
&= \lim_{h\to 0} \frac{1}{h}\Big[f_1(t+h)\hat{\boldsymbol{\imath}} + f_2(t+h)\hat{\boldsymbol{\jmath}} + f_3(t+h)\hat{\boldsymbol{k}} - f_1(t) - f_2(t) - f_3(t)\Big] \\
&= \lim_{h\to 0} \left[\frac{f_1(t+h)-f_1(t)}{h}\hat{\boldsymbol{\imath}} + \frac{f_2(t+h)-f_2(t)}{h}\hat{\boldsymbol{\jmath}} - \frac{f_3(t+h)-f_3(t)}{h}\hat{\boldsymbol{k}}\right] \\
&= \left\langle \lim_{h\to 0} \frac{f_1(t+h)-f_1(t)}{h}, \lim_{h\to 0} \frac{f_2(t+h)-f_2(t)}{h}, \lim_{h\to 0} \frac{f_3(t+h)-f_3(t)}{h} \right\rangle \\
&= \langle f_1'(t), f_2'(t), f_3'(t)\rangle
\end{aligned}$$

□

Example 2.2.2. Find the derivative of $\boldsymbol{r}(t) = \left\langle 1+t^3, te^{-t}, \sin 2t\right\rangle$. Then, find the unit tangent vector to $\boldsymbol{r}$ when $t=0$.

Solution: First, $\boldsymbol{r}'(t) = \left\langle 3t^2, e^{-t} - te^{-t}, 2\cos 2t\right\rangle$. Then, $\boldsymbol{r}'(0) = \langle 0,1,2\rangle$ and $||\boldsymbol{r}'(0)|| = \sqrt{5}$. So, the unit tangent vector is

$$\boldsymbol{T}(0) = \frac{\boldsymbol{r}'(0)}{||\boldsymbol{r}'(0)||} = \left\langle 0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right\rangle$$

△

Example 2.2.3. Find parametric equations for the tangent line to the helix with parametric equations

$$x = 2\cos t \quad y = \sin t \quad z = t$$

at the point $(0, 1, \pi/2)$.

Solution: The vector equation for the helix is $\boldsymbol{r}(t) = \langle 2\cos t, \sin t, t\rangle$. Then, differentiating that,

$$\boldsymbol{r}'(t) = \langle -2\sin t, \cos t, 1\rangle\,.$$

We then need to find the t-value that satisfies $\boldsymbol{r}(t) = \langle 0, 1, \pi/2\rangle$, which we can see by observation of the z-coordinate to be $t = \pi/2$. Then,

$$\boldsymbol{r}'\left(\frac{\pi}{2}\right) = \left\langle -2\sin\frac{\pi}{2}, \cos\frac{\pi}{2}, 1\right\rangle = \langle -2, 0, 1\rangle$$

Then, the tangent line is given by

$$\boldsymbol{L} = \langle -2, 0, 1\rangle\left(t - \frac{\pi}{2}\right) + \left\langle 0, 1, \frac{\pi}{2}\right\rangle$$

or, in parametric form,

$$x = -2t + \pi, \quad y = 1, \quad z = t$$

△

A curve given by a vector function $\boldsymbol{r}(t)$ on an interval I is called **smooth** if $\boldsymbol{r}'$ is continuous and $r'(t) \neq 0$ (except possibly at endpoints of I).

Example 2.2.4. Determine whether the semicubical parabola $\boldsymbol{r}(t) = \left\langle 1 + t^3, t^2 \right\rangle$ is smooth for $t \in \mathbb{R}$.

Solution: First, find $\boldsymbol{r}'(t) = \left\langle 3t^2, 2t \right\rangle$. Because $\boldsymbol{r}'(t) = \mathbf{0}$ when $t = 0$, $\boldsymbol{r}$ is not smooth. We can see from the below graph that there is a sharp corner–called a **cusp**–at the point $\boldsymbol{r}(0)$. Any curve with this behavior is not smooth. △

A curve with a finite number of smooth segments is called **piecewise smooth**. For example, the function in the previous example is piecewise smooth on the interval $(-\infty, 0) \cup (0, \infty)$.

2.2.2 Differentiation Rules

Theorem 2.2.5. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are differentiable vector functions, c is a scalar, and f is a differentiable real-valued function, then,

1. $\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{u}(t) + \boldsymbol{v}(t)] = \boldsymbol{u}'(t) + \boldsymbol{v}'(t)$
2. $\frac{\mathrm{d}}{\mathrm{d}t}[c\boldsymbol{u}(t)] = c\boldsymbol{u}'(t)$
3. $\frac{\mathrm{d}}{\mathrm{d}t}[f(t)\boldsymbol{u}(t)] = f'(t)\boldsymbol{u}(t) + f(t)\boldsymbol{u}'(t)$
4. $\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{u}(t) \cdot \boldsymbol{v}(t)] = \boldsymbol{u}'(t) \cdot \boldsymbol{v}(t) + \boldsymbol{u}(t) \cdot \boldsymbol{v}'(t)$
5. $\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{u}(t) \times \boldsymbol{v}(t)] = \boldsymbol{u}'(t) \times \boldsymbol{v}(t) + \boldsymbol{u}(t) \times \boldsymbol{v}'(t)$
6. $\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{u}(f(t))] = \boldsymbol{u}'(f(t))f'(t)$

Most of these rules are familiar, except for rules 4 and 5, which are similar to the product rule.

Example 2.2.6. Show that if $||\boldsymbol{r}(t)|| = c$ (a constant), then $\boldsymbol{r}'(t)$ is orthogonal to $\boldsymbol{r}(t)$ for all t.

Solution: Consider the relationship $\boldsymbol{r} \cdot \boldsymbol{r} = c^2$. Differentating both sides,

$$\boldsymbol{r}(t) \cdot \boldsymbol{r}'(t) + \boldsymbol{r}'(t) \cdot \boldsymbol{r}(t) = 2\boldsymbol{r}(t) \cdot \boldsymbol{r}'(t) = 0$$

Therefore, $\boldsymbol{r}'(t) \cdot \boldsymbol{r}(t)$ is zero for all t and $\boldsymbol{r}'$ is always orthogonal to $\boldsymbol{r}$. △

2.2.3 Integrals

The **definite integral** of a continuous vector function $\boldsymbol{r}(t)$ can be defined in much the same way as for real-valued functions except the integral is a vector. That is, for some $\boldsymbol{r}(t) = \langle f(t), g(t), h(t) \rangle$,

$$\int_a^b \boldsymbol{r}(t)\mathrm{d}t = \left\langle \int_a^b f(t)\mathrm{d}t, \int_a^b g(t)\mathrm{d}t, \int_a^b h(t)\mathrm{d}t \right\rangle$$

Further, we can extend the fundamental theorem of calculus to vector functions by defining $\boldsymbol{R}(t)$ to be an antiderivative of $\boldsymbol{r}(t)$. Then,

$$\int_a^b \boldsymbol{r}(t)\mathrm{d}t = \boldsymbol{R}(b) - \boldsymbol{R}(a).$$

Example 2.2.7. If $\boldsymbol{r}(t) = \langle 2\cos t, \sin t, 2t \rangle$, evaluate $\int \boldsymbol{r}(t)\mathrm{d}t$.

Solution: This is as simple as integrating each component,

$$\begin{aligned} \int \boldsymbol{r}(t)\mathrm{d}t &= \left\langle \int 2\cos t\mathrm{d}t, \int \sin t\mathrm{d}t, \int 2t\mathrm{d}t \right\rangle \\ &= \left\langle 2\sin t, -\cos t, t^2 \right\rangle + \boldsymbol{C} \end{aligned}$$

Where $\boldsymbol{C}$ is an arbitrary constant vector in $\mathbb{R}^3$. △

2.2.4 Arc Length and Curvature

Recall from Calculus 2 that the arc length of a curve in $\mathbb{R}^2$ is given by

$$S = \int_a^b \sqrt{\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2}\mathrm{d}t$$

The length of a space curve is defined similarly,

$$S = \int_a^b \sqrt{\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2}\mathrm{d}t$$

This can be expressed more compactly as

$$S = \int_a^b ||\boldsymbol{r}'(t)||\mathrm{d}t$$

Example 2.2.8. Find the arc length of the circular helix with vector equation $\boldsymbol{r}(t) = \langle \cos t, \sin t, t\rangle$ from the point $(1, 0, 0)$ to the point $(1, 0, 2\pi)$.

Solution: First, we can easily see from the z coordinate that our t interval is $t \in [0, 2\pi]$. Then, compute $\boldsymbol{r}'(t) = \langle -\sin t, \cos t, 1\rangle$ and

$$||\boldsymbol{r}'(t)|| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{2}$$

Then, compute the arc length,

$$L = \int_0^{2\pi} \sqrt{2}\mathrm{d}t = 2\sqrt{2}\pi$$

△

A single curve C can be represented by more than one vector function. For example, the curve $\boldsymbol{r}_1(t) = \langle t, t^2, t^3\rangle, t \in [1, 2]$ is identical to the curve $\boldsymbol{r}_2(t) = \langle e^u, e^{2u}, e^{3u}\rangle, u \in [0, \ln 2]$. These can be transformed back and forth with the relationship $t = e^u$. These two different (yet identical) curves are different **paramaterizations** of C. The arc length computed is identical regardless of what paramaterization you use.

Now, suppose that C is a piecewise-smooth curve on $I = [a, b]$ given by a differentiable vector function $\boldsymbol{r}(t)$, and C is traversed exactly once on I. Then, we can define the **arc length function** s with

$$s(t) = \int_a^t ||\boldsymbol{r}'(t)||\mathrm{d}t, \quad t \in (a, b]$$

If we differentiate both sides, we arrive at a useful result:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = ||\boldsymbol{r}'(t)||$$

It is often useful to paramaterize a curve with respect to arc length because arc length arises naturally from the shape of the curve and is independent of the coordinate system. If we are given a paramaterization $\boldsymbol{r}(t)$ and $s(t)$ is the arc length function, we may be able to solve for t as a function of s, so $\boldsymbol{r} = \boldsymbol{r}(t(s))$ can be used to find the position of the particle after it its path has reached a particular length.

Example 2.2.9. Reparamaterize the helix $\boldsymbol{r}(t) = \langle \cos t, \sin t, t\rangle$ with respect to arc length measured from $(1, 0, 0)$ in the direction of increasing t.

Solution: First, say $\boldsymbol{r}'(t) = \langle -\sin t, \cos t, 1\rangle$ and $||\boldsymbol{r}'(t)|| = \sqrt{2}$. Then, note that $\boldsymbol{r}(t) = \langle 1, 0, 0\rangle$ when $t = 0$ (so the starting t-value is 0), and

$$s(t) = \int_0^t \sqrt{2}\mathrm{d}t = \sqrt{2}t$$

From there, we can get t as a function of s:

$$t(s) = \frac{1}{\sqrt{2}}s$$

and plug that back into the original paramaterization of the helix:

$$\boldsymbol{r}_2(s) = \left\langle \cos\frac{s}{\sqrt{2}}, \sin\frac{s}{\sqrt{2}}, \frac{2}{\sqrt{2}} \right\rangle$$

$\triangle$

2.2.5 Curvature

If C is a smooth curve defined by the vector function $\boldsymbol{r}$, then $\boldsymbol{r}'(t) \neq \boldsymbol{0}$. Recall that the unit tangent vector given by

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

indicates the drection of the curve. When a curve is fairly straight, $\boldsymbol{T}$ changes direction very slowly. When a curve has sharp twists or bends, however, $\boldsymbol{T}$ changes direction quickly.

To put a number to this idea of changing direction "quickly" or "slowly", we will define a new quantity–curvature.

Definition 2.2.10. The **curvature** of a curve is

$$\kappa = \left\| \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}s} \right\|$$

where $\boldsymbol{T}$ is the unit tangent vector.

We use the derivative of $\boldsymbol{T}$ with respect to the arc length instead of t so that curvature will be independent of the paramaterization used. We can also express the curvature in another way with the chain rule. Because $\frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t}$, we can rewrite κ as

$$\kappa = \left\| \frac{\mathrm{d}\boldsymbol{T}/\mathrm{d}t}{\mathrm{d}s/\mathrm{d}t} \right\|$$

Example 2.2.11. Find the curvature of a circle of radius a.

Solution: First, we can paramaterize the circle as $\boldsymbol{r}(t) = \langle a\cos t, a\sin t\rangle$, where $t \in [0, 2\pi]$. Then, we can compute $\boldsymbol{r}'(t) = \langle -a\sin t, a\cos t\rangle$ and $\|\boldsymbol{r}'(t)\| = \sqrt{a^2\sin^2 t + a^2\cos^2 t} = a$. The tangent unit vector is

$$\boldsymbol{T}(t) = \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|} = \frac{1}{a}\langle -a\sin t, a\cos t\rangle = \langle -\sin t, \cos t\rangle .$$

Then, the arc length is

$$s(t) = \int_0^t \|\boldsymbol{r}'(\tau)\|\mathrm{d}\tau = \int_0^t a\mathrm{d}\tau = at$$

and the curvature is

$$\kappa = \left\| \frac{\mathrm{d}\boldsymbol{T}/\mathrm{d}t}{\mathrm{d}s/\mathrm{d}t} \right\| = \left\| \frac{\langle -\cos t, -\sin t\rangle}{a} \right\| = \frac{1}{a}$$

$\triangle$

Another way to compute the curvature that is often more convenient to apply is given by

$$\kappa = \frac{\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\|}{\|\boldsymbol{r}'(t)\|^3}$$

Proof. Since $\boldsymbol{T} = \boldsymbol{r}'/\|\boldsymbol{r}'\|$ and $\|\boldsymbol{r}'\| = \mathrm{d}s/\mathrm{d}t$, we have

$$\boldsymbol{T} = \frac{\mathrm{d}\boldsymbol{r}/\mathrm{d}t}{\mathrm{d}s/\mathrm{d}t} \implies \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \boldsymbol{T}\frac{\mathrm{d}s}{\mathrm{d}t}$$

Differentiating both sides with t,

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = \boldsymbol{T}\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} + \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t}\frac{\mathrm{d}s}{\mathrm{d}t}$$

Combining these two equations, we can see

$$\begin{aligned}
\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} \times \frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} &= \left[\boldsymbol{T}\frac{\mathrm{d}s}{\mathrm{d}t}\right] \times \left[\boldsymbol{T}\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} + \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t}\frac{\mathrm{d}s}{\mathrm{d}t}\right] \\
&= \left[\cancelto{0}{\frac{\mathrm{d}s}{\mathrm{d}t}\boldsymbol{T} \times \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\boldsymbol{T}}\right] + \left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]^2 \left[\boldsymbol{T} \times \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t}\right] \\
&= \left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]^2 \left[\boldsymbol{T} \times \frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t}\right]
\end{aligned}$$

Because $\|\boldsymbol{T}(t)\| = 1$ and $\boldsymbol{T}$ is orthogonal with $\boldsymbol{T}'$, we can say

$$\begin{aligned}
\left\|\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} \times \frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}\right\| &= \left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]^2 \|\boldsymbol{T} \times \boldsymbol{T}'\| \\
&= \left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]^2 \|\boldsymbol{T}'\| \cancelto{1}{\|\boldsymbol{T}\|} \\
&= \left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]^2 \|\boldsymbol{T}'\| \\
\left\|\frac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}t}\right\| &= \frac{\|\boldsymbol{r}(t) \times \boldsymbol{r}''(t)\|}{[s'(t)]^2}
\end{aligned}$$

Then, because $\kappa = \left\|\boldsymbol{T}'(t)/s'(t)\right\|$, we can rewrite as

$$\left\|\frac{\boldsymbol{T}'(t)}{s'(t)}\right\| = \frac{\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\|}{|s'(t)|^3} = \frac{\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\|}{\|\boldsymbol{r}'(t)\|^3}$$

□

Example 2.2.12. Find the curvature of the twisted cubic $\boldsymbol{r}(t) = \left\langle t, t^2, t^3 \right\rangle$ at a general point and at $(0,0,0)$.

Solution: Fist, compute $\boldsymbol{r}'$ and $\boldsymbol{r}''$.

$$\begin{aligned}
\boldsymbol{r}'(t) &= \left\langle 1, 2t, 3t^2 \right\rangle \\
\boldsymbol{r}''(t) &= \langle 0, 2, 6t \rangle
\end{aligned}$$

Then, compute $\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)$,

$$\boldsymbol{r}'(t) \times \boldsymbol{r}''(t) = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 1 & 2t & 3t^2 \\ 0 & 2 & 6t \end{vmatrix} = \left\langle 6t^2, -6t, 2 \right\rangle$$

Then, compute $\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\|$ and $\|\boldsymbol{r}'(t)\|$.

$$\begin{aligned}
\|\boldsymbol{r}'(t)\| &= \sqrt{1 + (2t)^2 + (3t^2)^2} \\
&= \sqrt{9t^4 + 4t^2 + 1} \\
\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\| &= \sqrt{(6t^2)^2 + (-6t)^2 + 2^2} \\
&= \sqrt{36t^4 + 36t + 4}
\end{aligned}$$

Finally, compute κ

$$\begin{aligned}\kappa &= \frac{\|\boldsymbol{r}'(t) \times \boldsymbol{r}''(t)\|}{\|r'(t)\|^3} \\ &= \frac{(36t^2 + 36t + 4)^{1/2}}{(9t^4 + 4t^2 + 1)^{3/2}}\end{aligned}$$

At $(0, 0, 0)$, the curvature is

$$\kappa(0) = \frac{\sqrt{4}}{1^{3/2}} = 2$$

△

For the special case of a curve in $\mathbb{R}^2$ with $y = f(x)$, we can choose x as the parameter and write $\boldsymbol{r}(x) = \langle x, f(x)\rangle$. Then, $\boldsymbol{r}'(x) = \langle 1, f'(x)\rangle$ and $\boldsymbol{r}''(x) = \langle 0, f''(x)\rangle$. Then, the curvature is given by

$$\kappa = \frac{|| \langle 1, f'(x), 0\rangle \times \langle 0, f''(x), 0\rangle ||}{|| \langle 1, f'(x)\rangle ||^3} = \frac{|f''(x)|}{[1 + (f'(x))^2]^{3/2}}$$

Example 2.2.13. Find the curvature of the parabola $y = ax^2$ at the point $x = x_0$.

Solution: First, compute the required components,

$$\begin{aligned}f'(x) &= 2ax \\ f''(x) &= 2a\end{aligned}$$

Then, compute the curvature,

$$\kappa = \frac{2|a|}{(1 + 4a^2x^2)^{3/2}}$$

And plug in the point $x = x_0$,

$$\kappa = \frac{2|a|}{(1 + 4a^2x_0^2)^{3/2}}$$

△

2.2.6 Normal and Binormal Vectors

At a given point on a smooth space curve $\boldsymbol{r}(t)$, there are many vectors that are orthogonal to the unit tangent vector $\boldsymbol{T}(t)$. We single out one by observing that $|\boldsymbol{T}(t)| = 1$ for all t, so $\boldsymbol{T}(t) \cdot \boldsymbol{T}'(t) = 0$ for all t. Therefore, $\boldsymbol{T}'$ is orthogonal to $\boldsymbol{T}$. We can define the **principal normal unit vector** $\boldsymbol{N}(t)$ (or simply unit normal) as

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

The vector $\boldsymbol{B}(t) = \boldsymbol{T}(t) \times \boldsymbol{N}(t)$ is also normal to both $\boldsymbol{T}$ and $\boldsymbol{N}$, and is called the **binormal vector**. Because $\|\boldsymbol{T}\| = \|\boldsymbol{N}\| = 1$ and they are orthogonal to each other, **B** is also a unit vector.

Example 2.2.14. Find the unit normal and binormal vectors for the circular helix

$$\boldsymbol{r}(t) = \langle \cos t, \sin t, t\rangle$$

Solution: First, compute the required components,

$$\begin{aligned}
\boldsymbol{r}'(t) &= \langle -\sin t, \cos t, 1\rangle \\
\boldsymbol{T}(t) = \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|} &= \left\langle -\frac{\sin t}{\sqrt{2}}, \frac{\cos t}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle \\
\boldsymbol{T}'(t) &= \left\langle -\frac{\cos t}{\sqrt{2}}, -\frac{\sin t}{\sqrt{2}}, 0 \right\rangle \\
\boldsymbol{N}(t) = \frac{\boldsymbol{T}'(t)}{\|\boldsymbol{T}'(t)\|} &= \boxed{\langle -\cos t, -\sin t, 0\rangle} \quad \text{(Normal Vector)} \\
\boldsymbol{B}(t) = \boldsymbol{T}\times\boldsymbol{N} &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ -\sin t/\sqrt{2} & \cos t/\sqrt{2} & 1/\sqrt{2} \\ -\cos t & -\sin t & 0 \end{vmatrix} \\
&= \frac{1}{\sqrt{2}}\left\langle \sin t, -\cos t, \sin^2 t + \cos^2 t\right\rangle = \boxed{\frac{1}{\sqrt{2}}\langle \sin t, -\cos t, 1\rangle} \quad \text{(Binormal Vector)}
\end{aligned}$$

△

The plane determined by $\boldsymbol{N}$ and $\boldsymbol{B}$ and a point P on a curve C is called the **normal plane** to C and P. Every line that lies on this plane is orthogonal to $\boldsymbol{T}$ at P.

The plane determined by $\boldsymbol{T}$ and $\boldsymbol{N}$ is called the **osculating plane** of C and P. This is the plane that comes closest to containing the part of the curve near P.

The circle that lies in the osculating plane of C and P, has the same tangent as C at P, lies on the concave side of C (i.e. in the direction that $\boldsymbol{N}$ points) and has radius $\rho = \kappa^{-1}$ is called the **osculating circle** (or **circle of curvature**) of C at P. It is the circle that best describes how C behaves near P; it shares the same tangent, normal, and curvature at P.

Example 2.2.15. Find the equations of the normal plane and osculating plane of the helix $\boldsymbol{r}(t) = \langle \cos t, \sin t, t\rangle$ at the point $P(0, 1, \pi/2)$.

Solution: From the previous example, we know that

$$\begin{aligned}
\boldsymbol{T}(t) &= \frac{1}{\sqrt{2}}\langle -\sin t, \cos t, 1\rangle \\
\boldsymbol{N}(t) &= \langle -\cos t, -\sin t, 0\rangle \\
\boldsymbol{B}(t) &= \frac{1}{\sqrt{2}}\langle \sin t, -\cos t, 1\rangle
\end{aligned}$$

The point P is found when $t = \pi/2$, so

$$\begin{aligned}
\boldsymbol{T}\left(\frac{\pi}{2}\right) &= \frac{1}{\sqrt{2}}\langle -1, 0, 1\rangle \\
\boldsymbol{N}\left(\frac{\pi}{2}\right) &= \langle 0, -1, 1\rangle \\
\boldsymbol{B}\left(\frac{\pi}{2}\right) &= \frac{1}{\sqrt{2}}\langle 1, 0, 1\rangle
\end{aligned}$$

Recall the equation of a plane, $\boldsymbol{n}\cdot(\boldsymbol{r}-\boldsymbol{r}_0) = 0$. So the equation for the normal plane of $\boldsymbol{r}$ at P is

$$0 = \boldsymbol{T}(\pi/2)\cdot(\boldsymbol{x}-\boldsymbol{p})$$

So

$$\begin{aligned}
0 &= \langle -1, 0, 1\rangle \cdot \langle x, y-1, z-\pi/2\rangle \\
&= -x + z - \pi/2
\end{aligned}$$

Therefore the normal plane is formed by the equation

$$z = x + \pi/2$$

The osculating plane has normal vector $\boldsymbol{B}(\pi/2)$, so its equation is

$$\begin{aligned} 0 &= \frac{1}{\sqrt{2}} \langle 1, 0, 1 \rangle \cdot \langle x, y - 1, z - \pi/2 \rangle \\ &= \frac{1}{\sqrt{2}} x + \frac{1}{\sqrt{2}} \left(z - \frac{\pi}{2} \right) \end{aligned}$$

Or,

$$z = \frac{\pi}{2} - x$$

△

Example 2.2.16. Find and graph the osculating circle of the parabola $y = x^2$ at the origin.

Solution: We already know that the curvature of the parabola $y = ax^2$ is given by

$$\frac{2|a|}{(1 + 4a^2x^2)^{3/2}}$$

Plugging in $a = 1$ and $x = 0$, we get $\kappa = 2$. Therefore, the radius of the circle is $1/\kappa = 1/2$. The circle must be tangent to $y = x^2$ in the direction of $\boldsymbol{N}$, so its center must be at $(0, 1/2)$. Therefore, the equation for the circle is

$$x^2 + \left(y - \frac{1}{2} \right)^2 = \frac{1}{4}$$

△

2.2.7 Motion in Space

Suppose a particle moves through space so that its position vector at time t is $\boldsymbol{r}(t)$, where $\boldsymbol{r}$ is twice-differentiable on the domain of $\boldsymbol{r}$. Then, its velocity vector is $\boldsymbol{r}'(t)$ and its acceleration vector is $\boldsymbol{r}''(t)$.

Additionally, the *speed* of the particle is $\|\boldsymbol{r}'(t)\| = \mathrm{d}s/\mathrm{d}t$, where s is the arclength (or, in the context of motion, distance) function.

Example 2.2.17. The position vector of an object moving in a plane is given by $\boldsymbol{r}(t) = \langle t^3, t^2 \rangle$. Find functions describing its velocity, speed, and acceleration.

Solution: This is as simple as plugging in:

$$\begin{aligned} \text{Velocity: } \boldsymbol{r}'(t) &= \langle 3t^2, 2t \rangle \\ \text{Speed: } \|\boldsymbol{r}'(t)\| &= \sqrt{9t^4 + 4t^2} = t\sqrt{9t^2 + 4} \\ \text{Acceleration: } \boldsymbol{r}''(t) &= \langle 6t, 2 \rangle \end{aligned}$$

△

Example 2.2.18. A moving particle starts at initial position $\boldsymbol{r}(0) = \langle s_x, s_y, s_z \rangle$ with initial velocity $\boldsymbol{v}(0) = \langle v_x, v_y, v_z \rangle$. Its acceleration as a function of time is given by $\boldsymbol{a}(t) = \langle 4t, 6t, 1 \rangle$. Find its position as a function of time.

Solution: First, we can find $\boldsymbol{v}(t)$,

$$\begin{aligned} \boldsymbol{v}(t) &= \int \boldsymbol{a}(t)\mathrm{d}t = \langle 2t^2, 3t^2, t \rangle + \boldsymbol{C} \\ \boldsymbol{v}(0) = \langle v_x, v_y, v_z \rangle \implies \boldsymbol{C} &= \langle v_x, v_y, v_z \rangle \\ \boldsymbol{v}(t) &= \langle 2t^2 + v_x, 3t^2 + v_y, t + v_z \rangle \end{aligned}$$

And then we can find $\boldsymbol{r}(t)$,

$$\begin{aligned}
\boldsymbol{x}(t) &= \int \boldsymbol{v}(t)\mathrm{d}t = \left\langle \frac{2}{3}t^3 + v_x t, t^3 + v_y t, \frac{1}{2}t^2 + v_z t \right\rangle + \boldsymbol{D} \\
\boldsymbol{x}(0) = \langle s_x, s_y, s_z \rangle \implies \boldsymbol{D} &= \langle s_x, s_y, s_z \rangle \\
\boldsymbol{x}(t) &= \left\langle \frac{2}{3}t^3 + v_x t + s_x, t^3 + v_y t + s_y, \frac{1}{2}t^2 + v_z t + s_z \right\rangle
\end{aligned}$$

These types of problems are basically three initial value problems (similar to the ones in Calculus 2) mushed into one. Treat the x, y, and z components separately, and you get three single-variable calculus problems. $\triangle$

If the force that acts on a particle and the particle's mass is known, then acceleration can be found with Newton's Second Law,

$$\boldsymbol{F} = m\boldsymbol{x}''$$

Example 2.2.19. An object with mass m that moves in a circular path with constant angular speed ω has position vector $\boldsymbol{r}(t) = \langle a\cos\omega t, a\sin\omega t\rangle$. Find the force acting on the object and show that is directed towards the origin.

Solution: First, find $\boldsymbol{a}(t)$,

$$\boldsymbol{a}(t) = \boldsymbol{r}''(t) = \frac{\mathrm{d}}{\mathrm{d}t}\langle -a\omega\sin\omega t, a\omega\cos\omega t\rangle = \left\langle -a\omega^2\cos\omega t, -a\omega^2\sin\omega t\right\rangle$$

The force is simply $\boldsymbol{F} = m\boldsymbol{a}$,

$$\boldsymbol{F} = \left\langle -am\omega^2\cos\omega t, -am\omega^2\sin\omega t\right\rangle.$$

Note that $\boldsymbol{F} = -m\omega^2\boldsymbol{r}$, the force acts in the opposite direction of the position vector, which will be pointed towards the origin. $\triangle$

Example 2.2.20. A projectile is fired with an angle of elevation θ and initial velocity $\boldsymbol{v}_0$. Assuming that air resistance is negligible, find the position function $\boldsymbol{r}(t)$ of the projectile. What value of θ maximizes the range (the horizontal distance traveled)?

Solution: First, we can find the components of the initial velocity,

$$\boldsymbol{v}_0 = \langle \|\boldsymbol{v}_0\|\cos\theta, \|\boldsymbol{v}_0\|\sin\theta\rangle$$

Then, use NSL to find the acceleration on the object,

$$\boldsymbol{F} = m\boldsymbol{a} = \langle 0, -mg\rangle \implies \boldsymbol{a} = \langle 0, -g\rangle$$

Then the velocity function,

$$\boldsymbol{v}(t) = \boldsymbol{v}_0 + \int_0^t \boldsymbol{a}(\tau)\mathrm{d}\tau = \langle \|\boldsymbol{v}_0\|\cos\theta, \|\boldsymbol{v}_0\|\sin\theta - gt\rangle$$

and finally the position function (assuming that $\boldsymbol{r}(0) = \boldsymbol{0}$),

$$\boldsymbol{r}(t) = \int_0^t \boldsymbol{v}(\tau)\mathrm{d}\tau = \left\langle \|\boldsymbol{v}_0\|t\cos\theta, \|\boldsymbol{v}_0\|t\sin\theta - \frac{1}{2}gt^2\right\rangle$$

The object will hit the ground when $\boldsymbol{r}(t)\cdot\boldsymbol{\jmath} = 0$, so

$$\begin{aligned}
0 &= \|\boldsymbol{v}_0\|t\sin\theta - \frac{1}{2}gt^2 \\
&= t\left(\|\boldsymbol{v}_0\|\sin\theta - \frac{1}{2}gt\right)
\end{aligned}$$

This will have two solutions: $t = 0$ and $t = 2g^{-1}\|\boldsymbol{v}_0\| \sin\theta$. The second solution is the one we're looking for. Then, we can plug that back into the x-component of $\boldsymbol{r}$ to find the range,

$$\begin{aligned} x_f &= \|\boldsymbol{v}_0\| \left[2g^{-1}\|\boldsymbol{v}_0\| \sin\theta\right] \cos\theta \\ &= \frac{2\|\boldsymbol{v}_0\|^2 \sin\theta \cos\theta}{g} \\ &= \frac{\|\boldsymbol{v}_0\|^2 \sin(2\theta)}{g} \end{aligned}$$

This will reach its max on the interval $[0, \pi/2]$ when $2\theta = \pi/2$ or $\theta = \pi/4$. △

2.2.8 Tangential and Normal Components of Acceleration

When we study the acceleration of a particle, it is often useful to resolve the acceleration into two components, one in the direction of the tangent and one in the direction of the normal. If we write $v = \|\boldsymbol{v}\|$ for the speed of the particle, then

$$\boldsymbol{T}(t) = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\boldsymbol{v}}{v}$$

So $\boldsymbol{v} = v\boldsymbol{T}$. Differentiating both sides of the equation with respect to t, we get

$$\boldsymbol{a} = v'\boldsymbol{T} + v\boldsymbol{T}'$$

We know that $\boldsymbol{T}$ is the tangent vector and $\boldsymbol{T}'/\left\|\boldsymbol{T}'\right\|$ is the normal vector, so we can resolve the acceleration into

$$\boldsymbol{a} = [v']\boldsymbol{T} + \left[v\left\|\boldsymbol{T}'\right\|\right]\boldsymbol{N}$$

where the first term is the tangential component and the second term is the normal component. Additionally, since we know that $\kappa = \left\|\boldsymbol{T}'\right\|/\|\boldsymbol{r}'\| = \left\|\boldsymbol{T}'\right\|/v$, we can rewrite as

$$\boldsymbol{a} = [v']\boldsymbol{T} + \left[\kappa v^2\right]\boldsymbol{N}$$

This formula tells us a couple of things about the behavior of acceleration functions

1. Due to the absence of a $\boldsymbol{B}$ term, we can conclude that $\boldsymbol{a}$ always lies in the osculating plane.
2. When the speed is constant, there is no tangential component of acceleration, such as in the case of uniform circular motion.
3. There will only be no normal component to acceleration if the speed is zero or the curvature is zero (i.e. the object is traveling in a straight line).

We can do some more rewriting to find requations that only depend on $\boldsymbol{r}$, $\boldsymbol{r}'$, and $\boldsymbol{r}''$. First, consider the product

$$\begin{aligned} \boldsymbol{v} \cdot \boldsymbol{v}' &= v\boldsymbol{T} \cdot \left[v'\boldsymbol{T} + \kappa v^2 \boldsymbol{N}\right] \\ &= v\boldsymbol{T} \cdot v'\boldsymbol{T} + v\boldsymbol{T} \cdot \kappa v^2 \boldsymbol{N} \end{aligned}$$

Because $\boldsymbol{T} \cdot \boldsymbol{T} = 1$ and $\boldsymbol{T} \cdot \boldsymbol{N} = 0$ (because $\boldsymbol{T}$ is orthogonal to $\boldsymbol{N}$),

$$\boldsymbol{v} \cdot \boldsymbol{v}' = vv' \implies v' = \frac{\boldsymbol{v} \cdot \boldsymbol{v}'}{v}$$

Plugging this back into the expresion for the tangential component,

$$v' = \frac{\boldsymbol{v} \cdot \boldsymbol{v}'}{v} = \frac{\boldsymbol{r}' \cdot \boldsymbol{r}''}{\|\boldsymbol{r}'\|}$$

and the normal component,

$$\kappa v^2 = \underbrace{\frac{\|\boldsymbol{r}' \times \boldsymbol{r}''\|}{\|\boldsymbol{r}'\|^3}}_{\kappa} \underbrace{\|\boldsymbol{r}'\|^2}_{v^2} = \frac{\|\boldsymbol{r}' \times \boldsymbol{r}''\|}{\|\boldsymbol{r}'\|}$$

So we can write $\boldsymbol{r}''$ as the linear combination of these,

$$\boldsymbol{r}'' = \left[\frac{\boldsymbol{r}' \cdot \boldsymbol{r}''}{\|\boldsymbol{r}'\|}\right]\boldsymbol{T} + \left[\frac{\|\boldsymbol{r}' \times \boldsymbol{r}''\|}{\|\boldsymbol{r}'\|}\right]\boldsymbol{N}$$

You may recognize the first term as the projection of $\boldsymbol{r}''$ onto $\boldsymbol{r}'$, which matches with our intuition of what the tangential component is.

Example 2.2.21. A particle moves with position function $\boldsymbol{r}(t) = \left\langle t^2, t^2, t^3 \right\rangle$. Find the tangential and normal components of acceleration.

Solution: First, find the required components,

$$\begin{aligned}
\boldsymbol{r}'(t) &= \left\langle 2t, 2t, 3t^2 \right\rangle \\
\|\boldsymbol{r}'(t)\| &= \sqrt{8t^2 + 9t^4} \\
\boldsymbol{r}''(t) &= \langle 2, 2, 6t \rangle \\
\boldsymbol{r}' \cdot \boldsymbol{r}'' &= 8t + 18t^3 \\
\|\boldsymbol{r}' \times \boldsymbol{r}''\| &= \left\|\begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 2t & 2t & 3t^2 \\ 2 & 2 & 6t \end{vmatrix}\right\| = \sqrt{(6t^2)^2 + (6t^2)^2 + (0)^2} \\
&= \sqrt{72t^4} = [6\sqrt{2}]t^2
\end{aligned}$$

Then, compute the components,

$$\begin{aligned}
\boldsymbol{r}'' &= \left[\frac{\boldsymbol{r}' \cdot \boldsymbol{r}''}{\|\boldsymbol{r}'\|}\right]\boldsymbol{T} + \left[\frac{\|\boldsymbol{r}' \times \boldsymbol{r}''\|}{\|\boldsymbol{r}'\|}\right]\boldsymbol{N} \\
&= \frac{8t + 18t^3}{t\sqrt{8 + 9t^2}}\boldsymbol{T} + \frac{6\sqrt{2}t^2}{t\sqrt{8 + 9t^2}}\boldsymbol{N} \\
&= \frac{8 + 18t^2}{\sqrt{9t^2 + 8}}\boldsymbol{T} + \frac{6\sqrt{2}t}{\sqrt{9t^2 + 8}}\boldsymbol{N}
\end{aligned}$$

△

3 Partial Derivatives

Chapter 14: Page 888

3.1 Functions of Several Variables

In this section we will study functions of multiple variables from several points of view:

1. verbally
2. numerically
3. algebraically
4. visually

3.1.1 Functions of Two Variables

The temperature T at a point on the surface of the Earth at a given time depends on the longitude x and latitude y of the point. We can think of T as being a function of the two variables x or y, or a function of the pair (x, y). We indicate this by writing $T = f(x, y)$, or with $f : \mathbb{R}^2 \to \mathbb{R}$, with the $\to$ ("to") operator indicating that f transforms from the domain of $\mathbb{R}^2$ to the codomain of $\mathbb{R}$.

The domains of functions of two variables will be some subset of $\mathbb{R}^2$, which can be written as some rule that x and y must follow.

Example 3.1.1. Find the domains of the two functions

$$f(x, y) = \frac{\sqrt{x + y + 1}}{x - 1} \quad g(x, y) = x \ln\big(y^2 - x\big)$$

Solution: Because the input of a square root must be positive, the domain of f is all two-tuples (x, y) such that $x + y + 1 \geq 0$. We also must have $x \neq 1$ because of the denominator. We can compactly write this as

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

The domain of g can be found through a similar process. The function ln necessitates a positive input, so our domain is

$$D = \{(x, y)|y^2 - x > 0\}$$

△

Example 3.1.2. Find the domain and range of

$$g(x, y) = \sqrt{9 - x^2 - y^2}$$

Solution: The domain is

$$D = \{(x, y)|9 - x^2 - y^2 \geq 0\}$$

The maximum value attained by g is 3, when $x = y = 0$. The minimum value of g is 0, when $9 - x^2 - y^2 = 0$. Therefore, the range is $[0, 3]$. We can write the range more generally as

$$R = \{z|z = g(x, y), (x, y) \in D\}$$

△

3.1.2 Graphs

Another way of visualizing the behavior of a function of two variables is with a graph.

Definition 3.1.3. If f is a function of two variables with domain D, then the **graph** of f is the set of all points (x, y, z) in $\mathbb{R}^3$ such that $z = f(x, y)$ and $(x, y) \in D$.

We call these graphs in $\mathbb{R}^3$ *surfaces*.

One special type of function are functions of the form

$$f(x, y) = ax + by + c.$$

These are called **linear functions**. The graphs of these types of functions will be planes in $\mathbb{R}^3$.

3.1.3 Level Curves

Another way to visualize functions of two variables is by using a *level curve*. These maps, also called contour curves, can visualize graphs within a plane. One place where you may have already seen a contour map is with graphs of equipotential lines in physics.

Definition 3.1.4. The **level curves** of a function $f : \mathbb{R}^2 \to \mathbb{R}$ with range R are the graphs of curves with equations $f(x, y) = k$, where $k \in R$ is a constant.

Example 3.1.5. Sketch the level curves of the function $f(x, y) = 6-3x-2y$ for the values $k \in \{-6, 0, 6, 12\}$.

Solution: The curves formed by $f(x, y) = k$ will be lines in $\mathbb{R}^2$,

$$\begin{aligned}
6 - 3x - 2y = -6 &\implies y = -\frac{3}{2}x + 6 \\
6 - 3x - 2y = 0 &\implies y = -\frac{3}{2}x + 3 \\
6 - 3x - 2y = 6 &\implies y = -\frac{3}{2}x \\
6 - 3x - 2y = 12 &\implies y = -\frac{3}{2}x - 3
\end{aligned}$$

Because these lines are equally spaced and parallel, we can reason that f describes a plane in $\mathbb{R}^3$. △

3.1.4 Functions of Three or More Variables

A **function of three variables**, f, is a rule that assigns a real number to each ordered triple (x, y, z) in a domain $D \subseteq \mathbb{R}^3$. For instance, if the temperature depends on the longitude x, latitude y, and time t, we can write $T = f(x, y, t)$, where $f : \mathbb{R}^3 \to \mathbb{R}$.

Example 3.1.6. Find the domain of f if

$$f(x, y, z) = \ln(z - y) + xy \sin z$$

Solution: The input of ln must be positive, so the domain is

$$D = \{(x, y, z) \in \mathbb{R}^3 | z - y > 0\}$$

△

It is very difficult to visualize functions of three dimensions, since that they would lie in a four-dimensional space. However, we can examine **level surfaces**, which are the surfaces generated by the equations $f(x, y, z) = k$ for a constant k.

Example 3.1.7. Find the level surfaces for the function

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

Solution: The level surfaces will be of the form $x^2 + y^2 + z^2 = k$, which we know to describe a sphere centered at $(0, 0, 0)$ with radius $\sqrt{k}$. △

3.1.5 Functions of Many Variables

Functions with more than three variables are also possible to consider, although they are exceedingly difficult to visualize. A function of n-variables is a rule that assigns a number $z = f(x_1, x_2, \ldots, x_n)$ in some domain $D \subseteq \mathbb{R}^n$. These (as well as functions of many variables in both the domain *and* codomain) will be explored more thoroughly in Linear Algebra.

3.2 Limits and Continuity

Let's compare the behavior of the functions

$$f(x, y) = \frac{\sin(x^2 + y^2)}{x^2 + y^2} \quad \text{and} \quad g(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$$

as x and y both approach 0.

Using the above table, we can see that as $(x, y) \to (0, 0)$, f approaches 1 from all directions, but g approaches different values depending on the direction you approach from. Thus, we write

$$\lim_{(x,y)\to(0,0)} f(x, y) = 1 \quad \text{and} \quad \lim_{(x,y)\to(0,0)} g(x, y) \text{ does not exist}$$

TABLE 1 Values of $f(x, y)$

$x \backslash y$	−1.0	−0.5	−0.2	0	0.2	0.5	1.0
−1.0	0.455	0.759	0.829	0.841	0.829	0.759	0.455
−0.5	0.759	0.959	0.986	0.990	0.986	0.959	0.759
−0.2	0.829	0.986	0.999	1.000	0.999	0.986	0.829
0	0.841	0.990	1.000		1.000	0.990	0.841
0.2	0.829	0.986	0.999	1.000	0.999	0.986	0.829
0.5	0.759	0.959	0.986	0.990	0.986	0.959	0.759
1.0	0.455	0.759	0.829	0.841	0.829	0.759	0.455

TABLE 2 Values of $g(x, y)$

$x \backslash y$	−1.0	−0.5	−0.2	0	0.2	0.5	1.0
−1.0	0.000	0.600	0.923	1.000	0.923	0.600	0.000
−0.5	−0.600	0.000	0.724	1.000	0.724	0.000	−0.600
−0.2	−0.923	−0.724	0.000	1.000	0.000	−0.724	−0.923
0	−1.000	−1.000	−1.000		−1.000	−1.000	−1.000
0.2	−0.923	−0.724	0.000	1.000	0.000	−0.724	−0.923
0.5	−0.600	0.000	0.724	1.000	0.724	0.000	−0.600
1.0	0.000	0.600	0.923	1.000	0.923	0.600	0.000

In general, the notation

$$\lim_{(x,y)\to(a,b)} f(x,y) = L \quad \text{or} \quad \lim_{\boldsymbol{x}\to\boldsymbol{v}} f(\boldsymbol{x}) = L$$

indicates that the values of f approach the number L as the input approaches (a, b) (or $\boldsymbol{v}$) from *any* path in the domain of f. A more precise definition follows.

Definition 3.2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function whose domain D includes points arbitrarily close to $\boldsymbol{a}$. Then we say that the **limit of f as $\boldsymbol{x}$ approaches $\boldsymbol{a}$ is** L, and we write

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = L$$

if for every number $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that

$$0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \delta, \boldsymbol{x} \in D \implies |f(\boldsymbol{x}) - L| < \epsilon$$

There is a similar definition for vector-valued functions.

Definition 3.2.2. Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a function whose domain D include points arbitrarily close to $\boldsymbol{a}$. Then,

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{L}$$

where $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{L} \in \mathbb{R}^m$ if and only if for every $\epsilon > 0$, there exists some $\delta > 0$ such that

$$0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \delta, \boldsymbol{x} \in D \implies \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{L}\| < \epsilon.$$

Example 3.2.3. Show that

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

does not exist.

Solution: First, set $y = 0$ and find that $\lim\limits_{x\to 0} f(x, 0) = \lim\limits_{x\to 0} \frac{x^2}{x^2} = 1$. Then, set $x = 0$ and find that $\lim\limits_{y\to 0} f(0, y) = \lim\limits_{y\to 0} \frac{-y^2}{y^2} = -1$. Because these values are different, f does not approach the same value from all paths and the limit does not exist. △

Example 3.2.4. If $f(x, y) = xy/(x^2 + y^2)$, does $\lim\limits_{(x,y)\to(0,0)} f(x, y)$ exist?

Solution: Analyzing the limit along the x and y axes, we can find that along the line $y = 0$,

$$\lim_{x\to 0} f(x, 0) = \lim_{x\to 0} \frac{0}{x^2} = 0$$

and

$$\lim_{y\to 0} f(0,y) = \lim_{y\to 0} \frac{0}{y^2} = 0$$

However, if analyze the limit along the line $y = x$, we can see

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

Because we've found two paths that give us a different value, the limit does not exist. △

Example 3.2.5. If $f(x,y) = \frac{xy^2}{x^2+y^4}$, does $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ exist?

Solution: First, we'll look at the limit along all lines $y = mx$ for $m \in \mathbb{R}$. This gives us

$$\lim_{x\to 0} f(x,mx) = \lim_{x\to 0} \frac{m^2x^3}{x^2+m^4x^4} = 0$$

So as (x,y) approaches $(0,0)$ along any *linear* path, $f(x,y)$ approaches 0. However, let's look at a nonlinear path, such as $y = \sqrt{x}$. Then,

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

So the limit does not exist. △

This process of looking at several possible paths is ultimately going to be fruitless, because we can come up with infinitely many unique ways to approach $(0,0)$. Instead, if we want to look at limits that do exist, we'll have to get more clever.

First up, let's look at some limit laws.

Theorem 3.2.6 (Properties of Multivariable Limits)**.** Let $f : \mathbb{R}^n : \mathbb{R}$ and $g : \mathbb{R}^n : \mathbb{R}$ be functions with domains including points arbitrarily close to $\boldsymbol{x}_0$. Let $h : \mathbb{R} \to \mathbb{R}$ be a function of a single variable that is continuous at $x = a$, let $c \in \mathbb{R}$, and let

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = a \quad \text{and} \quad \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g(\boldsymbol{x}) = b$$

Then the following equalities hold:

$$\begin{aligned}
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})g(\boldsymbol{x}) = \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g(\boldsymbol{x}) = ab\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) + g(\boldsymbol{x}) = \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g(\boldsymbol{x}) = a + b\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} [f(\boldsymbol{x})]^c = \left[\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})\right]^c = a^c\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} [f(\boldsymbol{x})]^{g(\boldsymbol{x})} = \left[\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})\right]^{\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g(\boldsymbol{x})} = a^b\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} cf(\boldsymbol{x}) = c \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = ca\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} [f(\boldsymbol{x}) + c] = \left[\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})\right] + c = a + c\\
&\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} h(f(\boldsymbol{x})) = h\left[\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})\right] = h(a)
\end{aligned}$$

Example 3.2.7. Find $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2}$ if it exists.

Solution: based on how the function looks, it seems that the numerator will shrink faster than the denominator and the limit will be 0. To prove this, we can use the epsilon-delta definition.

Suppose $\epsilon > 0$. Then, we need to find some $\delta > 0$ such that

$$\left|\frac{3x^2y}{x^2+y^2} - 0\right| < \epsilon \quad \text{whenever} \quad 0 < \sqrt{x^2+y^2} < \delta$$

We can rewrite this to say

$$\frac{3x^2|y|}{x^2+y^2} < \epsilon \quad \text{whenever} \quad 0 < \sqrt{x^2+y^2} < \delta$$

Because $x^2 \leq x^2 + y^2$ (as y^2 is always positive), we can say

$$\frac{3x^2|y|}{x^2+y^2} \leq \frac{3x^2|y|}{x^2} = 3|y| = 3\sqrt{y^2}$$

Because $\sqrt{y^2} \leq \sqrt{x^2+y^2}$, we can say

$$3\sqrt{y^2} \leq 3\sqrt{x^2+y^2}$$

If we choose $\delta = \epsilon/3$, then

$$|f(x,y) - 0| \leq 3\sqrt{x^2+y^2} \leq 3\delta = \epsilon$$

So the proof is complete and $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$. △

3.2.1 Continuity

Recall that evaluating limits of *continuous* functions of a single variable is easy, using direct substitution. This same property applies to continuous functions of multiple variables.

Definition 3.2.8. A function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is **continuous** at $\boldsymbol{x}_0 \in \mathbb{R}^n$ if and only if

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}_0)$$

If f is continuous on every point in a subset $S \subseteq \mathbb{R}^n$, we say that f is continuous on S.

Polynomial functions of several variables are functions of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_i c x_1^{m_{i,1}} x_2^{m_{i,2}} \cdots x_n^{m_{i,n}}$$

Where all exponents are non-negative integers.

Rational functions of several variables are functions of the form $f(x) = g(x)/h(x)$ where g and h are polynomial functions.

Because polynomial functions are only composed of addition and multiplication, limit properties can be used to find that all polynomials with the domain are continuous on their entire domain. From this property, we can also determine that all rational functions are continuous on their entire domain.

Example 3.2.9. Evaluate

$$\lim_{(x,y)\to(1,2)} \left[x^2y^3 - x^3y^2 + 3x + 2y\right]$$

Solution: We can just plug in, as this is a continuous polynomial function.

$$\lim_{(x,y)\to(1,2)} \left[x^2y^3 - x^3y^2 + 3x + 2y\right] = 1^2 \cdot 2^3 - 1^3 \cdot 2^2 + 3 \cdot 1 + 2 \cdot 2 = 11$$

△

Example 3.2.10. Let

$$f(x) = \begin{cases} \frac{x^2-y^2}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases} \quad \text{and} \quad g(x) = \begin{cases} \frac{3x^2y}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Is f continuous at $(0,0)$? Is g?

Solution: First, we can remember from the previous section that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist. Therefore, f is not continuous at $(0,0)$, even with the hole patched.

For g, recall that $\lim\limits_{(x,y)\to(0,0)} g(x,y) = 0$. Therefore, $g(x,y)$ is defined and has an existing limit at $(0,0)$, and those values are the same. Therefore, g is continuous. △

Example 3.2.11. Where is the function $h(x,y) = \arctan(y/x)$ continuous?

Solution: The function $f(x,y) = y/x$ is continuous for all $(x,y) \in \mathbb{R}^2, x \neq 0$. $\arctan u$ is continuous everywhere. Therefore, the composite function $\arctan(f(x,y))$ is continuous everywhere except where $x = 0$. △

3.2.2 Partial Derivatives

Partial derivatives give us a way to analyze the rates of change of functions when only one of their parameters is allowed to vary.

If $f(x,y)$ is a function of two variables, its *partial derivatives* are defined as

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

These are also sometimes written as f_x and f_y or $\partial_x f$ and $\partial_y f$. The values of these partial derivatives correspond to the rate of change of f when only one parameter (the one being differentiated with) is allowed to vary, and the other is held fixed.

To find partial derivatives, simply find the derivative as normal, but treat all variables besides the one being differentiated with as constants.

Example 3.2.12. Find f_x and f_y if $f(x,y) = x^3 + x^2y^3 - 2y^2$.

Solution: Differentiate.

$$\frac{\partial f}{\partial x} = 3x^2 + 2xy^3$$
$$\frac{\partial f}{\partial y} = 3x^2y^2 - 4y$$

△

3.2.3 Interpretations of Partial Derivatives

Let the equation $z = f(x,y)$ represent a surface S in $\mathbb{R}^3$. If $f(a,b) = c$, then the point $P(a,b,c)$ lies on S. By fixing $y = b$ and allowing x to vary, we are analyzing the change in the z-coordinate as x moves away from $x = a$. In other words, we are restricting our view only to the intersection of the plane $y = b$ with S.

Further, the partial derivatives $f_x(a,b)$ and $f_y(a,b)$ can be interpreted geometrically as the slopes of the tangent lines to the traces of S formed by the intersection of S with the planes $y = b$ and $x = a$.

Example 3.2.13. Find $\partial_x z$ and $\partial_y z$ if z is defined implicitly under the equation

$$x^3 + y^3 + z^3 + 6xyz = 1$$

Solution: First, implicitly (partial) differentiate with respect to x,

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

and solve for $\partial_x z$,

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

Then do the same for $\partial_y z$:

$$3y^2 + 3z^2\frac{\partial z}{\partial y} + 6xz + 6xy\frac{\partial z}{\partial y} = 0 \implies \frac{\partial z}{\partial y} = \frac{-y^2 - 2xz}{z^2 + 2xy}$$

△

3.2.4 Functions of more than two variables

Partial derivatives are defined in much the same way for functions of multiple variables.

Definition 3.2.14. If $f(x_1, x_2, \ldots, x_n)$ is differentiable, then $\partial_{x_i} f$ is defined as

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

3.2.5 Higher Derivatives

If $f(x, y)$ is a function of two variables, then its partial derivatives f_x and f_y are also functions of two variables, and can be partially differentiated as normal. Therefore, we must consider the **second partial derivatives** of f: $(f_x)_x, (f_y)_x, (f_x)_y$ and $(f_y)_y$. The following notation is used:

$$(f_a)_b = f_{ab} = \frac{\partial^2 f}{\partial a \partial b}$$

Example 3.2.15. Find all second partial derivatives of

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

Solution: First, find the first order partials,

$$f_x = 3x^2 + 2xy^3 \quad \text{and} \quad f_y = 3x^2y^2 - 4y$$

Then, find the second order partials

$$f_{xx} = 6x + 2y^3 \qquad f_{yy} = 6x^2y - 4$$
$$f_{xy} = 6xy^2 \qquad f_{yx} = 6xy^2$$

△

Note that $f_{xy} = f_{yx}$. This is not a coincidence.

Theorem 3.2.16 (Clairaut's Theorem)**.** Suppose f is defined on a disk D that contains the point (a, b). If the functions f_{xy} and f_{yx} are both continuous on D, then

$$f_{xy}(a, b) = f_{yx}(a, b)$$

This theorem will extend to functions of more than two variables and partial derivatives of higher order as well.

3.2.6 Partial Differential Equations

Any equation that contains a partial derivative is called a partial differential equation. For instance, the equation

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

is called **Laplace's equation**. Solutions to this equation are called **harmonic functions** and are important in many applications, especially in thermodynamics, fluids, and electricity.

The methods for solving these types of equations are outside the scope of this course. You'll learn more about it in the aptly named class "partial differential equations."

3.2.7 The Cobb-Douglas Production Problem

We're going to take a break from the pure math and look at an application for a bit.

Consider a production function $P(K, L)$ that describes how much of a good a society will produce. The parameters of this function are capital and labor for K and L respectively. The partial derivative $\partial_L P$ is the rate at which production changes with a change in labor. Economists refer to this as the *marginal productivity of labor*. Likewise, the partial derivative $\partial_K P$ is the *marginal productivity of capital.* The assumptions made by Cobb and Douglas to describe production are as follows:

1. If either capital or labor vanishes, then so will production
2. The marginal productivity of labor is proportional to the amount of production per unit of labor
3. The marginal productivity of capital is proportional to the amount of production per unit of capital.

Because the production per unit of labor is P/L, assumption 2 says

$$\frac{\partial P}{\partial L} = \alpha \frac{P}{L}$$

If we take capital to be constant at $K = K_0$, then this partial differential equation becomes a total differential equation

$$\frac{\mathrm{d}P}{\mathrm{d}L} = \alpha \frac{P}{L}$$

Solving this separable differential equation, we get

$$\begin{aligned} \frac{\mathrm{d}P}{P} &= \alpha \frac{\mathrm{d}L}{L} \\ \ln P &= \alpha \ln L + C(K_0) \\ P &= C_1(K_0)L^\alpha \end{aligned}$$

C_1 is a function of K_0 because the value of C will vary depending on what K_0 is.

Doing the same process, except taking labor to be constant at $L = L_0$, we find that

$$P = C_2(L_0)L^\beta$$

Combining these two equations, we get

$$P = bL^\alpha K^\beta$$

Where b is a positive constant independent of both L_0 and K_0. From assumption 1, we can guarantee that $\alpha, \beta > 0$. Further, if we increase both labor and capital by a constant factor m, then

$$\begin{aligned} P' &= b(L+m)^\alpha (K+m)^\beta \\ &= bm^{\alpha+\beta} L^\alpha K^\beta = m^{\alpha+\beta} P \end{aligned}$$

If $\alpha + \beta = 1$, the $P' = mP$. Because of this, Cobb and Douglas assumed $\alpha + \beta = 1$ and thus

$$P = bL^\alpha K^{1-\alpha}$$

This is the famous *Cobb-Douglas production function* that is widely used in the study of economics.

3.3 Tangent Planes and Linear Approximations

3.3.1 Tangent Planes

One of the most important results of single-variable calculus is the idea of using the tangent line to approximate the values of functions. In this section we will develop a similar idea for functions of multiple variables by using planes to approximate values.

Suppose a surface S has equation $z = f(x, y)$ where f has continuous partial derivative, and let $P_0(x_0, y_0, z_0)$ be a point on S. Then, let C_1 and C_2 be the curves formed by the intersection between S and the planes $y = y_0$ and $x = x_0$ respectively. Let T_1 and T_2 be tangent to these curves at P. Then the **tangent plane** to S at P is the plane containing both tangent lines T_1 and T_2.

In fact, all possible tangents to S will lie in this plane (we will show why this is true later). This tangent plane can be used to approximate values of f near P.

We know that the equation of a plane will be of the form

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

If we divide by C and define $a = -A/C$, $b = -B/C$, we get

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

setting $y = y_0$, we get the two lines

$$\begin{aligned} y &= y_0 \\ z - z_0 &= a(x - x_0) \end{aligned}$$

We can recognize this as the point-slope equation for a line, with slope a. However, we know that the T_1 must have slope $f_x(x_0, y_0)$. Therefore,

$$f_x(x_0, y_0) = a$$

We can repeat this process, putting $x = x_0$ into the equation for the plane to find

$$f_y(x_0, y_0) = b$$

Therefore, the equation for the tangent plane is

$$f_x(x_0, y_0)[x - x_0] + f_y(x_0, y_0)[y - y_0] = z - z_0$$

Example 3.3.1. Find the tangent plane to the elliptic paraboloid $z = 2x^2 + y^2$ at the point $(1, 1, 3)$.

Solution: First, compute $z_x(1, 1)$ and $z_y(1, 1)$.

$$\begin{aligned} z_x &= 4x \implies z_x(1, 1) = 4 \\ z_y &= 2y \implies z_y(1, 1) = 2 \end{aligned}$$

So the equation for the tangent plane is

$$z - 3 = 4(x - 1) + 2(y - 1)$$

△

3.3.2 Linear Approximations

Just as we used tangent lines to approximate values in single-variable calculus, we can use tangent planes to approximate values in multivariable functions. If $L(x, y)$ gives the z coordinate of the tangent plane to $f(x, y)$ at (x_0, y_0) in terms of the x and y coordinates, then L–called the *linearization* of f at (x_0, y_0)–is a good approximation of f near (x_0, y_0).

In the previous section, we found a tangent plane to $f(x,y)=2x^2+y^2$ at $(1,1)$ to be given by the equation

$$z-3=4(x-1)+2(y-1)$$

Rearranging this, we get

$$L(x,y)=4x+2y-3$$

Using this function, we can approximate values near $(1,1,3)$. For instance,

$$f(1.1,0.95)\approx L(1.1,0.95)=4.4+1.9-3=3.3$$

Compared to the true value of $f(1.1,0.95)$ of 3.3225, this is a pretty good approximation.

We have defined tangent planes for surfaces with continuous first partial derivatives, but what if this isn't the case? For instance, consider the surface described by the equation

$$f(x,y)=\begin{cases}\frac{xy}{x^2+y^2} & (x,y)\neq(0,0)\\ 0 & (x,y)=(0,0)\end{cases}$$

Computing the first partial derivatives, we get that $f_x(0,0)=f_y(0,0)=0$. However, they are not continuous, because $\lim\limits_{(x,y)\to(0,0)} f_x(x,y)$ does not exist (and likewise for f_y). The tangent plane approximation of f at $(0,0)$ would be $f(x,y)\approx 0$, but $f(x,y)=\frac{1}{2}$ for all points on the line $y=x$, so this tangent plane approximation is clearly inaccurate. To avoid cases like this, we will formula the idea of differentiability for multivariable functions.

Definition 3.3.2. A function $f(x,y)$ is **differentiable** at a point (a,b) if and only if, for all points (x,y) in a δ disk around (a,b), f can be expressed as

$$f(x,y)=f(a,b)+f_x(a,b)[x-a]+f_y(a,b)[y-a]+E(x,y)$$

where the error term E satisfies

$$\lim_{(x,y)\to(a,b)}\frac{E(x,y)}{\sqrt{([x-a]^2+[y-b]^2)}}=0$$

In English, this is essentially saying that if, for any point near (a,b), we can express $f(x,y)$ as the sum of the value of its tangent plane at (a,b) and some error term that becomes arbitrarily small as $(x,y)\to(a,b)$. In other words, f "looks like" its tangent plane near (a,b).

While the formal definition of differentiability is good to know, it is a very cumbersome and difficult definition to work with. Luckily, there is a much more simple way to check for differentiability.

Theorem 3.3.3. If $f:\mathbb{R}^n\to\mathbb{R}$ is a multivariable function, it is differentiable at a point (a,b) if all of its first partial derivatives exist near (a,b) and are continuous at (a,b).

Example 3.3.4. Show that $f(x,y)=xe^{xy}$ is differentiable at $(1,0)$ and find its linearization there. Then, use that to approximate $f(1.1,-0.1)$.

Solution: The partial derivatives are:

$$f_x(x,y)=e^{xy}+xye^{xy}\quad f_x(1,0)=1$$

$$f_y(x,y)=x^2e^{xy}\quad f_y(1,0)=1$$

Because both partial derivatives are continuous at $(1,0)$, f is continuous there as well. Further, the linearization is given by

$$L(x,y)=f(1,0)+f_x(1,0)(x-1)+f_y(1,0)y$$

or

$$L(x,y)=1+x-1+y=x+y$$

With this approximation, we find $f(1.1,-0.1)\approx L(1.1,-0.1)=1$. Compared to the actual value $f(1.1,-0.1)\approx=$ 0.98542, this is a pretty good approximation. △

3.3.3 Differentials

For functions of single variables, we defined the differential $\mathrm{d}x$ to be an independent variable. Then, $\mathrm{d}y$ can be defined in terms of $\mathrm{d}x$:

$$\mathrm{d}y = f'(x)\mathrm{d}x$$

Now, consider a function of multiple variables $z = f(x_1, x_2, \ldots, x_n)$. We can define $\mathrm{d}x_1$, $\mathrm{d}x_2$, and so on until $\mathrm{d}x_n$ to be independent variables, and then write

$$\mathrm{d}z = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \frac{\partial f}{\partial x_2}\mathrm{d}x_2 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n$$

Further, we can say that

$$f(x_1 + \mathrm{d}x_1, x_2 + \mathrm{d}x_2, \ldots, x_n + \mathrm{d}x_n) \approx f(x_1, x_2, \ldots, x_n) + \mathrm{d}z$$

As the size of the differentials increases, this approximation gets less and less accurate, since it only uses the tangent plane instead of the full surface.

Example 3.3.5. If $z = f(x, y) = x^2 + 3xy - y^2$, find an expression for $\mathrm{d}z$ in terms of x, y, $\mathrm{d}x$, and $\mathrm{d}y$. Then, with $x = 2$, $y = 3$, $\mathrm{d}x = 0.05$ and $\mathrm{d}y = -0.04$, compare the values of $\mathrm{d}z$ and $f(x + \mathrm{d}x, y + \mathrm{d}y) - f(x, y)$.

Solution: First, finding the differential,

$$\begin{aligned} \mathrm{d}z &= f_x(x, y)\mathrm{d}x + f_y(x, y)\mathrm{d}y \\ &= [2x + 3y]\mathrm{d}x + [3x - 2y]\mathrm{d}y \end{aligned}$$

and plugging in the values,

$$\mathrm{d}z = [2(2) + 3(3)]0.05 + [3(2) - 2(3)](-0.04) = 0.65$$

Comparing this to $f(x + \mathrm{d}x, y + \mathrm{d}y) - f(x, y)$,

$$\Delta z = f(2.05, 2.96) - f(2, 3) = 0.6449$$

So $\mathrm{d}z$ is a pretty good approximation for Δz. △

Example 3.3.6. The base radius of a right circular cone are measured as 10cm and 25cm respectively, with a possible error in each measurement of at most 0.1cm. Use differentials to approximate the maximum error in the calculated volume cone.

Solution: The volume V of a cone with radius r and height h is $V = \pi r^2 h/3$. Then, the differential $\mathrm{d}V$ is

$$\mathrm{d}V = \left[\frac{\pi r^2}{3}\right]\mathrm{d}h + \left[\frac{2\pi r h}{3}\right]\mathrm{d}r$$

Since the error is at most 0.1cm, $|\mathrm{d}h| \leq 0.1$ and $|\mathrm{d}r| \leq 0.1$. Taking the maximum error,

$$\mathrm{d}V \leq \left[\frac{\pi\left(10^2\right)}{3}\right]0.1 + \left[\frac{2\pi(10)(25)}{3}\right]0.1 = 20\pi$$

So the maximum possible error in the volume is $20\pi\text{cm}^3 \approx 63\text{cm}^3$ △

3.4 The Multivariable Chain Rule

Consider a single variable function $y = f(x(t))$. Then, the chain rule from calculus 1 states

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}$$

Now, in the case of multivariable functions, we can use the *multivariable chain rule.*

Theorem 3.4.1. If $z = f(x_1, x_2, \ldots, x_n)$ is a differentiable function, and $x_1 = x(t)$, $x_2 = x_2(t)$, and so on are all differentiable single variable functions, the derivative of z with respect to t is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\mathrm{d}x_i}{\mathrm{d}t}$$

Proof. Let $z = f(\boldsymbol{x})$ be a multivariable function where $\boldsymbol{x} = \langle x_1, x_2, \ldots, x_n \rangle$ that is differentiable at a point $\boldsymbol{u} = \langle u_0, u_1, \ldots, u_n \rangle$. Let $x_1 = x_1(t)$, $x_2 = x_2(t)$, and so on be differentiable functions with time. Then, a change in time Δt causes a change Δx_1 in x_1, Δx_2 in x_2, and so on. These, in turn, cause a change Δz in z. We can write

$$\Delta z = \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial f}{\partial x_n}\Delta x_n + E(\boldsymbol{x})$$

where the error term E satisfies

$$\lim_{\boldsymbol{x}\to\boldsymbol{u}} E(\boldsymbol{x}) = 0$$

If we divide both sides by Δt, we get

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x_1}\frac{\Delta x_1}{\Delta t} + \frac{\partial f}{\partial x_2}\frac{\Delta x_2}{\Delta t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\Delta x_n}{\Delta t} + \frac{E(\boldsymbol{x})}{\Delta t}$$

Now, if we let $\Delta t \to 0$,

$$\begin{aligned}
\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t} &= \lim_{\Delta t\to 0}\left[\frac{\partial f}{\partial x_1}\frac{\Delta x_1}{\Delta t} + \frac{\partial f}{\partial x_2}\frac{\Delta x_2}{\Delta t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\Delta x_n}{\Delta t} + \frac{E(\boldsymbol{x})}{\Delta t}\right] \\
&= \frac{\partial f}{\partial x_1}\lim_{\Delta t\to 0}\frac{\Delta x_1}{\Delta t} + \frac{\partial f}{\partial x_2}\lim_{\Delta t\to 0}\frac{\Delta x_2}{\Delta t} + \cdots + \frac{\partial f}{\partial x_n}\lim_{\Delta t\to 0}\frac{\Delta x_n}{\Delta t} + \lim_{\Delta t\to 0}\frac{E(\boldsymbol{x})}{\Delta t}
\end{aligned}$$

Now, because $\lim\limits_{\Delta t\to 0} \boldsymbol{x} = \boldsymbol{u}$, we can rewrite as

$$\begin{aligned}
\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t} &= \frac{\partial f}{\partial x_1}\lim_{\Delta t\to 0}\frac{\Delta x_1}{\Delta t} + \frac{\partial f}{\partial x_2}\lim_{\Delta t\to 0}\frac{\Delta x_2}{\Delta t} + \cdots + \frac{\partial f}{\partial x_n}\lim_{\Delta t\to 0}\frac{\Delta x_n}{\Delta t} + \cancelto{0}{\frac{\lim_{\boldsymbol{x}\to\boldsymbol{u}} E(\boldsymbol{x})}{\lim_{\Delta t\to 0}\Delta t}} \\
&= \frac{\partial f}{\partial x_1}\lim_{\Delta t\to 0}\frac{\Delta x_1}{\Delta t} + \frac{\partial f}{\partial x_2}\lim_{\Delta t\to 0}\frac{\Delta x_2}{\Delta t} + \cdots + \frac{\partial f}{\partial x_n}\lim_{\Delta t\to 0}\frac{\Delta x_n}{\Delta t}
\end{aligned}$$

Taking note that $\lim\limits_{\Delta t\to 0}\frac{\Delta z}{\Delta t} = \frac{\mathrm{d}z}{\mathrm{d}t}$ and $\lim\limits_{\Delta t\to 0}\frac{\Delta x_i}{\Delta t} = \frac{\mathrm{d}x_i}{\mathrm{d}t}$, rewrite:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial f}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial f}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\mathrm{d}x_n}{\mathrm{d}t} = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{\mathrm{d}x_i}{\mathrm{d}t}$$

And we've reached the desired result. □

Example 3.4.2. If $z = x^2 y + 3xy^4$, where $x = \sin 2t$ and $y = \cos t$, find $\partial z/\partial t$.

Solution: This is as simple as plugging in:

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} \\
&= 2\big[2xy + 3y^4\big]\cos 2t - \big[x^2 + 12xy^3\big]\sin t
\end{aligned}$$

We can also get an expression depending only on t if we plug back in for x and y,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = 2\big[2(\sin 2t)(\cos t) + 3(\cos t)^4\big]\cos 2t - \big[(\sin 2t)^2 + 12(\sin 2t)(\cos t)^3\big]\sin t$$

Note that we could have also plugged in the t-valued expressions for x and y *before* differentiating. We would get the same result, but it would be much more difficult and tedious to compute.

We can interpret this result as the rate of change of z as the point (x, y) moves across the curve described by parametric equations $x = \sin 2t$ and $y = \cos t$. △

Example 3.4.3. The pressure P (in pascals), volume V, and temperature T of an ideal gas are related by the equation $PV = nRT$ where n is the amount of substance (in moles) and R is the ideal gas constant, with approximate value 8.3144 $\text{m}^3 \cdot \text{Pa} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$.

If 1000 mol of a gas is in a 100 L container that is expanding at a rate of 0.2 Ls^{-1}, at a temperature of 300 K that is increasing at a rate of 0.1 Ks^{-1}, find the rate of change of the pressure of the gas.

Solution: First, isolate the pressure,

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

Then, apply the multivariable chain rule:

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}t} &= \frac{\partial P}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} + \frac{\partial P}{\partial V}\frac{\mathrm{d}V}{\mathrm{d}t} \\ &= \left[nRV^{-1}\right]\frac{\mathrm{d}T}{\mathrm{d}t} + \left[-nrTV^{-2}\right]\frac{\mathrm{d}V}{\mathrm{d}t} \\ &\approx -41.572 \ \text{Pa} \cdot \text{s}^{-1}\end{aligned}$$

△

Now, consider the case where the parameters of a multivariable function are also multivariable functions. This is another special case of the multivariable chain rule.

Theorem 3.4.4. Let $z = f(x_1, x_2, \ldots, x_n)$, with each parameter $x_i = g_i(t_1, t_2, \ldots, t_m)$ being a multivariable function. Then,

$$\frac{\partial z}{\partial t_a} = \sum_{i=1}^{n} \frac{\partial z}{\partial x_i}\frac{\partial x_i}{\partial t_a}$$

For each $a = 1, 2, \ldots, m$.

The proof of this theorem is pretty easy but i dont feel like writing it out so im just gonna kind of explain it

Find the same expression

$$\Delta z = \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial f}{\partial x_n}\Delta x_n + E(\boldsymbol{x})$$

and then recognize that each Δx_i can be shown to be equal to

$$\Delta x_i = \frac{\partial x_i}{\partial t_a}\Delta t_a + E_i(x_i)$$

Then plug these back into the original expression for Δz and take the limit as $\Delta t_a \to 0$.

The same ideas can be extended to multivariable functions with several layers of parameters.

Example 3.4.5. If $u = x^4 y + y^2 z^3$, where $x = rse^t$, $y = rs^2 e^{-t}$, and $z = r^2 s \sin t$, find an expression for $\partial u/\partial s$.

Solution: By the chain rule,

$$\begin{aligned}\frac{\partial u}{\partial s} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s} \\ &= \left[4x^3 y\right]re^t + \left[x^4 + 2yz^3\right]2rse^{-t} + \left[3y^2 z^2\right]r^2 \sin t\end{aligned}$$

△

Example 3.4.6. If $g(s,t) = f(s^2 - t^2, t^2 - s^2)$ and f is differentiable, show that g satisfies the equation

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 0.$$

Solution: First, define $u = s^2 - t^2$ and $v = t^2 - s^2$. Then, $g(s,t) = f(u,v)$. From there, we can compute each partial derivative:

$$\begin{aligned}
\frac{\partial g}{\partial s} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial s} \\
&= 2s\frac{\partial f}{\partial u} - 2s\frac{\partial f}{\partial v} \\
\frac{\partial g}{\partial t} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial t} \\
&= -2t\frac{\partial f}{\partial u} + 2t\frac{\partial f}{\partial v}
\end{aligned}$$

So the overall expression can be written as

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 2st\frac{\partial f}{\partial u} - 2st\frac{\partial f}{\partial v} - 2st\frac{\partial f}{\partial u} + 2st\frac{\partial f}{\partial v} = 0$$

△

Example 3.4.7. If $z = f(x,y)$ has continuous second-order partial derivatives, $x = r^2 + s^2$ and $y = 2rs$, find $\partial z/\partial r$ and $\partial^2 z/\partial r^2$.

Solution: The chain rule tells us

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial z}{\partial x}2r + \frac{\partial z}{\partial y}2s$$

Then, differentiating again,

$$\frac{\partial^2 z}{\partial r^2} = \left[2\frac{\partial z}{\partial x} + 2r\frac{\partial}{\partial r}\left(\frac{\partial z}{\partial x}\right)\right] + \left[2s\frac{\partial}{\partial r}\left(\frac{\partial z}{\partial y}\right)\right]$$

Applying the chain rule to the inner terms, we find that

$$\begin{aligned}
\frac{\partial}{\partial r}\left(\frac{\partial z}{\partial x}\right) &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right)\frac{\partial x}{\partial r} + \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right)\frac{\partial y}{\partial r} \\
&= 2r\frac{\partial^2 z}{\partial x^2} + 2s\frac{\partial^2 z}{\partial x \partial y} \\
\frac{\partial}{\partial r}\left(\frac{\partial z}{\partial y}\right) &= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right)\frac{\partial x}{\partial r} + \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right)\frac{\partial y}{\partial r} \\
&= 2r\frac{\partial^2 z}{\partial x \partial y} + 2s\frac{\partial^2 z}{\partial y^2}
\end{aligned}$$

Plugging these back into the expression for $\partial^2 z/\partial r^2$,

$$\begin{aligned}
\frac{\partial^2 z}{\partial r^2} &= 2\frac{\partial z}{\partial x} + 2r\left[2r\frac{\partial^2 z}{\partial x^2} + 2s\frac{\partial^2 z}{\partial x \partial y}\right] + 2s\left[2r\frac{\partial^2 z}{\partial x \partial y} + 2s\frac{\partial^2 z}{\partial y^2}\right] \\
&= 2\frac{\partial z}{\partial x} + 4r^2\frac{\partial^2 z}{\partial x^2} + 4s^2\frac{\partial^2 z}{\partial y^2} + 8rs\frac{\partial^2 z}{\partial x \partial y}
\end{aligned}$$

△

3.4.1 Implicit Differentiation

Suppose that an equation of the form $F(x,y) = 0$ defines y implicitly as a differentiable function of x. That is, $y = f(x)$, where $F(x, f(x)) = 0$ for all x in the domain of f. If F is differentiable, we can apply the multivariable chain rue to obtain

$$\frac{\partial F}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial F}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} = 0$$

Because $\mathrm{d}x/\mathrm{d}x$ is just 1, this simplifies to

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} = 0$$

Rearranging some terms, we obtain

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\partial F/\partial x}{\partial F/\partial y}$$

To derive this result, we assumed that $F(x, y)$ implicitly defines a function y of x. The **Implicit Function Theorem**, proved in advanced calculus, tells us when this assumption is valid. It states that if F is defined on a disk containing (a, b), where $F(a, b) = 0$, $F_y(a, b) \neq 0$, and F_x and F_y are continuous on the disk, then the equation $F(x, y)$ defines y as a function of x near (a, b) and the derivative of the function is given by the above formula.

Example 3.4.8. Find $\mathrm{d}y/\mathrm{d}x$ if $x^3 + y^3 = 6xy$.

Solution: First, move all terms over to one side, so $x^3 + y^3 - 6xy = 0$. Then, we can say that

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{3x^2 - 6y}{3y^2 - 6x} = -\frac{x^2 - 2y}{y^2 - 2x}$$

△

A similar rule also exists for an implicit definition of z as $F(x, y, z) = 0$. Partially differentiating both sides, we get one of the two following formulas,

$$\frac{\partial F}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0$$
$$\frac{\partial F}{\partial x}\frac{\partial x}{\partial y} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} = 0$$

Because $\partial y/\partial x = \partial x/\partial y = 0$ and $\partial/\partial x = \partial y/\partial y = 0$, these formulas become

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0$$
$$\frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} = 0$$

Which can be rearranged to give us

$$\frac{\partial z}{\partial x} = -\frac{\partial F/\partial x}{\partial F/\partial z} \quad \text{and} \quad \frac{\partial z}{\partial y} = -\frac{\partial F/\partial y}{\partial F/\partial z}$$

Again, we can use the implicit function theorem to say when the assumption that $F(x, y, z)$ implicitly defined z in terms of x and y is valid. If F is defined within a sphere containing (a, b, c), where $F(a, b, c) = 0$, $F_z(a, b, c) \neq 0$, and F_x, F_y, and F_z are all continuous inside the sphere, then the assumption is valid.

3.5 Directional Derivatives and the Gradient Vector

3.5.1 The Directional Derivative

Consider a differentiable function $f(x, y)$. Recall that f_x and f_y gave us the rate of change of f in the $+x$ and $+y$ directions respectively. Now, we will introduce a method to get the rates of change of functions in an *arbitrary* direction. If a unit vector $\boldsymbol{u} = \langle a, b\rangle$ is the direction we are trying to find the rate of change in, then the **directional derivative** of f in the direction of $\boldsymbol{u}$ tells us this. The x and y components of this will be the projection of the vector with components $\langle f_x, f_y\rangle$ onto $\boldsymbol{u}$. In equation form,

$$D_{\boldsymbol{u}}f = af_x + bf_y = \langle f_x, f_y\rangle \cdot \langle a, b\rangle$$

Definition 3.5.1. The **directional derivative** of f at (x_0, y_0) in the direction of a unit vector $\boldsymbol{u} - \langle a, b\rangle$ is

$$D_{\boldsymbol{u}} f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + a, y_0 + b) - f(x_0, y_0)}{h}$$

Note that if $\boldsymbol{u} = \hat{\boldsymbol{\imath}}$, we just get f_x, and if $\boldsymbol{u} = \hat{\boldsymbol{\jmath}}$, we just get f_y. This tells us that the partial derivatives of f are just special cases of the directional derivative.

3.5.2 The Gradient Vector

We will define a new quantity, the gradient, of a function.

Definition 3.5.2. If $f(x_1, x_2, \ldots, x_n)$ is a differentiable function, then the gradient of f is the vector function ∇f defined by

$$\nabla f = \left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right\rangle$$

Example 3.5.3. Find ∇f where $f(x, y) = \sin x + e^{xy}$.

Solution: Just plug in.

$$\begin{aligned} \nabla f &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \\ &= \langle \cos x + ye^{xy}, xe^{xy}\rangle \end{aligned}$$

△

With this new notation, we can rewrite our expression for the directional derivative.

Theorem 3.5.4. If $f(x_1, x_2, \ldots, x_n)$ is a differentiable function, and $\boldsymbol{u} = \langle u_1, u_2, \ldots, u_n\rangle$ is a unit vector, then the directional derivative of f in the direction of $\boldsymbol{u}$ at a point $\boldsymbol{t} = \langle t_0, t_1, \ldots, t_n\rangle$ is given by

$$D_{\boldsymbol{u}} f(\boldsymbol{t}) = \nabla f(\boldsymbol{t}) \cdot \boldsymbol{u}$$

Proof. First, define $g(a) = f(\boldsymbol{t} + a\boldsymbol{u})$. Then, we must have $x_i = t_i + au_i$. Further, by the definition of the derivative, we have

$$\begin{aligned} \frac{\mathrm{d}g}{\mathrm{d}a} &= \lim_{h\to 0} \frac{g(a+h) - g(a)}{h} \\ &= \lim_{h\to 0} \frac{f(\boldsymbol{t} + (a+h)\boldsymbol{u}) - f(\boldsymbol{t} + a\boldsymbol{u})}{h} \end{aligned}$$

Then, evaluating it at $a = 0$,

$$\left.\frac{\mathrm{d}g}{\mathrm{d}n}\right|_{a=0} = \lim_{h\to 0} \frac{f(\boldsymbol{t} + h\boldsymbol{u}) - f(\boldsymbol{t})}{h}$$

This definition may look similar. It is the same as the definition of the directional derivative. That is,

$$\left.\frac{\mathrm{d}g}{\mathrm{d}n}\right|_{a=0} = D_{\boldsymbol{u}}(\boldsymbol{t}) \tag{3.5.1}$$

We can also find the derivative of g with the chain rule, which tells us

$$\begin{aligned} \left.\frac{\mathrm{d}g}{\mathrm{d}a}\right|_{a=0} &= \left.\frac{\mathrm{d}f}{\mathrm{d}(\boldsymbol{t} + a\boldsymbol{u})}\right|_{a=0} \frac{\mathrm{d}}{\mathrm{d}a}[t + a\boldsymbol{u}] \\ &= \left.\frac{\mathrm{d}f}{\mathrm{d}(\boldsymbol{t} + a\boldsymbol{u})}\right|_{a=0} = \left.\frac{\mathrm{d}f}{\mathrm{d}v}\right|_{v=\boldsymbol{t}} = \partial_{x_1} f(\boldsymbol{t})\frac{\mathrm{d}x_1}{\mathrm{d}a} + \partial_{x_2} f(\boldsymbol{t})\frac{\mathrm{d}x_2}{\mathrm{d}a} + \cdots + \partial_{x_n} f(\boldsymbol{t})\frac{\mathrm{d}x_n}{\mathrm{d}a} \end{aligned}$$

Recall that $x_i = t_1 + au_1$, so $\mathrm{d}x_i/\mathrm{d}a = u_i$. Therefore,

$$\begin{aligned}
\left.\frac{\mathrm{d}g}{\mathrm{d}a}\right|_{a=0} &= \partial_{x_1} f(\boldsymbol{t})u_1 + \partial_{x_2} f(\boldsymbol{t})u_2 + \cdots + \partial_{\boldsymbol{t}} f(t_n)x_n u_n \\
&= \langle \partial_{x_1} f(\boldsymbol{t}), \partial_{x_2} f(\boldsymbol{t}), \ldots, \partial_{x_n} f(\boldsymbol{t})\rangle \cdot \langle u_1, u_2, \ldots, u_n\rangle \\
&= \nabla f(\boldsymbol{t}) \cdot \boldsymbol{u}
\end{aligned}$$

Finally, equating this result with Eq. 3.5.1, we get

$$D_{\boldsymbol{u}} f(\boldsymbol{t}) = \nabla f(\boldsymbol{t}) \cdot \boldsymbol{u}$$

□

Example 3.5.5. Find the directional derivative of the function $f(x, y) = x^2y^3 - 4y$ at the point $(2, -1)$ in the direction of $\boldsymbol{v} = \langle 2, 5\rangle$.

Solution: First, compute the gradient vector of f at $(2, -1)$:

$$\begin{aligned}
\nabla f &= \left\langle 2xy^3, 3x^2y^2 - 4\right\rangle \\
\nabla f(2, -1) &= \left\langle 2(2)(-1)^3, 3(2)^2(-1)^2 - 4\right\rangle \\
&= \langle -4, 8\rangle
\end{aligned}$$

Then, find the unit vector in the direction of $\boldsymbol{v}$,

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{1}{\sqrt{29}}\boldsymbol{v} = \left\langle \frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}} \right\rangle$$

Finally, compute the directional derivative,

$$D_{\boldsymbol{u}} f(2, -1) = \nabla f(2, -1) \cdot \boldsymbol{u} = \langle -4, 8\rangle \cdot \left\langle \frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}} \right\rangle = \frac{32}{\sqrt{29}}$$

△

Example 3.5.6. If $f(x, y, z) = x \sin yz$, find the gradient of f and the directional derivative of f at $(1, 3, 0)$ in the direction of $\boldsymbol{v} = \langle 1, 2, -1\rangle$.

Solution: First, compute the gradient,

$$\nabla f = \langle f_x, f_y, f_z\rangle = \langle \sin yz, zx \cos yz, yx \cos yz\rangle$$

Then, find $\nabla f(1, 3, 0)$,

$$\nabla f(1, 3, 0) = \langle \sin 0, 0 \cos 0, 3 \cos 0\rangle = \langle 0, 0, 3\rangle$$

Then, compute $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$,

$$\boldsymbol{u} = \frac{1}{\sqrt{6}}\boldsymbol{v}$$

And find the directional derivative,

$$D_{\boldsymbol{u}} f(1, 3, 0) = \langle 0, 0, 3\rangle \cdot \left\langle \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}} \right\rangle = -\frac{3}{\sqrt{6}}$$

△

3.5.3 Maximizing the Directional Derivative

Suppose we have a function f of n variables, and we consider all possible directional derivatives of f at a given point. In which direction would f change fastest, and how fast would that be? The answers are provided by the following theorem.

Theorem 3.5.7. Suppose f is a differentiable function of several variables. The maximum value of the directional derivative $D_{\boldsymbol{u}} f(\boldsymbol{x})$ is $\|\nabla f(\boldsymbol{x})\|$, and it occurs when $\boldsymbol{u}$ is in the same direction as ∇f.

Proof. Because $D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u}$, we can rewrite as

$$D_{\boldsymbol{u}} f = \|\nabla f\| \|\boldsymbol{u}\| \cos\theta$$

Where θ is the angle between $\boldsymbol{u}$ and ∇f. This expression reaches its maximum when $\cos\theta = 1$, or when $\theta = 0$. That is, when the direction of ∇f and $\boldsymbol{u}$ is the same.

Now, if we let $\boldsymbol{u} = \frac{\nabla f}{\|\nabla f\|}$ (the unit vector in the direction of ∇f), we can compute the value of the directional derivative,

$$D_{\boldsymbol{u}} f = \nabla f \cdot \boldsymbol{u} = \nabla f \cdot \frac{\nabla f}{\|\nabla f\|} = \frac{\|\nabla f\|^2}{\|\nabla f\|} = \|\nabla f\|$$

□

By the same reasoning, we can see that the directional derivative of f is 0 when $\boldsymbol{u}$ is orthogonal to ∇f, and that the directional derivative is at its minimum of $-\|\nabla f\|$ when ∇f and $\boldsymbol{u}$ are pointing in opposite directions.

Example 3.5.8. If $f(x, y) = xe^y$, find the rate of change of f at the point $P(2, 0)$ in the direction from P to $Q(0.5, 2)$. Then, find the direction where f has the greatest rate of change, and find the value of the greatest rate of change.

Solution: First, find the gradient of f at P,

$$\nabla f = \langle e^y, xe^y \rangle$$
$$\nabla f(2, 0) = \langle 1, 2 \rangle$$

And the vector from P to Q,

$$\overrightarrow{PQ} = \langle 0.5 - 2, 2 - 0 \rangle = \langle -1.5, 2 \rangle$$
$$\boldsymbol{u} = \frac{\overrightarrow{PQ}}{\left\|\overrightarrow{PQ}\right\|} = \left\langle -\frac{3/2}{5/2}, \frac{2}{5/2} \right\rangle = \left\langle -\frac{3}{5}, \frac{4}{5} \right\rangle$$

And the directional derivative,

$$D_{\boldsymbol{u}} f(2, 0) = \nabla f(2, 0) \cdot \boldsymbol{u} = -\frac{3}{5} + \frac{8}{5} = 1$$

To find the maximum rate of change, first normalize the gradient,

$$\boldsymbol{u}_2 = \frac{\nabla f(2, 0)}{\|\nabla f(2, 0)\|} = \left\langle \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right\rangle$$

This is the "direction of greatest ascent." Then, compute the directional derivative,

$$D_{\boldsymbol{u}_2} f(2, 0) = \|\nabla f(2, 0)\| = \sqrt{5}$$

△

Example 3.5.9. Suppose that the temperature at a point (x, y, z) in space is given by

$$T(x, y, z) = \frac{80}{1 + x^2 + 2y^2 + 3z^2},$$

where T is measured in degrees Celsius and x, y, z in meters. In which direction does temperature increase the fastest at the point $(1, 1, -2)$, and how fast does it change if you are moving at a speed of $v\,\text{ms}^{-1}$?

Solution: First, compute the gradient vector.

$$\begin{aligned}
\nabla T &= \left\langle -\frac{160x}{(1+x^2+2y^2+3z^2)^2}, -\frac{320y}{(1+x^2+2y^2+3z^2)^2}, -\frac{480z}{(1+x^2+2y^2+3z^2)^2} \right\rangle \\
\nabla T(1,1,-2) &= \left\langle -\frac{160}{(1+1^2+2(1)^2+3(-2)^2)^2}, -\frac{320}{(1+1^2+2(1)^2+3(-2)^2)^2}, \frac{960}{(1+1^2+2(1)^2+3(-2)^2)^2} \right\rangle \\
&= \left\langle -\frac{160}{16^2}, -\frac{320}{16^2}, \frac{960}{16^2} \right\rangle = \frac{5}{8}\langle -1, -2, 6\rangle
\end{aligned}$$

This is the direction of greatest ascent. The norm of it is that value of the directional derivative in that direction.

$$\|\nabla T(1,1,-2)\| = \frac{5}{8}\sqrt{(-1)^2 + (-2)^2 + 6^2} = \frac{5\sqrt{41}}{8}$$

This can be interpreted as the maximal rate of change of temperature per meter. If you are moving at v meters per second, the maximum rate of change will be the product of these two numbers.

So the maximum rate of increase of temperature is $5v\sqrt{41}/8 \approx 4v\,^\circ\text{Cs}^{-1}$. △

3.5.4 Tangent Planes to Level Surfaces

Suppose S is a surface with equation $F(x_1, x_2, \ldots, x_n) = k$, that is, it is a level surface of a function F of n variables. Let $P(p_1, p_2, \ldots, p_n)$ be a point on S, and let C be any curve that lies on S and passes through P. C can be parameterized by some vector function $\boldsymbol{r}(t) = \langle r_1(t), r_2(t), \ldots, r_n(t)\rangle$. Since C lies on S, any point $\boldsymbol{r}$ must satisfy the equation

$$F(\boldsymbol{r}) = k$$

If r_1, r_2, and so on are all differentiable functions of t and F is also differentiable, we can use the chain rule to differentiate both sides with respect to t,

$$\frac{\partial F}{\partial r_1}\frac{\mathrm{d}r_1}{\mathrm{d}t} + \frac{\partial F}{\partial r_2}\frac{\mathrm{d}r_2}{\mathrm{d}t} + \cdots + \frac{\partial F}{\partial r_n}\frac{\mathrm{d}r_n}{\mathrm{d}t} = 0$$

Which can be rewritten as

$$\nabla F(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) = 0$$

We can interpret this to say that the gradient vector at any point on a surface is orthogonal to the tangent vector of the curves that lies on that surface and pass through P. In this case, it tells us that the gradient vector of a surface is orthogonal to the tangent vector to level curve at a point shared between them. We can then define the **tangent plane to the level surface** $F(\boldsymbol{x}) = k$ **at** $P(\boldsymbol{p})$ as the plane that passes through P and has the normal vector $\nabla F(\boldsymbol{x})$. In equation form,

$$F_{x_1}(\boldsymbol{x})(x_1 - p_1) + F_{x_2}(\boldsymbol{x})(x_2 - p_2) + \cdots + F_{x_n}(\boldsymbol{x})(x_n - p_n) = 0$$

The **normal line** to S at P is the line that passes through P and is orthogonal to the tangent plane. So the direction of this line is given by $\nabla F(\boldsymbol{x})$ and has symmetric equations

$$\frac{x_1 - p_1}{F_{x_1}(\boldsymbol{x})} = \frac{x_2 - p_2}{F_{x_2}(\boldsymbol{x})} = \cdots = \frac{x_n - p_n}{F_{x_n}(\boldsymbol{x})}$$

3.5.5 Gradient Descent

Gradient descent is most well known in the context of machine learning applications, but it is a purely mathematical method of optimization at its core. Consider some multivariable function $f(\boldsymbol{x})$ and a particular point $\boldsymbol{a}_0$ on the domain of f. If we want to minimize the value of f, one way of doing that is via the method of gradient descent.

First, define some positive constant γ, which is called the *learning rate*. Then, because we know that the direction of fastest descent is in the direction of $-\nabla f$, we can let $\boldsymbol{a}_1 = \boldsymbol{a}_0 - \gamma \nabla f(\boldsymbol{a}_0)$. Subtracting the $\gamma \nabla f(\boldsymbol{a}_0)$ term will move us towards a local minimum of f. If we run this process repeatedly, we will (hopefully) converge to some local minimum. This will not always occur, due to potentially overshooting the minimum value with our step.

The nuances of picking an ideal learning rate (or an adaptable rate, in some cases) is more fit for a higher-level computer science class, but the basic ideas still apply.

To connect this to machine learning, consider some function $L(\boldsymbol{x})$, where L outputs the loss (error) of a layer in a neural network, and $\boldsymbol{x}$ represents the "settings" of that layer. Gradient descent will bring us towards the local minimums of L, and will (hopefully) give the ideal, or close to ideal, settings vector $\boldsymbol{x}$ to minimize the loss $L(\boldsymbol{x})$.

3.6 Maximum and Minimum Values

One of the primary applications of single-variable calculus is finding the maximum and minimum values of functions in $\mathbb{R}^2$. With the tools we've learned since then, we are now able to find the maximum and minimum values of multivariable functions.

Definition 3.6.1. A continuous function $f : \mathbb{R}^n \to \mathbb{R}$ has a **local maximum** at $\boldsymbol{a}$ if $f(\boldsymbol{x}) \leq f(\boldsymbol{a})$ for all $\boldsymbol{x}$ satisfying $\|\boldsymbol{x} - \boldsymbol{a}\| \leq \delta$ for some $\delta > 0$.

f has a **local minimum** at $\boldsymbol{b}$ if $f(\boldsymbol{x}) \geq f(\boldsymbol{b})$ for all $\boldsymbol{x}$ satisfying $\|\boldsymbol{x} - \boldsymbol{b}\| \leq \delta$ for some $\delta > 0$.

This definition can be interpreted as saying that $\boldsymbol{a}$ and $\boldsymbol{b}$ give the largest and smallest values of f for some disk/sphere/hypersphere around them.

Another theorem will also come in handy.

Theorem 3.6.2. If f has a local maximum or minimum at $\boldsymbol{a}$ and the first order partial derivatives of f exist there, then $\nabla f = \boldsymbol{0}$.

Proof. Let $g_i(x_i) = f(a_1, a_2, \ldots, x_i, \ldots, a_n)$ for all $1 \leq i \leq n$. If f has a local extremum at $\boldsymbol{x} = \boldsymbol{a}$, then g_i must as well at $x_i = a_i$, so $g_i'(a_i) = 0$. We can differentiate g_i to find that $g_i'(x_i) = f_{x_i}(a_1, a_2, \ldots, x_i, \ldots, a_n)$. Then, $g_i'(a_i) = 0$, and therefore $f_{x_i}(\boldsymbol{a}) = 0$ for all integers i on $[1, n]$. In other words, $\nabla f = 0$. $\square$

Let's consider the implications of this theorem on the tangent plane to surfaces in $\mathbb{R}^3$. If $f : \mathbb{R}^2 \to \mathbb{R}$ has a local extremum at (a, b), then $f_x(a, b) = f_y(a, b) = 0$. Then, the tangent plane to f at (a, b),

$$f_x(a, b)(x - a) + f_y(a, b)(y - a) + z - f(a, b)$$

can be rewritten as

$$z = f(a, b)$$

This plane will be flat in the xy plane, which matches with our intuition for what a local extremum should look like (similar to how the tangent line to local extrema in single-variable calculus is a horizontal line).

A point $\boldsymbol{x}$ is called a **critical point** (or *stationary point*) of f if $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$ or if any of the partial derivatives of f does not exist.

Example 3.6.3. Let $f(x, y) = x^2 + y^2 - 2x - 6y + 14$. Then,

$$f_x(x, y) = 2x - 2 \quad \text{and} \quad f_y(x, y) = 2y - 6$$

These partial derivatives are only both zero at $(x, y) = (1, 3)$, so that is the only critical point of f. Further, we can complete the square to rewrite f as

$$f(x, y) = (x - 1)^2 + (y - 3)^2 + 4$$

because $(x - 1)^2$ and $(y - 3)^2$ are both strictly increasing as x and y get further from $(1, 3)$, this point is a local minimum. In fact, this will be the absolute minimum of f, which turns out to be an elliptic paraboloid with vertex $(1, 3, 4)$. △

Example 3.6.4. Find the extreme values of $f(x, y) = y^2 - x^2$.

Solution Because $f_x = -2x$ and $f_y = 2y$, the only critical point is $(x, y) = (0, 0)$. Notice that if we affix $x = 0$, f is strictly increasing with y getting further from 0. However, affixing $y = 0$ shows us that f is strictly decreasing with x getting further from 0. This type of point is neither an absolute minimum nor maximum. In fact, it will be a *saddle point*. △

Saddle points, named after their resemblance to saddles, are any stationary points that are not local extrema.

In order to quickly determine what points are relative minima, maxima, or saddle points, we can use an extended version of the second derivative test from single-variable calculus.

Theorem 3.6.5. Suppose the second partial derivatives of some function $f : \mathbb{R}^n \to \mathbb{R}$ are continuous on a hypersphere with center $\boldsymbol{x}$, and suppose that $\nabla f = \mathbf{0}$. Let

$$H_f(\boldsymbol{x}) = \begin{bmatrix} f_{x_1x_1}(\boldsymbol{x}) & f_{x_1x_2}(\boldsymbol{x}) & \cdots & f_{x_1x_n}(\boldsymbol{x}) \\ f_{x_2x_1}(\boldsymbol{x}) & f_{x_2x_2}(\boldsymbol{x}) & \cdots & f_{x_2x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(\boldsymbol{x}) & f_{x_nx_2}(\boldsymbol{x}) & \cdots & f_{x_nx_n}(\boldsymbol{x}) \end{bmatrix}$$

This matrix is known as the *hessian matrix*. Then,

1. If $H_f(\boldsymbol{x})$ is positive definite (has all positive eigenvalues), then f has a local minimum at $\boldsymbol{x}$.
2. If $H_f(\boldsymbol{x})$ is negative definite (has all negative eigenvalues), then f has a local maximum at $\boldsymbol{x}$.
3. If $H_f(\boldsymbol{x})$ has both positive and negative eigenvalues, then f has a saddle point at $\boldsymbol{x}$.

If none of these conditions are met, the test fails, and we must use another method.

In the special case of two variables, the hessian becomes

$$H_f(a, b) = \begin{bmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{bmatrix}$$

and the conditions simplify. Define $D = \det H_f(a, b)$. Then,

1. If $D > 0$ and $f_{xx} > 0$, f has a local minimum at (a, b).
2. If $D > 0$ and $f_{xx} < 0$, f has a local maximum at (a, b).
3. If $D < 0$, f has a saddle point at (a, b).

If none of these conditions are met, the test fails, and we must use another method.

Now, let's justify the connection between these two. First, assume f and all of f's first partial derivatives are differentiable. Then, compute the eigenvalues of H:

$$\begin{aligned} \det(H_f - \mathbb{I}_2\lambda) &= \begin{vmatrix} f_{xx} - \lambda & f_{xy} \\ f_{yx} & f_{yy} - \lambda \end{vmatrix} \\ &= \lambda^2 - \lambda(f_{xx} + f_{yy}) - f_{xy}^2 + f_{xx}f_{yy} \end{aligned}$$

The quadratic formula tells us

$$\lambda = \frac{f_{xx} + f_{yy} \pm \sqrt{(f_{xx} + f_{yy})^2 - 4(-f_{xy}^2 + f_{xx}f_{yy})}}{2}$$
$$2\lambda = f_{xx} + f_{yy} \pm \sqrt{(f_{xx} + f_{yy})^2 - 4(f_{xx}f_{yy} - f_{xy}^2)}$$

If $D > 0$, we can guarantee that $f_{xx}f_{yy} - f_{xy}^2 > 0$. This gives us two results. First, we know that f_{xx} and f_{yy} are must have the same sign, because the product $f_{xx}f_{yy}$ must be positive. Then, we can also rewrite

$$2\lambda = f_{xx} + f_{yy} \pm \sqrt{(f_{xx} + f_{yy})^2 - c}$$

For some *positive* constant c. Note that

$$|f_{xx} + f_{yy}| > \sqrt{(f_{xx} + f_{yy})^2 - c} \tag{3.6.1}$$

for any positive c. Therefore, in the first case where f_{xx} and f_{yy} are positive, we have

$$2\lambda = f_{xx} + f_{yy} \pm \sqrt{(f_{xx} + f_{yy})^2 - c}$$

The right side is guaranteed to be positive for both solutions, because the $\sqrt{(f_{xx} + f_{yy})^2 - c}$ term cannot subtract more than the $f_{xx} + f_{yy}$ term adds, as per Eq. 3.6.1.

Similarly, in the case where f_{xx} and f_{yy} are negative, we have

$$2\lambda = -|f_{xx}| - |f_{yy}| \pm \sqrt{(f_{xx} + f_{yy})^2 - c}$$

The right side is guaranteed to be negative for both solutions, because the $\sqrt{(f_{xx} + f_{yy})^2 - c}$ term cannot add more than the $-|f_{xx}| - |f_{yy}|$ term subtracts, as per Eq. 3.6.1.

Therefore, if $D > 0$ and $f_{xx} > 0$, both eigenvalues must be positive. If $D > 0$ and $f_{xx} < 0$, both eigenvalues must be negative.

Example 3.6.6. Find the local maximum and minimum values and saddle points of $f(x, y) = x^4 + y^4 - 4xy + 1$.

Solution: First, compute each partial:

$$f_x = 4x^3 - 4y \quad \text{and} \quad f_y = 4y^3 - 4x$$

Then, setting these equal to zero, we get

$$x^3 - y = 0 \quad \text{and} \quad y^3 - x = 0$$

The first equation can be rearranged to say $y = x^3$, which can then be plugged into the second equation,

$$\begin{aligned} 0 &= (x^3)^3 - x = x^9 - x \\ &= x(x^8 - 1) = x(x^4 - 1)(x^4 + 1) \\ &= x(x^2 - 1)(x^2 + 1)(x^4 + 1) \end{aligned}$$

So our roots are $x = -1, 0, 1$, which give us the three critical points $(-1, -1)$, $(0, 0)$, and $(1, 1)$. Now, let's calculate the second partials and hessian determinant,

$$f_{xx} = 12x^2 \quad f_{yy} = 12y^2 \quad f_{xy} = -4$$
$$D = f_{xx}f_{yy} - f_{xy}^2 = 144x^2y^2 - 16$$

Then check each of the three points:

1. $(-1,-1)$: $D(-1,-1) = 128$, $f_{xx}(-1,-1) = 12$. So $(-1,-1)$ is a local minimum.
2. $(0,0)$: $D(0,0) = -16$. So $(0,0)$ is a saddle point.
3. $(1,1)$: $D(1,1) = 128$, $f_{xx}(1,1) = 12$. So $(1,1)$ is a local minimum.

△

Example 3.6.7. Find and identify the critical points of the function

$$f(x,y) = 10x^2y - 5x^2 - 4y^2 - x^4 - 2y^4$$

Solution: First, taking each partial,

$$f_x = 20xy - 10x - 4x^3 \quad f_y = 10x^2 - 8y - 8y^3$$

And setting them equal to zero,

$$20xy - 10x - 4x^3 = 0 \quad 10x^2 - 8y - 8y^3 = 0$$

We can immediately see one solution, $x = y = 0$. To find the others, we can start by solving for y in the first equation,

$$y = \frac{4x^3 + 10x}{20x} = \frac{2x^2 + 5}{10}$$

and then substituting it back into the second one,

$$10x^2 - 8\left[\frac{2x^2+5}{10}\right] - 8\left[\frac{2x^2+5}{10}\right]^3 = 0$$

Solving by calculator, we find the solutions to be $x = \pm 0.857, \pm 2.644$. Then, substitute that back into the equation for y and find the points to be $(-2.644, 1.898)$, $(-0.857, 0.647)$, $(0.857, 0.647)$, and $(-2.644, 1.898)$.

Now, taking the second partials,

$$f_{xx} = 20y - 10 - 12x^2 \quad f_{yy} = -8 - 24y^2 \quad f_{xy} = 20x$$

and calculating the hessian determinant,

$$D = f_{xx}f_{yy} - f_{xy}^2 = (-12x^2 + 20y - 10)(-8 - 24y^2) - (20x)^2$$

1. $(0,0)$: $D = 80$, and $f_{xx} = -10$, so it is a relative maximum.
2. $(\pm 2.644, 1.898)$: $D \approx 2486.614$, and $f_{xx} \approx -55.929$, so they are relative maximums
3. $(\pm 0.857, 0.647)$: $D \approx -187.785$, so they are saddle points.

△

Example 3.6.8. Find the shortest distance from the point $P(1,0,-2)$ to the plane $x + 2y + z = 4$.

Solution: First, because P does not satisfy the equation $x + 2y + z = 4$, it isn't on the plane. Now, let $Q(x,y,z)$ be a point on the plane. Then, the distance between Q and P is given by

$$\ell = \sqrt{(x-1)^2 + y^2 + (z+2)^2}$$

We can rewrite the equation of the plane as $z = 4 - x - 2y$, and substitute that for z in the equation for ℓ:

$$\ell = \sqrt{(x-1)^2 + y^2 + (6 - x - 2y)^2}$$

Now, to find the shortest value of ℓ, we need to find its absolute minimum. Our domain is $\mathbb{R}^3$, which normally presents an issue because it is an unbounded set (and thus does not always have absolute extrema). However, in this case, we can use our intuition for how distance works to say that there will definitely be an absolute minimum.

To find critical points, we can first redefine $f = \ell^2 = (x-1)^2 + y^2 + (6-x-2y)^2$. f will have the same critical points, but is simpler to differentiate.

$$f_x = 2(x-1) - 2(6-x-2y) = 4x + 4y - 14$$
$$f_y = 2y - 4(6-x-2y) = 10y + 4x - 24$$

These are both zero when the equations $10y+4x-24=0$ and $4x+4y-14=0$ are both satisfied. Rewriting this as a matrix,

$$\begin{bmatrix} 4 & 4 \\ 4 & 10 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 14 \\ 24 \end{bmatrix}$$

Which we can solve by row reduction or inverse matrices to find that the only critical point is $\left(\frac{11}{6}, \frac{5}{3}\right)$. Now, we can compute each second partial,

$$f_{xx} = 4 \quad f_{yy} = 10 \quad f_{xy} = 4$$

So the hessian determinant $f_{xx}f_{yy} - f_{xy}^2 = 26 > 0$ and $f_{xx} > 0$, so we have a local minimum, which also happens to be the absolute minimum.

To find the value of the smallest distance, simply plug in the point $\left(\frac{11}{6}, \frac{5}{3}\right)$ into ℓ,

$$\ell = \sqrt{\left(\frac{11}{6} - 1\right)^2 + \left(\frac{5}{3}\right)^2 + \left(6 - \frac{11}{6} - \frac{10}{3}\right)^2} = \frac{5\sqrt{6}}{6}$$

△

Example 3.6.9. A rectangular box with no lid is to be made from 12 m^2 of cardboard. Find the maximum volume of such a box.

Solution: Let x be the length of the box, y be the width of the box, and z be the height of the box. Then, the volume is given by $V = xyz$ and the surface area is given by $S = xy + 2xz + 2yz$. This must be equal to 12, so

$$xy + 2xz + 2yz = 12$$

Which can be rearranged to say

$$z = \frac{12 - xy}{2x + 2y}$$

Plugging this back into the expression for V,

$$V = xy\left[\frac{12 - xy}{2x + 2y}\right] = \frac{12xy - x^2y^2}{2x + 2y}$$

We can now find the partials of this.

$$\begin{aligned}
\frac{\partial V}{\partial x} &= \frac{(2x+2y)(12y - 2xy^2) - (12xy - x^2y^2)(2)}{(2x+2y)^2} \\
&= \frac{24xy + 24y^2 - 4x^2y^2 - 4xy^3 - 24xy + 2x^2y^2}{(2x+2y)^2} \\
&= \frac{-4xy^3 - 2x^2y^2 + 24y^2}{(2x+2y)^2} = \frac{-2y^2(x^2 + 2xy - 12)}{(2x+2y)^2} \\
\frac{\partial V}{\partial y} &= \frac{(2x+2y)(12x - 2yx^2) - (12xy - x^2y^2)(2)}{(2x+2y)^2} \\
&= \frac{24x^2 + 24xy - 4x^3y - 4x^2y^2 - 24xy + 2x^2y^2}{(2x+2y)^2} \\
&= \frac{-4x^3y - 2x^2y^2 + 24x^2}{(2x+2y)^2} = \frac{-2x^2(y^2 + 2xy - 12)}{(2x+2y)^2}
\end{aligned}$$

The first equation is zero when $y = 0$ or $x^2 + 2xy - 12 = 0$. The second equation is zero when $x = 0$ or $y^2 + 2xy - 12 = 0$. Notice that these two equations imply $x = y$, so we can substitute that in, and say

$$x^2 + 2x^2 - 12 = 0 \implies x = y = 2$$

Note that $x = y = 0$ is not a valid solution because V_x and V_y are not continuous there. Then, we can find the z,

$$z = \frac{12 - xy}{2x + 2y} = 1$$

We can pretty reasonably assume that this will be an absolute maximum due to the physical implication of this problem. $\triangle$

3.7 Absolute Minimum and Maximum Values

Previously, we've been kind of handwaving the issue of finding absolute extrema by using physical interpretations of problems. Now, we will introduce a mathematical way of finding the absolute extrema of functions.

First, we will introduce a multivariable equivalent of the extreme value theorem from single-variable calculus.

Theorem 3.7.1 (Extreme Value Theorem)**.** If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous on a closed, bounded set $D \subseteq \mathbb{R}^n$, f must have some absolute maximum and minimum value on D.

To find these extreme values, we must check not only the critical values of f, but also all points on the boundary of D. We can write this boundary of D as the set ∂D.

Definition 3.7.2. If X is a topological space, and $S \subseteq X$ is a set, then the **boundary set** ∂S is defined by

$$\partial S := \{p \in X : \text{for every neighborhood } O \text{ of } p, O \cap S \neq \emptyset, O \cap (X \setminus S) \neq \emptyset\}$$

This can be understood to say that the boundary set is the set of all elements that are infinitely close to an element of S and also infinitely close to an element of X that is not in S.

Example 3.7.3. Find the absolute maximum and minimum values of $f(x, y) = x^2 - 2xy + 2y$ on the rectangle $D = \{(x, y) | 0 \leq x \leq 3, 0 \leq y \leq 2\}$.

Solution: First, let's find the partials of f,

$$f_x = 2x - 2y \quad f_y = -2x + 2$$

So $(x, y) = (1, 1)$ is the only critical point, which will have a value of $f(1, 1) = 1$. We can break up the boundary set ∂D into four line segments.

1. $x = 0$ and $y \in [0, 2]$
2. $x = 3$ and $y \in [0, 2]$
3. $y = 0$ and $x \in [0, 3]$
4. $y = 2$ and $x \in [0, 3]$

Describing the behavior of f on these segments, we get

1. $f(0, y) = 2y$, $y \in [0, 2]$
2. $f(3, y) = 9 - 4y$, $y \in [0, 2]$
3. $f(x, 0) = x^2$, $x \in [0, 3]$
4. $f(x, 2) = x^2 - 4x + 4$, $x \in [0, 3]$

The absolute maximums of these four segments are $f(0,2) = 4$, $f(3,0) = 9$, $f(3,0) = 9$, and $f(0,2) = 4$.

The absolute minimums of these four segments are $f(0,0) = 0$, $f(3,2) = 1$, $f(0,0) = 0$, and $f(2,2) = 0$.

Therefore, the absolute minimum of f is the smallest value of f of these 10, which will be $f(0,0) = 0$. Similarly, the absolute maximum of f is the largest value of f of these 10, which will be $f(3,0) = 9$. △

3.8 Lagrange Multipliers

Lagrange multipliers give us an easier way to maximize or minimize a function subject to a constraint.

Consider some function $f(x,y,z)$ and some constraint $g(x,y,z) = k$. Geometrically, this is like trying to find the maximum value of f on the level surface described by $g(x,y,z) = k$. This can also be interepreted as finding the largest value c such that the level surfaec $f(x,y,z) = c$ intersects the level surface $g(x,y,z) = k$. We can reason that if we have reaches the highest c-value, then the two surfaces will be tangent to each other. Otherwise, c could be increased further. If these two surfaces are tangent, then their normal lines are identical at the point of tangency. Therefore, their gradient vectors must be parallel, which gives us the equation $\nabla f = \lambda \nabla g$ for some scalar λ.

The λ in this equation is known as the **lagrange multiplier**. The procedure to use the lagrange multiplier is as follows:

First, find all values of $\boldsymbol{x}$, and λ such that

$$\nabla f(\boldsymbol{x}) = \lambda \nabla g(\boldsymbol{x}) \quad \text{and} \quad g(\boldsymbol{x}) = k$$

Then, evaluate f at each of these points. The largest of these values is the absolute maximum of f constrained by g, and the smallest of these values is the absolute minimum of f constrained by g.

In the special case of functions of three variables, we can write $\nabla f = \lambda \nabla g$ as a system of four equations

$$f_x = \lambda g_x, \quad f_y = \lambda g_y, \quad f_z = \lambda g_z, \quad g(x,y,z) = k$$

Example 3.8.1. We will repeat an example from earlier, but now using the method of lagrange multipliers.

If a box with no top is to be made from 12 m^2 of cardboard, find the maximum volume.

Solution: First, write the equations $V = xyz$ and $S = 12 = xy + 2xz + 2yz$. Our constraint is S and our function is V, so we have

$$\nabla V = \lambda \nabla S$$

which becomes

$$\begin{bmatrix} yz \\ xz \\ xy \end{bmatrix} = \lambda \begin{bmatrix} y + 2z \\ x + 2z \\ 2x + 2y \end{bmatrix} \tag{3.8.1}$$

This, along with $S = 12$, gives us a system of four equations. There are no general rules for solving systems of (nonlinear) equations such as this one, so we might have to have some ingenuity. First, note that we can modify equation 3.8.1 to say

$$\begin{aligned} xyz &= \lambda(xy + 2xz) \\ xyz &= \lambda(xy + 2yz) \\ xyz &= \lambda(2xz + 2yz) \end{aligned}$$

Note that $xy + 2xz + 2yz$ necessitates that $\lambda \neq 0$ because that would cause $x = y = z = 0$.

Now, we can see from the first and second of the above equations that

$$xy + 2xz = xy + 2yz$$

which tells us that $xz = yz$, which implies $x = y$ since $z \neq 0$ (since this would give $V = 0$). Therefore, we can equate the second and third equations in our system to give

$$xy + 2yz = 2xz + 2yz$$

which we can substitute $x = y$ into to find

$$x^2 + 2xz = 4xz$$

and $x(x - 2z) = 0$, so $x = 2z$. We throw away $x = 0$ because that would give $V = 0$. So we can take our constraint $xy + 2xz + 2yz$ and substitute $x = y = 2z$ to find

$$12z^2 = 12$$

So $z = 1$. Then $x = y = 2z$, so our absolute maximum is given by $(x, y, z) = (2, 2, 1)$. △

Example 3.8.2. Find the extrema of $f(x, y) = x^2 + 2y^2$ subject to $x^2 + y^2 = 1$.

Solution: First, write out the lagrange multiplier system:

$$2x = \lambda 2x \tag{3.8.2}$$
$$4y = 2\lambda y \tag{3.8.3}$$
$$x^2 + y^2 = 1 \tag{3.8.4}$$

From 3.8.2, we can see that $\lambda = 1$ or $x = 0$. From 3.8.3, we can see that $\lambda = 2$ or $y = 0$. From 3.8.4, we see that if $x = 0$, then $y = \pm 1$ and if $y = 0$, then $y = \pm 1$. So we have four possible points that satisfy the system, $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$. Plugging each of these into f, we find

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

So our two absolute maximums are at $(x, y) = (0, \pm 1)$ and our two absolute minimum are at $(x, y) = (\pm 1, 0)$. △

Example 3.8.3. Find the extrema of $f(x, y) = x^2 + 2y^2$ subject to $x^2 + y^2 \leq 1$.

Solution: This is similar to the previous problem, but with a slight variation in the constraint. Now, the system becomes

$$2x = \lambda 2x \tag{3.8.5}$$
$$4y = 2\lambda y \tag{3.8.6}$$
$$x^2 + y^2 \leq 1 \tag{3.8.7}$$

Which is now satisfied by the points $P = \{(x, 0) | 0 \leq x \leq 1\} \cup \{(0, y) | 0 \leq y \leq 1\}$. We now want to compare both the points at the boundary of the constraint and all critical points of f.

Because $\nabla f = \langle 2x, 4y \rangle$, the only critical point is $f(0, 0) = 0$. The points on the boundary have already been evaluated in the previous example (because checking the boundary is equivalent to changing the constraint to $x^2 + y^2 = 1$, which we just examined).

Therefore, the absolute maximum is still $f(0, \pm 1) = 2$, but the absolute minimum is now $f(0, 0) = 0$. △

Example 3.8.4. Find the points on the sphere $x^2 + y^2 + z^2 = 4$ that are closest and farthest from the point $P(3, 1, -1)$.

Solution: Consider some arbitary point $\boldsymbol{v} = \langle x, y, z \rangle$. Then, the distance from $\boldsymbol{v}$ to P is

$$d = \sqrt{(x - 3)^2 + (y - 1)^2 + (z + 1)^2}$$

To make the math simpler, let's consider $f = d^2$, so

$$f(x, y, z) = (x-3)^2 + (y-1)^2 + (z+1)^2$$

Now, the gradient of f is given by

$$\nabla f = \langle 2(x-3), 2(y-1), 2(z+1) \rangle$$

The gradient of g is

$$\nabla g = \langle 2x, 2y, 2z \rangle$$

So our system of equations is

$$\begin{aligned} x - 3 &= \lambda x \\ y - 1 &= \lambda y \\ z + 1 &= \lambda z \\ x^2 + y^2 + z^2 &= 4 \end{aligned}$$

Rearranging the first three equations, we can equate them by isolating λ,

$$\frac{x-3}{x} = \frac{y-1}{y} = \frac{z+1}{z}$$

Which can all be rearranged to say

$$1 - \frac{3}{x} = 1 - \frac{1}{y} = 1 + \frac{1}{z}$$

Then, we can express y and z in terms of x:

$$\begin{aligned} y &= \frac{x}{3} \\ z &= -\frac{x}{3} \end{aligned}$$

Which we can substistute into equation 4,

$$\begin{aligned} 4 &= x^2 + \left(\frac{x}{3}\right)^2 + \left(-\frac{x}{3}\right)^2 \\ &= \frac{11}{9}x^2 \\ x &= \pm\frac{6}{\sqrt{11}} \end{aligned}$$

Which gives us two points,

$$\boldsymbol{r}_1 = \left\langle \frac{6}{\sqrt{11}}, \frac{2}{\sqrt{11}}, -\frac{2}{\sqrt{11}} \right\rangle \quad \text{and} \quad \boldsymbol{r}_2 = \left\langle -\frac{6}{\sqrt{11}}, -\frac{2}{\sqrt{11}}, \frac{2}{\sqrt{11}} \right\rangle$$

Clearly, $f(\boldsymbol{r}_2)$ will give the greater result, so $f(\boldsymbol{r}_1)$ is an absolute minimum (shortest distance) and $f(\boldsymbol{r}_2)$ is an absolute maximum (largest distance). △

3.8.1 Several Constraints

Suppose we want to find the extrema of a function $f : \mathbb{R}^n \to \mathbb{R}$ subject to m (assuming $m < n$) constraints of the form $g_i(\boldsymbol{x}) = k_i$, with each g_i taking an n-dimensional input and returning a scalar output. Geometrically, this means that we are looking for the minimum and maximum values of f on the surface formed by the intersection of each of the level surfaces formed by $g_1(\boldsymbol{x}) = k_1$, $g_2(\boldsymbol{x}) = k_2$, …, $g_m(\boldsymbol{x}) = k_m$. Let's call the surface formed by all of these intersections C.

Suppose that f has an extreme value at a point $\boldsymbol{u} \in \mathbb{R}^n$. We know that $\nabla f(\boldsymbol{u})$ must be orthogonal to the level surface $f(\boldsymbol{x}) = c$, which is tangent to C, so ∇f must be orthogonal to C at P. By the same logic, $\nabla g_i(\boldsymbol{u})$ is also orthogonal to C for all i. Then, f lies in the m dimensional subspace with basis elements $\nabla g_1(\boldsymbol{x}), \nabla g_2(\boldsymbol{x}), \ldots, \nabla g_m(\boldsymbol{x})$, and can be written as a linear combination of them,

$$\nabla f(\boldsymbol{u}) = \lambda_1 \nabla g_1(\boldsymbol{u}) + \lambda_2 \nabla g_2(\boldsymbol{u}) + \cdots + \lambda_m \nabla g_m(\boldsymbol{u}) \tag{3.8.8}$$

Where each λ_i is a constant.

This generates a system of $n+m$ equations: n from each component of 3.8.8, and m from the restrictions $g_i(\boldsymbol{x}) = k_i$.

Solving this system gives us the values of $\boldsymbol{x}$ that are candidates for the extrema of f subject to the m restrictions formed by $g_i = k_i$.

4 Multiple Integration

Chapter 15: Page 981

4.1 Double Integrals over Rectangles

Recall that definite integrals told us the area underneath the curve of a single-variable function between two boundary points a and b. Similarly, definite double integrals tell use the volume underneath the surface of a two-variable function along a boundary rectangle, defined by two points–a lower and upper x bound along with a lower and upper y bound.

This section assumes you understand single-variable definite integrals, and have a working knowledge of the definition of the integral

$$\int_a^b f(x)\mathrm{d}x = \lim_{n\to\infty} \sum_{i=1}^{n} f\left(a + \frac{i(b-a)}{n}\right)\frac{b-a}{n}$$

4.1.1 Volume and Double Integrals

The definition of the definite double integral is similar to the definition of the definite single integral,

$$\int_a^b \int_c^d f(x,y)\mathrm{d}y\mathrm{d}x = \lim_{n,m\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{m} f\left(a + i\frac{b-a}{n}, c + j\frac{d-c}{m}\right)\frac{b-a}{n}\frac{d-c}{m}$$

This represents finding the volume of infinitely many infinitesimal Riemann rectangular prism and adding them all up, with two summations, one having us crawl parallel to the x axis and the other having us crawl parallel to the y axis.

If we define the R as the rectangle $\{(x,y) | a \le x \le b, c \le y \le d\}$, we can more concisely write the double integral as

$$\iint\limits_R f(x,y)\mathrm{d}A = \lim_{n,m\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{m} f(x_j, y_i)\Delta A$$

Compute this sum with finite n, m is a way of approximating the volume underneath a curve in $\mathbb{R}^3$. We can also define similar definitions to compute more than two definite integrals, to integrate over curves in $\mathbb{R}^n$.

Example 4.1.1. Estimate the volume the lies above the square $R = [0,2] \times [0,2]$ and above the elliptic paraboloid $f(x,y) = 16 - x^2 - 2y^2$ by dividing R into four equal squares, with the sample points being the upper-right corner of each square.

Solution: Our four samples points are $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$. The values of f at these points are

$$f(1,1) = 13$$
$$f(1,2) = 7$$
$$f(2,1) = 10$$
$$f(2,2) = 4$$

And each $\Delta A = \Delta x \Delta y$ is $1 \cdot 1 = 1$. Therefore, the approximate volume is

$$V \approx f(1,1)(1) + f(1,2)(1) + f(2,1)(1) + f(2,2)(1) = 34$$

The actual number for V (which we will be able to compute soon) is 48, so this approximation isn't great. However, by increasing the number of squares, we can improve the accuracy. For example, with $m = n = 16$, the approximation becomes $V \approx 46.469$, which is much better. $\triangle$

Example 4.1.2. If $R = \{(x,y)|x \in [-1,1], y \in [-2,2]\}$, evaluate the integral

$$\iint_R \sqrt{1-x^2}\mathrm{d}A$$

Actually evaluating this integral is extremely difficult with our current tools, so we can instead look to interpret it geometrically. $z = \sqrt{1-x^2}$ can be rearranged to say $z^2 + x^2 = 1$ with $z \geq 0$, which is the equation describing a half-cylinder running along the y axis with radius 1.

So the integral can be interpreted as the volume of a half-cylinder with radius 1 and height 4,

$$\iint_R \sqrt{1-x^2}\mathrm{d}A = \frac{1}{2}\pi r^2 h = 2\pi$$

$\triangle$

4.1.2 The Midpoint Rule

When we want to approximate the volume of a double integral via a Riemann sum, we have to pick where we want our points to evaluate f at. For example, in the previous section, we chose the top-right corner of each interval. Another way of doing this is by choosing the midpoint of the interval.

Example 4.1.3. Use the Midpoint Rule with $m = n = 2$ to estimate the value of the integral $\iint_R (x - 3y^2)\mathrm{d}A$, where $R = \{(x,y)|x \in [0,2], y \in [1,2]\}$.

Solution: The midpoints of the x intervals are 0.5 and 1.5, while the midpoints of the y intervals are 1.25 and 1.75. The area of each sub-rectangle is $\Delta A = \Delta x \Delta y = 0.5$. So we can approximate the integral as

$$\begin{aligned}\iint_R (x-3y^2)\mathrm{d}A &\approx 0.5[f(0.5,1.25) + f(1.5,1.25) + f(0.5,1.75) + f(1.5,1.75)] \\ &= -11.875\end{aligned}$$

$\triangle$

4.1.3 Average Value

Recall that the average value of a single-variable function f on an interval $[a,b]$ is defined as

$$f_{\text{avg}} = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x$$

Similarly, the average value of a function of two-variables f on a rectangle R is

$$f_{\text{avg}} = \frac{1}{A(R)} \iint\limits_R f(x,y)\mathrm{d}A$$

where $A(R)$ is the area of the rectangle. If $R = \{(x,y)|x \in [a,b], y \in [c,d]\}$, $A(R) = (b-a)(d-c)$.

4.1.4 Properties of the Double Integral

I won't present a proof for these properties because I don't feel like writing it. But if you want to do it yourself, just use the definition of the integral and rearrange the terms a bit

Theorem 4.1.4 (Linearity of the Double Integral)**.** If $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ are both defined and integrable on a rectangle R, and $c \in \mathbb{R}$ is a constant, then

$$\iint\limits_R [f(x,y) + g(x,y)]\mathrm{d}A = \iint\limits_R f(x,y)\mathrm{d}A + \iint\limits_R g(x,y)\mathrm{d}A$$

and

$$\iint\limits_R cf(x,y)\mathrm{d}A = c\iint\limits_R f(x,y)\mathrm{d}A$$

Theorem 4.1.5. If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are both defined and integrable on a region R, and $f(\boldsymbol{x}) \geq g(\boldsymbol{x})$ for all $\boldsymbol{x} \in R$,

$$\int\limits_R f(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x} \geq \int\limits_R g(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x}$$

4.2 Iterated Integrals

Now let's learn how to actually compute these. Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ that is integrable on the rectangle $R = [a,b] \times [c,d]$. Then, the double integral can be written as

$$\int_a^b \int_c^d f(x,y)\mathrm{d}y\mathrm{d}x = \int_a^b \left[\int_c^d f(x,y)\mathrm{d}y\right]\mathrm{d}x$$

So we first integrate the inner function with respect to y, treating x as a constant. Then, we integrate that result with respect to x. This is an important result–in repeated integration, always work from the inside out.

Example 4.2.1. Evaluate the repeated integral

$$\int_0^3 \int_1^2 x^2y\mathrm{d}y\mathrm{d}x.$$

Solution: Integrate.

$$\begin{aligned} I = \int_0^3 \int_1^2 x^2y\mathrm{d}y\mathrm{d}x &= \int_0^3 \left[\frac{1}{2}x^2y^2\Big|_{y=1}^{y=2}\right]\mathrm{d}x \\ &= \int_0^3 \frac{3}{2}x^2\mathrm{d}x = \frac{1}{2}x^3\Big|_{x=0}^{x=3} = \frac{27}{2} \end{aligned}$$

△

Example 4.2.2. Evaluate the repeated integral

$$\int_1^2 \int_0^3 x^2 y \mathrm{d}x \mathrm{d}y$$

Solution: do the exact same thing (amazing!)

$$\begin{aligned} I = \int_1^2 \int_0^3 x^2 y \mathrm{d}x \mathrm{d}y &= \int_1^2 \left[\frac{1}{3} x^3 y \Big|_{x=0}^{x=3} \right] \mathrm{d}y \\ &= \int_1^2 9y \mathrm{d}y = \frac{9}{2} y^2 \Big|_{y=1}^{y=2} = \frac{27}{2} \end{aligned}$$

△

Notice that these two gave us the same thing. This is not just a coincidence.

Theorem 4.2.3 (Fubini's Theorem)**.** If f is continuous on the rectangle $R = \{(x, y)|x \in [a, b], y \in [c, d]\}$, then

$$\iint\limits_R f(x, y) \mathrm{d}A = \int_a^b \int_c^d f(x, y) \mathrm{d}y \mathrm{d}x = \int_c^d \int_a^b f(x, y) \mathrm{d}x \mathrm{d}y$$

The proof of this is very difficult and beyond the scope of this class, but rest assured that it's true.

Example 4.2.4. Evaluate the double integral $\iint\limits_R (x - 3y^2) \mathrm{d}A$, where $R = [0, 2] \times [1, 2]$.

Solution: Rewrite as a repeated integral,

$$\iint\limits_R (x - 3y^2) \mathrm{d}A = \int_0^2 \int_1^2 (x - 3y^2) \mathrm{d}y \mathrm{d}x$$

and evaluate.

$$\begin{aligned} \int_0^2 \int_1^2 (x - 3y^2) \mathrm{d}y \mathrm{d}x &= \int_0^2 \left[xy - y^3 \Big|_{y=1}^{y=2} \right] \mathrm{d}x \\ &= \int_0^2 x - 7 \mathrm{d}x = \frac{1}{2} x^2 - 7x \Big|_{x=0}^{x=2} = -12 \end{aligned}$$

△

Example 4.2.5. Evaluate $\iint\limits_R y \sin(xy) \mathrm{d}A$, where $R = [1, 2] \times [0\ \pi]$.

Solution: This is an example of Fubini's Theorem being extremely useful. By Fubini's theorem, this double integral can be written in one of two ways

$$\int_1^2 \int_0^\pi y \sin(xy) \mathrm{d}y \mathrm{d}x \quad \text{or} \quad \int_0^\pi \int_1^2 y \sin(xy) \mathrm{d}x \mathrm{d}y$$

The left repeated integral is much more difficult to evaluate than the right one, as it will require an application of integration by parts, and will just be generally more messy. Switching the order of the integrals will be an important problem-solving technique moving forward.

Continuing onwards with the easier repeated integral, we get

$$\begin{aligned} \int_0^\pi \int_1^2 y \sin(xy) \mathrm{d}x \mathrm{d}y &= \int_0^\pi \left[-\cos(xy) \Big|_{x=1}^{x=2} \right] \mathrm{d}y \\ &= \int_0^\pi [\cos y - \cos(2y)] \mathrm{d}y \\ &= \sin y - \frac{1}{2} \sin(2y) \Big|_{y=0}^{y=\pi} = 0 \end{aligned}$$

$\triangle$

Example 4.2.6. Find the volume of the solid S that is bounded by the elliptic paraboloid $x^2+2y^2+z=16$, the planes $x=2$ and $y=2$, and the three coordinate planes.

Solution: Our region of integration is the rectangle $R=[0,2]\times[0,2]$, and out function is $f(x,y)=16-x^2-2y^2$. Since this function is positive for all $(x,y)\in R$, the double integral $\iint\limits_R f(x,y)\mathrm{d}A$ gives the volume of S. We can write this as an iterated integral and solve,

$$\begin{aligned}\int_0^2\int_0^2[16-x^2-2y^2]\mathrm{d}x\mathrm{d}y &= \int_0^2\left[16x-\frac{1}{3}x^3-2y^2x\Big|_{x=0}^{x=2}\right]\\ &= \int_0^2\left[\frac{88}{3}-4y^2\right]\mathrm{d}y\\ &= \frac{88}{3}y-\frac{4}{3}y^3\Big|_{y=0}^{y=2}\\ &= 48\end{aligned}$$

$\triangle$

Theorem 4.2.7. Let $f:\mathbb{R}^2\to\mathbb{R}$ to be integrable on a rectangle $R=[a,b]\times[c,d]$, and let $f(x,y)=g(x)h(y)$ for all $(x,y)\in R$. Then, we can write

$$\iint\limits_R f(x,y)\mathrm{d}A=\left[\int_a^b g(x)\mathrm{d}x\right]\left[\int_c^d h(y)\mathrm{d}y\right]$$

Proof. Let

$$I=\iint\limits_R f(x,y)\mathrm{d}A$$

By Fubini's Theorem, we can then write this as

$$I=\int_a^b\int_c^d f(x,y)\mathrm{d}y\mathrm{d}x=\int_a^b\int_c^d g(x)h(y)\mathrm{d}y\mathrm{d}x$$

By the linearity of the integral, since $g(x)$ is constant, it can be pulled out. Then,

$$I=\int_a^b g(x)\int_c^d h(y)\mathrm{d}y\mathrm{d}x$$

Then, because the entire inner integral $\int_c^d h(y)\mathrm{d}y$ is constant under the outer integral (with x), it can be pulled out, leaving us with

$$I=\int_c^d h(y)\mathrm{d}y\int_a^b g(x)\mathrm{d}x$$

$\square$

Example 4.2.8. Evaluate $\int_0^{\pi/2}\int_0^{\pi/2}\sin x\cos y\mathrm{d}y\mathrm{d}x$.

Solution: Split into two integrals and evaluate.

$$\begin{aligned}\int_0^{\pi/2}\int_0^{\pi/2}\sin x\cos y\mathrm{d}y\mathrm{d}x &= \left[\int_0^{\pi/2}\sin x\mathrm{d}x\right]\left[\int_0^{\pi/2}\cos y\mathrm{d}y\right]\\ &= \left[-\cos x\Big|_0^{\pi/2}\right]\left[\sin y\Big|_0^{\pi/2}\right]=1\cdot 1=1\end{aligned}$$

$\triangle$

4.3 Double Integrals over General Regions

With single integrals, every region we could integrate over was an interval. However, with double integrals, we are not limited to only integrating over rectangles. Consider some arbitrary bounded region $D \subseteq \mathbb{R}^2$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be integrable for all $(x, y) \in D$. Then, define a rectangle $R \subseteq \mathbb{R}^2$ such that $D \subseteq R$. Then, we can define a new function $F : \mathbb{R}^2 \to \mathbb{R}$, where

$$F(x,y) = \begin{cases} f(x,y) & (x,y) \in D \\ 0 & (x,y) \notin D \end{cases}$$

So we can write

$$\iint\limits_D f(x,y)\mathrm{d}A = \iint\limits_R F(x,y)\mathrm{d}A$$

The right expression can be split into an iterated integral with Fubini's theorem. It is very likely that F will have discontinuities on ∂D. However, as long as f is "well-behaved" on D (i.e. the set of its discontinuities is of 2 dimensional Lebesgue measure zero), $\iint\limits_R F(x,y)\mathrm{d}A$ will still be defined.

We call D **type I** if it lies between the graphs of two continuous functions of x. That is, the bounds of y can be parameterized in terms of x, so

$$D = \{(x,y) | x \in [a,b], y \in [g_1(x), g_2(x)]\}.$$

To evaluate double integrals across type I regions, choose a rectangle $R = [a,b] \times [c,d]$ such that $c \leq g_1(x)$ and $d \geq g_2(x)$ for all $x \in [a,b]$. In other words, choose R such that $D \subseteq R$. Then, we can write

$$I = \iint\limits_D f(x,y)\mathrm{d}A = \iint\limits_R F(x,y)\mathrm{d}A = \int_a^b \int_c^d F(x,y)\mathrm{d}y\mathrm{d}x$$

Because $F(x,y) = 0$ if $(x,y) \notin D$, we can rewrite as

$$I = \int_a^b \int_{g_1(x)}^{g_2(x)} F(x,y)\mathrm{d}y\mathrm{d}x$$

Now, because we never leave D in our integration, we can once again rewrite, so:

$$\iint\limits_D f(x,y)\mathrm{d}A = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\mathrm{d}y\mathrm{d}x$$

Another possible region, **type II**, arises when D lies between the graphs of continuous functions of f, and the bounds of x can be parameterized in terms of y. Let

$$D = \{(x,y) | x \in [h_1(y), h_2(y)], y \in [a,b]\}$$

The derivation for this is similar to that of type I, so we will skip it and go straight to the result:

$$\iint\limits_D f(x,y)\mathrm{d}A = \int_a^b \int_{h_1(y)}^{h_2(y)} f(x,y)\mathrm{d}x\mathrm{d}y$$

We can create a more general version of this relationship:

Theorem 4.3.1. Let X be a topological space with $\dim X = n$, and D be a bounded subset of X. Suppose $f : X \to \mathbb{R}$ is integrable on D and that D can be parameterized in terms of some x_i, where i is an integer satisfying $i \in [1, n-1]$. That is,

$$D = \{(x_1, x_2, \ldots, x_i, \ldots, x_n) | x_1 \in [f_1(x_i), g_1(x_i)], \ldots, x_i \in [a,b], \ldots, x_n \in [f_n(x_i), g_n(x_i)]\}$$

Then, we have

$$\int\limits_D f(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x} = \int_a^b \int_{f_1(x_i)}^{g_1(x_i)} \cdots \int_{f_{i-1}(x_i)}^{g_{i-1}(x_i)} \int_{f_{i+1}(x_i)}^{g_{i+1}(x_1)} \cdots \int_{f_n(x_i)}^{g_n(x_i)} f(\boldsymbol{x})\mathrm{d}x_1 \ldots \mathrm{d}x_{i-1}\mathrm{d}x_{i+1} \ldots \mathrm{d}x_n\mathrm{d}x_i$$

Example 4.3.2. Evaluate $\iint\limits_D (x+2y)\mathrm{d}A$, where D is the region bounded by the parabolas $y = 2x^2$ and $y = 1 + x^2$.

Solution: This region has y-bounds that are paramaterized in terms of x. To find the x-bounds, we want to find the intersection of $y = 2x^2$ and $y = 1 + x^2$, which is at $x = \pm 1$. Then, our integral becomes

$$\begin{aligned}
\iint\limits_D (x+2y)\mathrm{d}A &= \int_{-1}^{1}\int_{2x^2}^{1+x^2} (x+2y)\mathrm{d}y\mathrm{d}x \\
&= \int_{-1}^{1} \left[xy + y^2 \Big|_{y=2x^2}^{y=1+x^2} \right] \mathrm{d}x \\
&= \int_{-1}^{1} \left[x + x^3 + x^4 + 2x^2 + 1 - 2x^3 - 4x^4 \right] \mathrm{d}x \\
&= \int_{-1}^{1} \left[-3x^4 - x^3 + 2x^2 + x + 1 \right] \mathrm{d}x \\
&= -\frac{3}{5}x^5 - \frac{1}{4}x^4 + \frac{2}{3}x^3 + \frac{1}{2}x^2 + x \Big|_{x=-1}^{x=1} = \frac{32}{15}
\end{aligned}$$

△

Example 4.3.3. Find the volume of the solid that lies under the paraboloid $z = x^2 + y^2$ and above the region D in the xy-plane bounded by the line $y = 2x$ and the parabola $y = x^2$.

Solution: Setting up our double integral, note that $y = 2x$ and $y = x^2$ intersect at $x = 0$ and $x = 2$, and that $2x > x^2$ on the interval $(0, 2)$. Therefore, the volume is

$$\begin{aligned}
\int_0^2 \int_{x^2}^{2x} \left[x^2 + y^2\right] \mathrm{d}y\mathrm{d}x &= \int_0^2 \left[x^2 y + \frac{1}{3}y^3 \Big|_{y=x^2}^{y=2x} \right] \mathrm{d}x \\
&= \int_0^2 \left[2x^3 + \frac{8}{3}x^3 - x^4 - \frac{1}{3}x^6 \right] \mathrm{d}x \\
&= \int_0^2 \left[-\frac{1}{3}x^6 - x^4 + \frac{14}{3}x^3 \right] \mathrm{d}x \\
&= -\frac{1}{27}x^7 - \frac{1}{5}x^5 + \frac{7}{6}x^4 \Big|_{x=0}^{x=2} = \frac{216}{35}
\end{aligned}$$

△

Example 4.3.4. Evaluate $\iint\limits_D xy\mathrm{d}A$, where D is the region bounded by the line $y = x - 1$ and the parabola $y^2 = 2x + 6$.

Solution: First, we can get our bounds in terms of y because getting them in terms of x would require casework (since the square root is not injective). So our bounds become $x = y + 1$ and $x = \frac{1}{2}y^2 - 3$. These intersect when $y + 1 = \frac{1}{2}y^2 - 3$.

$$\begin{aligned}
y + 1 &= \frac{1}{2}y^2 - 3 \\
-\frac{1}{2}y^2 + y + 4 &= 0
\end{aligned}$$

Then by quadratic formula,

$$\begin{aligned}
y &= \frac{-1 \pm \sqrt{1 - 4(-1/2)(4)}}{2(-1/2)} \\
&= -2, 4
\end{aligned}$$

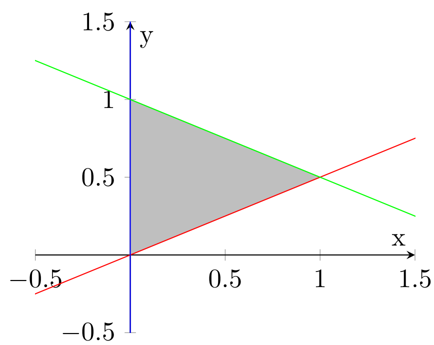
So our double integral is

$$\begin{aligned}
\iint_D xy\mathrm{d}A &= \int_{-8}^{10}\int_{\frac{1}{2}y^2-3}^{y+1} xy\mathrm{d}x\mathrm{d}y \\
&= \int_{-2}^{4} \frac{1}{2}y\left([y+1]^2 - \left[\frac{1}{2}y^2-3\right]^2\right)\mathrm{d}y \\
&= \int_{-2}^{4} \frac{1}{2}y\left[y^2+2y+1-\frac{1}{4}y^4+3y^2-9\right]\mathrm{d}y \\
&= \frac{1}{2}\int_{-2}^{4}\left[-\frac{1}{4}y^5+4y^3+2y^2-8y\right]\mathrm{d}y \\
&= \frac{1}{2}\left[-\frac{1}{24}y^6+y^4+\frac{2}{3}y^3-4y^2\right]_{-2}^{4} = 36
\end{aligned}$$

△

Example 4.3.5. Find the volume of the tetrahedron bounded by the planes $x+2y+z=2$, $x=2y$, $x=0$, and $z=0$.

Solution: First, it would do us some good to draw a picture of the region on the xy plane that we are integrating over. First, draw the lines $x=2y$ and $x=0$. Then, notice that the intersection of the planes $z=0$ and $x+2y+z=2$ is on the xy plane, and gives a curve described by $x+2y=2$.



The shaded region (we'll call it D) is what we want to integrate over. This region can be paramaterized:

$$D = \left\{(x,y) \quad | \quad 0\le x\le 1, \quad \frac{1}{2}x \le y \le 1-\frac{1}{2}x\right\}$$

Then, we can set up and compute our double integral,

$$\begin{aligned}
V &= \iint_D [2-x-2y]\mathrm{d}A = \int_0^1\int_{\frac{1}{2}x}^{1-\frac{1}{2}x}[2-x-2y]\mathrm{d}y\mathrm{d}x \\
&= \int_0^1 \left[2y-xy-y^2\right]_{y=x/2}^{y=1-x/2}\mathrm{d}x \\
&= \int_0^1\left[2\left(1-\frac{1}{2}x\right)-x\left(1-\frac{1}{2}x\right)-\left(1-\frac{1}{2}x\right)^2-2\left(\frac{x}{2}\right)+x\left(\frac{x}{2}\right)-\left(\frac{x}{2}\right)^2\right]\mathrm{d}x \\
&= \int_0^1[x^2-2x+1]\mathrm{d}x = \frac{x^3}{3}-x^2+x\Big|_0^1 = \boxed{\frac{1}{3}}
\end{aligned}$$

△

4.3.1 Properties of Double Integrals

First, let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ both be integrable on a region $D \subseteq \mathbb{R}^n$, and let $c \in \mathbb{R}$ be constant. Then,

$$\int_D [f + g]\mathrm{d}^n A = \int_D f\mathrm{d}^n A + \int_D g\mathrm{d}^n A \tag{4.3.1}$$

$$\int_D cf\mathrm{d}^n A = c\int_D f\mathrm{d}^n A \tag{4.3.2}$$

If $f \geq g$ on D, then

$$\int_D f\mathrm{d}^n A \geq \int_D g\mathrm{d}^n A \tag{4.3.3}$$

If D can be expressed as the union of several disjoint regions (except perhaps on their boundaries) $D_1, D_2, \dots, D_n$– that is, $D = \bigcup_{i=1}^n D_i$ where $\bigcap_{i=1}^n D_i \setminus \partial D_i = \varnothing$,

$$\int_D f\mathrm{d}^n A = \sum_{i=1}^{n} \int_{D_i} f\mathrm{d}^n A$$

The next property states that if we integrate the constant function $f = 1$, we get the area (or volume, or etc. in higher dimensions) of D.

$$\int_D \mathrm{d}^n A = A(D)$$

These properties can be combined to tell us,

Theorem 4.3.6. If $L, U \in \mathbb{R}$, and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is integrable on a region $D \subseteq \mathbb{R}^n$, with $L \leq f(\boldsymbol{x}) \leq U$ for all $x \in D$,

$$LA(D) \leq \int_D f\mathrm{d}^n A \leq UA(D)$$

Where $A(D)$ is the area (or n-dimensional equivalent) of D.

Example 4.3.7. Estimate the integral $\iint\limits_D e^{\sin x \cos y}\mathrm{d}A$ where D is the disk centered on the origin with radius 2.

Solution: First, recognize that the area of D is given by $\pi r^2 = 4\pi$, and that $-1 \leq \sin x \cos y \leq 1$, so $\frac{1}{e} \leq e^{\sin x \cos y} \leq e$ on D. Therefore,

$$4\pi e^{-1} \leq \iint\limits_D e^{\sin x \cos y}\mathrm{d}A \leq 4\pi e$$

Or,

$$4.623 \leq \iint\limits_D e^{\sin x \cos y}\mathrm{d}A \leq 34.159$$

△

4.4 Double Integrals in Polar Coordinates

Some regions are much easier to describe and integrate over with polar coordinates.

Consider a region $R = \{(r, \theta), a \leq r \leq b, \alpha \leq \theta \leq \beta\}$. This is called a *polar rectangle*. To define integration over this region, divide the interval $[a, b]$ into m equal sub-intervals where $\Delta r = (b - a)/m$. Similarly, divide the interval $[\alpha, \beta]$ into n equal sub-intervals where $\Delta\theta = (\beta - \alpha)/n$. Then, each infinitesimal segment of area A_{ij} is defined by the rays $r = r_i$ and $r = r_{i+1} = r_i + \Delta r$, along with $\theta = \theta_j$ and $\theta = \theta_{j+1} = \theta_j + \Delta\theta$.

Recall that the area of a sector of a circle with radius ℓ and central angle ϕ is $\frac{1}{2}\ell^2\phi$. Applying this to our scenario, each infinitesimal section of area is equivalent to the area of a sector formed by $\theta \in [\theta_j, \theta_{j+1}]$ and $r = r_{i+1}$ *minus* the area of the sector formed by $\theta \in [\theta_j, \theta_{j+1}]$ and $r = r_i$. That is,

$$\begin{aligned} A_{ij} &= \frac{1}{2}r_{i+1}^2(\theta_{j+1} - \theta_j) - \frac{1}{2}r_i^2(\theta_{j+1} - \theta_j) \\ &= \frac{1}{2}(\theta_{j+1} - \theta_j)(r_{i+1}^2 - r_i^2) \\ &= \frac{1}{2}\Delta\theta(r_{i+1} - r_i)(r_{i+1} + r_i) \end{aligned}$$

Noticing that the radius of the center of the sector is $r_i^* = \frac{1}{2}(r_{i+1} + r_i)$, we arrive at

$$A_{ij} = r_i^* \Delta r \Delta\theta$$

So the Riemann sum describing the area can be expressed as

$$\sum_{i=1}^{m}\sum_{j=1}^{n} r_i^* \Delta r \Delta\theta$$

Or, if we have a function $f(x, y)$, the double integral $\iint\limits_R f \mathrm{d}A$ is,

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f(r_i^* \cos\theta_j^*, r_i^* \sin\theta_j^*) r_i^* \Delta r \Delta\theta$$

Therefore, we have

$$\iint\limits_R f(x, y)\mathrm{d}A = \int_\alpha^\beta \int_a^b f(r\cos\theta, r\sin\theta) r \mathrm{d}r\mathrm{d}\theta$$

This can be thought of as integrating f over infinitely many polar rectangles, each one having area $\mathrm{d}A = r\mathrm{d}r\mathrm{d}\theta$ (because $r\mathrm{d}\theta$ is the arc length of the polar regions, which becomes closer to being straight and forming a rectangle with the side $\mathrm{d}r$ as $\mathrm{d}\theta \to 0$).

Example 4.4.1. Evaluate $\iint\limits_R (3x + 4y^2)\mathrm{d}A$, where R is the region bounded by $x^2 + y^2 = 1$, $x^2 + y^2 = 4$, and the line $y = 0$.

Solution: R can be expressed as $R = \{(r, \theta) | 1 \le r \le 2, 0 \le \theta \le \pi\}$. Therefore,

$$\begin{aligned} \iint\limits_R (3x + 4y^2)\mathrm{d}A &= \int_0^\pi \int_1^2 (3r\cos\theta + 4r^2\sin^2\theta) r \mathrm{d}r\mathrm{d}\theta \\ &= \int_0^\pi \int_1^2 (3r^2\cos\theta + 4r^3\sin^2\theta)\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^\pi \left(r^3\cos\theta + r^4\sin^2\theta \Big|_{r=1}^{r=2} \right)\mathrm{d}\theta \\ &= \int_0^\pi (8\cos\theta + 16\sin^2\theta - \cos\theta - \sin^2\theta)\mathrm{d}\theta \\ &= \int_0^\pi \left(7\cos\theta + \frac{15}{2}(1 - \cos(2\theta) \right)\mathrm{d}\theta \\ &= 7\sin\theta + \frac{15}{2}\theta - \frac{15}{4}\sin(2\theta)\Big|_{\theta=0}^{\theta=\pi} = \boxed{\frac{15\pi}{4}} \end{aligned}$$

△

The power reduction trig identities will be very helpful for these types of integrals, and are worth remembering.

$$\sin^2 x = \frac{1}{2}(1-\cos 2x), \quad \cos^2 x = \frac{1}{2}(1+\cos 2x)$$

Example 4.4.2. Find the volume of the solid bounded by the plane $z=0$ and the paraboloid $z = 1-x^2-y^2$.

Solution: First, we can find the intersection of the paraboloid with the plane $z=0$:

$$0 = 1-x^2-y^2 \quad \Longrightarrow \quad x^2+y^2=1$$

So the region is the unit circle on the xy-plane, or $R = \{(r,\theta)|0 \leq r \leq 1, 0 \leq \theta \leq 2\pi\}$. So the volume is given by

$$\begin{aligned} V &= \int_0^{2\pi}\int_0^1 (1-x^2-y^2)r\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^1 (1-r^2)r\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^1 r - r^3\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi}\left[\frac{1}{2}r^2 - \frac{1}{4}r^4\right]_{r=0}^{r=1}\mathrm{d}\theta \\ &= \int_0^{2\pi}\frac{1}{4}\mathrm{d}\theta = \boxed{\frac{\pi}{2}} \end{aligned}$$

This is a good example of a case where polar is a much easier method of integration. If we were to integrate in rectangular, we would have

$$V = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}[1-x^2-y^2]\mathrm{d}y\mathrm{d}x$$

Which would require some very advanced integration techniques, since you would have to find

$$\int\sqrt{1-x^2}\mathrm{d}x, \quad \int x^2\sqrt{1-x^2}\mathrm{d}x, \quad \text{and} \quad \int(1-x^2)^{3/2}\mathrm{d}x$$

△

Similar to double integrals over general rectangular regions, if we have a polar region that can be expressed as

$$R = \{(r,\theta)|\alpha \leq \theta \leq \beta, h_1(\theta) \leq r \leq h_2(\theta)\}$$

then $\iint\limits_R f(x,y)\mathrm{d}A$ can be written as

$$\iint\limits_R f(x,y)\mathrm{d}A = \int_\alpha^\beta\int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta)r\mathrm{d}r\mathrm{d}\theta$$

Choosing $f(x,y)=1$, we can get the area of a polar region:

$$A(R) = \iint\limits_R \mathrm{d}A = \int_\alpha^\beta\int_{h_1(\theta)}^{h_2(\theta)} r\mathrm{d}r\mathrm{d}\theta = \int_\alpha^\beta \frac{1}{2}(h_2(\theta)^2 - h_1(\theta)^2)\mathrm{d}\theta$$

Which you may recognize as the formula for the area between two polar graphs from Calculus 1.

Example 4.4.3. Use a double integral to find the area enclosed by one loop of the four-leaved rose $r = \cos(2\theta)$.

Solution: First, we want to find the bounds, which will occur at the θ values where $\cos(2\theta) = 0$. The two that we are concerned with are $\theta = -\pi/4$ and $\theta = \pi/4$. Therefore, our double integral is

$$\begin{aligned} A &= \int_{-\pi/4}^{\pi/4} \int_0^{\cos(2\theta)} r \mathrm{d}r \mathrm{d}\theta \\ &= \int_{-\pi/4}^{\pi/4} \frac{1}{2} \cos^2(2\theta) \mathrm{d}\theta \\ &= \int_{-\pi/4}^{\pi/4} \frac{1}{4}(1 + \cos(4\theta)) \mathrm{d}\theta \\ &= \frac{\theta}{4} + \frac{1}{16}\sin(4\theta)\Big|_{-\pi/4}^{\pi/4} = \boxed{\frac{\pi}{8}} \end{aligned}$$

△

Example 4.4.4. Find the volume of the solid that lies under the paraboloid $z = x^2 + y^2$ above the xy-plane, and inside the cylinder $x^2 + y^2 = 2x$.

Solution: First, we want to rewrite our bounds in a way that can be expressed in polar.

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

Or $(x-1)^2 + y^2 = 1$. This is a circle of radius 1 centered at the point $(1, 0)$. Now, to express this in polar, we will want to perform a coordinate shift. Note that if we let $u = x - 1$, then our circle equation becomes $u^2 + y^2 = 1$, which is centered at our "new origin." Our paraboloid equation then becomes $z = (u+1)^2 + y^2$, which we can now integrate.

$$\begin{aligned} \iint_R (x^2 + y^2) r \mathrm{d}r \mathrm{d}\theta &= \int_0^{2\pi} \int_0^1 [(u+1)^2 + y^2] r \mathrm{d}r \mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 r(r\cos\theta + 1)^2 + r^3 \sin^2\theta \mathrm{d}r \mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 r^3\cos^2\theta + 2r^2\cos\theta + r + r^3\sin^2\theta \mathrm{d}r \mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 r^3(\cos^2\theta + \sin^2\theta) + 2r^2\cos\theta + r \mathrm{d}r \mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 r^3 + 2r^2\cos\theta + r \mathrm{d}r \mathrm{d}\theta \\ &= \int_0^{2\pi} \left[\frac{1}{4}r^4 + \frac{2}{3}r^3\cos\theta + \frac{1}{2}r^2 \Big|_{r=0}^{r=1} \right] \\ &= \int_0^{2\pi} \frac{3}{4} + \frac{2}{3}\cos\theta = \frac{3}{2}\theta + \frac{2}{3}\sin\theta \Big|_0^{2\pi} = \boxed{\frac{3\pi}{2}} \end{aligned}$$

△

4.5 Applications of Double Integrals

We've already explored one application of double integrals in computing volume. Now, we will look at some other ones such as computing mass, electric charge, center of mass, moments, and probabilities.

4.5.1 Density and Mass

If we have an object that occupies a region D in the xy-plane, and has a variable area density function $\rho_A(x,y)$ that is continuous on D, then we can write the mass of the object as

$$m = \lim_{k,l\to\infty} \sum_{i=1}^{k}\sum_{j=1}^{l} \rho_A(x_i^*, y_j^*)\Delta A = \iint\limits_D \rho_A(x,y)\mathrm{d}A$$

Another similar application is using charge densities to find the total charge in a charge distribution. If the charge distribution is given by an integrable function $\sigma(x,y)$, then the total charge is given by

$$Q = \iint\limits_D \sigma(x,y)\mathrm{d}A$$

Example 4.5.1. Charge is distributed over a triangular region D enclosed by $y=1$, $x=1$, and $y=1-x$, where x and y are measured in meters. The charge density at a point (x,y) on D is given by $\sigma(x,y)=xy$, where σ is measured in coulombs per square meter. Find the total charge on D.

Solution: Our double integral can be written as

$$\begin{aligned}
\iint\limits_D \sigma(x,y)\mathrm{d}A &= \int_0^1 \int_{1-x}^1 xy\mathrm{d}y\mathrm{d}x \\
&= \int_0^1 \frac{1}{2}xy^2\Big|_{y=1-x}^{y=1} \mathrm{d}x \\
&= \frac{1}{2}\int_0^1 x - x(1-x)^2\mathrm{d}x \\
&= \frac{1}{2}\int_0^1 x - x^3 + 2x^2 - x\mathrm{d}x \\
&= \frac{1}{2}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\Big|_0^1\right] = \boxed{\frac{5}{24}}
\end{aligned}$$

△

4.5.2 Moments and Center of Mass

Suppose we have a lamina occupying a region D with a density function $\rho_A(x,y)$. Then, the **moment** of the lamina about the x-axis is

$$M_x = \iint\limits_D y\rho(x,y)\mathrm{d}A$$

and the moment of the lamina about the y-axis is

$$M_y = \iint\limits_D x\rho(x,y)\mathrm{d}A$$

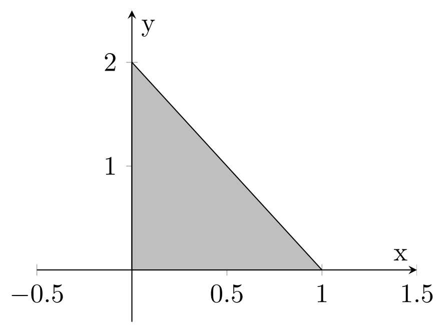
We define the center of mass $\langle \overline{x}, \overline{y}\rangle$ of the lamina so that if the total mass is m, then $m\overline{x} = M_x$ and $m\overline{y} = M_y$. That is,

$$\overline{x} = \frac{M_x}{m} = \frac{\iint\limits_D y\rho(x,y)\mathrm{d}A}{\iint\limits_D \rho(x,y)\mathrm{d}A}$$

$$\overline{y} = \frac{M_y}{m} = \frac{\iint\limits_D x\rho(x,y)\mathrm{d}A}{\iint\limits_D \rho(x,y)\mathrm{d}A}$$

Example 4.5.2. Find the mass and center of mass of a triangular lamina with vertices $(0,0)$, $(1,0)$, and $(0,2)$ if the density function is $\rho(x,y) = 1 + 3x + y$.

Solution: First, let's draw a picture.



Let's call the shaded region D. We can write D as

$$D = \{(x,y) | 0 \leq x \leq 1, 0 \leq y \leq 2 - 2x\}$$

So the mass of the lamina is

$$\begin{aligned} m = \iint_D \rho(x,y)\mathrm{d}A &= \int_0^1 \int_0^{2-2x} [1 + 3x + y]\mathrm{d}y\mathrm{d}x \\ &= \int_0^1 \left[y + 3xy + \frac{1}{2}y^2 \right]_{y=0}^{y=2-2x} \mathrm{d}x \\ &= \int_0^1 (2-2x) + 3x(2-2x) + \frac{1}{2}(2-2x)^2 \mathrm{d}x \\ &= \int_0^1 4 - 4x^2 \mathrm{d}x = 4\left[x - \frac{1}{3}x^3 \right]_0^1 = \boxed{\frac{8}{3}} \end{aligned}$$

△

Then, center of mass in the x direction:

$$\begin{aligned} \overline{x} = \frac{1}{m} \iint_D x\rho(x,y)\mathrm{d}A &= \frac{3}{8}\int_0^1 \int_0^{2-2x} [x + 3x^2 + xy]\mathrm{d}y\mathrm{d}x \\ &= \frac{3}{8}\int_0^1 \left[xy + 3x^2y + \frac{1}{2}xy^2 \right]_{y=0}^{y=2-2x} \mathrm{d}x \\ &= \frac{3}{8}\int_0^1 \left[2x - 2x^2 + 6x^2 - 6x^3 + 2x - 4x^2 + 2x^3 \right] \mathrm{d}x \\ &= \frac{3}{8}\int_0^1 \left[-4x^3 + 4x \right] \mathrm{d}x = \frac{3}{2}\int_0^1 (x - x^3)\mathrm{d}x \\ &= \frac{3}{2}\left[\frac{1}{2}x^2 - \frac{1}{4}x^3 \right]_0^1 = \boxed{\frac{3}{8}} \end{aligned}$$

And in the y direction:

$$\begin{aligned}
\overline{y} = \frac{1}{m}\iint\limits_D y\rho(x,y)\mathrm{d}A &= \frac{3}{8}\int_0^1\int_0^{2-2x}[y+3xy+y^2]\mathrm{d}y\mathrm{d}x \\
&= \frac{3}{8}\int_0^1\left[\frac{1}{2}y^2+\frac{3}{2}xy^2+\frac{1}{3}y^3\right]_0^{2-2x}\mathrm{d}x \\
&= \frac{3}{8}\int_0^1\left[\frac{1}{2}(2-2x)^2+\frac{3}{2}x(2-2x)^2+\frac{1}{3}(2-2x)^3\right]\mathrm{d}x \\
&= \frac{3}{8}\int_0^1\left[2x^2-4x+2+6x^3-12x^2+6x-\frac{8}{3}x^3+8x^2-8x+\frac{8}{3}\right]\mathrm{d}x \\
&= \frac{1}{4}\int_0^1\left[5x^3-3x^2-9x+7\right]\mathrm{d}x = \frac{1}{4}\left[\frac{5}{4}x^4-x^3-\frac{9}{2}x^2+7x\right]_0^1 \\
&= \frac{1}{4}\left[\frac{5}{4}-\frac{4}{4}-\frac{18}{4}+\frac{28}{4}\right] = \boxed{\frac{11}{16}}
\end{aligned}$$

Example 4.5.3. The density at a point on a semicircular lamina of radius a is proportional to the distance from the center of the circle. Find the center of mass of the lamina.

Solution: First, write out the density function:

$$\rho = k\sqrt{x^2+y^2} = kr$$

Then, integrate over the circle in polar coordinates:

$$\begin{aligned}
m = \iint\limits_D \rho\mathrm{d}A &= \int_0^\pi\int_0^a r\rho\mathrm{d}r\mathrm{d}\theta \\
&= \int_0^\pi\int_0^a kr^2\mathrm{d}r\mathrm{d}\theta \\
&= \int_0^\pi \frac{1}{3}ka^3\mathrm{d}\theta = \frac{ka^3\pi}{3}
\end{aligned}$$

Because of the symmetry of the lamina, we can reason that the center of mass in the x direction will be on the y axis (but we'll integrate it anyway just to verify)

$$\begin{aligned}
\overline{x} = \frac{1}{m}\iint\limits_D x\rho\mathrm{d}A &= \frac{3}{ka^3\pi}\int_0^\pi\int_0^a kr^3\cos\theta\mathrm{d}r\mathrm{d}\theta \\
&= \frac{3}{ka^3\pi}\int_0^\pi\frac{1}{3}kr^3\cos\theta\mathrm{d}\theta = \frac{3}{ka^3\pi}\left[\frac{1}{4}ka^4\sin\theta\right]_0^\pi = 0
\end{aligned}$$

and

$$\begin{aligned}
\overline{y} = \frac{1}{m}\iint\limits_D y\rho\mathrm{d}A &= \frac{3}{ka^3\pi}\int_0^\pi\int_0^a kr^3\sin\theta\mathrm{d}r\mathrm{d}\theta \\
&= \frac{3}{ka^3\pi}\int_0^\pi\frac{1}{4}ka^3\sin\theta\mathrm{d}\theta \\
&= -\frac{3}{ka^3\pi}\left[\frac{1}{4}ka^4\cos\theta\right]_0^\pi = \frac{3a}{2\pi}
\end{aligned}$$

△

4.5.3 Moments of Inertia

The moment of inertia (also called the second moment) of a particle with mass m about an axis is defined to be mr^2, where r is the distance between the particle and the axis. We can extend this definition to two-dimensional mass distributions.

Consider a lamina with density function $\rho(x,y)$ that occupies a region D. We can divide D into m sub-rectangles in the x direction and n sub-rectangles in the y direction, and let m and n tend to infinity. Then, we can choose two sample points $x_i^* \in [x_i, x_{i+1}]$ and $y_j^* \in [y_j, y_{j+1}]$ for each sub-rectangle (i,j). The distance between the x axis and the sample point is then y_j^*. Then, we sum up the contributions to the overall inertia from each sub-rectangle.

$$I_x = \lim_{m,n\to\infty} \sum_{i=1}^{m}\sum_{j=1}^{n} (y_j^*)^2 \rho(x_i^*, y_j^*)\Delta A = \iint_D y^2 \rho(x,y)\mathrm{d}A$$

The moment of inertia about the y axis is found similarly,

$$I_x = \lim_{m,n\to\infty} \sum_{i=1}^{m}\sum_{j=1}^{n} (x_i^*)^2 \rho(x_i^*, y_j^*)\Delta A = \iint_D x^2 \rho(x,y)\mathrm{d}A$$

We can also consider a moment of inertia about the origin (this can be thought of as an object rotating about what would be the z axis in three-dimensional space). We again split D into sub-rectangles, but now the distance to the sample point is given by $\sqrt{(x_i^*)^2 + (y_j^*)^2}$, so we find

$$I_0 = \lim_{m,n\to\infty} \sum_{i=1}^{m}\sum_{j=1}^{n} ((x_i^*)^2 + (y_j^*))\rho(x_i^*, y_j^*) = \iint_D (x^2+y^2)\rho(x,y)\mathrm{d}A$$

Note that the equivalency $I_0 = I_x + I_y$ holds because we can split up the integral as such:

$$\iint_D (x^2+y^2)\rho(x,y)\mathrm{d}A = \iint_D y^2\rho(x,y)\mathrm{d}A + \iint_D x^2\rho(x,y)\mathrm{d}A = I_x + I_y$$

Example 4.5.4. Find the moments of inertia I_x, I_y, and I_0 of a homogeneous disk D with density $\rho(x,y) = \rho$, center $(0,0)$ and radius a.

Solution: Let's first find I_0.

$$I_0 = \iint_D (x^2+y^2)\rho(x,y)\mathrm{d}A = \iint_D (x^2+y^2)\rho\mathrm{d}A$$

Note that D can be written as $\{(r,\theta) | 0 \le \theta \le 2\pi, 0 \le r \le a\}$, so we can rewrite

$$\begin{aligned} I_0 = \iint_D (x^2+y^2)\rho\mathrm{d}A &= \int_0^{2\pi}\int_0^a r^3\rho\mathrm{d}r\mathrm{d}\theta \\ &= \frac{\rho a^4}{4}\int_0^{2\pi}\mathrm{d}\theta = \frac{\rho\pi a^4}{2} \end{aligned}$$

Because of the symmetry of the problem, we can see that $I_x = I_y = \frac{1}{2}I_0$. Therefore,

$$\boxed{I_0 = \frac{\rho\pi a^4}{2}, \quad I_x = I_y = \frac{\rho\pi a^4}{4}}$$

If we notice that the mass of the disk will be $\rho\pi a^2$ (density $\times$ area), we can rewrite I_0 as $I_0 = \frac{1}{2}ma^2$, which you may recognize as the moment of inertia of a disk about its axle from physics. This is an alternate derivation for it. $\triangle$

We can also define a quantity called the **radius of gyration** about an axis. We define it as the number R such that

$$mR^2 = I$$

where m is the mass of the lamina and I is the moment of inertia about the given axis.

The radius of gyration can be thought of as the point where, if we were to concentrate the entire mass of the lamina there, would not affect the moment of inertia.

We can define the radius of gyration $\overline{\overline{y}}$ with respect to the x-axis and the radius of gyration $\overline{\overline{x}}$ with respect to the y-axis as

$$m\overline{\overline{y}}^2 = I_x, \quad m\overline{\overline{x}}^2 = I_y$$

We can also define the radius of gyration $\overline{\overline{r}}$ with respect to the origin as the number $\overline{\overline{r}}$ such that

$$m\overline{\overline{r}}^2 = I_0$$

Example 4.5.5. Find the radius of gyration about the x-axis of the disk in the previous example.

Solution: Recall that the moment of inertia about the x axis is $I_x = (1/4)\rho\pi a^4$ and the mass of the disk is $\rho\pi a^2$. So we can find $\overline{\overline{y}}$ with

$$\overline{\overline{y}} = \sqrt{\frac{I_x}{m}} = \sqrt{\frac{\rho\pi a^4}{4\rho\pi a^2}} = \frac{a}{2}$$

△

4.5.4 Probability

Let $f : \mathbb{R} \to \mathbb{R}$ be a probability density function of a continuous random variable X. This means that f must satisfy two criteria:

1. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.
2. the probability that $X \in [a, b]$ is found with $P(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x$.

One common probability density function is the normal distribution, defined with

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where σ is the mean of the dataset and μ is the standard deviation. It is left as an exercise to reader to verify that $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ (this will actually require some methods of multivariable calculus, despite it being a single integral).

Now, let $f : \mathbb{R}^2 \mapsto \mathbb{R}$ be a probability density function of a pair of continuous random variables (X, Y). This is known as a **joint density function**, and satisfies similar criteria to single-variable probability density functions.

1. $\iint\limits_{\mathbb{R}^2} f(x, y)\mathrm{d}A = 1$.
2. The probability that $(X, Y) \in D$ for some $D \subset \mathbb{R}^2$ is found with $P((X, Y) \in D) = \iint\limits_{D} f(x, y)\mathrm{d}A$.

Similarly, we can define a joint density function for a set of n continuous random variables $X_1, X_2, \ldots, X_n$ as a function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies

1. $\int\limits_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x} = 1$.
2. The probability that $(X_1, X_2, \ldots, X_n) \in D$ for some $D \subset \mathbb{R}^n$ is found with $P((X_1, X_2, \ldots, X_n) \in D) = \int\limits_{D} f(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x}$.

Example 4.5.6. If the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} C(x+2y) & x \in [0,10], y \in [0,10] \\ 0 & \text{otherwise} \end{cases}$$

find the value of the constant C. Then, find $P(X \leq 7, Y \geq 2)$.

Solution: First, let's find C. Remember that f must satisfy $\iint\limits_{\mathbb{R}^2} f(x,y)\mathrm{d}A = 1$. So,

$$\iint\limits_{\mathbb{R}^2} f(x,y)\mathrm{d}A = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x$$

But because $f(x,y) = 0$ for all (x,y) not in the square of sidelength 10 centered at the origin, we can rewrite

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x &= \int_0^{10}\int_0^{10} C(x+2y)\mathrm{d}y\mathrm{d}x \\
&= C\int_0^{10}\int_0^{10} (x+2y)\mathrm{d}y\mathrm{d}x \\
&= C\int_0^{10} \left[xy + y^2\right]_{y=0}^{y=10} \mathrm{d}x \\
&= C\int_0^{10} [10x + 100]dx \\
&= C\left[5x^2 + 100x\right]_0^{10} = 1500C
\end{aligned}$$

So because $1500C = 1$, $C = \frac{1}{1500}$.

Now, finding $P(X \leq 7, Y \geq 2)$, we can write it as an iterated integral

$$P(X \leq 7, Y \geq 2) = \int_{-\infty}^{7}\int_{2}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x$$

Again, since $f(x,y) = 0$ for all (x,y) not in the square of sidelength 10 centered at the origin, we can rewrite to

$$\begin{aligned}
P(X \leq 7, Y \geq 2) &= \int_0^{7}\int_2^{10} C(x+2y)\mathrm{d}y\mathrm{d}x \\
&= C\int_0^7 \left[xy + y^2\right]_{y=2}^{y=10} \mathrm{d}x \\
&= C\int_0^7 8x + 96\mathrm{d}x = C\left[4x^2 + 96x\right]_0^7 \\
&= 868C = \frac{868}{1500} \approx 0.579
\end{aligned}$$

△

In the above examples, the probaility for a given value of Y depended on the value of X. That is not always the case. When it isn't, we call X and Y **independent random variables**. Then, their joint density function is simply the product of their individual density functions.

$$f(x,y) = f_1(x)f_2(y)$$

It can be shown that if f_1 and f_2 satisfy the criteria for a probability density function, then f will satisfy the criteria for a joint probability density.

$$
\begin{aligned}
\iint\limits_{\mathbb{R}^2} f(x,y)\mathrm{d}A &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x)f_2(y)\mathrm{d}y\mathrm{d}x \\
&= \left[\int_{\infty}^{\infty} f_1(x)\mathrm{d}x\right]\left[\int_{\infty}^{\infty} f_2(y)\mathrm{d}y\right] \\
&= 1\cdot 1 = 1
\end{aligned}
$$

Consider the exponential density function

$$
f(t) = \begin{cases} 0 & t < 0 \\ \mu^{-1}e^{-t/\mu} & t \geq 0 \end{cases}.
$$

This can be used to model the probabilities of waiting times, where μ is the mean waiting time.

Example 4.5.7. The manager of a movie theater determines that the average time moviegoers wait in line to buy a ticket for this week's film is 10 minutes and the average time they wait to buy popcorn is 5 minutes. Assuming that the waiting times are independent, find the probability that a moviegoer waits a total of less than 20 minutes before taking their seat using the exponential density model.

Solution: The density functions for the movie ticket waiting time (x) and popcorn waiting time (y) respectively are

$$
f_1(x) = \begin{cases} 0 & x < 0 \\ e^{-x/10}/10 & x > 0 \end{cases} \quad \text{and} \quad f_2(y) = \begin{cases} 0 & y < 0 \\ e^{-y/5}/5 & y > 0 \end{cases}
$$

So the joint density function is

$$
f(x,y) = \begin{cases} \left[\frac{e^{-x/10}}{10}\right]\left[\frac{e^{-y/5}}{5}\right] & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}
$$

This simplifies to

$$
f(x,y) = \begin{cases} \frac{1}{50}e^{-x/10}e^{-y/5} & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}
$$

If the total time the moviegoer waits is less than 20 minutes, then we must have $x + y < 20$. Clearly, x and y must be positive because it is nonsensical to have negative waiting times. Therefore, the valid pairs (x, y) are described by the region

$$
R = \{(x,y) | 0 \leq x < 20, 0 \leq y < 20 - x\}
$$

So the probability that a moviegoer waits less than 20 minutes is given by

$$
P((x,y) \in R) = \iint\limits_{R} f(x,y)\mathrm{d}A = \int_0^{20}\int_0^{20-x} f(x,y)\mathrm{d}y\mathrm{d}x
$$

Because we never have x or y negative, we can rewrite as

$$\begin{aligned}
P((x,y)\in R) &= \int_0^{20}\int_0^{20-x}\frac{1}{50}e^{-x/10}e^{-y/5}\mathrm{d}y\mathrm{d}x \\
&= -\frac{1}{10}\int_0^{20}\left[e^{-x/10}e^{-y/5}\right]_{y=0}^{y=20-x}\mathrm{d}x \\
&= -\frac{1}{10}\int_0^{20}e^{-x/10}\left(e^{-(20-x)/5}-1\right)\mathrm{d}x \\
&= -\frac{1}{10}\int_0^{20}e^{-4+x/5-x/10}-e^{-x/10}\mathrm{d}x \\
&= -\frac{1}{10}\int_0^{20}e^{-4}e^{x/10}-e^{-x/10}\mathrm{d}x \\
&= -\left(e^{-4}e^{x/10}+e^{-x/10}\right)\Big|_0^{20} = 1+e^{-4}-2e^{-2}\approx 0.748
\end{aligned}$$

$\triangle$

4.5.5 Expected Value

The expected values (μ_1,μ_2) of a probability function $f(x,y)$ are the first moments of f with respect to x and y respectively. That is,

$$\mu_1 = \iint\limits_{\mathbb{R}^2} xf(x,y)\mathrm{d}A, \qquad \mu_2 = \iint\limits_{\mathbb{R}^2} yf(x,y)\mathrm{d}A$$

We can justify this by imagining the probability density function as describing a sort of "probability mass." The total mass is given by $\iint_{\mathbb{R}^2} f(x,y)\mathrm{d}A$, which we know to be 1 from the definition of a probability density function. Then, the expected value is analogous to the center of mass, which is qual to the moment divided by the mass. However, since the mass is 1, it is just equal to the moment.

4.6 Surface Area

Another application of double integrals is computing the area of a surface.

Consider a graph $z=f(x,y)$ of two variables that is differentiable along a region $D\subset\mathbb{R}^2$. Let U be the surface represented by $z=f(x,y)$ where $(x,y)\in D$. If we split D into m sub-rectangles in the x direction and n sub-rectangles in the y direction and choose a sample point $x_i^*\in[x_i,x_{i+1}]$, $y_j^*\in[y_j,y_{j+1}]$, we can use the area of the tangent plane of f at (x_i^*,y_j^*) along the sub-rectangle to approximate the area of the function.

The tangent vectors to f in the x and y direction are given by

$$\boldsymbol{t}_x = \left\langle \Delta x, 0, f_x(x_i^*,y_j^*)\right\rangle \quad \text{and} \quad \boldsymbol{t}_y = \left\langle 0, \Delta y, f_y(x_i^*,y_j^*)\right\rangle$$

and the area of the parallellogram formed by $\boldsymbol{t}_x$ and $\boldsymbol{t}_y$ is given by

$$\begin{aligned}
\mathrm{d}S = \|\boldsymbol{t}_x\times\boldsymbol{t}_y\| &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \Delta x & 0 & \Delta x f_x(x_i^*,y_j^*) \\ 0 & \Delta y & \Delta y f_y(x_i^*,y_j^*) \end{vmatrix} \\
&= \left\|\left\langle -\Delta y\Delta x f_x(x_i^*,y_j^*), -\Delta y\Delta x f_y(x_i^*,y_j^*), \Delta x\Delta y\right\rangle\right\| \\
&= \sqrt{\left[-\Delta y\Delta x f_x(x_i^*,y_j^*)\right]^2+\left[-\Delta y\Delta x f_y(x_i^*,y_j^*)\right]^2+\left[\Delta x+\Delta y\right]^2} \\
&= \sqrt{\left[f_x(x_i^*,y_j^*)\right]^2+\left[f_y(x_i^*,y_j^*)\right]^2+1}\,\Delta y\Delta x \\
&= \sqrt{\left[f_x(x_i^*,y_j^*)\right]^2+\left[f_y(x_i^*,y_j^*)\right]^2+1}\,\Delta A
\end{aligned}$$

So the total surface area is the sum over all of these subrectangles, or

$$S = \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{\left[f_x(x_i^*, y_j^*)\right]^2 + \left[f_y(x_i^*, y_j^*)\right]^2 + 1}\,\Delta A = \iint_D \sqrt{[f_x(x,y)]^2 + [f_y(x,y)]^2 + 1}\,\mathrm{d}A$$

Note the similarity between this formula and the arc-length formula from single-variable calculus:

$$S = \int_a^b \sqrt{1 + (f'(x))^2}\mathrm{d}x$$

Example 4.6.1. Find the surface area of the part of the surface $z = x^2 + 2y$ that lies above the triangular region T in the xy-plane with vertices $(0,0)$, $(1,0)$, and $(1,1)$.

Solution: We can describe the region we're integrating over as $T = \{(x,y)|0 \leq x \leq 1, 0 \leq y \leq x\}$. Then,

$$\begin{aligned} A &= \iint_T \sqrt{\left[\frac{\partial z}{\partial x}\right]^2 + \left[\frac{\partial z}{\partial y}\right]^2 + 1}\,\mathrm{d}A \\ &= \int_0^1 \int_0^x \sqrt{4x^2 + 5}\,\mathrm{d}y\mathrm{d}x \\ &= \int_0^1 x\sqrt{4x^2 + 5}\,\mathrm{d}x \\ &= \frac{1}{12}(4x^2 + 5)^{3/2}\Big|_0^1 = \boxed{\frac{1}{12}\left(27 - 5\sqrt{5}\right)} \end{aligned}$$

△

Example 4.6.2. Find the surface area of the part of the paraboloid $z = x^2 + y^2$ that lies under the plane $z = 9$.

Solution: The boundary of the region occurs when $x^2 + y^2 = 9$, so we can describe it as $R = \{(x,y)| -3 \leq x \leq 3, -\sqrt{9 - x^2} \leq y \leq \sqrt{9 - x^2}\}$. In polar coordinates, this becomes $R = \{(r,\theta)|0 \leq \theta \leq 2\pi, 0 \leq r \leq 3\}$. So our surface area integral is

$$\begin{aligned} A &= \iint_R \sqrt{\left[\frac{\partial z}{\partial x}\right]^2 + \left[\frac{\partial z}{\partial y}\right]^2 + 1}\,\mathrm{d}A \\ &= \iint_R \sqrt{4x^2 + 4y^2 + 1}\,\mathrm{d}A \\ &= \int_0^{2\pi} \int_0^3 r\sqrt{4r^2 + 1}\mathrm{d}r\mathrm{d}\theta \\ &= \frac{1}{12}\int_0^{2\pi} (4r^2 + 1)^{3/2}\Big|_{r=0}^{r=3} \mathrm{d}\theta \\ &= \frac{1}{12}\int_0^{2\pi} (37^{3/2} - 1)\mathrm{d}\theta = \boxed{\frac{\pi}{6}\left(37^{3/2} - 1\right)} \end{aligned}$$

△

4.7 Triple Integrals

We can define triple integrals of functions $f : \mathbb{R}^3 \to \mathbb{R}$ similarly to how we defined double integrals. We'll skip the derivation as it is nearly identical to the double integral derivation.

The **triple integral** of a function f over a box $R \subset \mathbb{R}^3$ is defined as

$$\iiint\limits_R f(x,y,z)\mathrm{d}V = \lim_{l,m,n\to\infty} \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} f(x_i^*, y_j^*, z_k^*)\Delta x\Delta y\Delta z$$

Where each (x_i^*, y_j^*, z_k^*) is a sample point in its respective sub-box B_{ijk}, assuming the limit exists.

Theorem 4.7.1 (Fubini's Theorem for Triple Integrals)**.** If $f : \mathbb{R}^3 \to \mathbb{R}$ is integrable on a box $B \subset \mathbb{R}^3$, where $B = \{(x,y,z)|x \in [a,b], y \in [c,d], z \in [e,g]\}$ then the triple integral of f can be written as an interated integral.

$$\iint\limits_B f(x,y,z)\mathrm{d}V = \int_a^b \int_c^d \int_e^g f(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x$$

The order of these integrals is interchangeable, provided the bounds have no dependency on any of the variables further in.

Example 4.7.2. Evaluate the triple integral $\iint_B f(x,y,z)\mathrm{d}V$ where B is the box given by

$$B = \{(x,y,z)|x \in [0,1], y \in [-1,2], z \in [0,3]\}$$

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

Solution: Just plug in.

$$\begin{aligned}
\iiint\limits_B f(x,y,z)\mathrm{d}V &= \int_0^1 \int_{-1}^2 \int_0^3 xyz^2\mathrm{d}z\mathrm{d}y\mathrm{d}x \\
&= \int_0^1 x \int_{-1}^2 y \int_0^3 z^2\mathrm{d}z\mathrm{d}y\mathrm{d}x \\
&= 9\int_0^1 x \int_{-1}^2 y\mathrm{d}y\mathrm{d}x \\
&= \frac{27}{2}\int_0^1 x\mathrm{d}x = \boxed{\frac{27}{4}}
\end{aligned}$$

△

By a similar derivation as the one for double integrals, we can define triple integrals over general bounded regions $E \subset \mathbb{R}^3$. First, consider a region of the form

$$E = \{(x,y,z)|(x,y) \in D, u_1(x,y) \le z \le u_2(x,y)\}$$

Where D is the projection of E onto the xy plane. This is known as a type I region. Similar definitions are given for type II and type III regions where $x \in [u_1(y,z), u_2(y,z)]$ or $y \in [u_1(x,z), u_2(x,z)]$. The triple integral of a function f over a type I region E is given by

$$\iint\limits_E f(x)\mathrm{d}V = \iint\limits_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\mathrm{d}z\right]\mathrm{d}A$$

If D can be written as

$$D = \{(x,y)|x \in [a,b], y \in [h_1(x), h_2(x)]\}$$

Then our triple integral becomes

$$\iiint\limits_E f(x)\mathrm{d}V = \int_a^b \int_{h_1(x)}^{h_2(x)} \int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\mathrm{d}z\mathrm{d}y\mathrm{d}x$$

Note that there are similar definitions for other types of regions. The general rule to follow is that the bounds with the most dependencies go on the inside.

Example 4.7.3. Evaluate $\iiint_E z\mathrm{d}V$ where E is the solid tetrahedron bounded by the four planes $x = 0$, $y = 0$, $z = 0$, and the $x + y + z = 1$.

Solution: First, we can find the upper x bound by setting $y = z = 0$, and solving to obtain $x = 1$. Then, the upper y bound as a function of x becomes $y = 1 - x$. And finally, the upper z bounds as a function of x and y is $z = 1 - y - x$. So our integral is,

$$\iiint\limits_E z\mathrm{d}V = \int_0^1 \int_0^{1-x} \int_0^{1-y-x} z\mathrm{d}z\mathrm{d}y\mathrm{d}x \tag{4.7.1}$$

$$= \frac{1}{2}\int_0^1 \int_0^{1-x} (1-y-x)^2\mathrm{d}y\mathrm{d}x \tag{4.7.2}$$

$$= -\frac{1}{6}\int_0^1 (1-y-x)^3\Big|_{y=0}^{y=1-x} \mathrm{d}x \tag{4.7.3}$$

$$= -\frac{1}{6}\int_0^1 (x-1)^3\mathrm{d}x \tag{4.7.4}$$

$$= -\frac{1}{24}(x-1)^4\Big|_0^1 = \boxed{\frac{1}{24}} \tag{4.7.5}$$

△

Example 4.7.4. Evaluate $\iiint_E \sqrt{x^2+z^2}\mathrm{d}V$ where E is the region bounded by the paraboloid $y = x^2 + z^2$ and the plane $y = 4$.

Solution: First, let's strategize on how we want to set up our region. Notice that the function we're looking at–$f(x, y, z) = \sqrt{x^2+z^2}$ is extremely difficult to integrate with x or z, but very easy to integrate with y. So we'll want to set up our integral in the form

$$I = \iiint\limits_E \sqrt{x^2+z^2}\mathrm{d}V = \iint\limits_D \left[\int_{h_1(x,z)}^{h_2(x,z)} \sqrt{x^2+z^2}\mathrm{d}y\right]\mathrm{d}A$$

Notice that y as a function of x and z is given by $y = x^2 + z^2$, which has a range of $[0, \infty]$ as z and x vary freely. Therefore, our lower bound is $x^2 + z^2$ and our upper bound is 4.

Now, since our region is symmetrical about x and z, we can decide to use them in either order for the outer two integrations. We'll arbitrarily choose z, but the exact same result can be achieved by choosing x.

Our xz slices of the paraboloid with $y = k$ are circles with radius $\sqrt{k}$. So x will vary at most between $-\sqrt{k}$ and $\sqrt{k}$, where k is the largest circle in all of our xz slices. The largest circle is $x^2 + z^2 = 4$ when $y = 4$, so our x bounds are $x \in [-2, 2]$.

Now, if $x^2 + z^2 = 4$, then $z = \pm\sqrt{4-x^2}$. Combining all of this, our triple integral is

$$I = \iiint\limits_E \sqrt{x^2+z^2}\mathrm{d}V = \int_{-2}^{2} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_{x^2+z^2}^{4} \sqrt{x^2+z^2}\mathrm{d}y\mathrm{d}z\mathrm{d}x$$

$$= \int_{-2}^{2} \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} (4-x^2-z^2)\sqrt{x^2+z^2}\mathrm{d}z\mathrm{d}x$$

This region is much easier to integrate using polar coordinates in the xz plane. With the substitution

$r = \sqrt{x^2+z^2}$, $x = r\cos\theta$, and $z = r\sin\theta$, we can write

$$\begin{aligned}
I &= \int_0^{2\pi}\int_0^2 (4-r^2)r^2 \mathrm{d}r\mathrm{d}\theta \\
&= \int_0^{2\pi} \mathrm{d}\theta \int_0^2 4r^2 - r^4 \mathrm{d}r \\
&= 2\pi\left[\frac{4}{3}r^3 - \frac{1}{5}r^5\right]_0^2 = 2\pi\left[\frac{32}{3}-\frac{32}{5}\right] = \boxed{\frac{128\pi}{15}}
\end{aligned}$$

△

4.8 Applications of Triple Integrals

Basically everything that has an application in double integrals has an equivalent with triple integrals.

If $f(x,y,z) \geq 0$ on a region $E \subset \mathbb{R}^3$, then the volume of the surface described by $f(x,y,z) = 1$ is given by $V = \iiint_E \mathrm{d}V$.

If the density function of a solid object occupying a region $E \subset \mathbb{R}^3$ is given by $\rho(x,y,z)$, then the mass of it is given by $m = \iiint_E \rho(x,y,z)\mathrm{d}V$. Its moments about the three coordinate planes are

$$M_{yz} = \iiint\limits_E x\rho(x,y,z)\mathrm{d}V, \quad M_{xz} = \iiint\limits_E y\rho(x,y,z)\mathrm{d}V, \quad M_{xy} = \iiint\limits_E z\rho(x,y,z)\mathrm{d}V$$

And the center of mass is at the point $(\overline{x},\overline{y},\overline{z})$ where

$$(\overline{x},\overline{y},\overline{z}) = \left(\frac{M_{yz}}{m}, \frac{M_{xz}}{m}, \frac{M_{xy}}{m}\right) = \left(\frac{\iiint_E x\rho\mathrm{d}V}{\iiint_E \rho\mathrm{d}V}, \frac{\iiint_E y\rho\mathrm{d}V}{\iiint_E \rho\mathrm{d}V}, \frac{\iiint_E z\rho\mathrm{d}V}{\iiint_E \rho\mathrm{d}V}\right)$$

And the moments of inertia about the three coordinate axes are

$$I_x = \iiint\limits_E (y^2+z^2)\rho(x,y,z)\mathrm{d}V, \quad I_y = \iiint\limits_E (x^2+z^2)\rho(x,y,z)\mathrm{d}V, \quad I_z = \iiint\limits_E (x^2+y^2)\rho(x,y,z)\mathrm{d}V$$

If we have three continuous random variables X, Y, and Z with a joint probability density function given by $f(x,y,z)$ where $f(x,y,z) \geq 0$ for all (x,y,z), f must satisfy

$$P((X,Y,Z) \in E) = \iiint\limits_E f(x,y,z)\mathrm{d}V$$

and

$$\iiint\limits_{\mathbb{R}^3} f(x,y,z)\mathrm{d}V = 1$$

We're not going to do examples for this because they're literally the exact same thing as the applications of double integrals, just with three integrals this time.

4.9 Triple Integrals in Cylindrical and Spherical Coordinates

4.9.1 Cylindrical Coordinates

Recall that the cylindrical coordinate system describes a point as a (r,θ,z) triple, where (r,θ) is the polar projection of the point onto the xy-plane, and z is the height above the xy plane.

Suppose a function $f : \mathbb{R}^3 \mapsto \mathbb{R}$ is continuous on a region

$$E = \{(x,y,z)|(x,y) \in D, z \in [u_1(x,y), u_2(x,y)]\}$$

where D is given by

$$D = \{(r,\theta) | \theta \in [\alpha,\beta], r \in [h_1(\theta), h_2(\theta)]\}$$

Then, the triple integral of f over E is given by

$$\iiint_E f(x,y,z)\mathrm{d}V = \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)\mathrm{d}z\right]\mathrm{d}A = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(r\cos\theta, r\sin\theta)}^{u_2(r\cos\theta, r\sin\theta)} f(r\cos\theta, r\sin\theta, z) r\mathrm{d}z\mathrm{d}r\mathrm{d}\theta$$

Please analyze this equation and see for yourself how it comes as a natural result from the one preceeding it (for the love of god don't just memorize it).

Example 4.9.1. A solid E lies within the cylinder $x^2 + y^2 = 1$, below the plane $z = 4$, and above the paraboloid $z = 1 - x^2 - y^2$. The density at and point in the cylinder is proportional to its distance from the axis of the cylinder. Find the mass of the solid.

Solution: First, let's rewrite all of our equations so that they are in cylindrical coordinates. The equation for the cylinder becomes $r = 1$. The equation for the plane remains the same as $z = 4$. The equation for the paraboloid becomes $z = 1 - r^2$.

We can also write an expression for the density function. The distance to the axis of the cylinder is given by $\sqrt{x^2 + y^2}$, so the density function is

$$\rho = k\sqrt{x^2 + y^2} = kr$$

For some $k \in \mathbb{R}$. Therefore, the mass is given by

$$\begin{aligned} m &= \iiint_E \rho \mathrm{d}V \\ &= \int_0^{2\pi} \int_0^1 \int_{1-r^2}^4 kr\, r\mathrm{d}z\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 kr^2(3 + r^2)\mathrm{d}r\mathrm{d}\theta \\ &= k\int_0^{2\pi} 3r^2 + r^4 \mathrm{d}r\mathrm{d}\theta \\ &= k\int_0^{2\pi} r^3 + \frac{1}{5}r^5\Big|_0^1 \mathrm{d}\theta = \boxed{\frac{12\pi k}{5}} \end{aligned}$$

△

Example 4.9.2. Evaluate

$$I = \int_{-2}^2 \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \int_{\sqrt{x^2+y^2}}^2 (x^2 + y^2)\mathrm{d}z\mathrm{d}y\mathrm{d}x.$$

Solution: First, we will want to rewrite in cylindrical coordinates. The z-coordinate lower bound becomes r, and the integrand becomes r^2. The x and y bounds together describe a circle of radius 2 centered around the origin, which we can rewrite in polar coordinates as $\{(r,\theta) | r \in [0,2], \theta \in [0, 2\pi]\}$. So our iterated integral becomes

$$\begin{aligned} I &= \int_0^{2\pi} \int_0^2 \int_r^2 r^2\, r\mathrm{d}z\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^2 2r^3 - r^4 \mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi} \left[\frac{1}{2}r^4 - \frac{1}{5}r^5\right]_0^2 \mathrm{d}\theta \\ &= 2\pi\left[\frac{16}{2} - \frac{32}{5}\right] = \boxed{\frac{16\pi}{5}} \end{aligned}$$

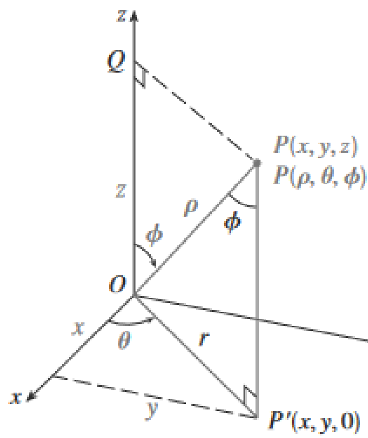
$\triangle$

4.9.2 Spherical Coordinates

Recall that we can use spherical coordinates to describe a point as $P(\rho, \theta, \phi)$, where

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

Take a moment to refresh yourself on how these relationships come from geometrically. Try and justify each equation to yourself.

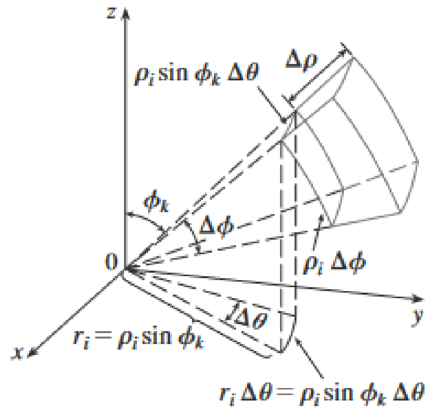


The spherical counterpart to the box is the **spherical wedge** given by

$$W = \{(\rho, \theta, \phi) | \rho \in [a, b], \theta \in [\alpha, \beta], \phi \in [\gamma, \delta]\}$$

Where $a \geq 0$, $\beta - \alpha \leq 2\pi$, and $\delta - \gamma \leq \pi$.

To come to a formula for the triple integral over this region, consider slicing the region into many mini-wedges with lower coordinates $(\rho_i, \theta_j, \phi_k)$ and upper coordinates $(\rho_i + \Delta\rho_i, \theta_j + \Delta\theta_j, \phi_k + \Delta\phi_k) = (\rho_{i+1}, \theta_{j+1}, \phi_{k+1})$. Let there be l partitions of ρ, m partitions of θ, and n partitions of ϕ. See the below figure.



First, let's analyze the inner sphere of the diagram.

The length of the top and bottom arcs is given by $\rho_i \sin \phi_k \Delta\theta_j$ and the length of the side arcs is given by $\rho_i \Delta\phi_k$. The area can then be approximated as

$$\Delta A_{ijk} \approx (\rho_i \sin \phi_k \Delta\theta_j)(\rho_i \Delta\phi_k) = \rho_i^2 \sin \phi_k \Delta\theta_j \Delta\phi_k$$

For the outer sphere, its top and bottom arcs are given by $\rho_{i+1} \sin \phi_{k+1} \Delta\theta_{j+1}$ and its side arcs are given by $\rho_{i+1} \Delta\phi_k$.

We can approximate a lower and upper limit for the volume between these two spheres by using the inner sphere's area (times the change in depth $\Delta\rho_i$) for the lower limit and the outer sphere's area (times the change in depth $\Delta\rho_i$) for the upper limit. Therefore,

$$\rho_i^2 \sin \phi_k \Delta\rho_i \Delta\theta_j \Delta\phi_k \leq \Delta V_{ijk} \leq \rho_{i+1}^2 \sin \phi_{k+1} \Delta\rho_i \Delta\theta_j \Delta\phi_k$$

We can then express the volume as a function of the particular sample point $\rho \in [\rho_i, \rho_{i+1}]$ and $\phi \in [\phi_k, \phi_{k+1}]$ we choose to construct our sphere (think of this as choosing some "slice" between the inner and outer sphere and extending that to have the correct depth):

$$\mathcal{V}(\rho, \phi) = \rho^2 \sin \phi \Delta\rho_i \Delta\theta_j \Delta\phi_k$$

We can then express our lower and upper limits for V_{ijk} in terms of this function.

$$\mathcal{V}(\rho_i, \phi_k) \leq \Delta V_{ijk} \leq \mathcal{V}(\rho_{i+1}, \phi_{k+1})$$

Now, because $\mathcal{V}$ is continuous, we can use the intermediate value theorem to state that there exists some ρ_i^*, ϕ_k^* on the square given by the cartesian product $[\phi_i, \phi_{i+1}] \times [\rho_i, \rho_{i+1}]$ such that $\mathcal{V}(\rho_i^*, \phi_k^*) = \Delta V_{ijk}$ and that

$$\Delta V_{ijk} = [\rho_i^*]^2 \sin \phi_k^* \Delta\rho_i \Delta\theta_j \Delta\phi_k$$

at that particular point.

Now, we can sum over all of those individual volumes to get the total volume given by our Riemann sum.

$$\begin{aligned}\iiint_W f(x,y,z)\mathrm{d}V &= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} f(x_{ijk}^*, y_{ijk}^*, z_{ijk}^*) \Delta V_{ijk} \\ &= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} f(\rho_i^* \cos\theta_j^* \sin\phi_k^*, \rho_i^* \sin\theta_j^* \sin\phi_k^*, \rho_i^* \cos\phi_k^*)[\rho_i^*]^2 \sin\phi_k^* \Delta\rho_i \Delta\theta_j \Delta\phi_k\end{aligned}$$

which we can recognize as representing the iterated integral

$$\iiint\limits_W f(x,y,z)\mathrm{d}V = \int_\gamma^\delta \int_\alpha^\beta \int_a^b f(\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)\rho^2 \sin\phi \,\mathrm{d}\rho\,\mathrm{d}\theta\,\mathrm{d}\phi$$

So the general process we will follow when setting up triple integrals in spherical coordinates is to convert the function from cartesian to spherical, setting up appropriate limits of integration, and replacing $\mathrm{d}V$ with $\rho^2 \sin\phi \,\mathrm{d}\rho\,\mathrm{d}\theta\,\mathrm{d}\phi$.

Similarly to in cylindrical coordinates, if ρ varies with ϕ and θ, we can just replace the ρ bounds a and b with the appropriate functions of ϕ and θ, with no other changes required.

Example 4.9.3. Evaluate $\iiint_B e^{(x^2+y^2+z^2)^{3/2}}\mathrm{d}V$ where B is the unit ball.

Solution: First, we can express our integrand in spherical coordinates. Recall that $\rho = (x^2+y^2+z^2)^{1/2}$, so we can rewrite

$$\exp\Big((x^2+y^2+z^2)^{3/2}\Big) = \exp\big(\rho^3\big)$$

Now, find that the unit ball can be expressed in spherical coordinates as

$$B = \{(\rho,\theta,\phi)|\rho \in [0,1],\, \theta \in [0,2\pi],\, \phi \in [0,\pi]\}$$

Therefore, we can write

$$\begin{aligned}
\iiint\limits_B e^{(x^2+y^2+z^2)^{3/2}}\mathrm{d}V &= \int_0^\pi \int_0^{2\pi} \int_0^1 (\rho^2\sin\phi)\, e^{\rho^3}\mathrm{d}\rho\mathrm{d}\theta\mathrm{d}\phi \\
&= \frac{1}{3}\int_0^\pi \int_0^{2\pi} (e-1)\sin\phi\mathrm{d}\theta\mathrm{d}\phi \\
&= \frac{2\pi}{3}(e-1)\int_0^\pi \sin\phi\,\mathrm{d}\phi = \boxed{\frac{4\pi}{3}(e-1)}
\end{aligned}$$

△

Example 4.9.4. Use spherical coordinates to find the volume of the solid that lies above the cone $z = \sqrt{x^2+y^2}$ and below the sphere $x^2+y^2+z^2 = z$.

Solution: Note that the sphere $x^2+y^2+z^2 = z$ is centered at $(0,0,1/2)$ and has a radius of $1/2$ (this can be verified by completing the square to obtain $x^2+y^2+(z-1/2)^2 = 1/4$). Writing it in polar coordinates, we find $\rho^2 = \rho\sin\phi$, or just $\rho = \sin\phi$.

The cone $z = \sqrt{x^2+y^2}$ can also be written as $z^2 = x^2+y^2$, or $\rho^2\sin^2\phi = \rho^2\cos^2\phi(\cos^2\theta + \sin^2\theta)$ this further reduces down to $\sin\phi = \cos\phi$. Therefore, the cone's equation becomes $\phi = \pi/4$ (as this is the point where sin and cos intersect on the first quadrant).

So our region of integration becomes

$$S = \{(\rho,\theta,\phi)|\rho \in [0,\cos\phi],\, \theta \in [0,2\pi],\, \phi \in [0,\pi/4]\}$$

and the iterated integral is

$$\begin{aligned}
I = \iiint\limits_S \mathrm{d}V &= \int_0^{\pi/4} \int_0^{2\pi} \int_0^{\cos\phi} \rho^2\sin\phi\,\mathrm{d}\rho\,\mathrm{d}\theta\,\mathrm{d}\phi \\
&= \frac{1}{3}\int_0^{\pi/4} \int_0^{2\pi} (\cos\phi)^3\sin\phi\,\mathrm{d}\theta\,\mathrm{d}\phi \\
&= \frac{2\pi}{3}\int_0^{\pi/4} (\cos\phi)^3\sin\phi\,\mathrm{d}\phi
\end{aligned}$$

With the substitution $u = \cos\phi$,

$$\begin{aligned} I &= -\frac{2\pi}{3}\int_1^{\sqrt{2}/2} u^4 \mathrm{d}u \\ &= \frac{\pi}{6}\left[1 - \frac{1}{4}\right] = \boxed{\frac{\pi}{8}} \end{aligned}$$

△

4.10 Change of Variables in Multiple Integrals

In single-variable calculus we often use a change of variable (substitution) to simplify an integral using the formula

$$\int_a^b f(x)\mathrm{d}x = \int_c^d f(g(u))g'(u)\mathrm{d}u$$

Applying change of variables is also often useful in multivariable calculus, such as when we convert double integrals into polar form with the formula

$$\iint\limits_R f(x,y)\mathrm{d}y\mathrm{d}x = \iint\limits_S f(r\cos\theta, r\sin\theta) r\mathrm{d}r\mathrm{d}\theta$$

You may notice that there are two things we have to change when we wish to apply change of variables. First, we have to change our bounds (which is shown above when the region of integration goes from R to S). And second is the extra factor of r on the end. This is known as the *Jacobian*, and is analogous to the factor of $g'(u)$ in single-variable u substitution.

In this section we will learn how to change the bounds of multiple integrals and also learn how to calculate the Jacobian for other changes of variables besides polar, spherical, and cylindrical coordinates.

4.10.1 Transformations

We can consider a general transformation T between a coordinate system U and the standard coordinate system:

$$T(u_1, \ldots, u_n) = (x_1, \ldots, x_n)$$

We can also relate x_1 through x_n to u and v with the equations

$$x_i = x_i(u_1, \ldots, u_n)$$

For integers $i \in [1, n]$. We could also write

$$T^{-1}(x_1, \ldots, x_n) = (u_1, \ldots, u_n)$$

Assuming that T is an invertible transformation.

We will usually assume that T is a C^1 transformation, which means that each $x_i(u_1, \ldots, u_n)$ has continuous first order partial derivatives.

A transformation T s really just a function whose domain and range are both subsets of $\mathbb{R}^n$. If $T(\boldsymbol{u}_1) = \boldsymbol{x}_1$, then $\boldsymbol{x}_1$ is called the **image** of $\boldsymbol{u}_1$ under T. If no two points have the same image, then T is called one-to-one or injective. All injective transformations have an inverse T^{-1}. If every point in $\mathbb{R}^n$ has a corresponding preimage under T, then T is said to be *onto* $\mathbb{R}^n$ or surjective. If T is both injective and surjective, then it is called **bijective**.

If a region R is formed by a list of points in the $\boldsymbol{x}$ world, then $T(R)$ refers to the region formed by applying T to every $\boldsymbol{x} \in R$. We also call $T(R)$ the image of R under T.

Example 4.10.1. A transformation T is defined by the equations

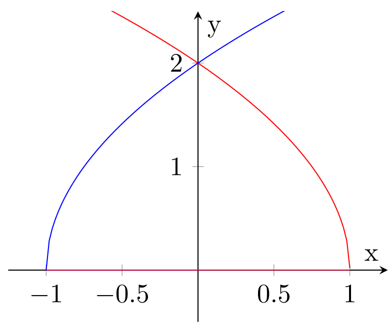
$$x = u^2 - v^2 \quad y = 2uv$$

Find the image of the square $S = [0,1] \times [0,1]$ which lies in the uv-plane.

Solution: We will try to split this into different segments by analyzing the four sides of the square.

1. $C_1 = \{T(u,v) | u \in [0,1], v = 0\}$. This gives us $(x,y) = T(u,0) = (u^2, 0)$ with $u \in [0,1]$. This will just be the line segment going along the x axis from 0 to 1.
2. $C_2 = \{T(u,v) | u = 1, v \in [0,1]\}$. This gives us $(x,y) = T(1,v) = (1 - v^2, 2v)$ with $v \in [0,1]$. We can turn this into solely an expression of x and y by rearranging to find $y = v/2$ and substituting in the x equation to get $x = 1 - \frac{1}{4}y^2$ with $y \in [0,2]$ and $x \in [0,1]$.
3. $C_3 = \{T(u,v) | u \in [0,1], v = 1\}$. This gives us $(x,y) = T(u,1) = (u^2 - 1, 2u)$ with $u \in [0,1]$. We can repeat the same process as with C_2 to obtain $x = \frac{1}{4}y^2 - 1$ with $y \in [0,2]$ and $x \in [-1,0]$.
4. $C_4 = \{T(u,v) | u = 0, v \in [0,1]\}$. This gives us $(x,y) = T(0,v) = (-v^2, 0)$ with $v \in [0,1]$, which is just the line segment going along the x axis from -1 to 0.

The total region will just be the area enclosed by these four line segments. So, let's graph them. △



4.10.2 Change of Variables in Multiple Integrals

To see how a change of variables affects a double integral, first consider a transformation $T : (u,v) \mapsto (x,y)$ whose domain includes a rectangle S in the uv plane with a lower left corner at (u_0, v_0) and an upper right corner at $(u_0 + \Delta u, v_0 + \Delta v)$.

The image of S is a region R in the xy plane with one boundary point at $(x_0, y_0) = T(u_0, v_0)$ and another at $\boldsymbol{x}_1 = T(\boldsymbol{u}_1)$. The vector $T(u,v) = \boldsymbol{r}(u,v) = \langle x(u,v), y(u,v) \rangle$ is the position vector of the image of the point $\boldsymbol{u}$.

We can obtain functions for the images of each of the edges of S that stem from the lower left corner by fixing one of either u or v at (u_0, v_0) and allowing the other to vary from the up to $(u_0 + \Delta u, v_0 + \Delta v)$. This gives us

$$\boldsymbol{r}_u = \boldsymbol{r}(u, v_0), \quad u \in [u_0, u_0 + \Delta u]$$

$$\boldsymbol{r}_v = \boldsymbol{r}(u_0, v), \quad v \in [v_0, v_0 + \Delta v]$$

Then, the tangent vector to each of these image curves is given by

$$\boldsymbol{T}_u = \langle x_u(u_0, v_0), y_u(u_0, v_0) \rangle$$

$$\boldsymbol{T}_v = \langle x_v(u_0, v_0), y_v(u_0, v_0) \rangle$$

We can then approximate the image of S under T as a parallelogram with its left side given by the difference vector $\boldsymbol{r}(u_0, v_0 + \Delta v) - \boldsymbol{r}(u_0, v_0)$ and the bottom side given by $\boldsymbol{r}(u_0 + \Delta u, v_0) - \boldsymbol{r}(u_0, v_0)$.

If you look closely at these vectors, they look quite similar to the definition of the partial derivative. That is,

$$\boldsymbol{T}_u = \lim_{\Delta u \to 0} \frac{\boldsymbol{r}(u_0 + \Delta u, v_0) - \boldsymbol{r}(u_0, v_0)}{\Delta u}$$

and

$$\boldsymbol{T}_v = \lim_{\Delta v \to 0} \frac{\boldsymbol{r}(u_0, v_0 + \Delta v) - \boldsymbol{r}(u_0, v_0)}{\Delta v}$$

From this, we can say

$$\boldsymbol{r}(u_0 + \Delta u, v_0) - \boldsymbol{r}(u_0, v_0) \approx \Delta u \boldsymbol{T}_u$$
$$\boldsymbol{r}(u_0, v_0 + \Delta v) - \boldsymbol{r}(u_0, v_0) \approx \Delta v \boldsymbol{T}_v$$

The area of $T(S)$ is approximately equal to the area of the parallelogram with sides approximately given by $\Delta u \boldsymbol{T}_u$ and $\Delta v \boldsymbol{T}_v$. The area of this can be computed using the cross product

$$\Delta A = \|\Delta u \boldsymbol{T}_u \times \Delta v \boldsymbol{T}_v\| = \|\boldsymbol{T}_u \times \boldsymbol{T}_v\| \Delta u \Delta v$$

Computing the cross product,

$$\begin{aligned} \|\boldsymbol{T}_u \times \boldsymbol{T}_v\| &= \left\| \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ x_u(u_0, v_0) & y_u(u_0, v_0) & 0 \\ x_v(u_0, v_0) & y_v(u_0, v_0) & 0 \end{vmatrix} \right\| \\ &= \begin{vmatrix} x_u(u_0, v_0) & y_u(u_0, v_0) \\ x_v(u_0, v_0) & y_v(u_0, v_0) \end{vmatrix} \end{aligned}$$

The determinant that arises in this calculation is known as the *Jacobian* of T and is given a special notation.

Definition 4.10.2. The **Jacobian** of a transformation T given by $x = x(u, v)$ and $y = y(u, v)$ is

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

With this notation, we can writ a concise approximation of ΔA:

$$\Delta A \approx \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \Delta u \Delta v$$

Where the Jacobian is evaluated at (u_0, v_0).

If we divide a region S in the uv plane into rectancles S_{ij} and call their images in the xy-plane R_{ij}, we can apply the approximation from the Jacobian and find

$$\begin{aligned} \iint\limits_R f(x, y) \mathrm{d}y \mathrm{d}x &= \lim_{m,n \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_i) \Delta y \Delta x \\ &= \lim_{m,n \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(x(u_i, v_j), y(u_i, v_j)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \Delta u \Delta v \\ &= \iint\limits_D f(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \mathrm{d}u \mathrm{d}v \end{aligned}$$

Theorem 4.10.3 (Change of Variables in a Multiple Integral)**.** Suppose that T is a C^1 transformation whose Jacobian is nonzero and that maps a region S in the preimage space U to a region R in the image space X. Suppose that f is continuous on R and that R and S can be described with one variable as a function of the others. Suppose also that T is one-to-one, except perhaps on the boundary of S. Then,

$$\int\limits_R f(x_1, \ldots, x_n) \mathrm{d}^n \boldsymbol{x} = \int\limits_S f(x_1(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n)) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} \right| \mathrm{d}^n \boldsymbol{u}$$

This theorem states that we can change an integral from the X world to the U world by writing

$$\mathrm{d}^n \boldsymbol{x} = \left|\frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_n)}\right| \mathrm{d}^n \boldsymbol{u}$$

and changing the bounds appropriately.

Example 4.10.4. Use the change of variables described by the transformation $T : (x, y) \mapsto (u, v)$ where $x = u^2 - v^2$ and $y = 2uv$ to evaluate the integral $\iint_R y\mathrm{d}A$. R is the region bounded by the x axis and the curves $x = 1 - \frac{1}{4}y^2$ and $x = \frac{1}{4}y^2 - 1$.

Solution: First, notice that R is the image of the square $[0, 1] \times [0, 1]$ under the transformation T, as we previously found in Example 4.10.1.

Now, we can compute the Jacobian for a transformation from the xy-plane to the uv-plane.

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2u & -2v \\ 2v & 2u \end{vmatrix} = 4u^2 + 4v^2 > 0$$

And plug in.

$$\begin{aligned} \iint\limits_R y\mathrm{d}A &= \iint\limits_S 2uv\left|\frac{\partial(x, y)}{\partial(u, v)}\right|\mathrm{d}u\mathrm{d}v \\ &= \int_0^1 \int_0^1 [8u^3v + 8uv^3]\mathrm{d}u\mathrm{d}v \\ &= \int_0^1 [2v + 4v^3]\mathrm{d}v = 2 \end{aligned}$$

△

Example 4.10.5. Find the Jacobian determinant $\left|\frac{\partial(x,y,z)}{\partial(\rho,\phi,\theta)}\right|$ for transforming between rectangular coordinates and spherical coordinates.

$$\begin{aligned} \frac{\partial(x, y, z)}{\partial(\rho, \phi, \theta)} &= \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta\sin\phi & \rho\cos\theta\cos\phi & -\rho\sin\theta\sin\phi \\ \sin\theta\sin\phi & \rho\sin\theta\cos\phi & \rho\cos\theta\sin\phi \\ \cos\phi & -\rho\sin\phi & 0 \end{vmatrix} \\ &= \begin{vmatrix} \cos\theta\sin\phi & \rho\cos\theta\cos\phi & -\rho\sin\theta\sin\phi \\ \sin\theta\sin\phi & \rho\sin\theta\cos\phi & \rho\cos\theta\sin\phi \\ \cos\phi & -\rho\sin\phi & 0 \end{vmatrix} \\ &= \cos\phi\begin{vmatrix} \rho\cos\theta\cos\phi & -\rho\sin\theta\sin\phi \\ \rho\sin\theta\cos\phi & \rho\cos\theta\sin\phi \end{vmatrix} + \rho\sin\phi\begin{vmatrix} \cos\theta\sin\phi & -\rho\sin\theta\sin\phi \\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi \end{vmatrix} \\ &= \cos\phi\big(\rho^2\cos^2\theta\sin\phi\cos\phi + \rho^2\sin^2\theta\sin\phi\cos\phi\big) + \rho\sin\phi\big(\rho\cos^2\theta\sin^2\phi + \rho\sin^2\theta\sin^2\phi\big) \\ &= \rho^2\sin\phi\cos^2\phi(\sin^2\theta + \cos^2\theta) + \rho^2\sin^3\phi(\sin^2\theta + \cos^2\theta) \\ &= \rho^2\sin\phi(\cos^2\phi + \sin^2\phi) = \rho^2\sin\phi \\ \left|\frac{\partial(x, y, z)}{\partial(\rho, \phi, \theta)}\right| &= \left|\rho^2\sin\phi\right| = \rho^2\sin\phi \end{aligned}$$

With the absolute value bars being able to be removed because $\sin\phi > 0$ in the domain of $\phi \in [0, \pi]$. △

5 Vector Calculus

5.1 Vector Fields

A vector field on a vector space V is a function $\boldsymbol{F} : V \to V$. In other words, it is a function that takes in a vector and spits out a vector.

We can picture a vector field on $\mathbb{R}^2$ as a set that takes in all of the points in two dimensional space and assigns an arrow with a particular magnitude and direction to each one. Similar visual intuitions works for vector fields on $\mathbb{R}^3$.

If $\boldsymbol{F}(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(\boldsymbol{x})\boldsymbol{b}_i$, where the set $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots \boldsymbol{b}_n\}$ forms a basis for V, then each f_i is called a **component function** of $\boldsymbol{F}$. In the two dimensional case, we have

$$\boldsymbol{F}(x, y) = \langle P(x, y), Q(x, y)\rangle$$

and in the three dimensional case, we have

$$\boldsymbol{F}(x, y, z) = \langle P(x, y, z), Q(x, y, z), R(x, y, z)\rangle$$

Functions like P, Q, and R which take in a vector input and have a scalar output are sometimes known as *scalar fields*. Notationally, these are denoted like $P : V \to \mathbb{F}$, where V is a vector space and $\mathbb{F}$ is a field (such as $\mathbb{R}$ or $\mathbb{C}$).

Definition 5.1.1. If $\boldsymbol{F} : V \to V$ is a vector field on a vector space V with basis $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$, and

$$\boldsymbol{F}(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(\boldsymbol{x})\boldsymbol{b}_i$$

then $\boldsymbol{F}$ is continuous on a region R if and only if each f_i is continuous on R.

One application of vector fields is the concept of a **velocity field**. Imagine a pipe with water flowing through it. At each point in the pipe, the water is flowing in a particular direction with a particular speed. Then, a function $\boldsymbol{V} : \mathbb{R}^3 \to \mathbb{R}^3$ assigns a velocity vector to the water at each point in the pipe. These types of fields have many applications in physics.

Another example is a **force field**, which describes a force being applied to an object at any particular point.

The famous equation *Newton's Law of Gravitation* describes the force between two masses m and M separated by a displacement vector $\boldsymbol{r}$:

$$\|\boldsymbol{F}\| = G\frac{mM}{\|\boldsymbol{r}\|^2}$$

this force will attract the two masses together, so its direction is given by $\boldsymbol{F}/\|\boldsymbol{F}\| = -\boldsymbol{r}/\|\boldsymbol{r}\|$, and we find

$$\boldsymbol{F} = G\frac{mM\boldsymbol{r}}{\|\boldsymbol{r}\|^3}$$

With this, we can also define the notion of a **gravitaional field**, which describes the force of gravity applied to a point mass with $m = 1$. This gravitational field is usually denoted by $\boldsymbol{g}$. If we set $m = 1$ in the above equation, we find

$$\boldsymbol{g} = -\frac{GM\boldsymbol{r}}{\|\boldsymbol{r}\|^3}$$

This gives us a convenient expression for the gravitational force on an object: $\boldsymbol{F}_g = m\boldsymbol{g}$. The strength of this gravitational field on the surface of the earth is the constant $g = 9.81$ meters per second squared, which gives us the famous result $\|\boldsymbol{F}_g\| = mg$.

Similarly, we can define the electric field from a particle with charge Q as

$$\boldsymbol{E} = \frac{1}{4\pi\epsilon_0}\frac{Q\boldsymbol{r}}{\|\boldsymbol{r}\|^3}$$

where ϵ_0 is a constant.

5.2 Gradient Fields

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, recall that the gradient of f is given by

$$\boldsymbol{\nabla} f = \left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right\rangle$$

Notice that $\boldsymbol{\nabla} f$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$, and is in fact a vector field. This is known as a **gradient vector field**.

A vector field $\boldsymbol{F}$ is called a **conservative vector field** if there exists some function f such that $\boldsymbol{\nabla} f = \boldsymbol{F}$. For instance, the gravitational field is conservative because if we define

$$f(\boldsymbol{x}) = \frac{MG}{\|\boldsymbol{x}\|}$$

we obtain

$$\begin{aligned}
\boldsymbol{\nabla} f(\boldsymbol{x}) &= \left\langle \frac{-MGx_1}{\|\boldsymbol{x}\|^3}, \frac{-MGx_2}{\|\boldsymbol{x}\|^3}, \dots, \frac{-MGx_n}{\|\boldsymbol{x}\|^3} \right\rangle \\
&= -G\frac{M}{\|\boldsymbol{x}\|^3} \langle x_1, x_2, \dots, x_n \rangle \\
&= -G\frac{M\boldsymbol{x}}{\|\boldsymbol{x}\|^3}
\end{aligned}$$

which is equivalent to our previous definition of $\boldsymbol{g}$. However, this process of verifying a field is conservative is arduous and difficult. We will find a new way soon.

5.3 Line Integrals

In this section, we will define a new form of integral that is similar to single variable integrals, except instead of integrating over a region $[a, b]$, we are integrating across a line.

Suppose C is a curve defined by the equation $\boldsymbol{r}(t) = \langle x_1(t), x_2(t), \dots, x_n(t) \rangle$ for $t \in [a, b]$. Assume that C is smooth (recall that this means that $\boldsymbol{r'}$ is continuous and $\boldsymbol{r'} \neq \boldsymbol{0}$). If we divide the interval $[a, b]$ into n sub-intervals $[t_i, t_{i+1}]$ and let $\boldsymbol{r}_i = \langle x_1(t_i), x_2(t_i), \dots, x_n(t_i) \rangle$. This divides C into n sub-arcs, with each one having length Δs_i. If we choose any point in time t_i^* on the interval $[t_i, t_{i+1}]$, we find $\boldsymbol{r}_i^* = \langle x_1(t_i^*), x_2(t_i^*), \dots, x_n(t_i^*) \rangle$.

Now, suppose we are integrating a function $f : \mathbb{R}^n \to \mathbb{R}$ whose domain includes all points on C. Then, our riemann sum becomes

$$\sum_{i=1}^{n} f(\boldsymbol{r}_i^*)\Delta s_i$$

Taking the limit of this expression, we get

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(\boldsymbol{r}_i^*)\Delta s_i = \int_C f(\boldsymbol{r})\mathrm{d}s$$

Recall that the arc length of a curve parameterized by time is

$$S = \int_a^b \|\boldsymbol{r'}(t)\|\mathrm{d}t$$

So the length of each of our mini arcs is

$$\mathrm{d}S = \|\boldsymbol{r'}(t)\|\mathrm{d}t$$

and we obtain

$$\int_C f(\boldsymbol{r})\mathrm{d}s = \int_a^b f(\boldsymbol{r}(t))\|\boldsymbol{r'}(t)\|\mathrm{d}t$$

Example 5.3.1. Evaluate $\int_C (2 + x^2 y)\mathrm{d}s$, where C is the top half of the unit circle.

Solution: The top half of the unit circle is parameterized by $(\cos t, \sin t)$ with $t \in [0, \pi]$. So, we find

$$\begin{aligned}
\int_C (2 + x^2 y)\mathrm{d}s &= \int_0^\pi 2 + (\cos t)^2 (\sin t)\sqrt{\sin^2 t + \cos^2 t}\ \mathrm{d}t \\
&= \int_0^\pi 2 + \sin t \cos^2 t\mathrm{d}t \\
&= 2t - \frac{1}{3}(\cos t)^3 \Big|_0^\pi \\
&= 2\pi + \frac{1}{3} + \frac{1}{3} = \boxed{2\pi + \frac{2}{3}}
\end{aligned}$$

△

If C is the union of finitely many disjoint curves $C_1, \ldots, C_n$ with each C_i's terminal point equal to each C_{i+1}'s initial point and each C_i being smooth, then we have

$$\int_C f(\boldsymbol{r})\mathrm{d}s = \sum_{i=1}^{n} \int_{C_i} f(\boldsymbol{r})\mathrm{d}s$$

The physical interpretation of line integrals depends on the function f that is being integrated. For instance, if f is a density function for a solid, then $\int_C f(\boldsymbol{r})\mathrm{d}s$ is the mass of a solid occupying C.

Example 5.3.2. A wire takes the shape of the semicircle $x^2 + y^2 = 1$ with $y \geq 0$, and is thicker near its base than its top. Find the mass and center of mass of the wire if its mass density is proportional to its distance from the line $y = 1$.

Solution: The mass density function is

$$\rho(x, y) = k|y - 1| = k(1 - y)$$

We can paramaterize the curve with $x = \cos t, y = \sin t$ with $t \in [0, \pi]$. Then,

$$m = \int_0^\pi k(1 - \sin t)\sqrt{\sin^2 t + \cos^2 t}\ \mathrm{d}t = k(\pi - 2)$$

By symmetry, the center of mass in the x direction is $\overline{x} = 0$. The center of mass in the y direction is given by

$$\begin{aligned}
\overline{y} &= \frac{1}{m}\int_C y\rho(x, y)\mathrm{d}s \\
&= \frac{1}{k(\pi - 2)}\int_0^\pi k\sin t - k\sin^2 t\ \mathrm{d}t \\
&= \frac{1}{\pi - 2}\left[-\cos t - \frac{1}{2}t + \frac{1}{4}\sin(2t)\right]_0^\pi \\
&= \frac{4 - \pi}{2\pi - 4}
\end{aligned}$$

△

So the center of mass is at the position

$$(\overline{x}, \overline{y}) = \left(0, \frac{4 - \pi}{2\pi - 4}\right)$$

We can obtain two other line integrals by replacing the Δs with either a Δx or Δy, giving us

$$\int_C f(x, y)\mathrm{d}x = \int_a^b f(x(t), y(t))x'(t)\mathrm{d}t$$

$$\int_C f(x,y)\mathrm{d}y = \int_a^b f(x(t),y(t))y'(t)\mathrm{d}t$$

These are known as the line integrals with respect to x and y, whereas our previous one is known as the line integral with respect to arclength.

Frequently, we see the line integrals with respect to x and y occurring together. In that case, we write

$$\int_C P(x,y)\mathrm{d}x + \int_C Q(x,y)\mathrm{d}y = \int_C P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y$$

Also note that we will frequently have to parameterize a line segment when doing line integrals. For this, it will be useful to remember the formula for a line segment going from $\boldsymbol{r}_0$ to $\boldsymbol{r}_1$:

$$\boldsymbol{r}(t) = (1-t)\boldsymbol{r}_0 + t\boldsymbol{r}_1 \quad t \in [0,1]$$

Example 5.3.3. Evaluate $\int_C y^2\mathrm{d}x + x\mathrm{d}y$ where:

1. $C = C_1$ is the line segment from $(-5,-3)$ to $(0,2)$.
2. $C = C_2$ is the arc of the parabola $x = 4 - y^2$ from $(-5,-3)$ to $(0,2)$.

Solution (1): We can parameterize C_1 with the function

$$\boldsymbol{r}_1(t) = (1-t)\langle -5,-3\rangle + t\langle 0,2\rangle = \langle 5t-5, 5t-3\rangle, \quad t \in [0,1]$$

Then, we obtain $\boldsymbol{r}_1'(t) = \langle 5,5\rangle$ and therefore,

$$\begin{aligned}
\int_{C_1} y^2\mathrm{d}x + x\mathrm{d}y &= \int_0^1 5(y(t))^2 + 5x(t)\mathrm{d}t \\
&= \int_0^1 5(5t-3)^2 + 5(5t-5)\mathrm{d}t \\
&= \int_0^1 125t^2 - 50t + 20\mathrm{d}t \\
&= \frac{125}{3}t^3 - 25t^2 + 20t\Big|_0^1 \\
&= \frac{125}{3} - 5
\end{aligned}$$

Solution (2): We can parameterize C_2 with the function

$$\boldsymbol{r}_2(t) = \langle 4-t^2, t\rangle, \quad t \in [-3,2]$$

Then, we obtain $\boldsymbol{r}_2'(t) = \langle -2t, 1\rangle$ and therefore

$$\begin{aligned}
\int_{C_2} y^2\mathrm{d}x + x\mathrm{d}y &= \int_{-3}^2 t^2(-2t) + (4-t^2)\mathrm{d}t \\
&= \int_{-3}^2 \left[-2t^3 - t^2 + 4\right]\mathrm{d}t \\
&= -\frac{1}{2}t^4 - \frac{1}{3}t^3 + 4t\Big|_{-3}^2 \\
&= 40 + \frac{5}{6}
\end{aligned}$$

△

Notice that despite these curves having the same endpoints, we got two different answers by following different paths. This indicates that the value of a line integral is not always dependent on solely the endpoints. We will explore this idea further in a little bit.

We can also notice that the answers will depend on the orientation of the curve. If we instead followed the curve going the opposite direction as C_1 (denoted by $-C_1$), we would instead get a result of 5/6. This, interestingly, happens to be off from $\int_{C_1} y^2\mathrm{d}x + x\mathrm{d}y$ by a factor of -1.

This fact will actually turn out to be generally true. That is,

$$\int_{-C} f(x,y)\mathrm{d}x + g(x,y)\mathrm{d}y = -\int_{C} f(x,y)\mathrm{d}x + g(x,y)\mathrm{d}y$$

However, if we instead integrate with respect to arc length, we get the *same* result:

$$\int_{-C} f(x,y)\mathrm{d}s = \int_{C} f(x,y)\mathrm{d}s$$

This is because the $\mathrm{d}x$s and $\mathrm{d}y$s may be negative or positive depending on the direction, but $\mathrm{d}s$, which represents a length, is always positive.

Example 5.3.4. Evaluate $\int_C y\sin z\mathrm{d}s$, where C is the circular helix given by the equations $x = \cos t$, $y = \sin t$, $z = t$, $t \in [0, 2\pi]$.

Solution: We haven't seen line integrals in space before. However, the process is much the same. First, we can arrange our equations for x, y, and z into a single vector-valued function to obtain

$$\boldsymbol{r}(t) = \langle \cos t, \sin t, t\rangle$$

Which tells us

$$\boldsymbol{r}'(t) = \langle -\sin t, \cos t, 1\rangle$$

and

$$\|\boldsymbol{r}'(t)\| = \sqrt{\sin^2 t + \cos^2 t + 1} = \sqrt{2}$$

Therefore, we find

$$\begin{aligned}
\int_C y\sin z\mathrm{d}s &= \int_0^{2\pi} y(t)\sin(z(t))\|\boldsymbol{r}'(t)\|\mathrm{d}t \\
&= \sqrt{2}\int_0^{2\pi} \sin^2 t \\
&= \frac{\sqrt{2}}{2}\int_0^{2\pi} [1 - \cos(2t)]\mathrm{d}t \\
&= \boxed{\sqrt{2}\pi}
\end{aligned}$$

△

Example 5.3.5. Evaluate $\int_C y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z$, where C consists of the line segment C_1 from $(2, 0, 0)$ to $(3, 4, 5)$ followed by the vertical line segment C_2 from $(3, 4, 5)$ to $(3, 4, 0)$.

Solution: The first observation we should make is that

$$\int_C y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z = \int_{C_1} y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z + \int_{C_2} y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z$$

This effectively splits the problem into two sub-problems. First, let's tackle the left integral that follows C_1. We can parameterize C_1 with the equation

$$\boldsymbol{r}(t) = (1-t)\langle 2, 0, 0\rangle + t\langle 3, 4, 5\rangle = \langle 2+t, 4t, 5t\rangle, \quad t \in [0, 1]$$

Therefore, we obtain $\mathrm{d}x = \mathrm{d}t$, $\mathrm{d}y = 4\mathrm{d}t$, and $\mathrm{d}z = 5\mathrm{d}t$. This allows us to rewrite our integral

$$\begin{aligned}\int_{C_1} y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z = \int_0^1 4t + 20t + 5(2+t)\mathrm{d}t \\ = \int_0^1 [29t + 10]\mathrm{d}t = \frac{29}{2}t^2 + 10t\Big|_0^1 = 24.5\end{aligned}$$

Now, we can tackle C_2. It is much the same process. We can find

$$\boldsymbol{r}(t) = (1-t)\langle 3,4,5\rangle + t\langle 3,4,0\rangle = \langle 3,4,5-5t\rangle, \quad t \in [0,1]$$

this is nice because the $\mathrm{d}x$ and $\mathrm{d}y$ terms will just go to zero. So our integral just becomes

$$\int_{C_2} y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z = \int_0^1 -15\mathrm{d}t = -15$$

So we just add these up to obtain the value of the integral over the whole path:

$$\int_C y\mathrm{d}x + z\mathrm{d}y + x\mathrm{d}z = 24.5 - 15 = 9.5$$

△

5.4 Line Integrals of Vector Fields

Recall that the work done on a particle moving in a straight line from $\boldsymbol{a}$ to $\boldsymbol{b}$ with a particular force $\boldsymbol{F}$ is given by $W = \boldsymbol{F} \cdot (\boldsymbol{b} - \boldsymbol{a})$. Suppose now that we wished to compute the work done on an object moving in some force field $\boldsymbol{F}(x,y,z) = \langle P, Q, R\rangle$ along an arbitrary path C. One way that we may be able to do this is by splitting C into infinitely many sub-arcs and adding up all of the differential works from each path.

This should be starting to sound a little familiar.

Say we split C into n sub-arcs C_iC_{i+1} with lengths Δs_i. Now, we choose any point $P_i^*(x_i^*, y_i^*, z_i^*)$ on each sub-arc and compute the work using the displacement from P_i to P_{i+1}.

$$W_i = \boldsymbol{F}(x_i^*, y_i^*, z_i^*) \cdot \Delta s_i \boldsymbol{T}_i$$

where $\boldsymbol{T}_i$ is the tangent vector to C at P_i^*. This can then be summed up over each sub-arc to obtain

$$W = \lim_{n\to\infty} \sum_{i=1}^n [\boldsymbol{F}(x_i^*, y_i^*, z_i^*) \cdot \boldsymbol{T}_i]\Delta s_i$$

which can then be easily turned into integral form:

$$W = \int_C \boldsymbol{F} \cdot \boldsymbol{T}\mathrm{d}s$$

If C is represented by some $\boldsymbol{r}(t)$ with $t \in [a,b]$, then we can recall that $\boldsymbol{T} = \boldsymbol{r}'/\|\boldsymbol{r}'\|$ and $\mathrm{d}s = \|\boldsymbol{r}'\|\mathrm{d}t$. This allows us to rewrite:

$$W = \int_a^b \left[\boldsymbol{F}(\boldsymbol{r}(t)) \cdot \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|}\right] \|\boldsymbol{r}'(t)\|\mathrm{d}t$$

and we obtain our final formula

$$W = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)\mathrm{d}t$$

which is also written as

$$W = \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

This new type of line integral shows up in many applications, especially in physics.

Example 5.4.1. Find the work done by the force field $\boldsymbol{F}(x,y) = \left\langle x^2, -xy \right\rangle$ in moving a particle counter-clockwise around the quarter of the unit circle in the first quadrant.

Solution: We can first parameterize the path as

$$\boldsymbol{r}(t) = \langle \cos t, \sin t \rangle, \quad t \in [0, \pi/2]$$

Then, we find $\boldsymbol{r}'(t) = \langle -\sin t, \cos t \rangle$. This will give us

$$\begin{aligned}
\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_0^{\pi/2} \left\langle x^2, -xy \right\rangle \cdot \langle x'(t), y'(t) \rangle \,\mathrm{d}t \\
&= \int_0^{\pi/2} \left[-2 \sin t \cos^2 t\right] \mathrm{d}t \\
&= \frac{2}{3} (\cos t)^3 \Big|_0^{\pi/2} = -\frac{2}{3}
\end{aligned}$$

△

Another thing to note is that $\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = -\int_{-C} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ because the unit tangent vector $\boldsymbol{T}$ is multiplied by -1 when reversing the path.

Example 5.4.2. Evaluate $\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ where $\boldsymbol{F}(x,y,z) = \langle xy, yz, zx \rangle$ and C is the twisted cubic given by $\boldsymbol{r}(t) = \left\langle t, t^2, t^3 \right\rangle$, $t \in [0,1]$.

Solution: Pretty simple–just plug in.

$$\begin{aligned}
\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_0^1 \left\langle t^3, t^5, t^4 \right\rangle \cdot \left\langle 1, 2t, 3t^2 \right\rangle \mathrm{d}t \\
&= \int_0^1 [5t^6 + t^3] \mathrm{d}t = \frac{5t^7}{7} + \frac{t^4}{4}\Big|_0^1 = \frac{5}{7} + \frac{1}{4}
\end{aligned}$$

△

One more thing to note is the connection between line integrals of vector fields and line integrals of scalar fields. Suppose $\boldsymbol{F} = \langle P, Q, R \rangle$. Then,

$$\begin{aligned}
\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) \mathrm{d}t \\
&= \int_a^b \langle P(x(t), y(t), z(t)), Q(x(t), y(t), z(t)), R(x(t), y(t), z(t)) \rangle \cdot \langle x'(t), y'(t), z'(t) \rangle \,\mathrm{d}t \\
&= \int_a^b x'(t) P(x(t), y(t), z(t)) + y'(t) Q(x(t), y(t), z(t)) + z'(t) R(x(t), y(t), z(t)) \mathrm{d}t \\
&= \int_C P \mathrm{d}x + Q \mathrm{d}y + R \mathrm{d}z
\end{aligned}$$

5.5 The Fundamental Theorem for Line Integrals

There is an analogue to the fundamental theorem of calculus for line integrals.

Definition 5.5.1. Let C be a smooth curve given by the vector function $\boldsymbol{r}(t)$ with $t \in [a,b]$. Let f be a differentiable function of multiple variables whose gradient vector $\boldsymbol{\nabla} f$ is continuous on C. Then,

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

Proof. Let f be a function from $\mathbb{R}^n$ to $\mathbb{R}$. Then,

$$\nabla f = \langle f_{x_1}, \ldots, f_{x_n} \rangle$$

and

$$\mathrm{d}\boldsymbol{r} = \langle x_1'(t), \ldots, x_n'(t) \rangle \, \mathrm{d}t$$

Therefore, we obtain

$$\begin{aligned}
\int_C \nabla f \cdot \mathrm{d}\boldsymbol{r} &= \int_a^b \langle f_{x_1}(\boldsymbol{r}(t)), \ldots, f_{x_n}(\boldsymbol{r}(t)) \rangle \cdot \langle x_1'(t), \ldots, x_n'(t) \rangle \, \mathrm{d}t \\
&= \int_a^b x_1'(t) f_{x_1}(\boldsymbol{r}(t)) + \cdots + x_n'(t) f_{x_n}(\boldsymbol{r}(t)) \mathrm{d}t \\
&= \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}[f(\boldsymbol{r}(t))] \mathrm{d}t \\
&= f(\boldsymbol{r}(b)) - f(\boldsymbol{r}(a))
\end{aligned}$$

□

Example 5.5.2. Find the work done by the gravitational field as a particle moves from $\boldsymbol{r}_0 = \langle x_0, y_0, z_0 \rangle$ to $\boldsymbol{r}_1 = \langle x_1, y_1, z_1 \rangle$

$$\boldsymbol{F}(x) = -\frac{mMG}{\|\boldsymbol{x}\|^3}\boldsymbol{x}$$

Solution: From earlier, we can recall that $\boldsymbol{F} = \nabla f$, where

$$f(\boldsymbol{x}) = -\frac{mMG}{\|\boldsymbol{x}\|}$$

Therefore, the work done is

$$\begin{aligned}
W &= \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_C \nabla f \cdot \mathrm{d}\boldsymbol{r} \\
&= f(\boldsymbol{r}_1) - f(\boldsymbol{r}_0) \\
&= -mMG\left(\frac{1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} - \frac{1}{\sqrt{x_0^2 + y_0^2 + z_0^2}}\right)
\end{aligned}$$

△

5.5.1 Independence of Path

Suppose C_1 and C_2 are two piecewise-smooth curves (paths) that have the same beginning and end points. We know that in general, $\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} \neq \int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$. However, one implication of the fundamental theorem for line integrals is that

$$\int_{C_1} \nabla f \cdot \mathrm{d}\boldsymbol{r} = \int_{C_2} \nabla f \cdot \mathrm{d}\boldsymbol{r}$$

Whenever ∇f is continuous. In other words, the line integral of a conservative vector field depends only on the initial and terminal point of the path.

5.5.2 Closed Curves

A curve is called **closed** if its terminal point is at the same location as its initial point. When the curve of a line integral is closed, we use a special form of notation:

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

Indicates a closed line integral. Another consequence of the fundamental theorem for line integrals is that the closed line integral of a conservative vector field is always zero.

$$\oint \boldsymbol{\nabla} f \cdot \mathrm{d}\boldsymbol{r} = f(\boldsymbol{r}_1) - f(\boldsymbol{r}_0)$$

But since $\boldsymbol{r}_1 = \boldsymbol{r}_0$,

$$f(\boldsymbol{r}_1) - f(\boldsymbol{r}_0) = 0$$

Theorem 5.5.3. $\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ is independent of path in some region D if and only if $\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = 0$ for every closed path C in D.

Theorem 5.5.4. Suppose $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field that is in an open connected region D. If $\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ is independent of path in D, then $\boldsymbol{F}$ is a conservative vector field–that is, there exists some $f : \mathbb{R} \to \mathbb{R}^n$ such that $\boldsymbol{\nabla} f = \boldsymbol{F}$.

Proof. Let $A(a_1, \ldots, a_n)$ be a fixed point in $B_1(b_1, \ldots, b_n)$ to P. We can define a function $f(b_1, \ldots, b_n)$ such that

$$f(b_1, \ldots, b_n) = \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}.$$

Additionally suppose $\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ is independent of path.

We will also introduce some new notation. Instead of writing a line integral with respect to a path, we can (in the case where the integral is independent of path) write a line integral with respect to only the beginning and end points, like so:

$$\int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_{(a_1,\ldots,a_n)}^{(b_1,\ldots,b_n)} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

Because D is open, then there must exist some hypersphere S with a nonzero radius entirely contained in D centered around A. We can now choose some new point in S: $B_2(b_1^*, \ldots, b_n)$ with $b_1^* < b_1$. This is the same point as B_1, except shifted down along only the first axis. We can then split C into two paths: C_1, which goes from A_1 to B_1. And C_2 which goes from B_1 to B_2. Then, we can write

$$f(b_1, \ldots, b_n) = \int_{(a_1,\ldots,a_n)}^{(b_1^*,\ldots,b_n)} + \int_{(b_1^*,\ldots,b_n)}^{(b_1,\ldots,b_n)} \boldsymbol{F} \cdot \mathrm{d}r$$

Because the first integral does not depend on b_1, we can find

$$\frac{\partial}{\partial b_1} f(b_1, \ldots, b_n) = 0 + \frac{\partial}{\partial b_1} \int_{(b_1^*,\ldots,b_n)}^{(b_1,\ldots,b_n)} \boldsymbol{F} \cdot \mathrm{d}r$$

If we write $\boldsymbol{F} = \langle F_1, \ldots F_n \rangle$, then

$$\int_{(b_1^*,\ldots,b_n)}^{(b_1,\ldots,b_n)} \boldsymbol{F} \cdot \mathrm{d}r = \int_{(b_1^*,\ldots,b_n)}^{(b_1,\ldots,b_n)} F_1 \mathrm{d}b_1 + \cdots + F_n \mathrm{d}b_n$$

On C_2, all of b_2 through b_n are constant, so $\mathrm{d}b_2 = \cdots = \mathrm{d}b_n = 0$. Therefore,

$$\frac{\partial}{\partial b_1} f(b_1, \ldots, b_n) = \frac{\partial}{\partial b_1} \int_{(b_1^*,\ldots,b_n)}^{(b_1,\ldots,b_n)} F_1 \mathrm{d}x_1 = \frac{\partial}{\partial b_1} \int_{b_1^*}^{b_1} F_1(t, b_2, \ldots, b_n) \mathrm{d}t = F_1(b_1, \ldots, b_n)$$

We can repeat this exact process for each b_2 through b_n to obtain

$$\frac{\partial}{\partial b_i} f(b_1, \ldots, b_n) = F_i(b_1, \ldots b_n)$$

Therefore,

$$\boldsymbol{F} = \langle F_1, \ldots, F_n \rangle = \left\langle \frac{\partial}{\partial b_1} f, \ldots, \frac{\partial}{\partial b_n} f \right\rangle$$

Which tells us that $\boldsymbol{F} = \boldsymbol{\nabla} f$, proving the theorem. □

5.6 Finding Potential Functions

The question still remains: How can we determine if a vector field is conservative or not?

Suppose it is known that $\boldsymbol{F} = \langle F_1, \dots F_n \rangle$ is conservative, where each F_i has continuous $n-1$th order partial derivatives. Then there exists some function $f : \mathbb{R} \to \mathbb{R}^n$ such that $\boldsymbol{F} = \boldsymbol{\nabla} f$. That is, each $F_i = \frac{\partial f}{\partial x_i}$.

From this, we can also obtain

$$\frac{\partial F_i}{\partial x_1 \cdots \partial x_{i-1} \partial_{x+1} \cdots \partial x_n} = \frac{\partial f}{\partial x_1 \cdots \partial x_n}$$

for all i between 1 and n. Therefore, by Clairaut's theorem, we find

$$\frac{\partial F_1}{\partial x_2 \cdots \partial x_n} = \frac{\partial F_2}{\partial x_1 \partial x_3 \cdots \partial x_n} = \cdots = \frac{\partial F_n}{\partial x_1 \cdots \partial x_{n-1}} = \frac{\partial f}{\partial x_1 \cdots \partial x_n}$$

Theorem 5.6.1. If $\boldsymbol{F} = \langle F_1, \dots F_n \rangle$ is a conservative vector field where each F_i has continuous $n-1$th order partial derivatives on a domain D, then throughout D we have

$$\frac{\partial F_1}{\partial x_2 \cdots \partial x_n} = \frac{\partial F_2}{\partial x_1 \partial x_3 \cdots \partial x_n} = \cdots = \frac{\partial F_n}{\partial x_1 \cdots \partial x_{n-1}}$$

However, this theorem is often too clunky to apply for any regions of a higher dimension than $\mathbb{R}^2$. We will instead focus solely on the two dimensional version of this theorem:

Theorem 5.6.2. If $\boldsymbol{F} = \langle F_1, F_2 \rangle$ is a conservative vector field where F_1 and F_2 have continuous first order partial derivatives on a domain D, then throughout D we have

$$\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$$

The converse of these theorems (which, you may notice, is the critical part for determining whether a field is conservative) only applies under a special type of region. To explain this, we must first introduce the idea of a **simple curve**. Simple curves are curves which do not intersect themselves at any point except perhaps at the endpoints.

The converse only applies over so-called **simply-connected regions**. A simply-connected region is a region D such that every simple closed curve in D encloses points that are in D. Physically speaking, this is a region with no "holes".

Therefore, we obtain

Theorem 5.6.3. If $\boldsymbol{F} = \langle F_1, \dots F_n \rangle$ is a vector field where each F_i has continuous $n-1$th order partial derivatives on an open simply-connected domain D and

$$\frac{\partial F_1}{\partial x_2 \cdots \partial x_n} = \frac{\partial F_2}{\partial x_1 \partial x_3 \cdots \partial x_n} = \cdots = \frac{\partial F_n}{\partial x_1 \cdots \partial x_{n-1}}$$

throughout D, then $\boldsymbol{F}$ is conservative.

And the appropriate version in $\mathbb{R}^2$:

Theorem 5.6.4. If $\boldsymbol{F} = \langle F_1, F_2 \rangle$ is a vector field where F_1 and F_2 have continuous first order partial derivatives on an open, simply-connected domain D and

$$\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$$

throughout D, then $\boldsymbol{F}$ is conservative.

We will skip the proofs of these theorems for now, as they become much simpler with the help of a future topic (Green's Theorem).

Example 5.6.5. Determine whether the vector field $\boldsymbol{F}(x, y) = \langle x - y, x - 2\rangle$ is conservative.

Solution: $\frac{\partial}{\partial y}(x - y) = -1$ and $\frac{\partial}{\partial x}(x - 2) = 1$. Because $\partial F_1/\partial y \neq \partial F_2/\partial x$, $\boldsymbol{F}$ is not conservative. △

Example 5.6.6. Determine whether the vector field $\boldsymbol{F}(x, y) = \left\langle 3 + 2xy, x^2 - 3y^2\right\rangle$ is conservative.

Solution: The domain of $\boldsymbol{F}$–$\mathbb{R}^2$–is open and simply connected. Further, $\frac{\partial}{\partial y}(3 + 2xy) = 2x$ and $\frac{\partial}{\partial x}\left(x^2 - 3y^2\right) = 2x$. Because $\partial F_1/\partial y = \partial F_2/\partial x$, $\boldsymbol{F}$ is conservative. △

Every function f such that $\boldsymbol{\nabla} f = \boldsymbol{F}$ is known as a potential function for $\boldsymbol{F}$. To find the potential function of a conservative field, we employ "partial integration" on each of the components of $\boldsymbol{F}$ to find the functions that are partially differentiated to become the original components. It is easiest to illustrate with an example.

Example 5.6.7. Find all potential functions of $\boldsymbol{F}(x, y) = \left\langle 3 + 2xy, x^2 - 3y^2\right\rangle$.

Solution: Let f be a function such that $\boldsymbol{\nabla} f = \boldsymbol{F}$. Then, we know

$$f_x = 3 + 2xy \quad \text{and} \quad f_y = x^2 - 3y^2.$$

If we perform partial integration on f_x, we find that

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

However, we also have to include constants of integration. Instead of just being one simple $+C$ on the end, however, we must notice that because our integration is only allowing x to change, all functions of purely y are also constant. Therefore,

$$f(x, y) = 3x + x^2 y + C_1(y)$$

Doing the same process on f_y, we find that

$$f(x, y) = x^2 y - y^3 + C_2(x)$$

We can notice that, when combined, these two functions describing f can be combined to tell us more about f. The first function, $C_1(y)$, can be matched with the $-y^3$ term. Similarly, $C_2(x)$ can be matched with $3x$. Overall, we obtain

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

Where C is a constant, not dependent on either x or y. △

Example 5.6.8. If $\boldsymbol{F}(x, y, z) = \left\langle y^2, 2xy + e^{3z}, 3ye^{3z}\right\rangle$ is conservative, then find all potential functions of $\boldsymbol{F}$.

Solution: We have

$$f_x(x, y, z) = y^2 \tag{5.6.1}$$
$$f_y(x, y, z) = 2xy + e^{3z} \tag{5.6.2}$$
$$f_z(x, y, z) = 3ye^{3z} \tag{5.6.3}$$

Which gives us

$$f(x, y, z) = xy^2 + C_1(y, z) \tag{5.6.4}$$
$$f(x, y, z) = xy^2 + ye^{3z} + C_2(x, z) \tag{5.6.5}$$
$$f(x, y, z) = ye^{3z} + C_3(x, y) \tag{5.6.6}$$

Combining these, we find

$$f(x, y, z) = xy^2 + ye^{3z} + C$$

△

5.7 Conservation of Energy

We can apply the ideas of this chapter to a physical situation. Consider a continuous force field $\boldsymbol{F}$ that moves an object along a path C given by $\boldsymbol{r}(t)$, $t \in [a, b]$, where $\boldsymbol{r}(a) = A$ is the initial point and $\boldsymbol{r}(b) = B$ is the terminal point. According to Newton's second law of motion, the force $\boldsymbol{F}(\boldsymbol{r}(t))$ at a point on C is related to the acceleration $\boldsymbol{a}(t) = \boldsymbol{r}''(t)$ by the equation

$$\boldsymbol{F}(\boldsymbol{r}(t)) = m\boldsymbol{r}''(t)$$

So the work done by the force on the object is

$$\begin{aligned} W &= \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)\mathrm{d}t \\ &= \int_a^b m\boldsymbol{r}''(t) \cdot \boldsymbol{r}'(t)\mathrm{d}t \\ &= \frac{1}{2}m \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{r}'(t) \cdot \boldsymbol{r}'(t)]\mathrm{d}t \\ &= \frac{1}{2}m \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{r}'(t)\|^2\mathrm{d}t \\ &= \frac{1}{2}m\|\boldsymbol{r}'(t)\|^2\Big|_a^b \\ &= \frac{1}{2}m\|\boldsymbol{r}'(b)\|^2 - \frac{1}{2}m\|\boldsymbol{r}'(a)\|^2 \end{aligned}$$

If you've taken a physics course, you may remember this formula: $W = \Delta K$, where $K = \frac{1}{2}mv^2$ is the kinetic energy.

Now, if we make the further assumption that $\boldsymbol{F}$ is a conservative force field, then we can write $\boldsymbol{F} = \boldsymbol{\nabla} f$. In physics, we define the **potential energy** of an object as $U(x, y, z) = -f(x, y, z)$. Therefore, we have $\boldsymbol{F} = -\boldsymbol{\nabla} U$ (This formula might also look familiar, although you may have learned it as $F = -\frac{\mathrm{d}U}{\mathrm{d}x}$).

From this, we can write

$$\begin{aligned} W &= -\int_C \boldsymbol{\nabla} U \cdot \mathrm{d}r \\ &= -[U(\boldsymbol{r}(b)) - U(\boldsymbol{r}(a))] \\ &= U(\boldsymbol{r}(a)) - U(\boldsymbol{r}(b)) \end{aligned}$$

Combining this with our previous equation, we find

$$U(\boldsymbol{r}(a)) - U(\boldsymbol{r}(b)) = K(\boldsymbol{r}(b)) - K(\boldsymbol{r}(a))$$

Which can be rearranged to obtain

$$K(\boldsymbol{r}(a)) + U(\boldsymbol{r}(a)) = K(\boldsymbol{r}(b)) + U(\boldsymbol{r}(b))$$

which is the famous expression of **the law of conservation of energy**–the sum of the kinetic and potential energies of an object moving through a conservative force field is constant.

5.8 Green's Theorem

Green's Theorem is one of the most important results of Calculus III, as it gives the relationship between the line integral around a simple closed curve C and the double integral of the plane region D bounded by C.

In stating Green's Theorem, we use the convention that the **positive orientation** of a simple closed curve C refers to single counterclockwise traversal of C. This, if C is given by the vector function $\boldsymbol{r}(t)$ with $t \in [a, b]$, then the region D appears on the left hand side of an observer following C along $\boldsymbol{r}$ (assuming the observer is looking in the direction that they are moving).

Theorem 5.8.1 (Green's Theorem)**.** Let C be the positively oriented, piece-wise smooth, simple closed curve in the plane and let D be the region bounded by C. If P and Q have continuous partial derivatives on an open region that contains D, then

$$\int_C P\mathrm{d}x + Q\mathrm{d}y = \iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A$$

The notations

$$\oint_C P\mathrm{d}x + Q\mathrm{d}y \qquad \circlearrowleft\!\!\!\!\!\!\int_C P\mathrm{d}x + Q\mathrm{d}y \quad \text{and} \quad \int_{\partial D} P\mathrm{d}x + Q\mathrm{d}y$$

is sometimes used in place of $\int_C P\mathrm{d}x + Q\mathrm{d}y$.

Green's Theorem can be regarded as the equivalent of the Fundamental Theorem of Calculus for double integrals.

In fact, if you continue to study calculus further, you'll find that Green's theorem, the Fundamental Theorem of Calculus, the Fundamental Theorem of Line Integrals, and Stokes' Theorem (we will learn about this one later) are all, in fact, one and the same. They are all special cases of the so-called *Generalized Stokes' Theorem.*

Proving Green's Theorem for the general case is extremely difficult, but we can give a proof in the special case where D can be expressed nicely as a region that can be expressed as *both* a type I and type II region. We will call such regions **simple regions**.

Proof of Green's Theorem for Simple Regions. Notice that if we are able to show that

$$\int_C P\mathrm{d}x = -\iint\limits_D \frac{\partial P}{\partial y}\mathrm{d}A$$

and

$$\int_C Q\mathrm{d}y = \iint\limits_D \frac{\partial Q}{\partial x}\mathrm{d}A$$

separately, then we have proven the whole statement. We can prove the first one of these by expressing the region D as a type I region:

$$D = \{(x,y) | x \in [a,b], y \in [g_1(x), g_2(x)]\}$$

Therefore,

$$\begin{aligned} -\iint\limits_D \frac{\partial P}{\partial y}\mathrm{d}A &= -\int_a^b \int_{g_1(x)}^{g_2(x)} \frac{\partial P}{\partial y}(x,y)\ \mathrm{d}y\mathrm{d}x \\ &= \int_a^b [P(x,g_1(x)) - P(x,g_2(x))]\mathrm{d}x \end{aligned}$$

We can also attempt to compute $\int_C P\mathrm{d}x$. Notice that C is actually the union of four piecewise-smooth curves, which we will denote by C_1, C_2, C_3, and C_4.

1. C_1 is the bottom curve where the x-coordinate ranges between a and b and the y coordinate is given by $g_1(x)$.
2. C_2 is the rightmost curve where the x-coordinate is constant at b and the y coordinate goes from $g_1(b)$ to $g_2(b)$.
3. C_3 is the top curve where the x-coordinate ranges between b and a and the y coordinate is given by $g_2(x)$.
4. C_4 is the rightmost curve where the x-coordinate is constant at a and the y coordinate goes from $g_2(a)$ to $g_1(a)$.

Therefore, $\int_C P\mathrm{d}x = \int_{C_1} P\mathrm{d}x + \int_{C_2} P\mathrm{d}x + \int_{C_3} P\mathrm{d}x + \int_{C_4} P\mathrm{d}x$.

For C_1, we can write

$$\int_{C_1} P(x,y)\mathrm{d}x = \int_a^b P(x, g_1(x))\mathrm{d}x$$

For C_2 and C_4, the x-coordinate is constant so they just go to zero.

For C_3, we can write

$$\int_{C_3} P(x,y)\mathrm{d}x = \int_b^a P(x, g_2(x))\mathrm{d}x = -\int_a^b P(x, g_2(x)\mathrm{d}x$$

Adding each of these up, we obtain the total line integral

$$\int_C P(x,y)\mathrm{d}x = \int_a^b P(x, g_1(x)) - P(x, g_2(x))\mathrm{d}x$$

This is exactly the expression for $-\iint\limits_D \frac{\partial P}{\partial y}\mathrm{d}A$ we found earlier, proving the first half of the theorem.

The second half is proven in pretty much the exact same manner, so we will omit it here. □

Example 5.8.2. Evaluate $\int_C x^4\mathrm{d}x + xy\mathrm{d}y$, where C is the triangular curve with vertices $(0,0)$, $(1,0)$, and $(0,1)$.

Solution: We're going to use Green's Theorem to solve this. Using Green's Theorem, we find

$$\int_C P\mathrm{d}x + Q\mathrm{d}y = \iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A$$

If we plug the functions in,

$$\begin{aligned}\int_C x^4\mathrm{d}x + xy\mathrm{d}y &= \int_0^1 \int_0^{1-x} y\mathrm{d}y\mathrm{d}x \\ &= \frac{1}{2}(1-x)^2\mathrm{d}x = -\frac{1}{6}(1-x)^3\bigg|_0^1 = \frac{1}{6}\end{aligned}$$

△

Example 5.8.3. Evaluate $\oint_C \left[3y - e^{\sin x}\right]\mathrm{d}x + \left[7x + \sqrt{y^4+1}\right]\mathrm{d}y$ where C is the circle $x^2 + y^2 = 9$.

Solution: We could just evaluate this normally but that looks, annoying, so we're going to use Green's Theorem instead. With Greens' Theorem, we find

$$\oint_C P\mathrm{d}x + Q\mathrm{d}y = \iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A$$

If we plug the functions in,

$$\begin{aligned}\oint_C \left(3y - e^{\sin x}\right)\mathrm{d}x + \left(7x + \sqrt{y^4+1}\right)\mathrm{d}y &= \int_0^{2\pi}\int_0^3 [7-3]r\mathrm{d}r\mathrm{d}\theta \\ &= 8\pi\int_0^3 r\mathrm{d}r = 36\pi\end{aligned}$$

△

Another application of Green's Theorem is in computing areas. Since the area of D is $\iint_D 1\mathrm{d}A$, we wish to choose P and Q so that

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$$

There are several possibilities, but a few of easier-to-compute ones are as follows:

$$\begin{array}{ccc} P(x,y) = 0 & P(x,y) = -y & P(x,y) = -\frac{1}{2}y \\ Q(x,y) = x & Q(x,y) = 0 & Q(x,y) = -\frac{1}{2}x \end{array}$$

Which gives us a few ways of finding the area of D:

$$A = \oint_C x\mathrm{d}y = -\oint_C y\mathrm{d}x = \frac{1}{2}\oint_C x\mathrm{d}y - y\mathrm{d}x$$

Example 5.8.4. Find the area enclosed by the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Solution: The ellipse is described by $\boldsymbol{r}(t) = \langle a\cos t, b\sin t\rangle$ with $t \in [0, 2\pi]$. We could find the area with a complicated, annoying double integral, or use Green's Theorem.

$$\begin{aligned} A = \frac{1}{2}\oint_C x\mathrm{d}y - y\mathrm{d}x &= \frac{1}{2}\int_0^{2\pi} [a\cos t][b\cos t] - [b\sin t][-a\sin t]\mathrm{d}t \\ &= \frac{1}{2}ab\int_0^{2\pi} [\cos^2 t + \sin^2 t]\mathrm{d}t = \pi ab \end{aligned}$$

△

Example 5.8.5. Evaluate $\oint_C y^2\mathrm{d}x + 3xy\mathrm{d}y$ where C is the boundary of the semicircular region D in the upper half of the xy plane between the circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 4$.

Solution: We can write D as the region

$$D = \{(r, \theta) | r \in [1, 2], \theta \in [0, \pi]\}$$

Therefore, Green's Theorem gives us

$$\begin{aligned} \oint_C y^2\mathrm{d}x + 3xy\mathrm{d}y &= \iint_D \left[\frac{\partial}{\partial x}3xy - \frac{\partial}{\partial y}y^2\right]\mathrm{d}A = \iint_D y\mathrm{d}A \\ &= \int_0^{\pi}\int_1^2 r^2\sin\theta\mathrm{d}r\mathrm{d}\theta \\ &= \frac{7}{3}\int_0^{\pi}\sin\theta\mathrm{d}\theta = \frac{14}{3} \end{aligned}$$

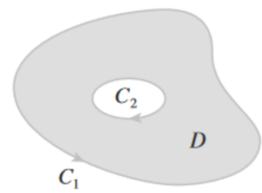
△

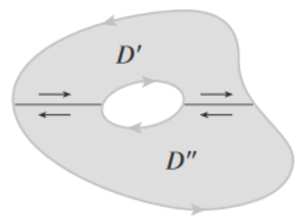
We can further Green's Theorem to two more types of regions. First, we have regions that are the finite union of simple regions, although we will not show that proof here.

Secondly, we have regions that are not simply-connected–that is, regions with holes.

Consider the above region D. Its boundary C can be split up into the union of the inner and outer boundaries C_1 and C_2. Because of our earlier assumption that positive orientation keeps the region on the left side, C_2 is oriented clockwise and C_1 is oriented counterclockwise.

We can draw two lines that can be used to split C in a different way that splits D into the union of two simply-connected regions D' and D'', as below:





Therefore, we are able to say:

$$\begin{aligned}\iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A &= \iint\limits_{D'} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A + \iint\limits_{D''} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A \\ &= \int_{\partial D'} P\mathrm{d}x + Q\mathrm{d}y + \int_{\partial D''} P\mathrm{d}x + Q\mathrm{d}y\end{aligned}$$

Then, we can notice that $\partial D'$ and $\partial D''$ share a common line segment, except running in opposite directions. Because it is known that

$$\int_C P\mathrm{d}x + Q\mathrm{d}y = -\int_{-C} P\mathrm{d}x + Q\mathrm{d}y$$

We can say that the line integral over these segments will cancel, and we can rewrite as

$$\iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}A = \int_{C_1 \cup C_2} P\mathrm{d}x + Q\mathrm{d}y = \int_C P\mathrm{d}x + Q\mathrm{d}y$$

Now, we can apply this to an example.

Example 5.8.6. If $\boldsymbol{F}(x, y) = (-y\hat{\boldsymbol{\imath}} + x\hat{\boldsymbol{\jmath}})/(x^2 + y^2)$, show that $\int_C \boldsymbol{F} \cdot \mathrm{d}r = 2\pi$ for *every* positively oriented simple closed path that encloses the origin.

Solution: Since C is an arbitrary closed path that encloses the origin, it will be difficult to evaluate the line integral directly.

Instead, we'll be taking a different approach via Green's Theorem. First, we will define two curves. C_1 is an arbitrary path surrounding the origin. C_2 is a circle of radius $a \in \mathbb{R}_+$ centered about the origin, with a being sufficiently small that C_2 is completely enclosed inside C_1. Both C_1 and C_2 are positively oriented–that is, the area enclosed by them (which we will call D) is always on the left of the path.

We can now use Green's Theorem to state:

$$\begin{aligned}\int_{C_1 \cup C_2} P\mathrm{d}x + Q\mathrm{d}y &= \int_{C_1} P\mathrm{d}x + Q\mathrm{d}y + \int_{C_2} P\mathrm{d}x + Q\mathrm{d}y \\ &= \iint\limits_D \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right]\mathrm{d}A \\ &= \iint\limits_D \left[\frac{y^2 - x^2}{(x^2 + y^2)^2} - \frac{y^2 - x^2}{(y^2 + x^2)^2}\right] = 0\end{aligned}$$

Because $\iint_D [Q_x - P_y]\mathrm{d}A = 0$, we must then have $\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} + \int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = 0$, or $\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = -\int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$.

Now, since C_2 is an easy-to-compute integral, we can find it which will then give us the value of $\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$.

Remember that C_2 is positively oriented with respect to D, which means that it is a *clockwise* circle. To replace it with the more familiar counter clockwise circle, we will instead integrate over $-C_2$, with the knowledge that $\int_{-C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = -\int_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$.

$$\begin{aligned}
\int_{-C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_0^{2\pi} \boldsymbol{F}(a\cos t, a\sin t) \cdot \boldsymbol{r}'(t)\mathrm{d}t \\
&= \int_0^{2\pi} \left[(-a\sin t)\frac{-a\sin t}{(a\cos t)^2 + (a\sin t)^2} + (a\cos t)\frac{a\cos t}{(a\cos t)^2 + (a\sin t)^2}\right]\mathrm{d}t \\
&= \int_0^{2\pi} \mathrm{d}t = 2\pi
\end{aligned}$$

Therefore, because $\int_{C_1} \boldsymbol{F} \cdot \mathrm{d}r = \int_{-C_2} \boldsymbol{F} \cdot \mathrm{d}r$, we have completed the problem. △

You may recall that we skipped the proof of Theorem 5.6.4, stating that it becomes easier with Green's Theorem. Now, it is time to come back to that.

As a refresher, the theorem stated that if $\boldsymbol{F} = \langle P, Q \rangle$ is a vector field in $\mathbb{R}^2$ with continuous first order partial derivatives, then $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$ for all points in a subset of D of $\boldsymbol{F}$'s domain if and only if $\boldsymbol{F}$ is conservative throughout D.

Proof of Theorem 5.6.4. Suppose that $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$. Then, Green's Theorem gives us that

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_D \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right]\mathrm{d}A = 0$$

For any closed simple path C in the domain of $\boldsymbol{F}$. Any non simple path can be split up into the finite union of simple curves.

Now, you may recall from Theorem 5.5.4 that the closed line integral of a vector field is zero if and only if that field is conservative.

To prove the inverse, suppose $\boldsymbol{F}$ is conservative. That means that $\boldsymbol{F}$ is also independent of path, and therefore $\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = 0$. Then, by Green's Theorem, we have

$$\iint\limits_D \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right]\mathrm{d}A = \oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = 0$$

Which can only be true if $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$.

Therefore, $\boldsymbol{F}$ is conservative if and only if $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$. □

5.9 Curl and Divergence

There are two more operations which will prove useful in our further study of vector fields–curl and divergence. These operations can both be thought of as different methods of "differentiation" for vector fields, although one will produce a vector field and the other will produce a scalar field.

5.9.1 Curl

Definition 5.9.1. If $\boldsymbol{F} = \langle P, Q, R \rangle$ is a vector field that has defined first order partial derivatives throughout some region $D \subset \mathbb{R}^3$, then the curl of $\boldsymbol{F}$ on D is defined by

$$\text{curl } \boldsymbol{F} = \left\langle \frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}, \frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}, \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right\rangle$$

One easy way to remember this definition is by defining the differential operator *del* ($\boldsymbol{\nabla}$) as

$$\boldsymbol{\nabla} = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$$

Then, we can remember that curl F as

$$\text{curl } \boldsymbol{F} = \boldsymbol{\nabla} \times \boldsymbol{F} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

This operator is actually useful in some other ways too. We can remember the gradient with

$$\boldsymbol{\nabla} f = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle$$

There will also be a similar trick for divergence which we will cover.

Example 5.9.2. If $\boldsymbol{F} = \left\langle xz, xyz, -y^2 \right\rangle$, find curl F.

Solution: This is as simple as plugging into the formula.

$$\begin{aligned} \boldsymbol{\nabla} \times \boldsymbol{F} &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xz & xyz & -y^2 \end{vmatrix} \\ &= \left\langle \frac{\partial}{\partial y}(-y^2) - \frac{\partial}{\partial z}(xyz), \frac{\partial}{\partial z}(xz) - \frac{\partial}{\partial x}(-y^2), \frac{\partial}{\partial x}(xyz) - \frac{\partial}{\partial y}(xz) \right\rangle \\ &= \langle -2y - xy, x, yz \rangle \end{aligned}$$

△

Another important fact about the curl is its relationship with the gradient.

Theorem 5.9.3. If $f : \mathbb{R}^3 \to \mathbb{R}$ has continuous partial derivatives over a region $D \subseteq \mathbb{R}^3$, then $\text{curl}\,(\boldsymbol{\nabla} f) = \boldsymbol{0}$.

Proof. Compute $\text{curl}\,(\boldsymbol{\nabla} f)$.

$$\begin{aligned} \text{curl}\,(\boldsymbol{\nabla} f) = \boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{vmatrix} \\ &= \left\langle \frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}, \frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}, \frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x} \right\rangle \\ &= \langle 0, 0, 0 \rangle = \boldsymbol{0} \end{aligned}$$

Recall that according to Clairaut's Theorem, the order in which mixed partials is applied does not matter. □

Because a vector field $\boldsymbol{F}$ is conservative if and only if there exists some f such that $\boldsymbol{F} = \boldsymbol{\nabla} f$, we can use the previous theorem to state that if $\boldsymbol{F}$ is conservative, we must have $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$.

The converse of this will only apply under more strict criteria, requiring that $\boldsymbol{F}$'s domain is simply connected. The proof of the converse will come later, once we have explored Stokes' Theorem.

Theorem 5.9.4. If $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a continuous vector field with a simply-connected domain whose component functions have continuous partial derivatives, then $\boldsymbol{F}$ is conservative if and only if $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$.

Example 5.9.5. Show that the vector field $\boldsymbol{F} = \left\langle xz, xyz, -y^2 \right\rangle$ is not conservative.

Solution: We earlier computed $\boldsymbol{\nabla} \times \boldsymbol{F} = \langle -2y - xy, x, yz \rangle$. This, of course, is not equal to $\boldsymbol{0}$, so $\boldsymbol{F}$ is not conservative. △

Example 5.9.6. Show that $\boldsymbol{F} = \left\langle y^2z^3, 2xyz^3, 3xy^2z^2 \right\rangle$ is conservative and find a potential function for it.

Solution: Clearly, the domain of $\boldsymbol{F}$ is simply connected because it is all of $\mathbb{R}^3$. Therefore, we can check if it is conservative by computing $\boldsymbol{\nabla} \times \boldsymbol{F}$.

$$\begin{aligned}\boldsymbol{\nabla} \times \boldsymbol{F} &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y^2z^3 & 2xyz^3 & 3xy^2z^2 \end{vmatrix} \\ &= \left\langle 6xyz^2 - 6xyz^2, 3y^2z^2 - 3y^2z^2, 2yz^3 - 2yz^3 \right\rangle = \mathbf{0}\end{aligned}$$

Therefore, $\boldsymbol{F}$ is conservative. To find a potential function, we will integrate each component function.

$$\begin{aligned}\int y^2z^3 \mathrm{d}x &= xy^2z^3 + f(y,z) \\ \int 2xyz^3 \mathrm{d}y &= xy^2z^3 + f(x,z) \\ \int 3xy^2z^2 \mathrm{d}z &= xy^2z^3 + f(x,y)\end{aligned}$$

Therefore, we have $f(x,y,z) = xy^2z^3 + C$. We can easily verify this as a potential function by computing its gradient and seeing that $\boldsymbol{\nabla} f = \boldsymbol{F}$. △

The reason for the name curl is that the curl vector is associated with the tendency of a vector field to *rotate*. If you imagine placing a freely rotating windmill blade in the middle of a vector field $\boldsymbol{F}$ at a point $\boldsymbol{a}$, then $\boldsymbol{\nabla} \times \boldsymbol{F}(\boldsymbol{a})$ represents the tendency for the windmill to rotate. The direction of rotation is given by the *right hand rule*. If you stick your right thumb in the direction of $\boldsymbol{\nabla} \times \boldsymbol{F}(\boldsymbol{a})$, then your fingers will curl in the direction that the field is causing rotation. The magnitude of the curl represents how fast this rotation will occur.

If $\boldsymbol{\nabla} \times \boldsymbol{F} = \mathbf{0}$, then $\boldsymbol{F}$ is called irrotational–that is, if you were to place a fan in the middle of $\boldsymbol{F}$ as a force field, it would not rotate at all. We will return to this concept later when we explore Stokes' Theorem.

5.9.2 Divergence

If $\boldsymbol{F} = \langle P, Q, R \rangle$ is a vector field with a domain $D \subseteq \mathbb{R}^3$ such that P_x, Q_y, and R_z exist on D, then the divergence of $\boldsymbol{F}$ is the scalar field defined by

$$\text{div } \boldsymbol{F} = \boldsymbol{\nabla} \cdot \boldsymbol{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

Similar to curl, we can remember this as

$$\boldsymbol{\nabla} \cdot \boldsymbol{F} = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle \cdot \langle P, Q, R \rangle$$

Example 5.9.7. If $\boldsymbol{F} = \left\langle xz, xyz, -y^2 \right\rangle$, find $\boldsymbol{\nabla} \cdot \boldsymbol{F}$.

Solution: Plug in to the formula.

$$\boldsymbol{\nabla} \cdot \boldsymbol{F} = \frac{\partial}{\partial x}(xz) + \frac{\partial}{\partial y}(xyz) + \frac{\partial}{\partial z}\left(-y^2\right) = z + xz$$

△

If $\boldsymbol{F}$ is a vector field on $\mathbb{R}^3$, then so is curl $\boldsymbol{F}$. Therefore, we can compute div (curl $\boldsymbol{F}$).

Theorem 5.9.8. If $\boldsymbol{F}$ is a vector field with a domain $D \subseteq \mathbb{R}^3$ whose component functions have continuous second-order partial derivatives, then

$$\text{div curl } \boldsymbol{F} = 0$$

Proof. Compute.

$$\begin{aligned}
\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{F}) &= \boldsymbol{\nabla}\cdot\begin{vmatrix}\hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R\end{vmatrix} \\
&= \boldsymbol{\nabla}\cdot\left\langle \frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}, \frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}, \frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right\rangle \\
&= \frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)+\frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)+\frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right) \\
&= \frac{\partial^2 R}{\partial x\partial y}-\frac{\partial^2 Q}{\partial x\partial z}+\frac{\partial^2 P}{\partial y\partial z}-\frac{\partial^2 R}{\partial y\partial x}+\frac{\partial^2 Q}{\partial x\partial z}-\frac{\partial^2 P}{\partial y\partial z}=0
\end{aligned}$$

□

Example 5.9.9. Show that $\boldsymbol{F}=\langle xz, xyz, -y^2\rangle$ cannot be written as the curl of another vector field.

Solution: If there is some $\boldsymbol{F}_2$ such that $\boldsymbol{F}=\boldsymbol{\nabla}\times\boldsymbol{F}_2$, then we must have $\boldsymbol{\nabla}\cdot\boldsymbol{F}=0$ by Theorem 5.9.8.

We previously computed $\boldsymbol{\nabla}\cdot\boldsymbol{F}=z+xz$, which is not zero. Therefore, $\boldsymbol{F}$ is not the curl of another vector field. △

Similar to curl, the name for divergence can be understood in the context of a velocity field. If $\boldsymbol{F}(x,y,z)$ denotes the velocity of a fluid at (x,y,z), then div $F(x,y,z)$ represents the net rate of change of mass of fluid flowing from the point (x,y,z) per unit volume. That is, a higher divergence means that fluid will "flee" the point (x,y,z) faster.

If $\boldsymbol{\nabla}\cdot\boldsymbol{F}=0$, then $\boldsymbol{F}$ is said to be **incompressible** or **solenoidal**. These types of fields have many applications in physics, particularly in magnetism, because all magnetic fields are incompressible.

Another useful differential operator occurs when we compute the divergence of a gradient vector field $\boldsymbol{\nabla}f$. If f is a function of three variables, we have

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

This expression is quite common, so we abbreviate it as $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}f=\boldsymbol{\nabla}^2 f$. The operator itself becomes

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}=\left\langle\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right\rangle\cdot\left\langle\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right\rangle=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$$

$\boldsymbol{\nabla}^2$ is called the **Laplace operator**, due to its relation to **Laplace's equation**: $\boldsymbol{\nabla}^2 f=0$.

We can also apply the Laplace operator to a vector field $\boldsymbol{F}=\langle P,Q,R\rangle$ and obtain

$$\boldsymbol{\nabla}^2\boldsymbol{F}=\left\langle\boldsymbol{\nabla}^2 P, \boldsymbol{\nabla}^2 Q, \boldsymbol{\nabla}^2 R\right\rangle$$

5.9.3 Vector Forms of Green's Theorem

Armed with our new knowledge of divergence and curl, we can revisit Green's Theorem and make a few new notes.

First, consider a vector field $\boldsymbol{F}=\langle P,Q\rangle$. We previously stated that

$$\oint_C \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}=\iint_D\left[\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right]\mathrm{d}A$$

However, now we can notice that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ can be rewritten.

$$\boldsymbol{\nabla} \times \boldsymbol{F} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & 0 \end{vmatrix} = \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right]\hat{\boldsymbol{k}}$$

Therefore, $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \hat{\boldsymbol{k}}$, and Green's Theorem can be rewritten as

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_D (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \hat{\boldsymbol{k}} \mathrm{d}A$$

Moreover, if the curve C is given by the vector equation

$$\boldsymbol{r}(t) = \langle x(t), y(t) \rangle, \quad t \in [a, b]$$

Then we can write the unit tangent vector $\boldsymbol{T}(t)$ as

$$\boldsymbol{T}(t) = \frac{\boldsymbol{r}'(t)}{\|\boldsymbol{r}'(t)\|} = \left\langle \frac{x'(t)}{\|\boldsymbol{r}'(t)\|}, \frac{y'(t)}{\|\boldsymbol{r}'(t)\|} \right\rangle$$

We can then recall that the outward unit normal vector is given by

$$\boldsymbol{n}(t) = \left\langle \frac{y'(t)}{\|\boldsymbol{r}'(t)\|}, \frac{-x'(t)}{\|\boldsymbol{r}'(t)\|} \right\rangle$$

Therefore, we can write

$$\begin{aligned} \oint_C \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}s &= \int_a^b (\boldsymbol{F} \cdot \boldsymbol{n})(t) \|\boldsymbol{r}'(t)\| \mathrm{d}t \\ &= \int_a^b \left[\frac{Py'(t)}{\|\boldsymbol{r}'(t)\|} - \frac{Qx'(t)}{\|\boldsymbol{r}'(t)\|}\right] \|\boldsymbol{r}'(t)\| \mathrm{d}t \\ &= \int_a^b [Py'(t) - Qx'(t)] \mathrm{d}t \\ &= \oint_C (-Q)\mathrm{d}x + P\mathrm{d}y = \iint\limits_D \left[\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right] \mathrm{d}A \end{aligned}$$

However, the integrand of this final double integral is simply just the divergence of $\boldsymbol{F}$. So we have a second version of Green's Theorem:

$$\oint_C \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}s = \iint\limits_D \boldsymbol{\nabla} \cdot \boldsymbol{F} \mathrm{d}A$$

This version states that the line integral of the normal component of $\boldsymbol{F}$ (as opposed to the tangential component with the curl form of Green's Theorem) across C is equal to the double integral of the divergence of $\boldsymbol{F}$ over the region enclosed by C.

Example 5.9.10. Use Green's Theorem to prove Green's first identity:

$$\iint\limits_D f \boldsymbol{\nabla}^2 g \mathrm{d}A = \oint_C f[\boldsymbol{\nabla} g \cdot \boldsymbol{n}] \mathrm{d}\boldsymbol{s} - \iint\limits_D \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g \mathrm{d}A$$

Solution: First, we should digest the problem a bit. If we want any chance of showing this, we should hope to get something of the form

$$\oint_C \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}s = \iint\limits_D \boldsymbol{\nabla} \cdot \boldsymbol{F} \mathrm{d}A$$

We can try and write the terms under the line integral in the form $\boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}s$. We can set $\boldsymbol{F} = f\boldsymbol{\nabla} g$ to achieve this. Now, we can notice that

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{F} = \boldsymbol{\nabla} \cdot f\nabla g &= \frac{\partial}{\partial x}\left(f\frac{\partial g}{\partial x}\right) + \frac{\partial}{\partial y}\left(f\frac{\partial g}{\partial y}\right) + \frac{\partial}{\partial z}\left(f\frac{\partial g}{\partial z}\right) \\
&= \frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + f\frac{\partial^2 g}{\partial x^2} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} + f\frac{\partial^2 g}{\partial y^2} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial z} + f\frac{\partial^2 g}{\partial z^2} \\
&= \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g + f\boldsymbol{\nabla}^2 g
\end{aligned}$$

So we can replace the $f\boldsymbol{\nabla}^2 g$ in the left surface integral with $\boldsymbol{\nabla} \cdot \boldsymbol{F} - \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g$. Therefore, our original expression becomes

$$\iint_D [\boldsymbol{\nabla} \cdot \boldsymbol{F} - \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g]\mathrm{d}A = \oint_C \boldsymbol{F} \cdot \boldsymbol{n}\mathrm{d}s - \iint_D \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g \mathrm{d}A$$

Adding $\iint_D \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g \mathrm{d}A$ to both sides, they cancel out and we get back to the original expression of Green's Theorem

$$\iint_D \boldsymbol{\nabla} \cdot \boldsymbol{F}\mathrm{d}A = \oint_C \boldsymbol{F} \cdot \boldsymbol{n}\mathrm{d}s$$

Thus proving the identity. △

Example 5.9.11. Use the result from the previous example to prove Green's second identity:

$$\iint_D (f\boldsymbol{\nabla}^2 g - g\boldsymbol{\nabla}^2 f)\mathrm{d}A = \oint_C (f\boldsymbol{\nabla} g - g\boldsymbol{\nabla} f) \cdot \boldsymbol{n}\mathrm{d}s$$

Solution: We'll first split the line integral, giving us

$$\oint_C (f\boldsymbol{\nabla} g - g\boldsymbol{\nabla} f) \cdot \boldsymbol{n}\mathrm{d}s = \oint_C f\boldsymbol{\nabla} g \cdot \boldsymbol{n}\mathrm{d}s - \oint_C g\boldsymbol{\nabla} f \cdot \boldsymbol{n}\mathrm{d}s$$

Then, we can compute $\boldsymbol{\nabla} \cdot (g\boldsymbol{\nabla} f)$ and find

$$\begin{aligned}
\boldsymbol{\nabla} \cdot (g\boldsymbol{\nabla} f) &= \frac{\partial}{\partial x}\left(g\frac{\partial f}{\partial x}\right) + \frac{\partial}{\partial y}\left(g\frac{\partial f}{\partial y}\right) + \frac{\partial}{\partial z}\left(g\frac{\partial f}{\partial z}\right) \\
&= \frac{\partial g}{\partial x}\frac{\partial f}{\partial x} + g\frac{\partial^2 f}{\partial x^2} + \frac{\partial g}{\partial y}\frac{\partial f}{\partial y} + g\frac{\partial^2 f}{\partial y^2} + \frac{\partial g}{\partial z}\frac{\partial f}{\partial z} + g\frac{\partial^2 f}{\partial z^2} \\
&= \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g + g\boldsymbol{\nabla}^2 f
\end{aligned}$$

This allows us to write

$$\oint_C g\boldsymbol{\nabla} f \cdot \boldsymbol{n}\mathrm{d}s = \iint_D (\boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g + g\boldsymbol{\nabla}^2 f)\mathrm{d}A$$

which can be substituted back into our original statement to find

$$\iint_D f\boldsymbol{\nabla}^2 g\mathrm{d}A - \iint_D g\boldsymbol{\nabla}^2 f\mathrm{d}A = \oint_C f\boldsymbol{\nabla} g \cdot \boldsymbol{n}\mathrm{d}s - \iint_D \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g\mathrm{d}A - \iint_D g\boldsymbol{\nabla}^2 f\mathrm{d}A$$

The $-\iint_D g\boldsymbol{\nabla}^2 f\mathrm{d}A$ terms on each side cancel, leaving us with

$$\iint_D f\boldsymbol{\nabla}^2 g\mathrm{d}A = \oint_C f\boldsymbol{\nabla} g \cdot n\mathrm{d}s - \iint_D \boldsymbol{\nabla} f \cdot \boldsymbol{\nabla} g\mathrm{d}A$$

which is precisely the statement of Green's first identity, thus proving the identity. △

5.10 Parametric Surfaces and Their Areas

In Calculus II we learned how to find the area of a surface of revolution, and earlier and Calculus III we learned how to find the area of a surface with the equation $z = f(x, y)$. Now, we will discuss surfaces described with parametric equations.

5.10.1 Parametric Surfaces

Similar to how we etched out space curves with a function $\boldsymbol{r}(t)$ of a single variable, we can etch out surfaces in space with a vector function $\boldsymbol{r}(u, v)$ of two variables. If

$$\boldsymbol{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle$$

is a vector-valued function defined on a region D of the uv plane, then we call the set of all points (x, y, z) such that $x = x(u, v)$, $y = y(u, v)$, and $z = z(u, v)$ the **parametric surface** described by $\boldsymbol{r}$.

Example 5.10.1. Identify and sketch the surface with vector equation

$$\boldsymbol{r}(u, v) = \langle 2\cos u, v, 2\sin u\rangle$$

Solution: We'll notice that the y component is free to vary as much as it pleases, but the x and z components are restricted to the circle of radius 2 inscribed in the xz plane. This describes a cylinder of radius 2 pointing in the y direction. △

We could have modified the shape formed in the previous example by placing restrictions on the values of u and v. For instance, if u was only allowed to vary between 0 and π, we would just get the upper half of the cylinder, and if v was only allowed to vary between -1 and 5, the cylinder would be 6 units long with endpoints at $y = -1$ and $y = 5$.

If a parametric surface S is given by $\boldsymbol{r}(u, v)$, then there are two useful families of curves that can help us classify S, one with u constant and the other with v constant. This turns our function of two variables into a function of one variable, effectively describing a space curve similar to the ones in the line integrals section. These curves are known as **grid curves**.

We use these in a similar manner to how we utilized traces when plotting the graphs of $z = f(x, y)$ functions by holding either x or y constant.

Example 5.10.2. Find a vector function to represent the plane that passes through the point P_0 with position vector $\boldsymbol{r}_0$ and contains two nonparallel vector $\boldsymbol{a}$ and $\boldsymbol{b}$.

Solution: From linear algebra, we can remember that every point on the plane formed by $\boldsymbol{a}$ and $\boldsymbol{b}$ that passes through the origin can be represented as a linear combination of $\boldsymbol{a}$ and $\boldsymbol{b}$–that is, the plane can be written as

$$u\boldsymbol{a} + v\boldsymbol{b}$$

If we want to offset the plane so that it contains $\boldsymbol{r}_0$, we simply add it:

$$\boldsymbol{r}(u, v) = \boldsymbol{r}_0 + u\boldsymbol{a} + v\boldsymbol{b}$$

Where $u, v \in \mathbb{R}$. △

Example 5.10.3. Find a parametric representation of the sphere

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

Solution: Recall that a sphere in spherical coordinates is given by $\{(\rho, \phi, \theta) : \rho = a, \phi \in [0, \pi], \theta \in [0, 2\pi]\}$. We can translate this into a parametric equation quite easily using our existing knowledge of the spherical coordinate system:

$$\boldsymbol{r}(\phi, \theta) = \langle a\cos\theta\sin\phi, a\sin\theta\sin\phi, a\cos\phi\rangle\,, \qquad \phi \in [0, \pi], \quad \theta \in [0, 2\pi]$$

△

Example 5.10.4. Find a parametric representation of the cylinder

$$x^2 + y^2 = a^2, \quad z \in [0, 1]$$

This is also quite simple. In cylindrical, we have $\{(r, \theta, z) : r = a, \theta \in [0, 2\pi], z \in [0, 1]\}$. So our parametric surface is given by

$$\boldsymbol{r}(\theta, z) = \langle a\cos\theta, a\sin\theta, z\rangle, \qquad \theta \in [0, 2\pi], \quad z \in [0, 1]$$

△

Example 5.10.5. Find a vector function that represents the elliptic paraboloid $z = x^2 + 2y^2$.

Solution: We can simply let x be one paramater and y be another parameter, and we get

$$\boldsymbol{r}(x, y) = \left\langle x, y, x^2 + 2y^2 \right\rangle$$

△

Example 5.10.6. Find a parametric representation for the surface $z = 2\sqrt{x^2 + y^2}$–that is, the top half of the cone $z^2 = 4x^2 + 4y^2$.

Solution: We could keep this in rectangular or move it to cylindrical. In rectangular, you would get

$$\boldsymbol{r}(x, y) = \left\langle x, y, 2\sqrt{x^2 + y^2} \right\rangle$$

and in cylindrical, you would get

$$\boldsymbol{r}(r, \theta) = \langle r\cos\theta, r\sin\theta, 2r\rangle, \qquad r \geq 0, \quad \theta \in [0, 2\pi]$$

△

5.10.2 Surfaces of Revolution

Surfaces of revolution can be represented parametrically and thus graphed with a computer. Consider revolving the surface $y = f(x)$ around the x axis with $x \in [a, b]$. We could represent this surface parametrically as

$$\boldsymbol{r}(x, \theta) = \langle x, f(x)\cos\theta, f(x)\sin\theta\rangle, \qquad x \in [a, b], \quad \theta \in [0, 2\pi]$$

Example 5.10.7. Obtain parametric equations for rotating the graph of $y = \sin x$ about the x axis with x ranging from a to b.

Solution: Use the previously mentioned formula.

$$\boldsymbol{r}(x, \theta) = \langle x, \sin x\cos\theta, \sin x\sin\theta\rangle, \qquad x \in [a, b], \quad \theta \in [0, 2\pi]$$

△

5.10.3 Tangent Planes

We can also represent tangent planes as parametric surfaces. Consider a surface S traced out by the parametric function

$$\boldsymbol{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle$$

The tangent plane to S at a point P_0 with a position vector $\boldsymbol{r}(u_0, v_0)$ can be found. If we keep u constant at $u = u_0$ and allow v to vary, we get a grid curve that lies on S. The tangent vector to this grid curve is obtained by taking the partial derivative of $\boldsymbol{r}$ with respect to v.

$$\boldsymbol{T}_v = \left\langle \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v} \right\rangle$$

Similarly, we can obtain the tangent vector to the grid curve obtained by setting $v = v_0$ by taking the partials with respect to u:

$$\boldsymbol{T}_u = \left\langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u} \right\rangle$$

If $\boldsymbol{T}_v \times \boldsymbol{T}_u \neq \boldsymbol{0}$ (that is, if the tangent vectors are not parallel), then the surface is called **smooth**. For a smooth surface, the tangent plane is the plane that contains P_0, as well as the vectors $\boldsymbol{T}_v$ and $\boldsymbol{T}_u$ and has a normal vector given by $\boldsymbol{T}_v \times \boldsymbol{T}_u$.

Example 5.10.8. Find the tangent plane to the surface with parametric equations

$$\boldsymbol{r}(t) = \left\langle u^2, v^2, u + 2v \right\rangle$$

at the point $(1, 1, 3)$.

Solution: First, we will notice that when $\langle x, y, z\rangle = \langle 1, 1, 3\rangle$, we must have $u = 1$ and $v = 1$. Then, we can compute the tangent vectors to the surface.

$$\begin{aligned}\boldsymbol{T}_v &= \left\langle \frac{\partial}{\partial v}\left(u^2\right)(1,1), \frac{\partial}{\partial v}\left(v^2\right)(1,1), \frac{\partial}{\partial v}(u+2v)(1,1) \right\rangle = \langle 0, 2, 2\rangle \\ \boldsymbol{T}_u &= \left\langle \frac{\partial}{\partial u}\left(u^2\right)(1,1), \frac{\partial}{\partial u}\left(v^2\right)(1,1), \frac{\partial}{\partial u}(u+2v)(1,1) \right\rangle = \langle 2, 0, 1\rangle\end{aligned}$$

So a normal vector to the tangent plane is

$$\boldsymbol{T}_v \times \boldsymbol{T}_u = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 0 & 2 & 2 \\ 2 & 0 & 1 \end{vmatrix} = \langle 2, 4, -4\rangle$$

and then the equation of the tangent plane is given by $\boldsymbol{n} \cdot (\boldsymbol{r} - \boldsymbol{r}_0) = 0$, or

$$2(x-1) + 4(y-1) - 4(z-3) = 0$$

△

5.10.4 Surface Area

Now we can define the surface area of a general parametric surface S given by

$$\boldsymbol{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle$$

For simplicity, let's consider a surface where the domain is a rectangle in the uv plane, which we can divide into many subrectangles R_{ij}. For each subrectangle, we can choose a sample point (u_i^*, v_j^*) to be at the *bottom left corner* of R_{ij}. The part S_{ij} of the surface formed by R_{ij} is known as a **patch** of S and has a point on it given by $\boldsymbol{r}(u_i^*, v_j^*)$.

We can also find that the tangent vectors to the sample point of P_{ij} are

$$\boldsymbol{r}_{u,i}^* = \frac{\partial \boldsymbol{r}}{\partial u}\left(u_i^*, v_j^*\right) \quad \text{and} \quad \boldsymbol{r}_{v,j}^* = \frac{\partial \boldsymbol{r}}{\partial v}\left(u_i^*, v_j^*\right)$$

We can construct a parallelogram with each of these tangent vectors to find the area of the subrectangle. We can approximate the area with

$$A_{ij} \approx \left\|(\Delta u \boldsymbol{r}_{u,i}^*) \times (\Delta v \boldsymbol{r}_{v,j}^*)\right\| = \left\|\boldsymbol{r}_{u,i}^* \times \boldsymbol{r}_{v,j}^*\right\| \Delta u \Delta v$$

As Δu and Δv get smaller, this becomes a more accurate approximation. So we can take a limit and sum up all of the areas of each subrectangle and we will get an integral expression:

$$A = \lim_{(m,n)\to\infty} \sum_{i=1}^{n} \sum_{j=1}^{m} \left\|\boldsymbol{r}_{u,i}^* \times \boldsymbol{r}_{v,j}^*\right\| \Delta u \Delta v = \iint_D \|\boldsymbol{r}_u \times \boldsymbol{r}_v\| \mathrm{d}A$$

This is how we find the surface area of a parametric surface.

Example 5.10.9. Find the surface area of a sphere of radius a using parametric surfaces.

Solution: We can represent the sphere using

$$\boldsymbol{r}(\phi,\theta) = \langle a\cos\theta\sin\phi, a\sin\theta\sin\phi, a\cos\phi\rangle, \quad \phi\in[0,\pi], \quad \theta\in[0,2\pi]$$

Then, we can take the partials to get

$$\boldsymbol{r}_\phi = \langle a\cos\theta\cos\phi, a\sin\theta\cos\phi, -a\sin\phi\rangle$$

and

$$\boldsymbol{r}_\theta = \langle -a\sin\theta\sin\phi, a\cos\theta\sin\phi, 0\rangle$$

Then we can compute $\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta$:

$$\begin{aligned}
\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ a\cos\theta\cos\phi & a\sin\theta\cos\phi & -a\sin\phi \\ -a\sin\theta\sin\phi & a\cos\theta\sin\phi & 0 \end{vmatrix} \\
&= \left\langle a^2\cos\theta\sin^2\phi, a^2\sin\theta\sin^2\phi, a^2\cos^2\theta\sin\phi\cos\phi + a^2\sin^2\theta\sin\phi\cos\phi\right\rangle \\
&= \left\langle a^2\cos\theta\sin^2\phi, a^2\sin\theta\sin^2\phi, a^2\sin\phi\cos\phi(\sin^2\theta+\cos^2\theta)\right\rangle \\
&= \left\langle a^2\cos\theta\sin^2\phi, a^2\sin\theta\sin^2\phi, a^2\sin\phi\cos\phi\right\rangle
\end{aligned}$$

Now, computing the magnitude of this:

$$\begin{aligned}
\|\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta\| &= \sqrt{a^4\cos^2\theta\sin^4\phi + a^4\sin^2\theta\sin^4\phi + a^4\sin^2\phi\cos^2\phi} \\
&= a^2\sin\phi\sqrt{\sin^2\phi(\cos^2\theta+\sin^2\theta)+\cos^2\phi} \\
&= a^2\sin\phi\sqrt{\sin^2\phi+\cos^2\phi} = a^2\sin\phi
\end{aligned}$$

Now, we can simply compute the surface area. Take special note that we are *not* performing any kind of coordinate transformation (in fact, we are integrating over what is essentially a rectangle in the $\phi\theta$ plane), so we do not need to place in the Jacobian.

$$A = \iint\limits_D \|\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta\| \mathrm{d}A = \int_0^{2\pi}\int_0^{\pi} a^2\sin\phi \,\mathrm{d}\phi\mathrm{d}\theta = 4\pi a^2$$

△

5.10.5 Surface Area of the Graph of a Function

For the special case of a surface S given by an equation $z = f(x,y)$ where (x,y) lies in D and f has continuous partials, we can define S as a parametric surface given by

$$\boldsymbol{r}(x,y) = \langle x, y, f(x,y)\rangle \quad (x,y)\in D$$

and find its surface area in that way.

Going about the normal process, we can take the partials to obtain

$$\boldsymbol{r}_x = \langle 1, 0, f_x(x,y)\rangle \quad \text{and} \quad \boldsymbol{r}_y = \langle 0, 1, f_y(x,y)\rangle$$

Then,

$$\begin{aligned}
\boldsymbol{r}_x \times \boldsymbol{r}_y &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 1 & 0 & f_x(x,y) \\ 0 & 1 & f_y(x,y) \end{vmatrix} \\
&= \langle -f_x(x,y), -f_y(x,y), 1\rangle
\end{aligned}$$

And

$$\|\boldsymbol{r}_x \times \boldsymbol{r}_y\| = \sqrt{[f_x(x,y)]^2 + [f_y(x,y)]^2 + 1}$$

Giving us the definition of surface area as

$$A = \iint\limits_D \sqrt{1 + \left[\frac{\partial z}{\partial x}\right]^2 + \left[\frac{\partial z}{\partial y}\right]^2}\,\mathrm{d}A$$

Which is identical to the one we found earlier, in the multiple integrals section.

We can also compare this equation to the surface area of a solid of revolution that we learned in single-variable calculus. Recall that the surface area of revolving a function $y = f(x)$ about the x axis from $x = a$ to $x = b$ was given by

$$2\pi \int_a^b f(x)\sqrt{1 + [f'(x)]^2}\mathrm{d}x$$

Now, we can show this using parametric surfaces. We can define our surface S with

$$\boldsymbol{r}(x,\theta) = \langle x, f(x)\cos\theta, f(x)\sin\theta\rangle \quad x \in [a,b] \quad \theta \in [0, 2\pi]$$

Then, our tangent vectors are

$$\boldsymbol{r}_x = \langle 1, f'(x)\cos\theta, f'(x)\sin\theta\rangle \quad \text{and} \quad \boldsymbol{r}_\theta = \langle 0, -f(x)\sin\theta, f(x)\cos\theta\rangle$$

We can compute the cross product $\boldsymbol{r}_x \times \boldsymbol{r}_\theta$:

$$\begin{aligned}
\boldsymbol{r}_x \times \boldsymbol{r}_\theta &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 1 & f'(x)\cos\theta & f'(x)\sin\theta \\ 0 & -f(x)\sin\theta & f(x)\cos\theta \end{vmatrix} \\
&= \left\langle f(x)f'(x)\cos^2\theta + f'(x)f(x)\sin^2\theta, -f(x)\cos\theta, -f(x)\sin\theta\right\rangle \\
&= \langle f(x)f'(x), -f(x)\cos\theta, -f(x)\sin\theta\rangle \\
\|\boldsymbol{r}_x \times \boldsymbol{r}_\theta\| &= \sqrt{[f(x)]^2[f'(x)]^2 + [f(x)]^2\cos^2\theta + [f(x)]^2\sin^2\theta} \\
&= f(x)\sqrt{[f'(x)]^2 + \cos^2\theta + \sin^2\theta} \\
&= f(x)\sqrt{[f'(x)]^2 + 1}
\end{aligned}$$

Then integrating, we obtain

$$\begin{aligned}
A = \iint\limits_D \|\boldsymbol{r}_x \times \boldsymbol{r}_\theta\|\mathrm{d}A &= \int_0^{2\pi}\int_a^b f(x)\sqrt{[f'(x)]^2 + 1}\,\mathrm{d}x\mathrm{d}\theta \\
&= 2\pi\int_a^b f(x)\sqrt{[f'(x)]^2 + 1}\,\mathrm{d}x
\end{aligned}$$

Which is precisely equal to the Calculus I formula.

5.11 Surface Integrals

The relationship between surface integrals and surface area is similar to the relationship between arclength and line integrals. Suppose f is a function of three variables whose domain includes a surface S. We divide S into patches S_{ij} of area ΔS_{ij}. We evaluate f at a sample point P_{ij}^* in each patch, multiply by the area, and form the sum

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f(P_{ij}^*)\Delta S_{ij}$$

If we take the limit of this, we get the definition of a surface integral:

$$\iint\limits_S f(x, y, z)\mathrm{d}S = \lim_{(m,n)\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{m} f(P_{ij}^*)\Delta S_{ij}$$

To evaluate this, we approximate the patch area ΔS_{ij} by the area ΔT_{ij} of the approximating parallelogram in the tangent plane. Then, this limit becomes a double integral. After this, the following steps will vary depending on if we are observing a graph or a parametric surface.

First, if we are observing a graph, then we have an equation of the form $z = g(x, y)$ with $(x, y) \in D$. The tangent plane to this surface is the plane that passes through P_{ij}^* and has normal vector $\boldsymbol{g}_x(P_{ij}^*) \times \boldsymbol{g}_y(P_{ij}^*)$. We will follow the same steps as the previous section to obtain

$$\Delta S_{ij} \approx \sqrt{\left[g_x(P_{ij}^*)\right]^2 + \left[g_y(P_{ij}^*)\right]^2 + 1}$$

Thus giving us

$$\iint\limits_S f(x, y, z)\mathrm{d}S = \iint\limits_D f(x, y, g(x, y))\sqrt{[g_x(x, y)]^2 + [g_y(x, y)]^2 + 1}\,\mathrm{d}A$$

If, instead, we are observing a parametric surface, we will have some equation of the form $\boldsymbol{r}(u, v) = \langle x(u, v), y(u, v), z(u, v)\rangle$ where $(u, v) \in D$. We will make the same approximation of surface area as we did in the previous section:

$$\Delta S_{ij} \approx \|\boldsymbol{r}_u \times \boldsymbol{r}_v\|$$

giving us an integral of

$$\iint\limits_S f(x, y, z)\mathrm{d}S = \iint\limits_D f(\boldsymbol{r}(u, v))\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|\,\mathrm{d}A$$

An easy way to remember this formula is to note the similarities between it and the formula for a scalar line integral: $\int_C f(x, y, z)\mathrm{d}s = \int_a^b f(\boldsymbol{r}(t))\|\boldsymbol{r}'(t)\|\mathrm{d}t$

Additionally, if a surface integral is done over a *closed* surface S, then we can use the notation

$$\oiint\limits_S f(x, y, z)\mathrm{d}S$$

To indicate that S is closed. We can think of closed surfaces as ones that enclose a region $E \subset \mathbb{R}^3$ in them and have no "holes" (the mathematical definition is more complicated, so we will not cover it).

Example 5.11.1. Evaluate $\iint_S y\mathrm{d}S$ where S is the surface $z = x + y^2$, $x \in [0, 1]$, $y \in [0, 2]$.

Solution: Plug in.

$$\begin{aligned}\iint\limits_S y\mathrm{d}S &= \iint\limits_D y\sqrt{2 + 4y^2}\,\mathrm{d}A \\ &= \sqrt{2}\int_0^1\int_0^2 y\sqrt{2y^2 + 1}\,\mathrm{d}y\mathrm{d}x \\ &= \frac{\sqrt{2}}{4}\cdot\frac{2}{3}\cdot\left(2y^2 + 1\right)^{3/2}\Big|_0^2 = \frac{13\sqrt{2}}{3}\end{aligned}$$

△

Example 5.11.2. Compute the surface integral $\oiint_S x^2\mathrm{d}S$ where S is the unit sphere $x^2 + y^2 + z^2 = 1$.

Solution: This surface is easiest to represent as the parametric surface described by

$$\boldsymbol{r}(\phi, \theta) = \langle\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi\rangle, \quad \phi \in [0, \pi], \quad \theta \in [0, 2\pi]$$

Then, we can compute the partials:

$$\begin{aligned} \boldsymbol{r}_\phi &= \langle \cos\theta\cos\phi, \sin\theta\cos\phi, -\sin\phi\rangle \\ \boldsymbol{r}_\theta &= \langle -\sin\theta\sin\phi, \cos\theta\sin\phi, 0\rangle \end{aligned}$$

And their cross product:

$$\begin{aligned} \boldsymbol{r}_\phi \times \boldsymbol{r}_\theta &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \cos\theta\cos\phi & \sin\theta\cos\phi & -\sin\phi \\ -\sin\theta\sin\phi & \cos\theta\sin\phi & 0 \end{vmatrix} \\ &= \left\langle \cos\theta\sin^2\phi, \sin\theta\sin^2\phi, \cos^2\theta\sin\phi\cos\phi + \sin^2\theta\cos\phi\sin\phi \right\rangle \\ &= \left\langle \cos\theta\sin^2\phi, \sin\theta\sin^2\phi, \cos\phi\sin\phi \right\rangle \\ \|\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta\| &= \sqrt{\cos^2\theta\sin^4\phi + \sin^2\theta\sin^4\phi + \cos^2\phi\sin^2\phi} \\ &= \sin\phi\sqrt{\sin^2\phi + \cos^2\phi} = \sin\phi \end{aligned}$$

And then finally the surface integral:

$$\begin{aligned} \oiint\limits_S x^2 \mathrm{d}S &= \int_0^{2\pi}\int_0^{\pi} [\cos\theta\sin\phi]^2 \sin\phi \mathrm{d}\phi\mathrm{d}\theta \\ &= \left[\int_0^{2\pi} \frac{1}{2}(1+\cos(2\theta)\mathrm{d}\theta\right]\left[\int_0^{\pi} \sin\phi[\sin^2\phi]\mathrm{d}\phi\right] \\ &= \pi\int_0^{\pi} \left[\sin\phi - \sin\phi\cos^2\phi\right]\mathrm{d}\phi \\ &= \pi(\cos 0 - \cos\pi)\frac{1}{3}(\cos^3\pi - \cos^3 0) = \frac{4\pi}{3} \end{aligned}$$

△

In the case where S is piecewise-smooth, that is, a finite union of smooth surfaces $S_1, \ldots, S_n$ that intersect only along their boundaries, then the surface integral over S is given by the sum of each of the surface integrals of component surfaces:

$$\iint\limits_S f(x,y,z)\mathrm{d}S = \sum_{i=1}^{n} \iint\limits_{S_i} f(x,y,z)\mathrm{d}S$$

Example 5.11.3. Evaluate $\oiint_S z\mathrm{d}S$ where $S = S_1 \cup S_2 \cup S_3$ has components given by:

1. S_1 is the disk $x^2 + y^2 \leq 1$ in the xy plane.
2. S_2 is the part of the plane $z = 1 + x$ that lies above S_1.
3. S_3 is the cylinder $x^2 + y^2 = 1$ which is bounded below by S_1 and above by S_2.

Solution: We can write

$$\oiint\limits_S z\mathrm{d}S = \iint\limits_{S_1} z\mathrm{d}S + \iint\limits_{S_2} z\mathrm{d}S + \iint\limits_{S_3} z\mathrm{d}S$$

and then solve each one individually. First, focusing on S_1, we can immediately see that it lies in the plane $z = 0$. Because the surface integral is solely depending on z, then it stands to reason that it will just go to zero.

Then, focusing on S_2, we can its surface integral with the standard formula

$$\begin{aligned}
\iint\limits_{S_2} \mathrm{d}S &= \iint\limits_{R} z(x,y)\big\|z_x^2 + z_y^2 + 1\big\|\mathrm{d}A \\
&= \int_0^{2\pi}\int_0^1 \sqrt{2}(1 + r\cos\theta)r\mathrm{d}r\mathrm{d}\theta \\
&= \sqrt{2}\int_0^{2\pi}\int_0^1 [r + r^2\cos\theta]\mathrm{d}r\mathrm{d}\theta \\
&= \sqrt{2}\int_0^{2\pi}\left[\frac{1}{2} + 0\right]\mathrm{d}\theta = \pi\sqrt{2}
\end{aligned}$$

Now, for S_3, we can write it as a parametric surface with $\boldsymbol{r}_1(z,\theta) = \langle\cos\theta, \sin\theta, z\rangle$ with $\theta \in [0, 2\pi]$ and $z \in [0, 1+x] = [0, 1+\cos\theta]$. Therefore, we obtain

$$\boldsymbol{r}_{1,z} = \langle 0, 0, 1\rangle \quad \text{and} \quad \boldsymbol{r}_{1,\theta} = \langle -\sin\theta, \cos\theta, 0\rangle$$

and then

$$\begin{aligned}
\boldsymbol{r}_{1,z} \times \boldsymbol{r}_{1,\theta} &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 0 & 0 & 1 \\ -\sin\theta & \cos\theta & 0 \end{vmatrix} \\
&= \langle -\cos\theta, -\sin\theta, 0\rangle \\
\|\boldsymbol{r}_{1,z} \times \boldsymbol{r}_{1,\theta}\| &= \sqrt{\sin^2\theta + \cos^2\theta} = 1
\end{aligned}$$

Then, we have

$$\begin{aligned}
\iint\limits_{S} z\mathrm{d}S &= \iint\limits_{D} z\|\boldsymbol{r}_{1,\theta} \times \boldsymbol{r}_{1,z}\|\mathrm{d}A \\
&= \int_0^{2\pi}\int_0^{1+\cos\theta} z\mathrm{d}z\mathrm{d}\theta \\
&= \frac{1}{2}\int_0^{2\pi}(1+\cos\theta)^2\mathrm{d}\theta \\
&= \frac{1}{2}\int_0^{2\pi}[\cos^2\theta + 2\cos\theta + 1]\mathrm{d}\theta \\
&= \frac{1}{2}\left[\frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + 2\sin\theta + \theta\right]_0^{2\pi} = \frac{3\pi}{2}
\end{aligned}$$

Now, since we have computed each component, we can simply add them up to obtain the total surface integral:

$$\oiint\limits_{S} z\mathrm{d}S = \iint\limits_{S_1} z\mathrm{d}S + \iint\limits_{S_2} z\mathrm{d}S + \iint\limits_{S_3} z\mathrm{d}S = 2\sqrt{\pi} + \frac{3\pi}{2}$$

△

5.11.1 Applications of Surface Integrals

Surface integrals also have applications, similar to the previous types of integrals we have gone over. For instance, if a thin sheet of a substance forms the shape of a surface S and the density of the sheet at (x, y, z) is given by $\rho(x, y, z)$, then the total mass is

$$m = \iint\limits_{S} \rho(x, y, z)\mathrm{d}S$$

and the center of mass is $(\overline{x}, \overline{y}, \overline{z})$ where

$$\overline{x} = \frac{1}{m} \iint\limits_S x\rho \mathrm{d}S \quad \overline{y} = \frac{1}{m} \iint\limits_S y\rho \mathrm{d}S \quad \overline{z} = \frac{1}{m} \iint\limits_S z\rho \mathrm{d}S$$

We can also define moments of inertia in much the same way, although we will omit here because it's nearly identical to the discussion of moments of inertia in the multiple integrals section.

5.11.2 Oriented Surfaces

In order to define surface integrals of vector fields, we need to get a sense of which way a surface is "facing." To do so, we need to rule out nonorientable surfaces such as Möbius strips and similar.

From now on, we will only consider orientable (two-sided) surfaces. We will start with a surface S that has a tangent plane at every point (x, y, z) on S except along its edges. Each tangent plane has two unit normal vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ with $\boldsymbol{n}_2 = -\boldsymbol{n}_1$. We will attempt to choose a unit normal vector $\boldsymbol{n}$ at every point such that it varies smoothly as you go from point to point (i.e. the normal vector stays on the same side of the surface). If this i possible, we call S an **oriented surface**, and the choice of $\boldsymbol{n}$ provides S with an **orientation**. Each orientable surface has two possible orientations.

For a surface $z = g(x, y)$ we can quite easily see that one of the normal vectors to the tangent plane is given by

$$\boldsymbol{n} = \frac{-g_x(x, y) - g_y(x, y) + 1}{\sqrt{[g_x(x, y)]^2 + [g_y(x, y)]^2 + 1}}$$

(this comes from defining a parametric surface by $\boldsymbol{r} = \langle x, y, g(x, y) \rangle$ and computing the cross product $\boldsymbol{r}_x \times \boldsymbol{r}_y$).

Because the z component of this is positive, we will call this the **upward** orientation.

For a more general parametric surface defined by $\boldsymbol{r}(u, v)$, we say that the upward orientation of the normal vector is

$$\boldsymbol{n} = \frac{\boldsymbol{r}_u \times \boldsymbol{r}_v}{\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|}$$

For example, in the parametric surfaces section, we found the representation for the ball $\rho^2 = a^2$ as a parametric surface to be

$$\boldsymbol{r}(\phi, \theta) = \langle a \sin\phi \cos\theta, a \sin\phi \sin\theta, a \cos\phi \rangle$$

Then, we later in the section found that

$$\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta = \left\langle a^2 \sin^2\phi \cos\theta, a^2 \sin^2\phi \sin\theta, a^2 \sin\phi \cos\phi \right\rangle$$

and

$$\|\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta\| = a^2 \sin\phi$$

So the positive unit normal vector of this surface is given by

$$\boldsymbol{n} = \left\langle \sin\phi \cos\theta, \sin\phi \sin\theta, a^{-1} \cos\phi \right\rangle$$

Observe that because the position vector points in the same direction as the normal vector (since it's a sphere), we have the equivalence

$$\boldsymbol{n} = \frac{1}{a} \boldsymbol{r}(\phi, \theta)$$

For a **closed surface**, that is, a surface that is the boundary of a solid region E, the convention is that the positive orientation is the one where the normal vectors point away from E.

5.11.3 Surface Integrals of Vector Fields

Suppose that S is an oriented surface with unit normal vector $\boldsymbol{n}$, and imagine a fluid with density $\rho(x,y,z)$ and velocity field $\boldsymbol{v}(x,y,z)$ flowing through S (think of S as an imaginary surface that doesn't impede the flow of the fluid). Then, the rate of flow (mass per unit time) at a particular point with position vector $\boldsymbol{x}$ is $\rho(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x})$. The amount of flow that goes in the direction of $\boldsymbol{n}$ can be given by $\rho(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x})\cdot\boldsymbol{n}$.

Now, if we divide S into many small patches S_{ij}, where each S_{ij} is nearly planar, we can approximate the total mass of fluid crossing S per unit time. First, choose a sample point P_{ij} within each S_{ij}. Then, we can write

$$\Phi_S \approx \sum_i \sum_j \Phi_{S_{ij}} = \sum_i \sum_j (\rho(P_{ij})\boldsymbol{v}(P_{ij})\cdot\boldsymbol{n})A(S_{ij})$$

where $A(S_{ij})$ is the area of S_{ij}.

We can convert this into an integral,

$$\Phi_S = \iint_S \rho\boldsymbol{v}\cdot\boldsymbol{n}\,\mathrm{d}S = \iint_S \rho(x,y,z)\boldsymbol{v}(x,y,z)\cdot\boldsymbol{n}(x,y,z)\,\mathrm{d}S$$

and this can be interpreted physically as the rate of flow through S.

If we write $\boldsymbol{F} = \rho\boldsymbol{v}$, then $\boldsymbol{F}$ is also a vector field on $\mathbb{R}^3$ and the previous integral becomes

$$\iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S$$

A surface integral of this form occurs frequently in physics, and is called the surface integral of $\boldsymbol{F}$ over S.

Definition 5.11.4. If $\boldsymbol{F}$ is a continuous vector field defined on an oriented surface S with unit normal vector $\boldsymbol{n}$, then the surface integral of $\boldsymbol{F}$ over S is

$$\iint_S \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{S} = \iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S$$

This integral is also called the **flux** of $\boldsymbol{F}$ across S.

In the case of a surface S given by a graph $z = g(x,y)$, we previously found $\boldsymbol{n}$ by using a parametric surface. Another way we could do it is by noting that S is also the graph of the level surface given by $f(x,y,z) = z - g(x,y) = 0$. The gradient $\boldsymbol{\nabla} f(x,y,z)$ is normal to S at (x,y,z), so the unit normal is

$$\boldsymbol{n} = \frac{\boldsymbol{\nabla} f(x,y,z)}{\|\boldsymbol{\nabla} f(x,y,z)\|} = \frac{\langle -g_x(x,y), -g_y(x,y), 1\rangle}{\sqrt{[g_x(x,y)]^2 + [g_y(x,y)]^2 + 1}}$$

This is, the upward unit normal because the z component is positive.

If we now plug this into the formula for the surface integral with $\boldsymbol{F} = \langle P, Q, R\rangle$, we obtain

$$\begin{aligned}
\iint_S \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{S} &= \iint_S \boldsymbol{F}\cdot\boldsymbol{n}\,\mathrm{d}S \\
&= \iint_D \langle P, Q, R\rangle \cdot \frac{\langle -g_x(x,y), -g_y(x,y), 1\rangle}{\sqrt{[g_x(x,y)]^2 + [g_y(x,y)]^2 + 1}} \sqrt{[g_x(x,y)]^2 + [g_y(x,y)]^2 + 1}\,\mathrm{d}A \\
&= \iint_D \left[-P\frac{\partial g}{\partial x} - Q\frac{\partial g}{\partial y} + R\right]\mathrm{d}A
\end{aligned}$$

Similar formulas can be found if we instead have $y = h(x,z)$ or $z = k(x,y)$.

Example 5.11.5. Evaluate $\iint_S \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S}$ where $\boldsymbol{F} = \langle y, x, z \rangle$ and S is the boundary of the solid region E enclosed by the paraboloid $z = 1 - x^2 - y^2$ and the plane $z = 0$.

Solution: S is the union of two parts–the paraboloid part S_1 and the part on the plane $z = 0$. We will focus first on S_1.

S_1 is easiest to describe in cylindrical coordinates, where we find that it is modeled by the equation $z = 1-r^2$, with $\theta \in [0, 2\pi]$ and $r \in [0, 1]$.

Therefore, we can solve the surface integral:

$$\begin{aligned}
\iint\limits_{S_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} &= \iint\limits_{D} \left[-P\frac{\partial g}{\partial x} - Q\frac{\partial g}{\partial y} + R\right]\mathrm{d}A \\
&= \iint\limits_{D} [2yx + 2xy + z]\mathrm{d}A \\
&= \iint\limits_{D} \left[4xy + 1 - x^2 - y^2\right]\mathrm{d}A \\
&= \int_0^{2\pi} \int_0^1 \left[4r^2 \cos\theta \sin\theta - r^2 + 1\right] r\mathrm{d}r\mathrm{d}\theta \\
&= \int_0^{2\pi} \left[\cos\theta \sin\theta + \frac{1}{4}\right]\mathrm{d}\theta \\
&= \frac{1}{2}\sin^2\theta + \frac{\theta}{4}\bigg|_0^{2\pi} = \frac{\pi}{2}
\end{aligned}$$

S_2 has the same projection D onto the xy plane, but its function is just $z = 0$, which gives us

$$\iint\limits_{S_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iint\limits_{D} [-0 - 0 + z]\mathrm{d}A = 0$$

Because z is always zero.

Then the total surface integral over S is just the sum of the integrals over S_1 and S_2, or

$$\oiint\limits_{S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iint\limits_{S_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} + \iint\limits_{S_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \frac{\pi}{2}$$

△

We can also evaluate surface integrals of vector fields over parametric surfaces. If S is given by a vector function $\boldsymbol{r}(u, v)$ with $(u, v) \in D$, then we have

$$\begin{aligned}
\iint\limits_{S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} &= \iint\limits_{S} \boldsymbol{F} \cdot \frac{\boldsymbol{r}_u \times \boldsymbol{r}_v}{\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|}\mathrm{d}S \\
&= \iint\limits_{D} \boldsymbol{F} \cdot \frac{\boldsymbol{r}_u \times \boldsymbol{r}_v}{\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|}\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|\mathrm{d}A \\
&= \iint\limits_{D} \boldsymbol{F} \cdot (\boldsymbol{r}_u \times \boldsymbol{r}_v)\mathrm{d}A
\end{aligned}$$

If we were to plug in $\boldsymbol{r}(x, y) = \langle x, y, g(x, y) \rangle$ into this equation, we would get the same formula as we previously derived for an integral over a graphed surface.

Example 5.11.6. Gauss' Law for electrostatics in integral form states that the closed surface integral of the electric field over a region S is equal to the amount of charge enclosed within S divided by the constant ε_0. As an equation, this states

$$\oiint_S \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S} = \frac{Q}{\varepsilon_0}$$

Evaluate the surface integral to verify Gauss' Law for the case where S is a sphere of radius a centered around the origin, where there is a positive charge of $+Q$ at the origin.

It will be important to remember that the formula for the electric field given by a point charge at the origin at a point with position vector $\boldsymbol{r}$ is given by

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0} \frac{Q\boldsymbol{r}}{\|\boldsymbol{r}\|^3}$$

Solution: We can write the sphere S as a parametric surface given by

$$\boldsymbol{r}(\phi, \theta) = \langle a\sin\phi\cos\theta, a\sin\phi\sin\theta, a\cos\phi\rangle, \qquad \phi \in [0, \pi] \quad \theta \in [0, 2\pi]$$

We previously found $\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta$ to be the vector

$$\boldsymbol{r}_\phi \times \boldsymbol{r}_\theta = \left\langle a^2\sin^2\phi\cos\theta, a^2\sin^2\phi\sin\theta, a^2\sin\phi\cos\phi\right\rangle$$

Now, we can simply plug in and evaluate the surface integral.

$$\Phi_E = \oiint_S \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S} = \frac{Q}{4\pi\varepsilon_0} \oiint_S \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|^3} \cdot \boldsymbol{n}\mathrm{d}S$$

Note that the magnitude of $\boldsymbol{r}$ is constant at $\|\boldsymbol{r}\| = a$, due to the nature of spheres.

$$\begin{aligned}
\Phi_E &= \frac{Q}{4\pi\varepsilon_0}\frac{1}{a^3} \iint_D \langle x, y, z\rangle \cdot \left\langle a^2\sin^2\phi\cos\theta, a^2\sin^2\phi\sin\theta, a^2\sin\phi\cos\phi\right\rangle \mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0}\frac{1}{a^3} \iint_D \langle a\sin\phi\cos\theta, a\sin\phi\sin\theta, a\cos\phi\rangle \cdot \left\langle a^2\sin^2\phi\cos\theta, a^2\sin^2\phi\sin\theta, a^2\sin\phi\cos\phi\right\rangle \mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0}\frac{1}{a^3} \iint_D \left[a^3\sin^3\phi\cos^2\theta + a^3\sin^3\phi\sin^2\theta + a^3\sin\phi\cos^2\phi\right] \mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0}\frac{a^3}{a^3} \iint_D \left[\sin^3\phi(\sin^2\theta + \cos^2\theta) + \sin\phi\cos^2\phi\right] \mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0} \iint_D [\sin\phi(1 - \cos^2\phi) + \sin\phi\cos^2\phi]\mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0} \iint_D \sin\phi\mathrm{d}A \\
&= \frac{Q}{4\pi\varepsilon_0} \int_0^{2\pi} \int_0^{\pi} \sin\phi\mathrm{d}\phi\mathrm{d}\theta \\
&= \left[\frac{Q}{4\pi\varepsilon_0}\right] [2\pi][\cos 0 - \cos\pi] = \frac{Q}{\varepsilon_0}
\end{aligned}$$

△

Another useful application of surface integrals occurs in the study of heat flow. Suppose the temperature at a point (x, y, z) in a body is $u(x, y, z)$. Then the **heat flow** is defined as the vector field

$$\boldsymbol{F} = -K\boldsymbol{\nabla}u$$

where K is an experimentally determined constant known as the **conductivity** of the substance. The rate of heat flow across the surface S in the body is then given by the flux of the heat flow.

$$R = -K \iint\limits_S \boldsymbol{\nabla} u \cdot \mathrm{d}\boldsymbol{S}$$

Example 5.11.7. The temperature u in a metal ball is proportional to the square of the distance from the center of the ball. Find the rate of heat flow across a sphere S of radius a centered at the center of the ball.

Solution: First, we can write the temperature as a function of the position:

$$u = C(x^2 + y^2 + z^2)$$

Then, the heat flow is given by

$$-K\boldsymbol{\nabla} u = \langle -2KCx, 2KCy, 2KCz\rangle$$

Recall that the normal vector to a sphere is given by

$$\boldsymbol{n} = \frac{1}{a}\langle x, y, z\rangle$$

so we find

$$\begin{aligned}\Phi_T = \oiint\limits_S -K\boldsymbol{\nabla} u \cdot \mathrm{d}\boldsymbol{S} &= \oiint\limits_S -K\boldsymbol{\nabla} u \cdot \boldsymbol{n}\mathrm{d}S \\ &= \frac{-2KC}{a} \oiint\limits_S \langle x, y, z\rangle \cdot \langle x, y, z\rangle \,\mathrm{d}S \\ &= \frac{-2KC}{a} \oiint\limits_S \rho^2 \mathrm{d}S\end{aligned}$$

But since ρ is constant at $\rho = a$,

$$\Phi_T = -2KCa \oiint\limits_S \mathrm{d}S$$

Remember that the surface integral of 1 is just the surface area of S. We know the surface area of a circle to be $4\pi a^2$ from geometry, so we have

$$\Phi_T = -8KC\pi a^3$$

△

5.12 Stokes' Theorem

Stokes' Theorem can be thought of as a higher-dimensional equivalent of Green's Theorem. While Green's Theorem relates the line integral over a the boundary of a plane region ∂D to the double integral over the whole region D, Stokes' Theorem relates a line integral over a surface S to the surface integral over a surface S.

Theorem 5.12.1 (Stokes' Theorem)**.** Let S be an oriented piecewise-smooth surface in $\mathbb{R}^3$ that is bounded by a simple, closed, piecewise-smooth boundary curve ∂S. Let $\boldsymbol{F}$ be a vector field whose components have continuous partial derivatives on an open region in $\mathbb{R}^3$ that contains S. Then,

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_S (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{S}$$

Because $\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \oint_{\partial S} \boldsymbol{F} \cdot \boldsymbol{T} \mathrm{d}s$ and $\iint_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iint_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}S$, Stokes' Theorem can be understood as saying that the line integral of the component of $\boldsymbol{F}$ tangent to ∂S is equal to the surface integral of the component of $\boldsymbol{\nabla} \times \boldsymbol{F}$ normal to S.

Green's Theorem can be thought of as a special case of Stokes' Theorem. In fact, when S is a region in the plane, the unit normal is $\hat{\boldsymbol{k}}$ we find

$$\iint_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}S = \iint_S (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \hat{\boldsymbol{k}} \mathrm{d}A$$

which is precisely the vector form of Green's Theorem.

Although it is too difficult for us to prove Stokes' Theorem generally, we can give a proof for the case when S is a graph and $\boldsymbol{F}$, S, and ∂S are all well-behaved.

Proof of a Special Case of Stokes' Theorem. Let S be defined by the graph $z = g(x, y)$ with $(x, y) \in D$, where g has continuous second-order partial derivatives. Let D be a simple plane region whose boundary curve ∂D corresponds to ∂S (in other words, D is the "projection" of S onto the plane).

Let $\boldsymbol{F} = \langle P, Q, R \rangle$ be a vector field with continuous first partials on S.

Because S is the graph of a function, we can compute $\iint_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S}$ as so:

$$\begin{aligned}
\iint\limits_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} &= \iint\limits_D \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \left\langle -\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1 \right\rangle \mathrm{d}A \\
&= \iint\limits_D \left[-\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \frac{\partial z}{\partial x} - \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \frac{\partial z}{\partial y} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \right] \mathrm{d}A \\
&= \iint\limits_D \left[\left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y} \right) \frac{\partial z}{\partial x} + \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} \right) \frac{\partial z}{\partial y} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \right] \mathrm{d}A
\end{aligned}$$

Now, we can evaluate the line integral $\int_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$. We can write a parametric representation of ∂S with

$$\boldsymbol{r}(t) = \langle x(t), y(t), g(x(t), y(t)) \rangle \quad t \in [a, b]$$

And then compute the line integral

$$\begin{aligned}
\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) \mathrm{d}t \\
&= \int_a^b \left[P \frac{\mathrm{d}x}{\mathrm{d}t} + Q \frac{\mathrm{d}y}{\mathrm{d}t} + R \frac{\mathrm{d}}{\mathrm{d}t} [g(x(t), y(t))] \right] \mathrm{d}t
\end{aligned}$$

By the chain rule, we can find that $\frac{\mathrm{d}}{\mathrm{d}t}[g(x(t), y(t))] = \frac{\partial g}{\partial x} \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial g}{\partial y} \frac{\mathrm{d}y}{\mathrm{d}t}$, which we can substitute into our previous expression to find

$$\begin{aligned}
\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} &= \int_a^b \left[P \frac{\mathrm{d}x}{\mathrm{d}t} + Q \frac{\mathrm{d}y}{\mathrm{d}t} + R \frac{\partial g}{\partial x} \frac{\mathrm{d}x}{\mathrm{d}t} + R \frac{\partial g}{\partial y} \frac{\mathrm{d}y}{\mathrm{d}t} \right] \mathrm{d}t \\
&= \int_a^b \left[\left(P + R \frac{\partial g}{\partial x} \right) \frac{\mathrm{d}x}{\mathrm{d}t} + \left(Q + R \frac{\partial g}{\partial y} \right) \frac{\mathrm{d}y}{\mathrm{d}t} \right] \mathrm{d}t \\
&= \int_{\partial D} \left(P + R \frac{\partial g}{\partial x} \right) \mathrm{d}x + \left(Q + R \frac{\partial g}{\partial y} \right) \mathrm{d}y
\end{aligned}$$

Using Green's Theorem, we can rewrite this as

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_D \left[\frac{\partial}{\partial x}\left(Q + R\frac{\partial g}{\partial y}\right) - \frac{\partial}{\partial y}\left(P + R\frac{\partial g}{\partial t}\right)\right]\mathrm{d}A$$
$$= \iint\limits_D \left[\frac{\partial}{\partial x}[Q(x,y,g(x,y)] + \frac{\partial}{\partial x}[R(x,y,g(x,y)]\frac{\partial g}{\partial y} + R\frac{\partial^2 g}{\partial x \partial y} - \frac{\partial}{\partial y}[P(x,y,g(x,y)] - \right.$$
$$\left. \frac{\partial}{\partial y}[R(x,y,g(x,y)]\frac{\partial g}{\partial x} - R\frac{\partial^2 g}{\partial x \partial y}\right]\mathrm{d}A$$

By the chain rule, we can say that $\frac{\partial}{\partial x}Q(x,y,g(x,y)) = \frac{\partial Q}{\partial x} + \frac{\partial Q}{\partial g}\frac{\partial g}{\partial x}$ (and similar for the other derivatives), which then gives us

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_D \left[\left(\frac{\partial Q}{\partial x} + \frac{\partial Q}{\partial g}\frac{\partial g}{\partial x}\right) + \left(\frac{\partial R}{\partial x}\frac{\partial g}{\partial y} + \frac{\partial R}{\partial g}\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\right) + \left(-\frac{\partial P}{\partial y} - \frac{\partial P}{\partial g}\frac{\partial g}{\partial y}\right) + \left(-\frac{\partial R}{\partial y}\frac{\partial g}{\partial x} - \frac{\partial R}{\partial g}\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\right)\right]\mathrm{d}A$$
$$= \iint\limits_D \left[\left(\frac{\partial Q}{\partial g} - \frac{\partial R}{\partial y}\right)\frac{\partial g}{\partial x} + \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial g}\right)\frac{\partial g}{\partial y} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\right]\mathrm{d}A$$

Recalling that $z = g(x,y)$, this can also be written as

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \iint\limits_D \left[\left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y}\right)\frac{\partial z}{\partial x} + \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right)\frac{\partial z}{\partial y} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\right]\mathrm{d}A$$

Which is precisely equal to $\iint\limits_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S}$, proving the theorem. □

Example 5.12.2. Evaluate $\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$, where $\boldsymbol{F} = \left\langle -y^2, x, z^2 \right\rangle$ and C is the curve of intersection of the plane $y + z = 2$ and the cylinder $x^2 + y^2 = 1$. Orient C so that it faces counterclockwise when viewed from above.

Solution: Although we could evaluate the line integral with the position function $\boldsymbol{r}(t) = \langle \cos t, \sin t, 2 - \sin t \rangle$, it will be much easier to compute by using Stokes' Theorem. First, compute $\boldsymbol{\nabla} \times F$.

$$\boldsymbol{\nabla} \times \boldsymbol{F} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y^2 & x & z^2 \end{vmatrix}$$
$$= \langle 0, 0, 1 + 2y \rangle$$

Now, C is the boundary of the surface S given by $z = 2 - y$ that is over the plane disk $x^2 + y^2 = 1$, so

$$\iint\limits_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iint\limits_R \langle 0, 0, 1 + 2y \rangle \cdot \left\langle -\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1 \right\rangle \mathrm{d}A$$
$$= \int_0^{2\pi} \int_0^1 [r + 2r^2 \sin\theta] \mathrm{d}r \mathrm{d}\theta$$
$$= \int_0^{2\pi} \left[\frac{1}{2} + \frac{2}{3}\sin\theta\right] \mathrm{d}\theta = \pi$$

△

Example 5.12.3. Use Stokes' Theorem to compute the integral $\iint_S \boldsymbol{\nabla} \times F \cdot \mathrm{d}\boldsymbol{S}$, where $\boldsymbol{F} = \langle xz, yz, xy \rangle$ and S is the part of the sphere $x^2 + y^2 + z^2 = 4$ that lies inside the cylinder $x^2 + y^2 = 1$ and above the xy-plane.

Solution: To find the boundary curve C, we need to find the set of all points that satisfy both $x^2 + y^2 + z^2 = 4$ and $x^2 + y^2 = 1$.

First, subtract the second equation from the first one to obtain $z^2 = 3$ so $z = \pm\sqrt{3}$. Because we're above the xy plane, we narrow this down to only $z = \sqrt{3}$.

The equations for x and y are easy to see from the boundary of the cylinder, so we get:

$$\boldsymbol{r}(t) = \left\langle \cos t, \sin t, \sqrt{3} \right\rangle$$

So we can rewrite

$$\iint\limits_S \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

where ∂S is the curve described by $\boldsymbol{r}(t)$ with $t \in [0, 2\pi]$. Computing this line integral, we find

$$\begin{aligned}
\int_0^{2\pi} \boldsymbol{F}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)\mathrm{d}t &= \int_0^{2\pi} \left\langle \sqrt{3}\cos t, \sqrt{3}\sin t, \sin t \cos t \right\rangle \cdot \langle -\sin t, \cos t, 0 \rangle \, \mathrm{d}t \\
&= \int_0^{2\pi} [\sqrt{3}\sin t \cos t - \sqrt{3}\sin t \cos t]\mathrm{d}t = 0
\end{aligned}$$

△

Stokes' Theorem has an interesting implication, similar to independence of path in line integrals. If there are two surfaces S_1 and S_2 with the same boundary curve C, then we have

$$\iint\limits_{S_1} \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iint\limits_{S_2} \boldsymbol{\nabla} \times \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$$

We can now use Stokes' Theorem to give some meaning to the curl vector. Suppose that C is an oriented closed curve and $\boldsymbol{v}(x, y, z)$ represents a velocity field of a fluid. Consider the line integral

$$\oint_C \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} = \oint_C \boldsymbol{v} \cdot \boldsymbol{T}\mathrm{d}s$$

Where $\boldsymbol{T}$ is the unit tangent vector to C. Recall that $\boldsymbol{v} \cdot \boldsymbol{T}$ is the component of $\boldsymbol{v}$ in the direction of $\boldsymbol{T}$ (because $\boldsymbol{T}$ is a unit vector).

When $\boldsymbol{v}$ is pointing in a direction similar to $\boldsymbol{T}$, $\boldsymbol{v} \cdot \boldsymbol{T}$ becomes larger. When $\boldsymbol{v}$ is pointing in a direction almost perpendicular to $\boldsymbol{T}$, $\boldsymbol{v} \cdot \boldsymbol{T}$ becomes closer to zero. When $\boldsymbol{v}$ is pointing close to the opposite direction of $\boldsymbol{T}$, then $\boldsymbol{v} \cdot \boldsymbol{T}$ is large and negative.

Therefore, we can say that $\oint_C \boldsymbol{v} \cdot \boldsymbol{T}\mathrm{d}s$ is a measure of the tendency of a fluid to move around C and is called the **circulation** of $\boldsymbol{v}$ around C.

Let P_0 be a point in the domain of $\boldsymbol{v}$ with position vector $\boldsymbol{p}_0$ and S_a be a small disk with radius a centered at P_0. Then, we can approximate the value of $(\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p})$ for any $\boldsymbol{p} \in S_a$ with

$$(\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}) \approx (\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0)$$

If we define C_a as the boundary circle of S_a, we then, by Stokes' Theorem, get

$$\begin{aligned}
\oint_{C_a} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r} &= \iint\limits_{S_a} (\boldsymbol{\nabla} \times \boldsymbol{v}) \cdot \mathrm{d}\boldsymbol{S} = \iint\limits_{S_a} (\boldsymbol{\nabla} \times \boldsymbol{v}) \cdot \boldsymbol{n}\mathrm{d}S \\
&\approx \iint\limits_{S_a} (\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0) \cdot \boldsymbol{n}(\boldsymbol{p}_0)\mathrm{d}S = (\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0) \cdot \boldsymbol{n}(\boldsymbol{p}_0) \iint\limits_{S_a} \mathrm{d}S \\
&= [(\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0) \cdot \boldsymbol{n}(\boldsymbol{p}_0)]\pi a^2
\end{aligned}$$

We can rearrange to find

$$(\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0) \cdot \boldsymbol{n}(\boldsymbol{p}_0) \approx \frac{1}{\pi a^2} \oint_{C_a} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}$$

As the disk gets smaller and $a \to 0$, this approximation gets better. So we then find

$$(\boldsymbol{\nabla} \times \boldsymbol{v})(\boldsymbol{p}_0) \cdot \boldsymbol{n}(\boldsymbol{p}_0) = \lim_{a \to 0} \frac{1}{\pi a^2} \oint_{C_a} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{r}$$

This gives us a relationship between the curl and the circulation of an infinitesimally small bounding disk. It shows that $(\boldsymbol{\nabla} \times \boldsymbol{v}) \cdot \boldsymbol{n}$ is a measure of the rotation effect of the fluid about an axis $\boldsymbol{n}$. The curling effect is greatest when $\boldsymbol{n}$ is parallel to $\boldsymbol{\nabla} \times \boldsymbol{v}$.

We can also use our knowledge of Stokes' Theorem to prove Theorem 5.9.4, which states that a vector field $\boldsymbol{F}$ is conservative if and only if $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$ across all of $\mathbb{R}^3$.

Proof of Theorem 5.9.4. Suppose $\boldsymbol{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field whose components each have continuous first order partial derivatives across all of $\mathbb{R}^3$. Let S be an arbitrary closed surface in $\mathbb{R}^3$ with a closed boundary ∂S.

In the forward direction, we must prove that $\boldsymbol{F}$ is conservative if $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$. If $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$, we have, by Stokes' Theorem,

$$\oint_{\partial S} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = \oiint_S (\boldsymbol{\nabla} \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{S} = \oiint_S \boldsymbol{0} \cdot \mathrm{d}\boldsymbol{S} = 0$$

Therefore, $\boldsymbol{F}$ is independent of path (all closed paths ∂S will have a corresponding closed surface S, although the proof of this fact is nontrivial) and conservative by Theorem 5.5.4.

In the reverse direction, we must prove that $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$ if $\boldsymbol{F}$ is conservative. If $\boldsymbol{F}$ is conservative, then $\boldsymbol{F} = \boldsymbol{\nabla} f$ for some scalar field $f : \mathbb{R}^3 \to \mathbb{R}$. Then, by Theorem 5.9.3, $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$.

Therefore, $\boldsymbol{\nabla} \times \boldsymbol{F} = \boldsymbol{0}$ holds if $\boldsymbol{F}$ is conservative, proving the theorem. □

5.13 The Divergence Theorem

Previously, we wrote Green's Theorem in a vector form as

$$\int_{\partial D} \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}s = \iint_D (\boldsymbol{\nabla} \cdot \boldsymbol{F}) \mathrm{d}A$$

For a plane region D. If we sought to extend this to $\mathbb{R}^3$, a good guess would be that we would obtain

$$\iint_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n} \mathrm{d}S = \iiint_V (\boldsymbol{\nabla} \cdot \boldsymbol{F}) \mathrm{d}V$$

For a solid region E. This will actually turn out to be the case. Once again, notice the similarity to the other theorems we've explored in this chapter. It relates the integral of a derivative-like function ($\boldsymbol{\nabla} \cdot \boldsymbol{F}$ in this case) over a region to the original function integrated only across the boundary of the region.

We will state and prove the divergence theorem for regions that are simultaneously what we called type I, II, and III in the triple integrals section. For review, the meanings of each type are below.

1. Type I: Regions that can be written as $\{(x, y, z) | (x, y) \in D_1, u_1(x, y) \leq z \leq u_2(x, y)\}$.
2. Type 2: Regions that can be written as $\{(x, y, z) | (y, z) \in D_2, u_1(y, z) \leq x \leq u_2(y, z)\}$.
3. Type 3: Regions that can be written as $\{(x, y, z) | (x, z) \in D_3, u_1(x, z) \leq y \leq u_2(x, z)\}/$

Where D_1, D_2, and D_3 are regions in the xy, yz, and xz planes respectively. Regions that are all of these types simultaneously are known as **simple solid regions**.

Simple solid regions have a closed boundary, with the convention that the normal vector points outwards.

Theorem 5.13.1 (The Divergence Theorem)**.** Let E be a simple solid region with a closed boundary ∂E. Let $\boldsymbol{F}$ be a vector field whose component functions have continuous partial derivatives on an open region that contains E. Then,

$$\oiint_{\partial E} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iiint_{E} (\boldsymbol{\nabla} \cdot \boldsymbol{F})\mathrm{d}V$$

Proof. Let $\boldsymbol{F} = \langle P, Q, R \rangle$. Then,

$$(\boldsymbol{\nabla} \cdot \boldsymbol{F}) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

which gives us

$$\iiint_{E} (\boldsymbol{\nabla} \cdot \boldsymbol{F})\mathrm{d}V = \iiint_{E} \frac{\partial P}{\partial x}\mathrm{d}V + \iiint_{E} \frac{\partial Q}{\partial y}\mathrm{d}V + \iiint_{E} \frac{\partial R}{\partial z}\mathrm{d}V$$

Now, we can rewrite the surface integral.

$$\begin{aligned}\oiint_{\partial E} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} &= \oiint_{\partial E} \boldsymbol{F} \cdot \boldsymbol{n}\mathrm{d}S \\ &= \oiint_{\partial E} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S + \oiint_{\partial E} (Q\hat{\boldsymbol{\jmath}} \cdot \boldsymbol{n})\mathrm{d}S + \oiint_{\partial E} (R\hat{\boldsymbol{k}} \cdot \boldsymbol{n})\mathrm{d}S\end{aligned}$$

This gives us three equalities we must show to prove the theorem:

$$\iiint_{E} \frac{\partial P}{\partial x}\mathrm{d}V = \oiint_{\partial E} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S$$

$$\iiint_{E} \frac{\partial Q}{\partial y}\mathrm{d}V = \oiint_{\partial E} (Q\hat{\boldsymbol{\jmath}} \cdot \boldsymbol{n})\mathrm{d}S$$

$$\iiint_{E} \frac{\partial R}{\partial z}\mathrm{d}V = \oiint_{\partial E} (R\hat{\boldsymbol{k}} \cdot \boldsymbol{n})\mathrm{d}S$$

To show the first equation, use the fact that E can be written as

$$E = \{(x, y, z) | (y, z) \in D, u_1(y, z) \leq x \leq u_2(y, z)\}$$

where D is the projection of E onto the yz plane. This allows us to use Fubini's Theorem to rewrite the triple integral as

$$\begin{aligned}\iiint_{E} \frac{\partial P}{\partial x}\mathrm{d}V &= \iint_{D} \left[\int_{u_1(y,z)}^{u_2(y,z)} \frac{\partial P}{\partial x}\mathrm{d}x\right]\mathrm{d}A \\ &= \iint_{D} [P(u_2(y, z), y, z) - P(u_1(y, z), y, z)]\mathrm{d}A\end{aligned}$$

Returning to the surface integral across ∂E, note that because E is a simple solid region, ∂E can be split up into two or three segments, regarding the yz plane as the "floor": ∂E_1 which is the bottom surface, ∂E_2 which is the top surface, and (in some cases) a vertical surface (with respect to the yz plane) ∂E_3, which connects ∂E_1 and ∂E_2. On ∂E_3, we have a normal vector that is orthogonal to the x axis, which tells us that $\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n}$ is zero along S_3. Thus, we can remove this part of the surface integral to find

$$\begin{aligned}\oiint_{\partial E} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S &= \iint_{\partial E_1} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S + \iint_{\partial E_2} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S + \iint_{\partial E_3} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S \\ &= \iint_{\partial E_1} (P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S + \iint_{\partial E_2} (P\hat{\boldsymbol{\jmath}} \cdot \boldsymbol{n})\mathrm{d}S\end{aligned}$$

The bottom surface E_1 is given by $\{(x,y,z)|(y,z) \in D, x = u_1(y,z)\}$ and the top surface E_2 is given by $\{(x,y,z)|(y,z) \in D, x = u_2(y,z)\}$. This allows us to write

$$\begin{aligned}\oiint_{\partial E}(P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S &= -\iint_D P(u_1(y,z),y,z)\mathrm{d}A + \iint_D P(u_2(y,z),y,z)\mathrm{d}A \\ &= \iint_D [P(u_2(y,z),y,z) - P(u_1(y,z),y,z)]\mathrm{d}A\end{aligned}$$

Because D is the projection of both ∂E_1 and ∂E_2 onto the yz plane, and the first surface integral is negative because the unit normal to ∂E_1 points downwards with respect to the yz plane.

Now, we've shown that

$$\iiint_E \frac{\partial P}{\partial x}\mathrm{d}V = \oiint_{\partial E}(P\hat{\boldsymbol{\imath}} \cdot \boldsymbol{n})\mathrm{d}S$$

Similar arguments using the fact that E is type II and type III can be used to show the other two equalities, thus proving the theorem. $\square$

Example 5.13.2. Find the flux of the vector field $\boldsymbol{F} = \langle z, y, x\rangle$ over the unit sphere $x^2 + y^2 + z^2 = 1$.

Solution: We will use the Divergence Theorem to save us a surface integral computation, and instead just do a triple integral.

First, we can compute $\boldsymbol{\nabla} \cdot \boldsymbol{F}$,

$$\boldsymbol{\nabla} \cdot \boldsymbol{F} = \frac{\partial}{\partial x}z + \frac{\partial}{\partial y}y + \frac{\partial}{\partial z}x = 1$$

and then,

$$\iiint_E (\boldsymbol{\nabla} \cdot \boldsymbol{F})\mathrm{d}V = \iiint_E \mathrm{d}V$$

This is just the volume of the sphere of radius 1, which is $\frac{4\pi}{3}$. By the Divergence Theorem, we then find

$$\oiint_{\partial E} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \frac{4\pi}{3}$$

$\triangle$

Example 5.13.3. Evaluate $\oiint_{\partial E} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S}$ where

$$\boldsymbol{F} = \left\langle xy, y^2 + e^{xz^2}, \sin(xy)\right\rangle$$

and E is the region bounded by the parabolic cylinder $z = 1 - x^2$, the planes $z = 0$, $y = 0$, and $z = 2 - y$.

Solution: First, computing $\boldsymbol{\nabla} \cdot \boldsymbol{F}$:

$$\boldsymbol{\nabla} \cdot \boldsymbol{F} = y + 2y = 3y$$

Solving for the intersection of $z = 1 - x^2$ and $z = 0$, we find it to be at $x = \pm 1$.

With this in mind, E can be written as

$$E = \{(x,y,z)|x \in [-1,1], y \in [0, 2-z], z \in [0, 1-x^2]\}$$

and then the triple integral can be computed.

$$\oiint\limits_{\partial E} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} = \iiint\limits_{E} (\boldsymbol{\nabla} \boldsymbol{\cdot} \boldsymbol{F})\mathrm{d}V = \int_{-1}^{1} \int_{0}^{1-x^2} \int_{0}^{2-z} [3y]\mathrm{d}y\mathrm{d}z\mathrm{d}x$$
$$= \frac{3}{2} \int_{-1}^{1} \int_{0}^{1-x^2} (2-z)^2 \mathrm{d}z\mathrm{d}x$$
$$= -\frac{1}{2} \int_{-1}^{1} [2-z]^3 \bigg|_{z=0}^{z=1-x^2} \mathrm{d}x$$
$$= -\frac{1}{2} \int_{-1}^{1} ([1+x^2]^3 - 8)\mathrm{d}x$$
$$= -\int_{-1}^{1} (x^6 + 3x^3 + 3x^2 - 7)\mathrm{d}x = \frac{184}{35}$$

△

Although we previously only proved the Divergence Theorem for simple solid regions, we can extend it to regions that are the finite union of simple solid regions using a similar procedure to the one we used to extend Green's Theorem to finite unions of simple regions.

For instance, consider a region E that lies between the closed boundary surfaces ∂E_1 and ∂E_2 where ∂E_1 lies inside ∂E_2. This gives us the region that lies between these two boundary surfaces. If $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are the outward unit normal vectors of the boundary surfaces, we can notice that $\boldsymbol{n}_1$ points *towards* E instead of away from it. This means that when we do our surface integral along ∂E_1, we want to use $-\boldsymbol{n}_1$ instead of $\boldsymbol{n}_1$.

Putting this into equation form, we find

$$\iiint\limits_{E} (\boldsymbol{\nabla} \boldsymbol{\cdot} \boldsymbol{F})\mathrm{d}V = \oiint\limits_{\partial E_1 \cup \partial E_2} \boldsymbol{F} \cdot \boldsymbol{n}\mathrm{d}S$$
$$= \oiint\limits_{\partial E_1} \boldsymbol{F} \cdot (-\boldsymbol{n}_1)\mathrm{d}S + \oiint\limits_{\partial E_2} \boldsymbol{F} \cdot \boldsymbol{n}_2\mathrm{d}S$$
$$= -\oiint\limits_{\partial E_1} \boldsymbol{F} \cdot \boldsymbol{n}_1\mathrm{d}S + \oiint\limits_{\partial E_2} \boldsymbol{F} \cdot \boldsymbol{n}_2\mathrm{d}$$

We can now apply this to the electric field, specifically relating to Gauss' Law for electricity. We previously showed that the electric flux through a sphere centered around the origin is Q/ε_0, where Q is the amount of charge at the origin and ε_0 is a constant.

Now, we will seek to apply this to a more general case. Let ∂E_1 be a sphere centered around the origin with radius a, let ∂E_2 be an arbitrary closed boundary, and let E be the region between them.

Recalling that the formula for the electric field is

$$\boldsymbol{E}(\boldsymbol{x}) = \frac{1}{4\pi\varepsilon_0} \frac{Q\boldsymbol{x}}{\|\boldsymbol{x}\|^3}$$

we can compute $\boldsymbol{\nabla} \cdot \boldsymbol{E}$ to find

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{E} &= \frac{\partial}{\partial x}\left[\frac{1}{4\pi\varepsilon_0}\frac{Qx}{\|\boldsymbol{x}\|^3}\right] + \frac{\partial}{\partial y}\left[\frac{1}{4\pi\varepsilon_0}\frac{Qy}{\|\boldsymbol{x}\|^3}\right] + \frac{\partial}{\partial z}\left[\frac{1}{4\pi\varepsilon_0}\frac{Qz}{\|\boldsymbol{x}\|^3}\right] \\
&= \frac{Q}{4\pi\varepsilon_0}\left[\frac{\partial}{\partial x}\left(\frac{x}{(x^2+y^2+z^2)^{3/2}}\right) + \frac{\partial}{\partial y}\left(\frac{y}{(x^2+y^2+z^2)^{3/2}}\right) + \frac{\partial}{\partial z}\left(\frac{z}{(x^2+y^2+z^2)^{3/2}}\right)\right] \\
&= \frac{Q}{4\pi\varepsilon_0}\left[(x^2+y^2+z^2)^{1/2}\right]\left[\frac{3x^2-(x^2+y^2+z^2)+3y^2-(x^2+y^2+z^2)+3z^2-(x^2+y^2+z^2)}{(x^2+y^2+z^2)^{3/2}}\right] \\
&= \frac{Q}{4\pi\varepsilon_0}\frac{3(x^2+y^2+z^2)-3(x^2+y^2+z^2}{x^2+y^2+z^2} = 0
\end{aligned}$$

This tells us that $\iint_E(\boldsymbol{\nabla} \cdot \boldsymbol{E}) = 0$. Using the divergence theorem, we can then state

$$\begin{aligned}
\iint\limits_E (\boldsymbol{\nabla} \cdot \boldsymbol{E}) &= -\oiint\limits_{\partial E_1} \boldsymbol{E} \cdot \boldsymbol{n}_1 \mathrm{d}S + \oiint\limits_{\partial E_2} \boldsymbol{E} \cdot \boldsymbol{n}_2 \mathrm{d}S \\
0 &= -\oiint\limits_{\partial E_1} \boldsymbol{E} \cdot \boldsymbol{n}_1 \mathrm{d}S + \oiint\limits_{\partial E_2} \boldsymbol{E} \cdot \boldsymbol{n}_2 \mathrm{d}S
\end{aligned}$$

Which can be rearranged to obtain

$$\oiint\limits_{\partial E_1} \boldsymbol{E} \cdot \boldsymbol{n}_1 \mathrm{d}S = \oiint\limits_{\partial E_2} \boldsymbol{E} \cdot \boldsymbol{n}_2 \mathrm{d}S$$

This effectively tells us that the flux across our arbitrary boundary curve ∂E_2 is equal to the flux across ∂E_1, which know to be the sphere of radius a centered around the origin. Now, we can compute the flux of $\boldsymbol{E}$ across ∂E_1.

$$\Phi_E = \oiint\limits_{\partial E_1} \boldsymbol{E} \cdot \boldsymbol{n} \mathrm{d}\boldsymbol{S} = \iint\limits_D \boldsymbol{E} \cdot \left\langle a^2 \sin^2\phi\cos\theta, a^2\sin^2\phi\sin\theta, a^2\sin\phi\sin\theta\right\rangle \mathrm{d}A$$

Which we already computed in a previous example to be equal to $\frac{Q}{\varepsilon_0}$.

Thus, we have shown a slightly more general case of Gauss' Law. The electric flux across any arbitrary surface that encloses the origin, where there is a point charge at the origin (and nowhere else), is equal to the charge divided by ε_0.

Another application of the Divergence Theorem occurs in fluid flow. Let $\boldsymbol{v}(x, y, z)$ be the velocity field of a fluid with constant density ρ. Then $\boldsymbol{F} = \rho\boldsymbol{v}$ is the rate of flow per unit area. If $P_0(x_0, y_0, z_0)$ is a point in the fluid and B_a is a ball with center P_0 and small radius a, then $(\boldsymbol{\nabla} \cdot \boldsymbol{F})(P) \approx (\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0)$ for all $P \in B_a$. We can then approximate the flux over the boundary sphere ∂B_a:

$$\begin{aligned}
\oiint\limits_{\partial B_a} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S} &= \iiint\limits_{B_a} (\boldsymbol{\nabla} \cdot \boldsymbol{F}) \mathrm{d}V \\
&\approx \iiint\limits_{B_a} (\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0) \mathrm{d}V \\
&= (\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0) \iiint\limits_{B_a} \mathrm{d}V \\
&= \frac{4}{3}\pi a^3 (\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0)
\end{aligned}$$

As $a \to 0$, this approximation becomes better, suggesting that

$$(\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0) = \lim_{a \to 0} \frac{3}{4\pi a^3} \oiint_{\partial B_a} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{S}$$

Because the flux is equal to the flow rate per unit area, This equation is saying that $(\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0)$ represents the flow rate per unit volume at P_0. If $(\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0) > 0$, then fluid is attempting to flow away from P_0 under $\boldsymbol{v}$, and P_0 is called as **source**. If $(\boldsymbol{\nabla} \cdot \boldsymbol{F})(P_0) < 0$, then fluid is attempting to flow into P_0 under $\boldsymbol{v}$, and P_0 is called a **sink**.

This is the end of the notes. banger