Calculus 101

alexander

July 25, 2025

Precalculus

the least common multiple of two (or more) numbers is the smallest number that is a multiple of both (or all) of them:

- multiples of 4: 4, 8, 12, 16, 20

- multiples of 6: 6, 12, 18, 24

the greatest common factor of two (or more) numbers is the largest number that divides both (or all) of them without a remainder:

- factors of 12: 1, 2, 3, 4, 6, 12

- factors of 18: 1, 2, 3, 6, 9, 18

the GCD is the same as the GCF they both refer to the largest number that evenly divides two or more integers.

prime factorization: GCD(72, 27)

- $\frac{72}{2} = 36$

- $\frac{36}{2} = 18$

- $\frac{18}{2} = 9$

- $\frac{9}{3} = 3$

- $\frac{3}{3} = 1$

- $72 = 2^3 \cdot 3^2$

- $\frac{27}{3} = 9$

- $\frac{9}{3} = 3$

- $\frac{3}{3} = 1$

- $27 = 3^3$

the common factor is $3^2 = 9$

euclidean algorithm: GCD(72, 27)
keep dividing and replacing until the remainder is 0. the last non-zero remainder is the GCD.

- $\frac{72}{27} = 2$ remainder 18

- $\frac{27}{18} = 1$ remainder 9
- $\frac{18}{9} = 2$ remainder 0

GCD = 9

inequalities and absolute value:

- $[a, b] = a \leq x \leq b$
- $(a, b) = a < x < b$
- $[a, b) = a \leq x < b$
- $(a, b] = a < x \leq b$
- $(-r, r) = |x| < r$
- $(c - a, c + a) = |x - c| < a = c - a < x < c + a$

triangle inequality

$|a + b| \leq |a| + |b| \rightarrow (|a + b|)^2 \leq (|a| + |b|)^2$

if you square any real number you will get a non-negative result $((x)^2 = (-x)^2)$

1. $(a + b)^2 \leq (|a| + |b|)^2$
2. $a^2 + 2ab + b^2 \leq (|a|)^2 + 2|a||b| + (b)^2$
3. $2ab \leq 2|a||b|$
4. $ab \leq |a||b|$

lines equations:

- $\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$
- $\frac{y_2 - y_1}{x_2 - x_1}$
- $y = mx + b$
- $y - y_1 = m(x - x_1)$
- $Ax + By = C$

eqautions forms:

- conic sections and lines: $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$
    - lines
    - parabolas
    - circles
    - ellipses
    - hyperbolas
- parametric form: $\begin{cases} x = f(t) \\ y = g(t) \end{cases}$ for $t \in [a, b]$

2

- polar form: $r = f(\theta)$

- matrix form:

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

$$-\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}}$$

distance:

- $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

pythagorean:

- $a^2 = b^2 + c^2$

triangle:

- $A = \frac{1}{2}bh$

- $A = \frac{1}{2}ab\sin(\theta)$

circle:

- $C = 2\pi r$

- $A = \pi r^2$

sector of circle:

- $A = \frac{1}{2}r^2\theta$

- $S = r\theta$

sphere:

- $V = \frac{4}{3}\pi r^3$

- $A_s = 4\pi r^2$

cylinder:

- $V = \pi r^2 h$

cone:

- $V = \frac{1}{3}\pi r^2 h$

- $A_s = \pi r \sqrt{r^2 + h^2}$

cone with arbitrary base: where A is the area of the base:

- $V = \frac{1}{3}Ah$

factoring:

1. $x^2 - 5x + 6$

2. $(x-3)(x-2)$

factoring by grouping:

1. $3x^2 - 8x + 4$

2. $3x^2 - 6x - 2x + 4$

3. $(3x^2 - 6x) + (-2x + 4)$

4. $3x(x-2) - 2(x-2)$

5. $(3x-2)(x-2)$

completing the square:

1. $3x^2 + 7x + 4$

2. $3(x^2 + \frac{7}{3}x + 4)$

3. $3((x^2 + \frac{7}{3}x) + 4)$

4. $3((x^2 + \frac{7}{3} + \frac{49}{36}) - \frac{49}{36}) + 4$

5. $3((x + \frac{7}{6})^2 - \frac{49}{36}) + 4$

6. $3(x + \frac{7}{6})^2 - \frac{49}{12} + 4$

7. $3(x + \frac{7}{6})^2 - \frac{49}{12} + 4$

8. $3(x + \frac{7}{6})^2 - \frac{1}{12}$

obtaining solutions:

1. $ax^2 + bx + c = 0$

2. $a(x-h)^2 + k = 0$

3. $a(x-h)^2 = k'$ (if you find that k prime is negative you will have complex solutions)

4. $(x-h)^2 = \frac{k'}{a}$

5. $x - h = \pm\sqrt{\frac{k'}{a}}$

6. $x = \pm\sqrt{\frac{k'}{a}} + h$

derivation of completing the square:

1. $ax^2 + bx + c = 0$

2. $x^2 + \frac{b}{a}x + \frac{c}{a} = 0$

3. $x^2 + \frac{b}{a}x + (\frac{b}{2a})^2 - (\frac{b}{2a})^2 + \frac{c}{a} = 0$

4. $(x + \frac{b}{2a})^2 = (\frac{b}{2a})^2 - \frac{c}{a}$

5. $(x + \frac{b}{2a})^2 = (\frac{b^2}{4a^2}) - \frac{4ac}{4a^2}$

6. $(x + \frac{b}{2a})^2 = \frac{b^2 - 4ac}{4a^2}$

7. $x + \frac{b}{2a} = \pm \frac{\sqrt{b^2 - 4ac}}{2a}$

8. $x = -b \pm \frac{\sqrt{b^2 - 4ac}}{2a}$

rational root theorem:

If a polynomial with integer coefficients has any rational roots, then those roots must be of the form: $\frac{p}{q}$ where p is a factor of the constant term and q is a factor of the leading term coefficient.

consider $f(x) = x^3 - 6x^2 + 11x - 6$
so, the possible values for p: ± 1, ± 2, ± 3, ± 6 and the possible values for q: ± 1
now form all possible $\frac{p}{q}$: $\pm\frac{1}{1}$, $\pm\frac{2}{1}$, $\pm\frac{3}{1}$, $\pm\frac{6}{1}$
You can test each value by plugging into the polynomial to see if it equals 0. If it does, it is a root, and you can factor out that term. Keep in mind that not all polynomials have rational roots. Some roots may be irrational (like $\sqrt{2}$) or complex (involving i)

$f(1) = 1^3 - 6(1)^2 + 11(1) - 6 = 0$ so now we now that $(x - 1)$ is a factor
we can now reduce the polynomial $\frac{x^3 - 6x^2 + 11x - 6}{(x-1)}$ and now we know that $x^3 - 6x^2 + 11x - 6 = (x - 1)(x^2 - 5x + 6)$ as you can see the solutions are $(x - 1)(x - 2)(x - 3)$

complex numbers $\sqrt{-1} = i$

a complex number is made up of a real part (a) and an imaginary part (b) like so: $z = a + bi$ a complex conjugate would be written as so: $\bar{z} = a - bi$...notice that $z + \bar{z} = 2Re(z)$ you can use the complex conguate to help in divition problems with complex numbers like so:

1. $\frac{1 + 2i}{4 - 5i}$

2. $\frac{1 + 2i}{4 - 5i} \frac{4 + 5i}{4 + 5i}$

3. $\frac{4 + 5i + 8i - 10}{16 - 20i + 20i - 25i^2}$

4. $\frac{-6 + 13i}{41}$

5. $\frac{-6}{41} + \frac{13i}{41}$

note that multiplying a complex number by its complex conjugate will give you a real number: $z \cdot \bar{z} = (a + bi)(a - bi) = (a^2) - (bi^2) = a^2 + b^2 = (|z|)^2$

difference of squares:

- $x^2 - y^2$

- $(x + y)(x - y)$

- $x^2 + xy - xy - y^2$

sum of squares:

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

2. $x^2 - i^2 y^2 = x^2 - (iy)^2$

3. $(x + iy)(x - iy)$

example:

1. $36a^8 + 2b^6$

2. $(6a^4)^2 + (\sqrt{2}b^3)^2$

3. $(6a^4)^2 - (-1)(\sqrt{2}b^3)^2$

4. $(6a^4)^2 - (i\sqrt{2}b^3)^2$

5. $(6a^4 + i\sqrt{2}b^3)(6a^4 - i\sqrt{2}b^3)$

rectangular form: $z = a + bi$

- $\cos(\theta) = \frac{a}{r}$ and $\sin(\theta) = \frac{b}{r}$

- $r\cos(\theta) = a$ and $r\sin(\theta) = b$

polar form: $z = r(\cos(\theta) + i\sin(\theta))$

- $r = \bar{z} = \sqrt{a^2 + b^2}$

- $\theta = \tan^{-1}(\frac{b}{a})$

eulers form: $re^{i\theta}$

given $z = -1 + i\sqrt{3}$ find z^4 in both polar and rectangular form

1. $|z| = \sqrt{((-1)^2 + (\sqrt{(3)})^2)} = \sqrt{1 + 3} = 2$ (notice how here we take the principal square root because magnitude (distance) has no direction)

2. $\theta = \tan^{-1}(\frac{\frac{\sqrt{3}}{2}}{\frac{-1}{2}}) = -60°$ (calculator will give you angles between $\frac{-\pi}{2}$ and $\frac{\pi}{2}$ because that is the period of arctan)

3. the vector is in Q2 so $-60° + 180° = 120°$

4. $z = 2(\cos(120°) + i\sin(120°))$

5. $z^2 = 4(\cos(240°) + i\sin(240°))$

6. $z^3 = 8(\cos(360°) + i\sin(350°))$

7. $z^4 = 16(\cos(120°) + i\sin(120°))$

8. $z^4 = 16(\frac{-1}{2}) + 16(\frac{\sqrt{3}}{2})i = -8 + 8\sqrt{3}i$

so there are three cube roots of one in the complex plane:

- $x^3 = 1 \rightarrow x^3 - 1 = 0$

- $1 = 1 + 0i \rightarrow 1 = 1e^{2\pi ni}$

- $x^3 = 1 \rightarrow x^3 = e^{2\pi ni}$

- $x = 1^{\frac{1}{3}} \rightarrow x = e^{\frac{2\pi n}{3}i}$

fundamental theorem of algebra:

Every non-zero polynomial equation ($f(x) \neq 0$) of degree n has exactly n complex roots including multiplicities.

$p(x) = ax^n = bx^{n-1} + \ldots + k$ (n complex roots)

conic sections:

Conic sections come from taking a three dimensional right circular double-napped cone and slicing it with a two dimensional plane. Right means the axis is perpendicular to the base. Circular means the base is a circle. Double-napped means the cone actually consists of two identical cones joined vertex to vertex.

A circle is the set of all points in a plane that are equidistant (radius) from a fixed point (center). $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \rightarrow r = \sqrt{(x - h)^2 + (y - k)^2} \rightarrow (x - h)^2 + (y - k)^2 = r^2$

$x^2 + 4x + y^2 - 6y = 23$:

1. $x^2 + 4x + y^2 - 6y = 23$

2. $x^2 + 4x + \frac{16}{4} - \frac{16}{4} + y^2 - 6y + \frac{36}{4} - \frac{36}{4} = 23$

3. $(x + 2)^2 - 4 + (y - 3)^2 - 9 = 23$

4. $(x + 2)^2 + (y - 3)^2 = 36$

An ellipse is the set of all point surrounding two foci such that $r_1 + r_2 = $ constant. The midpoint of the line segment connecting the two foci is the center. Ellipses have a major and minor axis of symmetry. At each end of the major axis we can find verticies.

$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$ (vertices: $(h \pm a, k)$)
$\frac{(x-h)^2}{b^2} + \frac{(y-k)^2}{a^2} = 1$ (vertices: $(h, k \pm a)$)

The major and minor axis has a length of 2a and 2b repectively. The distance from the center to the vertices are a, and the distance from the center to the covertices are b. The distance from the center to either foci is $a^2 = b^2 + c^2 \rightarrow c^2 = a^2 - b^2$
A parabola are the set of all point that are equidistant from a fixed point (the focus) and a fixed line (the directrix).

$f(x) = -x^2 - 2x + 1$:

1. $-(x^2 + 2x - 1)$

2. $-(x^2 + 2x + \frac{4}{4} - \frac{4}{4} - 1)$

3. $-((x + \frac{2}{2})^2 - 2)$

4. $-(x + 1)^2 + 2$

The definition of a hyperbola is similar to that of an ellipse, where there are two foci. While every point on a ellipse has distances to these foci that give a constant sum. Every point on a hyperbola has distances to these foci that give a constant difference. The foci sit on the outside of the branches with the two vertices between them. In

between those vertices we have the center which sits on the transverse axis connecting the vertices. The a, b, and c terms from the ellipse will pretty much mean the same things. a is the distance from the center to the vertex. c is the distance from the center to a focus. b is the distance from the center to the co-vertex. The co-vertex does not lie on the hyperbola like it does for an ellipse. However, b helps define the asymptotes, which guide the shape of the branches of the hyperbola. If you form a rectangle centered at (h, k) with dimensions: width $= 2a$ and height $= 2b$. The asymptotes pass through the diagonals of this rectangle.

let us find the asymptotes for the horizontal hyperbola:

1. $\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{b^2} = 1$

2. lets solve for y to understand the graphs shape $-\frac{(y-k)^2}{b^2} = -\frac{(x-h)^2}{a^2} + 1$

3. $\frac{(y-k)^2}{b^2} = \frac{(x-h)^2}{a^2} - 1$

4. $(y-k)^2 = b^2(\frac{(x-h)^2}{a^2} - 1)$

5. $y - k = \pm b\sqrt{\frac{(x-h)^2}{a^2} - 1}$ consider $x \to \infty$

6. $\sqrt{\frac{(x-h)^2}{a^2} - 1} \approx \sqrt{\frac{(x-h)^2}{a^2}} = \frac{|x-h|}{a} = \frac{x-h}{a}$

7. as $x \to \infty$ then $y \approx \pm b \cdot \frac{x-h}{a} \Rightarrow y \approx k \pm \frac{b}{a}(x-h)$

the discriminant is tradtionally introduced in the context of solving quadratic equations, but it is also very powerful in analyzing conic sections. let us begin with the general second-degree equation in two variables: $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$ this equation represents a conic section, which could be a parabola, ellipse, hyperbola, or (in degenerate cases) a point, line, ect. the discriminant used here is: $\Delta = B^2 - 4AC$. this is not the same as the quadratic discriminant: $b^2 - 4ac$ for single-variable quadratics. it is a different discriminant, but has an analogous role. depending on the value of $\Delta = B^2 - 4AC$, the conic is:

- ellipse (or circle if $A = C$ and $B = 0$ if $\Delta < 0$

- parabola if $\Delta = 0$

- hyperbola if $\Delta > 0$

it all comes down to the quadratic part of the conic: $Ax^2 + Bxy + Cy^2$ this defines the shape of the conic, while the $Dx + Ey + F$ part just translates it. you can think of this as a sort of matrix

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} A & \frac{B}{2} \\ \frac{B}{2} & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

this is a quadratic form, and the matrix's properties determine the geometry. we can rotate the coordinate system (diagonalize the matrix) to eliminate the cross-term Bxy. after rotation, the conic becomes: $A'x'^2 + C'y'^2 + \text{linear terms} = 0$ now the classification is straightforward:

- if A' and C' has the same sign $\to$ ellipse

- if one is zero and the other non-zero $\to$ parabola

- if they have opposite signs $\rightarrow$ hyperbola

and it turns out: $\Delta = B^2 - 4AC$ captures exactly that information about the signs of the eigenvalues of the matrix. so:

- $\Delta < 0 \Rightarrow$ same signs $\Rightarrow$ ellipse

- $\Delta = 0 \Rightarrow$ one zero eigenvalue $\Rightarrow$ parabola

- $\Delta > 0 \Rightarrow$ opposite signs $\Rightarrow$ hyperbola

basic classes of functions:

- polynomials: a sum of terms, where each term is made up of coefficient (a constant multiple) of a power function with a whole number exponent

- rational functions: a quotient of two polynomials

  - vertical asymptotes: vertical lines $x = a$ where the function grows without bound - usually where the denominator is zero and the numerator is not zero

  - removable discontinuity: points where the function is not defined due to a factor that cancels out from both the numerator and denominator

  - hotizontal asymptotes: horizontal lines $y = L$ where the function approaches as $x \rightarrow \infty$

    * deg P < deg Q ($y = 0$)
    * deg P = deg Q ($y = \frac{\text{leading coefficient of P}}{\text{leading coefficient of Q}}$)
    * deg P > deq Q (no horizontal asymptote look for slant instead)

  - slant asymptote: lines $y = mx + b$ that the function approaces as $x \rightarrow \infty$, when the numerator's degree is exactly one more than the denominator's degree...to find them perform polynomial long division and the quotient is the slant asymptotoe

- algebraic functions: produced by taking sums, products and quotients of roots or polynomials and rational functions

- exponential functions: $f(x) = b^x$ where $b > 0$...the inverse of which is $f(x) = log_b x$

- trigonometric functions: built from $\sin(x)$ and $\cos(x)$ are called trigonometric functions.

constructing new functions:
If f and g are functions, we may construct new functions by forming the sum, difference, product, and quotient functions: $(f + g)(x) = f(x) + g(x)$, $(f - g)(x) = f(x) - g(x)$, $(fg)(x) = f(x)g(x)$, $(\frac{f}{g})(x) = \frac{f(x)}{g(x)}$. We can also multiply functions by constants. We call this a linear combination: $c_1 f(x) + c_2 g(x)$. Composition is another important way of constructing new functions. The composition of f and g is the function $f \circ g$ defined by $(f \circ g)(x) = f(g(x))$.

invertable functions:
A function f is invertible if there exists another function $f^{-1}(x)$ such that: $f^{-1}(f(x)) = x$ and $f(f^{-1}(x)) = x$. This "inverse function" switches input and outputs - it reverses the effect of the original function.

A function is invertible if it is one-to-one (horizontal line test). So, this means that different input always produce different outputs.

special factorizations

- $x^2 - y^2 = (x + y)(x - y)$
- $x^3 + y^3 = (x + y)(x^2 - xy + y^2)$
- $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$

binomial

- $(x + y)^2 = x^2 + 2xy + y^2$
- $(x - y)^2 = x^2 - 2xy + y^2$
- $(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$
- $(x - y)^3 = x^3 - 3x^2y + 3xy^2 - y^3$
- $(x + y)^n = \sum_{k=0}^{n} = \binom{n}{k} x^{(n-k)} y^k$

scalars, vectors, and matrices

scalar multiplication:
$\vec{w} = (1, 2)$
$3\vec{w} = (3, 6)$

component form: $\vec{v} = (x, y)$
polar form: $\vec{v} = (r\cos(\theta), r\sin(\theta)) = r(\cos(\theta), \sin(\theta))$

When you graphically add two vectors you add them tail to head. Vector addition is commutative ($\vec{a} + \vec{b} = \vec{b} + \vec{a}$). Preforming both of these additions graphically will form a parallelogram and the vector that bisects this parallelogram is the solution.

A system of equations is a set of two or more equations that share the same variables. The goal is to find values for those variable that make all the equations in the system true at the same time. For example, in a system with two equations and two variables (like x and y), you are looking for a point (x, y) that satisfies both equations.

To solve a system of equations using substitution, you start by solving one of the equations for one variable in terms of the other. Then, you substitute that expression into one of the other equations. This gives you an equation with just one variable (assuming you started with two), which you can solve. Once you find that value, you plug it back into one of the original equations to find the second variable.

With elimination, the goal is to eliminate one variable by adding or subtracting the equations. You may need to multiply one or both equations first to make the coefficients of a variable match. Once a variable is eliminated, you solve the resulting equation for the remaining variable, then substitute that value back into one of the original equations to find the other variable.

You can only add or subtract matrices if they have the same number of rows and the same number of columns. This is because matrix addition and subtraction is an element-wise operation.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 8 & 9 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 8 & 10 & 12 \\ 4 & 6 & 6 \end{bmatrix}$$

$$A = \begin{bmatrix} -2 & 5 & 6 \\ 5 & 2 & 7 \end{bmatrix}$$

The dimensions of a matrix give the number of rows and columns of the matrix in that order. Since matrix A has 2 rows and 3 columns, it is called a 2×3 matrix.

A zero matrix is a matrix in which all of the entries are 0.

$$O_{3\times3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Zero matrices play a similar role in operations with matrices as the number zero plays in operations with real numbers. When we add an $m \times n$ zero matrix to any $m \times n$ matrix A, we get matrix A back. In other words, $A + O = A$ and $O + A = A$.

$$\begin{bmatrix} 4 & 1 \\ -6 & 2 \end{bmatrix} + \begin{bmatrix} -4 & -1 \\ 6 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

When we add any $m \times n$ matrix to its opposite, we get the $m \times n$ zero matrix. $A + (-A) = O$ and $-A + A = O$ It is also true that $A - A = O$. This is because subtracting a matrix is like adding its opposite.

matrix addition properties (A, B, and C are equal dimensions):

- commutative property of addition: $A + B = B + A$

- associative property of addition: $A + (B + C) = (A + B) + C$

- additive identity: For any matrix A, there is a unique matrix O such that $A + O = A$

- additive inverse property: For each A, there is a unique matrix $-A$ such that $A + (-A) = O$

- closure property of addition: $A + B$ is a matrix of the same dimensions A and B

matrix scalar multiplication properties (A and B are matrices of equal dimensions, c and d are scalars, and O is a zero matrix):

- associative property of multiplication: $(cd)A = c(dA)$

- distributive properties: $c(A + B) = cA + cB$

- multiplicative identity property: $1A = A$

- multiplicative properties of zero: $0 \cdot A = O$ and $c \cdot O = O$

- closure property of multiplication: cA is a matrix of the same dimensions as A

Matrices can be thought of as transformations of the plane, meaning they take points (or vectors) in a 2D space and move or change them in specific ways. Imagine every point in the plane as a vector $\begin{bmatrix} x \\ y \end{bmatrix}$. When you multiply that vector by a 2×2 matrix, the output is a new vector $\begin{bmatrix} x' \\ y' \end{bmatrix}$ that represents the transformed point.

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

the new vector would be $M \cdot v = v'$ or you could do $v_{\text{row}} M^T = v'$

What kinds of transformations can matrices do?

- scaling: stretching or shrinking the plane along the x- and/or y-axis

- rotation: turning all points around the origin by some angle

- reflection: flipping points over a line, like the x-axis or y-axis

- shearing: slanting the shape, pushing points sideways or vertically

- projection: flattening points onto a line or plane

A column vector and a row vector can contain the same numbers, but their orientation matters a lot, especially when it comes to matrix operations like multiplication.

$$\begin{pmatrix} 2 & 5 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix} = (2)(4) + (5)(0) + (-1)(3) = 8 + 0 - 3 = 5$$

$$\begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix} \begin{pmatrix} 2 & 5 & -1 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 6 \\ 20 & 0 & 15 \\ -4 & 0 & -3 \end{pmatrix}$$

The determinant of a square matrix is a special number that represents how the matrix transforms space—specifically, it measures how the matrix scales volume (or area in 2D). If the determinant is zero, it means the transformation squashes space into a lower dimension, losing information. A matrix with a zero determinant is called singular, which means it does not have an inverse. This happens when the matrix's columns (or rows) are linearly dependent, meaning one column can be expressed as a combination of others. In systems of equations, a singular (determinant-zero) matrix implies the system is either dependent—where the equations overlap and there are infinitely many solutions—or inconsistent, where the equations contradict and have no solution. A system is consistent if it has at least one solution, whether unique or infinite. Thus, the determinant and singularity give essential insight into whether a system can be solved uniquely, infinitely, or not at all.

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\det(A) = ad - bc$

If $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, then $\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$

What does it mean to compose matrices? It means that you can apply transformations in sequence by multiplying their matrices together.

$T = BA$
$v' = B(Av) = (BA)v$
$f(x) = Ax$
$g(x) = Bx$
$g(f(x)) = B(Ax) = (BA)x$

properties of matrix multiplication (A, B, and C are n x n matrices, I is the n x n identity matrix and O is the n x n zero matrix):

- the commutative property of multiplication does not hold! $AB \neq BA$

- associative property of multiplication: $(AB)C = A(BC)$

- distributive properties: $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$

- multiplicative identity property $IA = A$ and $AI = A$

- multiplicative property of zero: $OA = O$ and $AO = O$

- dimension property: the product of an m x n matrix and an n x k matrix is an m x k matrix

Consider a system of equations like so: $3x - 2y - z = -1$
$2x + 5y + 2 = 0$
$-4x - y = 8$
$\downarrow$

$$\begin{bmatrix} 3 & -2 & -1 \\ 2 & 5 & 1 \\ -4 & -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 8 \end{bmatrix}$$

$\uparrow$ generalized $A_{n\times n}\vec{x}_{n\times 1} = \vec{b}_{n\times 1}$

When the coefficient matrix A is not square, you cannot use the inverse to solve the system because only square matrices can have an inverse, and only non-singular square matrices (with $\det(A) \neq 0$) actually do have an inverse. In an underdetermined system (more variables than equations: $m < n$) you will usually have infintely many solutions, or sometimes none. Here you would use substitution, elimination, parametric solutions, or row-reduction (gaussian elimination). In a overdetermined system (more equations than variables: $m > n$) it might be inconsistent (no solution), or sometimes it has a best-fit solution. Here you would use gaussian elimination, least squares method, or pseudoinverses. There are other methods for square matrices like row-reduction or cramer's rule.

$f^{-1}(f(x)) = x$
$f(f^{-1}(x)) = x$

$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
$B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

let A represent a $90°$ clockwise rotation and b represent a $90°$ counter-clockwise rotation of the unit vectors. Visually speaking, thses two rotation matrices are inverses of each other. Multiplying them in either order gives the identity matrix.

to find the $\det(A)_{n\times n}$ use cofactor expansion or row reduction (gaussian elemination). you will need to understand minors and cofactors to compute the determinant this way.

the adjugate is the transpose of its cofactor matrix $\text{adj}(A) = [C_{ij}]^T$

$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A)$

gaussian elimination can also be used to find the inverse by augmenting the matrix

$[A \mid I] \to [I \mid A^{-1}]$ often faster for large matrices

advanced techniques: LU decomposition/row operations

probability and combinatorics consider a standard deck of cards (no jokers)
$P(\text{jack}) = \frac{4}{52} = \frac{1}{13}$
$P(\text{hearts}) = \frac{13}{52}$

now consider a venn diagram for the following: $P(\text{J} \cup \text{H}) = \frac{4+13-1}{52}$
since the probability of hearts overlaps with the probability of jacks (jack of hearts) youll
have to subtract out the double counting...we say that that are not mutually exclusive
otherwise you could just add $P(\text{J}) + P(\text{H})$ if they were mutually exclusive

addition rule:
$P(\text{A} \cup \text{B}) = P(\text{A}) + P(\text{B}) - P(\text{A} \cap \text{B})$

now consider mutually exclusive events:
$P(\text{A} \cap \text{B}) = 0$
$P(\text{A} \cup \text{B}) = P(\text{A}) + P(\text{B})$

multiplication rule for independent events:
if two events, A and B, are independent (meaning the occurrence of one does not affect
the other), then: $P(\text{A} \cap \text{B}) = P(\text{A}) \cdot P(\text{B})$

consider a fair coin:
$P(\text{back to back heads}) = \frac{1}{2} \cdot \frac{1}{2}$
now consider dependent events say you have a bag with 3 blue and 2 red balls...what is
the probability that the first pull is blue and the second is blue
$P(\text{1st blue}) \cdot P(\text{2nd blue} \mid \text{1st blue})$

$P(\text{A} \cap \text{B}) = P(\text{B}) \cdot P(\text{A} \mid \text{B}) = P(\text{A}) \cdot P(\text{B} \mid \text{A})$

permutations (order matters)
$P(n, k) = \frac{n!}{(n-k)!}$

combinations (order does not matter)
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

a probability distribution describes how likely different outcomes are for a random variable

- a random variable is something that can take on different values due to chance (like rolling a die or counting the number of heads in coin flips)

- a probability distribution assigns a probability to each possible value that the variable can take

the expected value of a random variable is the long-run average outcome you would expect if you repeated an experiment many times. think of it as the center or balance

point of a probability distribution.

$$E[X] = \sum_{i=1}^{n} x^i \cdot P(x_i)$$

series

A sequence is an ordered list of numbers following some rule or pattern. You can get each term from the previous one by repeated addition (arithmetic sequence) or repeated multiplication (geometric sequence), or some other rule.

examples of other rules:

- fibonacci: each term is the sum of the previous two terms: $F_n = F_{n-1} + F_{n-2}$
  so the sequence starts: $0, 1, 1, 2, 3, 5, 8, 13, \ldots$

- quadtratic:
  terms might increase by a pattern that fits a quadratic formula, like n^2:
  $1, 4, 9, 16, 25 \ldots$

- alternating:
  terms might alternate between positive and negative, like:
  $1, -1, 1, -1, 1, -1, \ldots$

- recursive:
  terms can be defined using more complex rules involoving previous terms

$a, ar, ar^2, ar^3, \ldots, ar^{n-1}$
$S_n = a + ar + ar^2 + \ldots + ar^{n-1}$

The first term is an expression where the variable r is raised to the zeroth power, so the nth term is an expression that has r raised to the n-1 power.

sum of a finite geometric series $(r \neq 1)$:

1. $S_n = a + ar + ar^2 + \ldots + ar^{n-1}$

2. $rS_n = ar + ar^2 + ar^3 + \ldots + ar^n$

3. $S_n - rS_n = (a + ar + ar^2 + \ldots + ar^{n-1}) - (ar + ar^2 + ar^3 + \ldots + ar^n)$

4. $S_n - rS_n = a - ar^n$

5. $S_n(1 - r) = a(1 - r^n)$

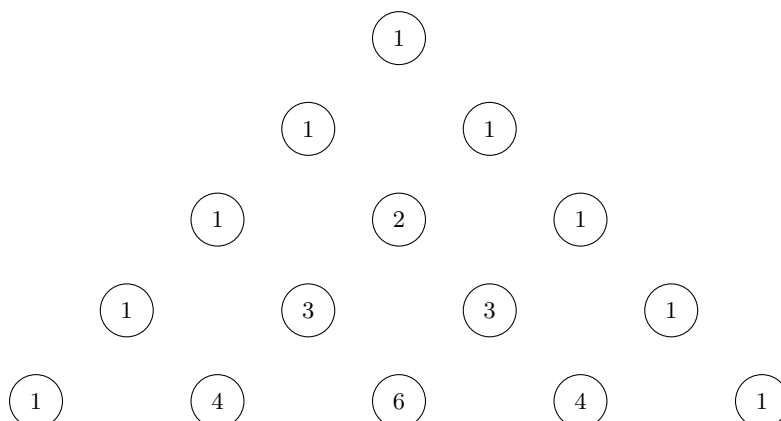6. $S_n = a\frac{1-r^n}{1-r}$ if $r \neq 1$

$$\sum_{k=1}^{n} ar^{k-1} = \frac{a(1-r^n)}{1-r}$$

The binomial theorem gives a formula for expanding powers of a binomial

$$(a + b)^n = \sum_{k=0}^{n} = \binom{n}{k} a^{(n-k)} b^k$$

Each term in the expansion represents on way to choose how many a's and how many b's you take from those n factors

Pascal's triangle gives you the binomial coeffiecients used in binomial expansion

$$
\begin{array}{ccccccccc}
& & & & 1 & & & & \\
& & & 1 & & 1 & & & \\
& & 1 & & 2 & & 1 & & \\
& 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}
$$

$(a+b)^4 = 1a^4b^0 + 4a^3b^1 + 6a^2b^2 + 4a^1b^3 + 1a^0b^4$
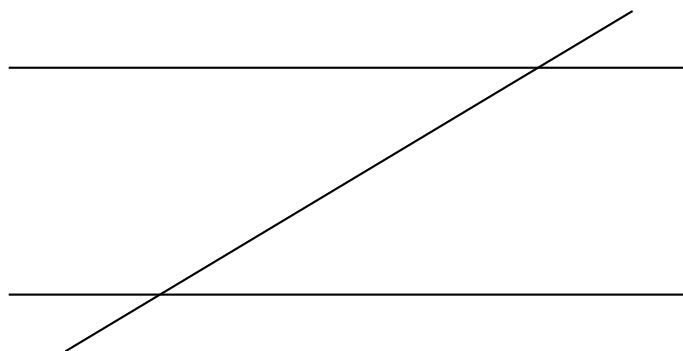
Each number $\binom{n}{k}$ counts the number of ways to chose k b's (and thus n-k a's) from n total factors of (a+b)

$a, a+d, a+2d, \ldots, a+(n-1)d$

sum of a finite arithmetic sequence:

1. $S_n = a + (a+d) + (a+2d) + \ldots + [a+(n-1)d]$
2. $S_n = [a+(n-1)d] + [a+(n-2)d] + [a+(n-3)d] + \ldots + a$
3. $2S_n = n[2a + (n-1)d]$
4. $S_n = \frac{n}{2}[2a + (n-1)d]$

$\sum_{k=1}^{n}[a+(k-1)d] = \frac{n}{2}[2a+(n-1)d]$

angle relationships formed when a transversal intersects parallel lines:

- complementary angles: two angles whose measures add up to 90°

- supplementary angles: two angles whose measures add up to 180°

- vertical angles: pairs of opposite angles formed when two lines intersect. they are always equal in measure.

- corresponding angles: when a transversal crosses two lines, corresponding angles are pairs located on the same side of the transversal and in matching corners (one interior, one exterior). If the lines are parallel, corresponding angles are equal.

- alternate interior angles: angles located between the two lines but on opposite sides of the transversal. if the lines are parallel, these angles are equal.

- alternate exterior angles: angles located outside the two lines and on opposite sides of the transversal. if the lines are parallel, these angles are equal.

- same-side (consecutive) interior angles: angles that lie between the two lines and on the same side of the transversal. if the lines are parallel, these angles are supplementary (add up to 180°)

circle angle theorems:

- inscribed angles subtended by the same arch are equal

- an angle insribed in a semi-circle equals 90°

- the angle subtendted by an arc at the center is double the angle subtended by the same arc at any point on the circumference
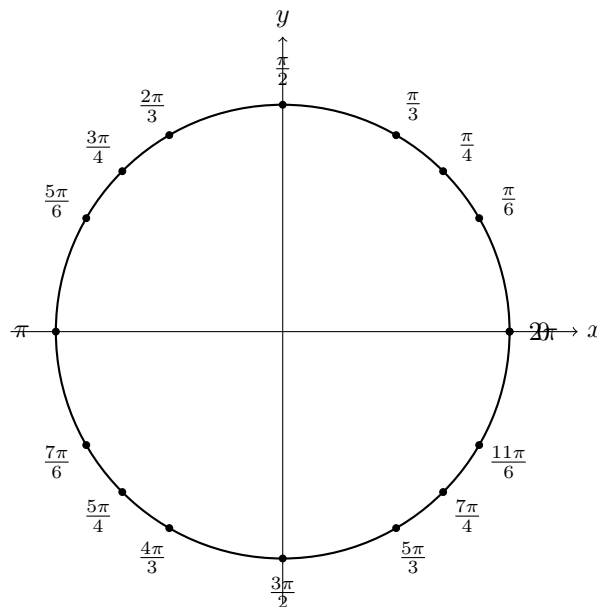
One radian is the angle made at the center of a circle when the arc length is equal to the radius of the circle.

θ radians $= \frac{s}{r} \Rightarrow s = r\theta$

$\pi = \frac{\text{circumference}}{\text{diameter}}$

$1° = \frac{\pi}{180}$ radians

Trigonometric functions are special mathematical functions that originally come from studying right triangles. They relate the size of an angle in a triangle to the ratios of the lengths of the triangle's sides.

SOH-CAH-TOA is a mnemonic device that expresses the relationship between the basic trigonometric functions and the ratios of the sides in a right triangle.

The triangle definition only works for angles between 0 and $\frac{\pi}{2}$. To extend trig functions to all angles (including negative angles and angles larger then (2π or $360°$), mathematicians use the unit circle. Thus allowing use to use trig functions on the coordinate plane, enabling graphing and calculus.

trigonometric identities:

- $\frac{1}{\cos(\theta)} = \sec(\theta)$

- $\frac{1}{\sin(\theta)} = \csc(\theta)$

- $\frac{1}{\tan(\theta)} = \cot(\theta)$

- $\sin(\theta + 2\pi) = \sin(\theta)$

- $\cos(\theta + 2\pi) = \cos(\theta)$

- $\tan(\theta + \pi) = \tan(\theta)$

- $\cos(\frac{\pi}{2} - \theta) = \sin(\theta)$

- $\sin(\frac{\pi}{2} - \theta) = \cos(\theta)$

- $\tan(\frac{\pi}{2} - \theta) = \cot(\theta)$

- $\sin(-\theta) = -\sin(\theta)$

- $\cos(-\theta) = \cos(\theta)$

- $\tan(\theta) = -\tan$

- $\sin^2(\theta) + \cos^2(\theta) = 1$

- $1 + \tan^2(\theta) = \sec^2(\theta)$

- $1 + \cot^2(\theta) = \csc^2(\theta)$

- $\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$

- $\sin(A \pm B) = \sin(A)\cos(B) \pm \cos(A)\sin(B)$

- $\tan(A \pm B) = \frac{\tan(A)\pm\tan(B)}{1\mp\tan(A)\tan(B)}$

- $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$

- $\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta)$

- $\cos(2\theta) = 2\cos^2(\theta) - 1$

- $\cos(2\theta) = 1 - 2\sin^2(\theta)$

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

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

- $\tan\left(\frac{x}{2}\right) = \pm\sqrt{\frac{1-\cos x}{1+\cos x}}$

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

- $\tan\left(\frac{x}{2}\right) = \frac{1 - \cos x}{\sin x}$

- $\cos A \cos B = \frac{1}{2}\left[\cos(A - B) + \cos(A + B)\right]$

- $\sin A \sin B = \frac{1}{2}\left[\cos(A - B) - \cos(A + B)\right]$

- $\sin A \cos B = \frac{1}{2}\left[\sin(A + B) + \sin(A - B)\right]$

- $\cos A \sin B = \frac{1}{2}\left[\sin(A + B) - \sin(A - B)\right]$

law of sines and cosines

$\frac{\sin(A)}{a} = \frac{\sin(B)}{b} \frac{\sin(C)}{c}$

$a^2 = b^2 + c^2 - 2bc\cos(A)$ (SAS or SSS)

Limits

An indeterminate form is a mathematical expression that arises in limits where the limit cannot be directly determined from the form itself because it is ambiguous or undefined in a straightforward way.

indeterminate forms:

- $\frac{0}{0}$

- $\frac{\infty}{\infty}$

- $0 \times \infty$

- $\infty - \infty$

- 0^0

- ∞^0

types of discontinuities:

- removeable: $f(x) = \frac{(x^2 - 1)}{x - 1}$

- jump: the left-hand and right-hand limit at the point exist but are not equal

- infinite (essential): $f(x) = \frac{1}{x}$

- oscillatory: $f(x) = \sin(\frac{1}{x})$

How do the values of a function behave when x approaches a number c, whether or not the function at c is defined?
$f(x) = \frac{\sin(x)}{x}$ $f(0) = \frac{\sin(0)}{0} = \frac{0}{0}$
Using a calculator you can numerically see that as $x \to 0+$ and $x \to 0-$ the function appears to approach one.

limit laws assume that $\lim_{x \to c} f(x)$ and $\lim_{x \to c} g(x)$ exist, then:

- sum law: $\lim x \to c(f(x) + g(x)) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$

- constant multiple law: for any number k, $\lim_{x \to c} kf(x) = k\lim_{x \to c} f(x)$

- product law: $\lim_{x \to c} f(x)g(x) = (\lim_{x \to c} f(x))(\lim_{x \to c} g(x))$

- quotient law: if $\lim_{x\to c} g(x) \neq 0$, then $\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$

continuity at a point: $\lim x \to c f(x) = f(c)$

laws of continuity assume that $f(x)$ and $g(x)$ are continuous at a point $x = c$. then the following functions area lso continuous at $x = c$:

- $f(x) + g(x)$ and $f(x) - g(x)$

- $kf(x)$ for any constant k

- $f(x)g(x)$

- $\frac{f(x)}{g(x)}$ if $g(c) \neq 0$

continuity of composite functions let $F(x) = f(g(x))$ be a composite function. If g is continuous at $x = c$ and f is continuous at $x = g(c)$, then $F(x)$ is continuous at $x = c$.

squeeze theorem assume that for $x \neq c$ (in some open interval containing c), $l(x) \leq f(x) \leq u(x)$ and $\lim_{x\to c} l(x) = \lim_{x\to c} u(x) = L$ then $\lim_{x\to c} f(x)$ exists and $\lim_{x\to c} f(x) = L$

important trigonometric limits:

- $\lim_{\theta\to 0} \frac{\sin(\theta)}{\theta} = 1$

- $\lim_{\theta\to 0} \frac{1-\cos(\theta)}{\theta} = 0$

intermediate value theorem: if $f(x)$ is continuous on a closed interval $[a, b]$ and $f(a) \neq f(b)$, then for every value M between $f(a)$ and $f(b)$, there exists at least one value $c \in (a, b)$ such that $f(c) = M$.

existence of zeros: if $f(x)$ is continuous on $[a, b]$ and if $f(a)$ and $f(b)$ are nonzero and have opposite signs, then $f(x)$ has a zero in (a, b).

we can locate zeroes of functions to arbitrary accuracy using the **bisection method**.
$f(x) = \cos^2(x) - 2\sin(\frac{x}{4})$
$f(0) = 1 > 0$, $f(2) \approx -0.786 < 0$
we can guarantee that $f(x) = 0$ has a solution in $(0, 2)$. we can locate a zero more accurately by dividing $[0, 2]$ into two intervals $[0, 1]$ and $[1, 2]$. one of these must contain a zero of $f(x)$. to determine which, we evalutate $f(x)$ at the midpoint $m = 1$. a calculator gives $f(1) \approx -0.203 < 0$, and since $f(0) = 1$, we see that $f(x)$ takes opposite signs at the endpoints of $[0, 1]$. therefore, $(0, 1)$ must contain a zero. we discard $[1, 2]$ because both $f(1)$ and $f(2)$ are negative. the bisection method consists of continuing this process until we narrow down the location of the zero to the desired accuracy.

the size of the gap:
recall that the distance from $f(x)$ to L is $|f(x) - L|$. it is convenient to refer to the quantity $|f(x) - L|$ as the gap between the value $f(x)$ and the limit L. lets reexamine the basic trigonometric limit $\lim_{x\to 0} \frac{\sin(x)}{x}$. so 1 tells us that the gap $|f(x) - 1|$ gets arbitrarily small when x is sufficiently close by not equal to 0. suppose we want the gap $|f(x) - 1|$ to be less than 0.2. how close to 0 must x be? $|f(x) - 1| < 0.2$ if $0 < |x| < 1$. if we insist instead that the gap be smaller than 0.004, we can check by zooming in $|f(x) - 1| < 0.004$ if $0 < |x| < 0.15$. it would seem that this process can be continued: by zooming in on the graph, we can find a small interval around $c = 0$ where

the gap $|f(x) - 1|$ is smaller than any prescribed positive number. to express this in a precise fasion, we follow time-honored tradition and use the greek letters ϵ (epsilon) and δ (delta to denote small numbers specifying the size of the gap and the quantity $|x - c|$, respectively. in our case, $c = 0$ and $|x - c| = |x - 0| = |x|$. The precise meaning is that for every choice of $\epsilon > 0$, there exists some δ (depending on ϵ) such that $|\frac{\sin(x)}{x} - 1| < \epsilon$ if $0 < |x| < \delta$. the number δ tells us how close is sufficiently close for a give ϵ.

formal definition of a limit:
suppose that $f(x)$ is defined for all x in an open interval containing c (but not necessarily at $x = c$). then

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

if for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - L| < \epsilon \text{ if } 0 < |x - c| < \delta$$

four rigorous definitions of the limit:

- $\lim_{x \to a} f(x) = L$ means $\forall \epsilon > 0, \exists \delta > 0 : 0 < |x - a| < \delta \Rightarrow |f(x) - L| < \epsilon$

- $\lim_{x \to \infty} f(x) = L$ means $\forall \epsilon > 0, \exists N > 0 : x > N \Rightarrow |f(x) - L| < \epsilon$

- $\lim_{x \to a} f(x) = \infty$ means $\forall M > 0, \exists \delta > 0 : 0 < |x - a| < \delta \Rightarrow f(x) > M$

- $\lim_{x \to \infty} f(x) = \infty$ means $\forall M > 0, \exists N > 0 : x > N \Rightarrow f(x) > M$

Differentiation

the derivative of a function f at $x = a$ is the limit of the difference quotients (if it exists):

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

when the limit exists, we say that f is differentiable at $x = a$. an equivalent definition of the derivative is

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

tangent lines:
assume that $f(x)$ is differentiable at $x = a$. the tangent line to the graph of $y = f(x)$ at $P = (a, f(a))$ is the line through P of slope $f'(a)$. the equation of the tangent line in point-slope form is

$$y - f(a) = f'(a)(x - a)$$

derivative of linear and constant functions:

- if $f(x) = mx + b$ is a linear function, then $f'(a) = m$ for all a

- if $f(x) = b$ is a constant function, then $f'(a) = 0$ for all a

proof of the power rule

1. assume that n is a whole number and let $f(x) = x^n$. then $f'(a) = \lim_{x \to a} \frac{x^n - a^n}{x - a}$

2. to simplify the difference quotient, we need to generalize the following identities:
   $x^2 - a^2 = (x - a)(x + a)$
   $x^3 - a^3 = (x - a)(x^2 + xa + a^2)$
   $x^4 - a^4 = (x - a)(x^3 + x^2a + xa^2 + a^3)$

3. the generalizaiton is $x^n - a^n = (x-a)(x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \ldots + xa^{n-2} + a^{n-1})$

4. to verify, observe that the right-hand side is equal to
$x(x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \ldots + xa^{n-2} + a^{n-1})$
$-a(x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \ldots + xa^{n-2} + a^{n-1})$

5. when we carry out the multiplications, al terms cancle except the first and the last, so $x^n - a^n$ remains, as required. the identity gives us
$\frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \ldots + xa^{n-2} + a^{n-1} (x \neq a)$

6. therefore,
$f'(a) = \lim_{x \to a}(x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \ldots + xa^{n-2} + a^{n-1})$
$a^{n-1} + a^{n-2}a + a^{n-3}a^2 + \ldots + aa^{n-2} + a^{n-1}$(n terms)
na^{n-1}

linearity rules assume that f and g are differentiable functions

- sum rule: the function $f + g$ is differentiable and $(f + h)' = f' + g'$

- constant multiple rule: for any constant c, cf is differentiable and $(cf)' = cf'$

differentiability implies continuity if f is differentiable at $x = c$, then f is continuous at $x = c$.
Proof:

1. by definition, if f is differentiable at $x = c$, then the following limit exists: $f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$

2. our goal is to prove that f is continuous at $x = c$, which means that $\lim_{x \to c} f(x) = f(c)$. To relate the two limits, consider the equation (valid for $x \neq c$) $f(x) - f(c) = (x - c)\frac{f(x) - f(c)}{x - c}$

3. both factors on the right approach a limit as $x \to c$, so we may apply the limit laws: $\lim_{x \to c}(f(x) - f(c)) = \lim_{x \to c}((x - c)\frac{f(x) - f(c)}{x - c})$

4. $\lim_{x \to c}(f(x) - f(c)) = (\lim_{x \to c}(x - c)(\lim_{x \to c}\frac{f(x) - f(c)}{x - c})$

5. $\lim_{x \to c}(f(x) - f(c)) = 0 \cdot f'(c) = 0$

6. therefore, $\lim_{x \to c} f(x) = \lim_{x \to c}(f(x) - f(c)) + \lim_{x \to c} f(c) = 0 + f(c) = f(c)$, as desired

proof of the product rule

1. we prove the product rule by writing the difference quotient for $f(x)g(x)$ in a vlever way as a sum of two terms. the limit definition of the derivative applied to the product function gives us
$(fg)'(x) = \lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}$

2. the trick is to subtract $f(x+h)g(x)$ and add it back again in the numerator of the difference quotient:
$f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)$

3. combing terms, we see that the numerator is equal to
$f(x+h)(g(x+h) - g(x)) + g(x)(f(x+h) - f(x))$

4. now write $(fg)'(x)$ as a sum of two limits:
$(fg)'(x) = \lim_{h \to 0} f(x+h)\frac{g(x+h) - g(x)}{h} + \lim_{h \to 0} g(x)\frac{f(x+h) - f(x)}{h}$

5. the use of the sum law is valid, provided that each limit on the right exists. to check that the first limit exist and evaluate it, we note that $f(x)$ is continuous (because it is differentiable) and that $g(x)$ is differentiable:
$\lim_{h \to 0} f(x+h) \frac{g(x+h)-g(x)}{h} = \lim_{h \to 0} f(x+h) \lim_{h \to 0} \frac{g(x+h)-g(x)}{h} = f(x)g'(x)$

6. the second limit is similar:
$\lim_{h \to 0} g(x) \frac{f(x+h)-f(x)}{h} = g(x) \lim_{h \to 0} \frac{f(x+h)-f(x)}{g} = g(x)f'(x)$

7. using the aforementioned we conclude that fg is differentiable and that $(fg)'(x) = f(x)g'(x) + g(x)f'(x)$ as desired.

proof of the quotient rule

1. start with the definition of the derivative
$(\frac{f}{g})'(x) = \lim_{h \to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h}$

2. combine the two fractions in the numerator
$(\frac{f}{g})'(x) = \lim_{h \to 0} \frac{1}{h} \cdot \frac{f(x+h)g(x)-f(x)g(x+h)}{g(x+h)g(x)}$

3. manipulate the numerator
$f(x+h)g(x) - f(x)g(x+h) = [f(x+h)g(x) - f(x)g(x)] - [f(x)g(x+h) - f(x)g(x)] = g(x)[f(x+h)-f(x)] - f(x)[g(x+h)-g(x)]$

4. lets put that back into the limit
$(\frac{f}{g})'(x) = \lim_{h \to 0} \frac{g(x)[f(x+h)-f(x)] - f(x)[g(x+h)-g(x)]}{h \cdot g(x+h)g(x)}$

5. split the limit
$\lim_{h \to 0} \left[\frac{g(x)[f(x+h)-f(x)]}{h \cdot g(x+h)g(x)} - \frac{f(x)[g(x+h)-g(x)]}{h \cdot g(x+h)g(x)} \right]$

6. $(\frac{f}{g})'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

proof of the derivative of sin

1. $\sin(x+h) = \sin(x)\cos(h) + \cos(x)\sin(h)$

2. $\sin(x+h) - \sin(x) = \sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)$

3. $\sin(x+h) - \sin(x) = (\sin(x)\cos(h) - \sin(x)) + \cos(x)\sin(h)$

4. $\sin(x+h) - \sin(x) = \sin(x)(\cos(h) - 1) + \cos(x)\sin(h)$

5. $\lim_{h \to 0} \frac{\sin(x)(\cos(h)-1)}{h} + \lim_{h \to 0} \frac{\cos(x)\sin(h)}{h}$

6. $\sin(x)(0) + \cos(x)(1)$

proof of the chain rule

1. $h(x) = f(g(x))$

2. $h'(x) = \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{h}$

3. the key idea is to manipulate the difference quotient to introduce $g(x+h) - g(x)$
$\frac{f(g(x+h)) - f(g(x))}{h} = \frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \cdot \frac{g(x+h) - g(x)}{h}$

4. $h'(x) = \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}$

5. this is legitimate only if the denominator $g(x+h-g(x)$ is nonzero. therefore, to continue our proof, we make the extra assumption that $g(x+h) - g(x) \neq 0$ for all h near but not equal to 0. this assumption is not necessary, but without it, the argument is more technical. the second limit on the right is $g'(x)$. the chain rule will follow if we show that the first limit equals $f'(g(x))$. to verify this, set $k = g(x+h) - g(x)$. then $g(x+h) = g(x) + k$ and
$\frac{f(g(x+h)-f(g(x))}{g(x+h)-g(x)} = \frac{f(g(x)+k)-f(g(x))}{k}$

6. since $g(x)$ is differentiable, it is also continuous. therefor $g(x+h)$ tends to $g(x)$ and $k = g(x+h) - g(x)$ tends to zero as $h \to 0$. thus, we may rewrite the limit in terms of k to obtain the desired result:
$\lim_{h \to 0} \frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)} = \lim_{k \to 0} \frac{f(g(x)+k-f(g(x))}{k} = f'(g(x))$

proof of the general power rule

1. we write $g(x)^n = f(g(x))$, with $f(u) = u^n$. then $f'(u) = nu^{n-1}$ and, by the chain rule,
$\frac{d}{dx}(g(x))^n = f'(g(x))g'(x) = n(g(x))^{n-1}g'(x)$

shifting and scaling rule if $f(x)$ is differentiable, then for any constants k and b,
$\frac{d}{dx}f(kx+b) = kf'(kx+b)$

implicit differentiation:
to differentiate using the methods covered thus far, we must have a formula for y in terms of x, for instance, $y = x^3 + 1$. but suppose that y is determined instead by an equation such as $y^4 + xy = x^3 - x + 2$. in this case, we say that y is defined impplicitly. how can we find the slope of the tangent line at a point on the graph? it may be inconvenient or even impossible to solve for y explicitly as a function of x, but we can find $\frac{dy}{dx}$ using the method of implicit differentiation. to illustrate, consider the equation of the unit circle:

1. $x^2 + y^2 = 1$

2. to compute $\frac{dy}{dx}$, first take the derivative of both sides of the equation and evaluate:
$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(1)$

3. $\frac{d}{dx}(x^2) + \frac{d}{dx}(y^2) = 0$

4. $2x + \frac{d}{dx}(y^2) = 0$

5. what should we do with the term $\frac{d}{dx}(y^2)$? we use the chain rule. if we think of y as a function $y = f(x)$, then $y^2 = f(x)^2$, and by the chain rule,
$\frac{d}{dx}y^2 = \frac{d}{dx}f(x)^2 = 2f(x)\frac{df}{dx} = 2y\frac{dy}{dx}$

6. $2x + \frac{d}{dx}(y^2) \to 2x + 2y\frac{dy}{dx}$, and we may solve for $\frac{dy}{dx}$ if $y \neq 0$:

7. $\frac{dy}{dx} = -\frac{x}{y}$

related rates:
in related rate problems, the goal is to calculate an unknown rate of change in terms of other rates of change that are known. one typical problem involves a ladder learning against a wall. the question is: how fast does the ladder move if the bottom of the ladder is pulled away from the wall at constant speed? what is interesting and perhaps surprising is that the top and bottom travel at different speeds.

linear approximation:
we can use the derivative to estimate Δf without computing it exactly. by definition,

the derivative is the limit
$f'(a) = \lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$
so when Δx is small, we have $\frac{\Delta f}{\Delta x} \approx f'(a)$, and thus, $\Delta f \approx f'(a)\Delta x$

the linear approximation if often called the tangent line approximation because of the following interpretation. the quantitiy Δf is the vertical change from $x = a$ to $a = a + \Delta x$ in the graph of $f(x)$. recall that for a nonvertical straight line, the vertical change is equal to the slope times the horizontal change. since the tangent line has slope $f'(a)$. the vertical change in the tangent line is $f'(a)\Delta x$. what the linear approximation does, therefore, is used the vertical change in the tangent line as an approximation to the vertical change in the graph of $f(x)$. when Δx is small, the two quantities are nearly equal.

Keep in mind the different roles played by Δf and $f'(a)\Delta x$. the quantity of interest is Δf and we estimate it by $f'(a)\Delta x$. most importantly, the linear approximation tells us that there is a nearly linear relationship between Δf and Δx when Δx is small.

linearization:
we can approximate $f(x)$ itself rather than the change Δf by rewriting the linear approximation in terms of the variable $x = a + \Delta x$. Then
$f(a + \Delta x) - f(a) \approx f'(a)\Delta x$
$f(x) - f(a) \approx f'(a)(x - a)$ (since $\Delta x = x - a$)
$f(x) \approx f(a) + f'(a)(x - a)$
the function on the right, denoted by $L(x)$, is called the linearization of $f(x)$ at $x = a$:
$L(x) = f'(a)(x - a) + f(a)$
we refer to $x = a$ as the center of the linearization. notice that $y = L(x)$ is the equation of the tangent line to the graph of $f(x)$ at $x = a$

approximating $f(x)$ by its linearization assume that f is differentiable at $x = a$ if x is close to a, then
$f(x) \approx L(x) = f'(a)(x - a) + f(a)$

extreme values on an interval let $f(x)$ be a function on an interval I and let $a \in I$. we say that $f(a)$ is the

- absolute minimum of $f(x)$ on I if $f(a) \leq f(x)$ for all $x \in I$

- absolute maximum of $f(x)$ on I if $f(a) \geq f(x)$ for all $x \in I$

existence of extrema on a closed interval if $f(x)$ is a continuous function on a closed (bounded) interval $I = [a, b]$, then $f(x)$ takes on a minimum and a maximum value on I.

local extrema we say that $f(x)$ has a

- at $x = c$ if $f(c)$ is the minimum value of f on some open interval (in the domain of f) containing c.

- at $x = c$ if $f(c)$ is the maximum value of $f(x)$ on some open interval (in the domain of f) containing c.

critical points a number c in the domain of f is called a critical point if either $f'(c) = 0$ or $f'(c)$ DNE.

fermat's theorem on local extrema if $f(c)$ is a local min or max, then c is a critical point of f.

extreme values on a closed interval assume that $f(x)$ is continuous on $[a, b]$ and let $f(c)$ be the minimum or maximum value on $[a, b]$. then c is either a critical point or one of the endpoints a or b.

rolle's theorem assume that $f(x)$ is continous on $[a, b]$ and differentiable on (a, b). if $f(a) = f(b)$, then there exists a number c between a and b such that $f'(c) = 0$.

mean value theorem assume that f is continuous on the closed interval $[a, b]$ and differentiable on (a, b). then there exists at least one value c in (a, b) such that
$f'(c) = \frac{f(b) - f(a)}{b - a}$

the sign of the derivative let f be a differentiable function on the open interval (a, b).

- if $f'(x) > 0$ for $x \in (a, b)$, then f is increasing on (a, b).
- if $f'(x) < 0$ for $x \in (a, b)$, then f is decreasing on (a, b).

first derivative test for critical points assume that $f(x)$ is differentiable and let c be a critical point of $f(x)$. then:

- $f'(x)$ changes from + to - at $c \Rightarrow f(c)$ is a local maximum.
- $f'(x)$ changes from - to + at $c \Rightarrow f(c)$ is a local minimum.

test for concavity suppose that $f''(x)$ exists for all $x \in (a, b)$.

- if $f''(x) > 0$ for all $x \in (a, b)$, then f is concave up on (a, b).
- if $f''(x) < 0$ for all $x \in (a, b)$, then f is concave down on (a, b).

test for inflection points assume that $f''(x)$ exists for all $x \in (a, b)$ and let $c \in (a, b)$. if $f''(c) = 0$ and $f''(x)$ changes sign at $x = c$, then $f(x)$ has a point of inflection at $x = c$.

second derivative test assume that $f(x)$ is differentiable and let c be a critical point. if $f''(c)$ exists, then

- $f''(c) > 0 \Rightarrow f(c)$ is a local minimum
- $f''(c) < 0 \Rightarrow f(c)$ is a local maximum
- $f''(c) = 0 \Rightarrow f(c)$ inconclusive: $f(c)$ may be a local min, max, or neither

newton's method:
this is a procedure for finding numerical approximations to zeros of functions. numerical approximations are important because it is often impossible to find the zeros exactly. for example, the polynomial $f(x) = x^5 - x - 1$ has one real root c, but we can prove, using an advanced branch of mathematics called Galois Theory, that there is no algebraic formula for this root. newtons method shows that $c \approx 1.1673$, and with enough computation, we can compute c to any desired degree of accuracy.

in newton's method, we begin by choosing a number x_0, which we believe is close to a root of $f(x)$. this starting value x_0 is called the initial guess. newton's method then produces a sequence $x_0, x_1, x_2, \ldots$ of successive approximations that, in favorable situations, converge to a root.

steps:

- choose initial guess x_0 (close to the desired root if possible).
- generate successive approximations $x_1, x_2, \ldots$ where
  $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

derivatives:

- $\frac{d}{dx}(c) = 0$
- $\frac{d}{dx}x = 1$
- $\frac{d}{dx}(x^n) = nx^{-1}$ (power rule)
- $\frac{d}{dx}[cf(x)] = cf'(x)$
- $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$
- $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$
- $\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$
- $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$
- $\frac{d}{dx}f(x)^n = nf(x)^{n-1}f'(x)$
- $\frac{d}{dx}\sin(x) = \cos(x)$
- $\frac{d}{dx}\cos(x) = -\sin(x)$
- $\frac{d}{dx}\tan(x) = \sec^2(x)$
- $\frac{d}{dx}\csc(x) = -\csc(x)\cot(x)$
- $\frac{d}{dx}\sec(x) = \sec(x)\tan(x)$
- $\frac{d}{dx}\cot(x) = -\csc^2(x)$
- $\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{(1-x^2)}}$
- $\frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{(1-x^2)}}$
- $\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2}$
- $\frac{d}{dx}(e^x) = e^x$
- $\frac{d}{dx}(a^x) = (\ln a)a^x$
- $\frac{d}{dx}\ln \mid x \mid = \frac{1}{x}$
- $\frac{d}{dx}\log_a x = \frac{1}{(\ln a)x}$

Integration

antiderivatives a function $F(x)$ is an antiderivative of $f(x)$ on (a, b) if $F'(x) = f(x)$ for all $x \in (a, b)$.

general antiderivative let $F(x)$ be an antiderivative of $f(x)$ on (a, b). then every other antiderivative on (a, b) is of the form $F(x) + C$ for some constant C.

indefinite integral the notion $\int f(x)\,dx = F(x) + C$ means that $F'(x) = f(x)$ we say that $F(x) + C$ is the antiderivative or indefinite integral of $f(x)$.

power rule for integrals $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$ for $n \neq -1$

linearity of the indefinite integral:

- $\int (f(x) + g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$

- $\int cf(x)\,dx = c \int f(x)\,dx$

basic trigonometric integrals:

- $\int \sin x\,dx = -\cos x + C$

- $\int \cos x\,dx = \sin x + C$

- $\int \sec^2 x\,dx = \tan x + C$

- $\int \csc^2 x\,dx = -\cot x + C$

- $\int \sec x \tan x\,dx = \sec x + C$

- $\int \csc x \cot x\,dx = -\csc x + C$

for the rest of this section assume that $f(x)$ is continuous and positive, so that the graph of $f(x)$ lies above the x-axis. as a first step, we approximate the area using rectangles. first, choose a whole number N and divide $[a, b]$ into N subintervals of equal width, each subinterval has width $\Delta x = \frac{b-a}{N}$ since $[a, b]$ has width $b - a$. the right endpoints of the subintervals are $a + \Delta x, a + 2\Delta x, \ldots, a + (N-1)\Delta x, a + N\Delta x$. notice that the last right enpoint is b because $a + N\Delta x = a + N(\frac{b-a}{N}) = b$. next, above each subinterval, construct the rectangle whose height is the value of $f(x)$ at the right enpoint of the subinterval. each rectangle has width Δx. the height of the first rectangle is $f(a + \Delta x)$ and its area is $f(a + \Delta x)\Delta x$. similarly, the second rectangle ahs a height $f(a + 2\Delta x)$ and area of $f(a + 2\Delta x)\Delta x$, ect. the sum of the areas of the rectangles is $f(a + \Delta x)\Delta x + f(a + 2\Delta x)\Delta x + \ldots + f(a + N\Delta x)\Delta x$. this sum is called the Nth right-endpoint approximation and is denoted R_N as $R_N = \Delta x[f(a + \Delta x) + f(a + 2\Delta x) + \ldots + f(a + N\Delta x)]$. in words, R_N is equal to Δx times the sum of the function values at the right endpoints of the subintervals.

linearity of summations

- $\sum_{j=m}^{n}(a_j + b_j) = \sum_{j=m}^{n} a_j + \sum_{j=m}^{n} b_j$

- $\sum_{j=m}^{n} Ca_j = C \sum_{j=m}^{n} a_j$ (C any constant)

- $\sum_{j=1}^{n} k = nk$ (k for any constant and $n \geq 1$)

$R_N = \Delta x[f(a + \Delta x) + f(a + 2\Delta x) + \ldots + f(a + N\Delta x)]$
$R_N = \Delta x \sum_{j=1}^{N} f(a + j\Delta x)$
$L_N = \Delta x[f(a) + f(a + \Delta x) + f(a + 2\Delta x) + \ldots + f(a + (N-1)\Delta x)]$
$L_N = \Delta x \sum_{j=0}^{N-1} f(a + j\Delta x)$
$M_N = \Delta x \sum_{j=1}^{N} f(a + (j - \frac{1}{2})\Delta x)$

if $f(x)$ is continuous on $[a, b]$, then the endpoint and midpoint approximations approach one and the same limit as as $N \to \infty$. in other words, there is a value L such that $\lim_{N\to\infty} R_N = \lim_{N\to\infty} L_N = \lim_{N\to\infty} M_N = L$

power sums
the kth power sum is the sum of the kth powers of the first N integers.

- $\sum_{j=1}^{N} j = 1 + 2 + \ldots + N = \frac{N(N+1)}{2} = \frac{N^2}{2} + \frac{N}{2}$

- $\sum_{j=1}^{N} j^2 = 1^2 + 2^2 + \ldots + N^2 = \frac{N(N+1)(2N+1)}{6} = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}$

- $\sum_{j=1}^{N} j^3 = 1^3 + 2^3 + \ldots + N^3 = \frac{N^2(N+1)^2}{4} = \frac{N^4}{4} + \frac{N^3}{2} + \frac{N^2}{4}$

in the endpoint and midpoint approximations, we approximate the area under the graph by rectangles of equal width. the heights of the rectangles are the values of $f(x)$ at the endpoints or midpoints of the subintervals. in riemann sum approximations, we relax these requirements; the rectangles need not have equal width and height of a rectangle may be any value of $f(x)$ within the subinterval. more formally, to define a riemann sum, we choose a partition and a set of intermediate points. a partition P of length N is any choice of points in $[a, b]$ that divides the interval into N subintervals:
partition P: $a = x_0 < x_1 < x_2 \ldots < x_N = b$
subintervals: $[x_0, x_1], [x_1, x_2], \ldots [x_{N-1}, x_N]$
we denote the length of the ith interval by $\Delta x_i = x_i - x_i - x_{i-1}$
the maximum of the lendths Δx_i is called the norm of the partition P and is denoted $\|P\|$. a set of intermediate points is a set of points $C = c_1, \ldots, c_N$, where c_i belongs to $[x_{i-1}, x_i]$

now, over each subinterval $[x_{i-1}, x_i]$ we constuct the rectangle of height $f(c_i)$ and base Δx_i. this rectangle has area $f(c_i)\Delta x_i$. if $f(c_i) < 0$, the rectangle extends below the x-axis and the "area" $f(c_i)\Delta x_i$ is negative. the riemann sum is the sum of these positive or negative areas: $R(f, P, C) = \sum_{i=1}^{N} f(c_i)\Delta x_i = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \ldots + f(c_N)\Delta x_N$

the endpoint and midpoint approximations are examples of riemann sums. they correspond to the partition of $[a, b]$ into N subintervals of equal length $\Delta x = \frac{b-a}{N}$ and the choice of right endpoints, left endpoints, or midpoints as intermediate points.

having introduced riemann sums, we are ready to make the following definition: a function is integrable of $[a, b]$ if all of the riemann sum (not just the endpoint and midpoint approximations) approach one and the same limit L as the norm of the partition tends to zero. more formally, we write
$L = \lim_{\|P\|\to 0} R(f, P, C) = \lim_{\|P\|\to 0} \sum_{i=1}^{N} f(c_i)\Delta x_i$
if $|R(f, P, C) - L|$ gets arbitrarily small as the norm $\|P\|$ tends to zero, no matter how we choose the partition and intermediate points. note that as $\|P\| \to 0$, the number N of intervals tends to ∞. the limit L is called the definite integral of $f(x)$ over $[a, b]$.

definite integral

the definite integral of $f(x)$ over $[a, b]$ is the limit of riemann sums and is denoted by the integral sign:

$\int_a^b f(x)\, dx = \lim_{\|P\| \to 0} R(f, P, C) = \lim_{\|P\| \to 0} \sum_{i=1}^N f(c_i) \Delta x_i$

when this limit exists, we say that $f(x)$ is integrable over $[a, b]$.

remarks

- we often refer to the definite integral more simply as the integral of f over $[a, b]$.

- the function $f(x)$ inside the integral sign is called integrand

- the numbers a and b representing the interval $[a, b]$ are called the limits of integration.

- any variable may be used as a variable of integration (this is a "dummy" variable). thus, the following three integrals all denote the same quantitiy: $\int_a^b f(x)\, dx$, $\int_a^b f(t)\, dt$, $\int_a^b f(u)\, du$

- recall that the indefinite integral $\int f(x)\, dx$ denotes the general antiderivative of $f(x)$

most functions that arise in practice are integrable. in particular, every continuous function is integrable, as stated in the following theorem.

if $f(x)$ is continuous on $[a, b]$, then $f(x)$ is integrable over $[a, b]$.

when $f(x)$ is continuous and positive, we interpret the definite integral as the area under the graph. if $f(x)$ is not positive, then the definite integral is not equal to an area in the usual sense, but we may interpret is as the signed area between the graph and the x-axis. signed area is defined as following:

signed area of a region = (area above x-axis) - (area below x-axis)

thus, signed area treats regions under the x-axis as "negative area". to see why the definit eintegral gives us signed area, suppose first that $f(x)$ is negative on $[a, b]$ and consider a riemann sum: $R(f, C, P) = f(c_1) \Delta x_1 + f(c_2) \Delta x_2 + \ldots + f(c_N) \Delta x_N$ since $f(c_i) < 0$, each term is equal to the negative of the area of a rectangle: $f(c_i) \Delta x_i = -(\text{area of rectangle of height } |f(c_i)|)$

therefore, the riemann sums $R(f, C, P)$ converge to the negative of the area between the graph and the x-axis. if $f(x)$ takes on both positive and negative values, then $f(c_i) \Delta x_i$ is positive or negative, depending on whether the corresponding rectangle lies above or below the x-axis. in this case, the riemann sums converge to the signed area. in summary,

$\int_a^b f(x)\, dx = $ signed area of region between the graph and x-axis over $[a, b]$

integral of a constant for any constant C, $\int_a^b C\, dx = C(b - a)$.

linearity of the definite integral if $f(x)$ and $g(x)$ are integrable over $[a, b]$, then

- $\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$

- $\int_a^b C f(x)\, dx = C \int_a^b f(x)\, dx$ for any constant C

reversing the limits of integration for $a < b$, we set $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$

additivity for adjacent intervals for $a \le b \le c$,
$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$

comparison theorem if $g(x) \le f(x)$ on an interval $[a, b]$, then
$\int_a^b g(x)\,dx \le \int_a^b f(x)\,dx$

the fundamental theorem of calculus, part I assume that $f(x)$ is continuous on $[a, b]$ and let $F(x)$ be an antiderivative of $f(x)$ on $[a, b]$. then
$\int_a^b f(x)\,dx = F(b) - F(a)$

part I of the fundamental theorem shows that we can compute definite integrals using antiderivatives. part II turns this relationship around: it tells us that we can use the definite integral to construct antiderivatives. to state part II, we introduce the area function (or cumulative area function) associated to a function $f(x)$ and lower limit a: $A(x) = \int_a^x f(t)\,dx =$ signed area from a to x in essence, we turn the definite integral into a function by trating the upper limit x as a variable rather than a constant. notice that $A(a) = 0$ because $A(a) = \int_a^a f(t)\,dt = 0$. although $A(x)$ is defined by an integral, in many cases we can find an explicit formula it.

fundamental theorem of calculus, part II let $f(x)$ be a continuous function on $[a, b]$. then $A(x) = \int_a^x f(t)\,dt$ is an antiderivative of $f(x)$, that is, $A'(x) = f(x)$, or equivalently,
$\frac{d}{dx} \int_a^x f(t)\,dt = f(x)$
furthermore, $A(x)$ satisfies the initial condition $A(a) = 0$.

integration (antidifferentiation) is generally more difficult than differentiation. there are no sure-fire methods for finding an antiderivative, and, in some cases, there is no elementary formula for the antiderivative. however, there are several general techniques that are widely applicable. one such technique is the substitution method, which uses the chain rule "in reverse".

substitution method if $F'(x) = f(x)$, then
$\int f(u(x))u'(x)\,dx = F(u(x)) + C$

there are two ways of using substitution to compute definite integrals. one way is to carry out the substitution method fully to find an antiderivative in terms of the x-variable. however, it is often more efficient to evaluate the definite integral directly in terms of the u-variable. this can be done provided that the limits of integration are changed as indicated in the next theorem.

change of variables formula for definite integrals if $u(x)$ is differentiable on $[a, b]$ and $f(x)$ is integrable on the range of $u(x)$, then
$\int_a^b f(u(x))u'(x)\,dx = \int_{u(a)}^{u(b)} f(u)\,du$

area between the graphs $= \int_a^b (y_{\text{top}} - y_{\text{bot}})\,dx = \int_a^b (f(x) - g(x))\,dx$

area between the graphs $= \int_c^d (g_2(y) - g_1(y))\,dy = \int_c^d (x_{\text{right}} - x_{\text{left}})\,dy$

volume as the integral of cross-sectional area

suppose that a solid body extends from height $y = a$ to $y = b$. let $A(y)$ be the area of the horizontal cross section at height y. then

volume of the solid body $= \int_a^b A(y)\,dy$

recall that the average of N numbers $a_1, a_2, \ldots, a_N$ is the sum divided by N:

$\frac{a_1 + a_2 + \ldots a_N}{N} = \frac{1}{N} \sum_{j=1}^N a_j$

we cannot define the average value of a function $f(x)$ on an interval $[a, b]$ as a sum because there are infinitely many values of x to consider. however, the right-endpoint appproximation may be interpreted as an average value:

$R_N = \frac{b-a}{N}(f(x_1) + f(x_2) + \ldots + f(x_N))$

where $x_i = a + i(\frac{b-a}{N})$. dividing by $(b - a)$, we obtain the average of the function values $f(x_i)$:

$\frac{1}{b-a} R_N = \frac{f(x_1) + f(x_2) + \ldots + f(x_N)}{N}$

if N is large, it is reasonable to think of this quantity as an approximation to the average of $f(x)$ on $[a, b]$. therefore, we define the average value itself as the limit:

average value $= \lim_{N \to \infty} \frac{1}{b-a} R_N(f) = \frac{1}{b-a} \int_a^b f(x)\,dx$ the average value is also called the mean value.

average value the average value of an integrable function $f(x)$ on $[a, b]$ is the quantitiy

average value $= \frac{1}{b-a} \int_a^b f(x)\,dx$

mean value theorem for integrals if $f(x)$ is continuous on $[a, b]$, then there exists a value $c \in [a, b]$ such that

$f(c) = \frac{1}{b-a} \int_a^b f(x)\,dx$

a solid of revolution is a solid obtained by rotating a region in the plane about an axis. the sphere and right circular cone are familiar examples of such solids.

volume of a solid of revolution: disk method if $f(x)$ is continuous and $f(x) \geq 0$ on $[a, b]$, then the volume V obtained by rotating the region under the graph about the x-axis is [with $R = f(x)$]:

$V = \pi \int_a^b R^2\,dx = \pi \int_a^b f(x)^2\,dx$

we now consider some variations on the formula for a volume of revolution. first suppose that the region between two curves $y = f(x)$ and $y = g(x)$, where $f(x) \geq g(x) \geq 0$, is rotated about the x-axis. then the vertical cross section of the solid at x is generated as the segment $\overline{AB}$ revolves around the x-axis. this cross section is a washer of outer radius $R = f(x)$ and inner radius $r = g(x)$. the area of this washer is $\pi R^2 - \pi r^2$ or $\pi(f(x)^2 - g(x)^2)$, and we obtain the volume of the solid as the integral of the cross-section area:

$V = \pi \int_a^b (R^2 - r^2)\,dx = \pi \int_a^b (f(x)^2 - g(x)^2)\,dx$

we computed volumes by integrating cross-section area. the shell method is based on a different idea and is more convenient in some cases. in the shell method, we use cylindrical shells to approximate a volume of revolution. to this end, let us first derive an approximation to the volume of a cylindrical shell of height h, outer radius R, and inner radius r. since the shell is obtained by removing a cylinder of radius r from the wider cylinder of radius R, its volume is equal to

$\pi R^2 h - \pi r^2 h = \pi h(R^2 - r^2) = \pi h(R + r)(R - r) = \pi h(R + r)\Delta r$

where $\Delta r = R - r$ is the shell's thickness. if the shell is very thin, then R and r are nearly equal and we may replace $(R + r)$ by $2R$ in the equation above to obtain the

approximation volume of shell $\approx 2\pi R h \Delta r$

now consider a solid obtained by rotating the region under $y = f(x)$ from $x = a$ to $x = b$ around the y-axis. the idea is to divide the solid into this concentric shells. more precisely, we divide $[a, b]$ into N subintervals of length $\Delta x = \frac{b-a}{N}$ with endpoints $x_0, x_1, \ldots, x_N$. when we rotate the thin strip of area above $[x_{i-1}, x_i]$ about the y-axis, we obtain a thin shell whose volume we denote by V_i. the total volume V of the solid is equal to $V = \sum_{i=1}^{N} V_i$.

the top rim of the ith shell is curved. however, when Δx is small we may approximate this thin shell by the cylindrical shell (with a flat rim) of height $f(x_i)$.
$V_i \approx$ (circumference)(height)(thickness) $= 2\pi x_i f(x_i) \Delta x$
therefore,
$V = \sum_{i=1}^{N} V_i \approx \sum_{i=1}^{N} 2\pi x_i f(x_i) \Delta x$

volume of a solid of revolution: the shell method the volume V of the solid obtained by rotating the region under the graph of $y = f(x)$ over the interval $[a, b]$ about the y-axis is equal to
$V = 2\pi \int_a^b x f(x) \, dx$

Exponential and Inverse Functions

an exponential function is a function of the form $f(x) = b^x$, where $b > 0$ and $b \neq 1$. the number b is called the base. some examples are 2^x, $(1.4)^x$, and 10^x. we exclude the case $b = 1$ because $f(x) = 1^x$ is a constant function. calculators give good decimal approximations to values of exponential functions: $2^4 = 16$, $2^{-3} = 0.125$, $(1.4)^3 = 2.744$, $10^{4.6} \approx 39,810.717$. three properties of exponential functions should be singled out from the start:

- exponential functions are positive: $b^x > 0$ for all x

- $f(x) = b^x$ is increasing if $b > 1$ and decreasing if $0 < b < 1$

- the range of $f(x) = b^x$ is the set of all positive real numbers

another important feature of $f(x) = b^x$ (for $b > 1$) is that it increases rapidly. although "rapid increase" is a subjective term, the following precise statement is true: $f(x) = b^x$ increases more rapidly than every polynomial function. for example, $f(x) = 3^x$ eventually overtakes and increases faster than the power functions x^3, x^4, and x^5.

laws of exponents $(b > 0)$

- $b^0 = 1$

- $b^x b^y = b^{x+y}$

- $\frac{b^x}{b^y} = b^{x-y}$

- $b^{-x} = \frac{1}{b^x}$

- $(b^x)^y = b^{xy}$

- $b^{\frac{1}{n}} = \sqrt[n]{b}$

inverse
let $f(x)$ have a domain D and range R. the inverse function $f^{-1}(x)$ (if it exists) is the function with domain R such that
$f^{-1}(f(x)) = x$ for $x \in D$ and $f(f^{-1}(x)) = x$ for $x \in R$
if f^{-1} exists, then f is called invertible.

one-to-one functions a function $f(x)$ is one-to-one (on its domain D) if for every value c, the equation $f(x) = c$ has at most one solution for $x \in D$

existence of inverses if $f(x)$ is one-to-one on its domain D, then f is invertible. furthermore,

- domain of f = range of f^{-1}

- range of f = domain of f^{-1}

horizontal line test a function $f(x)$ is one-to-one if and only if every horizontal line intersects the graph of f in at most one point.

derivative of the inverse assume that $f(x)$ is differentiable and one-to-one with inverse $g(x) = f^{-1}(x)$. if b belongs to the domain of $g(x)$ and $f'(g(b)) \neq 0$, then $g'(b)$ exists and $g'(b) = \frac{1}{f'(g(b))}$

the logarithm to the base b (where $b > 0$ and $b \neq 1$) is the inverse of the exponential function $f(x) = b^x$ and is denoted $\log_b x$. by definition, $b^{\log_b x} = x$ and $\log_b b^x = x$.

laws of logarithms

- $\log_b(1) = 0$

- $\log_b(b) = 1$

- $\log_b(xy) = \log_b x + \log_b y$

- $\log_b \frac{x}{y} = \log_b x - \log_b y$

- $\log_b(\frac{1}{x}) = -\log_b(x)$

- $\log_b(x^n) = n \log_b x$

we also note that all logarithm functions are proprtional. more precisely, the following change of base formula holds: $\log_b x = \frac{\log_a x}{\log_a b}$, $\log_b x = \frac{\ln(x)}{\ln(b)}$

derivative of $f(x) = \ln(x)$ $\frac{d}{dx} \ln(x) = \frac{1}{x}$ for $x > 0$

antiderivative of $y = \frac{1}{x}$ the function $f(x) = \ln|x|$ is an antiderivative of $y = \frac{1}{x}$ in the domain $x : x \neq 0$, that is, $\int \frac{dx}{x} = \ln|x| + C$

in this section we study quantities $P(t)$ that depend exponentially on time:

$$P(t) = P_0 e^{kt}$$

where P_0 and k are constants. if $k > 0$, then $P(t)$ grows exponentially with growth constant k, and if $k < 0$, then $P(t)$ decays exponentially with decay constant k. the coefficient P_0 is the initial size since $P(0) = P_0 e^{k \cdot 0} = P_0$. exponential growth or decay may

also be expressed by the formula $P(t) = P_0 b^t$ with $b = e^k$, since $P_0 b^t = P_0(e^k)^t = P_0 e^{kt}$.

the exponential functions $y = P_0 e^{kt}$ are the only functions that satisfy this differential equation.

theorem 1 if $y(t)$ is differentiable function satisfying the differential equation
$y' = ky$
then $y(t) = P_0 e^{kt}$, where P_0 is the initial value $P_0 = y(0)$

doubling time if $P(t) = P_0 e^{kt}$ with $k > 0$, then the doubling time of P is:
doubling time $= \frac{\ln 2}{k}$
if $P(t) = P_0 e^{-kt}$ then this is the formulae for half-life

exponential functions play an important role in financial calculations. in this section, we use these functions to compute compound interest and present value. the initial sum of money P_0 in an account or investment is called the principal. when a principal of P_0 dollars is deposited into a bank account, the amount or balance in the account at time t depends on two factors: the interest rate r and how often interest is compounded. if interest is paid out once a year at the end of the year, we say that the interest is compounded annually. if the principal and interest are retained in the account, then the balance grows exponentially.

compound interest if P_0 dollars are deposited into an account earning interest at an annual rate r, compounded M times yearly, then the value of the account after t years is $P(t) = P_0(1 + \frac{r}{M})^{Mt}$.

limit formula for e and e^x
$e = \lim n \to \infty (1 + \frac{1}{n})^n$ and $e^x = \lim_{n \to \infty} (1 + \frac{x}{n})^n$ for all x

continuously compounded interest if P_0 dollars are deposited into an account earning interest at an annual rate r, compounded continuously, then the value of the account after t years is $P(t) = P_0 e^{rt}$
the concept of present value (PV) is used in finance to compare the values of payments made at different times. assume that there is an interest rate r at which any investor can lend or borrow money and that interest is compounded continuously. then the PV of P dollars to be received t years in the future is defined to be Pe^{-rt}:
The PV of P dollars received at time t is Pe^{-rt}
what is the reasoning behind this definition? when you invest at the interest rate r for t years, your principal increases by a factor e^{rt}, so you invest Pe^{-rt}, dollars, your principal grows to $Pe^{-rt}e^{rt} = P$ dollars at time t. thus, the present value Pe^{-rt} is the amount you would have to invest today in order to have P dollars at time t
PV of an income stream the present value at interest rate r of an income stream paying out $R(t)$ dollars/year continuously for T years is $PV = \int_0^T R(t)e^{-rt} \, dt$

we have seen that a quantity grows or decays exponentially if its rate of change is proportional to the amount present. this characteristic property is expressed by the differential equation $y' = ky$. we now study a closely related differential equation:
$\frac{dy}{dt} = k(y - b)$
where k and b are constants and $k \neq 0$. this differential equation describes a quantity y whose rate of change is proportional to the size of the difference $y - b$. the general solution is $y(t) = b + Ce^{kt}$
where C is a constant determined by the initial condition. to very this, we note that $(y-b)' = y'$ since b is a constant, so the first equation may be rewritten $\frac{d}{dt}(y-b) = k(y-b)$

in other words, $y - b$ satisfies the differential equation of an exponential function. thus, $y - b = Ce^{kt}$ or $y = b + Ce^{kt}$, as claimed.

L'Hôpital's rule assume that $f(x)$ and $g(x)$ are differentiable on an open interval containing a and that
$$f(a) = g(a) = 0$$
also assume that $g'(x) \neq 0$ for x near but not equal to a. then
$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f(x)'}{g(x)'}$$
provided that the limit on the right exists. this conclusion also holds if $f(x)$ and $g(x)$ are differentiable for x near (but not equal to) a and
$$\lim_{x \to a} f(x) = \pm\infty \text{ and } \lim_{x \to a} f(x) = \pm\infty$$
furthermore, these limits may be replaced by one-sided limits.

sometimes, we are interested in determining which of two functions, $f(x)$ and $g(x)$, grows faster. for example, there are two standard computer algorithms for sorting data (alphabetizing, ordering according to rank, ect.): quick sort and bubble sort. the average time equired to sort a list of size n has order of magnitude $n\ln(n)$ for quick sort and n^2 for bubble sort. which algorithm is faster when the size n is large? although n is a whole number, this problem amounts to comparing the growth of $f(x) = x\ln(x)$ and $g(x) = x^2$ as $x \to \infty$.

we say that $f(x)$ grows faster than $g(x)$ if
$$\lim x \to \infty \frac{f(x)}{g(x)} = \infty \text{ or equivalently, } \lim x \to \infty \frac{g(x)}{f(x)} = 0$$
to compare the growth of functions, we need a version of L'Hôpital's rule that applies to limits as $x \to \infty$.

L'Hôpital's rule for limits as $x \to \infty$ assume that $f(x)$ and $g(x)$ are differentiable in an interval (b, ∞) and that $g'(x) \neq 0$ for $x > b$. if $\lim_{x \to \infty} f(x)$ and $\lim_{x \to \infty} g(x)$ both exists and either both are zero or both are infinite, then
$$\lim_{x \to infty} \frac{f(x)}{g(x)} = \lim_{x \to infty} \frac{f'(x)}{g'(x)}$$
provided that the limit on the right exists. a similar result holds for limits as $x \to -\infty$.

$\theta = \sin^{-1}(x)$ is the unique angle in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ such that $\sin(\theta) = x$
$\theta = \cos^{-1}(x)$ is the unique angle in $[0, \pi]$ such that $\cos(\theta) = x$

derivatives of arcsine and arccosine
$\frac{d}{dx} \sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$, $\frac{d}{dx} \cos^{-1}(x) = -\frac{1}{\sqrt{1-x^2}}$

derivatives of inverse trigonometric functions

- $\frac{d}{dx} \tan^{-1}(x) = \frac{1}{\sqrt{1+x^2}}$

- $\frac{d}{dx} \cot^{-1}(x) = -\frac{1}{\sqrt{1+x^2}}$

- $\frac{d}{dx} \sec^{-1}(x) = \frac{1}{|x|\sqrt{1-x^2}}$

- $\frac{d}{dx} \csc^{-1}(x) = -\frac{1}{|x|\sqrt{1-x^2}}$

the hyperbolic functions are certain special combinations of e^x and e^{-x} that play a role in engineering and physics. the hyperbolic sine and cosine, often called "cinch" and "cosh", are the functions defined as follows:
$\sinh(x) = \frac{e^x - e^{-x}}{2}$, $\cosh(x) = \frac{e^x + e^{-x}}{2}$

as the terminology suggest, there are similarities between the hyperbolic and trigono-
metric functions. the trigonometric functions and their hyperbolic analogs have the
same parity., thus, sin x and sing x are both odd, and cos x and cosh x are both
even. the basic trigonometric identity $\sin^2(x) + \cos^2(x) = 1$ has a hyperbolic analog:
$\sinh^2(x) - \cosh^2(x) = 1$. the addition formulas satisfied by $\sin(\theta)$ and $\cos(\theta)$ have hy-
perbolic analogs:
$\sinh(A + B) = \sinh(A)\cosh(B) + \cosh(A)\sinh(B)$
$\cosh(A + B) = \cosh(A)\cosh(B) + \sinh(A)\sinh(B)$

it follows from the definitions that $\cosh(x) + \sinh(x) = e^x$, $\cosh(x) - \sinh(x) = e^x$ we
obtain the following by multiplying these two equation together:
$\cosh^2(x) - \sinh^2(x) = (\cosh(x) + \sinh(x))(\cosh(x) - \sinh(x)) = e^x \cdot e^{-x} = 1$

hyperbola instead of the circle: the identity $\sin^2(t) + \cos^2(t) = 1$ tells us that
the point $(\cos(t), \sin(t))$ lies on the unit circle $x^2 + y^2 = 1$. similarly, the identity
$\cosh^2(t) - \sinh^2(t) = 1$ says that the point $(\cosh(t), \sinh(t))$ lies on the hyperbola
$x^2 - y^2 = 1$

derivative formulas: the formulas for derivatives of the hyperbolic functions are sim-
ilar to the formulas for the corresponsing trigonometric function, differing at most by a
minus sign: $\frac{d}{dx}\sinh(x) = \cosh(x)$, $\frac{d}{dx}\cosh(x) = \sinh(x)$

the terms "hyperbolic sine" and "hyperbolic cosine" suggest a connection between the
hyperbolic and trigonometric functions. this excursion explored the source of the con-
nection, which leads us to complex numbers and Euler's famous formula.
recall that $y = e^t$ satisfies the differential equation $y' = y$ and every solution is of the
form $y = Ce^t$ for some constant C. on the other hand, we can check that both $y = e^t$
and $y = e^{-t}$ satisfy the second-order differential equation
$y'' = y$
indeed, $(e^t)'' = e^t$ and $(e^{-t})'' = (-e^{-t}) = e^{-t}$. it can be shown that every solution of
this equation has the form $y = Ae^t + Be^{-t}$ for some constants A and B.
now let's examine what happens when we cahnge the equation by a minus sign:
$y'' = -y$
in this case, $y = \sin(t)$ and $y = \cos(t)$ are solutions because
$(\sin(t))'' = (\cos(t))' = -\sin(t)$, $(\cos(t))'' = (-\sin(t))' = -\cos(t)$
furthermore, it can be proved as before that every solution of the equation has the form
$y = A\cos(t) + B\sin(t)$
this might seem to be the end of the story. however, there is another way to write down
solutions to the equation. consider the exponential functions $y = e^{it}$ and $y = e^{-it}$,
where
$i = \sqrt{-1}$
the number i is an imaginary complex number satisfying $i^2 = -1$. since i is not a real
number, the exponential e^{it} is not defined without further explanation. but if we accept
that e^{it} can be defined and that it is legitimate to apply the usual rules of calculus to
it, we obtain
$\frac{d}{dt}e^{it} = ie^{it}$
differentiating a second time yields
$(e^{it})'' = (ie^{it})' = i^2 e^{it} = -e^{it}$
in other words, $y = e^{it}$ is a solution to the equation $y'' = -y$ and must therefore be
expressible in terms of $\sin(t)$ and $\cos(t)$. that is, for some constant A and B,
$e^{it} = A\cos(t) + B\sin(t)$
we determine A and B by considering initial conditions. first, set $t = 0$ in the equation

$1 = e^{i0} = \frac{d}{dt}e^{it} = A\cos(0) + B\sin(0) = A$

then take the derivative of the equation and set $t = 0$:

$ie^{it} = \frac{d}{dx}e^{it} = A\cos'(t) + B\sin'(t) = -A\sin(t) + B\cos(t)$

$i = ie^{i0} = -A\sin(0) + B\cos(0) = B$

thus, $A = 1$ and $B = i$, and the equation yields Euler's formula:

$e^{it} = \cos(t) + i\sin(t)$

Euler derived this result using power series, which allows us to define e^{it} in a rigorous fashion.

At $t = \pi$, Euler's formula yields

$e^{i\pi} = -1$

here we have a simple but surpring relation between the three important numbers e, π, and i.

euler's formula also reveals the source of the analogy between hyperbolic and trigonometric functions. let us calculate the hyperbolic cosine at $x = it$:

$\cosh(it) = \frac{e^{it}+e^{-it}}{2} = \frac{\cos(t)+\sin(t)}{2} + \frac{\cos(-t)+i\sin(-t)}{2} = \cos(t)$

a similar calculation shows that $\sinh(it) = i\sin(t)$. in other words, the hyperbolic and trigonometric functions are not merly analogous - once we introduce complex numbers, we see that they are very nearly the same functions.

Techniques of Integration

numerical integration is the process of approximating a definite integral using well-chosen sums of function values. we use numerical integration when we cannot find a formula for an antiderivative and thus cannot apply the FTC. for instanc, the gaussian function $f(x) = e^{\frac{-x^2}{2}}$, whose graph is a bell-shaped curve, does not have an elementary antiderivative. numerical methods are needed to approximate $\int_a^b e^{\frac{-x^2}{2}}\,dx$.

let N be a whole number. the trapezodial rule T_N consists of approximating $\int_a^b f(x)\,dx$ by the average of the left-and right-endpoint approximations: $T_N = \frac{1}{2}(R_N + L_N)$

this approximation was used informally in previous chapters, where we observed that the average was likely to give a better approximation than either of R_N or L_N alone.

to derive a formula for T_N, let us recall the formulas for the endpoint approximations. we divide $[a, b]$ into N subintervals of length $\Delta x = \frac{b-a}{N}$ with endpoints

$x_0 = a, x_1 = a + \Delta x, x_2 = a + 2\Delta x, \ldots, x_N = b$

for convenience, we denote the values of $f(x)$ at the endpoints by y_j:

$y_j = f(x_j) = f(a + j\Delta x)(j = 0, 1, 2, \ldots, N)$

in particular, $y_0 = f(a)$ and $y_N = f(b)$. then

$R_N = \Delta x \sum_{j=1}^{N} y_j$, $L_N = \Delta x \sum_{j=0}^{N-1} y_j$

when we form the sum $R_N + L_N$, each value y_j occurs twice except for the first y_0 (which only occursd in L_N) and the last y_N (which only occurs in R_N). therefore,

$R_N + L_N = \Delta x(y_0 + 2y_1 + 2y_2 + \ldots + 2y_{N-1} + y_N)$

dividing by 2 we obtain the formula for T_N

trapezodial rule the Nth trapezodial approximation to $int_a^b f(x)\,dx$ is

$T_N = \frac{1}{2}\Delta x(y_0 + 2y_1 + \ldots + 2y_{N-1} + y_N)$

where $\Delta x = \frac{b-a}{N}$ and $y_j = f(a + j\Delta x)$.

midpoint rule the Nth midpoint approximation to $\int_a^b f(x)\,dx$ is

$M_N = \Delta x(f(c_1) + f(c_2) + \ldots + f(c_N))$

where $\Delta x = \frac{b-a}{N}$ and $c_j = a + (j - \frac{1}{2})\Delta x)$ is the midpoint of the jth interval $[x_{j-1}, x_j]$.

in applications, it is important to know the accuracy of a numerical approximation. we define the error in T_N and M_N by

$\text{Error}(T_N) = |T_N - \int_a^b f(x)\,dx|$, $\text{Error}(M_N) = |M_N - \int_a^b f(x)\,dx|$

accoring to the following theorem, the magnitudes of these errors are related to the size of the second derivative $f''(x)$-here we assume $f''(x)$ exists and is continuous.

error bound for T_N and M_N let K_2 be a number such that $|f''(x)| \leq K_2$ for all $x \in [a,b]$. then

$\text{Error}(T_N) \leq \frac{K_2(b-a)^3}{12N^2}$, $\text{Error}(M_N) \leq \frac{K_2(b-a)^3}{24N^2}$

can we improve on the trapezoidal and midpoint rules? one clue is that when $f(x)$ is concave (up or down), then the exact value of the integral lies between T_N and M_N. in fact, we see geometrically that

- if $f(x)$ is concave down $\Rightarrow T_N \leq \int_a^b f(x)\,dx \leq M_N$

- if $f(x)$ is concave up $\Rightarrow M_N \leq \int_a^b f(x)\,dx \leq T_N$

this suggests that the average of T_N and M_N may be more accurate than either T_N or M_N alone.

simpson's rule S_N exploits this idea, but it takes into account that M_N is roughly twice as accurate as T_N. therefore, S_N uses a weighted average, which is more accurate than the ordinary average. for N even, we define

$S_N = \frac{1}{3}T_{\frac{N}{2}} + \frac{2}{3}M_{\frac{N}{2}}$

to derive a formula for S_N, let $\Delta x = \frac{b-a}{N}$ and, as before, set $x_j = a + j\Delta x$ and $y_j = f(a + j\Delta x$. if we use just the even-numbered endpoints, we obtain a partition of $[a,b]$ into $\frac{N}{2}$ subintervals (keep in mind that N is even):

$[x_0, x_2], [x_2, x_4], \ldots, [x_{N-2}, x_N]$

the endpoints of these intervals are used to compute $T_{\frac{N}{2}}$, and we use the midpoints $x_1, x_3, \ldots, x_{N-1}$ to compute $M_{\frac{N}{2}}$

$N_{\frac{N}{2}} = (\frac{1}{2})\frac{b-a}{\frac{N}{2}}(2y_0 + 2y_2 + 2y_4 + \ldots + 2y_{N-2} + y_N)$

$M_{\frac{N}{2}} = \frac{b-a}{\frac{N}{2}}(y_1 + y_3 + y_5 + \ldots + y_{N-1}) = \frac{b-a}{N}(2y_1 + 2y_3 + 2y_5 + \ldots + 2y_{N-1})$

thus,

$S_N = \frac{1}{3}T_{\frac{N}{2}} + \frac{2}{3}M_{\frac{N}{2}} = \frac{1}{3}\Delta x(y_0 + 2y_2 + 2y_4 + \ldots + 2y_{N-2} + y_N) + \frac{1}{3}\Delta x(4y_1 + 4y_3 + 4y_5 + \ldots + 4y_{n-1})$

simpson's rule assume that N is even. let $\Delta x = \frac{b-a}{N}$ and $y_j = f(a + j\Delta x)$.

the Nth approximation to $\int_a^b f(x)\,dx$ by simpson's rule the quantity

$S_N = \frac{1}{3}\Delta x[y_0 + 4y_1 + 2y_2 + \ldots + 4y_{N-3} + 2y_{N-2} + 4y_{N-1} + y_N]$

we now state (without proof) a bound for the error in simpson's rule, where

$\text{Error}(S_N) = |S_N(f) - \int_a^b f(x)\,dx$

the error involves the fourth derivative, which we assume exists and is continuous.

error bound for S_N let K_4 be a number such that $|f^{(4)}(x)| \leq K_4$ for all $x \in [a,b]$. then

$\text{Error}(S_N) \leq \frac{K_4(b-a)^5}{180N^4}$

the integration by parts formula is derived from the product rule:

$(u(x)v(x))' = u(x)v'(x) + u'(x)v(x)$

this formula states that $u(x)v(x)$ is an antiderivative of the right-hand side, so
$u(x)v(x) = \int u(x)v'(x)\,dx + \int u'(x)v(x)\,dx$
combining the term on the left with the second integral on the right, we obtain

integration by parts formula
$\int u(x)v'(x)\,dx = u(x)v(x) - \int u'(x)v(x)\,dx$

integration by parts can be used to derive reduction formulas for integrals that depend on a positive integer n such as $\int x^n e^x\,dx$ or $\int \ln^n x\,dx$. a reduction formula (also called a recursive formula) expresses the integral for a given value of n in terms of a similar integral for a smaller value of n. the desired integral is evaluated by applying the reduction formula repeatedly.

a reduction formula verify the reduction formula
$\int x^n e^x\,dx = x^n e^x - n \int x^{n-1} e^x\,dx$
then use the reduction formula to evaluate $\int x^3 e^x\,dx$. **solution** we derive the reduction formula using integration by parts once:
$u = x^n, v' = e^x, u' = nx^{n-1}, v = e^x$
$\int x^n e^x\,dx = uv - \int u'v\,dx = x^n e^x - n \int x^{n-1} e^x\,dx$
to evaluate $\int x^3 e^x\,dx$, we will need to use the reduction formula for $n = 3, 2, 1$:

1. $\int x^3 e^x\,dx = x^3 e^x - 3 \int x^2 e^x\,dx$

2. $= x^3 e^x - 3(x^2 e^x - 2 \int xe^x\,dx) = x^3 e^x - 3x^2 e^x + 6 \int xe^x\,dx$

3. $= x^3 e^x - 3x^2 e^x + 6(xe^x - \int e^x\,dx) = x^3 e^x - 3x^2 e^x + 6xe^x - 6e^x + C$

4. $= (x^3 - 3x^2 + 6x - 6)e^x + C$

in this section, we combine stubstitution and integration by parts with the appropriate identities to integrate various trigonometirc functions. first, consider
$\int \sin^m(x) \cos^n(x)\,dx$
where n, m are positive integers. the easier case occurs when one or both of the exponents m, n is odd.

in general, if n is odd, we may use the identity $\cos^2(x) = 1 - \sin^2(x)$ to write $\cos^n(x)$ as a power of $(1 - \sin^2(x))$ times $\cos(x)$ and subsitute $u = \sin(x), du = \cos(x)dx$. alternatively, if m is odd, we write $\sin^m(x)$ as a power of $(1 - \cos^2(x))$ times $\sin(x)$.

reduction formulas for sine and cosine

- $\int \sin^n(x)\,dx = -\frac{1}{n} \sin^{n-1} x \cos(x) + \frac{n-1}{n} \int \sin^{n-2}(x)\,dx$

- $\int \cos^n(x)\,dx = -\frac{1}{n} \cos^{n-1} x \sin(x) + \frac{n-1}{n} \int \cos^{n-2}(x)\,dx$

example 3 evaluate $\int \sin^4(x)\,dx$
by reduction formula with $n = 4$, we obtain
$\int \sin^4(x)\,dx = -\frac{1}{4} \sin^3(x) \cos(x) + \frac{3}{4} \int \sin^2(x)\,dx$
to evaluate the integral on the right, we apply the reduction formula again with $n = 2$:
$\int \sin^2(x)\,dx = -\frac{1}{2} \sin(x) \cos(x) + \frac{1}{2} \int dx = -\frac{1}{2} \sin(x) \cos(x) + \frac{1}{2}x + C$
$\int \sin^4(x)\,dx = -\frac{1}{4} \sin^3(x) \cos(x) - \frac{3}{8} \sin(x) \cos(x) + \frac{3}{8}x + C$
trigonometric integrals can be expressed in more than one way because of the large number of trigonometric identities. for example, a computer algebra system may yield the following evaluation of the integral in the previous example:

$\int \sin^4(x) \, dx = \frac{1}{32}(x - 8\sin(2x) + \sin(4x)) + C$

you can check that this agrees with the result. the following formulas are verified using some of those double angle identities.

- $\int \sin^2(x) \, dx = \frac{x}{2} - \frac{\sin(2x)}{4} + C = \frac{x}{2} - \frac{1}{2}\sin(x)\cos(x) + C$

- $\int \cos^2(x) \, dx = \frac{x}{2} + \frac{\sin(2x)}{4} + C = \frac{x}{2} + \frac{1}{2}\sin(x)\cos(x) + C$

more work is required to integrate $\sin^m(x)\cos^n(x)$ when both n and m are even:
if $m \leq n$, use the identity $\sin^2(x) = 1 - \cos^2(x)$ to replace $\sin^m(x)$ by $(1 - \cos^2(x))^{\frac{m}{2}}$:
$\int \sin^m(x)\cos^n(x) \, dx = \int (1 - \cos^2(x))^{\frac{m}{2}}\cos^n(x) \, dx$
expand the integral on the right to obtain a sum of integrals of powers of $\cos(x)$ and use reduction formula.
if $m \geq n$, replace $\cos^n(x)$ by $(1 - \sin^2(x))^{\frac{n}{2}}$:
$\int \sin^m(x)\cos^n(x) \, dx = \int (\sin^m(x))(1 - \sin^2(x))^{\frac{n}{2}} \, dx$
expand the integral on the right to obtain a sum of integrals of powers of $\sin(x)$ and again evaluate using reduction formula.

our next goal is to integrate functions involving one of the square root expressions:
$\sqrt{a^2 - x^2}, \sqrt{x^2 + a^2}, \sqrt{x^2 - a^2}$
in each case, a substitution transforms the integral into a trigonometric integral. for example, the substitution $x = a\sin(\theta)$ may be used when the integrand involves $\sqrt{a^2 - x^2}$.

integrals involving $\sqrt{a^2 - x^2}$ if $\sqrt{a^2 - x^2}$ occurs in an integral where $a > 0$, try the substitution
$x = a\sin(\theta), \, dx = a\cos(\theta)d\theta, \, \sqrt{a^2 - x^2} = a\cos(\theta)$

integrals involving $\sqrt{a^2 + x^2}$ if $\sqrt{a^2 + x^2}$ occurs in an integral where $a > 0$, try the substitution
$x = a\tan(\theta), \, dx = a\sec^2(\theta)d\theta, \, \sqrt{a^2 + x^2} = a\sec(\theta)$

integrals involving $\sqrt{x^2 - a^2}$ if $\sqrt{x^2 - a^2}$ occurs in an integral where $a > 0$, try the substitution
$x = a\sec(\theta), \, dx = a\sec(\theta)\tan(\theta)d\theta, \, \sqrt{x^2 - a^2} = a\tan(\theta)$

in this section, we introduce the methods of partial fractions, used to integrate rational functions $\frac{P(x)}{Q(x)}$, where $P(x)$ and $Q(x)$ are polynomials. as we will see the integral of a rational function can be expressed as a sum of three types of terms: rational functions, arctangents of linear or quadratic functions, and logarithms of polynomials. here is a typical example:
$\int \frac{(2x^3 + x^2 - 2x)dx}{(x^2 + 1)^2} = \frac{x+4}{2x^2 + 2} + \frac{3}{2}\tan^{-1}x + \ln(x^2 + 1) + C$
the method of partial fractions yields an explicit antiderivative of this type whenever we can factor the denominator $Q(x)$ into a product of linear and quadratic factors.
a rational function $\frac{P(x)}{Q(x)}$ is called proper if the degree of $P(x)$ is less than the degree of $Q(x)$.
suppose first that $\frac{P(x)}{Q(x)}$ is proper and that the denominator $Q(x)$ factors as a product of distinct linear factors. in other words,
$\frac{P(x)}{Q(x)} = \frac{P(x)}{(x-a_1)(x-a_2)...(x-a_n)}$
where the roots $a_1, a_2, \ldots, a_n$ are all distinct and $\deg(P) < n$. then there is a partial fraction decomposition:
$\frac{P(x)}{Q(x)} = \frac{A_1}{(x-a_1)} + \frac{A_2}{x-a_2} + \ldots + \frac{A_n}{x-a_n}$

a quadratic polynomial $ax^2 + bx + c$ is called irreducible if it cannot be written as a product of two linear factors (without using complex numbers). a power of an irreducible quadratic factor $(ax^2 + bx + c)^M$ contributes a sum of the following type to a partial fraction decomposition:

$\frac{A_1 x + B_1}{ax^2 + bx + c} + \frac{A_2 x + B_2}{(ax^2 + bx + c)^2} + \ldots + \frac{A_M x + B_M}{(ax^2 + bx + c)^M}$

for example,

$\frac{4 - x}{x(x^2 + 4x + 2)^2} = \frac{1}{x} - \frac{x + 4}{x^2 + 4x + 2} - \frac{2x + 9}{(x^2 + 4x + 2)^2}$

you may need to use trigonometric substitution to integrate these terms. in particular, the following result may be useful.

$\int \frac{dx}{x^2 + a} = \frac{1}{\sqrt{a}} \tan^{-1}\left(\frac{x}{\sqrt{a}}\right) + C$ (for $a > 0$)

sometimes, we are interested in integrating a function over an infinite interval. for example, in statistics, the probability integral

$\int_{\infty}^{\infty} e^{\frac{-x^2}{2}} \, dx$

represents the area under the bell-shaped curve for $-\infty < x < \infty$. this is an example of an improper integral. although the region under the curve extends infinitely to the left and right, the total area is finite.

we consider three types of improper integrals, depending on whether one or both endpoints are infinite:

$\inf_{-\infty}^{b} f(x) \, dx$, $\inf_{a}^{\infty} f(x) \, dx$, $\inf_{-\infty}^{\infty} f(x) \, dx$

improper integral fix a number a and assume that $f(x)$ is integrable over $[a, b]$ for all $b \geq a$. the improper integral of $f(x)$ over $[a, \infty)$ is defined as the limit (if it exists)

$\int_{a}^{\infty} f(x) \, dx = \lim_{R \to \infty} f(x) \, dx$

when the limit exists, we say that the improper integral converges. in a similar fashion we define $\int_{-\infty}^{a} f(x) \, dx = \lim_{R \to -\infty} \int_{R}^{a} f(x) \, dx$

a doubly infinite improper integral is defined as a sum (provided that both integrals on the right converge):

$\int_{-infty}^{\infty} f(x) \, dx = \int_{-\infty}^{0} f(x) \, dx + \int_{0}^{\infty} f(x) \, dx$