FSc/ICS Part-I (1st Year) · PECTAA Board · English Medium

Mathematics 1st Year

Complete Notes — All 14 Units

freebooks.pk

Table of Contents

Prepared by freebooks.pk — original student notes covering the PECTAA Board syllabus for FSc/ICS Part-I (1st Year) Mathematics. Each unit includes key concepts, definitions, formulas, diagrams, solved examples, short and long questions, MCQs, revision summaries, and exam tips.

Unit 1: Complex Numbers

The real number system has no solution for an equation as simple as $x^2 + 1 = 0$, because no real number squares to a negative value. Complex numbers extend the real numbers by introducing the imaginary unit $i = \sqrt{-1}$, so that $i^2 = -1$, and every polynomial equation -- not just some of them -- finally has a solution. This unit builds complex numbers from the ground up: their algebra (addition, subtraction, multiplication, division), their formal ordered-pair definition and field properties, how to plot and measure them on the Argand diagram, equality and square roots, factorization of complex polynomials, the cube and fourth roots of unity, and the polar form that makes multiplying and dividing complex numbers far simpler.

Complex numbers are not just an abstract algebra exercise -- the unit closes with two real-world uses: analysing alternating-current (AC) electrical circuits, where impedance is naturally expressed as a complex number, and cryptography, where multiplying a message by a complex "key" encrypts it, and dividing by that same key decrypts it.

Learning Objectives

- Define a complex number and identify its real and imaginary parts
- Perform addition, subtraction, multiplication, and division of complex numbers
- Define a complex number as an ordered pair of real numbers and verify its field properties
- Plot a complex number on the Argand diagram and find its modulus
- State the condition for the equality of two complex numbers
- Find the square root of a complex number using the general formula
- Apply the Fundamental Theorem of Algebra and the Rational Root Theorem to factor complex polynomials
- Solve quadratic equations with complex roots by completing the square
- Find the cube roots and fourth roots of unity and state their key properties
- Convert a complex number between rectangular form and polar form

- Multiply and divide complex numbers in polar form, and apply complex numbers to AC circuits and cryptography

Key Concepts

1.1 Complex Numbers and Their Algebra

A complex number is a number of the form $z = x + iy$, where x and y are real numbers and i is the imaginary unit defined by $i = \sqrt{-1}$, so that $i^2 = -1$. Here $x = \mathrm{Re}(z)$ is called the real part of z and $y = \mathrm{Im}(z)$ is called the imaginary part of z. If $y = 0$, z is simply a real number; if $x = 0$ and y is not 0, z is called purely imaginary.

Addition and subtraction of complex numbers is done by combining the real parts together and the imaginary parts together: $(a+ib) \pm (c+id) = (a \pm c) + i(b \pm d)$. Multiplication is done by distributing normally and then simplifying using $i^2 = -1$: $(a+ib)(c+id) = (ac-bd) + i(ad+bc)$. Division is done by multiplying both the numerator and the denominator by the conjugate of the denominator, which turns the denominator into a real number so the result can be written in $x+iy$ form.

1.2 Ordered Pair Definition and Field Properties

A complex number can also be defined formally as an ordered pair of real numbers (a, b), with addition defined as $(a,b) + (c,d) = (a+c, b+d)$ and multiplication defined as $(a,b)(c,d) = (ac-bd, ad+bc)$. Under this definition, i corresponds to the pair $(0,1)$, and $(0,1)(0,1) = (-1,0)$ matches $i^2 = -1$ exactly -- this is what makes $i^2 = -1$ rigorous rather than just a rule to memorize.

Complex numbers under these operations satisfy all the properties of a mathematical field: closure and commutativity of addition and multiplication, associativity, the distributive law, an additive identity $(0,0)$ and multiplicative identity $(1,0)$, an additive inverse $-z$ for every z, and a multiplicative inverse $z^{-1} = \bar{z} / |z|^2$ for every non-zero z.

1.3 The Argand Diagram and Modulus

A complex number z = x + iy can be represented as the point (x, y) on a coordinate plane called the Argand diagram (or complex plane), where the horizontal axis is the real axis and the vertical axis is the imaginary axis. This turns every complex number into a specific, plottable point, and every algebraic operation on complex numbers into a geometric operation on points.

The modulus of z, written |z|, is the distance of the point (x, y) from the origin: |z| = sqrt(x^2 + y^2). The conjugate of z, written z-bar, is x - iy -- the reflection of z across the real axis. Multiplying a complex number by its own conjugate always gives a real, non-negative result: z * z-bar = |z|^2.

1.4 Equality and Square Roots of Complex Numbers

Two complex numbers x + iy and a + ib are equal if and only if their real parts are equal AND their imaginary parts are equal separately: x = a and y = b. This single rule -- comparing real parts to real parts and imaginary parts to imaginary parts -- is the main tool used to solve equations where an unknown complex number appears.

The square root of a complex number z = x + iy is found using the general formula sqrt(x+iy) = +/-[sqrt((|z|+x)/2) + i*sgn(y)*sqrt((|z|-x)/2)], where |z| = sqrt(x^2+y^2) and sgn(y) is +1 if y is positive and -1 if y is negative. This formula comes from assuming the square root equals a+ib, squaring both sides, and matching real and imaginary parts to solve for a and b.

1.5 Complex Polynomials, the Fundamental Theorem of Algebra, and Completing the Square

The Fundamental Theorem of Algebra states that every polynomial of degree n (n >= 1) with complex coefficients has at least one complex root, and therefore has exactly n roots when counted with multiplicity -- meaning the polynomial can always be written as a product of n linear factors. The Rational Root Theorem narrows down which rational numbers could possibly be roots: any rational root must be of the form (a factor of the constant term) divided by (a factor of the leading coefficient), which gives a manageable list of candidates to test.

A quadratic equation $ax^2 + bx + c = 0$ whose discriminant $b^2 - 4ac$ is negative has no real roots, but it always has exactly two roots that are complex conjugates of each other. These roots are found using the same completing-the-square technique used for real quadratics -- the only difference is that the square root of a negative number is now expressed using i instead of being called "no solution".

1.6 Cube Roots and Fourth Roots of Unity

The cube roots of unity are the three solutions of the equation $x^3 = 1$. Rewriting this as $x^3 - 1 = 0$ and factoring using the difference-of-cubes identity gives $(x-1)(x^2+x+1) = 0$. One root is $x = 1$; the quadratic formula applied to $x^2+x+1=0$ gives the other two roots, usually denoted $\omega = (-1+i\sqrt{3})/2$ and $\omega^2 = (-1-i\sqrt{3})/2$.

The cube roots of unity satisfy two key properties that are used constantly to simplify algebraic expressions: $1 + \omega + \omega^2 = 0$ (their sum is zero), and $\omega^3 = 1$ (repeated multiplication cycles back to 1). The fourth roots of unity, solutions of $x^4 = 1$, are found by factoring $x^4 - 1 = (x-1)(x+1)(x^2+1) = 0$, giving the four roots 1, -1, i, and -i.

1.7 Polar Form and Real-World Applications

Any complex number $z = x + iy$ can also be written in polar form as $z = r(\cos(\theta) + i\sin(\theta))$, where $r = |z|$ is the modulus and $\theta = \text{Arg}(z) = \tan^{-1}(y/x)$, adjusted for the quadrant of (x,y), is the argument. Multiplying two complex numbers in polar form multiplies their moduli and ADDS their arguments; dividing them divides their moduli and SUBTRACTS their arguments -- this makes polar form far more efficient than rectangular form whenever several complex numbers need to be multiplied or divided together.

Complex numbers describe real systems, not just abstract algebra. In alternating-current (AC) electrical circuits, impedance $Z = R + iX$ combines a circuit's resistance R and reactance X into a single complex quantity, and Ohm's Law generalizes to $V = I \cdot Z$. In cryptography, a message's numeric encoding can be encrypted by multiplying it by a fixed complex "key", and decrypted by multiplying by the key's multiplicative inverse (equivalently, dividing by the key).

Important Definitions

What is a complex number?

A number of the form z = x + iy, where x and y are real numbers and i = sqrt(-1) is the imaginary unit; x is called the real part and y the imaginary part of z.

What is the imaginary unit i, and what is its key property?

i is defined by i = sqrt(-1), so that i^2 = -1; it is the building block that allows equations with no real solution, such as x^2+1=0, to be solved.

What are Re(z) and Im(z)?

For z = x+iy, Re(z) = x is the real part and Im(z) = y is the imaginary part; both Re(z) and Im(z) are themselves real numbers.

What is the conjugate of a complex number?

For z = x+iy, the conjugate is z-bar = x - iy, obtained by reversing the sign of the imaginary part; geometrically it is the reflection of z across the real axis.

What is the modulus of a complex number?

For z = x+iy, the modulus is |z| = sqrt(x^2+y^2), the distance of the point (x,y) from the origin on the Argand diagram.

What is the argument (Arg z) of a complex number?

The angle theta that the line from the origin to z makes with the positive real axis, given by theta = tan^(-1)(y/x) with the quadrant of (x,y) determining the correct angle, restricted to the principal range (-pi, pi].

What is an Argand diagram?

A coordinate plane used to plot complex numbers, where the horizontal axis represents the real part and the vertical axis represents the imaginary part.

What are the cube roots of unity?

The three solutions of x^3 = 1, namely 1, omega = (-1+i*sqrt(3))/2, and omega^2 = (-1-i*sqrt(3))/2, satisfying 1+omega+omega^2 = 0 and omega^3 = 1.

What is the polar form of a complex number?

z = r(cos(theta) + i*sin(theta)), where r = |z| is the modulus and theta = Arg(z) is the argument of z.

What is impedance in an AC circuit?

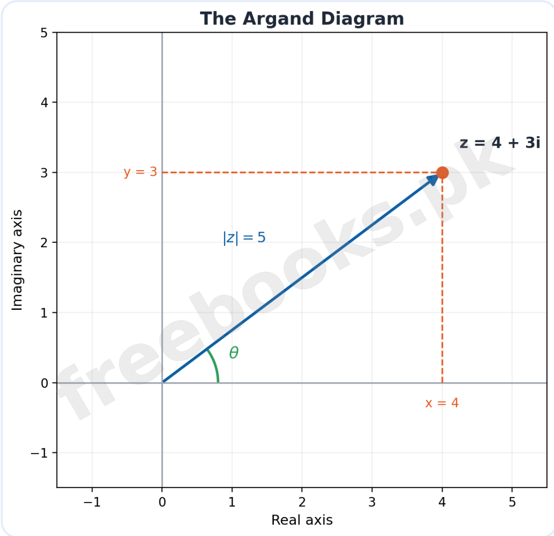
A complex quantity $Z = R + iX$ combining a circuit's resistance R and reactance X into one value, used in the AC form of Ohm's Law, $V = I*Z$.

Key Facts and Relations

Topic	Key Fact / Relation
Imaginary unit	$i^2 = -1$
Complex number	$z = x + iy$, $Re(z) = x$, $Im(z) = y$
Conjugate	z-bar = $x - iy$
Modulus	$\|z\| = sqrt(x^2 + y^2)$
Argument	theta = $tan^(-1)(y/x)$, adjusted for quadrant
Equality	$x + iy = a + ib$ if and only if $x = a$ and $y = b$
Square root formula	$sqrt(x+iy) = +/-[sqrt((\|z\|+x)/2) + i*sgn(y)*sqrt((\|z\|-x)/2)]$
Cube roots of unity	1, omega, omega^2, with 1+omega+omega^2 = 0 and omega^3 = 1
Fourth roots of unity	1, i, -1, -i
Polar form	z = r(cos theta + i sin theta), $r = \|z\|$
Multiplication in polar form	z1*z2 = r1*r2 [cos(theta1+theta2) + i sin(theta1+theta2)]
Division in polar form	z1/z2 = (r1/r2) [cos(theta1-theta2) + i sin(theta1-theta2)]
Ohm's Law for AC circuits	$V = I * Z$

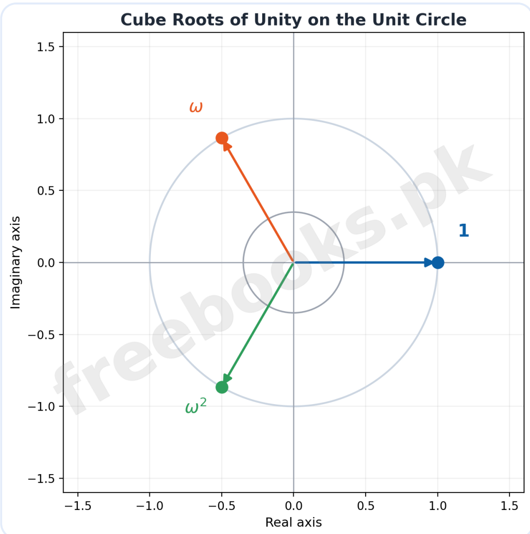
Diagrams

The Argand Diagram: A plotted point $z = 4+3i$ on the complex plane, with the real and imaginary axes, dashed projection lines onto each axis, the modulus shown as the length of the line from the origin to the point, and the argument shown as the angle this line makes with the positive real axis



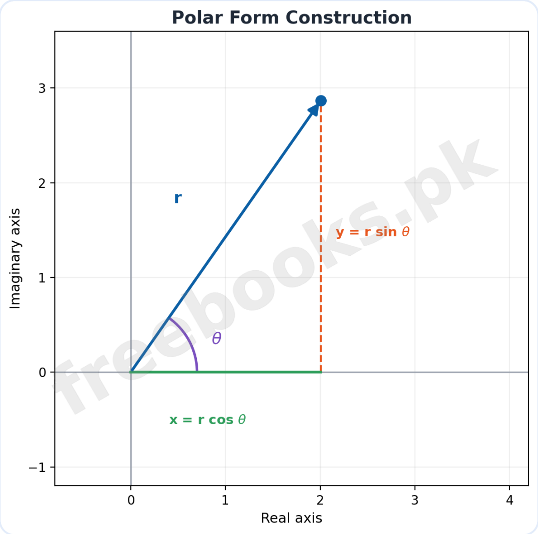
The Argand Diagram

Cube Roots of Unity on the Unit Circle: The three cube roots of unity -- 1, omega, and omega^2 -- plotted as three points spaced exactly 120 degrees apart on the unit circle, illustrating why they sum to zero and why each is a 120-degree rotation of the last



Cube Roots of Unity on the Unit Circle

Polar Form Construction: A right triangle in the complex plane showing how a point at modulus r and argument theta relates to its rectangular coordinates, with x = r*cos(theta) as the horizontal leg and y = r*sin(theta) as the vertical leg



Polar Form Construction

Solved Examples

Example 1: Operations on Complex Numbers

Problem: If z1 = 3+4i and z2 = 2-5i, find (a) z1+z2, (b) z1*z2, and (c) z1/z2.

1. Addition: combine real parts and imaginary parts separately -- z1+z2 = (3+2) + (4-5)i = 5 - i.
2. Multiplication: distribute and simplify using i^2 = -1 -- z1*z2 = (3)(2) + (3)(-5i) + (4i)(2) + (4i)(-5i) = 6 - 15i + 8i - 20i^2 = 6 - 7i + 20 = 26 - 7i.
3. Division: multiply numerator and denominator by the conjugate of the denominator, 2+5i -- z1/z2 = (3+4i)(2+5i) / [(2-5i)(2+5i)] = (6+15i+8i+20i^2) / (4+25) = (-14+23i) / 29.
4. Final answer: z1+z2 = 5-i, z1*z2 = 26-7i, z1/z2 = -14/29 + (23/29)i.

Example 2: Modulus and Argument on the Argand Diagram

Problem: Find the modulus and argument of z = -2+2i.

1. Modulus: |z| = sqrt((-2)^2 + 2^2) = sqrt(4+4) = sqrt(8) = 2*sqrt(2).
2. Since x = -2 is negative and y = 2 is positive, the point lies in the second quadrant.

3. Reference angle: tan^(-1)(2/2) = tan^(-1)(1) = 45 degrees. Because z is in the second quadrant, theta = 180 deg - 45 deg = 135 deg (that is, 3*pi/4 radians).
4. Final answer: |z| = 2*sqrt(2), Arg(z) = 135 degrees, plotted 2*sqrt(2) units from the origin at 135 degrees from the positive real axis.

Example 3: Square Root of a Complex Number

Problem: Find sqrt(3+4i) using the general square-root formula.

1. Here x = 3 and y = 4, so |z| = sqrt(3^2+4^2) = sqrt(25) = 5.
2. Apply sqrt(x+iy) = +/-[sqrt((|z|+x)/2) + i*sgn(y)*sqrt((|z|-x)/2)]: sqrt((5+3)/2) = sqrt(4) = 2, and sqrt((5-3)/2) = sqrt(1) = 1.
3. Since y = 4 is positive, sgn(y) = +1, so sqrt(3+4i) = +/-(2+i).
4. Check by squaring: (2+i)^2 = 4 + 4i + i^2 = 4 + 4i - 1 = 3+4i. Correct.
5. Final answer: sqrt(3+4i) = +/-(2+i).

Example 4: Cube Roots of Unity

Problem: Show that 1+omega+omega^2 = 0 and omega^3 = 1 for omega = (-1+i*sqrt(3))/2, then use this to find the three cube roots of 8.

1. omega = (-1+i*sqrt(3))/2 and omega^2 = (-1-i*sqrt(3))/2 (the other non-real root of x^3=1).
2. Adding: 1 + omega + omega^2 = 1 + (-1+i*sqrt(3))/2 + (-1-i*sqrt(3))/2 = 1 + (-2/2) = 1 - 1 = 0.
3. Multiplying: omega^3 = omega * omega^2 = [(-1)(-1) + (-1)(-i*sqrt(3)) + (i*sqrt(3))(-1) + (i*sqrt(3))(-i*sqrt(3))] / 4 = [1 + i*sqrt(3) - i*sqrt(3) + 3] / 4 = 4/4 = 1.
4. Since 8 = 2^3, and (r*omega^k)^3 = r^3 * omega^(3k) = r^3 for any k, the three cube roots of 8 are 2(1), 2(omega), and 2(omega^2), i.e. 2, -1+i*sqrt(3), and -1-i*sqrt(3).

5. Final answer: the three cube roots of 8 are 2, -1+i*sqrt(3), and -1-i*sqrt(3).

Example 5: Converting to Polar Form

Problem: Express z = -1 + i*sqrt(3) in polar form.

1. Modulus: |z| = sqrt((-1)^2 + (sqrt(3))^2) = sqrt(1+3) = sqrt(4) = 2.
2. x = -1 (negative) and y = sqrt(3) (positive), so z lies in the second quadrant.
3. Reference angle: tan^(-1)(sqrt(3)/1) = 60 degrees. In the second quadrant, theta = 180 deg - 60 deg = 120 degrees (2*pi/3 radians).
4. Final answer: z = 2(cos120 deg + i sin120 deg).

Example 6: Multiplying and Dividing in Polar Form

Problem: If z1 = 6(cos150 deg + i sin150 deg) and z2 = 3(cos30 deg + i sin30 deg), find z1*z2 and z1/z2 in rectangular (x+iy) form.

1. Multiplication: multiply the moduli and add the arguments -- z1*z2 = (6)(3) [cos(150+30) + i sin(150+30)] = 18[cos180 deg + i sin180 deg] = 18(-1 + 0i) = -18.
2. Division: divide the moduli and subtract the arguments -- z1/z2 = (6/3) [cos(150-30) + i sin(150-30)] = 2[cos120 deg + i sin120 deg] = 2(-1/2 + i*sqrt(3)/2) = -1 + i*sqrt(3).
3. Final answer: z1*z2 = -18, z1/z2 = -1 + i*sqrt(3).

Example 7: Application: Current in an AC Circuit

Problem: A circuit has impedance Z = 10(cos40 deg + i sin40 deg) ohms and is supplied with voltage V = 40(cos70 deg + i sin70 deg) volts. Find the current I using V = I*Z.

1. Rearranging Ohm's Law for AC circuits: I = V / Z.

2. Divide the moduli and subtract the arguments: I = (40/10)[cos(70-40) + i sin(70-40)] = 4[cos30 deg + i sin30 deg].
3. Convert to rectangular form: I = 4(sqrt(3)/2 + i/2) = 2*sqrt(3) + 2i.
4. Final answer: I = 4(cos30 deg + i sin30 deg), approximately 3.46 + 2i amperes.

Short Questions & Answers

What is the value of i^2?

i^2 = -1, since i is defined as sqrt(-1).

Define the modulus of a complex number z = x+iy.

|z| = sqrt(x^2+y^2), the distance of the point (x,y) from the origin on the Argand diagram.

How is a complex number divided by another complex number?

Multiply both the numerator and the denominator by the conjugate of the denominator; this turns the denominator into a real number, and the result can then be simplified into x+iy form.

State the four fourth roots of unity.

1, i, -1, and -i -- the four solutions of x^4 = 1.

What is the sum of the three cube roots of unity?

1 + omega + omega^2 = 0.

Write the polar form of a complex number with modulus r and argument theta.

z = r(cos theta + i sin theta).

When two complex numbers in polar form are multiplied, what happens to their moduli and arguments?

The moduli are multiplied together, and the arguments are added together.

Long Questions & Answers

Explain how a complex number is represented in rectangular form and in polar form, and describe with a worked example how to convert a complex number from rectangular to polar form.

What is the rectangular form of a complex number, and what do x and y represent?

$z = x + iy$ is the rectangular (or Cartesian) form of a complex number, where x is the real part and y is the imaginary part. It corresponds directly to plotting the point (x, y) on the Argand diagram, with x measured along the real axis and y measured along the imaginary axis.

How are the modulus and argument of a complex number calculated from its rectangular form?

The modulus is $|z| = \sqrt{x^2+y^2}$, the straight-line distance from the origin to the point (x,y). The argument $\theta = \tan^{-1}(y/x)$, but this basic inverse-tangent value only gives an angle between -90 degrees and 90 degrees, so the quadrant that (x,y) actually lies in must be checked and the angle adjusted accordingly.

What is the polar form of a complex number, and how is it written using the modulus and argument?

Once $r = |z|$ and $\theta = \text{Arg}(z)$ have been found, the complex number can be written in polar form as $z = r(\cos\theta + i\sin\theta)$. This form separates the "size" of the complex number (r) from its "direction" (θ), which is especially useful for multiplication and division.

Work through an example of converting a complex number to polar form.

To convert $z = -1 + i\sqrt{3}$: the modulus is $|z| = \sqrt{1+3} = 2$. Since x is negative and y is positive, z lies in the second quadrant; the reference angle is $\tan^{-1}(\sqrt{3}/1) = 60$ degrees, so $\theta = 180 - 60 = 120$ degrees. Hence $z = 2(\cos 120^\circ + i\sin 120^\circ)$.

Define the cube roots of unity and the fourth roots of unity, showing how each is obtained and stating their key properties.

How are the cube roots of unity obtained from the equation x^3 = 1?

Rewriting x^3 - 1 = 0 and factoring using the difference-of-cubes identity gives (x-1)(x^2+x+1) = 0. Setting each factor to zero gives x = 1 from the first factor, while the quadratic formula applied to x^2+x+1=0 gives the other two, non-real roots.

What are the three cube roots of unity, and how are they usually denoted?

The three roots are 1, omega = (-1+i*sqrt(3))/2, and omega^2 = (-1-i*sqrt(3))/2. Omega is the standard symbol used for the non-real cube root of unity, and the third root is simply its square, omega^2.

What key properties do the cube roots of unity satisfy?

They satisfy 1 + omega + omega^2 = 0, meaning their sum is always zero, and omega^3 = 1, meaning repeated multiplication by omega cycles back to 1 after three steps. These two identities are used constantly to simplify algebraic expressions involving omega without substituting its full value.

How are the fourth roots of unity obtained, and what are they?

Factoring x^4 - 1 = 0 as (x-1)(x+1)(x^2+1) = 0 gives four roots: x = 1 and x = -1 from the first two factors, and x = i and x = -i from x^2+1=0. So the four fourth roots of unity are 1, i, -1, and -i, corresponding to four points spaced 90 degrees apart on the unit circle.

Multiple Choice Questions (MCQs)

i^2 equals:

A. 1

B. -1

C. i

D. -i

Answer: B. -1

Explanation: By definition i = sqrt(-1), so squaring both sides gives i^2 = -1.

The modulus of z = x+iy is:

A. x+y

B. sqrt(x^2+y^2)

C. x^2+y^2

D. x-y

Answer: B. sqrt(x^2+y^2)

Explanation: The modulus is the distance from the origin to the point (x,y): |z| = sqrt(x^2+y^2).

The conjugate of a+bi is:

A. a+bi

B. -a-bi

C. a-bi

D. -a+bi

Answer: C. a-bi

Explanation: The conjugate reverses the sign of the imaginary part only, giving a-bi.

The sum of the three cube roots of unity equals:

A. 1

B. -1

C. 0

D. 3

Answer: C. 0

Explanation: 1 + omega + omega^2 = 0 is a key identity for the cube roots of unity.

The cube roots of unity are usually denoted:

A. 1, i, -i

B. 1, omega, omega^2

C. i, -i, 1

D. -1, 0, 1

Answer: B. 1, omega, omega^2

Explanation: The standard notation for the cube roots of unity is 1, omega, and omega^2.

omega^3 equals:

A. 0

B. omega

C. 1

D. -1

Answer: C. 1

Explanation: Since omega is a cube root of unity, omega^3 = 1 by definition.

The four fourth roots of unity are:

A. 1, -1, i, -i

B. 1, i, -1 only

C. 1, 2, 3, 4

D. i, -i only

Answer: A. 1, -1, i, -i

Explanation: Solving x^4=1 by factoring gives exactly four roots: 1, -1, i, and -i.

The polar form of a complex number with modulus r and argument theta is:

A. z = r + i*theta

B. z = r(cos theta + i sin theta)

C. z = r*theta

D. z = r/theta

Answer: B. z = r(cos theta + i sin theta)

Explanation: Polar form combines the modulus and argument as z = r(cos theta + i sin theta).

When two complex numbers in polar form are multiplied, the result's argument is:

A. the product of the two arguments

B. the sum of the two arguments

C. the difference of the two arguments

D. unchanged

Answer: B. the sum of the two arguments

Explanation: Multiplying complex numbers in polar form multiplies the moduli and ADDS the arguments.

Dividing z1 by z2 in polar form gives an argument equal to:

A. the sum of the two arguments

B. the product of the two arguments

C. the difference of the two arguments

D. the two arguments unchanged

Answer: C. the difference of the two arguments

Explanation: Division in polar form divides the moduli and SUBTRACTS the arguments: z1/z2 = (r1/r2)[cos(theta1-theta2)+i sin(theta1-theta2)].

Quick Revision Summary

- A complex number has the form z = x + iy, with i^2 = -1
- Re(z) = x and Im(z) = y are both real numbers
- The conjugate of z = x+iy is z-bar = x - iy
- Modulus: |z| = sqrt(x^2+y^2); Argument: theta = tan^(-1)(y/x), adjusted for quadrant
- Two complex numbers are equal only when both their real parts and imaginary parts match
- Square root formula: sqrt(x+iy) = +/-[sqrt((|z|+x)/2) + i*sgn(y)*sqrt((|z|-x)/2)]
- The Fundamental Theorem of Algebra guarantees every degree-n polynomial has exactly n complex roots
- Cube roots of unity: 1, omega, omega^2, with 1+omega+omega^2 = 0 and omega^3 = 1
- Fourth roots of unity: 1, i, -1, -i
- Polar form: z = r(cos theta + i sin theta), where r = |z|
- Multiplying in polar form: multiply moduli, add arguments; dividing: divide moduli, subtract arguments
- Complex numbers model real systems -- impedance in AC circuits (V = I*Z) and encryption keys in cryptography

Exam Tips

- Always rationalize the denominator by multiplying by its conjugate when dividing complex numbers
- Remember the principal argument Arg z is restricted to the range (-pi, pi]

- Sketch the Argand diagram whenever a problem involves modulus or argument -- it makes sign and quadrant errors obvious
- Memorize omega = (-1+i*sqrt(3))/2 and 1+omega+omega^2 = 0 -- they turn long cube-root algebra into instant answers
- Convert to polar form before multiplying or dividing several complex numbers together -- it is far less error-prone than expanding everything in rectangular form
- After finding the square root of a complex number, always square your answer back to check it matches the original number

Unit 2: Functions and Graphs

A function is a rule of correspondence that connects two sets so that every input produces exactly one output -- it is one of the most fundamental ideas in all of mathematics, describing how one quantity depends on another. This unit builds a precise understanding of a function's domain, codomain, and range, introduces the notation $y = f(x)$, and explains how to evaluate a function at a number or at an expression. It then covers the vertical line test for recognizing a function from its graph, and the three key types of functions -- one-to-one, onto, and bijective -- along with piecewise functions defined by different rules on different intervals.

The second half of the unit turns to graphs: finding where a line crosses the coordinate axes, where two lines cross each other, and where a line crosses a parabola -- with two, one, or no solutions depending on the situation. It closes with the characteristic shapes of square root and cube root function graphs, and two real-life applications of exponential growth and decay: predicting stock prices and modelling how a pollutant's concentration falls over time.

Learning Objectives

- Define a function, its domain, codomain, and range
- Evaluate a function at given values and expressions using function notation $f(x)$
- Find the domain and range of polynomial, rational, and radical functions
- Apply the vertical line test to determine whether a graph represents a function
- Define and identify one-to-one (injective) functions
- Define and identify onto (surjective) functions
- Define and identify bijective functions
- Define and evaluate piecewise functions
- Find the point(s) of intersection of a linear function with the coordinate axes graphically
- Find the point(s) of intersection of two linear functions, and of a linear and a quadratic function, graphically

- Graph square root and cube root functions and describe their domain and range
- Apply exponential growth and decay models to real-life problems such as stock prices and pollutant concentration

Key Concepts

2.1 Function, Domain, Codomain, and Range

The term function was introduced by the German mathematician Leibniz to describe the dependence of one quantity on another. Formally, a function f from a set X to a set Y is a rule of correspondence that assigns to each element x in X a unique element y in Y. The set X is called the domain of f -- the permitted inputs. The set Y is called the codomain -- the set the outputs are allowed to belong to. The range is the set of values actually produced as output, and the range is always a subset of the codomain (the two can be equal, but the codomain may contain values the function never actually reaches).

The Swiss mathematician Euler introduced the notation y = f(x), read "y equals f of x", to write the statement "y is a function of x". A function can be thought of as a machine that takes an input x, operates on it, and produces exactly one output f(x), called the value of f at x or the image of x under f. Here x is the independent variable and y is the dependent variable. Evaluating a function at a number or an expression is done by substituting that number or expression everywhere the variable appears in the formula.

2.2 Domain and Range of Common Function Types

For a polynomial function, the domain is always all real numbers, since any real number can be substituted into the formula without difficulty. For a rational function (a fraction with the variable in the denominator), every value of x that makes the denominator equal to zero must be excluded from the domain, since division by zero is undefined. For a function involving a square root, the expression under the root sign (the radicand) must be greater than or equal to zero, since the square root of a negative number is not a real number -- this condition is solved as an inequality, often using critical points to divide the number line into regions.

The range of a function is often found by setting y = f(x) and solving the equation for x in terms of y; the range is exactly the set of y-values for which that expression for x is defined (real-valued). For a rational function, this typically means checking which y-value makes the resulting denominator zero and excluding it; for a radical function, it means checking the smallest and largest values the expression under the root can produce as x ranges over the domain.

2.3 The Vertical Line Test and Types of Functions

The vertical line test is a graphical method for checking whether a curve represents a function: a graph represents a function if and only if no vertical line intersects it more than once. This works because a vertical line at a fixed x-value crosses the graph once for every output value paired with that x -- if it crosses more than once, that x-value has more than one output, which violates the very definition of a function.

A function f is one-to-one (injective) if different inputs always produce different outputs, i.e., f(x1) = f(x2) implies x1 = x2. A function f is onto (surjective) if every element of the stated codomain has at least one pre-image in the domain -- for every y in the codomain, there exists an x in the domain with f(x) = y. A function that is both one-to-one and onto is called bijective. A piecewise function is a function defined by different expressions over different intervals of its domain, such as f(x) = 2x+1 for x<0 and f(x) = x^2-1 for x>=0.

2.4 Graphical Intersection of Functions

The point of intersection of two graphs is the point where both graphs meet, and it represents the common solution to both equations at once. The intersection of a linear function with the coordinate axes is found by setting x=0 to get the y-intercept, and setting y=0 to get the x-intercept. The intersection of two linear functions is the single point (if any) where both lines have the same x and y values simultaneously -- found either by plotting both lines and reading off the crossing point, or by solving the two equations together algebraically.

A line and a parabola can intersect at two points, exactly one point (where the line is tangent to the parabola), or not at all, depending on the equations involved -- this can be seen by setting the two expressions for y equal to each other and examining the resulting quadratic equation's discriminant. Sketching the parabola's vertex (found using $h = -b/2a$, $k = f(h)$) alongside the line's intercepts makes it easy to see, and check, how many solutions exist.

2.5 Graphs of Square Root and Cube Root Functions

The square root function y = a*sqrt(x) (and its shifted forms) has domain restricted to x >= 0, since the square root of a negative number is not real; its graph is a curve that starts at a fixed point and rises (or falls) steadily, always staying on one side of the vertical line through its starting point. Its range depends on the sign of a and any vertical shift applied to the function.

The cube root function y = a*cbrt(x) (and its shifted forms) has domain and range equal to all real numbers, because the cube root of a negative number is a real (negative) number, unlike the square root. Its graph is an S-shaped curve that passes through the origin (or through the shifted centre point) and extends in both directions, growing more slowly as |x| increases.

2.6 Real-Life Applications: Growth and Decay

When a quantity increases over time -- money earning interest, a growing population -- it is called growth; when a quantity decreases over time -- a radioactive substance losing strength, a cooling cup of coffee -- it is called decay. The exponential growth model P(t) = P0 * e^(rt) describes quantities like stock prices, where P0 is the initial value, r is the growth rate, and t is time; the time needed for the quantity to double is found by setting 2 = e^(rt) and solving to get t = ln(2)/r.

Decay-type situations are modelled by similar functions, such as a pollutant's concentration C(t) = 100/sqrt(t+1) falling over time; these are solved the same way as any other function equation -- substituting a known time to find the concentration, or setting the concentration to a target value and solving algebraically for the time.

Important Definitions

What is a function?

A rule of correspondence from a set X to a set Y that assigns to each element of X exactly one element of Y.

What is the domain of a function?

The set X of all permitted input values of the function.

What is the codomain of a function?

The set Y in which the function's output values are allowed to lie; it may contain values that are never actually produced by the function.

What is the range of a function?

The set of actual output values produced by the function; the range is always a subset of the codomain.

What does the function notation y = f(x) mean?

A way of writing "y is a function of x", where f(x) denotes the value (or image) of the function f at the input x.

What is a one-to-one (injective) function?

A function in which different inputs always produce different outputs, i.e., f(x1) = f(x2) implies x1 = x2.

What is an onto (surjective) function?

A function in which every element of the codomain has at least one pre-image in the domain.

What is a bijective function?

A function that is both one-to-one and onto.

What is the vertical line test?

A graphical test stating that a curve represents a function if and only if no vertical line intersects it more than once.

What is a piecewise function?

A function defined by different expressions over different intervals of its domain.

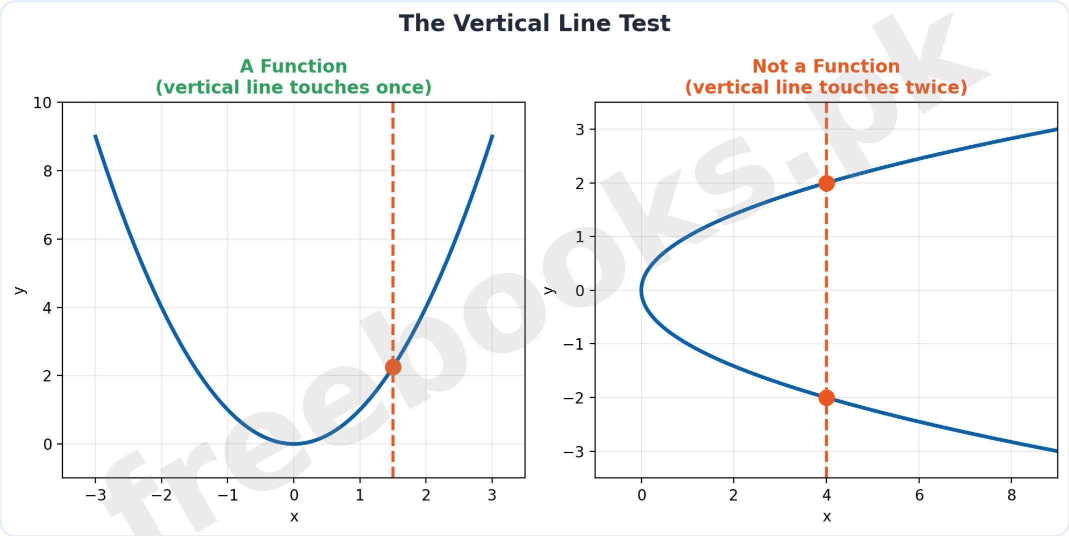
Key Facts and Relations

Topic	Key Fact / Relation
Function notation	y = f(x)
One-to-one condition	f(x1) = f(x2) implies x1 = x2
Onto condition	

Topic	Key Fact / Relation
	For every y in Y, there exists x in X such that f(x) = y
Domain rule (rational function)	Exclude every x that makes the denominator equal to 0
Domain rule (square root function)	The radicand (expression under the root) must be >= 0
Vertex of a parabola y = ax^2+bx+c	h = -b/2a, k = f(h)
Quadratic formula (for x-intercepts)	x = [-b +/- sqrt(b^2-4ac)] / 2a
Square root function y = a*sqrt(x+h)+k	Domain x >= -h; range depends on sign of a and value of k
Cube root function y = a*cbrt(x-h)+k	Domain and range = all real numbers
Exponential growth model	P(t) = P0 * e^(rt)
Doubling time	t = ln(2) / r

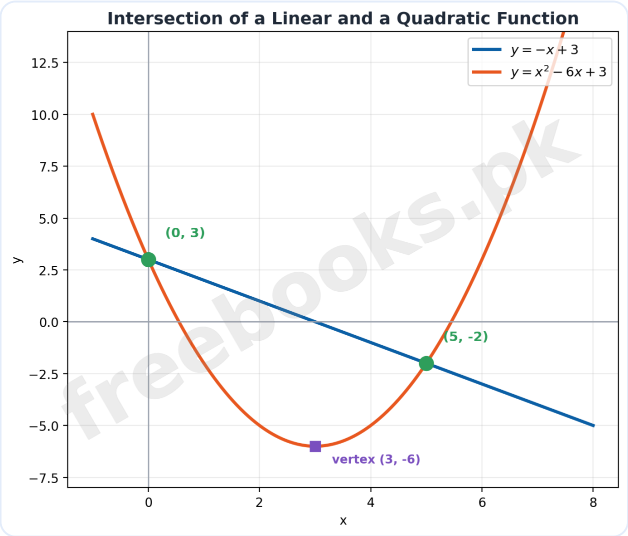
Diagrams

The Vertical Line Test: Two side-by-side graphs: a parabola y = x^2, where a vertical line touches the curve only once (a function), and a sideways parabola x = y^2, where a vertical line touches the curve twice (not a function)



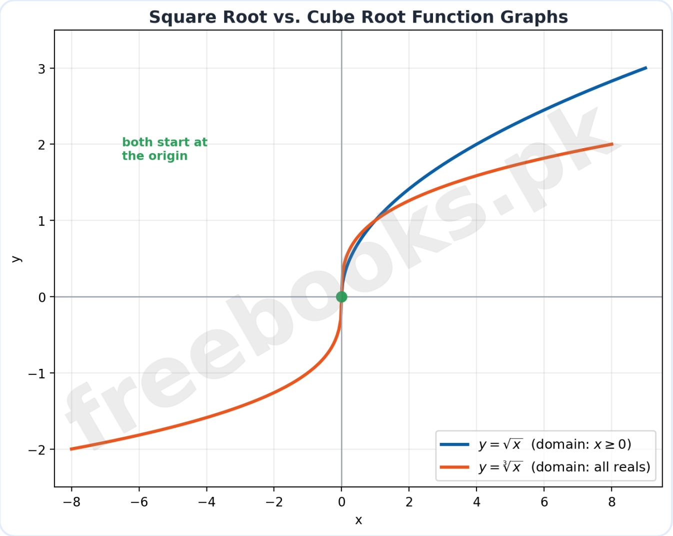
The Vertical Line Test

Intersection of a Linear and a Quadratic Function: The line y = -x+3 and the parabola y = x^2-6x+3 plotted together, with their two intersection points (0,3) and (5,-2) marked, and the parabola's vertex (3,-6) shown



Intersection of a Linear and a Quadratic Function

Square Root vs. Cube Root Function Graphs: The graphs of y = sqrt(x), which only exists for x >= 0, and y = cbrt(x), which exists for all real x, plotted on the same axes to compare their domains and characteristic shapes



Square Root vs. Cube Root Function Graphs

Solved Examples

Example 1: Evaluating a Function

Problem: If f(x) = 2x^2 - 3x + 5, find f(-2), f(3), and f(a+1).

1. f(-2): substitute x=-2 -- f(-2) = 2(-2)^2 - 3(-2) + 5 = 2(4) + 6 + 5 = 8 + 6 + 5 = 19.
2. f(3): substitute x=3 -- f(3) = 2(3)^2 - 3(3) + 5 = 2(9) - 9 + 5 = 18 - 9 + 5 = 14.
3. f(a+1): substitute x=(a+1) everywhere -- f(a+1) = 2(a+1)^2 - 3(a+1) + 5 = 2(a^2+2a+1) - 3a - 3 + 5 = 2a^2+4a+2-3a+2 = 2a^2+a+4.
4. Final answer: f(-2) = 19, f(3) = 14, f(a+1) = 2a^2+a+4.

Example 2: Domain and Range of a Rational Function

Problem: Find the domain and range of f(x) = (x+1)/(x-3).

1. Domain: the denominator x-3 equals 0 when x=3, so x=3 must be excluded. Domain = all real numbers except 3.
2. Range: let y = (x+1)/(x-3). Then y(x-3) = x+1, so xy - 3y = x+1, so x(y-1) = 3y+1, so x = (3y+1)/(y-1).

3. This expression for x is defined for every y except y=1 (which would make the denominator zero).
4. Final answer: Domain = all reals except 3; Range = all reals except 1.

Example 3: Domain and Range of a Square Root Function

Problem: Find the domain and range of f(x) = sqrt(x^2 - 16).

1. The radicand must be >= 0: x^2 - 16 >= 0, so x^2 >= 16, giving x <= -4 or x >= 4.
2. Domain = (-infinity, -4] union [4, infinity).
3. The smallest value the radicand can take is 0, at x = +/-4, which gives y = sqrt(0) = 0. As |x| grows beyond 4, x^2-16 grows without bound, so y grows without bound too.
4. Final answer: Domain = (-infinity,-4] union [4,infinity); Range = [0, infinity).

Example 4: Proving a Function Is Bijective

Problem: Show that f(x) = 3x - 5, with domain and codomain both the real numbers, is bijective.

1. One-to-one: assume f(x1) = f(x2). Then 3x1-5 = 3x2-5, so 3x1 = 3x2, so x1 = x2. So f is one-to-one.
2. Onto: for any real number y, solve y = 3x-5 for x, giving x = (y+5)/3, which is a real number for every real y, and f((y+5)/3) = 3*(y+5)/3 - 5 = y+5-5 = y.
3. Since f is both one-to-one and onto, f is bijective.
4. Final answer: f(x) = 3x-5 is bijective.

Example 5: Intersection with the Coordinate Axes

Problem: Find where the line y = -3x + 9 crosses the x-axis and the y-axis.

1. y-intercept: set x=0 -- y = -3(0)+9 = 9, giving the point (0, 9).
2. x-intercept: set y=0 -- 0 = -3x+9, so 3x=9, so x=3, giving the point (3, 0).
3. Final answer: the line crosses the axes at (0, 9) and (3, 0).

Example 6: Intersection of a Linear and a Quadratic Function

Problem: Find the points of intersection of y = x+1 and y = x^2-x-3.

1. Set the two expressions for y equal: x+1 = x^2-x-3.
2. Rearranging: 0 = x^2-2x-4, so x = [2 +/- sqrt(4+16)]/2 = [2 +/- sqrt(20)]/2 = 1 +/- sqrt(5).
3. Since sqrt(5) is approximately 2.236, x is approximately 3.236 or -1.236.
4. Substituting back into y=x+1: y is approximately 4.236 or -0.236.
5. Final answer: the graphs intersect at approximately (3.24, 4.24) and (-1.24, -0.24).

Example 7: Application: Exponential Growth

Problem: A bacteria population grows according to P(t) = P0*e^(rt), starting at P0 = 2000 with growth rate r = 8% per hour. Find the population after 5 hours, and the time for the population to double.

1. Population after 5 hours: P(5) = 2000*e^(0.08*5) = 2000*e^0.4. Using e^0.4 is approximately 1.4918, P(5) is approximately 2000*1.4918 = 2983.6, about 2984 bacteria.
2. Doubling time: set 2*P0 = P0*e^(rt), so 2 = e^(rt), so ln(2) = rt, so t = ln(2)/r.
3. t = 0.6931/0.08, which is approximately 8.66 hours.

4. Final answer: population after 5 hours is approximately 2984; doubling time is approximately 8.66 hours.

Short Questions & Answers

What is the domain of a function?

The set of all permitted input values of the function.

What is the difference between range and codomain?

The range is the actual set of outputs the function produces; the codomain is the set the outputs are allowed to belong to. The range is always a subset of the codomain.

State the vertical line test.

A curve represents a function if and only if no vertical line crosses it more than once.

When is a function called one-to-one?

When different inputs always give different outputs, i.e., $f(x1)=f(x2)$ implies $x1=x2$.

When is a function called onto?

When every element of the codomain has at least one pre-image in the domain.

What is a piecewise function?

A function defined by different expressions over different intervals of its domain.

What is the domain restriction for a square root function?

The expression under the square root (the radicand) must be greater than or equal to zero.

Long Questions & Answers

Define a function together with its domain, codomain, and range, and explain how the domain and range of a rational function are determined, using a worked example.

What is a function, and what are its domain, codomain, and range?

A function f from a set X to a set Y is a rule of correspondence that assigns to every element x in X exactly one element y in Y. X is the domain (the permitted inputs), Y is the codomain (the set the outputs are allowed to belong to), and the range is the actual subset of Y that is produced as output -- the range is always a subset of the codomain, though the two can be equal.

How is the domain of a rational function determined?

The domain of a rational function excludes every value of x that makes the denominator equal to zero, since division by zero is undefined; all other real numbers are permitted inputs.

How is the range of a rational function typically found?

Set y = f(x), solve the resulting equation for x in terms of y, and determine which values of y make that expression for x defined; the range is exactly the set of such y values.

Work through an example of finding the domain and range of a rational function.

For $f(x) = (x+1)/(x-3)$: the denominator is zero at $x=3$, so the domain is all real numbers except 3. Setting $y=(x+1)/(x-3)$ and solving gives $x=(3y+1)/(y-1)$, which is defined for every y except $y=1$, so the range is all real numbers except 1.

Explain the vertical line test and the three key types of functions -- one-to-one, onto, and bijective -- with the reasoning behind each definition.

What is the vertical line test, and what does it check?

The vertical line test says a graph represents a function if and only if no vertical line intersects it more than once; a vertical line at a fixed x-value crosses the graph once for each output value that corresponds to that x, so more than one crossing means that x-value has more than one output, which violates the definition of a function.

What makes a function one-to-one (injective)?

A function is one-to-one if different inputs always produce different outputs -- formally, $f(x1)=f(x2)$ implies $x1=x2$. Geometrically, a one-to-one function's graph passes a horizontal line test: no horizontal line crosses it more than once.

What makes a function onto (surjective)?

A function is onto if every element of the stated codomain is actually produced as an output by some input -- that is, for every y in the codomain, there exists at least one x in the domain with $f(x)=y$. This depends on how the codomain is defined, not just on the formula for f.

What makes a function bijective, and why does that matter?

A function is bijective if it is both one-to-one and onto. Bijective functions are important because they have a well-defined inverse function -- every output corresponds to exactly one input, so the process can be reversed uniquely.

Multiple Choice Questions (MCQs)

The domain of a function is:

A. The set of outputs
B. The set of permitted inputs
C. The graph of the function
D. The codomain

Answer: B. The set of permitted inputs

Explanation: The domain is the set X of all permitted input values of a function.

The range of a function is always:

A. Equal to the codomain

B. A subset of the codomain

C. Larger than the codomain

D. Unrelated to the codomain

Answer: B. A subset of the codomain

Explanation: The range is the actual set of outputs produced, and it is always a subset of the codomain.

A function is one-to-one if:

A. $f(x1)=f(x2)$ implies $x1=x2$

B. Every output has a pre-image

C. The graph is a straight line

D. The domain equals the range

Answer: A. $f(x1)=f(x2)$ implies $x1=x2$

Explanation: One-to-one (injective) means different inputs always give different outputs: $f(x1)=f(x2)$ implies $x1=x2$.

A function is onto if:

A. It passes the vertical line test

B. Every element of the codomain has a pre-image

C. It is one-to-one

D. Its domain is all real numbers

Answer: B. Every element of the codomain has a pre-image

Explanation: Onto (surjective) means every element of the codomain is actually produced as an output by some input.

A bijective function is:

A. Only one-to-one

B. Only onto

C. Both one-to-one and onto

D. Neither one-to-one nor onto

Answer: C. Both one-to-one and onto

Explanation: A bijective function satisfies both the one-to-one and onto conditions.

The vertical line test checks whether:

A. A graph is a function

B. A graph is one-to-one

C. A graph is onto

D. A graph is bijective

Answer: A. A graph is a function

Explanation: The vertical line test determines whether a curve represents a function at all, not whether it is one-to-one or onto.

The domain of f(x) = sqrt(x-5) is:

A. x <= 5

B. x >= 5

C. All real numbers

D. x is not equal to 5

Answer: B. x >= 5

Explanation: The radicand x-5 must be >= 0, so x >= 5.

The graph of a cube root function y = cbrt(x) has domain and range:

A. x >= 0 only

B. Both all real numbers

C. y >= 0 only

D. Both restricted to positive numbers

Answer: B. Both all real numbers

Explanation: Unlike the square root function, the cube root function is defined for negative inputs too, so both its domain and range are all real numbers.

In the exponential growth model P(t) = P0*e^(rt), the doubling time is found using:

A. t = 2/r

B. t = ln(2)/r

C. t = r/ln(2)

D. t = 2r

Answer: B. t = ln(2)/r

Explanation: Setting 2 = e^(rt) and solving gives t = ln(2)/r.

A piecewise function is:

A. A function with only one formula

B. A function defined by different expressions on different intervals

C. A function with no domain restrictions

D. A function that is always one-to-one

Answer: B. A function defined by different expressions on different intervals

Explanation: A piecewise function uses different formulas over different parts of its domain.

Quick Revision Summary

- A function assigns exactly one output to every input in its domain
- Domain = permitted inputs; Codomain = set outputs are allowed to belong to; Range = actual outputs produced (range is a subset of codomain)
- Function notation: y = f(x); f(x) is the value/image of f at x
- Vertical line test: a graph is a function if and only if no vertical line crosses it more than once
- One-to-one: f(x1)=f(x2) implies x1=x2
- Onto: every element of the codomain has at least one pre-image
- Bijective: both one-to-one and onto
- Domain of a rational function excludes values that make the denominator zero
- Domain of a square-root function requires the radicand to be >= 0
- Cube root functions have domain and range equal to all real numbers
- Linear-quadratic intersections can have two, one, or no solutions depending on the discriminant
- Exponential growth/decay models: P(t) = P0*e^(rt); doubling time t = ln(2)/r

Exam Tips

- Always check the denominator for zero and the radicand for negativity before stating a function's domain
- To prove a function is one-to-one, start by assuming f(x1)=f(x2) and show algebraically that x1 must equal x2
- To prove a function is onto, solve y=f(x) for x and confirm a valid real x exists for every y in the codomain
- Sketch intercepts and the vertex before graphing a linear-quadratic intersection -- it makes the number of solutions visually obvious
- Remember that square root graphs only exist for x >= some value, while cube root graphs exist for all real x

- In growth/decay problems, take the natural log of both sides to solve for time in the exponent

Unit 3: Theory of Quadratic Functions

A quadratic function f(x) = ax^2+bx+c graphs as a parabola, and this unit develops the tools needed to fully analyse that parabola: sketching it from its intercepts and vertex, and finding its exact maximum or minimum value by completing the square. It then tackles a subtlety that trips many students up -- a quadratic function is not one-to-one over its whole domain, so finding its inverse requires first restricting the domain to one side of the vertex.

The unit then moves into absolute value quadratic equations and inequalities, rational and radical equations that reduce to a quadratic equation once cleared of fractions or radicals, and closes with real-world word problems -- from room dimensions to braking distances -- that translate naturally into a quadratic equation or inequality once the given conditions are written symbolically.

Learning Objectives

- Define a quadratic function and identify its standard form
- Sketch a quadratic function's graph and identify its vertex, x-intercepts, and y-intercept
- Determine whether a quadratic function's vertex is a maximum or minimum point
- Find the maximum or minimum value of a quadratic function by completing the square
- Find the inverse of a quadratic function on a restricted domain, and determine its domain and range
- Solve absolute value quadratic equations, checking for extraneous roots
- Solve absolute value quadratic inequalities using critical values and test points
- Solve rational equations that reduce to a quadratic equation
- Solve radical equations that reduce to a quadratic equation, checking for extraneous roots
- Apply quadratic equations and inequalities to solve real-world word problems

Key Concepts

3.1 The Quadratic Function and Its Graph

A quadratic function is a polynomial function of degree two, written in standard form as $f(x) = ax^2+bx+c$, where a, b, c are real numbers and a is not equal to 0. Its graph is a parabola that opens upward if $a>0$ and downward if $a<0$. The tip of the parabola, called the vertex and labelled (h, k), is a turning point lying on the parabola's vertical axis of symmetry -- it is a minimum point when $a>0$ and a maximum point when $a<0$.

To sketch a quadratic function, find its y-intercept (set $x=0$), its x-intercepts (solve $f(x)=0$), and its vertex ($h=-b/2a$, $k=f(h)$). Once these are plotted, the direction the parabola opens shows immediately whether the function is increasing or decreasing on either side of the vertex.

3.2 Maximum and Minimum Values by Completing the Square

Completing the square rewrites $f(x)=ax^2+bx+c$ in vertex form, $f(x) = a(x-h)^2+k$, where $h=-b/2a$ and $k=c-b^2/4a$. This is done by factoring a out of the x^2 and x terms, adding and subtracting the square of half the new coefficient of x to form a perfect square trinomial, and simplifying.

In vertex form, the squared term $a(x-h)^2$ is always zero (at $x=h$) or has the same sign as a everywhere else. So if $a>0$, the smallest value f(x) can take is k, giving a minimum value of k at $x=h$; if $a<0$, the largest value f(x) can take is k, giving a maximum value of k at $x=h$ -- this finds the exact maximum or minimum without needing to sketch the graph first.

3.3 Inverse of a Quadratic Function

A quadratic function is generally not one-to-one over its entire domain, since two different x-values on opposite sides of the vertex can produce the same output -- it fails the horizontal line test. To find an inverse, the domain must first be restricted to one side of the vertex, commonly $x>=h$ or $x<=h$, which makes the function one-to-one on that restricted domain.

To find the inverse formula, swap x and y in $y=f(x)$, then solve the resulting equation for y using the quadratic formula -- this produces a "plus" branch and a "minus" branch. Since only one of these branches actually maps back onto the correct restricted domain, the domain and range of the original restricted function are used (often by testing a value) to identify and keep the correct branch, discarding the other as extraneous.

3.4 Absolute Value Quadratic Equations and Inequalities

The absolute value of x is defined as $|x|=x$ when $x>=0$ and $|x|=-x$ when $x<0$. An absolute value quadratic equation $|ax^2+bx+c|=d$ is solved by splitting it into two ordinary equations, $ax^2+bx+c=d$ and $ax^2+bx+c=-d$, solving both, and then substituting every candidate answer back into the original equation to discard any extraneous root.

An absolute value quadratic inequality such as $|ax^2+bx+c|<d$ is rewritten as the compound inequality $-d<ax^2+bx+c<d$, which splits into two separate quadratic inequalities to be solved using critical values and test points; the final solution set is the INTERSECTION of both regions' solutions. For $|ax^2+bx+c|>d$, the inequality splits into $ax^2+bx+c<-d$ OR $ax^2+bx+c>d$ instead, and the final solution set is the UNION of the two.

3.5 Rational and Radical Equations Reducible to a Quadratic Equation

A rational equation contains one or more rational expressions with the variable in the denominator. Multiplying every term of the equation by the least common denominator (LCD) clears all the fractions at once, typically leaving a quadratic equation that is solved by the usual methods -- while remembering that any value which was originally forbidden (because it made a denominator zero) must still be excluded from the final solution set.

A radical equation contains the variable under a root sign. It is solved by isolating a radical on one side and squaring both sides, repeating the process if a second radical remains, until a polynomial (usually quadratic) equation results. Because squaring can introduce solutions that do not satisfy the original equation, every candidate solution must be substituted back into the ORIGINAL radical equation, and any extraneous root discarded.

3.6 Real-World Applications

Many word problems translate directly into a quadratic equation: assign a variable to the unknown quantity, translate the given conditions into symbols to form an equation, solve it using the usual methods, and reject any solution that does not make physical sense -- such as a negative length or a negative amount of time.

Real-world quadratic inequalities work the same way: a threshold condition (such as a maximum safe braking distance, or a minimum required height) is set up as an inequality using the given function, solved for the variable using critical values and test points, and the resulting interval is interpreted in the context of the original problem.

Important Definitions

What is a quadratic function?

A polynomial function of degree two, written $f(x)=ax^2+bx+c$ with a, b, c real numbers and a not equal to 0.

What is the vertex of a parabola?

The turning point (h,k) of the parabola, lying on its axis of symmetry; it is a minimum point if $a>0$ and a maximum point if $a<0$.

What is the vertex form of a quadratic function?

$f(x) = a(x-h)^2+k$, obtained by completing the square, where (h,k) is the vertex.

What is the axis of symmetry of a parabola?

The vertical line $x=h$ passing through the vertex, about which the parabola is symmetric.

Why must the domain of a quadratic function be restricted to find its inverse?

Because a quadratic function is not one-to-one over its whole domain; restricting the domain to $x>=h$ or $x<=h$ makes it one-to-one and therefore invertible.

What is the absolute value of x?

$|x| = x$ if x is greater than or equal to 0, and $|x| = -x$ if x is less than 0.

What is an extraneous root?

A value obtained while solving an equation (for example after squaring, or from an absolute-value split) that satisfies the transformed equation but not the original equation, and must be rejected.

What is a rational equation?

An equation containing one or more rational expressions with the variable in the denominator.

What is a radical equation?

An equation in which the variable appears under a root (radical) sign.

How is a quadratic inequality's solution set typically found?

By finding the critical values (the roots of the corresponding equation), using them to divide the number line into regions, and testing a point from each region in the original inequality.

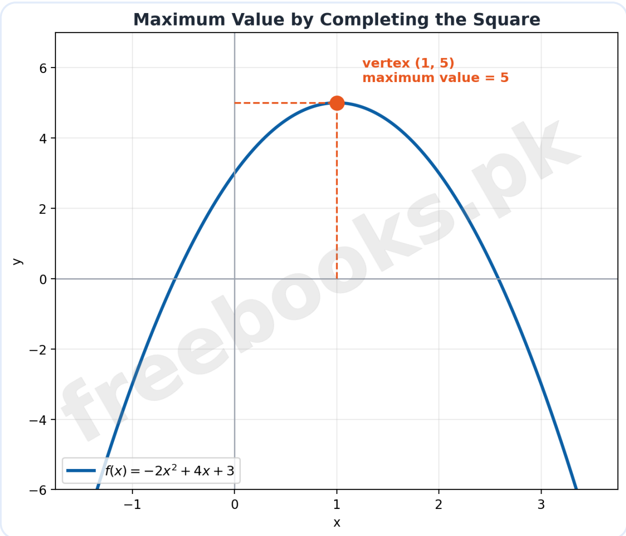
Key Facts and Relations

Topic	Key Fact / Relation
Standard form	f(x) = ax^2+bx+c, a not equal to 0
Vertex coordinates	h = -b/2a, k = f(h) = c - b^2/4a
Vertex form	f(x) = a(x-h)^2+k
Max/Min rule	a>0: minimum value k at x=h; a<0: maximum value k at x=h
Inverse domain restriction	Restrict to x>=h or x<=h before inverting
Absolute value definition	|x|=x (x>=0), |x|=-x (x<0)
Absolute value equation split	|E|=d => E=d or E=-d
Absolute value inequality (< form)	|E|<d => -d<E<d
Absolute value inequality (> form)	|E|>d => E<-d or E>d

Topic	Key Fact / Relation
Rational equation method	Multiply every term by the LCD to clear denominators
Radical equation method	Isolate a radical, square both sides, repeat if needed, then check for extraneous roots

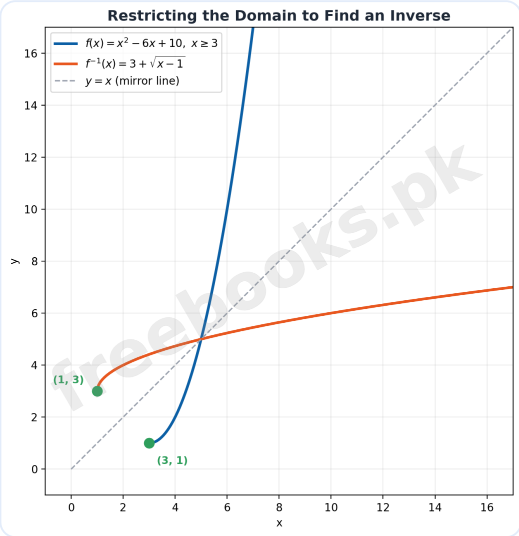
Diagrams

Maximum Value by Completing the Square: The graph of f(x) = -2x^2+4x+3, with its vertex at (1, 5) marked and dashed lines showing that the maximum value of the function is 5, occurring at x=1



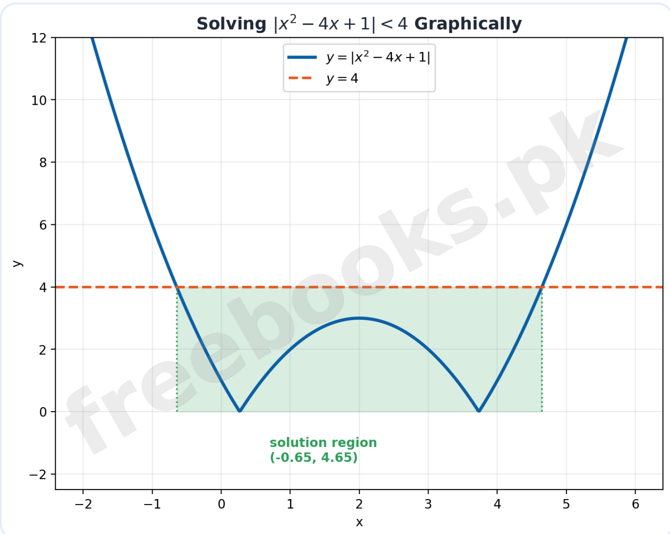
Maximum Value by Completing the Square

Restricting the Domain to Find an Inverse: The graph of f(x) = x^2-6x+10 restricted to x>=3, together with its inverse f^-1(x) = 3+sqrt(x-1) and the reference line y=x, showing the two graphs as mirror images of each other



Restricting the Domain to Find an Inverse

Solving an Absolute Value Quadratic Inequality Graphically: The graph of y = |x^2-4x+1| together with the horizontal line y=4, with the solution region where the curve lies below the line shaded and its endpoints marked



Solving an Absolute Value Quadratic Inequality Graphically

Solved Examples

Example 1: Sketching and Analyzing a Quadratic Function

Problem: Sketch and analyze y = x^2 - 4x - 5.

1. y-intercept: set x=0 -- y = 0-0-5 = -5, giving the point (0, -5).

2. x-intercepts: solve x^2-4x-5=0. Factoring gives (x-5)(x+1)=0, so x=5 or x=-1.
3. Vertex: h = -b/2a = -(-4)/2(1) = 2. k = (2)^2-4(2)-5 = 4-8-5 = -9. Vertex = (2, -9).
4. Since a=1>0, the vertex is a minimum; the function is decreasing on (-infinity, 2) and increasing on (2, infinity).
5. Final answer: vertex (2,-9) is a minimum point; x-intercepts (5,0) and (-1,0); y-intercept (0,-5).

Example 2: Maximum or Minimum Value by Completing the Square

Problem: Find the maximum or minimum value of f(x) = 3x^2 - 12x + 7 by completing the square.

1. Factor 3 out of the x-terms: f(x) = 3(x^2-4x) + 7.
2. Add and subtract (half of -4, squared) = 4 inside the brackets: f(x) = 3(x^2-4x+4-4) + 7 = 3[(x-2)^2-4] + 7.
3. Simplify: f(x) = 3(x-2)^2 - 12 + 7 = 3(x-2)^2 - 5.
4. Since a=3>0, the minimum value is -5, occurring at x=2.
5. Final answer: minimum value -5 at x=2.

Example 3: Finding the Inverse of a Quadratic Function

Problem: Find the inverse of f(x) = x^2 - 6x + 10, x >= 3, and state its domain and range.

1. Complete the square: f(x) = (x-3)^2 + 1. Domain f = [3, infinity); minimum value 1 at x=3, so Range f = [1, infinity).
2. Let y = x^2-6x+10. Swap x and y: x = y^2-6y+10, so y^2-6y+(10-x) = 0.
3. Apply the quadratic formula: y = [6 +/- sqrt(36-4(10-x))]/2 = [6 +/- sqrt(4x-4)]/2 = 3 +/- sqrt(x-1).
4. Since the original domain requires x>=3, the inverse must output values >=3, so the '+' branch is correct: f^-1(x) = 3 + sqrt(x-1).

5. Final answer: f^-1(x) = 3+sqrt(x-1); domain of f^-1 = [1,infinity); range of f^-1 = [3,infinity).

Example 4: Solving an Absolute Value Quadratic Equation

Problem: Solve |x^2 - 9| = 7.

1. Split into two equations: x^2-9 = 7 or x^2-9 = -7.
2. First equation: x^2 = 16, so x = +/-4.
3. Second equation: x^2 = 2, so x = +/- sqrt(2).
4. Check all four values in the original equation: |16-9|=7 (true for x=+/-4); |2-9|=|-7|=7 (true for x=+/-sqrt(2)). No extraneous roots.
5. Final answer: solution set = {-4, -sqrt(2), sqrt(2), 4}.

Example 5: Solving an Absolute Value Quadratic Inequality

Problem: Solve |x^2 - 4x + 1| < 4.

1. Rewrite as a compound inequality: -4 < x^2-4x+1 < 4.
2. Left part: x^2-4x+1 > -4, i.e. x^2-4x+5 > 0. Its discriminant is 16-20=-4<0, and since the leading coefficient is positive, this is TRUE for every real x.
3. Right part: x^2-4x+1 < 4, i.e. x^2-4x-3 < 0. Its roots are x = [4 +/- sqrt(16+12)]/2 = 2 +/- sqrt(7). Since the parabola opens upward, the inequality holds between the roots: (2-sqrt(7), 2+sqrt(7)).
4. Since the left part holds everywhere, the final solution is just the right part's interval.
5. Final answer: solution set = (2-sqrt(7), 2+sqrt(7)), approximately (-0.65, 4.65).

Example 6: Solving a Radical Equation Reducible to a Quadratic

Problem: Solve sqrt(x+5) - sqrt(x-3) = 2.

1. Isolate one radical: sqrt(x+5) = 2 + sqrt(x-3).
2. Square both sides: x+5 = 4 + 4*sqrt(x-3) + (x-3) = x+1+4*sqrt(x-3).
3. Simplify: x+5-x-1 = 4*sqrt(x-3), so 4 = 4*sqrt(x-3), so 1 = sqrt(x-3).
4. Square again: 1 = x-3, so x = 4.
5. Check in the original equation: sqrt(4+5)-sqrt(4-3) = sqrt(9)-sqrt(1) = 3-1 = 2. Correct -- no extraneous root.
6. Final answer: x = 4.

Example 7: Application: A Real-World Word Problem

Problem: The length of a rectangular garden is 4 metres more than its width. If the area of the garden is 96 square metres, find its length and width.

1. Let the width = x metres, so the length = x+4 metres.
2. Area condition: x(x+4) = 96, so x^2+4x-96 = 0.
3. Factor: (x+12)(x-8) = 0, so x = -12 or x = 8.
4. Since width cannot be negative, reject x=-12; take x=8. Then length = x+4 = 12.
5. Final answer: width = 8 metres, length = 12 metres.

Short Questions & Answers

What is the standard form of a quadratic function?

f(x) = ax^2+bx+c, where a, b, c are real numbers and a is not equal to 0.

When is the vertex of a parabola a maximum point?

When the leading coefficient a is negative (a<0).

State the vertex form of a quadratic function.

f(x) = a(x-h)^2+k, where (h,k) is the vertex.

Why must a quadratic function's domain be restricted before finding its inverse?

Because a quadratic function is not one-to-one over its whole domain.

How is |x^2-9|=7 split into two equations?

x^2-9 = 7 or x^2-9 = -7.

How is a rational equation typically solved?

By multiplying every term by the least common denominator (LCD) to clear the fractions, then solving the resulting polynomial equation.

Why must solutions of a radical equation be checked in the original equation?

Because squaring both sides can introduce extraneous roots that satisfy the squared equation but not the original one.

Long Questions & Answers

Explain how completing the square is used to find the maximum or minimum value of a quadratic function, and how the same idea is used to find its inverse on a restricted domain.

How does completing the square convert f(x)=ax^2+bx+c into vertex form?

Factor a out of the x^2 and x terms, add and subtract the square of half the new coefficient of x inside the brackets to form a perfect square trinomial, then simplify -- this produces f(x)=a(x-h)^2+k, where h=-b/2a and k=c-b^2/4a.

How does the vertex form reveal the maximum or minimum value?

In f(x)=a(x-h)^2+k, the squared term a(x-h)^2 is always zero or has the same sign as a; if a>0 the smallest possible value of f(x) is k (a minimum, at x=h), and if a<0 the largest possible value of f(x) is k (a maximum, at x=h).

Why is a quadratic function's domain restricted before finding its inverse?

A quadratic function fails the horizontal line test over its full domain -- two different x-values on opposite sides of the vertex can produce the same output --

so it is not one-to-one and has no inverse function unless its domain is restricted to one side of the vertex, x>=h or x<=h.

How is the correct branch of the inverse chosen after restricting the domain?

Solving the swapped equation with the quadratic formula produces a '+' branch and a '-' branch; only one of these branches actually maps the range of the restricted original function back onto its correct restricted domain, so a test value (or reasoning about which branch stays within the required domain) is used to discard the other, extraneous branch.

Explain the process for solving absolute value quadratic equations and inequalities, and describe how rational and radical equations are reduced to quadratic equations.

How is an absolute value quadratic equation |E|=d solved?

It is split into two ordinary equations, E=d and E=-d, both solved using standard quadratic methods; every resulting candidate value must then be substituted back into the original absolute value equation to confirm it is not extraneous.

How is an absolute value quadratic inequality such as |E|<d solved?

It is rewritten as the compound inequality -d<E<d, which splits into two separate quadratic inequalities; each is solved by finding its critical values and testing a point in each resulting region, and the final solution set is the intersection of both regions' solutions.

How is a rational equation reduced to a quadratic equation?

Every term is multiplied by the least common denominator of all the fractions present, which clears the denominators and typically leaves a quadratic (or lower-degree) polynomial equation, solved by the usual methods -- while remembering to exclude any value that was originally forbidden because it made a denominator zero.

How is a radical equation reduced to a quadratic equation, and what must be checked afterward?

A radical is isolated on one side and both sides are squared to remove it, repeating if a second radical remains, until a polynomial equation results; because squaring can introduce solutions that do not satisfy the original radical equation, every candidate solution must be substituted back into the ORIGINAL equation, and any extraneous root must be discarded.

Multiple Choice Questions (MCQs)

The standard form of a quadratic function is:

A. $f(x)=ax+b$

B. $f(x)=ax^2+bx+c$, a not equal to 0

C. $f(x)=ax^3+bx+c$

D. $f(x)=a/x+b$

Answer: B. $f(x)=ax^2+bx+c$, a not equal to 0

Explanation: A quadratic function is a degree-two polynomial function, $f(x)=ax^2+bx+c$ with a not equal to 0.

In vertex form $f(x)=a(x-h)^2+k$, if $a<0$ the vertex is a:

A. Minimum point

B. Maximum point

C. x-intercept

D. y-intercept

Answer: B. Maximum point

Explanation: When $a<0$, the parabola opens downward, so the vertex is the highest point -- a maximum.

The vertex coordinates of a parabola $y=ax^2+bx+c$ are given by:

A. $h=-b/2a$, $k=f(h)$

B. $h=b/2a$, $k=f(h)$

C. $h=-c/2a$, $k=f(b)$

D. $h=2a$, $k=b$

Answer: A. $h=-b/2a$, $k=f(h)$

Explanation: The vertex is found using $h=-b/2a$, then substituting back to get $k=f(h)$.

A quadratic function must have its domain restricted before finding its inverse because:

A. It is always negative

B. It is not one-to-one over its full domain

C. It has no y-intercept

D. It is not onto

Answer: B. It is not one-to-one over its full domain

Explanation: A quadratic function fails the horizontal line test over its whole domain, so it is not one-to-one until the domain is restricted.

$|x^2-9|=7$ splits into:

A. $x^2-9=7$ only

B. $x^2-9=7$ or $x^2-9=-7$

C. $x^2=7$ only

D. $x^2-9=-7$ only

Answer: B. $x^2-9=7$ or $x^2-9=-7$

Explanation: An absolute value equation $|E|=d$ splits into $E=d$ or $E=-d$.

The inequality $|E|<d$ is equivalent to:

A. $E<d$ only

B. $E>-d$ only

C. $-d<E<d$

D. $E<-d$ or $E>d$

Answer: C. $-d<E<d$

Explanation: $|E|<d$ means E lies strictly between -d and d, i.e. $-d<E<d$.

To clear denominators in a rational equation, you should:

A. Add the denominators

B. Multiply every term by the LCD

C. Ignore the denominators

D. Square both sides

Answer: B. Multiply every term by the LCD

Explanation: Multiplying every term by the least common denominator (LCD) clears all fractions at once.

Solving a radical equation by squaring both sides can introduce:

A. Extra domain restrictions

B. Extraneous roots

C. Negative coefficients

D. Imaginary vertices

Answer: B. Extraneous roots

Explanation: Squaring both sides of an equation can produce extraneous roots -- values that satisfy the squared equation but not the original.

In a word problem where x represents a length, a negative solution for x should be:

A. Accepted as valid

B. Rejected as not physically meaningful

C. Doubled

D. Rounded to zero

Answer: B. Rejected as not physically meaningful

Explanation: A length cannot be negative, so any negative solution is rejected as not physically meaningful.

The maximum value of $f(x) = -2(x-1)^2+5$ is:

A. 1

B. 2

C. 5

D. -2

Answer: C. 5

Explanation: In vertex form $f(x)=a(x-h)^2+k$ with $a<0$, the maximum value is k, which here is 5.

Quick Revision Summary

- Standard form: $f(x)=ax^2+bx+c$, a not equal to 0; the graph is a parabola
- Vertex (h,k): $h=-b/2a$, $k=f(h)$; minimum if $a>0$, maximum if $a<0$
- Vertex form: $f(x)=a(x-h)^2+k$, found by completing the square
- A quadratic function's domain must be restricted ($x>=h$ or $x<=h$) before it has an inverse
- $|x|=x$ for $x>=0$, $|x|=-x$ for $x<0$
- $|E|=d$ splits into $E=d$ or $E=-d$; always check for extraneous roots
- $|E|<d$ splits into $-d<E<d$; $|E|>d$ splits into $E<-d$ or $E>d$
- Rational equations are cleared of fractions by multiplying through by the LCD

- Radical equations are solved by isolating a radical, squaring, and checking for extraneous roots
- Real-world word problems: assign a variable, form an equation from the conditions, solve, reject physically impossible answers
- Quadratic inequalities are solved using critical values and test points in each region
- Every extraneous-root check applies to squared equations, absolute-value splits, and rational equations alike

Exam Tips

- Always find the y-intercept, x-intercepts, and vertex before sketching a parabola -- three points are enough to draw an accurate shape
- Remember: $a>0$ means the parabola opens upward and gives a minimum; $a<0$ opens downward and gives a maximum
- After finding an inverse using the quadratic formula, use the domain of the ORIGINAL restricted function to decide which +/- branch is correct
- When solving $|E|=d$, always substitute every candidate answer back into the ORIGINAL absolute value equation to catch extraneous roots
- For absolute value inequalities, solve the two split inequalities separately and only combine them at the very end
- In word problems, always check whether a negative solution makes physical sense before including it in your final answer

Unit 4: Matrices and Determinants

This is the largest and most foundational unit on linear algebra in the book. A matrix is simply an organized rectangular array of numbers, and this unit builds the full toolkit around it: the four basic matrix operations (addition, subtraction, scalar multiplication, and multiplication), and determinants -- computed through minors and cofactors -- which unlock the adjoint, the inverse of a matrix, and Cramer's Rule for solving systems of equations.

The unit also covers elementary row operations, echelon and reduced echelon forms, using row operations to find both the inverse and the rank of a matrix, and how to tell a consistent system of linear equations from an inconsistent one. It closes with homogeneous systems of equations (which always have at least the trivial all-zero solution) solved by Gaussian elimination, and two real-world applications: using transformation matrices to reflect shapes in computer graphics, and using matrix multiplication and inversion to encode and decode secret messages in cryptography.

Learning Objectives

- Define a matrix and identify its order, rows, columns, and special types (row, column, square, diagonal, scalar, identity, null)
- Find the transpose of a matrix
- Add, subtract, and scalar-multiply matrices
- Multiply two matrices when they are conformable for multiplication
- Evaluate the determinant of a 3x3 matrix using minors and cofactors
- Apply the key properties of determinants to evaluate them without full expansion
- Find the adjoint of a matrix and use it to find the inverse of a non-singular matrix
- Apply elementary row operations to reduce a matrix to echelon and reduced echelon form
- Find the inverse and the rank of a matrix using row operations
- Distinguish between consistent and inconsistent systems of linear equations

- Solve systems of non-homogeneous linear equations using reduced echelon form, matrix inversion, and Cramer's Rule
- Solve systems of homogeneous linear equations by Gaussian elimination, and apply matrices to real-world reflection and cryptography problems

Key Concepts

4.1 Matrices: Definitions and Special Types

A matrix is a rectangular array of numbers arranged in rows and columns and enclosed in brackets. A matrix with m rows and n columns has order m by n, and the element in row i and column j is written a_ij. A matrix with only one row (1xn) is a row matrix, and one with only one column (mx1) is a column matrix. If the number of rows does not equal the number of columns, it is a rectangular matrix; if they are equal, it is a square matrix.

In a square matrix, the entries a11, a22, ..., ann form the principal (main) diagonal. A diagonal matrix has every off-diagonal entry equal to zero; a scalar matrix is a diagonal matrix whose diagonal entries are all the same non-zero number; and the identity (unit) matrix I has every diagonal entry equal to 1 and every off-diagonal entry equal to 0. A null (zero) matrix has every entry equal to 0. Two matrices are equal if they have the same order and every corresponding entry matches. The transpose of a matrix A, written A^t, is formed by interchanging its rows and columns.

4.2 Matrix Operations

Addition and subtraction of matrices are defined entry-by-entry, and are only possible when both matrices have the exact same order. Scalar multiplication kA multiplies every single entry of A by the number k, so kA always has the same order as A.

Matrix multiplication AB is defined only when the number of columns of A equals the number of rows of B; each entry of the product is the sum of the products of a row of A with a column of B. Matrix multiplication is NOT commutative in general -- AB is usually not equal to BA -- although it is associative, A(BC)=(AB)C, and distributes over addition, A(B+C)=AB+AC.

4.3 Determinants: Minors, Cofactors, and Expansion

For a square matrix, the minor M_{ij} of an element a_{ij} is the determinant left after deleting that element's row and column; the cofactor A_{ij} attaches a sign to the minor, $A_{ij} = (-1)^{(i+j)} * M_{ij}$.

The determinant of a 3x3 matrix is found by cofactor expansion along any row or column: multiply each entry in that row (or column) by its own cofactor, and add the results -- the same value results no matter which row or column is chosen. Several properties make many determinants easy to evaluate without full expansion: interchanging two rows or columns negates the determinant; two identical rows or columns make the determinant zero; and the determinant of a triangular matrix is simply the product of its diagonal entries.

4.4 Adjoint and Inverse of a Matrix

The adjoint of a matrix A, written adj(A), is formed by first building the matrix of cofactors of A (replacing every entry with its own cofactor), then taking the transpose of that cofactor matrix.

For a non-singular matrix A (one where $|A|$ is not equal to 0), the inverse is $A^{-1} = (1/|A|) * \text{adj}(A)$. This satisfies $A * A^{-1} = A^{-1} * A = I$, the identity matrix; a matrix with $|A| = 0$ (a singular matrix) has no inverse at all.

4.5 Elementary Row Operations, Echelon Form, and Rank

Three elementary row operations transform a matrix into an equivalent matrix without changing the solution of the system it represents: interchanging two rows, multiplying a row by a non-zero number, and adding a multiple of one row to another. A matrix is in echelon form when each row's leading (first non-zero) entry is 1, later rows have more leading zeros than earlier ones, and any all-zero rows sit at the bottom.

Applying row operations to the combined matrix $[A \mid I]$ until the left half becomes I turns the right half into A^{-1}, giving $[I \mid A^{-1}]$. The rank of a matrix is the number of non-zero rows once it has been reduced to echelon form -- a useful, purely mechanical way to measure how much genuinely independent information a matrix contains.

4.6 Systems of Non-Homogeneous Linear Equations

A system of equations is consistent if it has a unique solution or infinitely many solutions, and inconsistent if it has no solution at all. Comparing the rank of the coefficient matrix to the rank of the augmented matrix (and to the number of variables) tells you which case applies: equal ranks equal to the number of variables gives a unique solution; equal ranks less than the number of variables gives infinitely many solutions; and unequal ranks means the system has no solution.

Three methods solve such a system: reducing the augmented matrix to reduced echelon form and reading off the solution (using back substitution if a variable is left arbitrary); the matrix inversion method, X = A^-1 B (valid only when |A| is not equal to 0); and Cramer's Rule, which finds each variable as a ratio of two determinants -- the denominator is always |A|, and the numerator is the determinant of A with the corresponding column replaced by the constants column B.

4.7 Systems of Homogeneous Linear Equations

A homogeneous system (every constant term equal to 0) is always consistent, because it is always satisfied by the trivial solution, where every variable equals zero. Any other solution is called a non-trivial solution.

Gaussian elimination solves such a system by reducing the augmented matrix to echelon form and then using back substitution. If the rank of the coefficient matrix equals the number of variables, the system has ONLY the trivial solution; if the rank is less than the number of variables, the system also has infinitely many non-trivial solutions.

4.8 Real-World Applications

A transformation (reflection) matrix reflects a point across an axis or line when it multiplies the point's coordinates written as a column vector. The matrix [[1,0],[0,-1]] reflects across the x-axis, [[-1,0],[0,1]] reflects across the y-axis, and [[0,1],[1,0]] reflects across the line y=x.

In cryptography, a message is converted into numbers (A=1, B=2, ..., Z=26), grouped into column vectors, and multiplied by an encoding matrix to encrypt it; multiplying the coded matrix by the encoding matrix's inverse decodes it back to the original message. This is a direct, practical use of matrix multiplication and matrix inversion outside of pure algebra.

Important Definitions

What is a matrix?

A rectangular array of numbers arranged in rows and columns, enclosed in brackets.

What is the order of a matrix?

If a matrix has m rows and n columns, its order is written m by n (m x n).

What is a square matrix?

A matrix in which the number of rows equals the number of columns.

What is the transpose of a matrix?

The matrix obtained by interchanging the rows and columns of the original matrix, denoted A^t.

What is an identity (unit) matrix?

A square matrix with 1s on the principal diagonal and 0s everywhere else, denoted I.

What is the minor of an element of a matrix?

The determinant formed by deleting the row and column containing that element.

What is the cofactor of an element a_ij?

A_ij = (-1)^(i+j) times the minor M_ij of that element.

What is a non-singular matrix?

A square matrix whose determinant is not equal to zero, meaning its inverse exists.

What is the rank of a matrix?

The number of non-zero rows when the matrix is reduced to echelon form.

What is the trivial solution of a homogeneous system?

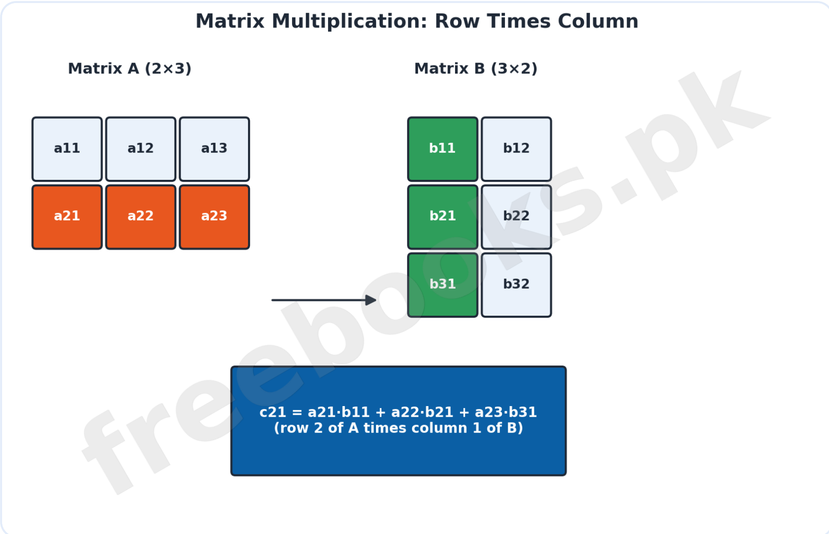
The solution where every variable equals zero, which always satisfies a homogeneous system.

Key Facts and Relations

Topic	Key Fact / Relation
Order of a matrix	m rows x n columns, written m x n
Matrix addition/ subtraction	Entrywise; requires both matrices to have the same order
Scalar multiplication	$kA = [k * a_{ij}]$
Matrix multiplication condition	Columns of A must equal rows of B
Cofactor	$A_{ij} = (-1)^{(i+j)} * M_{ij}$
Determinant of a 3x3 matrix	Expand along any row or column: sum of (entry x its cofactor)
Adjoint	adj(A) = transpose of the matrix of cofactors of A
Inverse via adjoint	$A^{-1} = (1/\|A\|) * \text{adj}(A)$, valid only if $\|A\|$ is not 0
Rank	Number of non-zero rows once the matrix is reduced to echelon form
Cramer's Rule	$x_i = \|A_i\| / \|A\|$, where A_i replaces column i of A with B
Matrix inversion method	$X = A^{-1} * B$ (solves $AX = B$)
Consistency rule	Equal ranks = number of variables: unique solution. Equal ranks < variables: infinite solutions. Unequal ranks: no solution
Homogeneous system rule	rank(A) = variables: trivial solution only. rank(A) < variables: non-trivial solutions exist too

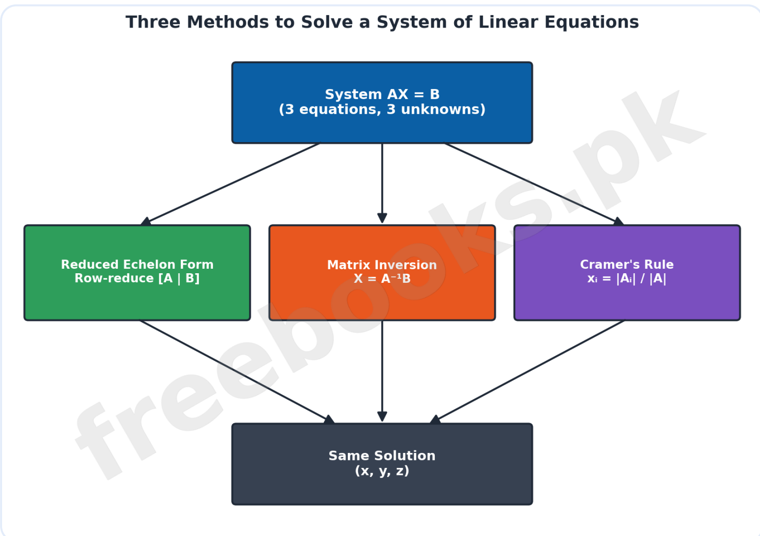
Diagrams

Matrix Multiplication: Row Times Column: A schematic showing a 2x3 matrix A and a 3x2 matrix B, with row 2 of A and column 1 of B highlighted, illustrating that the entry c21 of the product AB is formed by multiplying corresponding entries of that row and column and adding the results



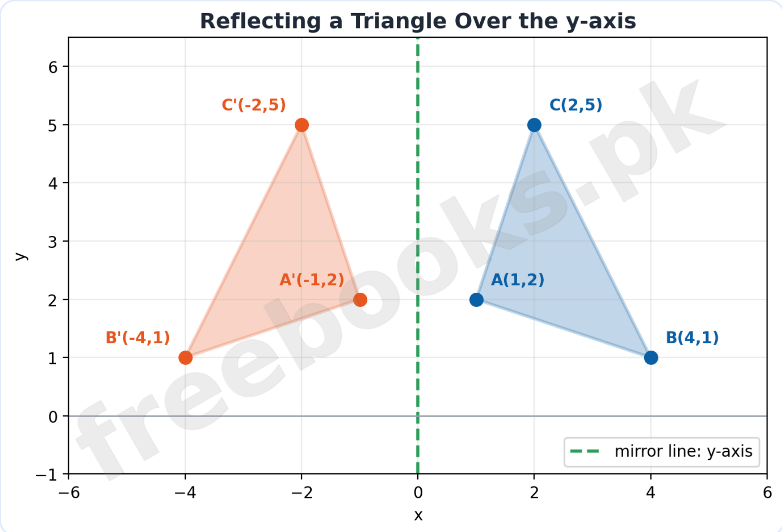
Matrix Multiplication: Row Times Column

Three Methods to Solve a System of Linear Equations: A flow diagram showing a system AX=B branching into three solution methods -- reduced echelon form, matrix inversion, and Cramer's Rule -- each converging on the same final solution (x, y, z)



Three Methods to Solve a System of Linear Equations

Reflecting a Triangle Over the y-axis: A coordinate-plane plot of a triangle with vertices A(1,2), B(4,1), C(2,5) alongside its reflection A'(-1,2), B'(-4,1), C'(-2,5) across the y-axis, produced by multiplying each vertex by the transformation matrix [[-1,0],[0,1]]



Reflecting a Triangle Over the y-axis

Solved Examples

Example 1: Matrix Addition, Subtraction, and Scalar Multiplication

Problem: If A = [[2,-1],[0,3]] and B = [[1,4],[-2,5]], find A+B, A-B, and 3A.

1. A+B: add corresponding entries -- [[2+1, -1+4], [0+(-2), 3+5]] = [[3, 3], [-2, 8]].
2. A-B: subtract corresponding entries -- [[2-1, -1-4], [0-(-2), 3-5]] = [[1, -5], [2, -2]].
3. 3A: multiply every entry of A by 3 -- [[6, -3], [0, 9]].
4. Final answer: A+B = [[3,3],[-2,8]], A-B = [[1,-5],[2,-2]], 3A = [[6,-3],[0,9]].

Example 2: Matrix Multiplication

Problem: If A = [[1,2],[3,4]] and B = [[2,0],[1,3]], find AB.

1. Entry (1,1): row 1 of A times column 1 of B -- 1*2 + 2*1 = 4.

2. Entry (1,2): row 1 of A times column 2 of B -- 1*0 + 2*3 = 6.
3. Entry (2,1): row 2 of A times column 1 of B -- 3*2 + 4*1 = 10.
4. Entry (2,2): row 2 of A times column 2 of B -- 3*0 + 4*3 = 12.
5. Final answer: AB = [[4, 6], [10, 12]].

Example 3: Evaluating a 3x3 Determinant by Cofactor Expansion

Problem: Evaluate |A| for A = [[2,-1,3],[0,2,-4],[1,3,1]] by expanding along the first row.

1. Minor M11 = |[2,-4],[3,1]| = (2)(1)-(-4)(3) = 2+12 = 14. Cofactor A11 = (+1)(14) = 14.
2. Minor M12 = |[0,-4],[1,1]| = (0)(1)-(-4)(1) = 4. Cofactor A12 = (-1)(4) = -4.
3. Minor M13 = |[0,2],[1,3]| = (0)(3)-(2)(1) = -2. Cofactor A13 = (+1)(-2) = -2.
4. |A| = a11*A11 + a12*A12 + a13*A13 = (2)(14) + (-1)(-4) + (3)(-2) = 28+4-6 = 26.
5. Final answer: |A| = 26.

Example 4: Finding the Inverse Using the Adjoint

Problem: Find A^-1 for A = [[1,2,0],[0,1,3],[1,0,2]] using the adjoint method.

1. Compute |A| by expanding along row 1: |A| = 1(1*2-3*0) - 2(0*2-3*1) + 0(0*0-1*1) = 1(2) - 2(-3) + 0 = 2+6 = 8. Since |A| is not 0, A is non-singular.
2. Compute all 9 cofactors: A11=2, A12=3, A13=-1, A21=-4, A22=2, A23=2, A31=6, A32=-3, A33=1.
3. adj(A) = transpose of the cofactor matrix = [[2,-4,6], [3,2,-3], [-1,2,1]].
4. A^-1 = (1/|A|) * adj(A) = (1/8) * [[2,-4,6], [3,2,-3], [-1,2,1]].
5. Final answer: A^-1 = (1/8)[[2,-4,6],[3,2,-3],[-1,2,1]].

Example 5: Finding the Rank of a Matrix

Problem: Find the rank of the matrix [[1,2,-1],[2,4,-2],[1,1,1]] by reducing it to echelon form.

1. Apply R2 -> R2-2R1: [2-2, 4-4, -2+2] = [0,0,0]. The matrix becomes [[1,2,-1],[0,0,0],[1,1,1]].
2. Apply R3 -> R3-R1: [1-1, 1-2, 1+1] = [0,-1,2]. The matrix becomes [[1,2,-1],[0,0,0],[0,-1,2]].
3. Swap R2 and R3 so the all-zero row sits at the bottom (as required for echelon form): [[1,2,-1],[0,-1,2],[0,0,0]].
4. The number of non-zero rows is 2.
5. Final answer: rank = 2.

Example 6: Solving a System by Cramer's Rule

Problem: Solve x+y+z=6, 2x-y+z=3, x+2y-z=2 using Cramer's Rule.

1. Write A = [[1,1,1],[2,-1,1],[1,2,-1]] and B = [6,3,2]. Compute |A| = 1(1-2) - 1(-2-1) + 1(4+1) = -1+3+5 = 7 (non-zero, so Cramer's Rule applies).
2. Replace column 1 of A with B to get Ax; expanding gives |Ax| = 7, so x = |Ax|/|A| = 7/7 = 1.
3. Replace column 2 of A with B to get Ay; expanding gives |Ay| = 14, so y = |Ay|/|A| = 14/7 = 2.
4. Replace column 3 of A with B to get Az; expanding gives |Az| = 21, so z = |Az|/|A| = 21/7 = 3.
5. Check: x+y+z=6, 2x-y+z=2-2+3=3, x+2y-z=1+4-3=2 -- all three original equations are satisfied.
6. Final answer: x=1, y=2, z=3.

Example 7: Solving a Homogeneous System by Gaussian Elimination

Problem: Solve x+y+z=0, 2x-y+z=0, x+4y+2z=0 by Gaussian elimination, and state whether the system has only the trivial solution.

1. Row reduce the augmented matrix (all constants are 0): R2 -> R2-2R1 gives [0,-3,-1 | 0]; R3 -> R3-R1 gives [0,3,1 | 0].
2. Add the new R2 to R3: R3 -> R3+R2 gives [0,0,0 | 0], an all-zero row.
3. The matrix is now in echelon form with only 2 non-zero rows, so rank(A) = 2, which is LESS than the number of variables (3).
4. By the rank rule for homogeneous systems, this means the system has non-trivial solutions, not just the trivial one.
5. Back substitution: from row 2, -3y-z=0 so z=-3y; from row 1, x+y+z=0 so x=2y. Letting y=t gives x=2t, y=t, z=-3t for any real t.
6. Final answer: infinitely many non-trivial solutions, (x,y,z) = (2t, t, -3t) for any real t.

Example 8: Application: Reflecting a Shape with a Transformation Matrix

Problem: A triangle has vertices A(1,2), B(4,1), C(2,5). Find the vertices of the triangle reflected over the y-axis using a transformation matrix.

1. The transformation matrix for reflection over the y-axis is [[-1,0],[0,1]].
2. A' = [[-1,0],[0,1]] * [1,2] = [-1, 2], giving A'(-1, 2).
3. B' = [[-1,0],[0,1]] * [4,1] = [-4, 1], giving B'(-4, 1).
4. C' = [[-1,0],[0,1]] * [2,5] = [-2, 5], giving C'(-2, 5).
5. Final answer: A'(-1,2), B'(-4,1), C'(-2,5).

Short Questions & Answers

What is the order of a matrix with 3 rows and 5 columns?

3 x 5.

What is the transpose of a matrix?

The matrix obtained by interchanging its rows and columns.

When are two matrices conformable for multiplication?

When the number of columns of the first matrix equals the number of rows of the second matrix.

What is the cofactor of an element a_{ij}?

$A_{ij} = (-1)^{(i+j)}$ times the minor of a_{ij}.

What condition must $|A|$ satisfy for A^{-1} to exist?

$|A|$ must not equal zero -- that is, A must be non-singular.

What is the rank of a matrix?

The number of non-zero rows when the matrix is reduced to echelon form.

What is the trivial solution of a homogeneous system?

The solution where every variable equals zero.

Long Questions & Answers

Explain how the determinant of a 3x3 matrix is evaluated using cofactor expansion, and how the adjoint is used to find the inverse of a non-singular matrix.

What is the minor and the cofactor of an element of a matrix?

The minor M_{ij} of an element a_{ij} is the determinant obtained by deleting the ith row and jth column of the matrix; the cofactor A_{ij} is the minor with a sign attached, $A_{ij} = (-1)^{(i+j)} * M_{ij}$.

How is the determinant of a 3x3 matrix computed using cofactor expansion?

Choose any row or column, multiply each entry in it by its own cofactor, and add the results; expanding along the first row gives $|A| = a_{11}*A_{11} + a_{12}*A_{12} + a_{13}*A_{13}$, and the same value results no matter which row or column is chosen.

What is the adjoint of a matrix, and how is it formed?

The adjoint, adj(A), is formed by first building the matrix of cofactors of A (replacing every entry with its own cofactor), then taking the transpose of that cofactor matrix.

How is the inverse of a non-singular matrix found using the adjoint?

For a matrix A with |A| not equal to 0 (non-singular), the inverse is A^-1 = (1/|A|) * adj(A); this satisfies A*A^-1 = A^-1*A = I, the identity matrix.

Describe the three methods for solving a system of non-homogeneous linear equations, and explain how the rank of a matrix determines whether a homogeneous system has only the trivial solution.

How is a system of equations solved using reduced echelon form?

Write the system as an augmented matrix, apply elementary row operations until the matrix reaches reduced echelon form, and read the solution directly from the resulting simplified equations, using back substitution if any variable is left arbitrary.

How is a system of equations solved using the matrix inversion method?

Write the system as AX=B, find A^-1 (provided |A| is not equal to 0), and compute X=A^-1 B to get the values of all the variables at once.

How does Cramer's Rule solve a system of equations?

Each variable x_i is found as the ratio of two determinants: the denominator is always |A|, and the numerator is the determinant of A with its ith column replaced by the constants column B.

How does the rank of the coefficient matrix determine whether a homogeneous system has only the trivial solution?

If the rank of the coefficient matrix equals the number of variables, the system has only the trivial (all-zero) solution; if the rank is less than the number of variables, the system has infinitely many non-trivial solutions as well.

Multiple Choice Questions (MCQs)

The order of a matrix with 2 rows and 4 columns is:

A. 4 x 2
B. 2 x 4
C. 8 x 1
D. 2 x 2
Answer: B. 2 x 4
Explanation: Order is written as (number of rows) x (number of columns), so 2 x 4.

The transpose of a matrix is formed by:

A. Adding a row of zeros
B. Interchanging rows and columns
C. Multiplying by -1
D. Deleting the last row
Answer: B. Interchanging rows and columns
Explanation: The transpose A^t is obtained by interchanging the rows and columns of A.

Two matrices can be multiplied only if:

A. They have the same order
B. The columns of the first equal the rows of the second
C. They are both square
D. They are both diagonal
Answer: B. The columns of the first equal the rows of the second
Explanation: Matrix multiplication AB requires the number of columns of A to equal the number of rows of B.

The cofactor of an element a_ij is:

A. Its minor only
B. (-1)^(i+j) times its minor
C. Its minor times -1 always
D. The element itself
Answer: B. (-1)^(i+j) times its minor
Explanation: The cofactor attaches an alternating sign to the minor: $A_{ij} = (-1)^{(i+j)} * M_{ij}$.

A square matrix A has an inverse only if:
A. |A| = 0
B. |A| is not equal to 0
C. A is diagonal
D. A is symmetric
Answer: B. |A| is not equal to 0
Explanation: A matrix is non-singular (invertible) only when its determinant is not equal to zero.

The rank of a matrix is:
A. The number of rows
B. The number of non-zero rows in echelon form
C. The determinant value
D. The number of columns
Answer: B. The number of non-zero rows in echelon form
Explanation: Rank is defined as the number of non-zero rows once the matrix is reduced to echelon form.

Cramer's Rule finds each variable as:
A. A sum of two determinants
B. The ratio of two determinants
C. The product of two determinants
D. A single determinant
Answer: B. The ratio of two determinants
Explanation: Cramer's Rule computes $x_i = |A_i| / |A|$, a ratio of two determinants.

A homogeneous system of equations is always:
A. Inconsistent
B. Consistent (satisfied by the trivial solution)

C. Undefined

D. Non-linear

Answer: B. Consistent (satisfied by the trivial solution)

Explanation: A homogeneous system is always consistent because the trivial (all-zero) solution always satisfies it.

A homogeneous system has only the trivial solution when:

A. rank(A) < number of variables

B. rank(A) = number of variables

C. rank(A) = 0

D. |A| = 0 always

Answer: B. rank(A) = number of variables

Explanation: Only the trivial solution exists when the rank of the coefficient matrix equals the number of variables.

To reflect a point over the x-axis using a transformation matrix, multiply it by:

A. [[1,0],[0,-1]]

B. [[-1,0],[0,1]]

C. [[0,1],[1,0]]

D. [[1,1],[1,1]]

Answer: A. [[1,0],[0,-1]]

Explanation: The matrix [[1,0],[0,-1]] keeps x unchanged and negates y, reflecting across the x-axis.

Quick Revision Summary

- A matrix is a rectangular array of numbers with order m x n (m rows, n columns)
- The transpose A^t is formed by interchanging rows and columns
- Matrix addition/subtraction requires the same order; multiplication requires columns of A = rows of B
- Matrix multiplication is NOT commutative in general: AB is not equal to BA
- Cofactor: $A_{ij} = (-1)^{(i+j)}$ x minor M_{ij}
- Determinant of a 3x3 matrix: expand along any row or column, using entries x their cofactors

- adj(A) = transpose of the cofactor matrix; A^-1 = (1/|A|) adj(A), valid only when |A| is not 0
- Row operations reduce a matrix to echelon form; rank = number of non-zero rows in echelon form
- A system is consistent (unique or infinite solutions) or inconsistent (no solution), based on comparing ranks
- Three methods to solve AX=B: reduced echelon form, matrix inversion X=A^-1B, and Cramer's Rule
- A homogeneous system is always consistent; rank(A) = variables gives only the trivial solution, rank(A) < variables gives non-trivial solutions too
- Matrices model real transformations: reflections in graphics, and encoding/decoding messages in cryptography

Exam Tips

- Always double-check matrix order before adding, subtracting, or multiplying -- mismatched orders mean the operation is undefined
- When expanding a 3x3 determinant, pick the row or column with the most zeros to minimize calculation
- Remember AB is generally not equal to BA -- never assume matrix multiplication commutes
- After finding an inverse, spot-check by multiplying A times A^-1 to confirm you get the identity matrix
- In Cramer's Rule, compute |A| first -- if it is zero, the rule cannot be used at all
- For homogeneous systems, always compare the rank of A to the number of variables to decide trivial vs non-trivial solutions

Unit 5: Partial Fractions

This unit teaches how to reverse a process students already know well -- combining two or more fractions into a single fraction -- by breaking a single, more complicated rational fraction back down into a sum of simpler ones. This skill, called partial fraction resolution, becomes essential later in calculus, where a complicated fraction is often far easier to integrate once it has been split into simple pieces.

The unit begins by distinguishing proper from improper rational fractions and showing how an improper fraction is reduced, using long division, to a polynomial plus a proper fraction before any decomposition can begin. It then works through all four standard cases of partial fraction decomposition, based on how the denominator factors: non-repeated linear factors, repeated linear factors, non-repeated irreducible quadratic factors, and repeated irreducible quadratic factors -- along with the two techniques (substitution and coefficient comparison) used to find the unknown constants in each case.

Learning Objectives

- Distinguish between a proper and an improper rational fraction
- Reduce an improper rational fraction to a polynomial plus a proper fraction using long division
- Distinguish between a conditional equation and an identity
- Apply the theorem that equal polynomials have equal coefficients of every matching power of x
- Resolve a rational fraction into partial fractions when the denominator has non-repeated linear factors
- Resolve a rational fraction into partial fractions when the denominator has repeated linear factors
- Resolve a rational fraction into partial fractions when the denominator has non-repeated irreducible quadratic factors
- Resolve a rational fraction into partial fractions when the denominator has repeated irreducible quadratic factors

- Apply both the substitution method and the coefficient-comparison method to find unknown constants, and verify a decomposition by recombining it

Key Concepts

5.1 Rational Fractions: Proper and Improper

A rational fraction is an expression of the form P(x)/Q(x), where P(x) and Q(x) are polynomials and Q(x) is not equal to zero. It is called proper if the degree of the numerator P(x) is less than the degree of the denominator Q(x), and improper if the degree of the numerator is equal to or greater than the degree of the denominator.

The method of partial fractions only works on proper fractions. If a fraction is improper, it must first be reduced by long division: dividing the numerator by the denominator produces a polynomial quotient plus a remainder over the original denominator, and since the remainder's degree is always less than the denominator's degree, this remainder fraction is now proper and ready for decomposition.

5.2 Equations, Identities, and the Coefficient-Comparison Theorem

A conditional equation is true only for certain particular values of the variable, while an identity is true for every value of the variable. Partial fraction decomposition always produces an identity -- once the constants are correctly found, the decomposed sum equals the original fraction for every value of x (except where the denominator is zero).

The key theorem used throughout this unit: if two polynomials are equal for all values of x, they have the same degree and the coefficients of every matching power of x must be equal. This theorem, combined with substituting convenient values of x that make a particular factor zero, is how every unknown constant in a partial fraction decomposition is found.

5.3 Case I: Non-Repeated Linear Factors

When the denominator factors as $Q(x) = (x-a1)(x-a2)\ldots(x-an)$, with every root distinct, the fraction decomposes as $P(x)/Q(x) = A1/(x-a1) + A2/(x-a2) + \ldots + An/(x-an)$ -- one constant-numerator term for each distinct linear factor.

Multiplying both sides by $Q(x)$ clears every denominator and produces an identity in x. Each constant can then be found either by equating the coefficients of matching powers of x on both sides, or -- much faster -- by substituting $x = a_i$ for each root in turn, which makes every term except the one containing A_i vanish completely.

5.4 Case II: Repeated Linear Factors

When the denominator contains a repeated linear factor $(x-a)^n$, the decomposition needs one term for every power from 1 up to n, not just a single term: $A1/(x-a) + A2/(x-a)^2 + \ldots + An/(x-a)^n$.

After clearing denominators, the highest-power constant An is usually found immediately by substituting $x=a$, since every other term vanishes at that value. The remaining constants are then found by equating the coefficients of the remaining powers of x.

5.5 Case III: Non-Repeated Irreducible Quadratic Factors

An irreducible quadratic factor is one that cannot be split into two real linear factors, such as x^2+1 or x^2+x+1. When the denominator contains such a factor (not repeated), its partial fraction term takes the form $(Ax+B)/(ax^2+bx+c)$ -- a linear expression in the numerator, not just a constant.

Any ordinary linear factors that also appear in the denominator still get standard constant-numerator terms exactly as in Case I. The unknown constants are usually found using a mix of convenient substitution for the linear factors and equating coefficients of matching powers of x for the quadratic term's A and B.

5.6 Case IV: Repeated Irreducible Quadratic Factors

When the denominator contains a repeated irreducible quadratic factor $(ax^2+bx+c)^n$, the decomposition needs one linear-numerator term for every power from 1 to n: $(A1x+B1)/(ax^2+bx+c) + (A2x+B2)/(ax^2+bx+c)^2 + \ldots + (Anx+Bn)/(ax^2+bx+c)^n$.

As with Case II, this almost always requires equating the coefficients of every matching power of x after clearing denominators, since there is no single convenient real substitution that isolates each unknown constant directly the way substituting a linear factor's root does.

Important Definitions

What is a rational fraction?

An expression $P(x)/Q(x)$ where $P(x)$ and $Q(x)$ are polynomials and $Q(x)$ is not equal to zero.

What is a proper rational fraction?

One in which the degree of the numerator $P(x)$ is less than the degree of the denominator $Q(x)$.

What is an improper rational fraction?

One in which the degree of the numerator is equal to or greater than the degree of the denominator.

What is the difference between a conditional equation and an identity?

A conditional equation is true only for particular values of the variable; an identity is true for every value of the variable.

What is partial fraction resolution?

The process of expressing a single rational fraction as a sum of two or more simpler rational fractions.

What is an irreducible quadratic factor?

A quadratic expression that cannot be written as the product of two linear factors with real coefficients.

What form does the partial fraction take for a non-repeated linear factor (x-a)?

$A/(x-a)$, where A is a constant to be found.

What form does the partial fraction take for a non-repeated irreducible quadratic factor?

(Ax+B)/(ax^2+bx+c), with a linear expression in the numerator.

How is an improper rational fraction prepared for partial fraction resolution?

By long division of the numerator by the denominator, writing it as a polynomial plus a proper fraction, then resolving the proper fraction.

What theorem allows coefficients to be equated in an identity?

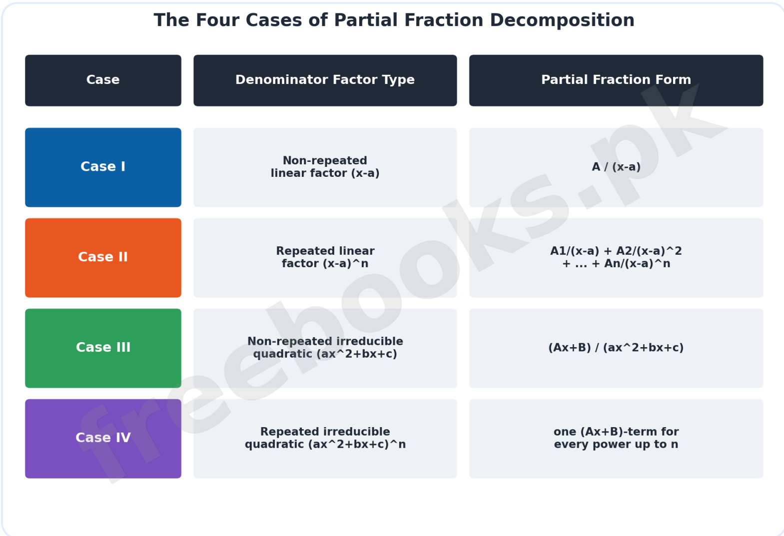
If two polynomials are equal for all values of the variable, they have the same degree and the coefficients of every matching power of the variable must be equal.

Key Facts and Relations

Topic	Key Fact / Relation
Proper fraction condition	degree of P(x) < degree of Q(x)
Improper fraction reduction	P(x)/Q(x) = polynomial quotient + remainder/Q(x), found by long division
Case I: non-repeated linear factors	P(x)/[(x-a1)...(x-an)] = A1/(x-a1) + A2/(x-a2) + ... + An/(x-an)
Case II: repeated linear factor (x-a)^n	A1/(x-a) + A2/(x-a)^2 + ... + An/(x-a)^n
Case III: non-repeated irreducible quadratic	(Ax+B) / (ax^2+bx+c)
Case IV: repeated irreducible quadratic (ax^2+bx+c)^n	(A1x+B1)/(ax^2+bx+c) + ... + (Anx+Bn)/(ax^2+bx+c)^n
Substitution shortcut	To find the constant for factor (x-a), substitute x=a after clearing denominators
Coefficient-comparison rule	Equal polynomials have equal coefficients for every matching power of x

Diagrams

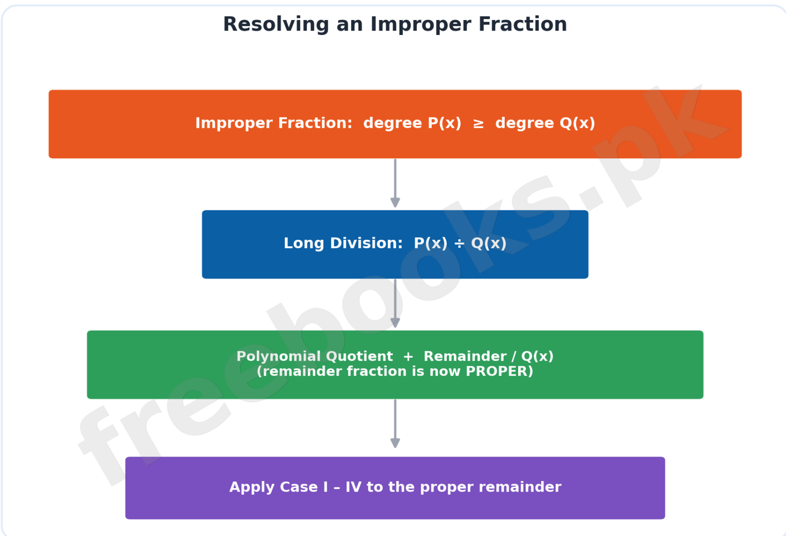
The Four Cases of Partial Fraction Decomposition: A summary table listing all four cases -- non-repeated linear, repeated linear, non-repeated irreducible quadratic, and repeated irreducible quadratic factors -- alongside the exact partial fraction form used for each


The Four Cases of Partial Fraction Decomposition

Case	Denominator Factor Type	Partial Fraction Form
Case I	Non-repeated linear factor (x-a)	A / (x-a)
Case II	Repeated linear factor (x-a)^n	A1/(x-a) + A2/(x-a)^2 + ... + An/(x-a)^n
Case III	Non-repeated irreducible quadratic (ax^2+bx+c)	(Ax+B) / (ax^2+bx+c)
Case IV	Repeated irreducible quadratic (ax^2+bx+c)^n	one (Ax+B)-term for every power up to n

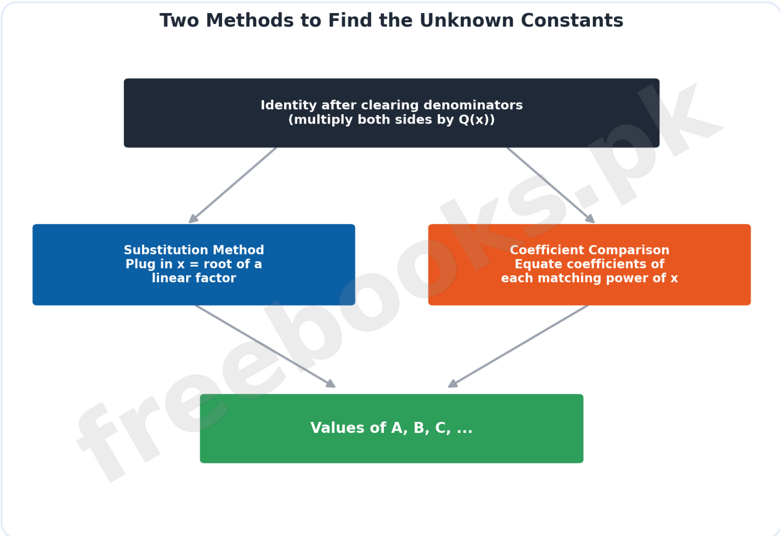
The Four Cases of Partial Fraction Decomposition

Resolving an Improper Fraction: A flow diagram showing an improper fraction being reduced by long division into a polynomial quotient plus a proper remainder fraction, which is then resolved using Case I-IV



Resolving an Improper Fraction

Two Methods to Find the Unknown Constants: A flow diagram showing the identity obtained after clearing denominators branching into two methods -- substitution (plugging in the roots of linear factors) and coefficient comparison (equating like powers of x) -- both converging on the values of the unknown constants



Two Methods to Find the Unknown Constants

Solved Examples

Example 1: Reducing an Improper Fraction

Problem: Express (2x^2+3)/(x-1) as a polynomial plus a proper fraction using long division.

1. Divide the leading term: 2x^2 divided by x gives 2x. Multiply 2x by (x-1) to get 2x^2-2x, and subtract from 2x^2+3 to leave 2x+3.
2. Divide the new leading term: 2x divided by x gives 2. Multiply 2 by (x-1) to get 2x-2, and subtract from 2x+3 to leave a remainder of 5.
3. The quotient is 2x+2 and the remainder is 5, so the division stops here.
4. Final answer: (2x^2+3)/(x-1) = 2x+2 + 5/(x-1).

Example 2: Case I: Non-Repeated Linear Factors (Substitution Method)

Problem: Resolve $(5x-1)/[(x-2)(x+1)]$ into partial fractions.

1. Suppose $(5x-1)/[(x-2)(x+1)] = A/(x-2) + B/(x+1)$.
2. Multiply both sides by $(x-2)(x+1)$ to clear denominators: $5x-1 = A(x+1) + B(x-2)$.
3. Substitute $x=2$ (making the B-term vanish): $5(2)-1 = A(3)$, so $9 = 3A$, giving $A = 3$.
4. Substitute $x=-1$ (making the A-term vanish): $5(-1)-1 = B(-3)$, so $-6 = -3B$, giving $B = 2$.
5. Final answer: $(5x-1)/[(x-2)(x+1)] = 3/(x-2) + 2/(x+1)$.

Example 3: Case II: Repeated Linear Factors

Problem: Resolve $(3x-1)/(x-1)^2$ into partial fractions.

1. Suppose $(3x-1)/(x-1)^2 = A/(x-1) + B/(x-1)^2$.
2. Multiply both sides by $(x-1)^2$ to clear denominators: $3x-1 = A(x-1) + B$.
3. Substitute $x=1$ (making the A-term vanish): $3(1)-1 = B$, so $B = 2$.
4. Equate the coefficients of x on both sides: $3 = A$.
5. Final answer: $(3x-1)/(x-1)^2 = 3/(x-1) + 2/(x-1)^2$.

Example 4: Case III: Non-Repeated Irreducible Quadratic Factor

Problem: Resolve $(2x+1)/[(x^2+1)(x-1)]$ into partial fractions.

1. Suppose $(2x+1)/[(x^2+1)(x-1)] = (Ax+B)/(x^2+1) + C/(x-1)$.
2. Multiply both sides by $(x^2+1)(x-1)$ to clear denominators: $2x+1 = (Ax+B)(x-1) + C(x^2+1)$.
3. Substitute $x=1$ (making the first term vanish): $2(1)+1 = C(2)$, so $3 = 2C$, giving $C = 3/2$.

4. Expand the right side and equate coefficients of x^2: $0 = A+C$, so $A = -3/2$. Equate the constant terms: $1 = -B+C$, so $B = C-1 = 1/2$.
5. Final answer: $(2x+1)/[(x^2+1)(x-1)] = (1-3x)/[2(x^2+1)] + 3/[2(x-1)]$.

Example 5: Case IV: Repeated Irreducible Quadratic Factor

Problem: Resolve $x^3/(x^2+1)^2$ into partial fractions.

1. Suppose $x^3/(x^2+1)^2 = (Ax+B)/(x^2+1) + (Cx+D)/(x^2+1)^2$.
2. Multiply both sides by $(x^2+1)^2$ to clear denominators: $x^3 = (Ax+B)(x^2+1) + (Cx+D) = Ax^3 + Bx^2 + (A+C)x + (B+D)$.
3. Equate coefficients of x^3: $A=1$. Equate coefficients of x^2: $B=0$.
4. Equate coefficients of x: $A+C=0$, so $C=-1$. Equate the constant terms: $B+D=0$, so $D=0$.
5. Final answer: $x^3/(x^2+1)^2 = x/(x^2+1) - x/(x^2+1)^2$.

Example 6: Combining Substitution and Coefficient Comparison

Problem: Resolve $(x^2+2x+3)/[(x-1)(x+2)^2]$ into partial fractions.

1. Suppose $(x^2+2x+3)/[(x-1)(x+2)^2] = A/(x-1) + B/(x+2) + C/(x+2)^2$.
2. Multiply both sides by $(x-1)(x+2)^2$ to clear denominators: $x^2+2x+3 = A(x+2)^2 + B(x-1)(x+2) + C(x-1)$.
3. Substitute $x=1$: $1+2+3 = A(9)$, so $A = 6/9 = 2/3$.
4. Substitute $x=-2$: $4-4+3 = C(-3)$, so $C = -1$.
5. Equate the coefficients of x^2: $1 = A+B$, so $B = 1 - 2/3 = 1/3$.
6. Final answer: $(x^2+2x+3)/[(x-1)(x+2)^2] = (2/3)/(x-1) + (1/3)/(x+2) - 1/(x+2)^2$.

Example 7: Verifying a Partial Fraction Decomposition

Problem: Verify that 4/(x+3) + 3/(x+4) recombines to give (7x+25)/[(x+3)(x+4)].

1. Find the common denominator (x+3)(x+4): 4/(x+3) + 3/(x+4) = [4(x+4) + 3(x+3)] / [(x+3)(x+4)].
2. Expand the numerator: 4(x+4) + 3(x+3) = 4x+16+3x+9 = 7x+25.
3. So 4/(x+3) + 3/(x+4) = (7x+25)/[(x+3)(x+4)], which matches the original fraction exactly.
4. This 'add back together' check is a reliable way to confirm any partial fraction answer, and is especially useful before using the decomposed form in a later step such as integration in calculus.
5. Final answer: the decomposition is verified -- 4/(x+3) + 3/(x+4) recombines to (7x+25)/[(x+3)(x+4)].

Short Questions & Answers

What is a proper rational fraction?

One in which the degree of the numerator is less than the degree of the denominator.

How is an improper rational fraction prepared for partial fractions?

By long division, reducing it to a polynomial plus a proper fraction.

What is an identity?

An equation that holds true for all values of the variable, not just particular ones.

What form is used for a non-repeated linear factor (x-a)?

A/(x-a).

What form is used for a repeated linear factor (x-a)^2?

A/(x-a) + B/(x-a)^2.

What makes a quadratic factor irreducible?

It cannot be factored into two real linear factors.

What is the fastest way to find the constant for a non-repeated linear factor?
Substitute the value of x that makes that factor equal to zero.

Long Questions & Answers

Explain the difference between a proper and an improper rational fraction, and describe how an improper fraction is prepared before it can be resolved into partial fractions.

What is a rational fraction, and what distinguishes proper from improper?

A rational fraction is $P(x)/Q(x)$ with $Q(x)$ not equal to zero; it is proper if the degree of $P(x)$ is less than the degree of $Q(x)$, and improper if the degree of $P(x)$ is equal to or greater than the degree of $Q(x)$.

Why can't partial fraction decomposition be applied directly to an improper fraction?

The standard partial fraction forms (constant or linear numerators over factors of the denominator) only ever add up to a proper fraction, so applying them directly to an improper fraction would not produce a valid identity.

How is an improper fraction converted into a form ready for partial fractions?

Long division is used to divide the numerator by the denominator, producing a polynomial quotient plus a remainder over the original denominator; since the remainder's degree is always less than the denominator's degree, this remainder fraction is now proper and can be decomposed normally.

Work through an example of this conversion.

For $(2x^2+3)/(x-1)$, dividing gives a quotient of $2x+2$ and a remainder of 5, so $(2x^2+3)/(x-1) = 2x+2 + 5/(x-1)$; the polynomial part $2x+2$ is left as is, and only the proper fraction $5/(x-1)$ would need further partial fraction work if its denominator had more than one factor.

Describe all four cases of partial fraction decomposition based on the type of factor found in the denominator, including the general form of the partial fractions used in each case.

What form is used when the denominator has only non-repeated linear factors?

Q(x) = (x-a1)(x-a2)...(x-an) with all different roots gives P(x)/Q(x) = A1/(x-a1) + A2/(x-a2) + ... + An/(x-an), one constant-numerator term per distinct linear factor.

What form is used when the denominator has a repeated linear factor?

A repeated factor (x-a)^n requires one term for every power from 1 up to n: A1/(x-a) + A2/(x-a)^2 + ... + An/(x-a)^n, not just a single term.

What form is used when the denominator has a non-repeated irreducible quadratic factor?

An irreducible quadratic factor ax^2+bx+c (one that cannot be split into real linear factors) gets a linear-numerator term, (Ax+B)/(ax^2+bx+c), instead of a constant-numerator term.

What form is used when the denominator has a repeated irreducible quadratic factor?

A repeated irreducible quadratic factor (ax^2+bx+c)^n requires one linear-numerator term for every power from 1 to n: (A1x+B1)/(ax^2+bx+c) + (A2x+B2)/(ax^2+bx+c)^2 + ... + (Anx+Bn)/(ax^2+bx+c)^n.

Multiple Choice Questions (MCQs)

A rational fraction P(x)/Q(x) is proper if:

A. degree P > degree Q
B. degree P < degree Q
C. degree P = degree Q always
D. Q(x) = 0
Answer: B. degree P < degree Q

Explanation: A proper fraction has numerator degree strictly less than denominator degree.

An improper rational fraction is first converted using:

A. Factoring only

B. Long division

C. Substitution only

D. Squaring

Answer: B. Long division

Explanation: Long division reduces an improper fraction to a polynomial plus a proper remainder fraction.

An identity is true for:

A. Only one value of x

B. A few particular values of x

C. All values of the variable

D. No values of x

Answer: C. All values of the variable

Explanation: An identity holds true for every value of the variable, unlike a conditional equation.

For a non-repeated linear factor (x-a), the partial fraction has the form:

A. A/(x-a)

B. A/(x-a)^2

C. (Ax+B)/(x-a)

D. Ax

Answer: A. A/(x-a)

Explanation: A non-repeated linear factor gets a single constant-numerator term, A/(x-a).

For a repeated linear factor (x-a)^2, the decomposition needs:

A. Only one term

B. Two terms: A/(x-a) and B/(x-a)^2

C. Three terms always

D. No terms

Answer: B. Two terms: A/(x-a) and B/(x-a)^2

Explanation: A repeated factor to the power 2 needs one term for each power up to 2.

An irreducible quadratic factor:

A. Can always be split into two real linear factors

B. Cannot be split into two real linear factors

C. Is always a perfect square

D. Has no real coefficients

Answer: B. Cannot be split into two real linear factors

Explanation: An irreducible quadratic cannot be factored into two real linear factors.

The partial fraction for a non-repeated irreducible quadratic factor has numerator:

A. A constant

B. A linear expression Ax+B

C. A quadratic expression

D. Always zero

Answer: B. A linear expression Ax+B

Explanation: Irreducible quadratic factors require a linear numerator, Ax+B, not just a constant.

The fastest way to find the constant for a non-repeated linear factor (x-a) is to:

A. Equate all coefficients

B. Substitute x = a into the cleared identity

C. Integrate both sides

D. Take the derivative

Answer: B. Substitute x = a into the cleared identity

Explanation: Substituting x=a makes every other term vanish, instantly giving that constant.

Partial fraction decomposition is especially useful in:

A. Geometry

B. Calculus (e.g. integration)

C. Trigonometry only

D. Probability

Answer: B. Calculus (e.g. integration)

Explanation: Decomposing a complicated fraction into simple pieces makes integration in calculus much easier.

To resolve $(3x-1)/(x-1)^2$, the correct form is:

A. $A/(x-1)$

B. $A/(x-1) + B/(x-1)^2$

C. $(Ax+B)/(x-1)^2$

D. $A/(x-1)^2$ only

Answer: B. $A/(x-1) + B/(x-1)^2$

Explanation: A repeated linear factor to the power 2 needs both $A/(x-1)$ and $B/(x-1)^2$.

Quick Revision Summary

- A rational fraction $P(x)/Q(x)$ is proper if degree $P(x)$ < degree $Q(x)$, improper otherwise
- An improper fraction is reduced by long division to a polynomial plus a proper fraction before resolving
- An identity holds for all values of the variable; a conditional equation holds only for particular values
- Case I (non-repeated linear factors): $P(x)/Q(x) = A_1/(x-a_1) + A_2/(x-a_2) + \ldots$
- Case II (repeated linear factor $(x-a)^n$): needs $A_1/(x-a) + A_2/(x-a)^2 + \ldots + A_n/(x-a)^n$
- Case III (non-repeated irreducible quadratic): uses $(Ax+B)/(ax^2+bx+c)$
- Case IV (repeated irreducible quadratic $(ax^2+bx+c)^n$): needs one $(Ax+B)$-type term per power up to n
- Substituting $x = a$ (the root of a linear factor) instantly finds that factor's constant
- Coefficients of matching powers of x must be equal in an identity -- used when substitution alone isn't enough
- Always factor $Q(x)$ completely into irreducible pieces before setting up the partial fraction form
- Multiply both sides by the full denominator $Q(x)$ to clear all fractions before solving for constants
- Recombining the partial fractions is a reliable way to check that a decomposition is correct

Exam Tips

- Always check whether a fraction is proper or improper FIRST -- improper fractions need long division before anything else
- Factor the denominator completely (into linear and irreducible quadratic factors) before writing the partial fraction form
- Use substitution (plugging in the root of a linear factor) whenever possible -- it is much faster than comparing coefficients
- For repeated factors, remember to include a term for EVERY power up to n, not just the highest one
- For irreducible quadratic factors, always use a linear numerator Ax+B, never just a constant
- After finding all constants, recombine the partial fractions to verify they add up to the original fraction

Unit 6: Sequences and Series

This unit studies patterns of numbers that follow a clear rule -- sequences -- and the sums formed by adding their terms together, called series. Three major types of progressions are covered in depth: arithmetic progressions, where each term is formed by adding a fixed constant to the one before it; geometric progressions, where each term is formed by multiplying by a fixed constant; and harmonic progressions, whose reciprocals form an arithmetic progression.

For each type of progression, the unit develops the formula for the general (nth) term, the arithmetic/geometric/harmonic mean between two numbers, and the formula for the sum of the corresponding series -- including the special case of the sum to infinity of a geometric series. It closes with sigma notation, key summation formulas, and real-world applications such as simple interest, compound interest, vehicle leasing, and depreciation, all of which turn out to be arithmetic or geometric progressions in disguise.

Learning Objectives

- Define a sequence and distinguish between finite and infinite sequences
- Identify an arithmetic progression (A.P.) and find its common difference, general term, and specific terms
- Find the arithmetic mean (A.M.) between two numbers, and insert n arithmetic means between two numbers
- Derive and apply the formula for the sum of an arithmetic series
- Identify a geometric progression (G.P.) and find its common ratio, general term, and specific terms
- Find the geometric mean (G.M.) between two numbers, and derive and apply the formula for the sum of a geometric series (finite and infinite)
- Identify a harmonic progression (H.P.) and find the harmonic mean (H.M.) between two numbers
- Use sigma notation to express series, and apply the summation formulas for the sums of the first n natural numbers, their squares, and their cubes

- Apply sequences and series to solve real-world problems involving simple interest, compound interest, and other practical scenarios

Key Concepts

6.1 Sequences: Definitions and Types

A sequence is a function whose domain is the set of natural numbers N, and whose range may be any subset of real (or complex) numbers; the numbers in a sequence are called its terms, with the nth term (or general term) denoted a_n. A sequence with a finite number of terms is a finite sequence; one that continues indefinitely is an infinite sequence.

If the general term a_n is known, any specific term of the sequence can be found by substitution; conversely, if enough terms are given, a formula for a_n can often be found. Sequences following a specific pattern are also called progressions, and this unit studies three major types: arithmetic, geometric, and harmonic progressions.

6.2 Arithmetic Progression (A.P.) and Arithmetic Mean (A.M.)

A sequence is an arithmetic progression if the difference between any term and the one before it is always the same constant, called the common difference d. The general term of an A.P. is $a_n = a_1 + (n-1)d$, and three numbers a, b, c are in A.P. if and only if $2b = a+c$.

A number A is the arithmetic mean between a and b if a, A, b are in A.P.; solving gives $A = (a+b)/2$. More generally, if n arithmetic means are to be inserted between a and b, the sequence a, A_1, A_2, ..., A_n, b is treated as an A.P. with n+2 terms, and the common difference is found using the general term formula.

6.3 Arithmetic Series

The sum of the terms of a sequence is called a series, and the sum of the first n terms is denoted S_n. The sum of an arithmetic series is given by $S_n = (n/2)(a_1+a_n)$, or equivalently $S_n = (n/2)[2a_1+(n-1)d]$ when a_n is not directly known.

These two forms of the sum formula are interchangeable -- the first is quicker when the last term a_n is already known, and the second is quicker when only a_1

and d are known -- and many real-life problems (like total lease payments or total interest) are solved by recognizing the pattern of amounts as an arithmetic series and applying one of these formulas.

6.4 Geometric Progression (G.P.) and Geometric Mean (G.M.)

A sequence is a geometric progression if each term after the first is found by multiplying the previous term by a fixed non-zero constant r, called the common ratio, found by dividing any term by the one before it. The general term of a G.P. is $a_n = a1 * r^{(n-1)}$, and three numbers a, b, c are in G.P. if and only if $b^2 = ac$.

A number G is the geometric mean between a and b if a, G, b are in G.P.; solving gives $G = +/- \sqrt{ab}$. If the positive numbers of a G.P. are replaced by their logarithms, the resulting sequence is always an A.P., and vice versa -- a useful link between the two progression types.

6.5 Geometric Series

The sum of the first n terms of a geometric series is $Sn = a1(1-r^n)/(1-r)$, for r not equal to 1 (and $Sn = n*a1$ when r=1); it is derived by subtracting rSn from Sn, which cancels all the middle terms.

When $|r| < 1$, r^n approaches 0 as n grows without bound, so the sum to infinity of a geometric series converges to $S_{infinity} = a1/(1-r)$ -- a result with no counterpart in arithmetic series, since arithmetic series never converge to a finite sum as n increases.

6.6 Harmonic Progression (H.P.) and Harmonic Mean (H.M.)

A sequence is a harmonic progression if the reciprocals of its terms form an arithmetic progression; since the reciprocal of zero is undefined, zero can never be a term of an H.P. The nth term of an H.P. is found by first finding the nth term of the corresponding A.P. of reciprocals, then taking its reciprocal.

A number H is the harmonic mean between a and b if a, H, b are in H.P., giving $H = 2ab/(a+b)$. To insert n harmonic means between a and b, first insert n arithmetic means between 1/a and 1/b, then take the reciprocal of each one.

6.7 Sigma Notation and Summation Formulas

The Greek letter sigma (Σ) is used as shorthand for a sum: the sum from k=1 to n of a_k means a1+a2+...+an, where k is the index of summation and 1 and n are its lower and upper limits. Summation obeys two useful properties: the sum of a sum equals the sum of the sums (term by term), and a constant multiplier can be pulled outside the sigma.

Using a telescoping identity, three key summation formulas can be derived: the sum of k (from 1 to n) = n(n+1)/2, the sum of k^2 = n(n+1)(2n+1)/6, and the sum of k^3 = [n(n+1)/2]^2 -- these let more complicated series (like sums of squares of odd numbers) be evaluated by expanding the general term and summing each power separately.

6.8 Real-Life Applications

Many real-world growth and payment patterns follow arithmetic or geometric progressions: a lease payment that increases by a fixed amount each month is an arithmetic sequence (solved with the an and Sn formulas), while an investment earning fixed-percentage compound interest each year is a geometric sequence (solved with An = P(1+rate)^n), since each year's amount is a constant multiple of the previous year's.

Recognizing which type of progression a real-world scenario follows -- constant addition (arithmetic) or constant multiplication (geometric) -- is the key first step in solving problems involving vehicle traffic counts, simple interest, compound interest, population growth, and similar applied scenarios.

Important Definitions

What is a sequence?

A function whose domain is the set of natural numbers N; its terms are a1, a2, a3, ..., with the nth term an called the general term.

What is a finite sequence?

A sequence with a finite number of terms.

What is an infinite sequence?

A sequence with an infinite number of terms.

What is an arithmetic progression (A.P.)?

A sequence in which the difference between any term and the one before it is always the same constant, called the common difference.

What is the arithmetic mean (A.M.) between two numbers?

The number A such that a, A, b are in A.P.; A = (a+b)/2.

What is a geometric progression (G.P.)?

A sequence in which each term after the first is found by multiplying the previous term by a fixed non-zero constant called the common ratio.

What is the geometric mean (G.M.) between two numbers?

The number G such that a, G, b are in G.P.; G = plus or minus the square root of ab.

What is a harmonic progression (H.P.)?

A sequence whose reciprocals form an arithmetic progression.

What is a series?

The sum of the terms of a sequence.

What does sigma notation represent?

A shorthand way of writing the sum of a sequence of terms, from k=1 to n of a_k equals a1+a2+...+an.

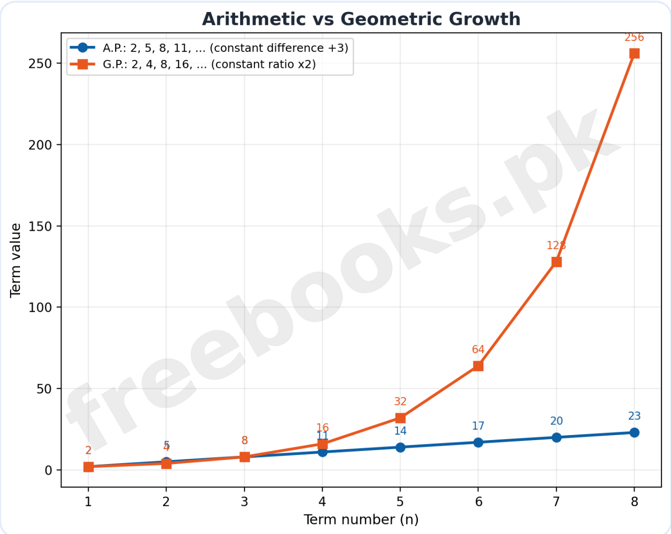
Key Facts and Relations

Topic	Key Fact / Relation
General term of an A.P.	an = a1 + (n-1)d
Arithmetic mean	A = (a+b) / 2
Sum of an arithmetic series	Sn = (n/2)(a1+an) = (n/2)[2a1+(n-1)d]
General term of a G.P.	an = a1 . r^(n-1)
Geometric mean	G = plus or minus square root of (ab)
Sum of a finite geometric series	Sn = a1(1-r^n)/(1-r), r not equal to 1
Sum to infinity of a geometric series	S(infinity) = a1/(1-r), valid only when |r| < 1

Topic	Key Fact / Relation
Harmonic mean	H = 2ab / (a+b)
Sum of first n natural numbers	sum of k = n(n+1)/2
Sum of squares of first n natural numbers	sum of k^2 = n(n+1)(2n+1)/6
Sum of cubes of first n natural numbers	sum of k^3 = [n(n+1)/2]^2
Compound interest as a G.P.	An = P(1 + rate)^n

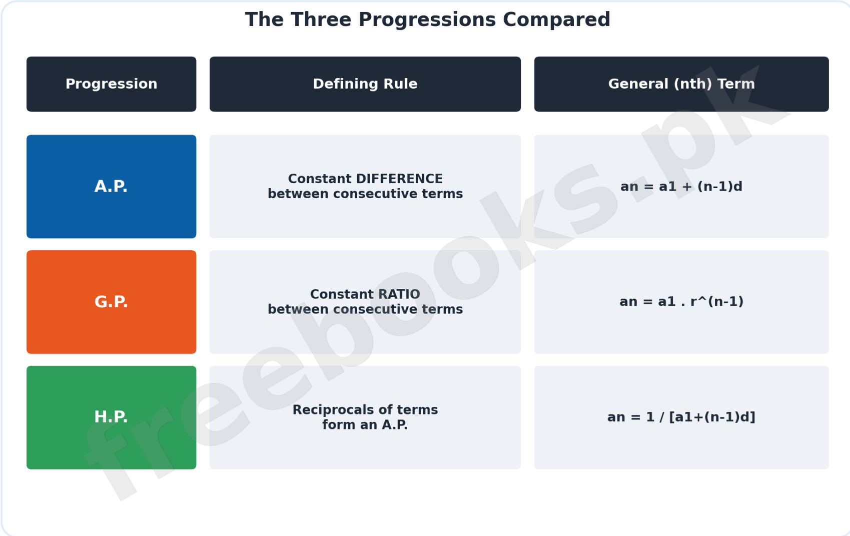
Diagrams

Arithmetic vs Geometric Growth: A line plot comparing an arithmetic progression (2, 5, 8, 11, ... growing by a constant difference of 3) against a geometric progression (2, 4, 8, 16, ... growing by a constant ratio of 2), visually showing linear growth versus exponential growth over 8 terms



Arithmetic vs Geometric Growth

The Three Progressions Compared: A summary table comparing arithmetic, geometric, and harmonic progressions, showing the defining rule and the general (nth) term formula for each

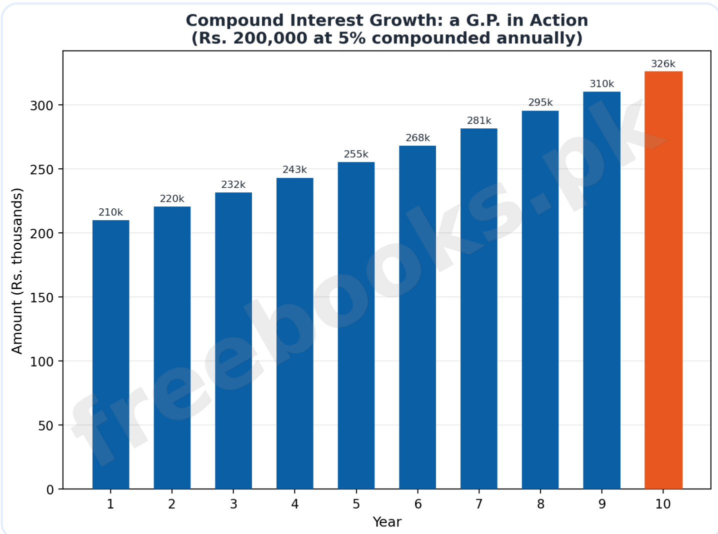


The Three Progressions Compared

Progression	Defining Rule	General (nth) Term
A.P.	Constant DIFFERENCE between consecutive terms	an = a1 + (n-1)d
G.P.	Constant RATIO between consecutive terms	an = a1 . r^(n-1)
H.P.	Reciprocals of terms form an A.P.	an = 1 / [a1+(n-1)d]

The Three Progressions Compared

Compound Interest Growth: a G.P. in Action: A bar chart showing an investment of Rs. 200,000 growing year by year at 5% compounded annually over 10 years, illustrating how compound interest forms a geometric progression



Compound Interest Growth: a G.P. in Action

Solved Examples

Example 1: Finding the General Term and a Specific Term of an A.P.

Problem: Find the general term and the 15th term of the A.P. whose first term is 4 and common difference is 5.

1. Use an = a1 + (n-1)d with a1=4, d=5: an = 4 + (n-1)(5) = 4 + 5n - 5.

2. Simplify: $a_n = 5n - 1$. This is the general term.
3. Substitute n=15 to find the 15th term: $a_{15} = 5(15) - 1 = 75 - 1 = 74$.
4. Final answer: general term $a_n = 5n-1$, and $a_{15} = 74$.

Example 2: Inserting Arithmetic Means

Problem: Insert three arithmetic means between 4 and 20.

1. Write 4, A_1, A_2, A_3, 20 as an A.P. with 5 terms, so $a_1=4$ and $a_5=20$.
2. Use $a_5 = a_1 + 4d$: $20 = 4 + 4d$, so $4d = 16$, giving $d = 4$.
3. $A_1 = a_1 + d = 4 + 4 = 8$. $A_2 = A_1 + d = 8 + 4 = 12$. $A_3 = A_2 + d = 12 + 4 = 16$.
4. Final answer: the three arithmetic means are 8, 12, 16.

Example 3: Sum of an Arithmetic Series

Problem: Find the sum of the first 30 terms of the arithmetic series $5+9+13+17+\ldots$

1. Identify $a_1=5$, $d=9-5=4$, $n=30$.
2. Use $S_n = (n/2)[2a_1+(n-1)d]$: $S_{30} = (30/2)[2(5)+(30-1)(4)] = 15[10+116]$.
3. Simplify inside the brackets: $10+116 = 126$.
4. Multiply: $S_{30} = 15(126) = 1890$.
5. Final answer: $S_{30} = 1890$.

Example 4: Finding a Specific Term of a G.P.

Problem: Find the 7th term of the G.P. 5, 15, 45, ...

1. Identify $a_1=5$ and $r=15/5=3$.
2. Use $a_n = a_1 \cdot r^{(n-1)}$: $a_7 = 5 \cdot 3^{(7-1)} = 5 \cdot 3^6$.
3. Compute $3^6 = 729$.

4. Multiply: a7 = 5(729) = 3645.
5. Final answer: a7 = 3645.

Example 5: Sum to Infinity of a Geometric Series

Problem: Find the sum to infinity of the G.P. 8, 4, 2, 1, ...

1. Identify a1=8 and r=4/8=1/2; since |r|<1, the sum to infinity exists.
2. Use S(infinity) = a1/(1-r): S(infinity) = 8/(1 - 1/2) = 8/(1/2).
3. Simplify: 8 divided by 1/2 equals 16.
4. Final answer: S(infinity) = 16.

Example 6: Finding the Harmonic Mean between Two Numbers

Problem: Find the harmonic mean between 6 and 10.

1. Use H = 2ab/(a+b) with a=6, b=10.
2. Compute the numerator: 2(6)(10) = 120.
3. Compute the denominator: 6+10 = 16.
4. Divide: H = 120/16 = 7.5.
5. Final answer: H = 7.5.

Example 7: Summing a Series Using a Sigma Notation Formula

Problem: Find the sum of the series 1^2+2^2+3^2+...+12^2 using the summation formula.

1. Use the formula sum of k^2 = n(n+1)(2n+1)/6 with n=12.
2. Substitute: sum = 12(13)(25)/6.
3. Compute the numerator: 12 x 13 = 156, then 156 x 25 = 3900.
4. Divide: 3900/6 = 650.

5. Final answer: the sum is 650.

Example 8: Application: Compound Interest as a Geometric Sequence

Problem: Farah invests Rs. 100,000 at 6% annual interest compounded annually. Find the total amount after 5 years.

1. This scenario is a G.P., since each year's amount is a constant multiple (1.06) of the previous year's amount.
2. Use the formula for the amount after n years: An = P(1+rate)^n = 100000(1.06)^5.
3. Compute (1.06)^5 by repeated multiplication: 1.06^2=1.1236, 1.06^3=1.19102, 1.06^4=1.26248, 1.06^5=1.33823 (approximately).
4. Multiply: A5 = 100000 x 1.33823 = Rs. 133,823 (approximately).
5. Final answer: the total amount after 5 years is approximately Rs. 133,823.

Short Questions & Answers

What is the general term formula for an A.P.?

an = a1 + (n-1)d.

What is the common ratio of a G.P.?

The constant multiplier between consecutive terms, found by dividing any term by the one before it.

What is the arithmetic mean between a and b?

A = (a+b)/2.

What is the geometric mean between a and b?

G = plus or minus the square root of ab.

What condition must |r| satisfy for a geometric series to have a sum to infinity?

|r| must be less than 1.

What is a harmonic progression?

A sequence whose reciprocals form an arithmetic progression.

What does the sigma symbol represent?

A shorthand notation for the sum of a sequence of terms.

Long Questions & Answers

Explain what an arithmetic progression is, and derive the formula for the sum of the first n terms of an arithmetic series.

What is an arithmetic progression, and what is its common difference?

An arithmetic progression (A.P.) is a sequence in which the difference between any term and the one immediately before it is always the same constant, called the common difference d, found as d = an - a(n-1) for any n greater than 1.

What is the formula for the nth term of an A.P.?

The nth (general) term is an = a1 + (n-1)d, where a1 is the first term and d is the common difference; this formula lets any specific term be found directly without listing every term before it.

How is the sum of the first n terms of an arithmetic series derived?

Writing Sn forwards and then backwards and adding the two versions term by term, every pair sums to (a1+an), and since there are n such pairs, 2Sn = n(a1+an), giving Sn = (n/2)(a1+an).

Why is there also a second version of the sum formula?

Substituting an = a1+(n-1)d into Sn=(n/2)(a1+an) gives Sn = (n/2)[2a1+(n-1)d], which is more convenient when the last term an is not directly known but a1 and d are.

Describe geometric progressions and geometric series, including how the sum to infinity is derived and when it exists.

What is a geometric progression, and what is its common ratio?

A geometric progression (G.P.) is a sequence in which each term after the first is found by multiplying the previous term by a fixed non-zero constant r, called the common ratio, found by dividing any term by the one immediately before it.

What is the formula for the nth term of a G.P.?

The nth term is an = a1 . r^(n-1), derived by repeatedly multiplying the first term by r one more time for each step further into the sequence.

How is the sum of the first n terms of a geometric series derived?

Writing Sn and then rSn (Sn multiplied by the common ratio), subtracting the two cancels every middle term, leaving Sn(1-r) = a1(1-r^n), so Sn = a1(1-r^n)/(1-r) for r not equal to 1.

When does the sum to infinity of a geometric series exist, and what is its formula?

The sum to infinity exists only when |r| < 1, because in that case r^n approaches 0 as n grows without bound, reducing the finite-sum formula to S(infinity) = a1/(1-r); if |r| is 1 or greater, the terms do not shrink and the series has no finite sum.

Multiple Choice Questions (MCQs)

The nth term of an A.P. is given by:

A. a1 . r^(n-1)
B. a1 + (n-1)d
C. a1 + nd
D. a1/(n-1)
Answer: B. a1 + (n-1)d
Explanation: The general term of an A.P. is an = a1 + (n-1)d, using the common difference d.

Three numbers a, b, c are in A.P. if and only if:

A. $b^2 = ac$

B. $2b = a+c$

C. $b = a+c$

D. $ac = a+c$

Answer: B. $2b = a+c$

Explanation: The middle term of three numbers in A.P. equals half the sum of the outer two: $2b = a+c$.

The common ratio of a G.P. is found by:

A. Subtracting consecutive terms

B. Dividing a term by the one before it

C. Adding consecutive terms

D. Multiplying all terms

Answer: B. Dividing a term by the one before it

Explanation: The common ratio r is found by dividing any term by its immediately preceding term.

The arithmetic mean between a and b is:

A. square root of ab

B. $2ab/(a+b)$

C. $(a+b)/2$

D. $a+b$

Answer: C. $(a+b)/2$

Explanation: The arithmetic mean is the simple average, $A = (a+b)/2$.

The geometric mean between a and b is:

A. $(a+b)/2$

B. plus or minus square root of ab

C. $2ab/(a+b)$

D. ab

Answer: B. plus or minus square root of ab

Explanation: The geometric mean satisfies $G^2=ab$, so G = plus or minus the square root of ab.

The sum to infinity of a geometric series exists only when:

A. $r > 1$

B. $|r| < 1$

C. r = 0

D. r is negative

Answer: B. |r| < 1

Explanation: The infinite sum formula only converges when the common ratio's absolute value is less than 1.

A harmonic progression is a sequence whose:

A. Terms are constant

B. Squares are in A.P.

C. Reciprocals are in A.P.

D. Reciprocals are in G.P.

Answer: C. Reciprocals are in A.P.

Explanation: A harmonic progression is defined as a sequence whose reciprocals form an arithmetic progression.

The formula for the sum of the first n natural numbers is:

A. n(n+1)/2

B. n(n+1)(2n+1)/6

C. [n(n+1)/2]^2

D. n^2

Answer: A. n(n+1)/2

Explanation: The sum of the first n natural numbers is n(n+1)/2.

The harmonic mean between a and b is:

A. (a+b)/2

B. plus or minus square root of ab

C. 2ab/(a+b)

D. a-b

Answer: C. 2ab/(a+b)

Explanation: The harmonic mean is H = 2ab/(a+b).

A fixed-percentage compound interest investment forms:

A. An A.P.

B. A G.P.

C. An H.P.

D. None of these

Answer: B. A G.P.

Explanation: Since each year's amount is a constant multiple of the previous year's, compound interest forms a geometric progression.

Quick Revision Summary

- A sequence is a function whose domain is the natural numbers; its terms are a1, a2, ..., an, ...
- An A.P. has a constant common difference d; general term an = a1 + (n-1)d
- Arithmetic mean: A = (a+b)/2; sum of an arithmetic series: Sn = (n/2)(a1+an) = (n/2)[2a1+(n-1)d]
- A G.P. has a constant common ratio r; general term an = a1 . r^(n-1)
- Geometric mean: G = plus or minus square root of ab; sum of a finite geometric series: Sn = a1(1-r^n)/(1-r), r not equal to 1
- Sum to infinity of a geometric series: S(infinity) = a1/(1-r), only when |r| < 1
- A harmonic progression is a sequence whose reciprocals form an A.P.
- Harmonic mean: H = 2ab/(a+b)
- Sigma notation is shorthand for a sum: sum from k=1 to n of ak = a1+a2+...+an
- Key summation formulas: sum of k = n(n+1)/2, sum of k^2 = n(n+1)(2n+1)/6, sum of k^3 = [n(n+1)/2]^2
- Constant addition each step signals an A.P.; constant multiplication each step signals a G.P.
- Real-world compound interest, population growth, and depreciation problems are typically G.P.s; fixed-increment payments and simple interest are typically A.P.s

Exam Tips

- Always identify whether a sequence has a constant DIFFERENCE (A.P.) or a constant RATIO (G.P.) before choosing a formula
- When the last term an is already known, use Sn=(n/2)(a1+an) -- it is faster than the expanded version
- Remember the sum to infinity only exists for a G.P. when |r| < 1 -- never apply it otherwise
- For H.P. problems, always work with the reciprocals (which form an A.P.) first, then take the reciprocal of your final answer

- In real-life applications, translate the wording carefully: 'increases by a fixed amount' means A.P., 'increases by a fixed percentage' means G.P.
- Double check by substituting your answer back into the original sequence or series to confirm it fits the given pattern

Unit 7: Permutations and Combinations

This unit builds the essential counting tools used throughout probability, computer science, and everyday decision-making: the fundamental principle of counting, factorial notation, permutations, and combinations. Permutations count the number of ways objects can be arranged in a specific order, while combinations count the number of ways objects can be selected without regard to order -- and knowing which one a real-world problem calls for is the single most important skill this unit develops.

Building on the basic nPr and nCr formulas, the unit also covers permutations of objects that are not all different (like the letters of a word with repeated letters), circular permutations (arranging objects around a table or a ring), and the complementary combination identity. It closes with real-life applications spanning cryptographic password counting, lottery odds, and selecting teams or committees -- all situations where recognizing whether order matters is the key first step.

Learning Objectives

- Apply the fundamental principle of counting to determine the total number of outcomes of combined events
- Define factorial notation and evaluate expressions involving factorials
- Define a permutation and evaluate nPr for arranging r objects out of n
- Find the number of permutations of objects when some objects are alike (not all different)
- Distinguish between the two cases of circular permutations and evaluate them
- Define a combination and evaluate nCr for selecting r objects out of n
- Apply the complementary combination property nCr = nC(n-r)
- Distinguish between permutations and combinations, and choose the correct one for a given real-world problem
- Apply permutations and combinations to real-life situations such as passwords, lotteries, and selecting teams or committees

Key Concepts

7.1 Fundamental Principle of Counting and Factorial Notation

If event A can occur in m ways and event B can occur in n ways, the two events together (one after the other) can occur in m x n ways; this extends to any number of events by multiplying the number of ways for each. This principle underlies both permutations and combinations.

The factorial of a natural number n, written n!, is the product n(n-1)(n-2)...3.2.1, with the special definition 0!=1 (chosen so that formulas like nPn=n!/0! remain consistent). Factorials appear throughout permutation and combination formulas, and the useful identity n! = n(n-1)! lets factorial expressions be simplified without full expansion.

7.2 Permutations

An arrangement of all or part of a set of objects in a specific order is called a permutation; the number of permutations of r objects chosen from a set of n distinct objects is nPr = n!/(n-r)!, valid for r less than or equal to n. Since order matters in a permutation, ab and ba count as two different permutations.

The formula follows directly from the fundamental principle of counting: the first of the r positions can be filled in n ways, the second in (n-1) ways (since one object is already used), and so on down to the rth position, which can be filled in (n-r+1) ways; multiplying these together and simplifying gives nPr = n!/(n-r)!.

7.3 Permutations of Objects Not All Different, and Circular Permutations

When some of the n objects being arranged are identical, straightforward n! overcounts the arrangements, since swapping identical objects produces no visibly different result. If there are n1 objects of one kind, n2 of a second kind, and so on, the number of distinct permutations of all n objects is n!/(n1! n2! n3! ...).

A circular permutation arranges objects around a circle rather than in a row; because rotating the whole circle produces no new arrangement, n objects in a circle give only (n-1)! distinct arrangements (not n!) when clockwise and anticlockwise orders are considered different. When a circular arrangement and its mirror-image reflection are considered the same (as with beads on a ring), the count is halved again to (n-1)!/2.

7.4 Combinations

A combination of r objects taken from a set of n objects is simply a subset of r objects -- order does not matter, so selecting {Ahmad, Sana} is the same combination as selecting {Sana, Ahmad}. The number of combinations of n objects taken r at a time is nCr = n!/[r!(n-r)!].

Since each combination of r objects can itself be internally arranged in r! different orders (each giving a different permutation), the relationship nCr x r! = nPr connects the two formulas directly, and dividing through by r! derives the combination formula from the permutation formula.

7.5 Complementary Combinations and Permutation vs Combination

The complementary combination identity nCr = nC(n-r) reflects a simple symmetry: choosing r objects to include is equivalent to choosing the remaining (n-r) objects to exclude. This identity is especially useful for evaluating nCr when r is more than half of n, since nC(n-r) with the smaller value of (n-r) is quicker to compute by hand.

The essential distinction between the two counting tools is whether order matters: use a permutation whenever the arrangement or sequence of the chosen objects matters (like arranging books on a shelf or people in a queue), and use a combination whenever only the makeup of the selected group matters, not its order (like choosing team members or committee members).

7.6 Real-Life Applications

Permutations and combinations appear throughout daily life and technology: permutations count ordered outcomes like passwords (when character order matters), rankings, and seating arrangements, while combinations count unordered selections like lottery number choices, team rosters, and committee memberships.

Recognizing whether a real-world scenario cares about order (calling for nPr, or a product of factorials if the count is built from independent stages) or does not care about order (calling for nCr) is the essential first step in solving any applied counting problem, from cryptographic password counting to estimating the odds of winning a lottery.

Important Definitions

What is the fundamental principle of counting?

If event A can occur in m ways and event B can occur in n ways, both events together can occur in m x n ways.

What is n factorial (n!)?

The product of all natural numbers from n down to 1: n! = n(n-1)(n-2)...3.2.1, with 0! defined as 1.

What is a permutation?

An arrangement of all or part of a set of objects in a specific order.

What is the formula for nPr?

nPr = n!/(n-r)!, the number of ways to arrange r objects chosen from n distinct objects.

What is a circular permutation?

An arrangement of objects around a circle rather than in a row.

How many distinct circular permutations exist for n objects when clockwise and anticlockwise arrangements are different?

(n-1)!

What is a combination?

A selection of r objects from a set of n objects, where order does not matter.

What is the formula for nCr?

nCr = n!/[r!(n-r)!], the number of ways to choose r objects from n distinct objects.

What is the complementary combination identity?

nCr = nC(n-r).

What is the key difference between a permutation and a combination?

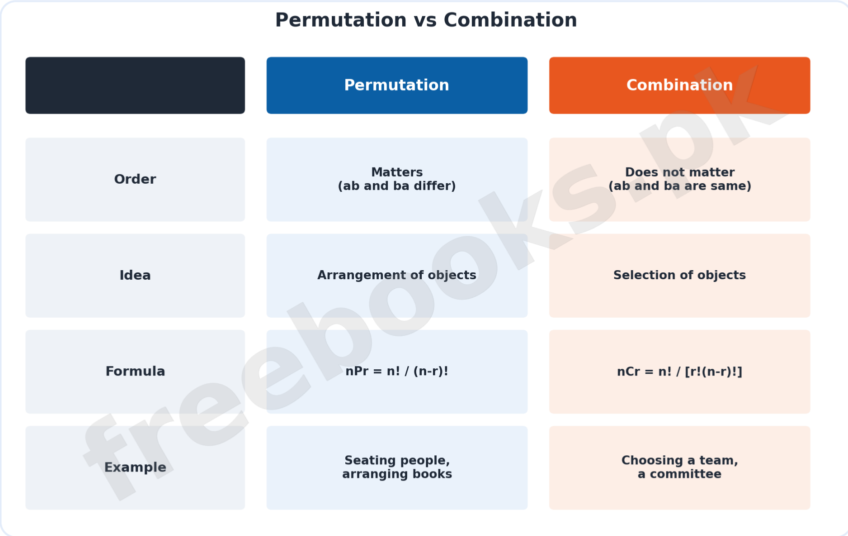
In a permutation, order matters; in a combination, order does not matter.

Key Facts and Relations

Topic	Key Fact / Relation
Fundamental principle of counting	m x n ways for two events; extends to m . n . k ... for more events
Factorial	n! = n(n-1)(n-2)...3.2.1, with 0! = 1
Useful factorial identity	n! = n(n-1)!
Permutation formula	nPr = n! / (n-r)!
Permutation with repeated objects	n! / (n1! n2! n3! ...)
Circular permutation (clockwise different from anticlockwise)	(n-1)!
Circular permutation (clockwise same as anticlockwise)	(n-1)! / 2
Combination formula	nCr = n! / [r!(n-r)!]
Relationship between nPr and nCr	nCr = nPr / r!
Complementary combinations	nCr = nC(n-r)

Diagrams

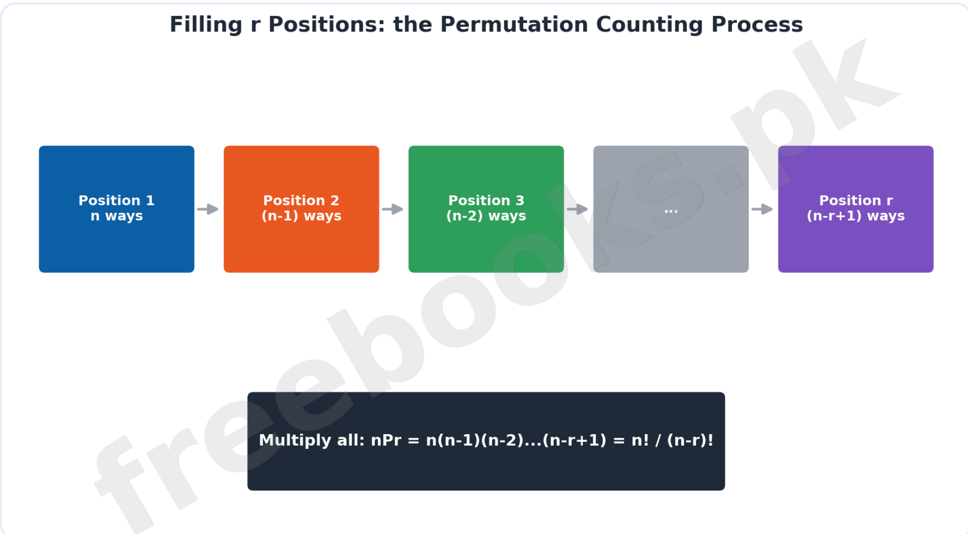
Permutation vs Combination: A summary table comparing permutations and combinations, showing whether order matters, the underlying idea, the formula, and a typical real-world example for each



	Permutation	Combination
Order	Matters (ab and ba differ)	Does not matter (ab and ba are same)
Idea	Arrangement of objects	Selection of objects
Formula	nPr = n! / (n-r)!	nCr = n! / [r!(n-r)!]
Example	Seating people, arranging books	Choosing a team, a committee

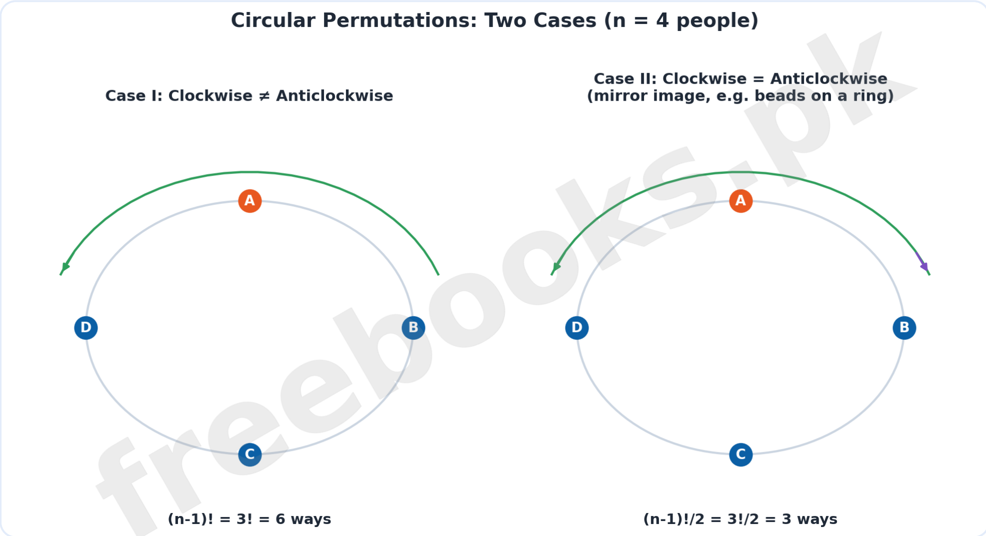
Permutation vs Combination

Filling r Positions: the Permutation Counting Process: A flow diagram showing how each of the r positions in a permutation is filled -- n ways for the first position, (n-1) ways for the second, and so on down to (n-r+1) ways for the rth position -- which multiply together to give nPr = n!/(n-r)!



Filling r Positions: the Permutation Counting Process

Circular Permutations: Two Cases: A schematic showing 4 people (A, B, C, D) seated around a circle, illustrating Case I (clockwise and anticlockwise arrangements considered different, giving (n-1)! = 6 ways) alongside Case II (clockwise and anticlockwise arrangements considered identical, as with beads on a ring, giving (n-1)!/2 = 3 ways)



Circular Permutations: Two Cases

Solved Examples

Example 1: Evaluating a Factorial Expression

Problem: Evaluate 9!/(6!3!).

1. Expand 9! as 9x8x7x6! so it cancels with the 6! in the denominator: 9!/(6! 3!) = (9x8x7x6!)/(6!x3!) = (9x8x7)/3!.
2. Compute the numerator: 9x8x7 = 504.
3. Compute 3! = 3x2x1 = 6.
4. Divide: 504/6 = 84.
5. Final answer: 9!/(6!3!) = 84.

Example 2: Counting Permutations (Order Matters)

Problem: How many different 4-digit numbers can be formed from the digits 1,2,3,4,5,6,7 when no digit is repeated?

1. Identify n=7 (total digits) and r=4 (digits used per number); since digits form an ordered number, this is a permutation.
2. Use nPr = n!/(n-r)!: 7P4 = 7!/3!.
3. Expand: 7!/3! = (7x6x5x4x3!)/3! = 7x6x5x4.

4. Multiply: 7x6=42, 42x5=210, 210x4=840.
5. Final answer: 840 different 4-digit numbers.

Example 3: Permutations of Objects Not All Different

Problem: In how many ways can the letters of the word ENGINEERING be arranged when all the letters are used?

1. Total letters = 11. Count repeated letters: E appears 3 times, N appears 3 times, G appears 2 times, I appears 2 times, R appears once.
2. Use n!/(n1! n2! n3! n4!) = 11!/(3! 3! 2! 2!).
3. Compute 11! = 39,916,800, and 3! x 3! x 2! x 2! = 6x6x2x2 = 144.
4. Divide: 39,916,800/144 = 277,200.
5. Final answer: 277,200 distinct arrangements.

Example 4: Circular Permutations

Problem: In how many ways can 6 people be seated at a round table if clockwise and anticlockwise arrangements are considered different?

1. Since clockwise and anticlockwise orders are different, use Case I: (n-1)!.
2. Substitute n=6: (6-1)! = 5!.
3. Compute 5! = 5x4x3x2x1 = 120.
4. Final answer: 120 ways.

Example 5: Counting Combinations (Order Doesn't Matter)

Problem: A committee of 4 students is to be chosen from a class of 10 students. In how many ways can this be done?

1. Since the committee is a selection with no regard to order, use nCr with n=10, r=4.
2. nCr = n!/[r!(n-r)!] = 10!/(4!6!).

3. Expand: 10!/(4!6!) = (10x9x8x7x6!)/(4!x6!) = (10x9x8x7)/4!.
4. Compute the numerator: 10x9=90, 90x8=720, 720x7=5040. Compute 4! = 24.
5. Divide: 5040/24 = 210.
6. Final answer: 210 ways.

Example 6: Complementary Combinations

Problem: Evaluate 14C11 using the complementary combination identity.

1. Use the identity nCr = nC(n-r): 14C11 = 14C(14-11) = 14C3.
2. Compute 14C3 = (14x13x12)/(3x2x1).
3. Numerator: 14x13=182, 182x12=2184. Denominator: 3x2x1=6.
4. Divide: 2184/6 = 364.
5. Final answer: 14C11 = 364.

Example 7: Combining Combinations for a Selection with Restrictions

Problem: A group has 7 boys and 5 girls. In how many ways can a team of 4 be selected with exactly 2 boys and 2 girls?

1. Select 2 boys from 7: 7C2 = 7!/(2!5!) = (7x6)/(2x1) = 21.
2. Select 2 girls from 5: 5C2 = 5!/(2!3!) = (5x4)/(2x1) = 10.
3. Since these two selections happen together, multiply by the fundamental principle of counting: 21 x 10 = 210.
4. Final answer: 210 ways.

Example 8: Application: Counting Cryptographic Passwords

Problem: A password must be exactly 5 characters long, using only lowercase letters (a-z, 26 letters). Find the number of possible passwords (i) if repetition of letters is not allowed, and (ii) if repetition is allowed.

1. (i) Without repetition, this is a permutation of 5 characters chosen from 26: 26P5 = 26!/(26-5)! = 26x25x24x23x22.
2. Compute: 26x25=650, 650x24=15,600, 15,600x23=358,800, 358,800x22=7,893,600.
3. (ii) With repetition allowed, each of the 5 positions independently has 26 choices, so by the fundamental principle of counting the total is 26^5.
4. Compute 26^5 = 11,881,376.
5. Final answer: without repetition, 7,893,600 passwords; with repetition, 11,881,376 passwords.

Short Questions & Answers

What is the fundamental principle of counting?

If one event can occur in m ways and another in n ways, both together can occur in m x n ways.

What is 0! defined as?

1.

What is the formula for nPr?

nPr = n!/(n-r)!.

What is the formula for nCr?

nCr = n!/[r!(n-r)!].

How many distinct circular permutations exist for n objects when clockwise and anticlockwise are different?

(n-1)!.

What is the complementary combination identity?

nCr = nC(n-r).

When should a permutation be used instead of a combination?

When the order of the arrangement matters.

Long Questions & Answers

Explain what a permutation is, derive the formula for nPr, and describe how the formula changes when some of the objects being arranged are identical.

What is a permutation, and how does it differ from a combination?

A permutation is an arrangement of all or part of a set of objects in a specific order, so ab and ba count as two different permutations; a combination, by contrast, is a selection where order does not matter.

How is the formula nPr = n!/(n-r)! derived?

By the fundamental principle of counting, the first of the r positions can be filled in n ways, the second in (n-1) ways, and so on down to the rth position in (n-r+1) ways; multiplying these together and rewriting the product using factorials gives nPr = n!/(n-r)!.

Why is 0! defined as 1?

When r=n, nPn should equal n! (all n objects arranged with no leftover), but the formula gives n!/(n-n)! = n!/0!; for this to equal n!, 0! must be defined as 1.

How does the permutation formula change when some objects are identical?

If among n objects there are n1 identical objects of one kind, n2 of a second kind, and so on, the count n! overcounts every group of identical objects' internal rearrangements, so the corrected formula divides by the factorial of each group's size: n!/(n1! n2! n3! ...).

Describe circular permutations, explaining both cases based on whether clockwise and anticlockwise arrangements are considered the same or different, with formulas.

Why does a circular arrangement have fewer distinct permutations than a linear one?

In a linear arrangement, every rotation gives a visibly different order, but in a circular arrangement, rotating the entire circle produces the same relative arrangement, so each set of n linear permutations that are rotations of one another collapses into just 1 circular permutation.

What is the formula when clockwise and anticlockwise arrangements are considered different (Case I)?

Since each circular arrangement corresponds to n rotations of a single linear arrangement, the n! linear permutations collapse to $n!/n = (n-1)!$ distinct circular permutations.

What is the formula when clockwise and anticlockwise arrangements are considered identical (Case II)?

When a circular arrangement and its mirror-image reflection (reversing the direction) are not considered different, the Case I count is further divided by 2, giving $(n-1)!/2$ distinct circular permutations.

Give an example illustrating the difference between the two cases.

For 4 people seated at a round table, Case I gives $(4-1)! = 3! = 6$ distinct seatings when direction matters, while Case II (such as 4 beads arranged on a ring, where flipping the ring over gives the same pattern) gives $(4-1)!/2 = 3!/2 = 3$ distinct arrangements.

Multiple Choice Questions (MCQs)

The fundamental principle of counting states that if event A occurs in m ways and event B in n ways, both together occur in:

A. m+n ways

B. m x n ways

C. m-n ways

D. m/n ways

Answer: B. m x n ways

Explanation: The events combine multiplicatively: m x n total ways.

0! is defined as:

A. 0

B. 1

C. Undefined

D. -1

Answer: B. 1

Explanation: 0! is defined as 1 so that formulas like nPn = n!/0! remain consistent.

The formula for nPr is:

A. n!/(n-r)!

B. n!/[r!(n-r)!]

C. n!/r!

D. r!/n!

Answer: A. n!/(n-r)!

Explanation: The permutation formula is nPr = n!/(n-r)!.

In a permutation, ab and ba are considered:

A. The same

B. Different

C. Undefined

D. Equal only if a=b

Answer: B. Different

Explanation: Order matters in a permutation, so ab and ba are two distinct permutations.

The number of distinct arrangements of n objects with n1 alike of one kind is:

A. n!

B. n!/n1!

C. n1!/n!

D. n! x n1!

Answer: B. n!/n1!

Explanation: Dividing n! by n1! corrects for the overcounting caused by the identical objects.

The number of distinct circular permutations of n objects (clockwise different from anticlockwise) is:

A. n!

B. (n-1)!

C. (n-1)!/2

D. n!/2

Answer: B. (n-1)!

Explanation: Rotating the circle gives no new arrangement, so the count reduces from n! to (n-1)!.

The formula for nCr is:

A. n!/(n-r)!

B. n!/[r!(n-r)!]

C. n!/r!

D. r!(n-r)!/n!

Answer: B. n!/[r!(n-r)!]

Explanation: The combination formula divides the permutation formula by r! to remove ordering.

In a combination, the order of the selected objects:

A. Matters

B. Does not matter

C. Always increases the count

D. Is always fixed

Answer: B. Does not matter

Explanation: A combination is a selection where order is irrelevant.

The complementary combination identity states that nCr equals:

A. nC(r-1)

B. nCn

C. nC(n-r)

D. nPr

Answer: C. nC(n-r)

Explanation: Choosing r objects to include is equivalent to choosing the remaining (n-r) objects to exclude: nCr = nC(n-r).

A committee selection problem (where order doesn't matter) should be solved using:

A. A permutation

B. A combination

C. A factorial alone

D. The fundamental principle of counting alone

Answer: B. A combination

Explanation: Since committee membership does not depend on order, a combination (nCr) is the correct tool.

Quick Revision Summary

- The fundamental principle of counting: if event A occurs in m ways and event B in n ways, both together occur in m x n ways
- n! = n(n-1)(n-2)...3.2.1, with 0! defined as 1
- Permutation: an arrangement where ORDER matters; nPr = n!/(n-r)!
- Permutations with repeated objects: n!/(n1! n2! n3! ...)
- Circular permutations (clockwise different from anticlockwise): (n-1)!
- Circular permutations (clockwise same as anticlockwise, e.g. a ring of beads): (n-1)!/2
- Combination: a selection where ORDER does not matter; nCr = n!/[r!(n-r)!]
- Relationship between the two: nCr = nPr / r!
- Complementary combinations: nCr = nC(n-r) -- useful when r is more than half of n
- Use a permutation for arrangements/orderings; use a combination for selections/groupings
- Multi-stage selections (e.g. choosing boys AND girls separately) are combined by multiplying the individual combination counts
- Real-life uses include passwords, lottery odds, seating arrangements, and committee selection

Exam Tips

- Before solving, always ask: does order matter? If yes, use a permutation; if no, use a combination
- Remember 0! = 1 -- a common point of confusion in factorial calculations
- For circular permutations, check whether the problem treats mirror-image (flipped) arrangements as the same or different
- Use the complementary identity nCr = nC(n-r) to simplify calculations when r is large (closer to n than to 0)
- When a problem has repeated identical objects, divide n! by the factorial of each group's size to avoid overcounting
- For multi-category selections, compute each category's combination separately, then multiply the results together

Unit 8: Mathematical Inductions and Binomial Theorem

This unit introduces mathematical induction, a powerful proof technique for statements involving natural numbers, and the binomial theorem, a formula for expanding a binomial raised to any power. Since a formula cannot be checked against infinitely many natural numbers one at a time, mathematical induction proves it true for all of them using just two finite steps: a base case, and a step showing the statement always carries forward from one integer to the next.

The unit then develops the binomial theorem for a positive integral index, including the general term, the middle term, and useful deductions like the sum of the binomial coefficients, along with Pascal's triangle as a quick way to find those coefficients. It closes with the binomial theorem for a negative or fractional index (an infinite series valid only for small x), and real-life applications: approximating roots, finding remainders and last digits, testing divisibility, comparing large numbers, and forecasting compound interest.

Learning Objectives

- State the principle of mathematical induction and use it to prove a formula true for all positive integers
- Apply the principle of extended mathematical induction to prove statements true for all integers n greater than or equal to some starting value i
- State and apply the binomial theorem to expand (a+b)^n for a positive integer index n
- Find a specified term, the general term, and the middle term(s) in a binomial expansion
- Derive and apply key deductions from the binomial expansion, including the sum of the binomial coefficients
- Use Pascal's triangle to find binomial coefficients and expand a binomial
- State and apply the binomial theorem when the index is a negative integer or a fraction

- Apply the binomial theorem to find approximate values, remainders, last digits, and to compare large numbers and solve compound interest problems

Key Concepts

8.1 The Principle of Mathematical Induction

Since a formula cannot be checked for every one of infinitely many natural numbers directly, the principle of mathematical induction proves it in two finite steps: verify the base case S(1) is true, then show that whenever S(k) is true for some positive integer k, S(k+1) must also be true (the induction of hypothesis). If both steps hold, S(n) is concluded true for every positive integer n.

A single case that fails a formula, called a counterexample, is enough to disprove it entirely -- which is exactly why observing a pattern in the first several cases is never a substitute for a real proof; the classic example is S(n)=n^2-n+41, which is prime for n=1 through 40 but fails at n=41.

8.2 The Principle of Extended Mathematical Induction

Some statements are not true starting from n=1 but from some other integer i (which may even be 0 or negative); the extended principle proves such statements by verifying the base case at S(i) instead of S(1), then showing S(k+1) follows from S(k) for every integer k >= i, concluding S(n) true for all integers n >= i.

The same two-method approach used in ordinary induction -- proving S(k+1) either by direct algebraic manipulation of S(k) or by adding/multiplying an extra term onto both sides of S(k) -- applies equally to extended induction; only the starting point of the base case changes.

8.3 The Binomial Theorem for a Positive Integral Index

For any positive integer n, (a+b)^n expands as the sum from r=0 to n of C(n,r) a^(n-r) b^r, where the C(n,r) values are called the binomial coefficients; this can itself be proved using mathematical induction (assuming it holds for n=k, then showing it holds for n=k+1 using Pascal's identity C(k,r)+C(k,r-1)=C(k+1,r)).

Several patterns hold in every binomial expansion: it has $n+1$ terms; the exponent of a decreases from n to 0 while the exponent of b increases from 0 to n (their sum is always n); and coefficients equidistant from the two ends are equal, since $C(n,r)=C(n,n-r)$. The $(r+1)$th term, $T(r+1) = C(n,r)\, a^{(n-r)}\, b^r$, is called the general term, since every term of the expansion can be generated from it.

8.4 The Middle Term and Deductions from the Binomial Expansion

A binomial expansion has $n+1$ terms; if n is even, there is exactly one middle term, the $(n/2 + 1)$th term, while if n is odd, there are two middle terms, the $((n+1)/2)$th and $((n+3)/2)$th terms.

Substituting particular values for a and b in the general expansion yields useful identities: setting $a=1$ gives the expansion of $(1+b)^n$ directly in terms of the binomial coefficients; setting $a=b=1$ shows the sum of all the binomial coefficients equals 2^n; and setting $a=1$, $b=-1$ shows the sum of the coefficients at odd positions equals the sum of the coefficients at even positions.

8.5 Pascal's Triangle

Pascal's triangle is a triangular arrangement of the binomial coefficients where the first row ($n=0$) is a single 1, every row begins and ends with 1, and every other interior entry is the sum of the two entries diagonally above it in the previous row -- this rule is expressed algebraically as Pascal's Rule, $C(n,k) = C(n-1,k-1) + C(n-1,k)$.

The entries in the nth row of Pascal's triangle are exactly the binomial coefficients $C(n,0), C(n,1), \ldots, C(n,n)$ needed to expand $(a+b)^n$, giving a quick way to find them without computing factorials directly, especially useful for smaller values of n.

8.6 The Binomial Theorem for a Negative or Fractional Index

When n is a negative integer or a fraction, $(1+x)^n$ no longer terminates after $n+1$ terms; instead it becomes an infinite series, valid only when $|x|<1$, given by $1 + nx + [n(n-1)/2!]x^2 + [n(n-1)(n-2)/3!]x^3 + \ldots$, called the binomial series. In this case the symbols $C(n,0)$, $C(n,1)$, etc. are not meaningful, since n is not a non-negative integer, so the coefficients must be written out using the general pattern of the numerator product instead.

Since $|x|<1$ means each successive power of x is numerically smaller than the one before it, the terms of the series shrink in size, which is what allows just the first few terms to give a good approximation of the expansion's true value -- the basis for every approximation and infinite-series-summation application in this unit.

8.7 Applications of the Binomial Theorem

Because the terms of a valid binomial series shrink rapidly, keeping only the first few terms gives an accurate approximate value for expressions like cube roots or square roots of numbers close to a perfect power (e.g. finding the cube root of 30 by writing it as 27(1+3/27) and expanding (1+1/9)^(1/3)).

Rewriting a large power as (1+k)^n (or (k-1)^n) and expanding via the binomial theorem is a powerful tool for finding remainders after division, the last digits of a large number, proving divisibility, comparing the size of two large numbers, and forecasting compound-interest growth -- since every term of the expansion after the first one or two is typically a clean multiple of some convenient number, isolating exactly the piece of the expansion that matters for the question being asked.

Important Definitions

What is a counterexample?
A single case that fails a mathematical statement, proving it false.

What is the principle of mathematical induction?
A two-step method to prove a statement true for all positive integers: verify the base case S(1), then show S(k+1) is true whenever S(k) is true.

What is the base case in a mathematical induction proof?
The first step, showing the statement is true for the starting value of n (usually n=1).

What is the induction of hypothesis?

Assuming the statement is true for n=k, then proving it must also be true for n=k+1.

What is the extended principle of mathematical induction used for?

Proving a statement true for all integers n greater than or equal to some starting integer i, not necessarily 1.

What is the binomial theorem?

A formula for expanding (a+b)^n as a sum of terms involving binomial coefficients C(n,r).

What is the general term of a binomial expansion?

T(r+1) = C(n,r) a^(n-r) b^r, from which every term of the expansion can be generated.

What are binomial coefficients?

The numbers C(n,0), C(n,1), ..., C(n,n) that appear in the expansion of (a+b)^n.

What is Pascal's Rule?

C(n,k) = C(n-1,k-1) + C(n-1,k), the rule that generates each entry of Pascal's triangle from the row above.

What condition must x satisfy for the binomial series expansion of (1+x)^n (n negative or fractional) to be valid?

|x| must be less than 1.

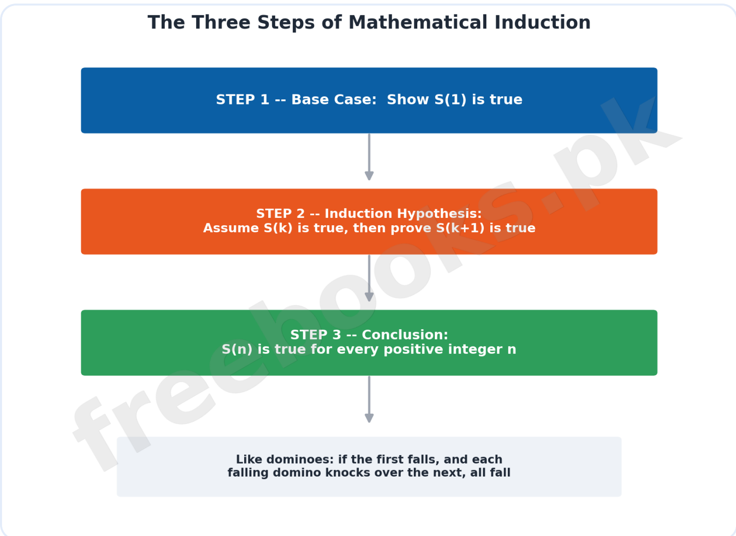
Key Facts and Relations

Topic	Key Fact / Relation
Mathematical induction steps	Base case S(1) true; S(k) true implies S(k+1) true; conclusion S(n) true for all n
Binomial theorem	(a+b)^n = sum from r=0 to n of C(n,r) a^(n-r) b^r
Binomial coefficient	C(n,r) = n! / [r!(n-r)!]
General term	T(r+1) = C(n,r) a^(n-r) b^r
Middle term (n even)	the (n/2 + 1)th term
Middle terms (n odd)	the ((n+1)/2)th and ((n+3)/2)th terms

Topic	Key Fact / Relation
Sum of binomial coefficients	C(n,0)+C(n,1)+...+C(n,n) = 2^n
Sum of odd/even coefficients	sum of odd-position coefficients = sum of even-position coefficients
Pascal's Rule	C(n,k) = C(n-1,k-1) + C(n-1,k)
Binomial series (n negative/ fractional)	(1+x)^n = 1+nx+[n(n-1)/2!]x^2+..., valid only for |x| < 1

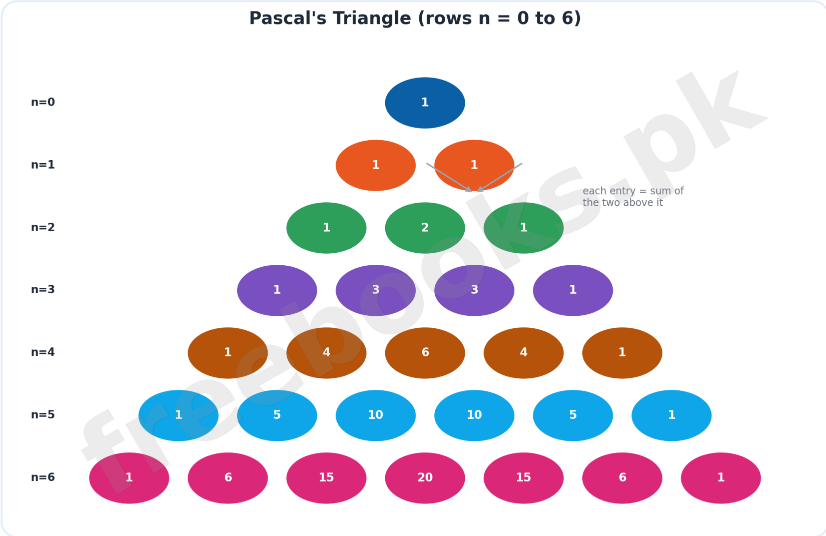
Diagrams

The Three Steps of Mathematical Induction: A flow diagram showing the base case (S(1) true), the induction hypothesis (assume S(k) true, prove S(k+1) true), and the conclusion (S(n) true for all positive integers n), compared to a chain of falling dominoes



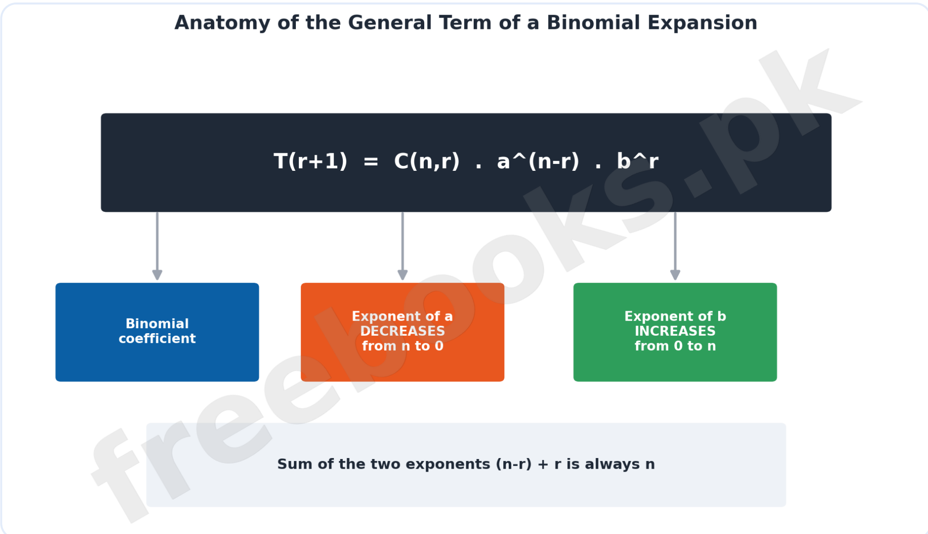
The Three Steps of Mathematical Induction

Pascal's Triangle: Rows n=0 through n=6 of Pascal's triangle, showing how each interior entry is the sum of the two entries directly above it, and how the entries of each row give the binomial coefficients for that value of n



Pascal's Triangle

Anatomy of the General Term of a Binomial Expansion: A labeled breakdown of the general term formula T(r+1) = C(n,r) a^(n-r) b^r, showing the binomial coefficient, the decreasing exponent of a, the increasing exponent of b, and the fact that the two exponents always sum to n



Anatomy of the General Term of a Binomial Expansion

Solved Examples

Example 1: Proving a Sum Formula by Mathematical Induction

Problem: Use mathematical induction to prove that 2+4+6+...+2n = n(n+1) for every positive integer n.

1. Base case: for n=1, LHS=2 and RHS=1(1+1)=2, so S(1) is true.
2. Induction hypothesis: assume S(k): 2+4+...+2k = k(k+1) is true.
3. Add 2(k+1) to both sides: 2+4+...+2k+2(k+1) = k(k+1)+2(k+1) = (k+1)(k+2), which matches the required form of S(k+1).
4. Conclusion: since both conditions hold, S(n) is true for every positive integer n.
5. Final answer: the formula 2+4+6+...+2n = n(n+1) is proven true for all positive integers n.

Example 2: Proving Divisibility by Mathematical Induction

Problem: Prove that 5^n - 1 is divisible by 4 for all positive integers n.

1. Base case: for n=1, 5^1-1=4, which is divisible by 4, so S(1) is true.
2. Induction hypothesis: assume 5^k - 1 = 4m for some integer m.
3. For n=k+1: 5^(k+1)-1 = 5.5^k-1 = 5(4m+1)-1 = 20m+4 = 4(5m+1), which is divisible by 4.
4. Conclusion: since both conditions hold, 5^n-1 is divisible by 4 for all positive integers n.
5. Final answer: proven true by mathematical induction.

Example 3: Extended Mathematical Induction with an Inequality

Problem: Prove that n! > 2^n for all integers n >= 4.

1. Base case: for n=4, 4!=24 and 2^4=16; since 24>16, S(4) is true.
2. Induction hypothesis: assume k! > 2^k is true for some integer k >= 4.

3. For n=k+1: (k+1)! = (k+1).k! > (k+1).2^k (by the hypothesis); since k>=4, (k+1)>=5>2, so (k+1).2^k > 2.2^k = 2^(k+1).
4. Therefore (k+1)! > 2^(k+1), so S(k+1) is true whenever S(k) is true.
5. Final answer: by the principle of extended mathematical induction, n! > 2^n is true for all integers n >= 4.

Example 4: Expanding a Binomial Using the Binomial Theorem

Problem: Expand (2x - 1/x)^5 using the binomial theorem.

1. Identify a=2x, b=-1/x, n=5; use T(r+1) = C(5,r)(2x)^(5-r)(-1/x)^r for r=0,1,2,3,4,5.
2. r=0: C(5,0)(2x)^5 = 32x^5. r=1: C(5,1)(2x)^4(-1/x) = 5(16x^4)(-1/x) = -80x^3.
3. r=2: C(5,2)(2x)^3(-1/x)^2 = 10(8x^3)(1/x^2) = 80x. r=3: C(5,3)(2x)^2(-1/x)^3 = 10(4x^2)(-1/x^3) = -40/x.
4. r=4: C(5,4)(2x)(-1/x)^4 = 5(2x)(1/x^4) = 10/x^3. r=5: C(5,5)(-1/x)^5 = -1/x^5.
5. Final answer: (2x - 1/x)^5 = 32x^5 - 80x^3 + 80x - 40/x + 10/x^3 - 1/x^5.

Example 5: Finding a Specific Term in a Binomial Expansion

Problem: Find the term involving x^4 in the expansion of (x^2 + 3/x)^8.

1. Use T(r+1) = C(8,r)(x^2)^(8-r)(3/x)^r = C(8,r) . 3^r . x^(16-3r).
2. Set the exponent of x equal to 4: 16-3r=4, so 3r=12, giving r=4.
3. Substitute r=4: T5 = C(8,4) . 3^4 . x^4.
4. Compute C(8,4)=70 and 3^4=81; multiply 70x81=5670.
5. Final answer: the term involving x^4 is 5670x^4.

Example 6: Finding the Middle Term

Problem: Find the middle term in the expansion of (x/3 + 3/x)^8.

1. Since n=8 is even, there is exactly one middle term: the (8/2+1)=5th term, T5.
2. Use T5 = C(8,4)(x/3)^4(3/x)^4.
3. Simplify (x/3)^4(3/x)^4 = (x^4/81)(81/x^4) = 1.
4. Compute C(8,4)=70, so T5 = 70 x 1 = 70.
5. Final answer: the middle term is 70.

Example 7: Using the Binomial Series for Approximation

Problem: Use the binomial theorem to find the approximate value of 1/sqrt(1.02), correct to three decimal places.

1. Write 1/sqrt(1.02) = (1+0.02)^(-1/2), so n=-1/2 and x=0.02 (since |x|<1, the expansion is valid).
2. Use (1+x)^n = 1+nx+[n(n-1)/2!]x^2+...: term1=1, term2 = (-1/2)(0.02) = -0.01.
3. term3 = [(-1/2)(-3/2)/2](0.02)^2 = [0.75/2](0.0004) = 0.00015.
4. Add the first three terms: 1 - 0.01 + 0.00015 = 0.99015.
5. Final answer: 1/sqrt(1.02) is approximately 0.990 (correct to three decimal places).

Example 8: Application: Finding a Remainder Using the Binomial Theorem

Problem: Using the binomial theorem, find the remainder when 7^50 is divided by 4.

1. Write 7^50 = (8-1)^50 and expand using the binomial theorem: the sum from r=0 to 50 of C(50,r) . 8^(50-r) . (-1)^r.

2. Every term with r<50 contains a positive power of 8, and since 8 is a multiple of 4, every such term is divisible by 4.
3. The only term without a factor of 8 is the last one (r=50): C(50,50) . 8^0 . (-1)^50 = 1.
4. So 7^50 = (a multiple of 4) + 1.
5. Final answer: the remainder when 7^50 is divided by 4 is 1.

Short Questions & Answers

What are the three steps of mathematical induction?

Base case, induction of hypothesis, and conclusion.

What is a counterexample?

A single case that disproves a mathematical statement.

What is the binomial theorem used for?

Expanding (a+b)^n as a sum of terms involving binomial coefficients.

What is the general term of a binomial expansion?

T(r+1) = C(n,r) a^(n-r) b^r.

How many middle terms does a binomial expansion have when n is odd?

Two.

What is Pascal's Rule?

C(n,k) = C(n-1,k-1) + C(n-1,k).

What condition must |x| satisfy for the binomial series of (1+x)^n (n negative or fractional) to be valid?

|x| must be less than 1.

Long Questions & Answers

Explain the principle of mathematical induction, including its three steps, and describe how the extended principle differs from it.

What problem does mathematical induction solve?

Since a formula involving natural numbers cannot be checked for infinitely many values one by one, mathematical induction proves it is true for every natural number using just two finite steps instead.

What is the base case, and why is it necessary?

The base case verifies that the statement S(n) is true for the starting value (usually n=1); without it, the induction hypothesis alone would have nothing to build from.

What is the induction of hypothesis step?

Assuming S(k) is true for some arbitrary positive integer k, this step proves that S(k+1) must also be true, showing the statement can always be extended one step further.

How does the extended principle of mathematical induction differ from the ordinary principle?

The extended principle allows the base case to start at any integer i (not necessarily 1, and possibly 0 or negative), and then proves S(k+1) follows from S(k) for every integer k greater than or equal to i, concluding S(n) is true for all integers n greater than or equal to i.

Describe the binomial theorem for a positive integral index, including the general term, and explain how the middle term is found.

What does the binomial theorem state?

For any positive integer n, (a+b)^n expands as the sum from r=0 to n of C(n,r) a^(n-r) b^r, where C(n,r) are the binomial coefficients.

What is the general term, and why is it useful?

The general term is $T(r+1) = C(n,r)\, a^{(n-r)}\, b^r$; since substituting $r=0,1,2,\ldots,n$ generates every term of the expansion, it lets any specific term be found directly without writing out the whole expansion.

How many terms does a binomial expansion have, and what patterns do the exponents follow?

A binomial expansion of $(a+b)^n$ has exactly $n+1$ terms; the exponent of a decreases from n to 0 while the exponent of b increases from 0 to n, and the two exponents in any term always add up to n.

How is the middle term found, and why does the method differ based on whether n is even or odd?

If n is even, $n+1$ (the total number of terms) is odd, so there is exactly one middle term, the $(n/2+1)$th term; if n is odd, $n+1$ is even, so there are two middle terms, the $((n+1)/2)$th and $((n+3)/2)$th terms -- the difference arises because an odd total number of terms has one true center, while an even total has two.

Multiple Choice Questions (MCQs)

The first step of mathematical induction is called the:

A. Conclusion
B. Base case
C. Hypothesis test
D. Counterexample
Answer: B. Base case
Explanation: The base case verifies the statement is true for the starting value, usually $n=1$.

A counterexample is:

A. A proof of a statement
B. A single case that disproves a statement
C. An extra base case
D. A binomial coefficient
Answer: B. A single case that disproves a statement

Explanation: A counterexample is one failing case that is enough to disprove an entire statement.

The binomial theorem expands $(a+b)^n$ as a sum involving:

A. Logarithms

B. Binomial coefficients $C(n,r)$

C. Trigonometric ratios

D. Matrices

Answer: B. Binomial coefficients $C(n,r)$

Explanation: The expansion of $(a+b)^n$ is a sum of terms, each with a binomial coefficient $C(n,r)$.

The general term of a binomial expansion is given by:

A. $C(n,r)\ a^r\ b^{(n-r)}$

B. $C(n,r)\ a^{(n-r)}\ b^r$

C. $n!\ a^n$

D. a^n+b^n

Answer: B. $C(n,r)\ a^{(n-r)}\ b^r$

Explanation: The general term is $T(r+1) = C(n,r)\ a^{(n-r)}\ b^r$.

A binomial expansion of $(a+b)^n$ has how many terms?

A. n

B. $n+1$

C. $n-1$

D. $2n$

Answer: B. $n+1$

Explanation: The expansion always has $n+1$ terms, one more than the index.

If n is even, the number of middle terms in a binomial expansion is:

A. 0

B. 1

C. 2

D. n

Answer: B. 1

Explanation: An even n gives an odd total number of terms $(n+1)$, so there is exactly one true middle term.

Pascal's Rule states that $C(n,k)$ equals:

A. C(n-1,k) x C(n-1,k-1)

B. C(n-1,k-1) + C(n-1,k)

C. C(n-1,k) - C(n-1,k-1)

D. C(n,k-1)

Answer: B. C(n-1,k-1) + C(n-1,k)

Explanation: Each entry of Pascal's triangle is the sum of the two entries above it: C(n-1,k-1)+C(n-1,k).

The sum of all binomial coefficients C(n,0)+C(n,1)+...+C(n,n) equals:

A. n!

B. 2^n

C. n^2

D. 2n

Answer: B. 2^n

Explanation: Setting a=b=1 in the binomial expansion gives (1+1)^n = 2^n as the sum of all coefficients.

The binomial series expansion of (1+x)^n for n negative or fractional is valid only when:

A. x > 1

B. |x| < 1

C. x = 0

D. x is a whole number

Answer: B. |x| < 1

Explanation: The infinite binomial series only converges when the absolute value of x is less than 1.

The binomial theorem can be used to find the remainder of a large power when divided by a number by writing the base as:

A. A logarithm

B. A sum or difference matching a multiple of the divisor

C. A matrix

D. A trigonometric identity

Answer: B. A sum or difference matching a multiple of the divisor

Explanation: Rewriting the base as (multiple of the divisor +/- 1) makes every expanded term but the last a clean multiple of the divisor.

Quick Revision Summary

- Mathematical induction proves a statement true for all n by two steps: base case ($S(1)$ true) and induction of hypothesis ($S(k)$ true implies $S(k+1)$ true)
- A counterexample -- a single failing case -- is enough to disprove any mathematical statement
- Extended mathematical induction proves statements true for all integers $n \geq i$, starting the base case at $S(i)$ instead of $S(1)$
- Binomial theorem: $(a+b)^n$ = sum from $r=0$ to n of $C(n,r)\, a^{(n-r)}\, b^r$, for positive integer n
- General term: $T(r+1) = C(n,r)\, a^{(n-r)}\, b^r$
- A binomial expansion of $(a+b)^n$ has $n+1$ terms; exponents of a and b always sum to n
- Middle term: one middle term ($(n/2+1)$th) if n is even; two middle terms if n is odd
- Sum of binomial coefficients: $C(n,0)+C(n,1)+\ldots+C(n,n) = 2^n$
- Pascal's Rule: $C(n,k) = C(n-1,k-1) + C(n-1,k)$, generating Pascal's triangle
- For n negative or fractional, $(1+x)^n$ becomes an infinite binomial series, valid only when $|x| < 1$
- Since terms shrink when $|x| < 1$, the first few terms of a binomial series give good approximations
- Rewriting a large power as $(\text{base} \pm k)^n$ and expanding is a powerful tool for remainders, last digits, divisibility, and comparisons

Exam Tips

- Always write out the base case explicitly and show it equals both sides -- don't skip this step
- In the induction step, add or manipulate $S(k)$ algebraically to reach exactly the form of $S(k+1)$ -- work toward the target expression
- Remember $C(n,r) = C(n,n-r)$ -- use it to quickly find coefficients equidistant from the ends of an expansion
- To find a specific term (like the one involving x^5), set the exponent of x in $T(r+1)$ equal to the target power and solve for r
- For approximation problems, only the first 3-4 terms of a binomial series are usually needed since later terms shrink rapidly when $|x| < 1$

- For remainder/divisibility problems, rewrite the base as (multiple of divisor +/- 1) before expanding, so every term but the last is a clean multiple of the divisor

Unit 9: Division of Polynomials

This unit explores how one polynomial divides into another to produce a quotient and a remainder, and develops two powerful theorems that make working with polynomial division far more efficient. The remainder theorem lets the remainder of a division by (x-a) be found by a single evaluation, f(a), without performing the division at all; the factor theorem then uses this idea to test whether a given linear expression is a factor of a polynomial simply by checking whether that evaluation equals zero.

The unit also introduces synthetic division, a compact shortcut method for dividing by (x-a) using only coefficients, and closes with real-world applications in polynomial regression (statistical modeling) and digital signal processing, where the poles and zeros of a system's transfer function -- found using exactly these same theorems -- determine whether a real engineering system is stable.

Learning Objectives

- Define a polynomial and its degree
- Divide one polynomial by another using long division to find the quotient and remainder
- State and apply the remainder theorem to find a remainder without performing division
- State and apply the factor theorem to test whether a linear polynomial is a factor of a given polynomial
- Use synthetic division as a shortcut method for dividing a polynomial by x-a
- Apply the remainder and factor theorems to solve for unknown coefficients in a polynomial
- Apply polynomial division concepts to real-world applications in polynomial regression and digital signal processing

Key Concepts

9.1 Polynomials and Polynomial Division

A polynomial in x is an expression of the form $a_n x^n + a_{(n-1)} x^{(n-1)} + \dots + a_1 x + a_0$, where n is a non-negative integer and the coefficients are real numbers; the highest power of x with a non-zero coefficient is called the degree of the polynomial.

Dividing one polynomial by another (of lower or equal degree) using long division produces a unique quotient $q(x)$ and remainder R, so that the dividend $f(x) =$ (divisor)(quotient) + remainder -- exactly analogous to dividing whole numbers, and division continues until the remainder's degree is less than the divisor's degree.

9.2 The Remainder Theorem

If a polynomial $f(x)$ of degree $n>=1$ is divided by $x-a$, the remainder is simply $f(a)$ -- the value of the polynomial evaluated at $x=a$. This is proved by substituting $x=a$ into the division identity $f(x) = (x-a)q(x) + R$, which makes the $(x-a)q(x)$ term vanish, leaving $f(a) = R$.

The remainder theorem is a major shortcut: instead of performing the entire long division just to find the remainder, only a single substitution and evaluation is needed -- this also makes it easy to solve for an unknown coefficient in a polynomial when the remainder of a particular division is given.

9.3 The Factor Theorem

The polynomial $x-a$ is a factor of $f(x)$ if and only if $f(a)=0$ -- in other words, $x=a$ is a root of $f(x)=0$ exactly when $x-a$ divides $f(x)$ with no remainder. This follows directly from the remainder theorem: since the remainder of dividing by $x-a$ is $f(a)$, a remainder of zero means $x-a$ divides evenly.

The factor theorem is used both to test whether a given linear expression is a factor of a polynomial (by checking whether $f(a)=0$) and, when multiple factors are known, to set up a system of equations to solve for unknown coefficients in the polynomial.

9.4 Synthetic Division

Synthetic division is a shortcut method for dividing a polynomial by a linear divisor of the form x-a, using only the coefficients of the dividend (with 0 inserted for any missing power) arranged in a compact numerical layout, avoiding the need to write out the full long-division process with variables.

The method uses two repeating patterns: working diagonally, each number is multiplied by a and placed under the next coefficient; working vertically, each column is added straight down -- the final row gives the coefficients of the quotient (one degree lower than the dividend) followed by the remainder.

9.5 Real-Life Applications: Polynomial Regression

In polynomial regression, a statistical model fits an nth-degree polynomial P(x) to a set of data points; the remainder theorem is used to evaluate the model's predicted value at any given point x (by direct substitution), and the difference between an observed value and the predicted value gives the prediction error, often expressed as a percentage of the observed value.

The factor theorem also helps simplify or reduce a regression model's degree when a known root (a point where the fitted curve crosses zero) is identified, since that root corresponds to a linear factor that can be divided out.

9.6 Real-Life Applications: Digital Signal Processing

In digital signal processing, a system's transfer function H(z) is a ratio of two polynomials, B(z)/A(z); the roots of the numerator B(z) are called the zeros of the system, and the roots of the denominator A(z) are called the poles, both found using the factor theorem by solving B(z)=0 or A(z)=0.

A system is stable only if every one of its poles lies strictly inside the unit circle in the complex plane, meaning the absolute value (modulus) of each pole is less than 1; finding the poles via the factor theorem and checking their magnitudes is therefore how the stability of a signal-processing system is determined.

Important Definitions

What is a polynomial in x?

An expression a_n x^n + a_(n-1)x^(n-1) + ... + a_1 x + a_0, where n is a non-negative integer and the coefficients are real numbers.

What is the degree of a polynomial?

The highest power of x with a non-zero coefficient.

What does the remainder theorem state?

If a polynomial f(x) is divided by x-a, the remainder equals f(a).

What does the factor theorem state?

x-a is a factor of f(x) if and only if f(a)=0.

What is synthetic division?

A shortcut method for dividing a polynomial by x-a using only its coefficients.

What is a root (or zero) of a polynomial equation f(x)=0?

A value of x that makes f(x) equal to zero.

In the division identity f(x)=(x-a)q(x)+R, what are q(x) and R called?

q(x) is the quotient and R is the remainder.

What is polynomial regression?

A statistical modeling technique that fits an nth-degree polynomial to a data set.

In a signal processing transfer function H(z)=B(z)/A(z), what are the zeros and poles?

The zeros are the roots of the numerator B(z); the poles are the roots of the denominator A(z).

What condition on the poles determines whether a signal processing system is stable?

Every pole must lie inside the unit circle, meaning its absolute value (modulus) is less than 1.

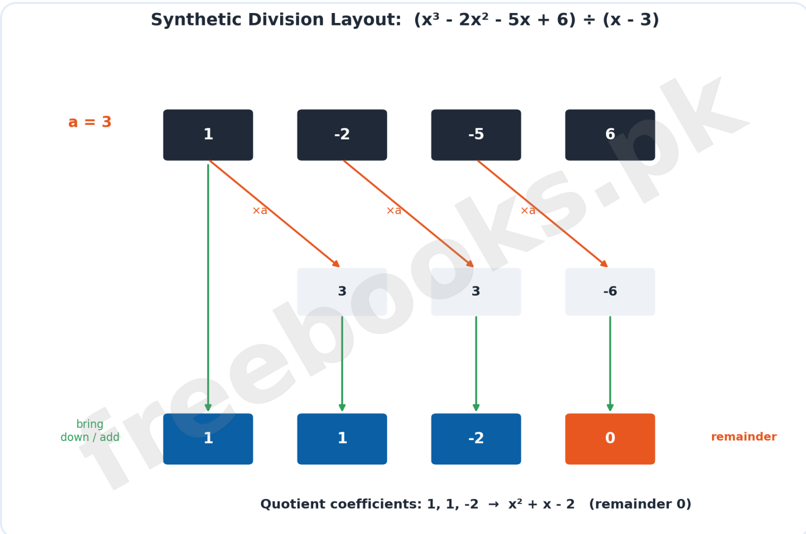
Key Facts and Relations

Topic	Key Fact / Relation
Division identity	f(x) = (x-a) . q(x) + R
Remainder theorem	remainder of f(x) divided by (x-a) = f(a)

Topic	Key Fact / Relation
Factor theorem	(x-a) is a factor of f(x) iff f(a) = 0
Synthetic division layout	diagonal = multiply by a, vertical = add
Percentage error	|observed - predicted| / observed x 100
Transfer function	H(z) = B(z) / A(z)
Stability condition	every pole p0 satisfies |p0| < 1

Diagrams

Synthetic Division Layout: A step-by-step schematic of synthetic division for (x^3-2x^2-5x+6) divided by (x-3), showing the diagonal 'multiply by a' pattern and the vertical 'add' pattern, ending in the quotient coefficients and remainder



Synthetic Division Layout

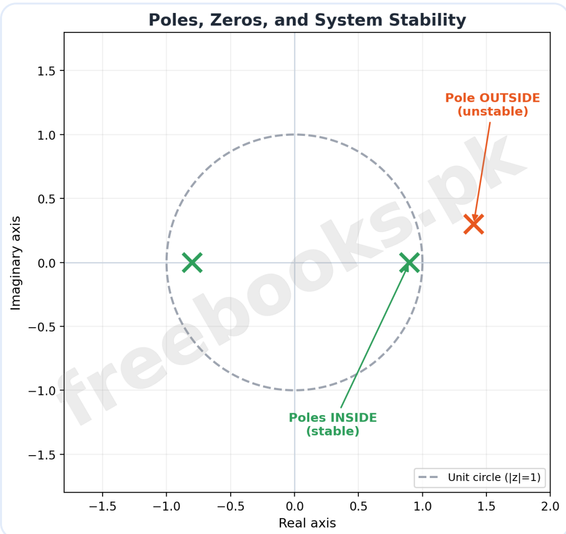
Remainder Theorem vs Factor Theorem: A comparison table showing the statement, typical use, and key check for the remainder theorem alongside the factor theorem

Remainder Theorem vs Factor Theorem

	Remainder Theorem	Factor Theorem
Statement	Remainder of f(x)÷(x-a) = f(a)	(x-a) is a factor of f(x) iff f(a) = 0
Used to	Find a remainder WITHOUT performing division	Test if a linear expression IS a factor
Key check	Evaluate f(a) directly	Check whether f(a) = 0

Remainder Theorem vs Factor Theorem

Poles, Zeros, and System Stability: A complex-plane plot showing the unit circle, with example poles inside the circle (stable) marked in green and a pole outside the circle (unstable) marked in orange



Poles, Zeros, and System Stability

Solved Examples

Example 1: Long Division of Polynomials

Problem: Divide $2x^3 - 3x^2 - 8x + 3$ by $x-3$, finding the quotient and remainder.

1. Divide the leading term: $2x^3$ divided by x gives $2x^2$. Multiply $2x^2(x-3)=2x^3-6x^2$, and subtract from the dividend to get $3x^2-8x$ (bringing down the $-8x$ term).
2. Divide the new leading term: $3x^2$ divided by x gives $3x$. Multiply $3x(x-3)=3x^2-9x$, and subtract to get $x+3$ (bringing down the $+3$ term).
3. Divide again: x divided by x gives 1. Multiply $1(x-3)=x-3$, and subtract to get a final remainder of 6.
4. Final answer: quotient = $2x^2+3x+1$, remainder = 6.

Example 2: Using the Remainder Theorem

Problem: Find the remainder when $f(x)=2x^3+3x^2-5x+7$ is divided by $x+2$, without performing division.

1. Since the divisor is $x+2 = x-(-2)$, identify $a=-2$.
2. By the remainder theorem, the remainder equals $f(-2)$.
3. Compute $f(-2) = 2(-2)^3+3(-2)^2-5(-2)+7 = -16+12+10+7$.
4. Add the terms: $-16+12=-4$, $-4+10=6$, $6+7=13$.
5. Final answer: the remainder is 13.

Example 3: Finding an Unknown Coefficient Using the Remainder Theorem

Problem: Find the value of k so that the polynomial $2x^3+kx^2-5x+8$ leaves a remainder of 6 when divided by $x-1$.

1. Since the divisor is $x-1$, use the remainder theorem with $a=1$: remainder = $f(1)$.
2. Compute $f(1) = 2(1)^3+k(1)^2-5(1)+8 = 2+k-5+8 = k+5$.
3. Set this equal to the given remainder: $k+5=6$.
4. Solve: $k=1$.
5. Final answer: $k=1$.

Example 4: Applying the Factor Theorem

Problem: Show that $x+3$ is a factor of $f(x)=x^3+2x^2-5x-6$, without factorizing.

1. Since the factor is $x+3 = x-(-3)$, identify $a=-3$.
2. Compute $f(-3) = (-3)^3+2(-3)^2-5(-3)-6 = -27+18+15-6$.
3. Add the terms: $-27+18=-9$, $-9+15=6$, $6-6=0$.
4. Since $f(-3)=0$, by the factor theorem $x+3$ is a factor of $f(x)$.
5. Final answer: $x+3$ is confirmed to be a factor of $f(x)$.

Example 5: Finding Unknown Coefficients from Two Known Factors

Problem: If $x-1$ and $x+2$ are factors of x^3+px^2+qx-6, find the values of p and q.

1. Since $x-1$ is a factor, $f(1)=0$: $1+p+q-6=0$, giving $p+q=5$. (i)
2. Since $x+2$ is a factor, $f(-2)=0$: $-8+4p-2q-6=0$, giving $2p-q=7$. (ii)
3. From (i), $q=5-p$. Substitute into (ii): $2p-(5-p)=7$, so $3p-5=7$, giving $p=4$.
4. Substitute $p=4$ back into $q=5-p$ to get $q=1$.

5. Final answer: p=4, q=1.

Example 6: Synthetic Division

Problem: Use synthetic division to divide f(x)=x^3-2x^2-5x+6 by x-3, and state the quotient and remainder.

1. Write the coefficients of f(x) in order: 1, -2, -5, 6, with a=3 (since the divisor is x-3).
2. Bring down the first coefficient (1). Multiply by a: 1x3=3, add to the next coefficient: -2+3=1.
3. Multiply the new value by a: 1x3=3, add to the next coefficient: -5+3=-2. Multiply again: -2x3=-6, add to the last coefficient: 6+(-6)=0.
4. The bottom row 1, 1, -2, 0 gives the quotient x^2+x-2 with remainder 0.
5. Final answer: quotient = x^2+x-2, remainder = 0 (so x-3 is a factor).

Example 7: Application: Polynomial Regression Prediction Error

Problem: A polynomial regression model P(x)=2x^3-x^2+3x-4 is fitted to monthly sales data. If the observed sales in month 4 are 130 units, find the percentage error.

1. Use the remainder theorem to evaluate the model at x=4: P(4)=2(4)^3-(4)^2+3(4)-4 = 128-16+12-4.
2. Simplify: 128-16=112, 112+12=124, 124-4=120. So the predicted value is 120.
3. Compute the error: Error = Observed - Predicted = 130-120 = 10.
4. Percentage error = |10/130| x 100, approximately 7.7%.
5. Final answer: the percentage error is approximately 7.7%.

Example 8: Application: Poles and Stability of a Signal Processing System

> **Problem:** A signal processing system has a transfer function with denominator A(z)=z^2-0.1z-0.72. Use the factor theorem to find the poles and determine if the system is stable.
>
> 1. Poles occur where A(z)=0: z^2-0.1z-0.72=0.
> 2. Factor: looking for two numbers that multiply to -0.72 and combine to match the -0.1z coefficient: 0.9 and -0.8 work, since 0.9x(-0.8)=-0.72 and -0.9+0.8=-0.1.
> 3. So z^2-0.1z-0.72 = (z-0.9)(z+0.8) = 0, giving poles z=0.9 and z=-0.8.
> 4. Check stability: |0.9|=0.9<1 and |-0.8|=0.8<1 -- both poles lie inside the unit circle.
> 5. Final answer: the poles are z=0.9 and z=-0.8; since both lie inside the unit circle, the system is stable.

Short Questions & Answers

What is the degree of a polynomial?

The highest power of x with a non-zero coefficient.

What does the remainder theorem state?

The remainder when f(x) is divided by x-a equals f(a).

What does the factor theorem state?

x-a is a factor of f(x) if and only if f(a)=0.

What is synthetic division used for?

A shortcut method for dividing a polynomial by x-a using only coefficients.

In H(z)=B(z)/A(z), what are the zeros of the system?

The roots of the numerator B(z).

In H(z)=B(z)/A(z), what are the poles of the system?

The roots of the denominator A(z).

What condition makes a signal processing system stable?

Every pole must have an absolute value less than 1 (lie inside the unit circle).

Long Questions & Answers

Explain the remainder theorem and the factor theorem, including how the factor theorem follows directly from the remainder theorem.

What does the remainder theorem state, and how is it proved?

If a polynomial f(x) is divided by x-a, the remainder equals f(a); this is proved by substituting x=a into the division identity f(x)=(x-a)q(x)+R, which makes the (x-a)q(x) term vanish, leaving f(a)=R.

Why is the remainder theorem useful in practice?

It lets the remainder of a division be found with a single substitution and evaluation, avoiding the need to perform the entire long division just to learn the remainder.

What does the factor theorem state?

The polynomial x-a is a factor of f(x) if and only if f(a)=0, meaning x=a is a root of the equation f(x)=0.

How does the factor theorem follow from the remainder theorem?

Since the remainder theorem says the remainder of dividing f(x) by x-a is f(a), a remainder of exactly zero means x-a divides f(x) with nothing left over -- which is precisely the definition of x-a being a factor.

Describe synthetic division as a method for dividing a polynomial by x-a, and explain its real-world use in evaluating signal processing transfer functions.

What is synthetic division, and what does it require as input?

Synthetic division is a shortcut method for dividing a polynomial by a linear divisor x-a, using only the polynomial's coefficients (in order, with 0 inserted for any missing power) rather than writing out the full division with variables.

What are the two repeating patterns used in synthetic division?

Working diagonally, each number in the bottom row is multiplied by a and placed under the next coefficient; working vertically, each column is then added straight down to produce the next entry in the bottom row.

What does the final row of a synthetic division represent?

The final row gives the coefficients of the quotient (one degree lower than the original polynomial), followed by the remainder as the very last entry.

How is this method applied in digital signal processing?

A system's transfer function $H(z)=B(z)/A(z)$ is a ratio of polynomials; finding the roots of $A(z)$ (the poles) using the factor theorem, and checking whether each pole's absolute value is less than 1, determines whether the physical system the transfer function models is stable.

Multiple Choice Questions (MCQs)

The degree of a polynomial is:

A. The number of terms
B. The highest power of x with a non-zero coefficient
C. The value of the leading coefficient
D. Always an even number
Answer: B. The highest power of x with a non-zero coefficient
Explanation: The degree is the highest power of x whose coefficient is not zero.

By the remainder theorem, the remainder when f(x) is divided by (x-a) equals:

A. f(0)
B. f(a)
C. a
D. f(x)/a
Answer: B. f(a)
Explanation: The remainder theorem states the remainder equals f(a), the value of the polynomial at x=a.

By the factor theorem, (x-a) is a factor of f(x) if and only if:

A. f(a)=1

B. f(0)=a

C. f(a)=0

D. f(x)=0 for all x

Answer: C. f(a)=0

Explanation: x-a divides f(x) exactly (with zero remainder) precisely when f(a)=0.

Synthetic division is a shortcut method for dividing a polynomial by:

A. Any polynomial

B. A linear divisor x-a

C. A quadratic divisor

D. A constant

Answer: B. A linear divisor x-a

Explanation: Synthetic division only works for a linear divisor of the form x-a.

In synthetic division, the diagonal pattern means:

A. Add the two numbers

B. Multiply by a

C. Subtract the numbers

D. Divide by a

Answer: B. Multiply by a

Explanation: The diagonal step multiplies the previous bottom-row entry by a.

In a transfer function H(z)=B(z)/A(z), the roots of the denominator A(z) are called:

A. Zeros

B. Poles

C. Coefficients

D. Remainders

Answer: B. Poles

Explanation: The roots of the denominator are called the poles of the system.

A signal processing system is stable if all its poles:

A. Lie outside the unit circle

B. Lie exactly on the unit circle

C. Lie inside the unit circle

D. Equal zero

Answer: C. Lie inside the unit circle

Explanation: Stability requires every pole's absolute value to be less than 1, i.e. inside the unit circle.

If f(x) is divided by (x-a) and the remainder is 0, then:

A. x-a is not a factor

B. a is not a root

C. x-a is a factor of f(x)

D. f(x) is undefined

Answer: C. x-a is a factor of f(x)

Explanation: A zero remainder means x-a divides f(x) exactly, so x-a is a factor.

Percentage error in a regression model is calculated as:

A. Observed x Predicted

B. |Observed - Predicted| / Observed x 100

C. Predicted / Observed

D. Observed / Predicted x 100

Answer: B. |Observed - Predicted| / Observed x 100

Explanation: Percentage error is the absolute difference between observed and predicted, divided by the observed value, times 100.

To find the remainder of a polynomial divided by x+3 using the remainder theorem, substitute:

A. x=3

B. x=-3

C. x=0

D. x=1/3

Answer: B. x=-3

Explanation: Since x+3 = x-(-3), the value to substitute is a=-3.

Quick Revision Summary

- A polynomial's degree is the highest power of x with a non-zero coefficient
- Dividing f(x) by a divisor gives f(x) = (divisor)(quotient) + remainder
- Remainder theorem: the remainder when f(x) is divided by (x-a) equals f(a)
- Factor theorem: (x-a) is a factor of f(x) if and only if f(a)=0

- Synthetic division divides a polynomial by (x-a) using only coefficients: diagonal = multiply by a, vertical = add
- The final row of synthetic division gives the quotient's coefficients, then the remainder
- To find an unknown coefficient, set f(a) equal to the given remainder and solve
- To find unknown coefficients from two known factors, set up f(a1)=0 and f(a2)=0 as simultaneous equations
- Polynomial regression fits an nth-degree polynomial to data; the remainder theorem evaluates predictions at any point
- Percentage error = |observed - predicted| / observed x 100
- In signal processing, zeros are roots of the numerator B(z); poles are roots of the denominator A(z)
- A system is stable only if every pole's absolute value (modulus) is less than 1 (inside the unit circle)

Exam Tips

- Always rewrite the divisor as x-a first (e.g. x+3 means a=-3) before applying the remainder or factor theorem
- Use the remainder theorem to check answers from long division -- f(a) should match the remainder you found
- For unknown-coefficient problems, translate remainder theorem facts into simple linear equations and solve
- In synthetic division, always insert 0 for any missing power of x in the coefficient row
- Remember: a zero remainder from the factor theorem confirms a factor; it does NOT mean the polynomial itself is zero
- For stability problems, compute the absolute value of every pole and compare it to 1 -- all poles must qualify for the system to be stable

Unit 10: Trigonometric Identities

This unit builds the whole machinery of trigonometric identities from a single geometric fact: the Fundamental Law of Trigonometry, cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta, proved by equating two equal chord lengths on the unit circle. From this one law, a chain of deductions produces the co-function identities, the sum and difference formulas for sine, cosine, and tangent, and the rules for evaluating trigonometric ratios of allied angles (angles that differ from a multiple of a right angle).

Building further on the sum-angle formulas, the unit develops double angle, half angle, and triple angle identities (by setting beta=alpha or solving for the half-angle), and finally the product-to-sum and sum-to-product identities, which convert between a product of sines/cosines and a sum or difference, and back again -- tools that are essential for simplifying complex trigonometric expressions and proving identities without a calculator.

Learning Objectives

- State and prove the Fundamental Law of Trigonometry using the unit circle
- Derive the sum and difference formulas for sine, cosine, and tangent from the fundamental law
- Evaluate trigonometric ratios of allied angles using the co-ratio and sign rules
- Derive and apply double angle, half angle, and triple angle identities
- Convert products of sines and cosines into sums or differences, and vice versa
- Simplify trigonometric expressions and prove trigonometric identities without using tables or a calculator
- Apply sum-angle and double-angle identities to solve for unknown trigonometric ratios given partial information

Key Concepts

10.1 The Fundamental Law of Trigonometry

Using the distance formula between two points on a unit circle, the Fundamental Law of Trigonometry is derived: cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta. Two angles alpha and beta placed in standard position cut the unit circle at points A(cos alpha, sin alpha) and B(cos beta, sin beta); rotating triangle AOB to an equivalent triangle COD (where C is at angle alpha-beta and D is at (1,0)) gives |AB|=|CD|, and equating the squared distances yields the law.

Although proved for alpha > beta > 0, the law holds for all real values of alpha and beta. This single identity is the seed from which every other sum, difference, double angle, half angle, and product-sum identity in this unit is systematically derived.

10.2 Deductions from the Fundamental Law

Substituting specific values for alpha or beta into the fundamental law produces the co-function identities -- cos(pi/2 - beta) = sin beta and sin(pi/2 + alpha) = cos alpha -- and, by replacing beta with -beta, the cosine sum formula cos(alpha+beta) = cos alpha cos beta - sin alpha sin beta.

Continuing the chain of substitutions produces the sine sum and difference formulas, sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta and sin(alpha-beta) = sin alpha cos beta - cos alpha sin beta, and dividing the sine formulas by the corresponding cosine formulas (then dividing every term by cos alpha cos beta) produces the tangent sum and difference formulas.

10.3 Trigonometric Ratios of Allied Angles

Two angles alpha and beta are called allied if alpha +/- beta = n(90 degrees) for some integer n; common allied angles of theta include 90 degrees +/- theta, 180 degrees +/- theta, 270 degrees +/- theta, and 360 degrees +/- theta. Using the fundamental law and its deductions, a full set of identities can be derived for the sine, cosine, and tangent of every allied angle.

These identities follow a memorable pattern: if theta is added to or subtracted from an odd multiple of a right angle, the ratio changes to its co-ratio (sin to cos, tan to cot, and vice versa); if theta is added to or subtracted from an even multiple, the ratio stays the same. In both cases, the sign of the result is decided by which quadrant the resulting angle's terminal side falls in.

10.4 Double Angle Identities

Setting beta = alpha in the sum formulas for sine, cosine, and tangent produces the double angle identities: sin 2 alpha = 2 sin alpha cos alpha, cos 2 alpha = cos^2 alpha - sin^2 alpha (which can also be written as 2 cos^2 alpha - 1 or 1 - 2 sin^2 alpha using the Pythagorean identity), and tan 2 alpha = 2 tan alpha / (1 - tan^2 alpha).

The three equivalent forms of cos 2 alpha are especially useful, since each expresses the double angle purely in terms of a single ratio (cos alpha alone, or sin alpha alone) -- this flexibility is what makes the half angle identities in the next section possible.

10.5 Half Angle and Triple Angle Identities

Solving the double angle identity cos alpha = 2 cos^2(alpha/2) - 1 for cos(alpha/2) gives the half angle identity cos(alpha/2) = +/- sqrt((1+cos alpha)/2); the same approach applied to cos alpha = 1 - 2 sin^2(alpha/2) gives sin(alpha/2) = +/- sqrt((1-cos alpha)/2), and dividing the two gives tan(alpha/2) = +/- sqrt((1-cos alpha)/(1+cos alpha)).

Writing 3 alpha as 2 alpha + alpha and expanding with the sum and double angle formulas produces the triple angle identities sin 3 alpha = 3 sin alpha - 4 sin^3 alpha and cos 3 alpha = 4 cos^3 alpha - 3 cos alpha, both useful for reducing higher powers of sine and cosine to expressions in multiples of the original angle.

10.6 Product-to-Sum and Sum-to-Product Identities

Adding and subtracting the sine sum/difference formulas (and separately the cosine sum/difference formulas) converts a product of a sine and a cosine, or two cosines, or two sines, into a sum or difference: for example, 2 sin alpha cos beta = sin(alpha+beta) + sin(alpha-beta).

Substituting P = alpha+beta and Q = alpha-beta into these four product-to-sum identities and solving for alpha and beta in terms of P and Q produces the reverse set -- the sum-to-product identities, such as sin P + sin Q = 2 sin((P+Q)/2) cos((P-Q)/2) -- which are essential for simplifying sums of sines/cosines and proving identities involving multiple angles.

Important Definitions

What is the Fundamental Law of Trigonometry?

cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta, proved by equating two equal chord lengths on the unit circle.

What are allied angles?

Two angles alpha and beta such that alpha +/- beta = n(90 degrees) for some integer n.

What is the sum formula for sin(alpha+beta)?

sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta.

What is the double angle identity for sin 2 alpha?

sin 2 alpha = 2 sin alpha cos alpha.

What are the three equivalent forms of cos 2 alpha?

cos 2 alpha = cos^2 alpha - sin^2 alpha = 2 cos^2 alpha - 1 = 1 - 2 sin^2 alpha.

What is the half angle identity for cos(alpha/2)?

cos(alpha/2) = +/- sqrt((1+cos alpha)/2).

What is the triple angle identity for sin 3 alpha?

sin 3 alpha = 3 sin alpha - 4 sin^3 alpha.

What is the triple angle identity for cos 3 alpha?

cos 3 alpha = 4 cos^3 alpha - 3 cos alpha.

State one product-to-sum identity.

2 sin alpha cos beta = sin(alpha+beta) + sin(alpha-beta).

What memory device helps recall how allied angles change trigonometric ratios?

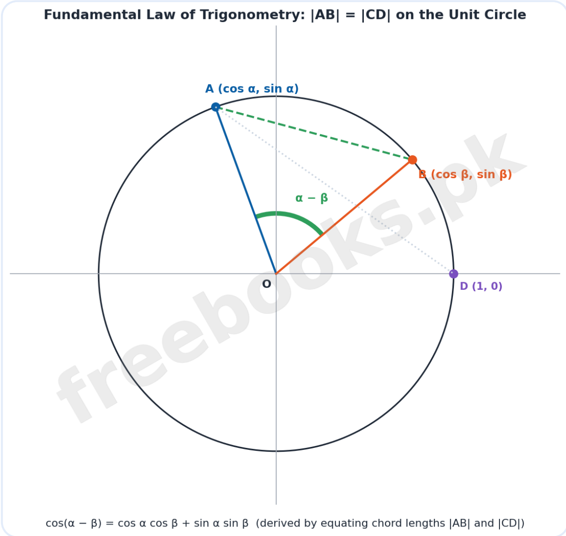
If theta is added to or subtracted from an odd multiple of a right angle, the ratio changes to its co-ratio; if from an even multiple, the ratio stays the same; the sign is decided by the quadrant.

Key Facts and Relations

Topic	Key Fact / Relation
Fundamental Law	cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta
Cosine sum/difference	cos(alpha+beta) = cos alpha cos beta - sin alpha sin beta
Sine sum/difference	sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta
Tangent sum/difference	tan(alpha+beta) = (tan alpha + tan beta) / (1 - tan alpha tan beta)
Double angle (sine)	sin 2 alpha = 2 sin alpha cos alpha
Double angle (cosine, 3 forms)	cos 2 alpha = cos^2 alpha - sin^2 alpha = 2 cos^2 alpha - 1 = 1 - 2 sin^2 alpha
Double angle (tangent)	tan 2 alpha = 2 tan alpha / (1 - tan^2 alpha)
Half angle (cosine)	cos(alpha/2) = +/- sqrt((1+cos alpha)/2)
Triple angle	sin 3 alpha = 3 sin alpha - 4 sin^3 alpha; cos 3 alpha = 4 cos^3 alpha - 3 cos alpha
Sum-to-product (sine)	sin P + sin Q = 2 sin((P+Q)/2) cos((P-Q)/2)

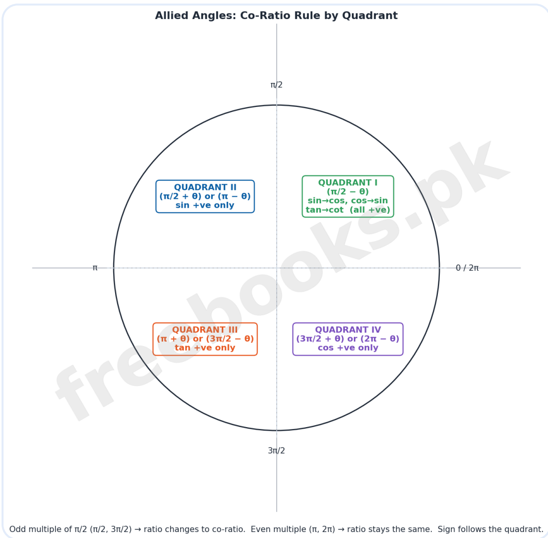
Diagrams

Fundamental Law of Trigonometry: A unit-circle diagram showing points A and B at angles alpha and beta, the chord AB, and the angle alpha-beta at the origin, illustrating the geometric basis of the fundamental law



Fundamental Law of Trigonometry

Allied Angles: Co-Ratio Wheel: A four-quadrant wheel showing which allied-angle forms (pi/2-theta, pi+theta, 3pi/2-theta, 2pi-theta, etc.) fall in each quadrant, and the co-ratio/sign rule that applies



Allied Angles: Co-Ratio Wheel

Product-to-Sum and Sum-to-Product Identities: A side-by-side comparison of the four product-to-sum identities and their corresponding four sum-to-product identities

Product-to-Sum and Sum-to-Product Identities

Product → Sum/Difference		Sum/Difference → Product	
$2 \sin\alpha \cos\beta$	$= \sin(\alpha+\beta) + \sin(\alpha-\beta)$	$\sin P + \sin Q$	$= 2 \sin(P+Q)/2 \cdot \cos(P-Q)/2$
$2 \cos\alpha \sin\beta$	$= \sin(\alpha+\beta) - \sin(\alpha-\beta)$	$\sin P - \sin Q$	$= 2 \cos(P+Q)/2 \cdot \sin(P-Q)/2$
$2 \cos\alpha \cos\beta$	$= \cos(\alpha+\beta) + \cos(\alpha-\beta)$	$\cos P + \cos Q$	$= 2 \cos(P+Q)/2 \cdot \cos(P-Q)/2$
$-2 \sin\alpha \sin\beta$	$= \cos(\alpha+\beta) - \cos(\alpha-\beta)$	$\cos P - \cos Q$	$= -2 \sin(P+Q)/2 \cdot \sin(P-Q)/2$

Product-to-Sum and Sum-to-Product Identities

Solved Examples

Example 1: Using the Fundamental Law to Find sin(5pi/12)

Problem: Find the value of sin(5pi/12) without using tables.

1. Rewrite 5pi/12 as a sum of standard angles: 5pi/12 = 75 degrees = 45 degrees + 30 degrees = pi/4 + pi/6.
2. Apply the sine sum formula: sin(pi/4 + pi/6) = sin(pi/4)cos(pi/6) + cos(pi/4)sin(pi/6).
3. Substitute known values: (1/sqrt(2))(sqrt(3)/2) + (1/sqrt(2))(1/2).
4. Combine over a common denominator: (sqrt(3)+1) / (2 sqrt(2)).
5. Final answer: sin(5pi/12) = (sqrt(3)+1) / (2 sqrt(2)).

Example 2: Evaluating Allied Angles Without Tables

Problem: Without using tables, find the values of sin 225 degrees and tan 600 degrees.

1. For sin 225 degrees: rewrite as sin(180 degrees + 45 degrees). Since (180 degrees + theta) is in Quadrant III where sine is negative, and even multiples of 90 degrees keep the same ratio, sin(180+45) = -sin 45 degrees.

2. So sin 225 degrees = -1/sqrt(2).
3. For tan 600 degrees: rewrite as tan(540 degrees + 60 degrees) = tan(6x90 degrees + 60 degrees). Since 540 degrees is a multiple of 180 degrees (period of tangent), this reduces to tan 60 degrees.
4. So tan 600 degrees = sqrt(3).
5. Final answer: sin 225 degrees = -1/sqrt(2), tan 600 degrees = sqrt(3).

Example 3: Simplifying a Trigonometric Expression Using Allied Angles

Problem: Simplify: [sin(180 deg - theta) cos(360 deg - theta) tan(90 deg + theta)] / [sin(90 deg - theta) cos(180 deg + theta) tan(270 deg - theta)].

1. Convert each allied-angle term using the co-ratio/sign rules: sin(180-theta)=sin theta, cos(360-theta)=cos theta, tan(90+theta)=-cot theta.
2. Convert the denominator terms similarly: sin(90-theta)=cos theta, cos(180+theta)=-cos theta, tan(270-theta)=cot theta.
3. Substitute into the original expression: [sin theta . cos theta . (-cot theta)] / [cos theta . (-cos theta) . cot theta].
4. Cancel the common cos theta and cot theta factors, leaving (-sin theta) / (-cos theta).
5. Final answer: the expression simplifies to tan theta.

Example 4: Finding All Trigonometric Ratios of 105 Degrees

Problem: Without using tables, find sin 105 degrees, cos 105 degrees, and tan 105 degrees.

1. Rewrite 105 degrees as a sum of standard angles: 105 degrees = 60 degrees + 45 degrees.
2. Apply the sine sum formula: sin 105 = sin 60 cos 45 + cos 60 sin 45 = (sqrt(3)/2)(1/sqrt(2)) + (1/2)(1/sqrt(2)) = (sqrt(3)+1)/(2sqrt(2)).

3. Apply the cosine sum formula: cos 105 = cos 60 cos 45 - sin 60 sin 45 = (1/2)(1/sqrt(2)) - (sqrt(3)/2)(1/sqrt(2)) = (1-sqrt(3))/(2sqrt(2)).
4. Apply the tangent sum formula: tan 105 = (tan 60 + tan 45)/(1 - tan 60 tan 45) = (sqrt(3)+1)/(1-sqrt(3)).
5. Final answer: sin 105 = (sqrt(3)+1)/(2sqrt(2)), cos 105 = (1-sqrt(3))/(2sqrt(2)), tan 105 = (sqrt(3)+1)/(1-sqrt(3)).

Example 5: Finding sin(alpha+beta) and cos(alpha+beta) from Given Ratios and Quadrants

Problem: If cos alpha = -7/25 with alpha's terminal side in Quadrant II, and tan beta = 12/5 with beta's terminal side in Quadrant III, find sin(alpha+beta) and cos(alpha+beta), and state which quadrant (alpha+beta) lies in.

1. Find sin alpha using sin^2 alpha + cos^2 alpha = 1: sin alpha = +/- sqrt(1-(-7/25)^2) = +/- 24/25. Since alpha is in Quadrant II where sine is positive, sin alpha = 24/25.
2. Find cos beta and sin beta from tan beta = 12/5: sec beta = +/- sqrt(1+(12/5)^2) = +/- 13/5. Since beta is in Quadrant III where secant is negative, sec beta = -13/5, so cos beta = -5/13, and sin beta = -12/13 (sine is also negative in Quadrant III).
3. Apply the sine sum formula: sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta = (24/25)(-5/13) + (-7/25)(-12/13) = (-120+84)/325 = -36/325.
4. Apply the cosine sum formula: cos(alpha+beta) = cos alpha cos beta - sin alpha sin beta = (-7/25)(-5/13) - (24/25)(-12/13) = (35+288)/325 = 323/325.
5. Since sin(alpha+beta) is negative and cos(alpha+beta) is positive, (alpha+beta) lies in Quadrant IV.

Example 6: Finding Double Angle Ratios from a Given Sine Value

Problem: If sin alpha = 3/5 with 0 < alpha < pi/2, find sin 2 alpha, cos 2 alpha, and tan 2 alpha.

1. Find cos alpha using sin^2 alpha + cos^2 alpha = 1: cos alpha = sqrt(1-(3/5)^2) = 4/5 (positive since alpha is in Quadrant I).
2. Apply the double angle formula for sine: sin 2 alpha = 2 sin alpha cos alpha = 2(3/5)(4/5) = 24/25.
3. Apply the double angle formula for cosine: cos 2 alpha = cos^2 alpha - sin^2 alpha = (4/5)^2 - (3/5)^2 = 16/25 - 9/25 = 7/25.
4. Find tan 2 alpha as the ratio: tan 2 alpha = sin 2 alpha / cos 2 alpha = (24/25)/(7/25) = 24/7.
5. Final answer: sin 2 alpha = 24/25, cos 2 alpha = 7/25, tan 2 alpha = 24/7.

Example 7: Expressing 3 sin theta + 4 cos theta in the Form r sin(theta + phi)

Problem: Express 3 sin theta + 4 cos theta in the form r sin(theta + phi), where the terminal side of phi is in Quadrant I.

1. Let 3 = r cos phi and 4 = r sin phi (matching coefficients with the expanded form of r sin(theta+phi)).
2. Square and add both equations: 3^2 + 4^2 = r^2(cos^2 phi + sin^2 phi) = r^2, so r^2 = 25 and r = 5.
3. Divide the two equations to find phi: tan phi = 4/3, so phi = arctan(4/3), with phi in Quadrant I since both 3 and 4 are positive.
4. Substitute back: 3 sin theta + 4 cos theta = r cos phi sin theta + r sin phi cos theta = r sin(theta+phi).
5. Final answer: 3 sin theta + 4 cos theta = 5 sin(theta + phi), where r=5 and tan phi = 4/3.

Example 8: Proving a Product Identity Using Product-to-Sum Formulas

Problem: Show that cos 20 degrees cos 40 degrees cos 80 degrees = 1/8.

1. Multiply and divide by 4: cos20 cos40 cos80 = (1/4)(4 cos20 cos40 cos80) = (1/4)[(2cos40cos20)(2cos80)].
2. Apply the product-to-sum identity to 2cos40cos20: 2cos40cos20 = cos60+cos20 = 1/2+cos20.
3. Substitute back: (1/4)[(1/2+cos20)(2cos80)] = (1/4)[cos80 + 2cos80cos20].
4. Apply the product-to-sum identity to 2cos80cos20: 2cos80cos20 = cos100+cos60 = -cos80+1/2 (using cos100=cos(180-80)=-cos80).
5. Substitute and simplify: (1/4)[cos80 - cos80 + 1/2] = (1/4)(1/2) = 1/8. Final answer: cos20 cos40 cos80 = 1/8.

Short Questions & Answers

What is the Fundamental Law of Trigonometry?

cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta.

What are allied angles?

Two angles alpha and beta such that alpha +/- beta = n(90 degrees) for some integer n.

What is the double angle formula for sin 2 alpha?

sin 2 alpha = 2 sin alpha cos alpha.

Give the three forms of cos 2 alpha.

cos^2 alpha - sin^2 alpha, 2 cos^2 alpha - 1, and 1 - 2 sin^2 alpha.

What is the half angle identity for sin(alpha/2)?

sin(alpha/2) = +/- sqrt((1-cos alpha)/2).

What is the triple angle identity for cos 3 alpha?

cos 3 alpha = 4 cos^3 alpha - 3 cos alpha.

How does an odd multiple of a right angle affect an allied-angle ratio?

It changes the ratio to its co-ratio (sin to cos, tan to cot, and so on).

Long Questions & Answers

Derive the Fundamental Law of Trigonometry using the unit circle, and explain how it leads to the sum and difference formulas for sine, cosine, and tangent.

What is the Fundamental Law of Trigonometry and how is it proved geometrically?

It states cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta. It is proved by placing angles alpha and beta in standard position on a unit circle at points A and B, noting that triangle AOB is congruent to a triangle COD (where C is at angle alpha-beta and D is at (1,0)), so |AB|=|CD|; equating the squared distances using the distance formula and simplifying gives the law.

How is the formula for cos(alpha+beta) derived from the fundamental law?

By replacing beta with -beta in the fundamental law and using cos(-beta)=cos beta and sin(-beta)=-sin beta, the law becomes cos(alpha+beta) = cos alpha cos beta - sin alpha sin beta.

How are the sine sum and difference formulas derived?

Using the co-function identity sin(pi/2+alpha)=cos alpha (itself derived from the fundamental law) together with the cosine sum formula, substitution produces sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta, and replacing beta with -beta gives the difference formula sin(alpha-beta) = sin alpha cos beta - cos alpha sin beta.

How is the tangent sum formula tan(alpha+beta) derived?

By writing tan(alpha+beta) as sin(alpha+beta)/cos(alpha+beta), substituting the sine and cosine sum formulas, and dividing every term in the numerator and denominator by cos alpha cos beta, which produces tan(alpha+beta) = (tan alpha+tan beta)/(1-tan alpha tan beta).

Explain the double angle, half angle, and triple angle identities for sine and cosine, and describe how each is derived from the sum-angle formulas.

How are the double angle identities for sin 2 alpha and cos 2 alpha derived?

By setting beta = alpha in the sine and cosine sum formulas: sin(alpha+alpha) gives sin 2 alpha = 2 sin alpha cos alpha, and cos(alpha+alpha) gives cos 2 alpha = cos^2 alpha - sin^2 alpha.

Why does cos 2 alpha have three equivalent forms?

Because the Pythagorean identity sin^2 alpha + cos^2 alpha = 1 lets cos^2 alpha - sin^2 alpha be rewritten by substituting sin^2 alpha = 1-cos^2 alpha (giving 2cos^2 alpha - 1) or cos^2 alpha = 1-sin^2 alpha (giving 1-2sin^2 alpha) -- all three forms are algebraically equal, just expressed in different single ratios.

How is the half angle identity for cos(alpha/2) obtained from the double angle identity?

Starting from cos alpha = 2 cos^2(alpha/2) - 1 (substituting alpha/2 for alpha in the 2cos^2-1 form), solving for cos(alpha/2) gives cos^2(alpha/2) = (1+cos alpha)/2, and taking the square root gives cos(alpha/2) = +/- sqrt((1+cos alpha)/2).

How is the triple angle identity for sin 3 alpha derived?

By writing 3 alpha as 2 alpha + alpha and expanding sin(2alpha+alpha) = sin 2 alpha cos alpha + cos 2 alpha sin alpha using the double angle formulas, then simplifying with the Pythagorean identity to get sin 3 alpha = 3 sin alpha - 4 sin^3 alpha.

Multiple Choice Questions (MCQs)

The Fundamental Law of Trigonometry states cos(alpha-beta) equals:

A. sin alpha sin beta - cos alpha cos beta

B. cos alpha cos beta + sin alpha sin beta

C. cos alpha sin beta + sin alpha cos beta

D. cos alpha - cos beta

Answer: B. cos alpha cos beta + sin alpha sin beta

Explanation: The fundamental law states cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta.

Two angles alpha and beta are called allied if:

A. alpha = beta

B. alpha + beta = 180 degrees only

C. alpha +/- beta = n(90 degrees), n an integer

D. alpha - beta = 0

Answer: C. alpha +/- beta = n(90 degrees), n an integer

Explanation: Allied angles satisfy alpha +/- beta = n(90 degrees) for some integer n.

sin(pi/2 + theta) equals:

A. sin theta

B. -sin theta

C. cos theta

D. -cos theta

Answer: C. cos theta

Explanation: sin(pi/2 + theta) = cos theta, a direct deduction from the fundamental law.

cos(pi - theta) equals:

A. cos theta

B. -cos theta

C. sin theta

D. -sin theta

Answer: B. -cos theta

Explanation: cos(pi - theta) = -cos theta, since pi is an even multiple of pi/2 (ratio unchanged) but the terminal side is in Quadrant II (cosine negative).

The double angle identity sin 2 alpha equals:

A. sin^2 alpha - cos^2 alpha

B. 2 sin alpha cos alpha

C. sin alpha + cos alpha

D. 2 cos^2 alpha - 1

Answer: B. 2 sin alpha cos alpha

Explanation: sin 2 alpha = 2 sin alpha cos alpha, obtained by setting beta=alpha in the sine sum formula.

Which of these is NOT a valid form of cos 2 alpha?

A. cos^2 alpha - sin^2 alpha

B. 2 cos^2 alpha - 1

C. 1 - 2 sin^2 alpha

D. 2 sin alpha cos alpha

Answer: D. 2 sin alpha cos alpha

Explanation: 2 sin alpha cos alpha is the formula for sin 2 alpha, not cos 2 alpha.

The double angle identity for tangent, tan 2 alpha, equals:

A. tan alpha / 2

B. 2 tan alpha / (1 - tan^2 alpha)

C. tan^2 alpha - 1

D. 2 tan alpha + 1

Answer: B. 2 tan alpha / (1 - tan^2 alpha)

Explanation: tan 2 alpha = 2 tan alpha / (1 - tan^2 alpha).

The triple angle identity for cosine, cos 3 alpha, equals:

A. 4 cos^3 alpha - 3 cos alpha

B. 3 cos alpha - 4 cos^3 alpha

C. cos^3 alpha - 3 cos alpha

D. 4 cos alpha - 3 cos^3 alpha

Answer: A. 4 cos^3 alpha - 3 cos alpha

Explanation: cos 3 alpha = 4 cos^3 alpha - 3 cos alpha.

The product-to-sum identity for 2 sin alpha cos beta is:

A. cos(alpha+beta) + cos(alpha-beta)

B. sin(alpha+beta) + sin(alpha-beta)

C. sin(alpha+beta) - sin(alpha-beta)

D. cos(alpha-beta) - cos(alpha+beta)

Answer: B. sin(alpha+beta) + sin(alpha-beta)

Explanation: 2 sin alpha cos beta = sin(alpha+beta) + sin(alpha-beta).

The sum-to-product identity for sin P + sin Q is:

A. 2 sin((P+Q)/2) cos((P-Q)/2)

B. 2 cos((P+Q)/2) sin((P-Q)/2)

C. 2 cos((P+Q)/2) cos((P-Q)/2)

D. -2 sin((P+Q)/2) sin((P-Q)/2)

Answer: A. 2 sin((P+Q)/2) cos((P-Q)/2)

Explanation: sin P + sin Q = 2 sin((P+Q)/2) cos((P-Q)/2).

Quick Revision Summary

- The Fundamental Law of Trigonometry: cos(alpha-beta) = cos alpha cos beta + sin alpha sin beta, proved via the unit circle
- Cosine sum formula: cos(alpha+beta) = cos alpha cos beta - sin alpha sin beta
- Sine sum/difference: sin(alpha+beta) = sin alpha cos beta + cos alpha sin beta; sin(alpha-beta) = sin alpha cos beta - cos alpha sin beta
- Tangent sum/difference: tan(alpha+beta) = (tan alpha+tan beta)/(1-tan alpha tan beta)
- Allied angles satisfy alpha +/- beta = n(90 degrees); odd multiples of 90 degrees change the ratio to a co-ratio, even multiples keep it the same
- The sign of an allied-angle result is determined by the quadrant of the resulting angle's terminal side
- Double angle: sin 2 alpha = 2 sin alpha cos alpha; cos 2 alpha = cos^2 alpha - sin^2 alpha = 2cos^2 alpha-1 = 1-2sin^2 alpha
- Double angle for tangent: tan 2 alpha = 2 tan alpha / (1 - tan^2 alpha)
- Half angle: cos(alpha/2) = +/- sqrt((1+cos alpha)/2); sin(alpha/2) = +/- sqrt((1-cos alpha)/2)
- Triple angle: sin 3 alpha = 3 sin alpha - 4 sin^3 alpha; cos 3 alpha = 4 cos^3 alpha - 3 cos alpha
- Product-to-sum identities convert a product of two sines/cosines into a sum or difference
- Sum-to-product identities convert a sum or difference of two sines/cosines into a product

Exam Tips

- Memorize the fundamental law first -- every other identity in this unit can be re-derived from it if forgotten
- For allied angles, first check whether the multiple of 90 degrees involved is odd or even to know if the ratio changes to its co-ratio
- Always determine the sign of an allied-angle result last, by checking which quadrant the terminal side falls in
- When solving for missing trigonometric ratios (like sin alpha from cos alpha), always use the given quadrant to pick the correct sign
- For product-to-sum/sum-to-product proofs, look for terms that can be paired and multiplied/divided by 2 or 4 to match a known identity pattern
- Practice rewriting non-standard angles (105, 15, 75 degrees, etc.) as a sum or difference of two standard angles (30, 45, 60, 90 degrees) before applying any formula

Unit 11: Trigonometric Functions and their Graphs

This unit examines trigonometric functions as functions in their own right: their domains and ranges (derived from the unit circle), whether each is even or odd, and the periodicity that makes them repeat their values at fixed intervals. Building on this foundation, the unit shows how to construct and read the graphs of sine, cosine, and tangent, and lists the key properties (domain, range, continuity, symmetry, intercepts, amplitude) that each graph displays.

The unit then develops the general sinusoidal function a + b sin(c theta + d), showing how the constants a, b, c, and d control vertical shift, amplitude, period, and phase shift respectively, and how to find the maximum and minimum values of such functions (and their reciprocals) without graphing. It closes with real-world applications -- Ferris wheels, tidal water levels, temperature cycles, and angle-of-elevation/depression problems -- that show sinusoidal functions modeling genuinely periodic real-world phenomena.

Learning Objectives

- Determine the domain and range of each of the six trigonometric functions
- Determine whether a given trigonometric function is even, odd, or neither
- Find the period of a trigonometric function, including functions of the form sin(cx), tan(cx), etc.
- Construct and interpret the graphs of y = sin x, y = cos x, and y = tan x, and state their key properties
- Identify the vertical shift, amplitude, period, and phase shift of a general sinusoidal function a + b sin(c theta + d)
- Find the maximum and minimum values of sinusoidal functions and their reciprocals
- Apply sinusoidal functions to model real-world periodic phenomena such as Ferris wheels, tides, and temperature cycles

Key Concepts

11.1 Domains and Ranges of Trigonometric Functions

For a point P(x,y) on the unit circle with m angle XOP = theta in standard position, cos theta = x and sin theta = y; since every real theta gives exactly one such point, both sine and cosine are functions with domain R (all real numbers), and since -1<=x<=1 and -1<=y<=1 on the unit circle, both have range [-1,1].

Tangent (=y/x) and secant (=1/x) are undefined wherever x=0, i.e. at odd multiples of pi/2, so their domain excludes those points; cotangent (=x/y) and cosecant (=1/y) are undefined wherever y=0, i.e. at integer multiples of pi, so their domain excludes those points instead. The range of tangent and cotangent is all of R, while the range of secant and cosecant is R minus the open interval (-1,1), since |1/x|>=1 whenever 0<|x|<=1.

11.2 Even and Odd Functions

A function f is even if f(-x)=f(x) for every x in its domain (graph symmetric about the y-axis); it is odd if f(-x)=-f(x) for every x in its domain (graph symmetric about the origin). Since cos(-theta)=cos theta, cosine is an even function; since sin(-theta)=-sin theta, sine is an odd function -- and this same test classifies every other trigonometric function and combination of them.

Recognizing whether a trigonometric expression is even, odd, or neither is often the fastest way to predict symmetry in its graph before plotting a single point, and is a routine first step when simplifying or analyzing a new trigonometric function.

11.3 Period of Trigonometric Functions

A function f is periodic if f(theta+p)=f(theta) for all theta in its domain, and the smallest positive such p is called the period. By substituting theta=0 and testing candidate values, sine and cosine are proved to have period 2 pi, while tangent and cotangent have the smaller period pi (since tan(theta+pi)=tan theta, unlike sin(theta+pi)=-sin theta).

For a function of the form sin(cx) or tan(cx), the period scales inversely with |c|: if the period of sin x is 2pi, then sin(cx) repeats every 2pi/|c|, since sin(c(x+2pi/c))=sin(cx+2pi)=sin(cx). Adding or subtracting a constant (like the '3' in 3+tan(x/3)) never affects the period, only the vertical position of the graph.

11.4 Graphs of Trigonometric Functions

To graph a trigonometric function, a table of ordered pairs (x, y) is built using standard angle values, the points are plotted with angle measure on the x-axis and function value on the y-axis, and the points are joined with a smooth curve; because sine, cosine, and tangent are periodic, the resulting curve repeats indefinitely in both directions once one full period is drawn.

The graphs of y=sin x and y=cos x are smooth, continuous waves confined between y=-1 and y=1, with sine passing through the origin and cosine starting at its maximum; the graph of y=tan x, by contrast, is discontinuous, breaking into separate branches at each odd multiple of pi/2 where the function is undefined, with the curve shooting toward positive or negative infinity as it approaches each break.

11.5 Maximum and Minimum Values of Sinusoidal Functions

For a general sinusoidal function f(theta) = a + b sin(c theta + d) (or the cosine equivalent), a is the vertical shift (moves the whole graph up/down without changing its shape), |b| is the amplitude (maximum height of the wave from its midline), the period is 2pi/c, and d is the phase shift (moves the wave left or right along the x-axis).

Since sine and cosine never exceed 1 or fall below -1, the maximum value of f(theta) is M = a + |b| (attained when sin/cos equals 1) and the minimum is m = a - |b| (attained when sin/cos equals -1); for the reciprocal of such a function, the maximum of the reciprocal is 1/m and the minimum is 1/M, since taking a reciprocal reverses which extreme is 'largest'.

11.6 Real-World Applications of Sinusoidal Functions

Sinusoidal functions naturally model any quantity that oscillates in a regular, repeating cycle: the height of a rider on a Ferris wheel over time, the rise and fall of tidal water levels through the day, daily or seasonal temperature swings, and population cycles driven by seasonal migration -- in every case, the period matches the physical cycle time, the amplitude matches half the total swing, and the vertical shift matches the average/center value.

Angle of elevation and depression problems (such as finding the height of a distant peak from two different angles) use the sum and difference identities from Unit 10 together with basic right-triangle trigonometry, often solved elegantly using the algebraic technique of componendo and dividendo on a ratio of two tangent expressions.

Important Definitions

What is the domain of the sine and cosine functions?

The set of all real numbers, R, since every real angle theta gives exactly one point (cos theta, sin theta) on the unit circle.

What is the range of the sine and cosine functions?

The closed interval [-1, 1].

What is the domain of the tangent function?

All real numbers except odd multiples of pi/2, i.e. R - {x : x = (2n+1)pi/2, n in Z}.

What does it mean for a function f to be even?

f(-x) = f(x) for every x in the domain of f; its graph is symmetric about the y-axis.

What does it mean for a function f to be odd?

f(-x) = -f(x) for every x in the domain of f; its graph is symmetric about the origin.

What is the period of a function?

The smallest positive number p such that f(theta+p) = f(theta) for all theta in the domain of f.

What are the periods of sine, cosine, tangent, and cotangent?

Sine and cosine have period 2 pi; tangent and cotangent have period pi.

In the sinusoidal function a + b sin(c theta + d), what does 'a' represent?

The vertical shift -- the upward or downward translation of the graph's midline.

In the sinusoidal function a + b sin(c theta + d), what does '|b|' represent?

The amplitude -- the maximum height of the wave measured from its midline.

In the sinusoidal function a + b sin(c theta + d), what does 'd' represent?

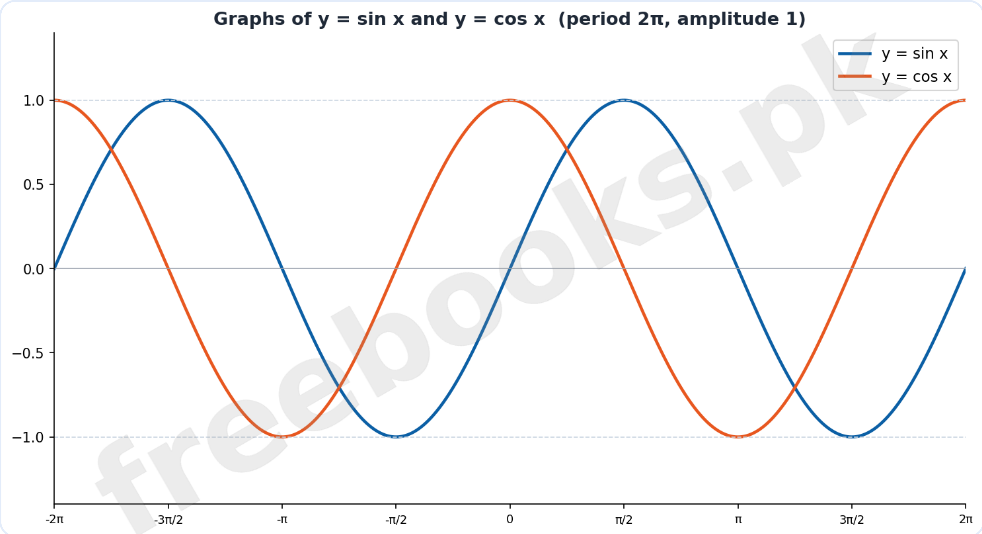
The phase shift -- the horizontal translation of the graph along the x-axis.

Key Facts and Relations

Topic	Key Fact / Relation
Domain/Range of sin, cos	Domain = R; Range = [-1, 1]
Domain of tan, sec	R - {(2n+1) pi/2 : n in Z}
Domain of cot, csc	R - {n pi : n in Z}
Period of sin x, cos x	2 pi
Period of tan x, cot x	pi
Period of sin(cx) or cos(cx)	2 pi / |c|
Period of tan(cx) or cot(cx)	pi / |c|
Sinusoidal function	f(theta) = a + b sin(c theta + d): vertical shift a, amplitude |b|, period 2pi/c, phase shift d
Maximum / minimum value	M = a + |b|; m = a - |b|
Max/min of reciprocal 1/f(theta)	max = 1/m; min = 1/M

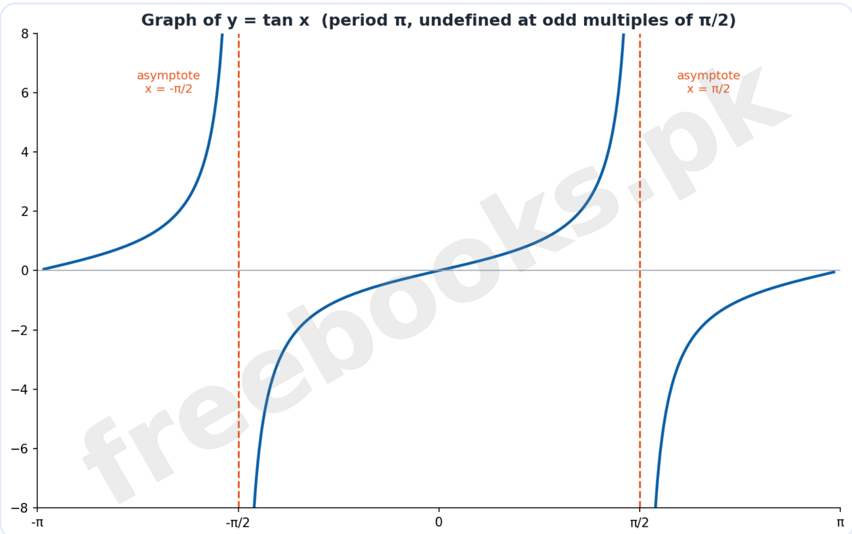
Diagrams

Graphs of y = sin x and y = cos x: The sine and cosine curves plotted together from -2pi to 2pi, showing their shared period of 2pi, amplitude of 1, and the horizontal shift between them



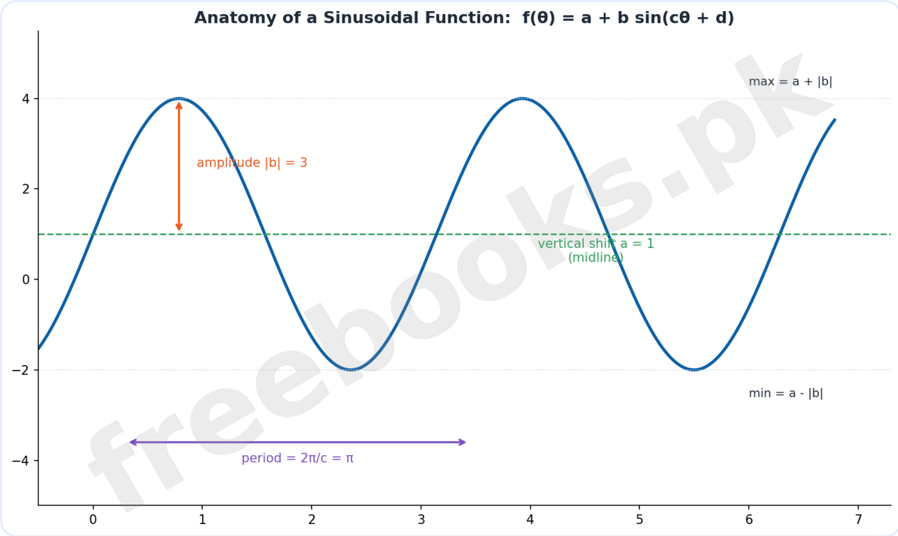
Graphs of y = sin x and y = cos x

Graph of y = tan x with Asymptotes: The tangent curve from -pi to pi, showing its period of pi and the vertical asymptotes at x = -pi/2 and x = pi/2 where the function is undefined



Graph of y = tan x with Asymptotes

Anatomy of a Sinusoidal Function: A labeled sine wave for f(theta) = 1 + 3 sin(2 theta), showing the vertical shift (midline), amplitude, and period



Anatomy of a Sinusoidal Function

Solved Examples

Example 1: Determining Whether a Function is Even or Odd

Problem: Determine whether f(x) = sin^2 x is even, odd, or neither.

1. Compute f(-x): f(-x) = sin^2(-x) = (sin(-x))^2 = (-sin x)^2 = sin^2 x.
2. Compare f(-x) to f(x): f(-x) = sin^2 x = f(x).
3. Since f(-x) = f(x) for every x, the function satisfies the definition of an even function.
4. Final answer: f(x) = sin^2 x is an even function (its graph is symmetric about the y-axis).

Example 2: Finding the Period of Trigonometric Functions

Problem: Find the periods of (i) sin 2x and (ii) 3 + tan(x/3).

1. For sin 2x: the period of sin x is 2pi, so sin(2x + 2pi) = sin 2x means sin 2(x+pi) = sin 2x -- the value repeats when x increases by pi.
2. So the period of sin 2x is pi (using period = 2pi/|c| with c=2: 2pi/2 = pi).

3. For 3 + tan(x/3): the period of tan x is pi, so tan(x/3 + pi) = tan(x/3) means tan(1/3)(x+3pi) = tan(x/3) -- the value repeats when x increases by 3pi.
4. So the period of tan(x/3) is 3pi (using period = pi/|c| with c=1/3: pi/(1/3) = 3pi); adding the constant 3 does not change the period.
5. Final answer: period of sin 2x is pi; period of 3+tan(x/3) is 3pi.

Example 3: Finding Maximum and Minimum Values of Sinusoidal Functions

Problem: Find the maximum and minimum values of (i) 2 + 3 sin x and (ii) 5 - 2 cos 3x.

1. For 2 + 3 sin x: here a=2, b=3. Maximum M = a + |b| = 2+3 = 5, occurring when sin x = 1.
2. Minimum m = a - |b| = 2-3 = -1, occurring when sin x = -1.
3. For 5 - 2 cos 3x: here a=5, b=-2. Maximum M = a + |b| = 5 + 2 = 7, occurring when cos 3x = 1.
4. Minimum m = a - |b| = 5 - 2 = 3, occurring when cos 3x = -1.
5. Final answer: (i) max=5, min=-1; (ii) max=7, min=3.

Example 4: Finding Maximum and Minimum Values of a Reciprocal Sinusoidal Function

Problem: Find the maximum and minimum values of the reciprocal of 5 - 2 cos 3x.

1. First find the maximum and minimum of 5 - 2 cos 3x itself: from the previous example, M=7 and m=3.
2. For a reciprocal function g(x) = 1/(5-2cos3x), the maximum of g occurs where the original function is at its minimum, and vice versa.
3. Maximum of g(x) = 1/m = 1/3, approximately 0.33.
4. Minimum of g(x) = 1/M = 1/7, approximately 0.14.

5. Final answer: maximum = 1/3 (approximately 0.33), minimum = 1/7 (approximately 0.14).

Example 5: Modeling a Ferris Wheel with a Sinusoidal Function

Problem: A Ferris wheel with a radius of 45 feet has its lowest point 5 feet above the ground and completes one revolution every 60 seconds. Model the height of a rider as a function of time, and find the height after 40 seconds.

1. The period equals the time for one revolution: 2pi/c = 60, so c = 2pi/60 = pi/30.
2. The amplitude b equals the radius: b = 45. The vertical shift a is the height of the wheel's center: since the lowest point is 5 feet up, a = 5 + 45 = 50.
3. Since the rider starts at the lowest point (minimum), model with a negative cosine: h(t) = -b cos(ct) + a = -45 cos(pi t/30) + 50.
4. Substitute t=40: h(40) = -45 cos(pi(40)/30) + 50 = -45 cos(4pi/3) + 50.
5. Since cos(4pi/3) = -0.5, h(40) = -45(-0.5) + 50 = 22.5 + 50 = 72.5. Final answer: the rider is 72.5 feet above the ground after 40 seconds.

Example 6: Modeling Tidal Water Level with a Sinusoidal Function

Problem: The water level of a tidal river is modeled by L(t) = 8 + 4 sin(pi t/6), where t is in hours. Find the water level at t=3 hours and the minimum water level.

1. Substitute t=3 into the model: L(3) = 8 + 4 sin(pi(3)/6) = 8 + 4 sin(pi/2).
2. Since sin(pi/2) = 1, L(3) = 8 + 4(1) = 12. So the water level at t=3 hours is 12 feet.
3. For the minimum water level, use the minimum value of sine: sin(pi t/6) = -1.
4. Substitute: L(t) = 8 + 4(-1) = 8 - 4 = 4.

5. Final answer: water level at t=3 hours is 12 feet; the minimum water level is 4 feet.

Example 7: Finding the Height of a Peak Using Angles of Elevation and Depression

Problem: From a point 100 m above a lake's surface, the angle of elevation of a cliff peak is 15 degrees and the angle of depression of its reflection is 30 degrees. Find the height of the peak.

1. Let h = height of the peak above the lake, y = horizontal distance to the peak. From the elevation angle: tan 15 = (h-100)/y. From the depression angle to the reflection (which appears 100m below the lake surface): tan 30 = (h+100)/y.
2. Divide the two equations: tan15/tan30 = (h-100)/(h+100).
3. Apply componendo and dividendo: (tan30+tan15)/(tan30-tan15) = [(h+100)+(h-100)] / [(h+100)-(h-100)] = 2h/200 = h/100.
4. Substitute known values tan15=0.2679 and tan30=0.5774: h = 100 x (0.5774+0.2679)/(0.5774-0.2679) = 100 x 0.8453/0.3095.
5. Final answer: h is approximately 273 metres.

Example 8: Finding the Period and Maximum Height of a Swing

Problem: The height of a swing seat is modeled by h(t) = 1.5 + 1.2 sin(3 pi t), where t is in seconds. Find the maximum height and the period of the motion.

1. Here a=1.5, b=1.2, c=3pi. Maximum height M = a + |b| = 1.5 + 1.2 = 2.7 metres, occurring when sin(3 pi t) = 1.
2. Minimum height m = a - |b| = 1.5 - 1.2 = 0.3 metres, occurring when sin(3 pi t) = -1.
3. The period is given by 2 pi / c = 2 pi / (3 pi) = 2/3 seconds.

4. This means the swing completes one full back-and-forth cycle every 2/3 of a second.
5. Final answer: maximum height = 2.7 m, minimum height = 0.3 m, period = 2/3 second.

Short Questions & Answers

What is the domain of the sine function?

All real numbers, R.

What is the range of the tangent function?

All real numbers, R (unbounded).

Is the cosine function even or odd?

Even, since cos(-theta) = cos theta.

Is the sine function even or odd?

Odd, since sin(-theta) = -sin theta.

What is the period of tan x?

pi.

What does the amplitude of a sinusoidal function represent?

The maximum height of the wave measured from its midline, equal to |b| in a + b sin(c theta + d).

What is the formula for the maximum value of a + b sin(c theta + d)?

M = a + |b|, occurring when sin(c theta + d) = 1.

Long Questions & Answers

Explain how the domains and ranges of the six trigonometric functions are derived from the unit circle, and describe the graphs of sine, cosine, and tangent.

How are the domain and range of sine and cosine determined from the unit circle?

For a point P(x,y) on the unit circle with angle theta, cos theta = x and sin theta = y; since every real theta produces exactly one such point, the domain of both is R, and since -1<=x,y<=1 on the unit circle, the range of both is [-1,1].

Why are tangent and secant undefined at odd multiples of pi/2?

Because tan theta = y/x and sec theta = 1/x, both of which require x not equal to 0; x=0 occurs exactly when the terminal side lies on the y-axis, i.e. at odd multiples of pi/2, so those values are excluded from the domain.

What are the key visual properties of the graphs of y=sin x and y=cos x?

Both are smooth, continuous waves confined between y=-1 and y=1 with period 2pi; sine passes through the origin and is symmetric about the origin (odd), while cosine starts at its maximum value 1 and is symmetric about the y-axis (even).

How does the graph of y=tan x differ from the graphs of sine and cosine?

Unlike the continuous sine and cosine curves, the tangent graph is discontinuous: it breaks into separate branches at each odd multiple of pi/2 (where tangent is undefined), with the curve increasing or decreasing without bound as it approaches each break, and it has a shorter period of pi instead of 2pi.

Describe the general sinusoidal function a + b sin(c theta + d), explain what each constant controls, and show how to find its maximum and minimum values.

What does the constant 'a' control in a + b sin(c theta + d)?

It controls the vertical shift -- translating the entire graph up or down along the y-axis to set the midline of the wave, without changing its shape, amplitude, or period.

What does the constant 'b' control?

Its absolute value |b| is the amplitude -- the maximum height of the wave measured from its midline; a larger |b| makes the wave taller.

What do the constants 'c' and 'd' control?

c determines the period, equal to 2pi/c -- a larger c compresses the wave horizontally, fitting more cycles into the same interval; d is the phase shift, translating the wave left or right along the x-axis without changing its shape or period.

How are the maximum and minimum values of the function found?

Since sine (and cosine) never exceed 1 or fall below -1, the maximum value is M = a + |b| (when sin/cos equals 1) and the minimum is m = a - |b| (when sin/cos equals -1); for a reciprocal of such a function, the maximum becomes 1/m and the minimum becomes 1/M.

Multiple Choice Questions (MCQs)

The domain of the sine and cosine functions is:

A. Only positive real numbers

B. All real numbers, R

C. [-1, 1]

D. Only integers

Answer: B. All real numbers, R

Explanation: Every real number theta produces exactly one point on the unit circle, so the domain of sin and cos is all of R.

The range of the sine function is:

A. R

B. [0, 1]

C. [-1, 1]

D. (-infinity, infinity)

Answer: C. [-1, 1]

Explanation: Since sin theta equals the y-coordinate on the unit circle, and -1<=y<=1, the range is [-1,1].

A function f is even if:

A. f(-x) = -f(x)

B. f(-x) = f(x)

C. f(x) = 0

D. f(x) = x

Answer: B. f(-x) = f(x)

Explanation: An even function satisfies f(-x) = f(x) for every x in its domain.

The cosine function is:

A. Odd

B. Even

C. Neither odd nor even

D. Undefined at x=0

Answer: B. Even

Explanation: cos(-theta) = cos theta, so cosine is an even function.

The period of the tangent function is:

A. 2 pi

B. pi/2

C. pi

D. 4 pi

Answer: C. pi

Explanation: tan(theta+pi) = tan theta, and pi is the smallest such positive value, so the period of tangent is pi.

The period of sin(3x) is:

A. 2 pi

B. 2 pi / 3

C. 3 pi

D. pi / 3

Answer: B. 2 pi / 3

Explanation: Using period = 2pi/|c| with c=3, the period is 2pi/3.

The graph of y = tan x is:

A. Continuous everywhere

B. Discontinuous at odd multiples of pi/2

C. A straight line

D. Bounded between -1 and 1

Answer: B. Discontinuous at odd multiples of pi/2

Explanation: The tangent graph breaks into separate branches at odd multiples of pi/2, where the function is undefined.

In f(theta) = a + b sin(c theta + d), the amplitude is:

A. a

B. |b|

C. c

D. d

Answer: B. |b|

Explanation: The amplitude is the absolute value of b, the maximum height of the wave from its midline.

The maximum value of f(theta) = a + b sin(c theta + d) is:

A. a - |b|

B. a + |b|

C. a x |b|

D. |b| - a

Answer: B. a + |b|

Explanation: The maximum value M = a + |b|, occurring when sin(c theta + d) = 1.

For a reciprocal function 1/f(theta), if f(theta) has maximum M and minimum m, the reciprocal's maximum is:

A. M

B. m

C. 1/M
D. 1/m
Answer: D. 1/m
Explanation: Taking a reciprocal reverses the extremes: the reciprocal's maximum is 1/m (using the original minimum).

Quick Revision Summary

- Sine and cosine have domain R and range [-1,1]; tangent and cotangent have range R
- Tangent and secant are undefined at odd multiples of pi/2; cotangent and cosecant are undefined at integer multiples of pi
- A function is even if f(-x)=f(x) (symmetric about y-axis); odd if f(-x)=-f(x) (symmetric about origin)
- Cosine is even; sine is odd
- Sine and cosine have period 2 pi; tangent and cotangent have period pi
- The period of sin(cx) or cos(cx) is 2pi/|c|; the period of tan(cx) is pi/|c|
- The graphs of sin x and cos x are smooth, continuous waves between -1 and 1; the graph of tan x is discontinuous with asymptotes
- In f(theta) = a + b sin(c theta + d): a = vertical shift, |b| = amplitude, 2pi/c = period, d = phase shift
- Maximum value M = a + |b|; minimum value m = a - |b|
- For a reciprocal function, maximum = 1/m and minimum = 1/M (using the original function's extremes)
- Sinusoidal functions model periodic real-world phenomena: Ferris wheels, tides, temperature cycles, population cycles
- Angle of elevation/depression problems often use componendo and dividendo to solve for an unknown height

Exam Tips

- Always identify a, b, c, d first when working with a sinusoidal function before computing max/min or period
- Remember that adding a constant to a trigonometric function shifts the graph vertically but never changes its period

- For reciprocal functions, find the max/min of the original function first, then swap and invert to get the reciprocal's min/max
- When modeling real-world periodic phenomena, match the period to the physical cycle time and the vertical shift to the average value
- Check whether the phenomenon starts at a maximum/minimum (use cosine) or at the midline (use sine) when choosing which function to model with
- For angle of elevation/depression problems with two ratios, try componendo and dividendo before attempting to solve directly

Unit 12: Limit and Continuity

This unit introduces the limit of a function -- the value a function approaches as its input gets arbitrarily close to a given number -- which is the foundational idea on which all of calculus rests. Starting from the informal meaning of phrases like 'x approaches a', the unit builds up to a precise definition, a set of theorems that let limits be evaluated by direct substitution in most cases, and algebraic techniques (factoring, rationalizing) for the 0/0 indeterminate form that substitution alone cannot resolve.

The unit then extends these ideas to limits at infinity, the limit of a sequence, the important number e as a limit, and the Sandwich Theorem (used to prove the famous result lim(theta->0) sin(theta)/theta = 1). Finally, it defines one-sided limits and uses them to give a precise, three-condition definition of continuity at a point, closing with real-world applications of limits in radioactive decay, compound interest, population growth, and other transcendental-function models.

Learning Objectives

- Explain the meaning of the phrases 'x approaches a', 'x approaches zero', and 'x approaches infinity'
- State and apply the theorems on limits of functions to evaluate limits by direct substitution
- Evaluate limits that produce a 0/0 indeterminate form using factoring and rationalization
- Evaluate limits at infinity, including limits of sequences and limits involving the number e
- Apply the Sandwich Theorem to prove and use lim(theta->0) sin(theta)/theta = 1
- Determine left-hand and right-hand limits, and use them to test whether a limit exists at a point
- Determine whether a function is continuous or discontinuous at a point, and apply limits to real-world models

Key Concepts

12.1 The Meaning of a Limit

The phrase 'x approaches a' (written x -> a) means x gets arbitrarily close to a from both the left and right sides, without ever actually equaling a; this is different from x = a, which means x is actually equal to a. If a function f(x) approaches a specific number L as x approaches a from both sides, L is called the limit of f(x) as x approaches a, written lim(x->a) f(x) = L.

This idea can be seen numerically: for f(x) = x^3, as x takes values closer and closer to 2 from both the left (1.9, 1.99, 1.999, ...) and the right (2.1, 2.01, 2.001, ...), f(x) gets closer and closer to 8 -- so lim(x->2) x^3 = 8, even though evaluating this way for every limit is impractical, which is why the theorems on limits are needed.

12.2 Theorems on Limits of Functions

If lim(x->a) f(x) = L and lim(x->a) g(x) = M, a set of theorems establishes that the limit of a sum, difference, constant multiple, product, quotient (with nonzero denominator limit), or integer power of functions equals the same operation applied to their individual limits -- for example, lim(x->a)[f(x)+g(x)] = L+M and lim(x->a)[f(x)g(x)] = LM.

The practical consequence of these theorems is that most limits of polynomial, rational, and root functions can be evaluated by simple direct substitution of the value x approaches into the function -- no special technique is needed unless substitution produces an indeterminate form like 0/0.

12.3 Evaluating 0/0 Forms and Limits at Infinity

When direct substitution produces 0/0, the function must first be simplified -- typically by factoring and cancelling a common factor (using identities like x^n - a^n = (x-a)(x^(n-1)+ax^(n-2)+...+a^(n-1))) or by rationalizing a numerator or denominator containing a square root -- before the limit can be evaluated.

Limits at infinity (x -> +infinity or x -> -infinity) describe the long-term behavior of a function; a key result is that lim(x->+-infinity) a/x^p = 0 for any real a and positive rational p. For a ratio of polynomials, dividing every term of the numerator and denominator by the highest power of x in the denominator reduces the limit to a form this result can evaluate directly.

12.4 The Number e and the Sandwich Theorem

The important limit lim(n->infinity) (1+1/n)^n = e (approximately 2.718281) arises naturally from the binomial theorem, and its counterpart lim(x->0) (1+x)^(1/x) = e is obtained by substituting n=1/x; both forms are used to express related limits (like (1+3/n)^(2n)) in terms of e through algebraic manipulation of the exponent.

The Sandwich Theorem states that if f(x) <= g(x) <= h(x) near c (except possibly at c) and both f and h approach the same limit L as x -> c, then g must also approach L. This theorem is used geometrically (comparing the areas of a triangle, a circular sector, and a larger triangle) to prove the cornerstone trigonometric limit lim(theta->0) sin(theta)/theta = 1.

12.5 One-Sided Limits and the Criterion for Existence

The left-hand limit lim(x->c-) f(x) = L describes the value f(x) approaches as x approaches c only from values less than c; the right-hand limit lim(x->c+) f(x) = M describes the same from values greater than c. The two-sided limit lim(x->c) f(x) exists, and equals a value L, if and only if both one-sided limits exist and are equal to that same L.

This criterion is especially useful for piecewise-defined functions, where the formula for f(x) changes at the point c -- the left-hand and right-hand limits must be computed separately using whichever piece of the definition applies on each side, and then compared.

12.6 Continuity, Discontinuity, and Real-World Applications

A function f is continuous at a number c if and only if three conditions all hold: f(c) is defined, lim(x->c) f(x) exists, and lim(x->c) f(x) = f(c). If any one of these three conditions fails, f is discontinuous at c -- this might show up as a removable break (an undefined or mismatched point), or a jump (where the one-sided limits disagree, as commonly happens with piecewise functions).

Limits of transcendental functions (especially e^x and its variants) model many real-world long-term behaviors: radioactive decay A(t)=A0*e^(-kt) approaches 0 as t->infinity, continuously compounded investments A(t)=P0*e^(rt) grow without bound, and logistic population models approach a fixed carrying capacity -- in each case, evaluating lim(t->infinity) of the model reveals its eventual real-world behavior.

Important Definitions

What does the phrase 'x approaches a' mean?

x gets arbitrarily close to a from both sides, without x ever actually equaling a.

What is the limit of a function f(x) as x approaches a?

The single number L that f(x) approaches as x approaches a from both sides, written lim(x->a) f(x) = L.

What is an indeterminate form?

An expression like 0/0 obtained by direct substitution that does not by itself reveal the value of the limit, requiring further algebraic simplification.

What is the limit of a sequence?

The value L that the terms of a sequence {a_n} approach as n approaches infinity; if such an L exists, the sequence converges to L.

What is a divergent sequence?

A sequence that does not converge to any single value -- either increasing/ decreasing without bound, or oscillating between values.

What does the Sandwich Theorem state?

If f(x) <= g(x) <= h(x) near c (except possibly at c) and both f and h approach the same limit L as x->c, then g(x) also approaches L.

What is the left-hand limit of f(x) as x approaches c?

The value f(x) approaches as x approaches c only through values less than c.

What is the right-hand limit of f(x) as x approaches c?

The value f(x) approaches as x approaches c only through values greater than c.

What is the criterion for the existence of lim(x->c) f(x)?

The limit exists and equals L if and only if the left-hand and right-hand limits both exist and are equal to L.

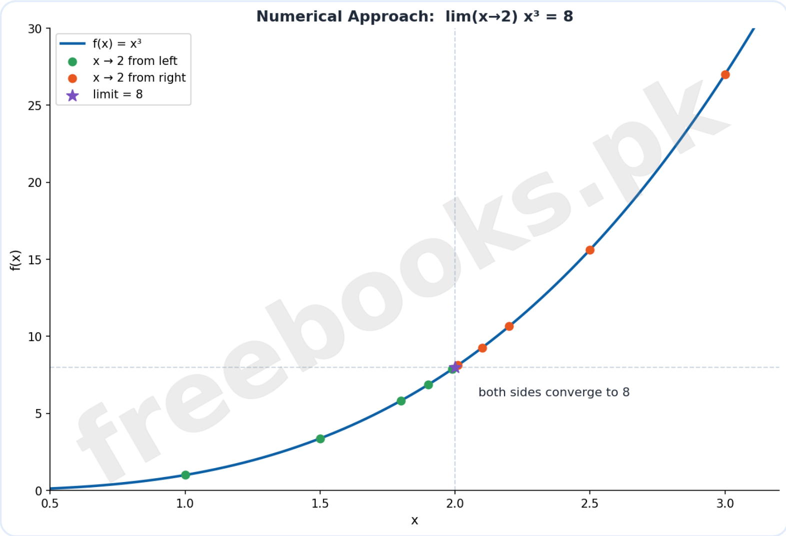
What are the three conditions for f to be continuous at c?
f(c) is defined; lim(x->c) f(x) exists; and lim(x->c) f(x) = f(c).

Key Facts and Relations

Topic	Key Fact / Relation
Basic limit theorems	lim[f+g] = L+M; lim[fg] = LM; lim[f/g] = L/M (M not 0)
Important algebraic limit	lim(x->a) (x^n - a^n)/(x-a) = n a^(n-1)
Rationalization limit	lim(x->0) (sqrt(x+a) - sqrt(a))/x = 1/(2 sqrt(a))
Limit at infinity	lim(x-> +-infinity) a/x^p = 0, p > 0
The number e	lim(n->infinity) (1 + 1/n)^n = e; lim(x->0) (1+x)^(1/x) = e
Exponential limit	lim(x->0) (a^x - 1)/x = log_e a; lim(x->0) (e^x - 1)/x = 1
Sandwich theorem result	lim(theta->0) sin(theta)/theta = 1
Criterion for limit existence	lim(x->c) f(x) = L iff lim(x->c-) f(x) = lim(x->c+) f(x) = L
Continuity at a point	f continuous at c iff f(c) defined, lim(x->c) f(x) exists, and lim(x->c) f(x) = f(c)

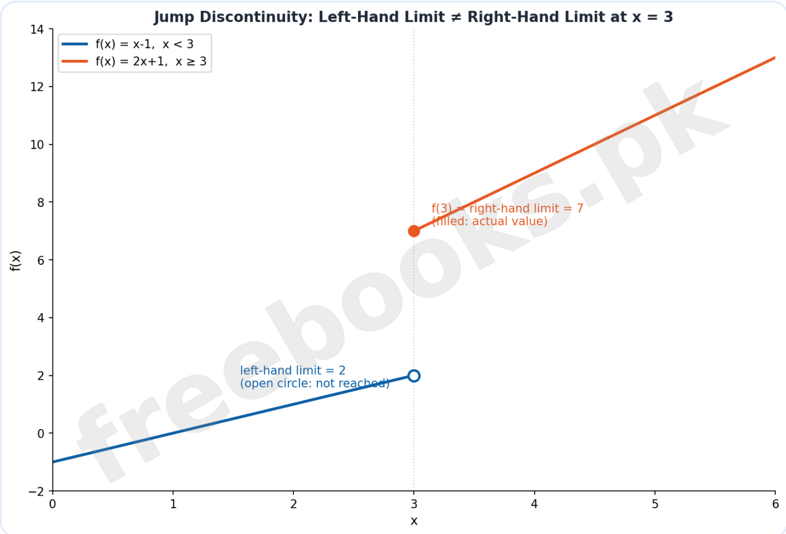
Diagrams

Numerical Approach to a Limit: A graph of f(x)=x^3 with points approaching x=2 from the left and right, both converging to the limit value 8



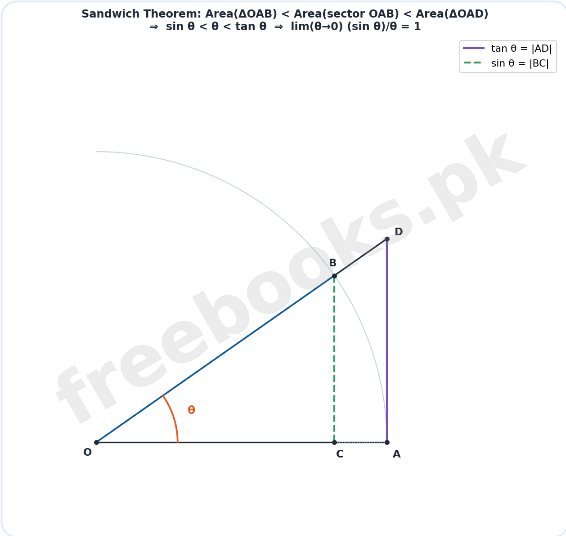
Numerical Approach to a Limit

Jump Discontinuity: A piecewise graph showing a function with different left-hand and right-hand limits at x=3, illustrating a jump discontinuity



Jump Discontinuity

Geometric Proof: lim(theta->0) sin(theta)/theta = 1: A unit-circle sector diagram showing triangle OAB, sector OAB, and triangle OAD, used with the Sandwich Theorem to prove the limit



Geometric Proof: lim(theta->0) sin(theta)/theta = 1

Solved Examples

Example 1: Evaluating a 0/0 Limit by Factoring

Problem: Evaluate lim(x->1) (x^2-1)/(x^2-x).

1. Substitute x=1 directly: (1-1)/(1-1) = 0/0, an indeterminate form, so factoring is needed.
2. Factor the numerator and denominator: x^2-1 = (x-1)(x+1), and x^2-x = x(x-1).
3. Cancel the common factor (x-1): (x-1)(x+1) / [x(x-1)] = (x+1)/x.
4. Now substitute x=1 into the simplified expression: (1+1)/1 = 2/1 = 2.
5. Final answer: lim(x->1) (x^2-1)/(x^2-x) = 2.

Example 2: Evaluating a 0/0 Limit by Rationalizing

Problem: Evaluate lim(x->3) (x-3)/(sqrt(x)-sqrt(3)).

1. Substitute x=3 directly: (3-3)/(sqrt3-sqrt3) = 0/0, an indeterminate form.
2. Multiply numerator and denominator by the conjugate (sqrt(x)+sqrt(3)): the denominator becomes (sqrt(x)-sqrt(3))(sqrt(x)+sqrt(3)) = x-3.

3. The expression becomes (x-3)(sqrt(x)+sqrt(3)) / (x-3), and the common factor (x-3) cancels, leaving sqrt(x)+sqrt(3).
4. Substitute x=3: sqrt(3)+sqrt(3) = 2 sqrt(3).
5. Final answer: lim(x->3) (x-3)/(sqrt(x)-sqrt(3)) = 2 sqrt(3).

Example 3: Finding the Limit of a Sequence

Problem: Find the limit of the sequence a_n = (2n+3)/(n+1) as n -> infinity.

1. Divide every term in the numerator and denominator by n, the highest power of n present: a_n = (2 + 3/n) / (1 + 1/n).
2. As n -> infinity, both 3/n and 1/n approach 0.
3. Substitute these limits: a_n -> (2+0)/(1+0) = 2/1 = 2.
4. Since the terms approach a single finite value, the sequence converges.
5. Final answer: lim(n->infinity) (2n+3)/(n+1) = 2.

Example 4: Evaluating a Limit at Infinity for a Rational Function

Problem: Evaluate lim(x->+infinity) (5x^4-10x^2+1) / (-3x^3+10x^2+50).

1. Identify the highest power of x in the denominator: x^3.
2. Divide every term in the numerator and denominator by x^3: numerator becomes 5x - 10/x + 1/x^3; denominator becomes -3 + 10/x + 50/x^3.
3. As x -> +infinity, the terms 10/x, 1/x^3, 10/x, and 50/x^3 all approach 0, while 5x grows without bound.
4. The expression approaches (infinity - 0 + 0) / (-3 + 0 + 0), i.e. infinity divided by a negative constant.
5. Final answer: since the numerator grows without bound while the denominator approaches -3, the limit is -infinity.

Example 5: Expressing a Limit in Terms of e

Problem: Express lim(n->+infinity) (1 + 3/n)^(2n) in terms of e.

1. Recognize the resemblance to lim(n->infinity) (1+1/n)^n = e; rewrite the exponent to match this pattern.
2. Write (1+3/n)^(2n) = [(1+3/n)^(n/3)]^6, since (n/3) x 6 = 2n.
3. Let m = n/3; as n -> infinity, m -> infinity also, so (1+3/n)^(n/3) = (1+1/m)^m -> e.
4. Substitute back: [(1+1/m)^m]^6 -> e^6.
5. Final answer: lim(n->+infinity) (1+3/n)^(2n) = e^6.

Example 6: Applying the Sandwich Theorem Result to Trigonometric Limits

Problem: Evaluate (i) lim(theta->0) sin(7 theta)/theta and (ii) lim(theta->0) (1-cos theta)/theta.

1. For (i): let x=7 theta, so theta = x/7; as theta->0, x->0 also. Rewrite: sin(7theta)/theta = sin(x)/(x/7) = 7 sin(x)/x.
2. Using lim(x->0) sin(x)/x = 1: lim(theta->0) sin(7theta)/theta = 7(1) = 7.
3. For (ii): multiply numerator and denominator by (1+cos theta): (1-cos theta)/theta = (1-cos^2 theta) / [theta(1+cos theta)] = sin^2(theta) / [theta(1+cos theta)].
4. Rewrite as sin(theta) . [sin(theta)/theta] . [1/(1+cos theta)], and take the limit of each factor: 0 . 1 . (1/2) = 0.
5. Final answer: (i) lim(theta->0) sin(7theta)/theta = 7; (ii) lim(theta->0) (1-cos theta)/theta = 0.

Example 7: Discussing Continuity of a Piecewise Function

Problem: Discuss the continuity of f(x) = x-1 if x<3, and f(x) = 2x+1 if x>=3, at x=3.

1. Check condition (i): f(3) = 2(3)+1 = 7, so f(3) is defined.
2. Find the left-hand limit: lim(x->3-) f(x) = lim(x->3-) (x-1) = 3-1 = 2.
3. Find the right-hand limit: lim(x->3+) f(x) = lim(x->3+) (2x+1) = 2(3)+1 = 7.
4. Since the left-hand limit (2) does not equal the right-hand limit (7), lim(x->3) f(x) does not exist -- condition (ii) fails.
5. Final answer: f(x) is discontinuous at x=3, because the two-sided limit does not exist (a jump discontinuity), even though f(3) itself is defined.

Example 8: Applying Limits to a Real-World Population Model

Problem: A town's population is modeled by P(t) = 100000 / (1 + 9 e^(-0.5t)). Find the long-term population as t -> infinity.

1. As t -> infinity, the exponent -0.5t -> -infinity, so e^(-0.5t) -> 0 (since e raised to a very negative power approaches 0).
2. Substitute this into the denominator: 1 + 9(0) = 1.
3. So the model approaches P(t) -> 100000/1 = 100000.
4. This means the population grows but levels off, approaching a fixed maximum rather than growing without bound.
5. Final answer: the long-term population as t -> infinity is 100,000 (the model's carrying capacity).

Short Questions & Answers

What does x -> a mean?

x gets arbitrarily close to a from both sides, without ever equaling a.

What is an indeterminate form?

An expression like 0/0 that does not directly reveal a limit's value and requires further simplification.

What is the value of lim(n->infinity) (1+1/n)^n?

e, approximately 2.718281.

What does the Sandwich Theorem allow you to conclude?

If a function is squeezed between two others that both approach the same limit L, it also approaches L.

What is the value of lim(theta->0) sin(theta)/theta?

1.

What is the criterion for lim(x->c) f(x) to exist?

The left-hand limit and right-hand limit must both exist and be equal.

What are the three conditions for continuity of f at x=c?

f(c) is defined; lim(x->c) f(x) exists; and lim(x->c) f(x) equals f(c).

Long Questions & Answers

Explain how the theorems on limits allow most limits to be evaluated by substitution, and describe the algebraic techniques needed when substitution produces a 0/0 form.

What do the theorems on limits state about sums, products, and quotients of functions?

If lim(x->a) f(x)=L and lim(x->a) g(x)=M, then the limit of their sum, difference, product, and quotient (with M not 0) equals the same operation applied to L and M -- for example, lim[f(x)g(x)] = LM.

Why do these theorems mean most limits can be found by direct substitution?

Since sums, differences, products, and quotients of limits reduce to the same operations on the individual function values at a, evaluating the limit of a polynomial or rational function (with nonzero denominator) is the same as simply substituting the approached value into the function.

What should be done when direct substitution produces a 0/0 form?

The function must be algebraically simplified first -- typically by factoring both the numerator and denominator to reveal and cancel a common factor, using identities like $x^n-a^n=(x-a)(x^{(n-1)}+\ldots+a^{(n-1)})$.

How is rationalization used when the 0/0 form involves a square root?

Multiplying the numerator and denominator by the conjugate of the square-root expression eliminates the root from the problematic term, turning the difference of square roots into a difference of squares that can be cancelled with the denominator.

Describe how left-hand and right-hand limits are used to test whether a two-sided limit exists, and explain the three conditions required for a function to be continuous at a point.

What is the difference between a left-hand limit and a right-hand limit?

The left-hand limit $\lim_{x\to c^-} f(x)$ considers only values of x approaching c from below (less than c), while the right-hand limit $\lim_{x\to c^+} f(x)$ considers only values approaching from above (greater than c).

What is the criterion for the two-sided limit $\lim_{x\to c} f(x)$ to exist?

The two-sided limit exists, and equals L, if and only if both the left-hand and right-hand limits exist and are equal to that same value L.

Why is this criterion especially important for piecewise-defined functions?

Because the formula defining f(x) changes at the boundary point c, the one-sided limits often must be computed using different pieces of the definition on each side, and a mismatch between them (a jump) is a common source of discontinuity.

What three conditions must all hold for f to be continuous at c?

$f(c)$ must be defined, $\lim_{x\to c} f(x)$ must exist, and the two must be equal ($\lim_{x\to c} f(x) = f(c)$); if any one of these three conditions fails, f is discontinuous at c.

Multiple Choice Questions (MCQs)

The phrase 'x approaches a' means:

A. x equals a

B. x gets arbitrarily close to a without equaling a

C. x is always greater than a

D. x is undefined

Answer: B. x gets arbitrarily close to a without equaling a

Explanation: x -> a means x gets arbitrarily close to a from both sides, but x never actually equals a.

lim(x->a) [f(x) + g(x)] equals:

A. lim f(x) - lim g(x)

B. lim f(x) + lim g(x)

C. lim f(x) x lim g(x)

D. Cannot be determined

Answer: B. lim f(x) + lim g(x)

Explanation: The limit of a sum equals the sum of the individual limits.

When direct substitution gives 0/0, the correct next step is to:

A. Conclude the limit is 0

B. Conclude the limit does not exist

C. Simplify algebraically (factor or rationalize) and retry

D. Conclude the limit is 1

Answer: C. Simplify algebraically (factor or rationalize) and retry

Explanation: A 0/0 result is indeterminate and requires algebraic simplification before the limit can be evaluated.

lim(n->infinity) (1 + 1/n)^n equals:

A. 1

B. 0

C. e

D. infinity

Answer: C. e

Explanation: This is the defining limit of the number e, approximately 2.718281.

The Sandwich Theorem is used to prove which important limit?

A. lim(x->0) x^2 = 0

B. lim(theta->0) sin(theta)/theta = 1

C. lim(x->infinity) 1/x = 0

D. lim(x->a) c = c

Answer: B. lim(theta->0) sin(theta)/theta = 1

Explanation: The Sandwich Theorem is used geometrically to prove lim(theta->0) sin(theta)/theta = 1.

lim(x-> +infinity) a/x^p, where p>0, equals:

A. a

B. infinity

C. 0

D. 1/a

Answer: C. 0

Explanation: As x grows without bound, a/x^p approaches 0 for any real a and positive rational p.

The two-sided limit lim(x->c) f(x) exists if and only if:

A. f(c) is defined

B. The left-hand and right-hand limits both exist and are equal

C. f is a polynomial

D. f(x) is always positive

Answer: B. The left-hand and right-hand limits both exist and are equal

Explanation: The two-sided limit exists exactly when the one-sided limits both exist and agree.

A function f is continuous at c if:

A. f(c) is defined only

B. lim(x->c) f(x) exists only

C. f(c) is defined, the limit exists, and they are equal

D. f is a straight line

Answer: C. f(c) is defined, the limit exists, and they are equal

Explanation: Continuity requires all three conditions to hold simultaneously.

A jump discontinuity occurs when:

A. f(c) is undefined

B. The left-hand and right-hand limits disagree

C. f(x) is a constant function

D. The domain of f is all real numbers

Answer: B. The left-hand and right-hand limits disagree

Explanation: A jump discontinuity happens when the one-sided limits exist but are not equal to each other.

For the radioactive decay model A(t) = A0 e^(-kt), the limit as t -> infinity is:

A. A0

B. infinity

C. 0

D. k

Answer: C. 0

Explanation: Since e^(-kt) -> 0 as t -> infinity (for k>0), the amount of substance approaches 0.

Quick Revision Summary

- x -> a means x gets arbitrarily close to a without ever equaling a
- lim(x->a) f(x) = L means f(x) approaches L as x approaches a from both sides
- The theorems on limits let sums, differences, products, quotients, and powers of functions be evaluated by direct substitution
- A 0/0 result requires algebraic simplification -- factoring or rationalizing -- before the limit can be found
- lim(x-> +-infinity) a/x^p = 0 for any real a and positive rational p
- lim(n->infinity) (1+1/n)^n = e; lim(x->0) (1+x)^(1/x) = e
- The Sandwich Theorem: if f(x)<=g(x)<=h(x) near c and f, h both approach L, then g also approaches L
- lim(theta->0) sin(theta)/theta = 1, proved geometrically via the Sandwich Theorem
- lim(x->c) f(x) exists iff the left-hand and right-hand limits both exist and are equal
- f is continuous at c iff f(c) is defined, lim(x->c) f(x) exists, and the two are equal
- A discontinuity can arise from an undefined point, a mismatched value, or a jump (unequal one-sided limits)

- Limits of exponential/transcendental functions model real-world long-term behavior: decay to 0, unbounded growth, or leveling off at a carrying capacity

Exam Tips

- Always try direct substitution first -- only switch to factoring or rationalizing if you get an indeterminate form like 0/0
- When rationalizing, multiply by the conjugate of whichever part of the expression contains the square root
- For limits at infinity in a rational function, divide every term by the highest power of x in the denominator
- Recognize the pattern (1+1/n)^n before assuming a limit needs a totally new technique -- many 'in terms of e' problems are just algebraic rewrites of this one pattern
- For piecewise functions, always compute the left-hand and right-hand limits separately using the matching piece of the definition on each side
- Remember continuity needs all three conditions -- a function can have a limit at a point without being continuous there if f(c) is undefined or does not match the limit

Unit 13: Differentiation

This unit introduces differentiation, the process at the heart of differential calculus, developed independently by Newton and Leibniz to study instantaneous rates of change. It begins geometrically -- the tangent line to a curve at a point is defined as the limiting position of a secant line as a second point slides toward the first -- and this same limiting idea gives the derivative, the instantaneous rate of change of a function, whether that rate describes a slope, a velocity, or any other quantity that varies continuously.

The unit then develops the formal definition of the derivative (differentiation 'by first principles' or 'ab-initio'), proves the power rule and the key theorems on differentiation (constant multiple, sum/difference, product, and quotient rules), and examines the important connection between differentiability and continuity. It closes with real-world applications of derivatives in business, physics, and engineering -- marginal profit, velocity and acceleration, investment growth, and structural stress.

Learning Objectives

- Explain the tangent line to a curve as the limiting position of a secant line
- Define the derivative of a function as the limit of a difference quotient
- Find derivatives of functions by definition (first principles / ab-initio method)
- State and apply the power rule for differentiation, including for rational exponents
- Explain the connection between differentiability and continuity, including cases where a function is continuous but not differentiable
- State and apply the constant multiple, sum/difference, product, and quotient rules of differentiation
- Apply differentiation to real-world problems involving velocity, acceleration, marginal profit, and rates of change

Key Concepts

13.1 Tangent to a Curve and the Derivative as a Limit

For two points P(x, f(x)) and Q(x+delta x, f(x+delta x)) on a curve y=f(x), the secant line through P and Q has slope [f(x+delta x)-f(x)]/delta x. As Q slides along the curve toward P (delta x -> 0), the secant line rotates into the tangent line at P, and its slope approaches the limit m_tan = lim(delta x->0) [f(x+delta x)-f(x)]/delta x.

This limiting slope is exactly the derivative of f at x. The same idea -- an average rate of change over a shrinking interval settling down to an instantaneous rate -- appears throughout calculus, whether the 'curve' represents a graph, a position-versus-time relationship, or any other continuously varying quantity.

13.2 The Derivative as Instantaneous Rate of Change

For a quantity changing over an interval, the average rate of change is the difference quotient [f(x1)-f(x)]/(x1-x); as x1 approaches x, this ratio approaches the instantaneous rate of change, written f'(x) -- the derivative of f at x. The process of finding f' is called differentiation, and f is differentiable at x whenever this limit exists.

A common physical interpretation is velocity: if s(t) gives a particle's position at time t, the average velocity over [t, t1] is [s(t1)-s(t)]/(t1-t), and as t1 -> t, this approaches the instantaneous velocity at t -- the derivative of the position function with respect to time.

13.3 Finding the Derivative by Definition and Notation

The derivative f'(x) can be found from its limit definition using four steps: find f(x+delta x); simplify f(x+delta x)-f(x); divide by delta x and simplify; then take the limit as delta x -> 0. This process is called differentiation by definition, by first principles, or ab-initio.

Several equivalent notations for the derivative of y=f(x) are in common use: Leibniz's dy/dx, Newton's f'(x) or y', Lagrange's f'(x), and Euler's Df(x); the value of the derivative at a specific point a is written f'(a) or dy/dx evaluated at x=a.

13.4 The Power Rule and Connection to Continuity

The power rule states d/dx(x^n) = n x^(n-1) for any real number n -- proved for positive and negative integers using the binomial theorem, and holding also for rational exponents. This single rule, combined with the constant rule d/dx(c)=0, handles the derivative of any power of x directly, without needing to return to first principles each time.

A key theorem connects the two central ideas of this and the previous unit: if a function is differentiable at a point, it must also be continuous there -- but the converse is false. The function f(x)=|x| is continuous at x=0 but not differentiable there, because its left-hand derivative (-1) and right-hand derivative (+1) disagree, leaving no single, well-defined tangent line at that sharp corner.

13.5 Theorems on Differentiation

Beyond the constant and power rules, four further theorems make differentiation of combined functions efficient: the constant multiple rule d/dx[cf(x)]=c f'(x); the sum/difference rule d/dx[f(x)+-g(x)]=f'(x)+-g'(x); the product rule d/dx[f(x)g(x)]=f'(x)g(x)+f(x)g'(x); and the quotient rule d/dx[f(x)/g(x)]=[f'(x)g(x)-f(x)g'(x)]/[g(x)]^2 (for g(x) not 0).

Each of these theorems is proved directly from the limit definition of the derivative, using the same difference-quotient technique as first principles -- but once proved, they let derivatives of complicated combinations of functions be found quickly, without repeating the limit process every time.

13.6 Applications of Differentiation

Because the derivative measures an instantaneous rate of change, it applies wherever a real-world quantity varies continuously: marginal profit (the derivative of a profit function with respect to units produced), velocity and acceleration (the first and second derivatives of a position function with respect to time), the rate of growth of an investment, and the rate of change of structural stress along a beam.

In every such application, the same pattern applies: model the quantity as a function of the relevant variable, differentiate using the rules of this unit, then substitute the specific value of interest to find the instantaneous rate at that point.

Important Definitions

What is the slope of the secant line through P(x,f(x)) and Q(x+delta x, f(x+delta x))?

[f(x+delta x) - f(x)] / delta x.

What is the derivative of a function f at x?

f'(x) = lim(delta x->0) [f(x+delta x) - f(x)] / delta x, provided the limit exists.

What does it mean for f to be differentiable at x?

The limit defining f'(x) exists at that point.

What is differentiation?

The process of finding the derivative of a function.

What is meant by 'differentiation by first principles' or 'ab-initio'?

Finding a derivative directly from its limit definition, using the four-step process, rather than by applying a shortcut rule.

What does the power rule state?

d/dx(x^n) = n x^(n-1), for any real number n.

What is the key relationship between differentiability and continuity?

If a function is differentiable at a point, it must be continuous there; however, a function can be continuous at a point without being differentiable there.

What does the product rule state?

d/dx[f(x)g(x)] = f'(x)g(x) + f(x)g'(x).

What does the quotient rule state?

d/dx[f(x)/g(x)] = [f'(x)g(x) - f(x)g'(x)] / [g(x)]^2, for g(x) not equal to 0.

What is instantaneous velocity?

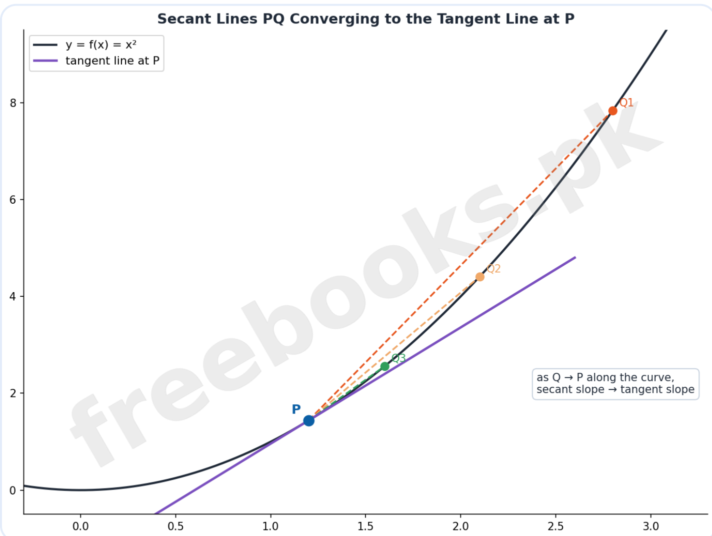
The derivative of the position function s(t) with respect to time t, i.e. the limit of average velocity as the time interval shrinks to zero.

Key Facts and Relations

Topic	Key Fact / Relation
Derivative (limit definition)	f'(x) = lim(delta x->0) [f(x+delta x) - f(x)] / delta x
Alternative form at x=a	f'(a) = lim(x->a) [f(x) - f(a)] / (x-a)
Constant rule	d/dx (c) = 0
Power rule	d/dx (x^n) = n x^(n-1)
Constant multiple rule	d/dx [c f(x)] = c f'(x)
Sum / difference rule	d/dx [f(x) +- g(x)] = f'(x) +- g'(x)
Product rule	d/dx [f(x) g(x)] = f'(x) g(x) + f(x) g'(x)
Quotient rule	d/dx [f(x)/g(x)] = [f'(x)g(x) - f(x)g'(x)] / [g(x)]^2
Instantaneous velocity	v(t) = ds/dt; acceleration a(t) = dv/dt

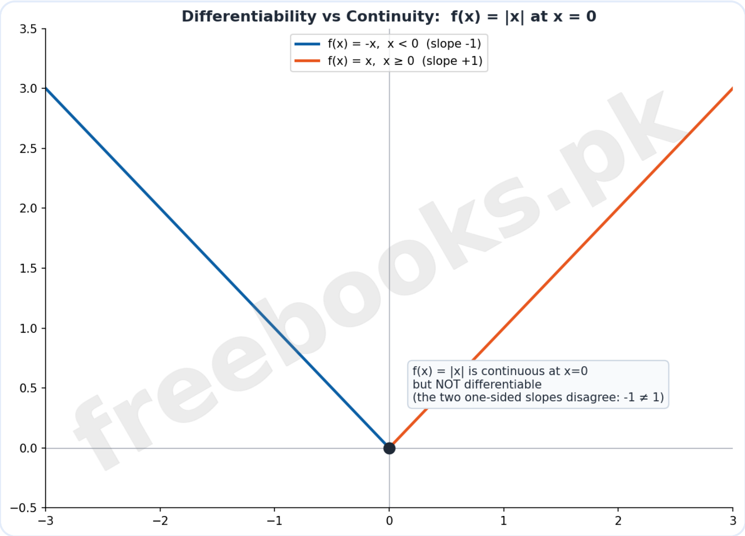
Diagrams

Secant Lines Converging to the Tangent Line: A curve y=f(x) with a fixed point P and several secant lines PQ shown converging to the tangent line at P as Q approaches P



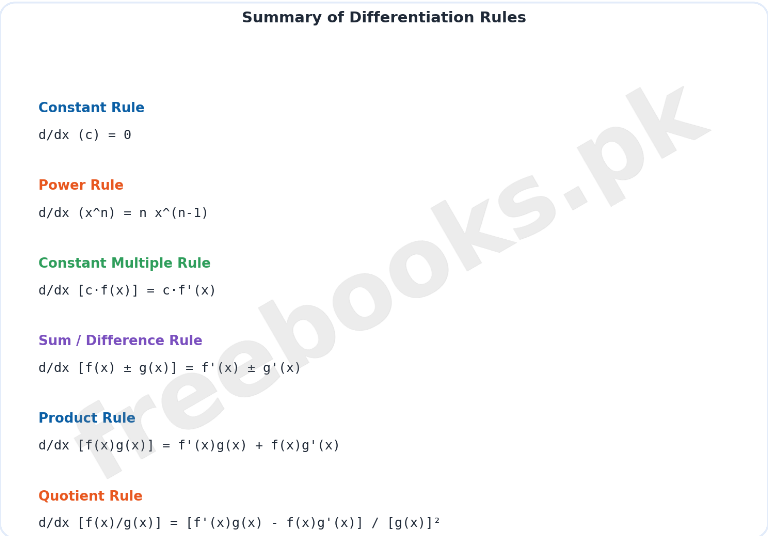
Secant Lines Converging to the Tangent Line

Differentiability vs Continuity: f(x) = |x|: A graph of f(x)=|x| showing the sharp corner at x=0, where the function is continuous but the left-hand and right-hand slopes disagree, so it is not differentiable



Differentiability vs Continuity: f(x) = |x|

Summary of Differentiation Rules: A reference table listing the constant, power, constant multiple, sum/difference, product, and quotient rules of differentiation


Summary of Differentiation Rules

Constant Rule
d/dx (c) = 0

Power Rule
d/dx (x^n) = n x^(n-1)

Constant Multiple Rule
d/dx [c·f(x)] = c·f'(x)

Sum / Difference Rule
d/dx [f(x) ± g(x)] = f'(x) ± g'(x)

Product Rule
d/dx [f(x)g(x)] = f'(x)g(x) + f(x)g'(x)

Quotient Rule
d/dx [f(x)/g(x)] = [f'(x)g(x) - f(x)g'(x)] / [g(x)]²

Summary of Differentiation Rules

Solved Examples

Example 1: Finding the Gradient and Tangent Line by Definition

Problem: Find the gradient and an equation of the tangent line to $f(x)=x^2-2$ at the point $P(-1,-1)$.

1. Apply the limit definition at $x=-1$: $m_{tan} = \lim_{\Delta x \to 0} [f(-1+\Delta x)-f(-1)]/\Delta x$.
2. Expand: $f(-1+\Delta x) = (-1+\Delta x)^2-2 = 1-2\Delta x+(\Delta x)^2-2$, and $f(-1) = 1-2 = -1$.
3. Simplify the difference quotient: $[1-2\Delta x+(\Delta x)^2-2-(-1)]/\Delta x = (-2\Delta x+(\Delta x)^2)/\Delta x = -2+\Delta x$.
4. Take the limit as $\Delta x \to 0$: $m_{tan} = -2$.
5. Using point-slope form with slope -2 and point $(-1,-1)$: $y+1 = -2(x+1)$, giving the final answer $y = -2x-3$.

Example 2: Average and Instantaneous Velocity

Problem: A particle's position is $s(t) = 4t^2+2t+1$ (miles). Find (a) the average velocity over $[2,5]$ and (b) the instantaneous velocity at $t=3$.

1. (a) Average velocity = $[s(5)-s(2)]/(5-2)$. Compute $s(5)=4(25)+10+1=111$ and $s(2)=4(4)+4+1=21$.
2. Average velocity = $(111-21)/3 = 90/3 = 30$ miles/hour.
3. (b) Instantaneous velocity = $\lim_{\Delta t \to 0} [s(3+\Delta t)-s(3)]/\Delta t$. Expand $s(3+\Delta t) = 4(3+\Delta t)^2+2(3+\Delta t)+1 = 36+24\Delta t+4(\Delta t)^2+6+2\Delta t+1$.
4. Simplify the difference quotient: $[43+26\Delta t+4(\Delta t)^2-43]/\Delta t = 26+4\Delta t$.
5. Take the limit as $\Delta t \to 0$: instantaneous velocity = 26 miles/hour.

Example 3: Derivative of sqrt(x) by First Principles

Problem: Find the derivative of f(x) = sqrt(x) at x=a from first principles.

1. Form the difference f(x+delta x)-f(x) = sqrt(x+delta x)-sqrt(x), and rationalize by multiplying by the conjugate sqrt(x+delta x)+sqrt(x).
2. This gives f(x+delta x)-f(x) = delta x / [sqrt(x+delta x)+sqrt(x)].
3. Divide by delta x: [f(x+delta x)-f(x)]/delta x = 1/[sqrt(x+delta x)+sqrt(x)].
4. Take the limit as delta x -> 0: f'(x) = 1/[sqrt(x)+sqrt(x)] = 1/(2 sqrt(x)).
5. Final answer: at x=a, f'(a) = 1/(2 sqrt(a)).

Example 4: Derivative of 1/x^2 by Ab-Initio Method

Problem: If y = 1/x^2, find dy/dx at x=-1 by the ab-initio method.

1. Form delta y = 1/(x+delta x)^2 - 1/x^2 = [x^2-(x+delta x)^2] / [x^2(x+delta x)^2].
2. Factor the numerator as a difference of squares: [x+(x+delta x)][x-(x+delta x)] = (2x+delta x)(-delta x).
3. Divide by delta x: delta y/delta x = -(2x+delta x) / [x^2(x+delta x)^2].
4. Take the limit as delta x -> 0: dy/dx = -2x / (x^2 . x^2) = -2/x^3.
5. Substitute x=-1: dy/dx = -2/(-1)^3 = -2/-1 = 2. Final answer: the gradient at x=-1 is 2.

Example 5: Applying the Power Rule and Sum Rule

Problem: Find the derivative of y = (3/4)x^4 + (2/3)x^3 + (1/2)x^2 + 2x + 5 with respect to x.

1. Apply the sum rule, differentiating each term separately.
2. Differentiate each power term using the power rule: d/dx[(3/4)x^4] = (3/4)(4x^3) = 3x^3; d/dx[(2/3)x^3] = (2/3)(3x^2) = 2x^2.

3. Continue: d/dx[(1/2)x^2] = (1/2)(2x) = x; d/dx[2x] = 2; d/dx[5] = 0 (constant rule).
4. Add all the terms together: 3x^3 + 2x^2 + x + 2.
5. Final answer: dy/dx = 3x^3 + 2x^2 + x + 2.

Example 6: Applying the Product Rule

Problem: Find the derivative of y = (x^2+5)(x^3+7) with respect to x, using the product rule.

1. Let f(x) = x^2+5 and g(x) = x^3+7, so f'(x) = 2x and g'(x) = 3x^2.
2. Apply the product rule: dy/dx = f'(x)g(x) + f(x)g'(x) = 2x(x^3+7) + (x^2+5)(3x^2).
3. Expand each term: 2x(x^3+7) = 2x^4+14x; (x^2+5)(3x^2) = 3x^4+15x^2.
4. Combine like terms: 2x^4+14x+3x^4+15x^2 = 5x^4+15x^2+14x.
5. Final answer: dy/dx = 5x^4 + 15x^2 + 14x.

Example 7: Applying the Quotient Rule

Problem: Differentiate phi(x) = (2x^3-3x^2+5)/(x^2+1) with respect to x, using the quotient rule.

1. Let f(x) = 2x^3-3x^2+5 and g(x) = x^2+1, so f'(x) = 6x^2-6x and g'(x) = 2x.
2. Apply the quotient rule: phi'(x) = [f'(x)g(x) - f(x)g'(x)] / [g(x)]^2.
3. Expand the numerator: (6x^2-6x)(x^2+1) - (2x^3-3x^2+5)(2x) = (6x^4-6x^3+6x^2-6x) - (4x^4-6x^3+10x).
4. Simplify: 6x^4-6x^3+6x^2-6x-4x^4+6x^3-10x = 2x^4+6x^2-16x.
5. Final answer: phi'(x) = (2x^4+6x^2-16x) / (x^2+1)^2.

Example 8: Real-World Application: Velocity and Acceleration of a Particle

> **Problem:** A particle's position is s(t) = 4t^3-3t^2+2t (metres). Find the velocity and acceleration at t=2 seconds.
>
> 1. Velocity is the derivative of position: v(t) = d/dt(4t^3-3t^2+2t) = 12t^2-6t+2.
> 2. Substitute t=2: v(2) = 12(4)-6(2)+2 = 48-12+2 = 38 m/s.
> 3. Acceleration is the derivative of velocity: a(t) = d/dt(12t^2-6t+2) = 24t-6.
> 4. Substitute t=2: a(2) = 24(2)-6 = 48-6 = 42 m/s^2.
> 5. Final answer: velocity at t=2 is 38 m/s; acceleration at t=2 is 42 m/s^2.

Short Questions & Answers

What is the slope of a tangent line defined as?

The limit of the slope of the secant line as the second point approaches the point of tangency.

What is the derivative of a constant function?

Zero: d/dx(c) = 0.

What does the power rule state?

d/dx(x^n) = n x^(n-1), for any real number n.

Is every continuous function differentiable?

No -- a function can be continuous at a point without being differentiable there, as with f(x)=|x| at x=0.

Is every differentiable function continuous?

Yes -- differentiability at a point always implies continuity at that point.

What is the product rule?

d/dx[f(x)g(x)] = f'(x)g(x) + f(x)g'(x).

What is acceleration in terms of derivatives?

The derivative of velocity with respect to time (the second derivative of position).

Long Questions & Answers

Explain how the tangent line to a curve is defined using secant lines and limits, and describe the four-step process for finding a derivative by definition.

What is a secant line, and what does its slope represent?

A secant line is a line joining two points P(x,f(x)) and Q(x+delta x, f(x+delta x)) on a curve; its slope, [f(x+delta x)-f(x)]/delta x, represents the average rate of change of f between those two points.

How does the tangent line arise from secant lines?

As the point Q slides along the curve toward P (i.e. as delta x -> 0), the secant line PQ rotates continuously until it becomes the tangent line at P; the tangent's slope is the limit of the secant's slope as delta x -> 0.

What are the four steps for finding a derivative by definition?

Step I: find f(x+delta x). Step II: simplify f(x+delta x)-f(x). Step III: divide by delta x and simplify. Step IV: take the limit of the resulting expression as delta x -> 0.

Why is this method called 'first principles' or 'ab-initio'?

Because it derives the result directly from the fundamental limit definition of the derivative, without relying on any shortcut differentiation rule -- 'ab-initio' is Latin for 'from the beginning'.

Describe the theorems on differentiation (constant multiple, sum/difference, product, and quotient rules), and explain the relationship between differentiability and continuity.

What do the constant multiple and sum/difference rules state?

The constant multiple rule states d/dx[c f(x)] = c f'(x); the sum/difference rule states d/dx[f(x)+-g(x)] = f'(x)+-g'(x) -- both let derivatives of scaled or combined functions be found term by term.

What does the product rule state, and why is it not simply f'(x)g'(x)?

The product rule states d/dx[f(x)g(x)] = f'(x)g(x)+f(x)g'(x); this is proved by adding and subtracting a cross term in the difference quotient, and it is more complex than simply multiplying the two derivatives because both functions are changing simultaneously.

What does the quotient rule state?

d/dx[f(x)/g(x)] = [f'(x)g(x)-f(x)g'(x)] / [g(x)]^2, valid wherever g(x) is not zero; like the product rule, it accounts for the way both the numerator and denominator change together.

What is the relationship between differentiability and continuity?

Differentiability at a point implies continuity at that point, but the converse does not hold -- a function such as f(x)=|x| is continuous everywhere, including at x=0, yet is not differentiable at x=0 because its left-hand slope (-1) and right-hand slope (+1) disagree there.

Multiple Choice Questions (MCQs)

The slope of the tangent line at P is defined as:

A. The slope of any line through P

B. The limit of the secant slope as Q approaches P

C. Always equal to 1

D. The y-coordinate of P

Answer: B. The limit of the secant slope as Q approaches P

Explanation: The tangent slope is the limit of the secant slope PQ as Q approaches P along the curve.

The derivative of a constant function c is:

A. c

B. 1

C. 0

D. x

Answer: C. 0

Explanation: d/dx(c) = 0, since a constant function never changes.

The power rule states d/dx(x^n) equals:

A. x^n

B. n x^(n-1)

C. n x^n

D. x^(n-1)

Answer: B. n x^(n-1)

Explanation: The power rule gives d/dx(x^n) = n x^(n-1) for any real number n.

If a function is differentiable at a point, then it is:

A. Not necessarily continuous there

B. Always continuous there

C. Always discontinuous there

D. Undefined there

Answer: B. Always continuous there

Explanation: Differentiability at a point always implies continuity at that point.

The function f(x) = |x| at x = 0 is:

A. Differentiable but not continuous

B. Continuous but not differentiable

C. Neither continuous nor differentiable

D. Both continuous and differentiable

Answer: B. Continuous but not differentiable

Explanation: f(x)=|x| is continuous at x=0, but its left-hand and right-hand derivatives disagree (-1 and +1), so it is not differentiable there.

The product rule states d/dx[f(x)g(x)] equals:

A. f'(x)g'(x)

B. f'(x)g(x) + f(x)g'(x)

C. f'(x)g(x) - f(x)g'(x)

D. f(x)g(x)

Answer: B. f'(x)g(x) + f(x)g'(x)

Explanation: The product rule is f'(x)g(x) + f(x)g'(x), not simply the product of the two derivatives.

The quotient rule for d/dx[f(x)/g(x)] requires:

A. f(x) not equal to 0

B. g(x) not equal to 0

C. f(x) = g(x)

D. No conditions

Answer: B. g(x) not equal to 0

Explanation: The quotient rule requires g(x) to be non-zero, since division by zero is undefined.

Instantaneous velocity is defined as:

A. The average velocity over a large interval

B. The derivative of the position function with respect to time

C. A constant value

D. The derivative of velocity with respect to time

Answer: B. The derivative of the position function with respect to time

Explanation: Instantaneous velocity is the derivative of position with respect to time, v(t) = ds/dt.

Acceleration is the derivative of:

A. Position with respect to time

B. Velocity with respect to time

C. Time with respect to velocity

D. A constant

Answer: B. Velocity with respect to time

Explanation: Acceleration is the derivative of velocity with respect to time, i.e. the second derivative of position.

Differentiation 'by first principles' means finding the derivative using:

A. The power rule only

B. A table of known derivatives

C. The direct limit definition of the derivative

D. Guesswork

Answer: C. The direct limit definition of the derivative

Explanation: First principles (ab-initio) means computing the derivative directly from its limit definition, without using shortcut rules.

Quick Revision Summary

- The tangent slope at P is the limit of the secant slope PQ as Q approaches P
- f'(x) = lim(delta x->0) [f(x+delta x)-f(x)]/delta x is the derivative of f at x

- Differentiation by first principles (ab-initio): find f(x+delta x), simplify the difference, divide by delta x, take the limit
- Common notations for the derivative: dy/dx (Leibniz), f'(x) or y' (Newton/Lagrange), Df(x) (Euler)
- Power rule: d/dx(x^n) = n x^(n-1), valid for any real number n
- Differentiability implies continuity, but continuity does not imply differentiability (e.g. f(x)=|x| at x=0)
- Constant multiple rule: d/dx[c f(x)] = c f'(x)
- Sum/difference rule: d/dx[f(x)+-g(x)] = f'(x)+-g'(x)
- Product rule: d/dx[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)
- Quotient rule: d/dx[f(x)/g(x)] = [f'(x)g(x)-f(x)g'(x)] / [g(x)]^2
- Velocity is the derivative of position; acceleration is the derivative of velocity
- Derivatives model real-world rates of change: marginal profit, investment growth, structural stress, and more

Exam Tips

- When finding a derivative by first principles, always simplify f(x+delta x)-f(x) fully before dividing by delta x -- factor out delta x wherever possible
- For square-root expressions in first-principles problems, rationalize the numerator using the conjugate before taking the limit
- Apply the power rule directly for any term of the form x^n -- there is no need to use first principles once the rule is known
- Remember 'differentiable implies continuous' only works in one direction -- always check both one-sided derivatives at a sharp corner or cusp before assuming differentiability
- In product and quotient rule problems, write out f(x), g(x), f'(x), and g'(x) separately first, then substitute into the rule to avoid sign errors
- For real-world applications, always identify what the derivative represents (velocity, marginal profit, rate of stress, etc.) before differentiating

Unit 14: Vectors in Space

This final unit extends vector ideas from the plane into three-dimensional space. Using the rectangular coordinate system in space and the right-hand rule, a point P(a,b,c) is described by its position vector OP, and vectors are written in component form using the standard unit vectors i, j, and k. From here the unit develops the magnitude of a vector, unit vectors, the distance between two points in space, and direction angles and direction cosines, culminating in the identity cos^2(alpha)+cos^2(beta)+cos^2(gamma)=1.

The unit then introduces the two major vector products. The dot (scalar) product u.v = |u||v| cos(theta), computed equally from components as u1 v1+u2 v2+u3 v3, gives projections, angles between vectors, orthogonality tests, and work done by a constant force. The cross (vector) product u x v, a vector perpendicular to both u and v found via the determinant formula, gives the area of a parallelogram or triangle and the moment (torque) of a force. Finally, the scalar triple product u.(v x w) combines both products to compute the volume of a parallelepiped or tetrahedron and to test whether three vectors are coplanar -- with real-world applications spanning engineering, physics, and business.

Learning Objectives

- Describe the rectangular coordinate system in space and locate a point using the right-hand rule
- Write a vector in space in component form using the unit vectors i, j, and k
- Find the magnitude of a vector in space and construct a unit vector in a given direction
- Find the distance between two points in space and the direction angles/ direction cosines of a vector
- Compute the dot product of two vectors both geometrically and from components, and use it to find projections, angles, and orthogonality
- Apply the dot product to compute the work done by a constant force
- Compute the cross product of two vectors using the determinant formula and use it to find areas and the moment of a force

- Compute the scalar triple product of three vectors and use it to find volumes and test for coplanarity

Key Concepts

14.1 Rectangular Coordinate System in Space and Vectors

Space is described using three mutually perpendicular axes (x, y, z) meeting at the origin O, oriented according to the right-hand rule: if the fingers of the right hand curl from the positive x-axis toward the positive y-axis, the thumb points along the positive z-axis. Any point P in space is located by an ordered triple (a,b,c), and the position vector OP = [a,b,c] is the vector from the origin to P.

A vector in space is written in component form as u = [u1,u2,u3] or, using the standard unit vectors i=[1,0,0], j=[0,1,0], k=[0,0,1], as u = u1 i + u2 j + u3 k. The fundamental operations -- addition, scalar multiplication, the negative of a vector, the difference of two vectors, and the zero vector -- extend from the plane to space by simply applying the operation to each of the three components in turn.

14.2 Magnitude, Unit Vectors, and Distance in Space

The magnitude (length) of u = [u1,u2,u3] is |u| = sqrt(u1^2+u2^2+u3^2), a direct extension of the Pythagorean theorem into three dimensions. A unit vector has magnitude 1; the unit vector in the direction of any nonzero vector u is found by scaling: u-hat = u/|u|. Two vectors are parallel exactly when one is a nonzero scalar multiple of the other.

The distance between two points P1(x1,y1,z1) and P2(x2,y2,z2) in space is the magnitude of the vector between them: d = sqrt[(x2-x1)^2+(y2-y1)^2+(z2-z1)^2]. This same formula underlies problems phrased as physical displacement -- the straight-line distance between a start point and an end point in three-dimensional space.

14.3 Direction Angles and Direction Cosines

For a nonzero vector u, the direction angles alpha, beta, and gamma are the angles u makes with the positive x-, y-, and z-axes respectively, and the direction cosines are $\cos(\alpha)$, $\cos(\beta)$, and $\cos(\gamma)$ -- each equal to the corresponding component of u divided by $|u|$. These three cosines always satisfy the identity $\cos^2(\alpha)+\cos^2(\beta)+\cos^2(\gamma) = 1$, which follows directly from the magnitude formula.

Direction cosines give a compact, coordinate-independent way to describe the orientation of a vector in space, and the unit vector in the direction of u can always be written as $\hat{u} = [\cos(\alpha), \cos(\beta), \cos(\gamma)]$.

14.4 The Dot (Scalar) Product

The dot product of two nonzero vectors u and v is defined geometrically as $u.v = |u||v| \cos(\theta)$, where theta is the angle between them, and equivalently in component form as $u.v = u_1 v_1+u_2 v_2+u_3 v_3$ -- the two definitions can be shown equivalent using the Law of Cosines. Because the dot product of two vectors is always a scalar (not a vector), it is also called the scalar product.

The dot product gives the projection of one vector along another: the (scalar) projection of v along u is $(u.v)/|u|$. Two nonzero vectors are orthogonal (perpendicular) exactly when their dot product is zero, and the angle between any two nonzero vectors can be recovered as $\theta = \cos^{-1}[(u.v)/(|u||v|)]$. A key physical application is the work done by a constant force F moving an object through a displacement d: $W = F.d$.

14.5 The Cross (Vector) Product

The cross product of two nonzero, non-parallel vectors u and v is the vector $u \times v = (|u||v| \sin(\theta))\, n$, where n is the unit vector perpendicular to both u and v with direction fixed by the right-hand rule; because the result is itself a vector, it is also called the vector product. In component form, $u \times v$ is computed as a 3x3 determinant with i, j, k in the first row, the components of u in the second row, and the components of v in the third row. Two nonzero vectors are parallel exactly when their cross product is the zero vector.

The magnitude $|u \times v|$ equals $|u||v| \sin(\theta)$, which is exactly the area of the parallelogram with u and v as adjacent sides; half of this, $(1/2)|u \times v|$, gives the area of the triangle with u and v as two sides. In mechanics, the moment (torque) of a force F applied at a point with position vector r (relative to the pivot) is the vector M = r x F, whose magnitude measures the turning effect of the force.

14.6 The Scalar Triple Product: Volume and Coplanarity

The scalar triple product of three vectors u, v, w is the scalar u.(v x w), often written [u v w], and computed directly as a 3x3 determinant of the components of u, v, w. It is cyclic -- u.(v x w) = v.(w x u) = w.(u x v) -- and its absolute value gives the volume of the parallelepiped with u, v, w as edges from a common vertex; one-sixth of that same absolute value gives the volume of the tetrahedron with the same three edges.

Three vectors are coplanar (lie in the same plane, or are linearly dependent) exactly when their scalar triple product is zero -- a direct volume-based test, since three coplanar edges enclose zero volume. This same combination of dot and cross product underlies real-world torque, force-magnitude, and even revenue calculations, wherever a single scalar summary of several vector quantities is needed.

Important Definitions

What is the position vector of a point P(a,b,c) in space?

The vector OP = [a,b,c] from the origin O to the point P.

What is a unit vector, and how is it constructed from a nonzero vector u?

A vector of magnitude 1; the unit vector in the direction of u is u-hat = u/|u|.

What is the magnitude of u = [u1,u2,u3]?

|u| = sqrt(u1^2+u2^2+u3^2).

What are the direction cosines of a vector u?

The cosines of the angles u makes with the positive x-, y-, and z-axes, equal to the corresponding components of u divided by |u|; they satisfy cos^2(alpha) +cos^2(beta)+cos^2(gamma)=1.

How is the dot product of two vectors defined?

u.v = |u||v| cos(theta) geometrically, or u1 v1+u2 v2+u3 v3 in component form; the result is always a scalar.

When are two vectors orthogonal?

When their dot product is zero, i.e. u.v=0.

What is the scalar projection of v along u?

(u.v)/|u|.

How is the cross product of two vectors defined?

u x v = (|u||v| sin(theta)) n, a vector perpendicular to both u and v with direction given by the right-hand rule; computed from components using a 3x3 determinant.

What does the scalar triple product u.(v x w) measure?

Its absolute value equals the volume of the parallelepiped with u, v, w as edges from a common vertex; it equals zero exactly when u, v, w are coplanar.

What is the moment (torque) of a force F applied at position r relative to a pivot?

M = r x F, a vector whose magnitude measures the turning effect of the force about the pivot.

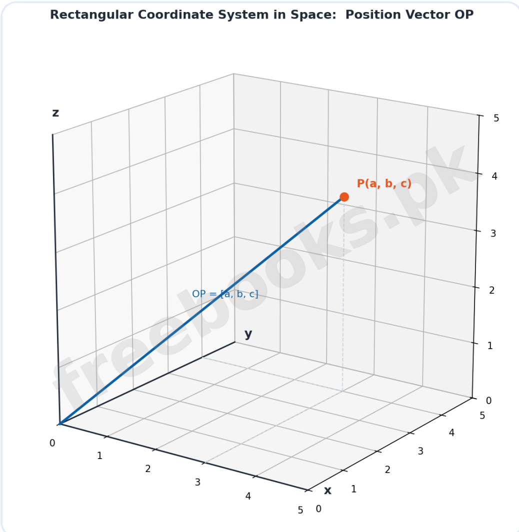
Key Facts and Relations

Topic	Key Fact / Relation
Magnitude of a vector in space	|u| = sqrt(u1^2+u2^2+u3^2)
Unit vector in the direction of u	u-hat = u / |u|
Distance between two points in space	d = sqrt[(x2-x1)^2+(y2-y1)^2+(z2-z1)^2]
Direction cosine identity	cos^2(alpha) + cos^2(beta) + cos^2(gamma) = 1
Dot product (geometric and component form)	u.v = |u||v| cos(theta) = u1 v1 + u2 v2 + u3 v3
Angle between two vectors	theta = cos^-1[(u.v) / (|u||v|)]
Work done by a constant force	W = F . d

Topic	Key Fact / Relation
Cross product magnitude / area of parallelogram	|u x v| = |u||v| sin(theta) = Area of parallelogram with sides u, v
Moment of a force	M = r x F
Scalar triple product / volume of parallelepiped	u.(v x w) = [u v w]; Volume = |u.(v x w)|; Volume of tetrahedron = (1/6)|u.(v x w)|

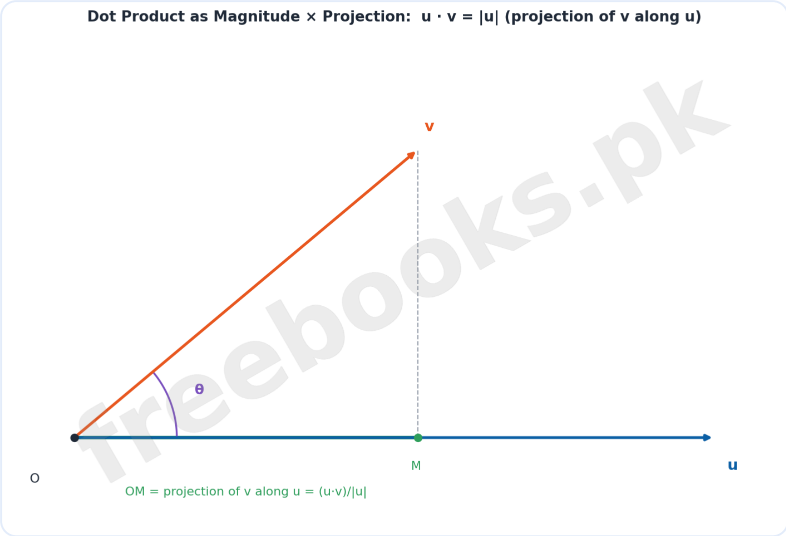
Diagrams

Rectangular Coordinate System in Space: Position Vector OP: A 3D diagram of the x-, y-, z-axes meeting at the origin, with a point P(a,b,c) and its position vector OP shown, illustrating the right-hand rule orientation



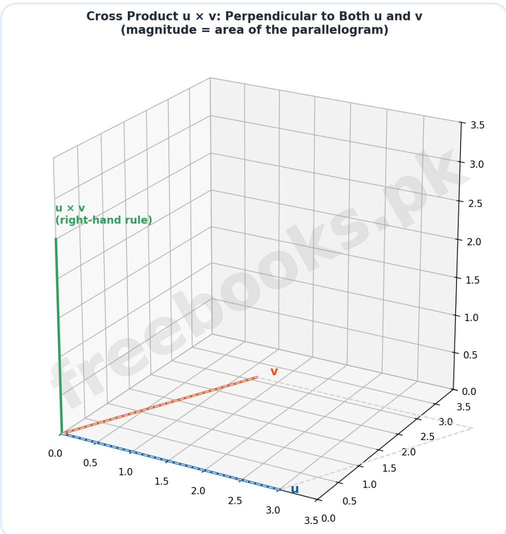
Rectangular Coordinate System in Space: Position Vector OP

Dot Product as Magnitude times Projection: A 2D diagram showing two vectors u and v with the angle theta between them, and the projection of v along u (segment OM) shaded, illustrating u.v = |u| times the projection of v along u



Dot Product as Magnitude times Projection

Cross Product and the Right-Hand Rule: A 3D diagram showing vectors u and v with their cross product u x v drawn perpendicular to both, following the right-hand rule, with the parallelogram spanned by u and v outlined to show that |u x v| is its area



Cross Product and the Right-Hand Rule

Solved Examples

Example 1: Vector Arithmetic in Space

Problem: If u = [2,-1,3] and w = [1,4,-2], find (a) u+w, (b) 3w, and (c) |u|.

1. (a) Add corresponding components: u+w = [2+1, -1+4, 3+(-2)] = [3,3,1].

2. (b) Multiply every component of w by the scalar 3: 3w = [3(1), 3(4), 3(-2)] = [3,12,-6].
3. (c) Apply the magnitude formula: |u| = sqrt(2^2+(-1)^2+3^2) = sqrt(4+1+9) = sqrt(14).
4. Final answers: u+w = [3,3,1]; 3w = [3,12,-6]; |u| = sqrt(14).

Example 2: Finding a Unit Vector

Problem: Find the unit vector in the direction of v = [3,-4,12].

1. Compute the magnitude: |v| = sqrt(3^2+(-4)^2+12^2) = sqrt(9+16+144) = sqrt(169) = 13.
2. Divide each component of v by |v|: v-hat = v/|v| = [3/13, -4/13, 12/13].
3. Check the result has magnitude 1: (3/13)^2+(-4/13)^2+(12/13)^2 = (9+16+144)/169 = 169/169 = 1. Correct.
4. Final answer: v-hat = [3/13, -4/13, 12/13].

Example 3: Distance Between Two Points in Space

Problem: A drone flies in a straight line from point A(1,2,5) to point B(7,-2,9), measured in kilometres. Find the straight-line distance it travels.

1. Apply the distance formula with (x1,y1,z1)=(1,2,5) and (x2,y2,z2)=(7,-2,9): d = sqrt[(7-1)^2+(-2-2)^2+(9-5)^2].
2. Compute each square: (7-1)^2=36, (-2-2)^2=16, (9-5)^2=16.
3. Add and take the square root: d = sqrt(36+16+16) = sqrt(68) = 2 sqrt(17).
4. Final answer: the drone travels 2 sqrt(17) is approximately 8.25 kilometres in a straight line.

Example 4: Dot Product and the Angle Between Two Vectors

Problem: Find the angle between u = [1,1,0] and v = [0,1,1].

1. Compute the dot product: u.v = (1)(0)+(1)(1)+(0)(1) = 0+1+0 = 1.
2. Compute the magnitudes: |u| = sqrt(1+1+0) = sqrt(2); |v| = sqrt(0+1+1) = sqrt(2).
3. Apply theta = cos^-1[(u.v)/(|u||v|)] = cos^-1[1/(sqrt(2).sqrt(2))] = cos^-1(1/2).
4. Solve: cos^-1(1/2) = 60 degrees.
5. Final answer: the angle between u and v is 60 degrees.

Example 5: Work Done by a Constant Force

Problem: A constant force F = [4,3,0] Newtons moves an object along a displacement d = [5,2,0] metres. Find the work done.

1. Work done by a constant force is the dot product of force and displacement: W = F.d.
2. Compute the dot product: F.d = (4)(5)+(3)(2)+(0)(0) = 20+6+0 = 26.
3. Units: since F is in Newtons and d is in metres, W is in Newton-metres (Joules).
4. Final answer: W = 26 Joules.

Example 6: Cross Product and Area of a Triangle

Problem: Find the area of the triangle with vertices A(0,0,0), B(2,0,0), and C(0,3,0), using u=AB and v=AC.

1. Form the vectors: u = AB = [2,0,0] and v = AC = [0,3,0].
2. Compute the cross product using the determinant formula: u x v = [(0)(0)-(0)(3), (0)(0)-(2)(0), (2)(3)-(0)(0)] = [0,0,6].
3. Find the magnitude: |u x v| = sqrt(0^2+0^2+6^2) = 6.

4. Area of triangle = (1/2)|u x v| = (1/2)(6) = 3.
5. Final answer: the area of triangle ABC is 3 square units.

Example 7: Moment of a Force (Torque)

Problem: A wrench applies a force F = [0,-20,0] Newtons at a point with position vector r = [0.3,0,0] metres from the bolt (the pivot). Find the moment of the force about the bolt.

1. The moment of a force is M = r x F, with r=[0.3,0,0] and F=[0,-20,0].
2. Compute using the determinant formula: M = [(0)(0)-(0)(-20), (0)(0)-(0.3)(0), (0.3)(-20)-(0)(0)] = [0,0,-6].
3. So M = -6k Newton-metres -- a vector along the negative z-axis with magnitude 6.
4. Final answer: the moment of the force about the bolt is 6 N.m in magnitude, directed along -k (a clockwise turning effect when viewed from the +z direction).

Example 8: Scalar Triple Product: Volume of a Parallelepiped

Problem: Find the volume of the parallelepiped with edges u = [1,0,0], v = [0,2,0], and w = [0,0,3] from a common vertex.

1. Compute the scalar triple product u.(v x w) as the determinant of the three vectors' components (rows u, v, w).
2. First find v x w = [(2)(3)-(0)(0), (0)(0)-(0)(3), (0)(0)-(2)(0)] = [6,0,0].
3. Then dot with u: u.(v x w) = (1)(6)+(0)(0)+(0)(0) = 6.
4. Volume = |u.(v x w)| = |6| = 6.
5. Final answer: the volume of the parallelepiped is 6 cubic units (matching 1x2x3, since these edges are mutually perpendicular).

Short Questions & Answers

What rule fixes the orientation of the x-, y-, and z-axes in space?

The right-hand rule: curling the fingers of the right hand from the positive x-axis to the positive y-axis, the thumb points along the positive z-axis.

How is the magnitude of u = [u1,u2,u3] computed?

|u| = sqrt(u1^2+u2^2+u3^2).

What condition makes two vectors parallel?

One vector is a nonzero scalar multiple of the other.

What identity do direction cosines always satisfy?

cos^2(alpha)+cos^2(beta)+cos^2(gamma) = 1.

What does a dot product of zero indicate about two nonzero vectors?

That the vectors are orthogonal (perpendicular).

What type of quantity does the cross product of two vectors produce?

A vector, perpendicular to both original vectors, with direction given by the right-hand rule.

What does a scalar triple product of zero indicate about three vectors?

That the three vectors are coplanar (lie in the same plane).

Long Questions & Answers

Explain how a vector in space is described using coordinates and unit vectors, and how its magnitude, unit vector, and direction cosines are found.

How is a point located in the rectangular coordinate system in space?

By an ordered triple (a,b,c) relative to three mutually perpendicular axes meeting at the origin, oriented by the right-hand rule; the position vector OP = [a,b,c] runs from the origin to the point.

How is a vector written using the unit vectors i, j, and k?

As u = u1 i + u2 j + u3 k, where i=[1,0,0], j=[0,1,0], k=[0,0,1] are the standard unit vectors along the three axes and u1, u2, u3 are the components of u.

How are the magnitude and unit vector of u found?

The magnitude is |u| = sqrt(u1^2+u2^2+u3^2), the three-dimensional Pythagorean theorem; the unit vector in the same direction is u-hat = u/|u|, obtained by dividing each component by the magnitude.

What are direction cosines, and what identity do they satisfy?

The cosines of the angles a vector makes with the positive x-, y-, and z-axes; each equals the corresponding component divided by the magnitude, and they always satisfy cos^2(alpha)+cos^2(beta)+cos^2(gamma) = 1.

Compare the dot product and the cross product of two vectors: how each is defined, computed, and applied.

How is the dot product defined and computed?

Geometrically as u.v = |u||v| cos(theta); in components as u1 v1+u2 v2+u3 v3. The result is always a scalar, so the dot product is also called the scalar product.

How is the cross product defined and computed?

As the vector u x v = (|u||v| sin(theta)) n, where n is a unit vector perpendicular to both u and v fixed by the right-hand rule; in components it is computed via a 3x3 determinant with i, j, k in the first row. The result is always a vector, so it is also called the vector product.

What real-world quantity does each product naturally compute?

The dot product computes work done by a constant force, W=F.d, and gives projections and angles between vectors. The cross product computes the area of a parallelogram or triangle, |u x v| or (1/2)|u x v|, and the moment (torque) of a force, M=r x F.

How does each product test a special geometric relationship?

Two nonzero vectors are orthogonal exactly when their dot product is zero; two nonzero vectors are parallel exactly when their cross product is the zero vector --

each product vanishes under the opposite geometric extreme from the one it naturally measures.

Multiple Choice Questions (MCQs)

The position vector of the point P(a,b,c) is:

A. [a,b]

B. [a,b,c]

C. a+b+c

D. sqrt(a^2+b^2+c^2)

Answer: B. [a,b,c]

Explanation: The position vector OP of a point P(a,b,c) in space is the vector [a,b,c] from the origin to P.

The magnitude of u = [u1,u2,u3] is:

A. u1+u2+u3

B. sqrt(u1^2+u2^2+u3^2)

C. u1 u2 u3

D. u1^2+u2^2+u3^2

Answer: B. sqrt(u1^2+u2^2+u3^2)

Explanation: The magnitude extends the Pythagorean theorem to three dimensions: |u| = sqrt(u1^2+u2^2+u3^2).

Direction cosines of a vector always satisfy:

A. cos(alpha)+cos(beta)+cos(gamma)=1

B. cos^2(alpha)+cos^2(beta)+cos^2(gamma)=1

C. cos(alpha)cos(beta)cos(gamma)=1

D. cos^2(alpha)-cos^2(beta)=1

Answer: B. cos^2(alpha)+cos^2(beta)+cos^2(gamma)=1

Explanation: The direction cosine identity is cos^2(alpha)+cos^2(beta)+cos^2(gamma) = 1.

The dot product of two vectors is:

A. Always a vector

B. Always a scalar

C. Sometimes a scalar, sometimes a vector

D. Undefined in space

Answer: B. Always a scalar

Explanation: The dot product u.v is always a scalar, which is why it is also called the scalar product.

Two nonzero vectors u and v are orthogonal exactly when:

A. u x v = 0

B. u.v = 0

C. |u| = |v|

D. u = v

Answer: B. u.v = 0

Explanation: Orthogonality (perpendicularity) of two nonzero vectors is equivalent to their dot product being zero.

The cross product u x v is:

A. Always a scalar

B. Always a vector perpendicular to both u and v

C. Always parallel to u

D. Always equal to u.v

Answer: B. Always a vector perpendicular to both u and v

Explanation: The cross product produces a vector perpendicular to both u and v, with direction fixed by the right-hand rule.

The magnitude |u x v| equals:

A. The volume of a parallelepiped

B. The area of the parallelogram with sides u and v

C. The dot product u.v

D. Zero always

Answer: B. The area of the parallelogram with sides u and v

Explanation: |u x v| = |u||v| sin(theta), which is exactly the area of the parallelogram spanned by u and v.

The moment of a force F applied at position r about a pivot is:

A. r.F

B. r x F

C. F/r

D. r+F

Answer: B. r x F

Explanation: The moment (torque) of a force is the cross product M = r x F.

Three vectors u, v, w are coplanar exactly when:

A. u.(v x w) = 0

B. u.(v x w) = 1

C. u x v = w

D. u+v+w = 0

Answer: A. u.(v x w) = 0

Explanation: A scalar triple product of zero means the three vectors enclose zero volume, i.e. they are coplanar.

The volume of a tetrahedron with edges u, v, w from a common vertex is:

A. |u.(v x w)|

B. (1/2)|u.(v x w)|

C. (1/6)|u.(v x w)|

D. (1/3)|u.(v x w)|

Answer: C. (1/6)|u.(v x w)|

Explanation: The tetrahedron's volume is one-sixth of the parallelepiped's volume: (1/6)|u.(v x w)|.

Quick Revision Summary

- A point P(a,b,c) in space has position vector OP=[a,b,c]; axes are oriented by the right-hand rule
- Magnitude: |u| = sqrt(u1^2+u2^2+u3^2); unit vector: u-hat = u/|u|
- Distance between two points: d = sqrt[(x2-x1)^2+(y2-y1)^2+(z2-z1)^2]
- Direction cosines satisfy cos^2(alpha)+cos^2(beta)+cos^2(gamma) = 1
- Dot product: u.v = |u||v| cos(theta) = u1 v1+u2 v2+u3 v3 -- always a scalar
- Two vectors are orthogonal exactly when u.v = 0; angle between vectors: theta = cos^-1[(u.v)/(|u||v|)]
- Work done by a constant force: W = F.d
- Cross product u x v is a vector perpendicular to both u and v, computed via a 3x3 determinant; |u x v| = |u||v| sin(theta)
- |u x v| gives the area of the parallelogram with sides u,v; (1/2)|u x v| gives the area of the triangle
- Moment of a force: M = r x F

- Scalar triple product u.(v x w) is cyclic; its absolute value gives the volume of the parallelepiped with edges u,v,w
- Three vectors are coplanar exactly when their scalar triple product is zero; tetrahedron volume is (1/6) of the parallelepiped's volume

Exam Tips

- Always sketch or at least mentally place the right-hand rule before assigning axis directions -- sign errors in 3D problems are the most common mistake
- When finding a unit vector, always compute the magnitude first, then divide every component by it -- and check the result really has magnitude 1
- Remember: dot product answers 'how much' (a scalar -- projection, angle, work), cross product answers 'perpendicular to what' (a vector -- area, torque)
- Set up the cross product determinant carefully with i, j, k in the first row and the two vectors' components in the next two rows -- a transposed row is the most common cross-product error
- To test orthogonality, parallelism, or coplanarity, compute the dot product, cross product, or scalar triple product respectively and check whether it equals zero
- For real-world force/torque problems, first identify which vector is the force and which is the position vector relative to the pivot before applying M = r x F