1. Matrix

Matrix

A Matrix consists of a rectangular array of elements represented by a single symbol. [A] is the notation for the matrix and a_{ij} designates individual element of the matrix.

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \dots & a_{ij} & \dots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

A horizontal set of elements is called a row and a vertical set is called a column. The first subscript *i* always designates the number of the row in which the element lies. The second subscript *j* designates the column. For example, element a_{23} is in row 2 and column 3.

The matrix has n rows and m columns and is said to have a dimension of n by m or n x m. It is referred to as an n by m matrix.

Special types of matrices

Diagonal Matrix

A diagonal matrix is the a square matrix where all elements off the main diagonal are equal to zero, as in

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

Square Matrix

Matrices where n = m are called square matrices. For example, a 3 by 3 matrix is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Triangular Matrices

An upper triangular matrix is the one where all the elements below the main diagonal are zero

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

A lower triangular matrix is the one where all elements above the main diagonal are zero

$$\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Symmetric Matrix

A symmetric matrix is the one where a_{ij}= a_{ji} for all i's and j's.

Example

$$\begin{bmatrix} 5 & 1 & 2 \\ 1 & 3 & 7 \\ 2 & 7 & 8 \end{bmatrix}$$

Square Matrix

Matrices where n = m are called square matrices.

Example, a 3 by 3 matrix is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Block or Partitioned Matrix

A key feature of our work with matrices has been the ability to regard a matrix A as a list of column vectors rather than just a rectangular array of numbers. This point of view has been so useful that we wish to consider other partitions of A, indicated by horizontal and vertical dividing rules, as in Example below.

Example

$$\left[\begin{array}{cc|cc|cc} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \end{array}\right]$$

Augmented Matrix

Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{2n} \\ a_{m1} & a_{m2} & a_{m3} & a_{mn} \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_n \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ b_m \end{bmatrix}$

Then, the system of equations

$$a_{11}x_1 + a_{12}x_2 + \ \dots\dots \ + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ \dots\dots \ + a_{2n}x_n = b_2$$
$$a_{m1}x_1 + a_{m2}x_2 + \ \dots\dots \ + a_{mn}x_n = b_m$$

can be rewritten as Ax = b, where A is called the coefficient matrix and the matrix [A | b] is called the augmented matrix.

Example

Given the system

$$\begin{aligned} x_1 - 2x_2 \ + \ \ x_3 &= 0 \\ 2x_2 \ - \ 8x_3 &= 8 \\ 5x_1 \qquad - \ \ x_3 &= 10 \end{aligned}$$

The matrix $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & 8 \\ 5 & 0 & -1 \end{bmatrix}$ is called the coefficient matrix

$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -1 & 10 \end{array}\right]$ is called the augmented matrix of the system.

Diagonal of a matrix

Let A= $[a_{ij}]$ be a n-square matrix. The diagonal or main diagonal of A consists of the elements with the same subscripts i.e. a_{11}, a_{22},…. a_{nn}. It is denoted by diag(A)

Example

$$diag\left(\begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 7 \\ 5 & 2 & 6 \end{bmatrix}\right) = [1,\ 2,\ 6]$$

Trace of a matrix
Trace of matrix A, denoted by trA, is the sum of the elements along the diagonal of A
i.e. $trA = a_{11} + a_{22} + \; \ldots . \; + a_{nn}$
Example
For $A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 7 \\ 5 & 2 & 6 \end{bmatrix}$
trA= 1 + 2 + 6 = 9

Matrix Operations

Scalar Multiplication
The product of the matrix A by a scalar k, written k.a or simply kA, is the matrix obtained by multiplying each element of A by k.

kA = $\begin{bmatrix} ka_{11} & ka_{12} & .. & ka_{13} \\ ka_{21} & ka_{22} & .. & ka_{23} \\ .. & .. & .. & .. \\ ka_{m1} & ka_{m2} & .. & ka_{mn} \end{bmatrix}$

Example
Let A = $\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ Then, 3.A = $\begin{bmatrix} 3(2) & 3(2) & 3(0) \\ 3(1) & 3(0) & 3(0) \\ 3(0) & 3(0) & 3(3) \end{bmatrix}$ = $\begin{bmatrix} 6 & 6 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 9 \end{bmatrix}$

Matrix Addition and Subtraction
Addition of two matrices, A and B, is accomplished by adding corresponding terms in each matrix. The elements of the resulting matrix C are computed,
c_{ij}= a_{ij} + b_{ij} for i = 1, 2, . . ., n and j = 1, 2, . . ., m.

Similarly, the subtraction of two matrices, A minus B, is obtained by subtraction corresponding terms,
c_{ij}= a_{ij} - $b_{ij}$for i = 1, 2, . . ., n and j = 1, 2, . . ., m.

It follows directly from the above definition that addition and subtraction can be performed only between matrices having the same dimensions

Properties of Matrix Addition and Scalar Multiplication
Let A, B, and C be matrices of the same size, and let r and s be scalars.

- A + B = B + A
- (A + B) + C = A + (B + C)
- A + 0 = A
- r(A + B)= rA + rB
- (r + s)A= rA + sA
- r(sA) = (rs)A

Matrix Multiplication

The product of two matrices is represented a C = A . B, where the elements of C are defined as $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$
where n= the column dimension of A and the row dimension of B.
That is, the c_{ij} element is obtained by adding the product of individual elements from the i^{th}row of the first matrix, in this case A, by the j^{th}column of the second matrix B
The product *AB* (or *A* · *B*) of two matrices *A* and *B* is defined only when the number of columns in *A* is equal to the number of rows in *B*.

Example

Suppose A =$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ $B = \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix}$

$$\text{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} . \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 5+0 & 6-4 \\ 15+0 & 18-8 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 15 & 10 \end{bmatrix}$$

Properties of Matrix Multiplication

- Associative : Matrix Multiplication is Associative
  (AB)C = A(BC)
- Distributive : Matrix Multiplication is Distributive
  A(B + C) = AB + BC
- Commutative : Matrix Multiplication is not commutative.
  A B ≠ BA
- r(AB) = (rA)B = A(Br), where r is a scalar
- IA = A = AI

Example

Suppose A =$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ $B = \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix}$

AB = $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} . \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 5+0 & 6-4 \\ 15+0 & 18-8 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 15 & 10 \end{bmatrix}$
and
BA = $\begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix} . \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5+18 & 10+24 \\ 0-6 & 0-8 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ -6 & -8 \end{bmatrix}$
The above example shows that matrix multiplication is not commutative.
i.e. A B ≠ BA

Powers of a Matrix

If A is a n×n matrix and if k is a positive integer, then A^k denotes the product of k copies of A:
$A^k = A \dots \dots A(k\ times)$
If A is nonzero and if x is in R^n, then $A^k\ x$ is the result of left multiplying x by A repeatedly k times.
If k = 0, then $A^0\ x$ should be x itself. Thus A^0 is interpreted as the identity matrix.

Hadamard Product

Hadamard product or element wise product is denoted $\odot$ or O. The Hadamard product is only defined over matrices of equal size and returns a matrix of the same size

Example

$$\begin{bmatrix} 5 & 10 \\ -2 & 0 \\ 1 & -1 \end{bmatrix} \odot \begin{bmatrix} 2 & -1 \\ -1 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 \times 2 & 10 \times -1 \\ -2 \times -1 & 0 \times 5 \\ 1 \times 0 & -1 \times 1 \end{bmatrix} = \begin{bmatrix} 10 & -10 \\ 2 & 0 \\ 0 & -1 \end{bmatrix}$$

Properties of Hadamard product

- Associative : Hadamard product is Associative

  (AB)C = A(BC)
- Distributive : Hadamard product is Distributive

  A(B + C) = AB + BC
- Commutative : Hadamard product is commutative.

  A B = BA

Kronecker Product

Definition Let A be a $K \times L$ matrix and B an $M \times N$ matrix. Then, the Kronecker product between A and B is the $(KM \times LN)$ block matrix

$$A \otimes B = \begin{bmatrix} A_{11}B & \dots & A_{1L}B \\ \vdots & \ddots & \vdots \\ A_{K1}B & \dots & A_{KL}B \end{bmatrix}$$

where A_{kl} denotes the (k,l)-th entry of A.

In other words, the Kronecker product $A \otimes B$ is a block matrix whose (k,l)-th block is equal to the (k,l)-th entry of A multiplied by the matrix B.

Note that, unlike the ordinary product between two matrices, the Kronecker product is defined regardless of the dimensions of the two matrices A and B.

Example

Find the Kronecker product of the following matrices

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & -1 \\ -1 & 4 \end{bmatrix}$$

Solution

$$A \otimes B = \begin{bmatrix} 2B & 0B \\ 1B & 3B \end{bmatrix}$$

$$= \begin{bmatrix} 10 & -2 & 0 & 0 \\ -2 & 8 & 0 & 0 \\ 5 & -1 & 15 & -3 \\ -1 & 4 & -3 & 12 \end{bmatrix}$$

Example

Find the Kronecker product of the following matrices

$$A = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 2 & 7 \end{bmatrix}$$

Solution

$$A \otimes B = \begin{bmatrix} 3B \\ 2B \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 \\ 6 & 21 \\ 2 & 0 \\ 4 & 14 \end{bmatrix}$$

Example

Find the Kronecker product of the following matrices

$$A = \begin{bmatrix} 1 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 5 \end{bmatrix}$$

Solution

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

$$= \begin{bmatrix} 1 & 0 & 5 & -1 & 0 & -5 \end{bmatrix}$$

Example

Find the Kronecker product of the following matrices

$$A = \begin{bmatrix} 3 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Solution

$$A \otimes B = \begin{bmatrix} 3B & 6B \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 0 & 0 \\ 3 & 6 \end{bmatrix}$$

Properties of Kronecker product

- Associative: Kronecker product is Associative
  $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
- Distributive : Kronecker product is Distributive
  $A \otimes (B + C) = A \otimes B + B \otimes C$
  $(A + B) \otimes C = A \otimes C + B \otimes C$
- Commutative : Kronecker product is not commutative.
  $(A \otimes B) \neq (B \otimes A)$
- $r \otimes B = rB$ where r is a scalar
- $(rA) \otimes (sB) = (rs)(A \otimes B)$ where r,s are scalars
- $A \otimes 0 = 0$ and $0 \otimes B = 0$

Transpose of a matrix

Transpose of a matrix is obtained by switching the row elements with the column elements. We denote the transpose of a matrix A by A^T

Example

A = $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

Properties of Transpose

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- For any scalar r, $(rA)^T = rA^T$
- $(AB)^T = B^T A^T$

Length or Norm of a vector

The length or norm of v is the nonnegative scalar $||v||$ defined by

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

and $||v||^2 = v.v$

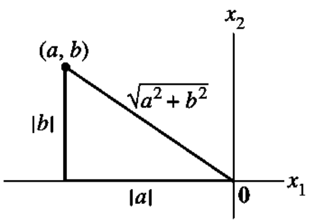


FIGURE 1

Interpretation of $\|\mathbf{v}\|$ as length.

Norm of a matrix

For the rectangular n × d matrix A with (i,j)th entry denoted by a_{ij}, its Frobenius norm is defined as follows:

$$||A||_F = ||A^T||_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} a_{ij}^2}$$

Note the use of ‖ · ‖F to denote the Frobenius norm. The squared Frobenius norm is the sum of squares of the norms of the row-vectors (or, alternatively, column vectors) in the matrix. It is invariant to matrix transposition.

Energy of a matrix

The energy of a matrix A is an alternative term used in machine learning community for the squared Frobenius norm.

The energy of a rectangular matrix A is equal to the trace of either AA^T or A^TA

$$||A||_F^2 = Energy(A) = tr(AA^T) = tr(A^T A)$$

Orthonormal Vectors

Let q_1, q_2, ……. , q_n be vectors, they are said to be orthonormal

if $q_i^T q_j = \begin{cases} 0 & if\ 1 \neq j \\ 1 & if\ 1 = j \end{cases}$

In other words, the length of each vector is 1

Orthogonal Matrix

An orthogonal matrix is a square matrix with orthonormal columns

$$\begin{bmatrix} - & q_1^T & - \\ - & q_i^T & - \\ - & q_n^T & - \end{bmatrix} \begin{bmatrix} | & | & | \\ q_1 & q_j & q_n \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathrm{I}$$

When row i of Q^T multiplies column j of Q, the result is

$q_j^T q_j = 0$

On the diagonals where i =j, we have $q_j^T q_j = 1$, ie. The normalization to unit vector of length i

An orthonormal matrix is a type of square matrix whose columns and rows are orthonormal unit vector, eg. Perpendicular and have a length or magnitude of 1.

Then Q^T is Q^{-1}

ie. $Q^T Q = Q Q^T = I$

Computing of Q^T is more time efficient as compared to computing Q^{-1}

Example

$$Q = \begin{bmatrix} \cos\theta & -sin\theta \\ sin\theta & \cos\theta \end{bmatrix}$$

$$Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & sin\theta \\ -sin\theta & \cos\theta \end{bmatrix}$$

Questions

1. Compute matrix sum or product if it is defined. If an expression is undefined, explain why.
   Let

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -5 & 2 \end{bmatrix}, B = \begin{bmatrix} 7 & -5 & 1 \\ 1 & -4 & -3 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}, E = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

a.	-2A	b.	B-2A	c.	AC	d.	CD	e.	A+2B	f.	3C-E
g.	CB^T	h.	EB	i.	trD	j.	A⊙B	k.	$C \odot D$	l.	C⊗E

2. If a matrix A is 5×3 and the product AB is 5×7, what is the size of B ?
3. How many rows does B have if BC is a 3×4 matrix?

Answers

1a) $\begin{bmatrix} -4 & 0 & 2 \\ -8 & 10 & -4 \end{bmatrix}$ 1b) $\begin{bmatrix} 3 & -5 & 3 \\ -7 & 6 & -7 \end{bmatrix}$

1c) The product AC is not defined because the number of columns of A does not match the number of rows of C.

1d) $\begin{bmatrix} 1 & 13 \\ -7 & -6 \end{bmatrix}$ 1e) $\begin{bmatrix} 16 & -10 & 1 \\ 6 & -13 & -4 \end{bmatrix}$

1f) The expression 3C – E is not defined because 3C has 2 columns and –E has only 1 column.

1g) $\begin{bmatrix} 9 & -13 \\ -13 & 6 \\ -5 & -5 \end{bmatrix}$

1h) The product EB is not defined because the number of columns of E does not match the number of rows of R.

1i) 7 1j) $\begin{bmatrix} 14 & 0 & -1 \\ 4 & 20 & -6 \end{bmatrix}$ 1k) $\begin{bmatrix} 3 & 10 \\ 2 & 4 \end{bmatrix}$ 1l) $\begin{bmatrix} -5 & -10 \\ 3 & 6 \\ 10 & -5 \\ -6 & 3 \end{bmatrix}$

2) Since A has 3 columns, B must match with 3 rows. Otherwise, AB is undefined. Since AB has 7 columns, so does B. Thus, B is 3×7.

3) The number of rows of B matches the number of rows of BC, so B has 3 rows.

2 Determinants

Determinant of 2 x 2 matrix

For a 2x2 Matrix A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is denoted by |A| or detA

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

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is |det A|.

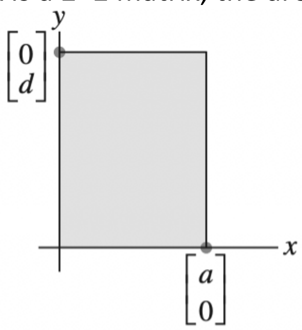


FIGURE 1
Area $= |ad|$.

Example

Let A = $\begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}. Find\ |A|$

Solution

$|A| = \begin{vmatrix} 2 & -3 \\ 1 & -2 \end{vmatrix}$ = 2(−2) – 1(-3) = -1

Minor of a matrix

Minor of matrix A is the determinant of some smaller square matrix cut down from A by removing one or more of its rows and columns.

The minor M_{ij} is the determinant of the submatrix obtained by the deleting the i^{th} row and j^{th} column.

Example

Let A be a 3 x 3 matrix

$A = \begin{bmatrix} 1 & 4 & 4 \\ 7 & 2 & 8 \\ -1 & 9 & 0 \end{bmatrix}$. Find minor $M_{2,\,3}$

Solution

The minor $M_{2,\,3}$ of A can be obtained as the determinant of the above matrix with row 2 and column 3 removed.

$$M_{2,\,3} = det\begin{bmatrix} 1 & 4 & \\ & & \\ -1 & 9 & \end{bmatrix} = det\begin{bmatrix} 1 & 4 \\ -1 & 9 \end{bmatrix} = 9 - (-4) = 13$$

Co-factor of a matrix
The cofactor C_{ij} is obtained by multiplying the minor M_{ij} by $(-1)^{i+j}$

Example
For the above example,
$C_{2,3} = (-1)^{2+3} M_{2,3} = -13$

Adjoint of matrix
Let A = $[a_{ij}]$ be a nxn matrix and let C_{ij} denote the cofactor of $a_{ij.}$ The classical adjoint or adjugate of A, denoted by the adjA, is the transpose of the matrix of cofactors of A
$adjA = [C_{ij}]^T$

Determinant of a 3x3 Matrix
For a 3x3 Matrix
A = $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

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

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is |det A|. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is |det A|.

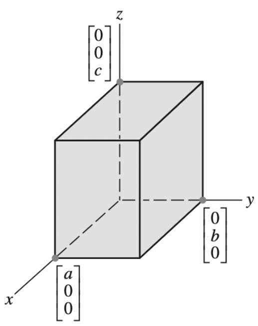


FIGURE 3
Volume = $|abc|$.

Determinant of a nxn Matrix
For a nxn Matrix
A= $\begin{bmatrix} a_{11} & a_{12} & .. & a_{1n} \\ a_{21} & a_{22} & .. & a_{2n} \\ .. & .. & .. & .. \\ a_{n1} & a_{n2} & .. & a_{nn} \end{bmatrix}$

$|A| = \sum_{\substack{1 \le i \le n \\ 1 \le j \le n}} a_{ij} C_{ij}$, where C_{ij} is the co-factor of a_{ij}

C_{ij} = $(-1)^{i+j} M_{ij,}$ where M_{ij} is M_{ij}which is the determinant of the reduced matrix obtained by removing the row i and column j

The determinant of an n×n matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactors in (4) is
$|A| = a_{i1}C_{i1} + a_{i2}C_{i2}$+… + $a_{in}C_{in}$, where $C_{ij}$is the co-factor of a_{ij}
The cofactor expansion down the j th column is
$|A| = a_{1j}C_{1j} + a_{2j}C_{2j}$+… + $a_{nj}C_{nj}$, where C_{ij} is the co-factor of a_{ij}

Example 1
Find the determinant for a 3x3 Matrix

$$A = \begin{bmatrix} 7 & 2 & 1 \\ 0 & 3 & -1 \\ -3 & 4 & -2 \end{bmatrix}$$

Solution
$|A| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 7C_{11} + 2C_{12} + 1C_{13}$

$$C_{11} = (-1)^{1+1}M_{11} = (-1)2\begin{vmatrix} 3 & -1 \\ 4 & -2 \end{vmatrix} = -6+4 = -2$$

$$C_{12} = (-1)^{1+2}M_{12} = (-1)3\begin{vmatrix} 0 & -1 \\ -3 & -2 \end{vmatrix} = (-1)(0-3) = 3$$

$$C_{13} = (-1)^{1+3}M_{13} = (-1)4\begin{vmatrix} 0 & 3 \\ -3 & 4 \end{vmatrix} = 0+9= 9$$

$|A| = 7(-2) + 2(3) +9 =1$

Example 2
Find the determinant for a 3x3 Matrix

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

Solution
$|A| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 2C_{11} + 6C_{12} + 3C_{13}$

$$C_{11} = (-1)^{1+1}M_{11} = (-1)2\begin{vmatrix} 0 & 2 \\ 0 & -2 \end{vmatrix} = 0$$

$$C_{12} = (-1)^{1+2}M_{12} = (-1)3\begin{vmatrix} 1 & 2 \\ 2 & -2 \end{vmatrix} = 6$$

$$C_{13} = (-1)^{1+3}M_{13} = (-1)4\begin{vmatrix} 1 & 0 \\ 2 & 0 \end{vmatrix} = 0$$

$|A| = 2(0) + 6(1) + 3(0) = 6$

Properties of Determinant

- If A is a triangular matrix, then detA is the product of entries on the main diagonal of A.
- The determinant of identity matrix is 1.
- A square matrix A is invertible if and only if det A $\neq$ 0.
- If A has a row(column) of zeros, then |A| = 0
- If A has two identical rows (columns) then |A| = 0
- If a multiple of one row of A is added to another row to get matrix B, then det B = det A.
- If two rows of A are interchanged to produce B, then det B = −det A.
- If one row of A is multiplied by k to produce B, then det B = k · det A.
- The determinant of matrix A and its transpose are equal
  $|A| = |A^T|$
- If A and B are n×n matrices, then det AB = (det A)(det B)

Questions

1. Find the determinant of the following matrices $A = \begin{bmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{bmatrix}, B = \begin{bmatrix} 0 & 5 & 1 \\ 4 & -3 & 0 \\ 2 & 4 & 1 \end{bmatrix}, C = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$

2. Each equation given below illustrates a property of determinants. State the property.

a) $\begin{vmatrix} 0 & 5 & -2 \\ 1 & -3 & 6 \\ 4 & -1 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 6 \\ 0 & 5 & -2 \\ 4 & -1 & 8 \end{vmatrix}$ b) $\begin{vmatrix} 2 & -6 & 4 \\ 3 & 5 & -2 \\ 1 & 6 & 3 \end{vmatrix} = 2 \begin{vmatrix} 1 & -3 & 2 \\ 3 & 5 & -2 \\ 1 & 6 & 3 \end{vmatrix}$

c) $\begin{vmatrix} 1 & 3 & -4 \\ 2 & 0 & -3 \\ 5 & -4 & 7 \end{vmatrix} = \begin{vmatrix} 1 & 3 & -4 \\ 0 & -6 & 5 \\ 5 & -4 & 7 \end{vmatrix}$ d) $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \\ 3 & 7 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \\ 0 & 1 & -5 \end{vmatrix}$

3. Let A and B be 3×3 matrices, with det A =4 and det B =−3. Use properties of to compute:

a) det AB b) det B^T c) det A^{-1} d) det A^3

4. Let A and B be 4×4 matrices, with det A = −1 and det B = 2. Compute:

a) det AB b) det B^5 c) det $A^T A$ d) det $B^{-1}AB$

5. Verify that det AB = (det A)(det B) for the matrices given below

a) $A = \begin{bmatrix} 3 & 0 \\ 6 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 2 & 0 \\ 5 & 4 \end{bmatrix}$

b) $A = \begin{bmatrix} 3 & 6 \\ -1 & -2 \end{bmatrix}$ and $A = \begin{bmatrix} 4 & 2 \\ -1 & -1 \end{bmatrix}$

Answers

1) |A| = 1, |B| = 2, |C| = -5

2a) Rows 1 and 2 are interchanged, so the determinant changes sign

2b) The constant 2 may be factored out of the Row 1

2c) The row replacement operation does not change the determinant

2d) The row replacement operation does not change the determinant

3a) det AB = (det A)(det B) = (4)(–3) = –12.

3b) det B^T = det B = −3.

3c) det A^{-1} = 1/det A=1/4.

3d) det A^3 = $(det\ A)^3$ = 4^3 = 64.

4a) det AB = (det A)(det B) = (–1)(2) = –2.

4b) det B^5 = $(det\ B)^5$ = 2^5 = 32.

4c) det $A^T A$ = (det A^T)(det A) = (det A)(det A) = (−1)(−1) = 1.

4d) det $B^{-1}AB$ = (det B^{-1}) (det A) (det B) = (1 / det B) (det A) (det B) = det A = − 1 .

5a) det AB = 24 = (3)(8) = (det A)(det B).

5b) det AB = 0 = (0)(–2) = (det A)(det B).

3. Inverse using determinants

Let square matrices A and B have the property
AB = BA = I
then B is called the inverse of A and denoted by A^{-1}
Not every matrix A possesses an inverse.
If this inverse does exist, A is called regular/invertible/non-singular.
If this inverse does not exist, A is called singular.
If matrix inverse exists, it is unique.

Properties of inverse

- $A^{-1}A = I \text{ and } AA^{-1} = I$
- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}\,A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$ Transpose of inverse is inverse of the transpose

Inverse using determinant

$$A^{-1} = \frac{1}{|A|}\,(adjA)$$
$|A| \neq 0$

Example Find inverse of A = $\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ using determinant

Solution
$|A|$ = $6 - 6 = 0$
Hence, A is not invertible

Example Show that given matrices are inverses of each other

$$A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} \text{ and } C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$$

Solution

$$AC = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}\begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and}$$
$$CA = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, C = A^{-1}

Example Find inverse of A = $\begin{bmatrix} 2 & 3 & -4 \\ 0 & -4 & 2 \\ 1 & -1 & 5 \end{bmatrix}$ using determinant

Solution

$$C_{11} = +\begin{vmatrix} -4 & 2 \\ -1 & 5 \end{vmatrix} = -18,\ C_{12} = -\begin{vmatrix} 0 & 2 \\ 1 & 5 \end{vmatrix} = 2,\ C_{13} = +\begin{vmatrix} 0 & -4 \\ 1 & -1 \end{vmatrix} = 4$$

$$C_{21} = -\begin{vmatrix} 3 & -4 \\ -1 & 5 \end{vmatrix} = -11,\ C_{22} = +\begin{vmatrix} 2 & -4 \\ 1 & 5 \end{vmatrix} = 14,\ C_{23} = -\begin{vmatrix} 2 & 3 \\ 1 & -1 \end{vmatrix} = 5$$

$C_{31} = +\begin{vmatrix} 3 & -4 \\ -4 & 2 \end{vmatrix} = -10,\ C_{32} = -\begin{vmatrix} 2 & -4 \\ 0 & 2 \end{vmatrix} = -4,\ C_{33} = +\begin{vmatrix} 2 & 3 \\ 0 & -4 \end{vmatrix} = -8$

$C_{ij} = \begin{bmatrix} -18 & 2 & 4 \\ -11 & 14 & 5 \\ -10 & -4 & -8 \end{bmatrix}$

The transpose of the above matrix of cofactors yields the adjoint of A

$adjA = \begin{bmatrix} -18 & -11 & -10 \\ 2 & 14 & -4 \\ 4 & 5 & -8 \end{bmatrix}$

$det(A) = -40 + 6 + 0 - 16 + 4 + 0 = -46$

$det(A) \neq 0$

$A^{-1} = \frac{1}{|A|}\ (adjA) = \frac{-1}{46}\begin{bmatrix} -18 & -11 & -10 \\ 2 & 14 & -4 \\ 4 & 5 & -8 \end{bmatrix}$

$= \begin{bmatrix} 9/23 & 11/46 & 5/23 \\ -1/23 & -7/23 & 2/23 \\ -2/23 & -5/46 & 4/23 \end{bmatrix}$

Example Find inverse of A = $\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix}$

Solution

C_{11} = -1	C_{12} = 8	C_{13} = -5
C_{21} = 1	C_{22} = -6	C_{23} = 3
C_{31} = -1	C_{32} = 2	C_{33} = -1

$adjA = \begin{bmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{bmatrix}$

$|A| = 0 - 1(11 - 9) + 2(1 - 6) = 8 - 10$ = -2

$A^{-1} = \frac{1}{-2}\begin{bmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ -4 & 3 & -1 \\ 5/2 & -3/2 & 1/2 \end{bmatrix}$

Example Find inverse of A = $\begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$

Solution

|A| = 9

Co-factor matrix = $\begin{bmatrix} 1 & 4 & -3 \\ -6 & 3 & 0 \\ 2 & -1 & 3 \end{bmatrix}$

$$A^{-1} = \begin{bmatrix} 1/9 & -6/9 & 2/9 \\ 4/9 & 3/9 & -1/9 \\ -3/9 & 0 & 3/9 \end{bmatrix}$$

Questions

1. Use determinants to determine which of the following matrices are invertible.

$a)\ \begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix}$ $b)\ \begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix}$ $c)\ \begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix}$

2. Find the inverses of the matrices

$a)\ \begin{bmatrix} 8 & 6 \\ 5 & 4 \end{bmatrix}$ $b)\ \begin{bmatrix} 3 & 2 \\ 7 & 4 \end{bmatrix}$ $c)\ \begin{bmatrix} 8 & 5 \\ -7 & -5 \end{bmatrix}$ $d)\ \begin{bmatrix} 3 & -4 \\ 7 & -8 \end{bmatrix}$

Answers

1a) det = 36, invertible 1b) det = 20, invertible 1c) det = 0, non-invertible

2.

$a)\ \begin{bmatrix} 2 & -3 \\ -5/2 & 4 \end{bmatrix}$ $b)\ \begin{bmatrix} -2 & 1 \\ 7/2 & -3/2 \end{bmatrix}$ $c)\ \begin{bmatrix} 1 & 1 \\ -1.4 & -1.6 \end{bmatrix}$ $d)\ \begin{bmatrix} -2 & 1 \\ -7/4 & 3/4 \end{bmatrix}$

4. Elementary Matrix

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

Example

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

Elementary Row Operations on a matrix

The following operations can be performed on the rows of the matrices

1. (Replacement) Replace one row by the sum of itself and a multiple of another row
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a non-zero constant.

- Row operations can be applied to any matrix. Two matrices are called row equivalent if there is a sequence of elementary row operations that transforms one matrix into the other.
- It is important to note that row operations are *reversible*. If two rows are interchanged, they can be returned to their original positions by another interchange. If a row is scaled by a nonzero constant c, then multiplying the new row by 1/c produces the original row.
- A *nonzero* row or column in a matrix means a row or column that contains at least one nonzero entry;
- A leading entry of a row refers to the leftmost nonzero entry (in a nonzero row).

Row echelon form

A rectangular matrix is in echelon form (or row echelon form) if it has the following three properties:

- All nonzero rows are above any rows of all zeros.
- Each leading entry of a row is in a column to the right of the leading entry of the row above it.
- All entries in a column below a leading entry are zeros.

Example

The following matrices are in echelon form.

$$\begin{bmatrix} 0 & 2 & 1 & 0 \\ 0 & 0 & -8 & 8 \\ 0 & 0 & 0 & 10 \end{bmatrix} \quad and \quad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & -5 & 10 \end{bmatrix}$$

Reduced row echelon form

If a matrix in echelon form satisfies the following additional conditions, then it is in reduced echelon form (or reduced row echelon form):

The leading entry in each nonzero row is 1

Each leading 1 is the only nonzero entry in its column.

Example

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0's below *and above* each leading 1.

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad and \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 10 \end{bmatrix}$$

Any nonzero matrix may be row reduced (that is, transformed by elementary row operations) into more than one matrix in echelon form, using different sequences of row operations. However, the reduced echelon form one obtains from a matrix is unique.

Pivot Position

A pivot position in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A.

A pivot column is a column of A that contains a pivot position.

Eg, $A = \begin{bmatrix} 0 & 3 & 4 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$

Here, the entries a_{12} and a_{a23} are pivots and columns 2 and 3 are pivotal columns.

Questions

1. Determine which matrices are in reduced echelon form and which others are only in echelon form.

a) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ $b)\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $c)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $d)\begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ $f)\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ $g)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ $h)\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Answers

a) Reduced echelon form
b) Reduced echelon form
c) Not Echelon form
d) Echelon form
e) Reduced echelon form
f) Echelon form
g) Not Echelon form
h) Echelon form

5. Inverse using Elementary Row Operations (Gauss Jordan Method)

If a matrix A is invertible, there are a set of steps to reduce it to the identity matrix, which also means that we have some set of elementary matrices such that

$E_n E_{n-1} \ldots E_2 E_1 A$ = I

However, by right-multiplying by A^{-1} (since A is invertible), we get $E_n E_{n-1} \ldots E_2 E_1$ I= A^{-1}

So by performing the steps to reduce A to the identity matrix, those same steps performed on the identity matrix create the inverse of A.

If we start with [A | I] and reduce the left side to the identity matrix, then we would end up with [I | A^{-1}]

Example Find the inverse of the matrix A = $\begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{bmatrix}$

Solution

Form the block matrix M = [A | I] and row reduce M to an echelon form

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{array}\right]$$

Subtract the two times reduced first row from the second row and also multiply the first row by 4 and then subtract from the third, gives us

Replace R_2 by $R_2 - 2R_1$ and R_3 by $R_3 - 4R_1$ to get

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right]$$

The next step is to multiply the second row by -1 gives us

Replace R_2 by (-1) R_2

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right]$$

We reduce the second column to [0,1,0] by row operations

Replace R_3 by $R_3 - R_2$ to get

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{array}\right]$$

The third step is to multiply the third row by -1 gives us

Replace R_3 by (-1) R_3

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{array}\right]$$

Subtract the from first row two times the third row

Replace R_1 by $R_1 - 2R_3$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{array}\right]$$

The inverse of the matrix is $\begin{bmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{bmatrix}$

Example Find the inverse of the matrix A = $\begin{bmatrix} 2 & 3 & -1 \\ 4 & 4 & -3 \\ 2 & -3 & 1 \end{bmatrix}$

Solution

Form the block matrix M = [A | I] and row reduce M to an echelon form

$$\left[\begin{array}{ccc|ccc} 2 & 3 & -1 & 1 & 0 & 0 \\ 4 & 4 & -3 & 0 & 1 & 0 \\ 2 & -3 & 1 & 0 & 0 & 1 \end{array}\right]$$

The next step is to multiply the first row by 1/2 gives us

Replace R_1 by (1/2) R_1

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 4 & 4 & -3 & 0 & 1 & 0 \\ 2 & -3 & 1 & 0 & 0 & 1 \end{array}\right]$$

Subtract the two times reduced first row from the second row and also multiply the first row by 4 and then subtract from the third, gives us

Replace R_2 by $R_2 - 4R_1$ and R_3 by $R_3 - 2R_1$ to get

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 0 & -2 & -1 & -2 & 1 & 0 \\ 0 & -6 & 2 & -1 & 0 & 1 \end{array}\right]$$

The next step is to multiply the second row by -1/2 gives us

Replace R_2 by (-1/2) R_2

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 1/2 & 1 & -1/2 & 0 \\ 0 & -6 & 2 & -1 & 0 & 1 \end{array}\right]$$

We reduce the second column to [0,1,0] by row operations

Replace R_1 by $R_1 - (3/2)R_2$, R_3 by $R_3 + 6R_2$ to get

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -5/4 & -1 & 3/4 & 0 \\ 0 & 1 & 1/2 & 1 & -1/2 & 0 \\ 0 & 0 & 5 & 5 & -3 & 1 \end{array}\right]$$

The third step is to multiply the third row by 1/5 gives us

Replace R_3 by (1/5) R_3

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -5/4 & -1 & 3/4 & 0 \\ 0 & 1 & 1/2 & 1 & -1/2 & 0 \\ 0 & 0 & 1 & 1 & -3/5 & 1/5 \end{array}\right]$$

Replace R_1 by $R_1 + (5/4)R_3$, R_2 by $R_2 - (1/2)R_3$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 1/2 & -1/5 & -1/10 \\ 0 & 0 & 1 & 1 & -3/5 & 1/5 \end{array}\right]$$

The inverse of the matrix is $\begin{bmatrix} 1/4 & 0 & 1/4 \\ 1/2 & -1/5 & -1/10 \\ 1 & -3/5 & 1/5 \end{bmatrix}$

The Invertible Matrix Theorem

Let A be a square n×n matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

- A is an invertible matrix.
- A is row equivalent to the n×n identity matrix.
- A has n pivot positions.
- There is an n×n matrix C such that CA = I.
- There is an n×n matrix D such that AD = I
- A^T is an invertible matrix.

Questions

1. Find the inverses of the matrices using Gauss Jordan Method

$a)\ \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \quad b)\ \begin{bmatrix} 5 & 10 \\ 4 & 7 \end{bmatrix} \quad c)\ \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix} \quad d)\ \begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$

Answers

1. Find the inverses of the matrices using Gauss Jordan Method

$a)\ \begin{bmatrix} -7 & 2 \\ 4 & -1 \end{bmatrix} \quad b)\ \begin{bmatrix} -7/5 & 2 \\ 4/5 & -1 \end{bmatrix} \quad c)\ \begin{bmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ 7/2 & 3/2 & 1/2 \end{bmatrix}$

d) The matrix is not invertible

6. Application of Matrices

Graphics

One important use of matrices is in the digital representation of images.

- A digital camera or a scanner converts an image into a matrix by dividing the image into a rectangular array of elements called pixels.
- Each pixel is assigned a value that represents the color, brightness, or some other feature of that location.

One Hot Encoding

Sometimes in datasets, we encounter columns that contain categorical features (string values)

Consider the data where fruits and their corresponding categorical value and prices are given.

Original data

Fruit	Category	Price
Apple	1	5
Mango	2	10
Apple	1	15
Orange	3	20

One hot encoded data

Apple	Mango	Orange	Price
1	0	0	5
0	1	0	10
1	0	0	15
0	0	1	20

Though this approach eliminates the hierarchy/order issues but does have the downside of adding more columns to the data set. It can cause the number of columns to expand greatly if you have many unique values in a category column. In the above example, it was manageable, but it will get really challenging to manage when encoding gives many columns.

Sparse Matrix

A sparse matrix is a matrix that is comprised of mostly zero values. Sparse matrices are distinct from matrices with mostly non-zero values, which are referred to as dense matrices.

A matrix is sparse if many of its coefficients are zero. The interest in sparsity arises because its exploitation can lead to enormous computational savings and because many large matrix problems that occur in practice are sparse.

The sparsity of a matrix can be quantified with a score, which is the number of zero values in the matrix divided by the total number of elements in the matrix.

Sparsity = $\frac{\text{count of zero elements}}{\text{total elements}}$

Below is an example of a small 3 × 6 sparse matrix.

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

The example has 13 zero values of the 18 elements in the matrix, giving this matrix a sparsity score of 0.722 or about 72%.

In practice, most large matrices are sparse — almost all entries are zeros.

An example of a very large matrix that is too large to be stored in memory is a link matrix that shows the links from one website to another.

An example of a smaller sparse matrix might be a word or term occurrence matrix for words in one book against all known words in English.
In both cases, the matrix contained is sparse with many more zero values than data values.
The problem with representing these sparse matrices as dense matrices is that memory is required and must be allocated for each 32-bit or even 64-bit zero value in the matrix. This is clearly a waste of memory resources as those zero values do not contain any information.
Assuming a very large sparse matrix can be fit into memory, we will want to perform operations on this matrix. Simply, if the matrix contains mostly zero-values, i.e. no data, then performing operations across this matrix may take a long time where the bulk of the computation performed will involve adding or multiplying zero values together.
This is a problem of increased time complexity of matrix operations that increases with the size of the matrix. This problem is compounded when we consider that even trivial machine learning methods may require many operations on each row, column, or even across the entire matrix, resulting in vastly longer execution times.
Sparse matrices turn up a lot in applied machine learning. Some common examples to motivate you to be aware of the issues of sparsity.

Data

Sparse matrices come up in some specific types of data, most notably observations that record the occurrence or count of an activity. Three examples include:
Whether or not a user has watched a movie in a movie catalogue.
Whether or not a user has purchased a product in a product catalogue.
Count of the number of listens of a song in a song catalogue.

Data Preparation

Sparse matrices come up in encoding schemes used in the preparation of data. Three common examples include:
One hot encoding, used to represent categorical data as sparse binary vectors.
Count encoding, used to represent the frequency of words in a vocabulary for a document
TF-IDF encoding, used to represent normalized word frequency scores in a vocabulary.

Area of Study

Some areas of study within machine learning must develop specialized methods to address sparsity directly as the input data is almost always sparse. Three examples include:
Natural language processing for working with documents of text.
Recommender systems for working with product usage within a catalog.
Computer vision when working with images that contain lots of black pixels.
If there are 100,000 words in the language model, then the feature vector has length 100,000, but for a short email message almost all the features will have count zero.

Working with Sparse Matrices

The solution to representing and working with sparse matrices is to use an alternate data structure to represent the sparse data. The zero values can be ignored and only the data or non-zero values in the sparse matrix need to be stored or acted upon. There are multiple data structures that can be used to efficiently construct a sparse matrix; three common examples are listed below.

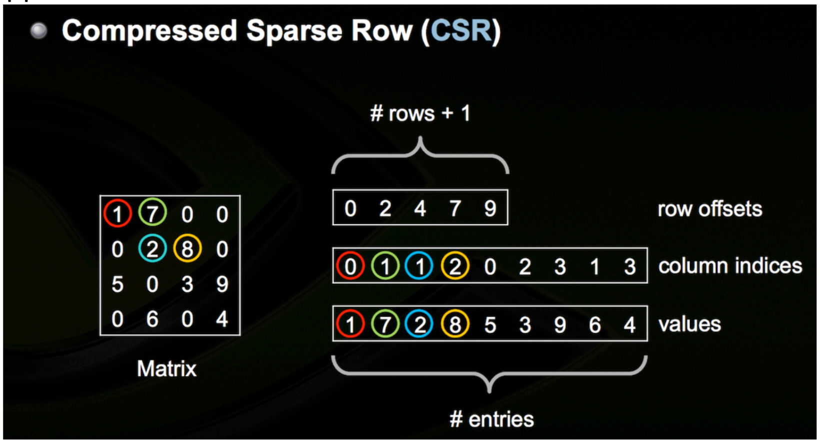
- Dictionary of Keys. A dictionary is used where a row and column index is mapped to a value.
- List of Lists. Each row of the matrix is stored as a list, with each sublist containing the column index and the value.
- Coordinate List. A list of tuples is stored with each tuple containing the row index, column index, and the value.

There are also data structures that are more suitable for performing efficient operations; two commonly used examples are listed below.

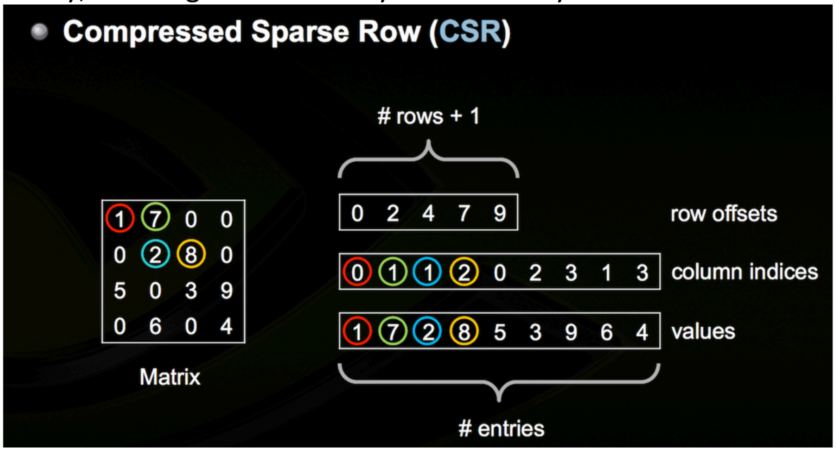
- Compressed Sparse Row. The sparse matrix is represented using three one-dimensional arrays for the non-zero values, the extents of the rows, and the column indexes.
- Compressed Sparse Column. The same as the Compressed Sparse Row method except the column indices are compressed and read first before the row indices.
-

Compressed Sparse Row

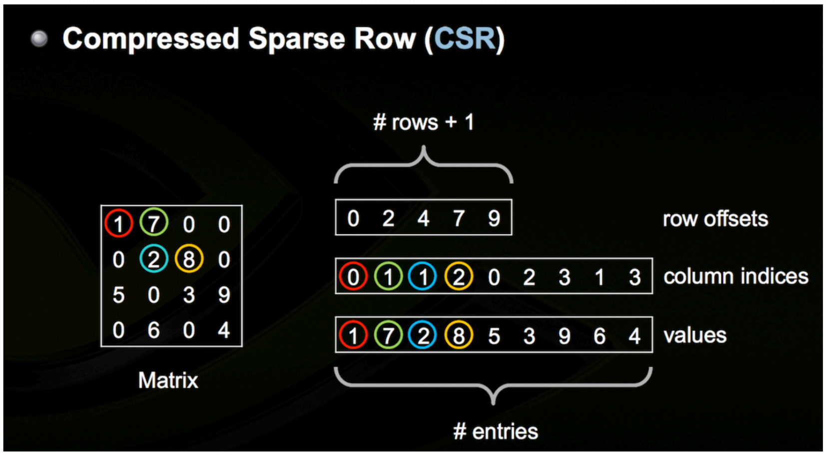
The Compressed Sparse Row, also called CSR for short, is often used to represent sparse matrices in machine learning given the efficient access and matrix multiplication that it supports.



The first step is to populate the first array. It always starts with 0. Since there are two nonzero values in row 1, we update our array like so [0 2]. There are 2 nonzero values in row 2, so we update our array to [0 2 4]. Doing that for the remaining rows yields [0 2 4 7 9]. By the way, the length of this array should always be the number of rows + 1.



Step two is populating the second array of column indices. Note that the columns are zero-indexed. The first value, 1, is in column 0. The second value, 7, is in column 1. The third value, 2, is in column 1. And so on. The result is the array [0 1 1 2 0 2 3 1 3].

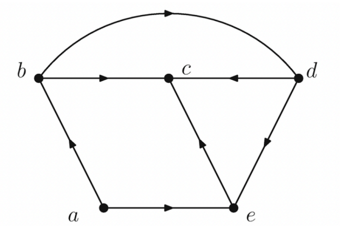


Finally, we populate the third array which looks like this [1 7 2 8 5 3 9 6 4]. Again, we are only storing nonzero values.

Matrix Algorithms in the Graph Theory

Graphs

A graph consists of a finite set of points (called vertices) and a finite set of edges, each of which connects two (not necessarily distinct) vertices.
Graph is a pair G = (V, E) , where V is the set of vertices , while E is the set of edges , connecting some pairs of vertices.
In directed graphs , the edges are the ordered pair of vertices, i.e. it is of importance which vertex is the beginning of the edge and which one is the end.
Directed graphs are also referred to as digraphs .
A drawing where the graph vertex is shown as points and the edges are shown as segments or arcs is called a graph diagram
Consider a digraph D(V, E), the set of vertices V and the set of edges E of which are specified as follows:
V = {a,b,c,d,e}
E= {ab,ae,bc,bd,dc,de,ec}



Two vertices v_i and v_j of the graph are **adjacent**, if they are connected by the edge *r* = $v_i\, v_j$.
In this case it is said that the vertices v_i and v_j are the **endpoints** of the edge *r*.
If the vertex *v* is the endpoint of the edge *r*, then *v* and *r* are considered to be **incident**
The number of elements (**cardinality**) of any set, for example *V* , is denoted as $|V|$.

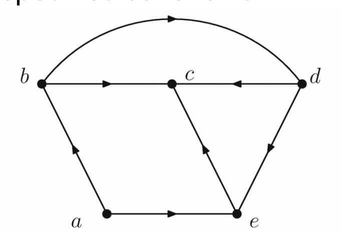
Adjacency matrix

Adjacency matrix *A* is a binary matrix of a relation over the set of vertices of the graph *G*(*V*, *E*), which is specifies by its edges. The adjacency matrix as the size $|V|\times|V|$, and its elements are determined in accordance with the rule

$$a(i,j) = \begin{cases} 1, \text{ if there is an edge between vertices } i \text{ and } j \\ 0, \ otherwise. \end{cases}$$

A path of length k in the graph G is a sequence of vertices $v_0, v_1, \ldots v_k$ such that $\forall\, i = 1, \ldots .\, k$ the vertices v_{i-1} and v_i are adjacent.
For undirected graphs , paths are also called routes .
The length of the path is the number of edges in it, taking into account the iterations.

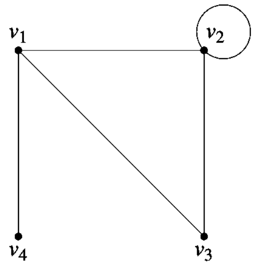
Example Consider a digraph D(V, E), the set of vertices V and the set of edges E of which are specified as follows:



The adjacency matrix A of the digraph D has the form:

$$A = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

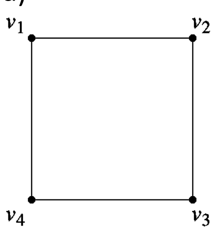
A diagonal entry a_{ij} of *A* is zero unless there is a loop at vertex *i*. In some situations, a graph may have more than one edge between a pair of vertices. In such cases, it may make sense to modify the definition of the adjacency matrix so that a_{ij} equals the *number* of edges between vertices *i* and *j*.



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

Example Determine the adjacency matrix of the given graph.

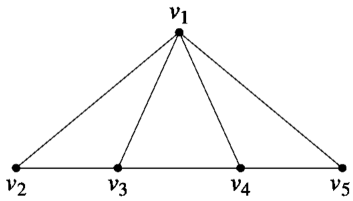
a)



Solution

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

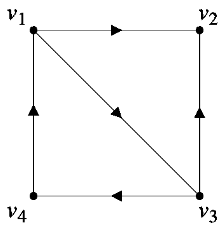
b)



Solution

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

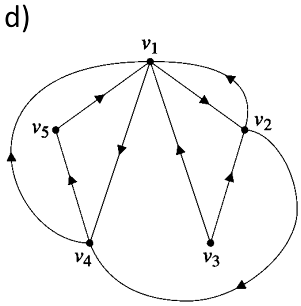
c)



Solution

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

d)



Solution

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

Length of Path

What do the entries of A^2 represent? Look at the (2, 3) entry.

$$A^2 = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 2 & 3 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

From the definition of matrix multiplication, we know that

$$(A^2)_{23} = a_{21}a_{13} + a_{22}a_{23} + a_{23}a_{33} + a_{24}a_{43}$$

The only way this expression can result in a nonzero number is if at least one of the products $a_{2k}\ a_{k3}$ that make up the sum is nonzero. But $a_{2k}a_{k3}$is nonzero if and only if both a_{2k} and a_{k3} are nonzero, which means that there is an edge between v_2 and v_k as well as an edge between v_k and v_3. Thus, there will be a 2-path between vertices 2 and 3 (via vertex *k*).

$$\begin{aligned}(A^2)_{23} &= a_{21}a_{13} + a_{22}a_{23} + a_{23}a_{33} + a_{24}a_{43} \\ &= 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 + 0 \cdot 0 \\ &= 2\end{aligned}$$

which tells us that there are two 2-paths between vertices 2 and 3.

If *A* is the adjacency matrix of a graph *G*, then the (*i*, *j*) entry of A^k is equal to the number of *k*-paths between vertices *i* and *j*.

Reachability matrix A^*

Reachability matrix A^* of the digraph *D*(*V*, *E*) stores the information about the existence of paths between the digraph vertices: at the intersection of the *i*-th row and the *j*-the column stands 1 when and only when there exists a path from the vertex v_i to v_j

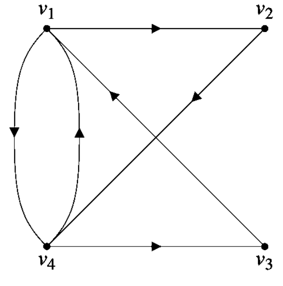
$\boldsymbol{A}^*$ may be calculated by a formula using the logical operation **or**

$\boldsymbol{A}^* = \boldsymbol{A}\ \boldsymbol{or}\ \boldsymbol{A}^2$ or …. $\boldsymbol{A}^n$

where *n* is the number of vertices of the directed graph, i.e. *n* = |*V* |

Note, that determining the elements of the matrix $\boldsymbol{A}^*$ by formula is associated with a considerable volume of calculations, this is why for the digraphs with a great number of vertices, the **Warshall algorithm** is used, also known as the **algorithm of Roy–Warshall**

Example How many 3-paths are there between v_1 and v_2 ?



Solution

The adjacency matrix is given by

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

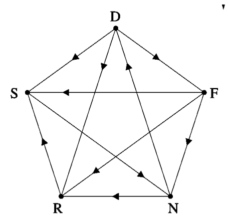
We need the (1, 2) entry of A^3, which is the dot product of row 1 of A^2 and column 2 of A.

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

$$A^3 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

So, there is one path between v_1 and v_2

Example Five tennis players (Djokovic, Federer, Nadal, Roddick, and Safin) compete in a round-robin tournament in which each player plays every other player once. The digraph in Figure summarizes the results. A directed edge from vertex *i* to vertex *j* means that player *i* defeated player *j*. (A digraph in which there is exactly one directed edge between every pair of vertices is called a *tournament*.)



Solution

The adjacency matrix for the digraph is

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

where the order of the vertices (and hence the rows and columns of A) is determined alphabetically. Thus, Federer corresponds to row 2 and column 2, for example.

Suppose we wish to rank the five players, based on the results of their matches. One way to do this might be to count the number of wins for each player. Observe that the number of wins each player had is just the sum of the entries in the corresponding row; equivalently, the vector containing all the row sums is given by the product Aj, where

$$\mathbf{j} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$A\mathbf{j} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

which produces the following ranking:

- First: Djokovic, Federer (tie)
- Second: Nadal
- Third: Roddick, Safin (tie)

Are the players who tied in this ranking equally strong?
Djokovic might argue that since he defeated Federer, he deserves first place. Roddick would use the same type of argument to break the tie with Safin. However, Safin could argue that he has two "indirect" victories because he beat Nadal, who defeated *two* others; furthermore, he might note that Roddick has only *one* indirect victory (over Safin, who then defeated Nadal).

Since in a group of ties there may not be a player who defeated all the others in the group, the notion of indirect wins seems more useful. Moreover, an indirect victory corresponds to a 2-path in the digraph, so we can use the square of the adjacency matrix.

To compute both wins and indirect wins for each player, we need the row sums of the matrix $A + A^2$, which are given by

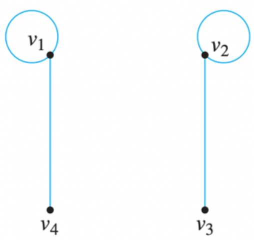
$$(A + A^2)\mathbf{j} = \left(\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 2 & 1 & 2 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

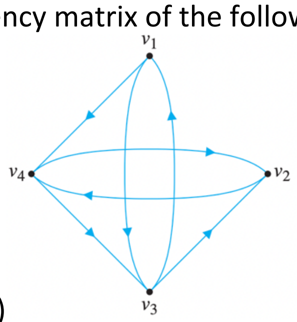
$$= \begin{bmatrix} 0 & 1 & 2 & 2 & 3 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 6 \\ 2 \\ 3 \end{bmatrix}$$

Thus, we would rank the players as follows: Djokovic, Federer, Nadal, Safin, Roddick.
Unfortunately, this approach is not guaranteed to break all ties.

Questions

1. Determine the adjacency matrix of the following.

a) 

b) 

2. Draw a digraph that has the given adjacency matrix.

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

3. Use powers of adjacency matrices to determine the number of paths of the length 3 between the vertices v_4 to v_1.

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

4. Let *A* be the adjacency matrix of a graph *G*.

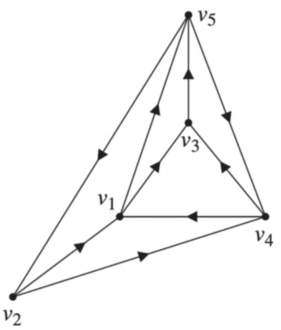
(a) If row *i* of *A* is all zeros, what does this imply about *G*?

(b) If column *j* of *A* is all zeros, what does this imply about *G*?

Answers

1.

$$a)\ \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad b)\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$



2.

3. 3

4a) There is no edge Vertex i to the other vertices.

4b) There is no edge the other vertices to Vertex j.

7. Solution of linear Equations

Linear Equations

A linear equation in the variables $x_1, x_2, \dots \dots., x_n$ is an equation that can be written in the form

$a_1x_1 + a_2x_2 + a_3x_3 + \cdots \dots. + a_nx_n = b$

where b and the coefficients of $x_1, x_2, \dots \dots., x_n$ are real or complex numbers

Example

$7x_1 + 5x_2 - 12x_3 = 4.5$

Homogeneous linear equations

A system of linear equations (or a linear system) is a collection of one or more linear equations involving the same variables $x_1, x_2, \dots \dots., x_n$

$a_{11}x_1 + a_{12}x_2 + \dots \dots. + a_{1n}x_n = b_1$

$a_{21}x_1 + a_{22}x_2 + \dots \dots. + a_{2n}x_n = b_2$

$a_{m1}x_1 + a_{m2}x_2 + \dots \dots. + a_{mn}x_n = b_m$

This linear system is called homogeneous if $b_1 = b_2 = \cdots = b_m = 0$ and nonhomogeneous, otherwise.

Example

$$\begin{aligned} x_1 - x_2 + {} & x_3 = 8 \\ x_1 \quad - {} & 4x_3 = 7 \end{aligned}$$

Solution of the system

A solution of the system is a list $\{s_1, s_2, \dots \dots., s_n\}$ of numbers that makes each equation a true statement when the values $s_1, s_2, \dots \dots., s_n$ are substituted for $x_1, x_2, \dots \dots., x_n$ respectively.

Example

{11, 4, 1} is a solution of the above equations because, when these values are substituted for $x_1, x_2, \dots \dots., x_n$, respectively, the equations simplify to 8 = 8 and 7 = 7.

The set of all possible solutions is called the solution set of the linear system.

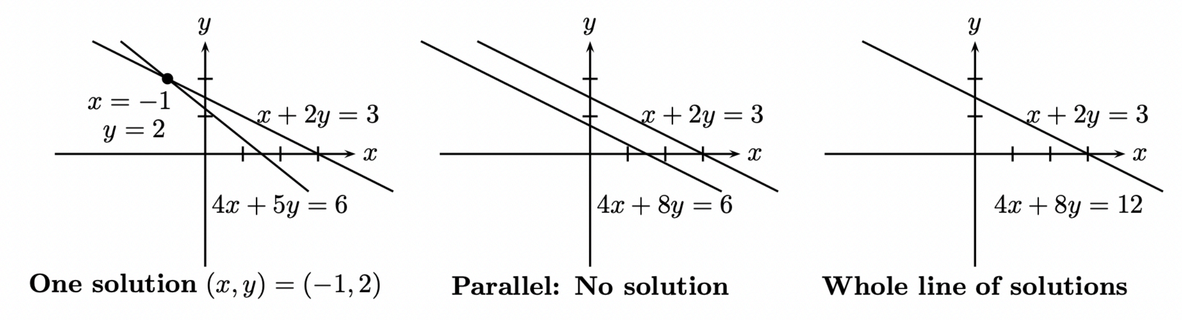
Two linear systems are called equivalent if they have the same solution set. That is, each solution of the first system is a solution of the second system, and each solution of the second system is a solution of the first.

A system of linear equations has

1. no solution
2. exactly ones solution
3. infinitely many solutions.

A system of linear equations is said to be consistent if it has either one solution or infinitely many solutions; a system is inconsistent if it has no solution.

Geometric representation of the solution of the system



Cramer's Rule

Let the system be given by

$a_{11}x_1 + a_{12}x_2 = b_1$ - (1)

$a_{21}x_1 + a_{22}x_2 = b_2$ -(2)

To solve for x_1

Multiply first equation by a_{22} and second by a_{12} we get

$a_{22}a_{11}x_1 + a_{22}a_{12}x_2 = a_{22}b_1$

$a_{12}a_{21}x_1 + a_{12}a_{22}x_2 = a_{12}b_2$

Therefore, we get

$(a_{22}a_{11} - a_{12}a_{21})x_1 = a_{22}b_1 - a_{12}b_2$

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{a_{22}a_{11} - a_{12}a_{21}}$$

$$x_1 = \frac{D_{x1}}{D} = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

D_{x1} : Determinant of the numerator in the solution of x_1

If we are solving for x_1, the column 1 is replaced with constants

To solve for x_2

Multiply first equation by a_{21} and second by a_{11} we get

$a_{21}a_{11}x_1 + a_{21}a_{12}x_2 = a_{21}b_1$

$a_{11}a_{21}x_1 + a_{11}a_{22}x_2 = a_{11}b_2$

Therefore, we get

$(a_{21}a_{12} - a_{11}a_{22})x_2 = a_{21}b_1 - a_{11}b_2$

$$x_2 = \frac{a_{21}b_1 - a_{11}b_2}{a_{21}a_{12} - a_{11}a_{22}}$$

$$x_2 = \frac{D_{x2}}{D} = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

D_{x2} : determinant of the numerator in the solution of x_2

If we are solving for x_2, the column 2 is replaced with constants

If $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \ and \ x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

Solving for 3 equations

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

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

$$\mathrm{D}\, x_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}$$
$$= b_1 \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} b_2 & a_{23} \\ b_3 & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} b_2 & a_{22} \\ b_3 & a_{32} \end{vmatrix}$$

$\mathrm{D}\, x_1$ is determinant of the numerator in the solution of x_1

$$\mathrm{D}\, x_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}$$
$$= a_{11} \begin{vmatrix} b_2 & a_{23} \\ b_3 & a_{33} \end{vmatrix} - b_1 \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & b_2 \\ a_{31} & b_3 \end{vmatrix}$$

$\mathrm{D}\, x_2$ is determinant of the numerator in the solution of x_2

$$\mathrm{D}\, x_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}$$
$$= a_{11} \begin{vmatrix} a_{22} & b_2 \\ a_{32} & b_3 \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & b_2 \\ a_{31} & b_3 \end{vmatrix} + b_1 \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$\mathrm{D\,x_3}$ is determinant of the numerator in the solution of $\mathrm{x_3}$

$$x_1 = \frac{D\,x_1}{D}, \ x_2 = \frac{D\,x_2}{D}, \ x_3 = \frac{D\,x_3}{D}$$

Example

$$x_1 + x_2 - x_3 = 6$$

$$3x_1 - 2x_2 + x_3 = -5$$

$$x_1 + 3x_2 + 2x_3 = 14$$

$$\mathrm{D} = \begin{vmatrix} 1 & 1 & -1 \\ 3 & -2 & 1 \\ 1 & 3 & -2 \end{vmatrix}$$

$$= 1\begin{vmatrix} -2 & 1 \\ 3 & -2 \end{vmatrix} - 1\begin{vmatrix} 3 & 1 \\ 1 & -2 \end{vmatrix} - 1\begin{vmatrix} 3 & -2 \\ 1 & 3 \end{vmatrix}$$

=1(4-3) – (-6-1) – (9+2) = 1+7-11 = -3

$$\mathrm{D}x1 = \begin{vmatrix} 6 & 1 & -1 \\ -5 & -2 & 1 \\ 14 & 3 & -2 \end{vmatrix}$$

$$= 6\begin{vmatrix} -2 & 1 \\ 3 & -2 \end{vmatrix} - 1\begin{vmatrix} -5 & 1 \\ 14 & -2 \end{vmatrix} - 1\begin{vmatrix} -5 & -2 \\ 14 & 3 \end{vmatrix}$$

= 6(4-3) – (10-14) – (-15+28)

= 6+4-13 = -3

$$\mathrm{D}x2 = \begin{vmatrix} 1 & 6 & -1 \\ 3 & -5 & 1 \\ 1 & 14 & -2 \end{vmatrix}$$

$$= 1\begin{vmatrix} -5 & 1 \\ 14 & -2 \end{vmatrix} - 6\begin{vmatrix} 3 & 1 \\ 1 & -2 \end{vmatrix} - 1\begin{vmatrix} 3 & -5 \\ 1 & 14 \end{vmatrix}$$

= 1(10-14)- 6(-6-1) – (42+5)

= -4+ 42 - 47= -9

$$\mathrm{D}x3 = \begin{vmatrix} 1 & 1 & 6 \\ 3 & -2 & -5 \\ 1 & 3 & 14 \end{vmatrix}$$

$$= 1\begin{vmatrix} -2 & -5 \\ 3 & 14 \end{vmatrix} - 1\begin{vmatrix} 3 & -5 \\ 1 & 14 \end{vmatrix} + 6\begin{vmatrix} 3 & -2 \\ 1 & 3 \end{vmatrix}$$

=1(-28+15) – (42+5) + 6(9 + 2)

= -13 – 47 +66 = 6

$$x_1 = \frac{\mathrm{D}\,x_1}{\mathrm{D}},\ x_2 = \frac{\mathrm{D}\,x_2}{\mathrm{D}},\ x_3 = \frac{\mathrm{D}\,x_3}{\mathrm{D}}$$

$$x_1 = \frac{-3}{-3},\ x_2 = \frac{-9}{-3},\ x_3 = \frac{6}{-3}$$

$$x_1 = 1,\ x_2 = 3,\ x_3 = -2$$

Gauss Elimination Process

We now start with solving a systems of linear equations. The idea is to manipulate the rows of the augmented matrix in place of the linear equations themselves. Since, multiplying a matrix on the left corresponds to row operations, we left multiply by certain matrices to the augmented matrix so that the final matrix is in row echelon form . The process of obtaining the row echelon form of a matrix is called the Gauss Elimination method.

The general Gaussian elimination procedure is applied to the linear systems:

$R_1 : a_{11}x_1 + a_{12}x_2 + \ldots \ldots . + a_{1n}x_n = b_1$

$R_2 : a_{21}x_1 + a_{22}x_2 + \ldots \ldots . + a_{2n}x_n = b_2$

…. : ….. + ……. = ..

$R_n : a_{n1}x_1 + a_{n2}x_2 + \ldots \ldots . + a_{nn}x_n = b_n$

Form the augmented matrix from the system of equations
The unknowns are eliminated to obtain an upper-triangular matrix.

To eliminate x_1 from R_2, we multiply R_1 by (-a_{21}/a_{11}) and obtain

$$-a_{21}x_1 - a_{12}\left(\frac{a_{21}}{a_{11}}\right)x_2 - \ldots \ldots . - a_{1n}\left(\frac{a_{21}}{a_{11}}\right)x_n = -b_1\left(\frac{a_{21}}{a_{11}}\right)$$

Adding the above equation to R_2 we obtain

$$\left(a_{22} - a_{12}\frac{a_{21}}{a_{11}}\right)x_2 - \left(a_{23} - a_{13}\frac{a_{21}}{a_{11}}\right)x_3 \ldots \ldots . - \left(a_{2n} - a_{1n}\frac{a_{21}}{a_{11}}\right)x_n = b_2 - b_1\left(\frac{a_{21}}{a_{11}}\right)$$

R_2 can be rewritten as

$R_2 : a'_{22}x_2 + a'_{23}x_3 + \cdots \ldots . + a'_{2n}x_n = b'_2$

Where $a'_{22} = \left(a_{22} - a_{12}\frac{a_{21}}{a_{11}}\right)$ and so on.

In a similar fashion, we can eliminate x_1 from the remaining equations and after eliminating x_1 from the last row Rn, we obtain the system

$R_1 : a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots \ldots . + a_{1n}x_n = b_1$

$R_2 : \quad a'_{22}x_2 + a'_{23}x_3 + \cdots \ldots . + a'_{2n}x_n = b'_2$

$R_n : \quad a'_{n2}x_2 + a'_{n3}x_3 + \cdots \ldots . + a'_{nn}x_n = b'_n$

In the process of obtaining the above system, we have multiplied the first row by (-a_{21}/a_{11}), i.e. we have divided it bya_{11} which is therefore assumed to be nonzero. For this reason, the first row R_1 in is called the pivot equation, and a_{11} is called the pivot or pivotal element. The method obviously fails if a_{11}= 0.

Similarly, we eliminate the variables will be obtain the upper-triangular matrix in the form:

$R_1 : a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots \ldots . + a_{1n}x_n = b_1$

$R_2 : \qquad a'_{22}x_2 + a'_{23}x_3 + \cdots \ldots . + a'_{2n}x_n = b'_2$

$R_3 : \qquad a''_{33}x_3 + \cdots \ldots . + a''_{3n}x_n = b''_3$

$R_n : \qquad a_{nn}^{(n-1)}x_n = b_n^{(n-1)}$

where $a_{nn}^{(n-1)}$ indicates the element a_{nn} has changed (n-1) times.

From R_n : $a_{nn}^{(n-1)}x_n = b_n^{(n-1)}$

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}}$$

This is then substituted in the $R_{(n-1)}$ to obtain x_{n-1} and the process is repeated to compute the other unknowns. We have therefore first computed x_n *then* $x_{n-1}, \ldots \ldots x_2, x_1$ in that order. Due to this reason, the process is called **back substitution**.

Example

Find the solution of the following system of equations using Gauss Elimination

$$\begin{aligned} x_2 + x_3 &= 2 \\ 2x_1 + 3x_3 &= 5 \\ x_1 + x_2 + x_3 &= 3 \end{aligned}$$

Solution

The augmented matrix can be written as

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{array}\right]$$

Interchange R_2 and R_1 to get

$$\left[\begin{array}{ccc|c} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{array}\right]$$

Replace R_3 by $R_3 - {}^1\!/_2 R_1$ to get

$$\left[\begin{array}{ccc|c} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1-({}^1\!/_2)2 & 1-({}^1\!/_2)0 & 1-({}^1\!/_2)3 & 3-({}^1\!/_2)5 \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & -{}^1\!/_2 & {}^1\!/_2 \end{array}\right]$$

Replace R_3 by $R_3 - R_2$ to get

$$\left[\begin{array}{ccc|c} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 0-0 & 1-1 & -(1/2)-1 & (1/2)-2 \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -3/2 & -3/2 \end{array}\right]$$

The matrix is in row echelon form. Using the last row we get $x_3 = 1$
Second row of the matrix gives us $x_2 + x_3 = 2$ So, $x_2 = 1$
First row gives us $2x_1 + 3x_3 = 5$ So $x_1 = 1$

Free Variables and General Solutions

Example

The augmented matrix of a linear system has been changed into the equivalent *reduced* echelon form

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned} x_1 \qquad - 5x_3 &= 1 \\ x_2 + \ x_3 &= 4 \\ 0 &= 0 \end{aligned}$$

The variables x_1 and x_2 corresponding to pivot columns in the matrix are called **basic variables**. The other variable , x_3 , is called a **free variable**.
We can write the above equations in the form

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases}$$

By saying that x_3 is “free,” we mean that we are free to choose any value for , x_3. Once that is done, we determine the values for x_1 and x_2. For instance, when , x_3 = 0, the solution is (1, 4, 0); when , x_3 = 1, the solution is (6, 3, 1). *Each different choice of* , x_3 *determines a (different) solution of the system, and every solution of the system is determined by a choice of* , x_3.
The solution is called a **general solution** of the system because it gives an explicit description of *all* solutions.

Example

$$\begin{aligned} x_1 + 6x_2 \qquad + 3x_4 \qquad &= 0 \\ x_3 - 4x_4 \qquad &= 5 \\ x_5 &= 7 \end{aligned}$$

The pivot columns of the matrix are 1, 3, and 5, so the basic variables are x_1, x_3, and x_5 The remaining variables, x_2 and x_4, must be free.
We Obtain the general solution as

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases}$$

Example

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 + 2x_3 &= 5 \\ 3x_1 + 4x_2 + 4x_3 &= 11 \end{aligned}$$

Solution

The augmented matrix can be written as

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 5 \\ 3 & 4 & 4 & 11 \end{array}\right]$$

Replace R_2 by $R_2 - R_1$ and R_3 by $R_3 - 3R_1$ to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1-1 & 2-1 & 2-1 & 5-3 \\ 3-3 & 4-3 & 4-3 & 11-3(3) \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{array}\right]$$

Replace R_3 by R_3 - R_2 to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Now back substitution gives us

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_2 + x_3 &= 2 \end{aligned}$$

Since there are 3 unknowns but only 2 constraints
The system has infinite number of solutions

Consistent systems

A linear system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column—that is, if and only if an echelon form of the augmented matrix has *no* row of the form
[0 ⋯ 0 b] with b nonzero

If a linear system is consistent, then the solution set contains either (i) a unique solution, when there are no free variables, or (ii) infinitely many solutions, when there is at least one free variable.

Inconsistent Systems

A system is inconsistent if it has no solution.

Example

$$\begin{aligned} x_1 - 2x_2 - x_3 + 3x_4 &= 0 \\ -2x_1 + 4x_2 + 5x_3 - 5x_4 &= 3 \\ 3x_1 - 6x_2 - 6x_3 + 8x_4 &= 2 \end{aligned}$$

The echelon form can be derived as

$$\left[\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -6 & 8 & 2 \end{array}\right] \sim \left[\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & -3 & -1 & 2 \end{array}\right]$$

$$\sim \left[\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{array}\right]$$

This echelon matrix shows that the system is *inconsistent*, because its rightmost column is a pivot column; the third row corresponds to the equation 0 = 5. There is no need to perform any more row operations. Note that the presence of the free variables in this problem is irrelevant because the system is inconsistent.

Pivoting

Complete Pivoting

We now come to the important case of the pivot being zero or very close to zero. If the pivot is zero, the entire process fails and if it is close to zero, round-off errors may occur. These problems can be avoided by adopting a procedure called pivoting. We search both columns and rows for the largest element, the procedure is called complete pivoting. It is obvious that complete pivoting involves more complexity in computations since interchange of columns means change of order of unknowns which invariably requires more programming effort.

Partial Pivoting

If a_{11} is either zero or very small compared to the other coefficients of the equation, then we find the largest available coefficient in the columns below the pivot equation and then interchange the two rows. In this way, we obtain a new pivot equation with a nonzero pivot. Such a process is called partial pivoting, since in this case we search only the columns below for the largest element In comparison to complete pivoting, partial pivoting, i.e. row interchanges, is easily adopted in programming. Due to this reason, complete pivoting is rarely used.

Example

Solve the following system of equations with and without pivoting

$0.0003120x_1$ +$0.006032x_2$ = 0.003328

$0.500000x_1$ + 0.89420 x_2= 0.9471

Solution
The exact solution is x_1 = 1 and x_2 = 0.5

We first solve the system with pivoting. We write the given system as

$$\left[\begin{array}{cc|c} 0.500000 & 0.89420 & 0.9471 \\ 0.0003120 & 0.006032 & 0.003328 \end{array}\right]$$

Replace R₂ by $R_2 - \left(\frac{0.500000}{0.0003120}\right) R_1$ to get

$$\left[\begin{array}{cc|c} 0.500000 & 0.89420 & 0.9471 \\ 0 & 0.005474 & 0.002737 \end{array}\right]$$

Back substitution gives us x_1 = 1 and x_2 = 0.5

Without pivoting, Gauss elimination gives

$$\left[\begin{array}{cc|c} 0.0003120 & 0.006032 & 0.003328 \\ 0 & -8.77725 & -5.3300 \end{array}\right]$$

Back substitution gives us x_1 = -1.0803 and x_2 = 0.6076

Gauss Seidel

We shall now describe the iterative or indirect methods, which start from an approximation to the true solution and, if convergent, derive a sequence of closer approximations- the cycle of computations being repeated till the required accuracy is obtained. This means that in a direct method the amount of computation is fixed, while in an iterative method the amount of computation depends on the accuracy required.

In general, one should prefer a direct method for the solution of a linear system, but in the case of matrices with a large number of zero elements, it will be advantageous to use iterative methods which preserve these elements.

Let the system be given by

$$R_1 : a_{11}x_1 + a_{12}x_2 + \ \ldots\ldots. + a_{1n}x_n = b_1$$

$$R_2 : a_{21}x_1 + a_{22}x_2 + \ \ldots\ldots. + a_{2n}x_n = b_2$$

$$R_n : a_{n1}x_1 + a_{n2}x_2 + \ \ldots\ldots. + a_{nn}x_n = b_n$$

in which the diagonal elements a_{ii} do not vanish. If this is not the case, then the equations should be rearranged so that this condition is satisfied.

Suppose $x_1^{(1)}, x_2^{(1)}, \ldots \ldots\ x_n^{(1)}$ are any first approximations to the unknowns $x_1, x_2, \ldots\ldots., x_n$

We rewrite the equations as

$$x_1^{(1)} = \left(\frac{b_1}{a_{11}}\right) - \left(\frac{a_{12}}{a_{11}}\right)x_2 - \left(\frac{a_{13}}{a_{11}}\right)x_3 \ldots\ldots. - \left(\frac{a_{1n}}{a_{11}}\right)x_n$$

$$x_2^{(1)} = \left(\frac{b_2}{a_{22}}\right) - \left(\frac{a_{21}}{a_{22}}\right)x_1 - \left(\frac{a_{23}}{a_{22}}\right)x_3 \ldots\ldots. - \left(\frac{a_{2n}}{a_{22}}\right)x_n$$

$$x_3^{(1)} = \left(\frac{b_3}{a_{33}}\right) - \left(\frac{a_{31}}{a_{33}}\right)x_1 - \left(\frac{a_{32}}{a_{33}}\right)x_2 \ldots\ldots. - \left(\frac{a_{2n}}{a_{33}}\right)x_n$$

$$x_n^{(1)} = \left(\frac{b_n}{a_{nn}}\right) - \left(\frac{a_{n1}}{a_{nn}}\right)x_1 - \left(\frac{a_{n2}}{a_{nn}}\right)x_2 \ldots\ldots. - \left(\frac{a_{n,n-1}}{a_{nn}}\right)x_{n-1}$$

We get the second approximations as

$$x_1^{(2)} = \left(\frac{b_1}{a_{11}}\right) - \left(\frac{a_{12}}{a_{11}}\right) x_2^{(1)} - \left(\frac{a_{13}}{a_{11}}\right) x_3^{(1)} \dots \dots . - \left(\frac{a_{1n}}{a_{11}}\right) x_n^{(1)}$$

Since, we already have $x_1^{(2)}$, the second estimates of x_1, we can write second estimate of x_2 as

$$x_2^{(2)} = \left(\frac{b_2}{a_{22}}\right) - \left(\frac{a_{21}}{a_{22}}\right) x_1^{(2)} - \left(\frac{a_{23}}{a_{22}}\right) x_3^{(1)} \dots \dots . - \left(\frac{a_{2n}}{a_{22}}\right) x_n^{(1)}$$

Since, we already have$x_1^{(2)}, x_2^{(2)}$ the second estimates of x_1 and x_2, we can write second estimate of x_3 as

$$x_3^{(2)} = \left(\frac{b_3}{a_{33}}\right) - \left(\frac{a_{31}}{a_{33}}\right) x_1^{(2)} - \left(\frac{a_{32}}{a_{33}}\right) x_2^{(2)} \dots \dots . - \left(\frac{a_{2n}}{a_{33}}\right) x_n$$

Since, we already have$x_1^{(2)}, x_2^{(2)}, x_{n-1}^{(2)}$ the second estimate of x_1 and x_2,….., we can write second estimate of $x_n$as

$$x_n^{(2)} = \left(\frac{b_n}{a_{nn}}\right) - \left(\frac{a_{n1}}{a_{nn}}\right) x_1^{(2)} - \left(\frac{a_{n2}}{a_{nn}}\right) x_2^{(2)} \dots \dots . - \left(\frac{a_{n,n-1}}{a_{nn}}\right) x_{n-1}^{(n-1)}$$

In this manner, we complete the first stage of iteration and the entire process is repeated till the values of $x_1, x_2, \dots \dots ., x_n$ are obtained to the accuracy required.
It is clear, therefore, that this method uses an improved component as soon as it is available and it is called the method of successive displacements, or the Gauss-Seidel method.

Diagonal Dominance

The Gauss-Seidel methods converge, for any choice of the first approximation $x_j^{(1))}$ (*j* = 1, 2, ..., *n*), if every equation of the system satisfies the condition that the sum of the absolute values of the coefficients $\left(\frac{a_{ij}}{a_{ii}}\right)$ is almost equal to, or in at least one equation less than unity, i.e. provided that

$$\sum_{j=1, j\neq 1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \leq 1, (i = 1, 2, \dots \dots, n)$$

where the < sign should be valid in the case of ‘at least’ one equation.

Stopping criteria for iterations

1. Number of iterations
2. Calculate the Absolute Relative Approximate Error

$$|\epsilon_a| = \left|\frac{x_i^{new} - x_i^{old}}{x_i^{new}}\right| * 100$$

The iterations are stopped when the absolute relative approximate error is less than a prespecified tolerance for all unknowns

Example

Solve using Gauss Seidel method. Take initial approx. as $x_1 = 1.5, x_2 = 2, x_3 = 3$

$$\begin{aligned} 20x_1 + 2x_2 + x_3 &= 30 \\ x_1 - 40x_2 + 3x_3 &= -75 \\ 2x_1 - x_2 + 10x_3 &= 30 \end{aligned}$$

Solution

Check for diagonal dominance

$|a_{11}| = 20 \geq |a_{12}| + |a_{13}| = 2 + 1 = 3$
$|a_{22}| = 40 \geq |a_{21}| + |a_{23}| = 1 + 3 = 4$
$|a_{33}| = 10 \geq |a_{31}| + |a_{32}| = 2 + 1 = 3$

We rewrite the equations as

$$x_1 = \frac{1}{20}(30 - 2x_2 - x_3)$$
$$x_2 = \frac{1}{40}(75 + x_1 + 3x_3)$$
$$x_3 = \frac{1}{10}(30 - 2x_1 + x_2)$$

First iteration gives us

$$x_1^{(1)} = \frac{1}{20}(30 - 2(2) - 3) = 1.15$$
$$x_2^{(1)} = \frac{1}{40}(75 + 1.15 + 3(3)) = 2.14$$
$$x_3^{(1)} = \frac{1}{10}(30 - 2(1.15) + 2.14) = 2.98$$

Second iteration gives us

$$x_1^{(2)} = \frac{1}{20}(30 - 2(2.14) - 2.98) = 1.137$$
$$x_2^{(2)} = \frac{1}{40}(75 + 1.137 + 3(2.98)) = 2.127$$
$$x_3^{(2)} = \frac{1}{10}(30 - 2(1.137) + 2.127) = 2.986$$

Third iteration gives us

$$x_1^{(3)} = \frac{1}{20}(30 - 2(2.127) - 2.986) = 1.138$$
$$x_2^{(3)} = \frac{1}{40}(75 + 1.138 + 3(2.986)) = 2.127$$
$$x_3^{(3)} = \frac{1}{10}(30 - 2(1.138) + 2.127) = 2.985$$

The solution can be written as $x_1 = 1.14$, $x_2 = 2.13$, $x_3 = 2.98$

Example

$$10x_1 - 2x_2 - x_3 - x_4 = 3$$
$$-2x_1 + 10x_2 - x_3 - x_4 = 15$$
$$-x_1 - x_2 + 10x_3 - 2x_4 = 27$$
$$-x_1 - x_2 - 2x_3 + 10x_4 = -9$$

Take initial approx. as $x_1 = 0.3$, $x_2 = 1.5$, $x_3 = 2.7$, $x_4 = -0.9$

Solution
Check for diagonal dominance
$|a_{11}| = 10 \geq |a_{12}| + |a_{13}| + |a_{14}| = 2 + 1 + 1 = 4$

$|a_{22}| = 10 \geq |a_{21}| + |a_{23}| + |a_{24}| = 2 + 1 + 1 = 4$

$|a_{33}| = 10 \geq |a_{31}| + |a_{32}| + |a_{34}| = 1 + 1 + 2 = 4$

$|a_{44}| = 10 \geq |a_{41}| + |a_{42}| + |a_{43}| = 1 + 1 + 2 = 4$

We rewrite the equations as
$x_1 = (0.30 - 0.2x_2 + 0.1x_3 + 0.1x_4)$

$x_2 = (1.5 + 0.2x_1 + 0.1x_3 + 0.1x_4)$

$x_3 = (2.7 + 0.1x_1 + 0.1x_2 + 0.2x_4)$

$x_4 = (-0.9 + 0.1x_1 + 0.1x_2 + 0.2x_3)$

$|\epsilon_a| = \left|\frac{x_i^{new} - x_i^{old}}{x_i^{new}}\right|$ * 100

Iteration	x_1	$\|\epsilon_a\|$	x_2	$\|\epsilon_a\|$	x_3	$\|\epsilon_a\|$	x_4	$\|\epsilon_a\|$
1	0.72		1.824		2.774		-0.0196	
2	0.9403	23%	1.9635	7.15%	2.9864	7.1%	-0.0125	56.8%
3	0.9901	5%	1.9954	1.6%	2.9960	0.32%	-0.0023	443%
4	0.9984	0.83%	1.9990	0.18%	2.9993	0.11%	-0.0004	475%
5	0.9997	0.13%	1.9998	0.04%	2.9998	0.016%	-0.0003	33%
6	0.9998	0.01%	1.9998	0	2.9998	0	-0.0003	0
7	1.0000	0	2.0000	0.01%	3.0000	0.006%	0.0000	-

The solution can be written as $x_1 = 1.,\ x_2 = 2,\ x_3 = 3,\ x_4 = 0$

Gauss –Jordan Elimination Process

The Gauss-Jordan method is a variation of the Gauss Elimination method. In this method, the augmented coefficient matrix is transformed by row operations such that the coefficient matrix reduces to the Identity matrix. The solution of the system is then directly obtained as the reduced augmented column of the transformed augmented matrix. The process of obtaining the row reduced echelon form of a matrix is called the Gauss-Jordan Elimination method.
The Gaussian-Jordan elimination procedure is applied to the linear systems:
$R_1 : a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$

$R_2 : a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$

$R_3 : a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$

Form the augmented matrix from the system of equations

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_3 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right]$$

We assume that a_{11} is non-zero. If a_{11} =0, we can interchange rows so that a_{11} is non-zero in the resulting system.

The first step is to divide the first row by a_{11}

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right] \xrightarrow{\frac{R_1}{a_{11}}} \left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right]$$

$$a'_{12} = {a_{12}}/{a_{11}}, a'_{13} = {a_{13}}/{a_{11}}, b'_1 = {b_1}/{a_{11}}$$

Eliminate x_1 from 2nd and 3rd equation by row operations of

Replace R_2 by multiplying the reduced first row by a_{21} and subtracting from the second and

Replace R_3 by multiplying the reduced first row by a_{31} and subtracting from the third row.

$$\left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right] \xrightarrow{R_2 - R_1 a_{21}, R_3 - R_1 a_{31}} \left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ 0 & a'_{22} & a'_{23} & b'_2 \\ 0 & a'_{32} & a'_{33} & b'_3 \end{array}\right]$$

$$a'_{22} = a_{22} - a_{21}a'_{12}, and\ so\ on, b'_2 = b_2 - a_{21}b'_1\ and\ so\ on$$

Now considering a'_{22} as the non-zero pivot, we first divide the second row by a'_{22}

$$\left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ 0 & a'_{22} & a'_{23} & b'_2 \\ 0 & a'_{32} & a'_{33} & b'_3 \end{array}\right] \xrightarrow{{R_2}/{a'_{22}}} \left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & a'_{32} & a'_{33} & b'_3 \end{array}\right]$$

Where $a''_{23} = {a'_{23}}/{a'_{22}}, b''_2 = {b'_2}/{a'_{22}}$

Replace R_1 by Multiplying the reduced second row by a'_{12} and subtract it from the first row and

Replace R_3 by multiplying the reduced second row by a'_{32} and subtract it from the third row

$$\left[\begin{array}{ccc|c} 1 & a'_{12} & a'_{13} & b'_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & a'_{32} & a'_{33} & b'_3 \end{array}\right] \xrightarrow{R_1 - R_2 a'_{12}, R_3 - R_2 a'_{32}} \left[\begin{array}{ccc|c} 1 & 0 & a''_{13} & b''_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & 0 & a''_{33} & b''_3 \end{array}\right]$$

$$a''_{13} = a'_{13} - a'_{12}a''_{23}, and\ so\ on, b''_1 = b'_1 - a'_{12}b''_2\ and\ so\ on$$

Now considering a''_{33} as the non-zero pivot, we first divide the third row by a''_{33}

$$\left[\begin{array}{ccc|c} 1 & 0 & a''_{13} & b''_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & 0 & a''_{33} & b''_3 \end{array}\right] \xrightarrow{{R_3}/{a''_{33}}} \left[\begin{array}{ccc|c} 1 & 0 & a''_{13} & b''_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & 0 & 1 & b'''_3 \end{array}\right]$$

Where $b'''_3 = {b''_3}/{a''_{33}}$

Replace R_1 by Multiplying the reduced third row by a''_{13} and subtract it from the first row and Replace R_2 by multiply the reduced third row by a''_{23} and subtract it from the second row

$$\left[\begin{array}{ccc|c} 1 & 0 & a''_{13} & b''_1 \\ 0 & 1 & a''_{23} & b''_2 \\ 0 & 0 & 1 & b'''_3 \end{array}\right] \xrightarrow{R_1 - R_3 a''_{13}, R_2 - R_3 a''_{23}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & b'''_1 \\ 0 & 1 & 0 & b'''_2 \\ 0 & 0 & 1 & b''_3 \end{array}\right]$$

$b'''_1 = b''_1 - a''_{13} b'''_3$ *and* $b'''_2 = b''_2 - a''_{23} b'''_3$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & b'''_1 \\ 0 & 1 & 0 & b'''_2 \\ 0 & 0 & 1 & b''_3 \end{array}\right]$$

Finally the solution of the system is given by the reduced augmented column
i.e. $x_1 = b'''_1, x_2 = b'''_2, x_3 = b'''_3$

The advantage of using Gauss Jordan method is that it involves no labour of back substitution. Back substitution has to be done while solving linear equations formed during solving the problem.

The difference between Gaussian elimination and the Gaussian Jordan elimination is that one produces a matrix in row echelon form while the other produces a matrix in row reduced echelon form.

Solution of system of equations

While solving a system of equations three outcomes are possible

Unique Solution

If the reduced row echelon form has no free variables, then it looks like:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & b_1 \\ 0 & 1 & 0 & b_2 \\ 0 & 0 & 1 & b_3 \end{array}\right]$$

and there is a unique solution, namely, $x_1 = b_1$, $x_2 = b_2$, $x_3 = b_3$.

Example

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

A unique solution exists

$x_1 = 5$, $x_2 = 2$, $x_3 = 4$

Infinite Solutions (dependent system)

If the reduced row echelon form has free variables, then there are an infinite number of solutions. The parameter assigned to any one free variable can take on an infinite number of values.

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & -3 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The equations are $x_1 + 2x_3 = 3$ and $x_2 - 3x_3 = 4$

Solve for x_1 and x_2

$x_1 = 3 - 2x_3$ and $x_2 = 4 + 3x_3$
Thus, the solution is
$(3 - 2x_3 , 4 + 3x_3 , x_3)$

No Solution (inconsistent system)
If the reduced row echelon form has a row of the form [0,0,…,0,b] then the system of linear equations has no solution.

Example
$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 4 \end{array}\right]$$
Here we get 0 = 4 which is false. Hence, there is no solution for the system of equations

Example
$$2x_1 + 2x_2 + 4x_3 = 18$$
$$x_1 + 3x_2 + 2x_3 = 13$$
$$3x_1 + x_2 + 3x_3 = 14$$
The augmented matrix can be written as
$$\left[\begin{array}{ccc|c} 2 & 2 & 4 & 18 \\ 1 & 3 & 2 & 13 \\ 3 & 1 & 3 & 14 \end{array}\right]$$

The first step is to divide the first row by 2 gives us
Replace R_1 by (1/2) R_1
$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 1 & 3 & 2 & 13 \\ 3 & 1 & 3 & 14 \end{array}\right]$$

Subtract the reduced first row from the second row and also multiply the first row by 3 and then subtract from the third, gives us
Replace R_2 by $R_2 - R_1$ and R_3 by $R_3 - 3R_1$ to get
$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 2 & 0 & 4 \\ 0 & -2 & -3 & -13 \end{array}\right]$$

The second step is to divide the second row by 2 gives us
Replace R_2 by (1/2) R_2
$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & 0 & 2 \\ 0 & -2 & -3 & -13 \end{array}\right]$$

We reduce the second column to [0,1,0] by row operations
Replace R_1 by $R_1 - R_2$ and R_3 by $R_3 + 2R_2$ to get
$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 7 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & -3 & -9 \end{array}\right]$$

The third step is to divide the third row by -3 gives us
Replace R_3 by (-1/3) R_2

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 7 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Subtract from the first row, elements of the third row multiplied by 2
Replace R_1 by $R_1 - 2R_3$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

The solution is $x_1 = 1,\ x_2 = 2,\ x_3 = 3$

Comparing Iterative and reduction methods

When are iterative methods (Gauss-Seidel) useful? A major advantage of Gauss-Seidel is that roundoff errors are not given a chance to "accumulate," as they are in Gaussian elimination and the Gauss Jordan Method, because each iteration essentially creates a new approximation to the solution. The only roundoff error that we need to consider with Gauss-Seidel method is the error involved in the most recent step.

Also, in many applications, the coefficient matrix for a given system contains a large number of zeroes (sparse matrix). When a linear system has a sparse matrix, each equation in the system may involve very few variables. If so, each step of the Gauss-Seidel process is relatively easy. However, neither the Gauss-Jordan Method nor Gaussian elimination would be very attractive in such a case because the cumulative effect of many row operations would tend to replace the zero coefficients with nonzero numbers. But even if the coefficient matrix is not sparse, Gauss-Seidel methods often give more accurate answers when large matrices are involved because fewer arithmetic operations are performed overall.

On the other hand, when Gauss-Seidel method take an extremely large number of steps to stabilize or do not stabilize at all (absence of diagonal dominance), it is much better to use the Gauss-Jordan Method or Gaussian elimination.

Ill-conditioned matrix

System of equations, in which a very small change in a coefficient leads to a very large change in the solution set, are called ill-conditioned systems.
Ill- conditioned or Nearly singular matrix—an invertible matrix that can become singular if some of its entries are changed ever so slightly. In this case, row reduction may produce fewer than n pivot positions, as a result of roundoff error. Also, roundoff error can sometimes make a singular matrix appear to be invertible.

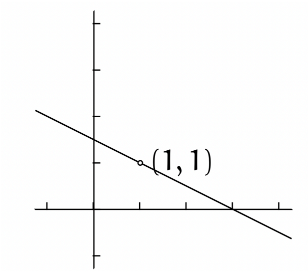
Example

Consider the following equations(the extra digits are in the ninth significant place)

$$\begin{aligned} x_1 + \quad 2x_2 &= 3 \\ 1.000\ 000\ 01x_1 + \quad 2x_2 &= 3.000\ 000\ 01 \end{aligned}$$

The solution is x_1 = 1, x_2 = 1. A computer has more trouble. If it represents real numbers to eight significant places, called single precision, then it will represent the second equation internally as
1000 000 0 x_1 + 2 x_2 = 3.000 000 0,
losing the digits in the ninth place. Instead of reporting the correct solution, this computer will think that the two equations are equal and it will report that the system is singular.

Consider this graph of this system.



We cannot tell the two lines apart; this system is nearly singular in the sense that the two lines are nearly the same line. This gives the system the property that a small change in an equation can cause a large change in the solution.
For instance, changing the 3.000 000 01 to 3.000 000 03 changes the intersection point from (1, 1) to (3, 0). The solution changes radically depending on the ninth digit, which explains why an eight-place computer has trouble. A problem that is very sensitive to inaccuracy or uncertainties in the input values is ill-conditioned
The above example gives one way in which a system can be difficult to solve on a computer. It has the advantage that the picture of nearly-equal lines gives a memorable insight into one way for numerical difficulties to happen. Unfortunately this insight isn't useful when we wish to solve some large system. We typically will not understand the geometry of an arbitrary large system .

Example
Consider the following equations

$$\begin{aligned} 0.001x_1 + \quad x_2 &= 1 \\ x_1 - \quad x_2 &= 0 \end{aligned}$$

The second equation gives x_1 = x_2 , so x_1 = x_2 = 1/1.001 and thus both variables have values that are just less than 1. A computer using two digits represents the system internally in this way (we will do this example in two-digit floating point arithmetic for clarity but inventing a similar one with eight or more digits is easy).

$$(1.0 \times 10^{-3})x_1 + (1.0 \times 10^{0})x_2 = (1.0 \times 10^{0})$$

$$(1.0 \times 10^{0})x_1 - (1.0 \times 10^{0})x_2 = (1.0 \times 10^{0})$$

The row reduction step −1000 R_1 + R_2 produces a second equation −1001 x_2 = −1000, which this computer rounds to two places as

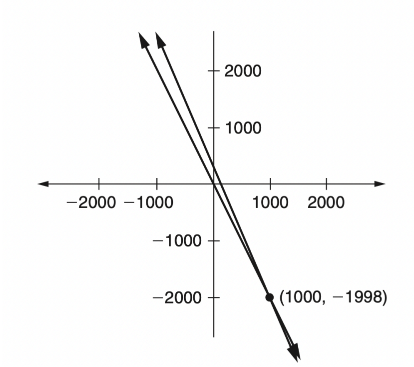
$$(-1.0 \times 10^{3})x_2 = (-1.0 \times 10^{3})$$

The computer decides from the second equation that x_2 = 1 and with that it concludes from the first equation that x_1 = 0.
The x_2 value is close but the x_1 is incorrect

Another cause of unreliable output is the computer's reliance on floating point arithmetic when the system-solving code leads to using leading entries that are small.
An experienced programmer may respond by using double precision, which retains sixteen significant digits, or perhaps using some even larger size. This will indeed solve many problems. However, double precision has greater memory requirements and besides we can obviously tweak the above to give the same trouble in the seventeenth digit, so double precision isn't a panacea. We need a strategy to minimize numerical trouble as well as some guidance about how far we can trust the reported solutions.

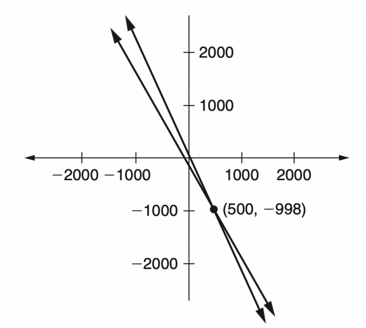
Example
Consider the following similar systems
A:

$$\begin{aligned} 2x_1 + \quad x_2 &= 2 \\ 2.005x_1 + \quad x_2 &= 7 \end{aligned}$$



B:

$$\begin{aligned} 2x_1 + \quad x_2 &= 2 \\ 2.01x_1 + \quad x_2 &= 7 \end{aligned}$$



Even though the coefficients of systems (A) and (B) are almost identical, the solutions to the systems are very different.
Solution to (A) = (1000, -1998) and solution to (B) = (500, -998).
In this case, there is a geometric way to see that these systems are ill-conditioned; the pair of lines in each system are almost parallel. Therefore, a small change in one line can move the point of intersection very far along the other line.
Suppose the coefficients in system (A) had been obtained after a series of long calculations.

Condition Number

A slight difference in the roundoff error of those calculations could have led to a very different final solution set. Thus, we need to be very careful when working with ill-

conditioned systems. Special methods have been developed for recognizing ill-conditioned systems, and a technique known as iterative refinement is used when the coefficients are known only to a certain degree of accuracy.
The degree of ill-conditioning of a matrix is measured by its *condition number.*
Some matrix programs will compute a condition number for a square matrix. The larger the condition number, the closer the matrix is to being singular. The condition number of the identity matrix is 1.
A singular matrix has an infinite condition number. In extreme cases, a matrix program may not be able to distinguish between a singular matrix and an ill-conditioned matrix.

Why do we need these methods in machine learning?

Machine learning deals a lot with data
This data is at times represented in the form of equations
We need to solve these equations to arrive at a solution of our problem
Many applications require the inverse of matrices

Questions

1. Solve using Gauss Elimination

$$\begin{aligned} x_1 - 2x_2 - x_3 + 3x_4 &= 0 \\ -2x_1 + 4x_2 + 5x_3 - 5x_4 &= 3 \\ 3x_1 - 6x_2 - 6x_3 + 8x_4 &= 2 \end{aligned}$$

2. Solve using Gauss Seidel Method

$a)$ $7x_1 - x_2 = 5$
$3x_1 + 5x_2 = -7$
Starting from (0,0)

$b)$ $x_1 - x_2 = 1$
$2x_1 + x_2 = 5$
Starting from (0,0)

3. The augmented matrix of a linear system has been transformed by row operations into the form below. Determine if the system is consistent.

$$\begin{bmatrix} 1 & 5 & 2 & -6 \\ 0 & 4 & -7 & 2 \\ 0 & 0 & 5 & 0 \end{bmatrix}$$

4. Solve using Gauss Jordan Method

a) $\begin{bmatrix} 0 & 1 & -6 & 5 \\ 1 & -2 & 7 & -6 \end{bmatrix}$

$b)$

$$\begin{aligned} x_1 + 2x_2 + 4x_3 &= -2 \\ x_2 + 5x_3 &= 2 \\ -2x_1 - 4x_2 - 3x_3 &= 9 \end{aligned}$$

5. Use Cramer's rule to compute the solutions of the systems

$a)$ $5x_1 + 7x_2 = 3$
$2x_1 + 4x_2 = 1$

$b)$ $$\begin{aligned} 2x_1 + x_2 + x_3 &= 4 \\ -x_1 + 2x_3 &= 2 \\ 3x_1 + x_2 + 3x_3 &= -2 \end{aligned}$$

Answers

1.

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

The system is inconsistent

2.

a) (1,2) b) (2,1)

3. Yes. A solution exists, and it is unique.

4a) $$\begin{cases} x_1 = 4 + 5x_3 \\ x_2 = 5 + 6x_3 \\ x_3 \text{ is free} \end{cases}$$

4b) (0, -3, 1)

5a) (5/6, -1/6) 5b) (-4, 13, -1)

8. Matrix Subspaces

Span of a set of Matrices

Generally, then, if $A_1, A_2, \ldots A_k$ are matrices of the same size and $c_1, c_2, \ldots c_k$ are scalars, we may form the ***linear combination***

$c_1A_2 + \ c_2A_2, \ldots c_kA_k$

We will refer to $c_1, c_2, \ldots c_k$ as the ***coefficients*** of the linear combination

The ***span*** of a set of matrices is the set of all linear combinations of the matrices.

Example

Let $A_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $A_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

(a) Is $B = \begin{bmatrix} 1 & 4 \\ 2 & 1 \end{bmatrix}$ a linear combination of A_1, A_2, and A_3?

(b) Is $C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ a linear combination of A_1, A_2, and A_3?

Solution

(a) We want to find scalars $c_1, c_2, \ c_3$ such that

$c_1A_1 + \ c_2A_2 + \ c_3A_3 = B$

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

This can be written as

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

Now,

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

Gives $c_1 = 1, \ c_3 = 3$

$c_2 + c_3 = 1$ gives $c_2 = -2$

B is a linear combination of A_1 , A_2 , A_3

i.e. $1 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 1 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 1 \end{bmatrix}$

(b) We want to find scalars $c_1, c_2, \ c_3$ such that

$c_1A_1 + \ c_2A_2 + \ c_3A_3 = C$

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

This can be written as

$$\begin{aligned} c_2 + c_3 &= 1 \\ c_1 \quad + c_3 &= 2 \\ -c_1 \quad + c_3 &= 3 \end{aligned}$$

$c_2 \quad + \quad c_3 = 4$

Now, $c_2 \quad + c_3 = 1$
$and \quad c_2 + \quad c_3 = 4$
Gives, there is no solution for this system of equation.
C is not a linear combination of A_1 , A_2 , A_3

Subspace of R^n

A **subspace** of R^n is any set H in R^n that has three properties:

- The zero vector is in H.
- For each **u** and **v** in H, the sum **u** + **v** is in H.
- For each **u** in H and each scalar c, the vector c**u** is in H.

In words, a subspace is *closed* under addition and scalar multiplication.

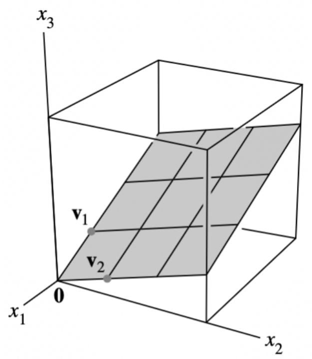
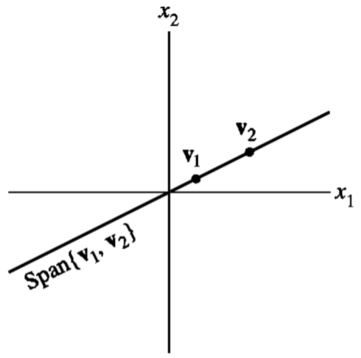


FIGURE 1
Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ as a plane through the origin.

Example 1
If v_1 and v_2 are in R^n and H = Span$\{v_1, v_2\}$, then H is a subspace of Rn.
Solution
To verify this statement, note that the zero vector is in H (because 0**v** + 0**u** is a linear combination of **u** and **v**).

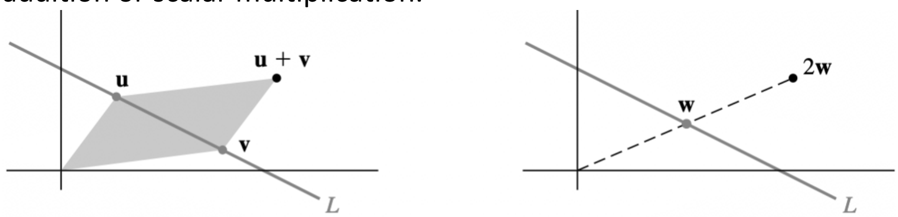
Now take two arbitrary vectors in H , say,
$\boldsymbol{u} = s_1v_1 + s_2v_2 \; and \; \boldsymbol{v} = t_1v_1 + \; t_2v_2$
Then
$u + v = (s_1 + t_1)v_1 + (s_2 + \; t_2)v_2$
which shows that **u** + **v** is a linear combination of v_1 and v_2 and hence is in H .
Also, for any scalar c, the vector c**u** is in H , because
$\boldsymbol{cu} = c(s_1v_1 + s_2v_2) = (cs_1)v_1 + (cs_2)v_2)$

Example
If v_1 is not zero and if v_2 is a multiple of v_1 , then v_1 and v_2 simply span a *line* through the origin. So a line through the origin is another example of a subspace.



$\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2 = k\mathbf{v}_1$.

Example
A line L *not* through the origin is *not* a subspace,
Solution
Because it does not contain the origin, as required. Also, Fig. 2 shows that L is not closed under addition or scalar multiplication.



$\mathbf{u} + \mathbf{v}$ is not on L $\qquad$ $2\mathbf{w}$ is not on L

FIGURE 2

Example
Let $v_1 = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}$, $v_2 = \begin{bmatrix} -4 \\ -5 \\ 8 \end{bmatrix}$, and w = $\begin{bmatrix} 8 \\ 2 \\ -9 \end{bmatrix}$. Determine if w is in the. subspace of R^3 generated by v_1 and v_2.
Solution
The vector w is in the subspace generated by v_1 and v_2 if and only if the vector equation $x_1 v_1 + x_2 v_2$ = w is consistent.

$[\,v_1 \quad v_2 \quad w\,] = \begin{bmatrix} 2 & -4 & 8 \\ 3 & -5 & 2 \\ -5 & 8 & -9 \end{bmatrix}$

Replace R_2 by R_2 - (3/2) R_1 and R_3 by R_3 + (5/2) R_1

$\begin{bmatrix} 2 & -4 & 8 \\ 0 & 1 & -10 \\ 0 & -2 & 11 \end{bmatrix}$

Replace R_3 by R_3 + 2 R_2

$\begin{bmatrix} 2 & -4 & 8 \\ 0 & 1 & -10 \\ 0 & 0 & -9 \end{bmatrix}$

The row operations above show that the solution is not consistent
Hence, w is not in the subspace generated by v_1 and v_2.

Example

Let $v_1 = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 3 \end{bmatrix}$, $v_2 = \begin{bmatrix} 4 \\ -7 \\ 9 \\ 7 \end{bmatrix}$, $v_3 = \begin{bmatrix} 5 \\ -8 \\ 6 \\ 5 \end{bmatrix}$ and u = $\begin{bmatrix} -4 \\ 10 \\ -7 \\ -5 \end{bmatrix}$. Determine if u is in the subspace of R^3 generated by v_1 , v_2 and v_3.

Solution

The vector w is in the subspace generated by v_1 , v_2 and v_3 if and only if the vector equation $x_1 v_1 + x_2 v_2 + x_3 v_3$ = u is consistent

$[\,v_1 \quad v_2 \quad v_3 \quad u]$ = $\begin{bmatrix} 1 & 4 & 5 & -4 \\ -2 & -7 & -8 & 10 \\ 4 & 9 & 6 & -7 \\ 3 & 7 & 5 & -5 \end{bmatrix}$

Replace R_2 by R_2 + 2 R_1 , R_3 by R_3 - 4 R_1 and R_4 by R_4 - 3 R_1

$$\begin{bmatrix} 1 & 4 & 5 & -4 \\ 0 & 1 & 2 & 2 \\ 0 & -7 & -14 & 9 \\ 0 & -5 & -10 & 7 \end{bmatrix}$$

Replace R_3 by R_3 + 7 R_2

$$\begin{bmatrix} 1 & 4 & 5 & -4 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 23 \\ 0 & 0 & 0 & 17 \end{bmatrix}$$

The row operations above show that the solution is not consistent
Hence, u is not in the subspace generated by v_1 $v_2$and v_3.

Column Space of a Matrix

The column space of a matrix A is the set Col A of all linear combinations of the columns of A.
If A = $[a_1 \quad \cdots \quad a_n]$ with the columns in R^m, then Col A is the same as Span $\{a_1 \quad \cdots \quad a_n\}$
In other words, for an n × d matrix A, its column space is defined as the vector space spanned by its columns, and it is a subspace of R^n.
The column space of an m×n matrix is a subspace of R^m .
A plane through the origin is the standard way to visualize the subspace

Example.

Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$, and b = $\begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Determine if b is in the the column space of A.

Solution

The vector b is a linear combination of the columns of A if and only if b can be written as Ax for some x, that is, if and only if the equation Ax = b has a solution.
Row reducing the augmented matrix [A b],

$$\left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ -4 & 6 & -2 & 3 \\ -3 & 7 & 6 & -4 \end{array}\right]$$

Replace R_2 by R_2 + 4 R_1 and R_3 by R_3 + 3 R_1

$$\left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & -2 & -6 & 5 \end{array}\right]$$

Replace R_3 by R_3 - (1/3) R_2

$$\left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

we conclude that Ax = b is consistent and b is in Col A.

Example

Let $v_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}$, $v_2 = \begin{bmatrix} -3 \\ 8 \\ -7 \end{bmatrix}$, $v_3 = \begin{bmatrix} -4 \\ 6 \\ -7 \end{bmatrix}$ and p = $\begin{bmatrix} 6 \\ -10 \\ 11 \end{bmatrix}$.

A = [$v_1 \quad v_2 \quad v_3$]

a. How many vectors are in { $v_1, \quad v_2, \quad v_3$ }?

b. How many vectors are in Col A?

c. Is p in Col A? Why or why not?

Solution

a. There are three vectors: v1, v2, and v3 in the set { $v_1, \quad v_2, \quad v_3$ }.

b. There are infinitely many vectors in Span { $v_1, \quad v_2, \quad v_3$ } = Col A.

c. The vector p is a linear combination of the columns of A if and only if p can be written as Ax for some x, that is, if and only if the equation Ax = p has a solution.

Row reducing the augmented matrix [A p],

$$\left[\begin{array}{ccc|c} 2 & -3 & -4 & 6 \\ -8 & 8 & 6 & -10 \\ 6 & -7 & -7 & 11 \end{array}\right]$$

Replace R_2 by R_2 + 4 R_1 and R_3 by R_3 - 3 R_1

$$\left[\begin{array}{ccc|c} 2 & -3 & -4 & 6 \\ 0 & -4 & -10 & 14 \\ 0 & 2 & 5 & -7 \end{array}\right]$$

Replace R_3 by R_3 + (1/2) R_2

$$\left[\begin{array}{ccc|c} 2 & -3 & -4 & 6 \\ 0 & -4 & -10 & 14 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

we conclude that Ax = p is consistent and p is in Col A.

Row Space

For an n × d matrix A, its row space is defined as the vector space spanned by the columns of A^T (which are simply the transposed rows of A). The row space of A is a subspace of R^d.

Example

Consider the matrix $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 3 & -3 \end{bmatrix}$.

a. Determine if b = $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is in the the column space of A.

b. Determine if b = $\begin{bmatrix} 4 & 5 \end{bmatrix}$ is in the the row space of A.

Solution

a. The vector b is a linear combination of the columns of A if and only if b can be written as Ax for some x, that is, if and only if the equation Ax = b has a solution.

Row reducing the augmented matrix [A b],

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 3 & -3 & 3 \end{array}\right]$$

Replace R_1 by $R_1 + R_2$, R_3 by R_3 - 3 R_1

$$\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right]$$

we conclude that Ax = b is consistent and b is in Col A.

b. If the vector w is in row(A), then w is a linear combination of the rows of A

Row reducing the augmented matrix $\begin{bmatrix} A \\ w \end{bmatrix}$

$$\left[\begin{array}{cc} 1 & -1 \\ 0 & 1 \\ 3 & -3 \\ \hline 4 & 5 \end{array}\right]$$

Replace R_3 by R_3 - 3 R_1 , R_4 by R_4 - 4 R_1

$$\left[\begin{array}{cc} 1 & -1 \\ 0 & 1 \\ 0 & 0 \\ \hline 0 & 9 \end{array}\right]$$

Replace R_4 by R_4 - 9 R_2

$$\left[\begin{array}{cc} 1 & -1 \\ 0 & 1 \\ 0 & 0 \\ \hline 0 & 0 \end{array}\right]$$

Therefore, w is a linear combination of the rows of A

Null Space of a matrix

The null space of a matrix A is the set Nul A of all solutions to the homogeneous equation Ax = 0.

The null space of an m×n matrix A is a subspace of R^n . Equivalently, the set of all solutions to a system Ax = 0 of m homogeneous linear equations in n unknowns is a subspace of R^n .

Example

Let $v_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}$, $v_2 = \begin{bmatrix} -3 \\ 8 \\ -7 \end{bmatrix}$, $v_3 = \begin{bmatrix} -4 \\ 6 \\ -7 \end{bmatrix}$ and p = $\begin{bmatrix} 6 \\ -10 \\ 11 \end{bmatrix}$.

A = $[\, v_1 \quad v_2 \quad v_3]$

Determine if p is in Nul A.

Solution

p is in Nul A if Ap = 0

Ap = $\begin{bmatrix} 2 & -3 & -4 \\ -8 & 8 & 6 \\ 6 & -7 & -7 \end{bmatrix}\begin{bmatrix} 6 \\ -10 \\ 11 \end{bmatrix} = \begin{bmatrix} -2 \\ -62 \\ 29 \end{bmatrix}$

Since Ap $\neq$ 0, p is not in Nul A.

Example

Let u = $\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$ and A = $\begin{bmatrix} -3 & -2 & 0 \\ 0 & 2 & -6 \\ 6 & 3 & 3 \end{bmatrix}$

Determine if u is in Nul A.

Solution

u is in Nul A if Au = 0

Au = $\begin{bmatrix} -3 & -2 & 0 \\ 0 & 2 & -6 \\ 6 & 3 & 3 \end{bmatrix}\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Since Ap $=$ 0, u is in Nul A.

Example

Let A = $\begin{bmatrix} 3 & 2 & 1 & -5 \\ -9 & -4 & 1 & 7 \\ 9 & 2 & -5 & 1 \end{bmatrix}$

Find a nonzero vector in Nul A

A nonzero vector in Col A.

b. To produce a vector in Col A, select any column of A.

Solution

u is in Nul A if Au = 0

Au = $\begin{bmatrix} 3 & 2 & 1 & -5 & 0 \\ -9 & -4 & 1 & 7 & 0 \\ 9 & 2 & -5 & 1 & 0 \end{bmatrix}$

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

The general solution is x_1 = x_3 $-x_4$, and x_2 = –2 x_3 + 4 x_4, with x_3 and x_4 free. The general solution in parametric vector form is not needed. All that is required here is one nonzero vector. So choose any values for x_3 and x_4 (not both zero). For instance, set x_3 = 1 and x_4 = 0 to obtain the vector (1, –2, 1, 0) in Nul A.

Left and Right Null Space

The notion of a null space refers to a right null space by default. This is because the vector x occurs on the right side of matrix A in the product Ax, which must evaluate to the zero vector. Similar to the definition of a right null space, one can define the left null space of a matrix, which is the orthogonal complement of the vector space spanned by the columns of the matrix.

The left null space of an n × d matrix A is the sub- space of R^n containing all column vectors $x \in R^n$, such that $A^T x = 0$. The left null space of A is the orthogonal complementary subspace of the column space of A. (Let V be a vector space and W be a subspace of V. Then the orthogonal complement of W in V is the set of vectors u such that u is orthogonal to all vectors in W.)

Alternatively, the left null space of a matrix A contains all vectors x satisfying $x^T A = 0^T$.

The row space, column space, the right null space, and the left null space are referred to as the four fundamental subspaces of linear algebra.

Basis for a Subspace

A **basis** for a subspace H of R^n is a linearly independent set in H that spans H .

The columns of an invertible n×n matrix form a basis for all of R^n because they are linearly independent and span R^n , by the Invertible Matrix Theorem. One such matrix is the n×n identity matrix. Its columns are denoted by $e_1 \quad \cdots \quad e_n$:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ .. \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ .. \\ 0 \end{bmatrix}, \ldots\ldots\; e_n = \begin{bmatrix} 0 \\ .. \\ 0 \\ 1 \end{bmatrix}$$

The set $\{e_1 \quad \cdots \quad e_n\}$ is called the **standard basis** for R^n .

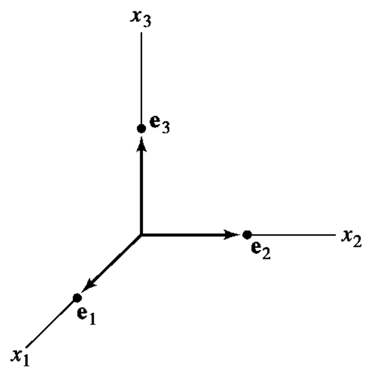


FIGURE 3
The standard basis for $\mathbb{R}^3$.

Example

Determine if the set is bases for R^2 or R^3.

$$\begin{bmatrix} 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 10 \\ -3 \end{bmatrix}$$

Solution

Yes. Let A be the matrix whose columns are the vectors given. Then A is invertible because its determinant is nonzero, and so its columns form a basis for R^2 , by the Invertible Matrix Theorem

Example
Determine if the set is bases for R^2 or R^3.
$\begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ -7 \\ 4 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}$

Solution
Let A be the matrix whose columns are the vectors given.
A = $\begin{bmatrix} 0 & 5 & 6 \\ 1 & -7 & 3 \\ -2 & 4 & 5 \end{bmatrix}$

Row reduce the matrix A to get
Interchange R_2 and R_1
$\begin{bmatrix} 1 & -7 & 3 \\ 0 & 5 & 6 \\ -2 & 4 & 5 \end{bmatrix}$

Replace R_3 by R_3 + 2 R_1
$\begin{bmatrix} 1 & -7 & 3 \\ 0 & 5 & 6 \\ 0 & -10 & 11 \end{bmatrix}$

Replace R_3 by R_3 + 2 R_2
$\begin{bmatrix} 1 & -7 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 23 \end{bmatrix}$

The matrix A has three pivots, so A is invertible by the Invertible Matrix Theorem and its columns form a basis for R^3

Example
Given matrix A and an echelon form of A.
a. Find a basis for Col A and
b. Find a basis for Nul A.

$$A = \begin{bmatrix} 4 & 5 & 9 & -2 \\ 6 & 5 & 1 & 12 \\ 3 & 4 & 8 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 6 & -5 \\ 0 & 1 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution
a. Basis for Col A

$$\begin{bmatrix} 1 & 2 & 6 & -5 \\ 0 & 1 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The echelon form identifies columns 1 and 2 as the pivot columns.
A basis for Col A uses columns 1 and 2 of A:

$$\begin{bmatrix} 4 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 4 \end{bmatrix}$$

b. Basis for Nul A

For Nul A, obtain the reduced (and augmented) echelon form for Ax = 0

$$\left[\begin{array}{cccc|c} 1 & 2 & 6 & -5 & 0 \\ 0 & 1 & 5 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Replace R_1 by R_1 - $2R_2$

$$\left[\begin{array}{cccc|c} 1 & 0 & -4 & 7 & 0 \\ 0 & 1 & 5 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1 & 0 & -4 & 7 & 0 \\ 0 & 1 & 5 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]. \text{ This corresponds to: } \begin{array}{rl} x_1 \quad - 4x_3 + 7x_4 &= 0 \\ x_2 + 5x_3 - 6x_4 &= 0. \\ 0 &= 0 \end{array}$$

Solve for the basic variables and write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4x_3 - 7x_4 \\ -5x_3 + 6x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 4 \\ -5 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -7 \\ 6 \\ 0 \\ 1 \end{bmatrix}. \text{ Basis for Nul } A\text{: } \begin{bmatrix} 4 \\ -5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ 6 \\ 0 \\ 1 \end{bmatrix}$$

Example

Given matrix A and an echelon form of A.

a. Find a basis for Col A and

b. a basis for Nul A.

$$A = \begin{bmatrix} -3 & 9 & -2 & -7 \\ 2 & -6 & 4 & 8 \\ 3 & -9 & -2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 6 & 9 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution

a. Basis for Col A

$$\begin{bmatrix} 1 & -3 & 6 & 9 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The echelon form identifies columns 1 and 3 as the pivot columns.

A basis for Col A uses columns 1 and 3 of A:

$$\begin{bmatrix} -3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix}$$

b. Basis for Nul A

For Nul A, obtain the reduced (and augmented) echelon form for Ax = 0

$$\left[\begin{array}{cccc|c} 1 & -3 & 6 & 9 & 0 \\ 0 & 0 & 4 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Replace R_1 by R_1 - $(\frac{6}{4})R_2$,

$$\begin{bmatrix} 1 & -3 & 0 & 1.5 & 0 \\ 0 & 0 & 4 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For Nul A, obtain the reduced (and augmented) echelon form for $A\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & -3 & 0 & 1.50 & 0 \\ 0 & 0 & 1 & 1.25 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \text{ This corresponds to: } \begin{aligned} x_1 - 3x_2 \quad + 1.50x_4 &= 0 \\ x_3 + 1.25x_4 &= 0. \\ 0 &= 0 \end{aligned}$$

Solve for the basic variables and write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3x_2 - 1.5x_4 \\ x_2 \\ -1.25x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1.5 \\ 0 \\ -1.25 \\ 1 \end{bmatrix}. \text{ Basis for Nul } A: \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1.5 \\ 0 \\ -1.25 \\ 1 \end{bmatrix}.$$

Example

Given matrix A and an echelon form of A.

a. Find a basis for Col A and

b. a basis for Nul A.

$$A = \begin{bmatrix} 1 & 4 & 8 & -3 & -7 \\ -1 & 2 & 7 & 3 & 4 \\ -2 & 2 & 9 & 5 & 5 \\ 3 & 6 & 9 & -5 & -2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution

a. Find a basis for Col A

$$\begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \text{ Basis for Col } A: \begin{bmatrix} 1 \\ -1 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 2 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \\ 5 \\ -5 \end{bmatrix}$$

b. Find a basis for Nul A

For Nul A, obtain the reduced (and augmented) echelon form for $A\mathbf{x} = \mathbf{0}$:

$$[A \quad \mathbf{0}] \sim \begin{bmatrix} 1 & 0 & -2 & 0 & 7 & 0 \\ 0 & 1 & 2.5 & 0 & -.5 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad \begin{aligned} x_1 \quad - 2x_3 \quad + 7x_5 &= 0 \\ x_2 + 2.5x_3 \quad - .5x_5 &= 0 \\ x_4 + 4x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The solution of $A\mathbf{x} = 0$ in parametric vector form: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_3 - 7x_5 \\ -2.5x_3 + .5x_5 \\ x_3 \\ -4x_5 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} 2 \\ -2.5 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -7 \\ .5 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$

Basis for Nul A: $\begin{bmatrix} 2 \\ -2.5 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -7 \\ .5 \\ 0 \\ -4 \\ 1 \end{bmatrix}$

The Dimension of a Subspace

The **dimension** of a nonzero subspace H, denoted by dimH, is the number of vectors in any basis for H . The dimension of the zero subspace {**0**} is defined to be zero.
The space R^n has dimension n. Every basis for R^n consists of n vectors. A plane through **0** in R^3 is two-dimensional, and a line through **0** is one-dimensional.

Example
Find the dimension of the matrix

$$A = \begin{bmatrix} 3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

Solution
First, write the solution of Ax = 0 in parametric vector form:

$$A \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{aligned} x_1 - 2x_2 \qquad - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

The null space of the matrix A has a basis of 3 vectors. So the dimension of Nul A in this case is 3. Observe how each basis vector corresponds to a free variable in the equation A**x** = **0**. Our construction always produces a basis in this way. So, to find the dimension of Nul A, simply identify and count the number of free variables in A**x** = **0**.

Rank of a matrix

The **rank** of a matrix A, denoted by rank A, is the dimension of the column space of A.

Since the pivot columns of A form a basis for Col A, the rank of A is just the number of pivot columns in A.

Example

Find the rank of the following matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$$

Solution

The matrix has a reduced row echelon form of

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

The matrix has 2 pivot columns. So, the rank of the matrix is 2

Questions

1. Given matrix A and an echelon form of A.

a. Find a basis for Col A and

b. A basis for Nul A.

c. state the dimensions of these subspaces

$$\begin{bmatrix} 1 & -3 & 2 & -4 \\ -3 & 9 & -1 & 5 \\ 2 & -6 & 4 & -3 \\ -4 & 12 & 2 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 & -4 \\ 0 & 0 & 5 & -7 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2. Let $v_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $v_3 = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$ and w = $\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$, u = $\begin{bmatrix} 8 \\ 4 \\ 7 \end{bmatrix}$.

a. Determine if w is in the subspace of R^3 generated by v_1 , v_2 and v_3.

b. Determine if u is in the subspace of R^3 generated by v_1 , v_2 and v_3.

3. Let u = $\begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$ and A = $\begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}$

Determine if u is in Nul A.

4. Let $A = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & -7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix}$, and b = $\begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}$, $u = \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix}$.

a. Determine if b is in the the column space of A.

b. Determine if u is in Nul A.

Answers

1a) $\begin{bmatrix} 1 \\ -3 \\ 2 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ -3 \\ 7 \end{bmatrix}.$ 1b) $\begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ 1c) dim Col A = 3, dim Nul A = 3

2a) $\begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & 1 & 2 & 1 \\ -1 & 3 & 6 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

The row operations above show that the system has a solution

Hence, w is in the subspace generated by v_1 , v_2 and v_3.

2b) $\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 2 & 4 \\ -1 & 3 & 6 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

The row operations above show that the solution is not consistent

Hence, w is not in the subspace generated by v_1 , v_2 and v_3.

3) Yes, u is in Nul A

4a) Yes 4b) No

9. Linear Transformation

A transformation (or function or mapping) T from R^n to R^m is a rule that assigns to each vector x in R^n a vector T(x) in R^m. The set R^n is called the domain of T, and R^m is called the codomain of T .

The notation T : $R^n \rightarrow R^m$ indicates that the domain of T is R^n and the codomain is R^m . For x in R^n , the vector T(x) in R^m is called the image of x (under the action of T). The set of all images T (x) is called the range of T.

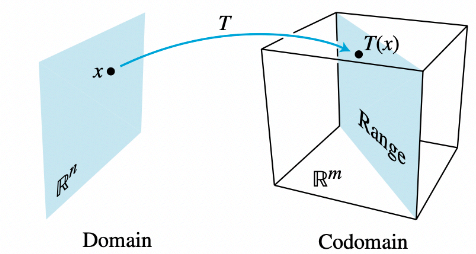


FIGURE 2 Domain, codomain, and range of $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Example

Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $u = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $b = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $c = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$,and define a transformation T : $R^2 \rightarrow R^3$ by T (**x**) = A**x**, so that

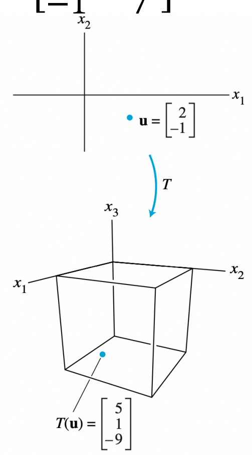
$$T(x) = Ax = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}. \text{ - (1)}$$

a. Find T(**u**), the image of **u** under the transformation T.
b. Find an **x** in R^2 whose image under T is **b**.
c. Is there more than one **x** whose image under T is **b**?
d. Determine if **c** is in the range of the transformation T.

Solution

a. Compute $T(u) = Au$

$$= \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$



b. Solve T(**x**) = **b** for **x**. That is, solve A**x** = **b**,

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$$

Row reduced augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & -0.5 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1.5 \\ 0 & 1 & -0.5 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence x1 = 1.5, x2 = −0.5,and **x** = $\begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix}$. The image of this **x** under T is the given vector **b**.

c. Any **x** whose image under T is **b** must satisfy (1). From b. it is clear that equation (1) has a unique solution. So there is exactly one **x** whose image is **b**.
d. The vector **c** is in the range of T if **c** is the image of some **x** in R^2, that is, if **c** = T(**x**) for some **x**. This is just another way of asking if the system A**x** = **c** is consistent. To find the answer, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 14 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -35 \end{bmatrix}$$

The third equation, 0 = −35, shows that the system is inconsistent.
So **c** is *not* in the range of T .

Example

Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$.

Find the images under T of $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$.

Solution

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$$

$$T(\mathbf{v}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a \\ 2b \end{bmatrix}$$

Example

Let $A = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

Define a transformation T : $R^2 \rightarrow R^3$ by T (**x**) = A**x**. Find T(**u**), T(**v**)

Solution

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix} = \begin{bmatrix} .5 \\ 0 \\ -2 \end{bmatrix}$$

$$T(\mathbf{v}) = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} .5a \\ .5b \\ .5c \end{bmatrix}$$

Example
With T defined by T (**x**) = A**x**, find a vector **x** whose image under T is **b**, and determine whether **x** is unique.

$$A = \begin{bmatrix} 1 & 0 & -2 \\ -2 & 1 & 6 \\ 3 & -2 & -5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ -3 \end{bmatrix}$$

Solution

$$[A \quad \mathbf{b}] = \begin{bmatrix} 1 & 0 & -2 & -1 \\ -2 & 1 & 6 & 7 \\ 3 & -2 & -5 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 2 & 5 \\ 0 & -2 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & 5 & 10 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \text{ unique solution}$$

Linear Transformation

A transformation (or mapping) T is linear if:
(i) T(u + v)=T(u) + T(v) for all u, v in the domain of T
(ii) T (cu) = cT(u) for all u and all scalars c.

If T is a linear transformation, then
T(0) = 0 and T (cu + dv) = cT (u) + dT (v)
for all vectors u, v in the domain of T and all scalars c, d.

Shear transformation

The transformation T : $R^2 \rightarrow R^2$ defined by T (**x**) = A**x** is called a **shear transformation**. It can be shown that if T acts on each point in the 2×2 square shown in Fig. 4, then the set of images forms the shaded parallelogram. The key idea is to show that T maps line segments onto line segments and then to check that the corners of the square map onto the vertices of the parallelogram.

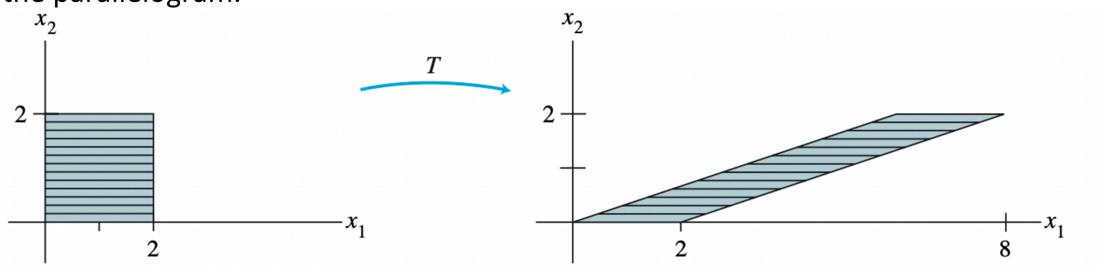


FIGURE 4 A shear transformation.

Example

Let $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$, $u = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\mathrm{v} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

Solution

$T(u) = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ $and\ T(v) = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$

T deforms the square as if the top of the square were pushed to the right while the base is held fixed.

Contraction and Dilation

Given a scalar r, define T : $R^2 \to R^2$ by T(x) = rx. T is called a contraction when 0 ≤ r ≤ 1 and a dilation when r > 1.

Example

Let r = 3, show that T is a linear transformation.

Solution

Let u, v be in R^2 and let c, d be scalars. Then

T(cu + dv) = 3(cu + dv)

= 3cu + 3dv

= c(3u) + d(3v) = cT(u)+dT(v)

Thus T is a linear transformation

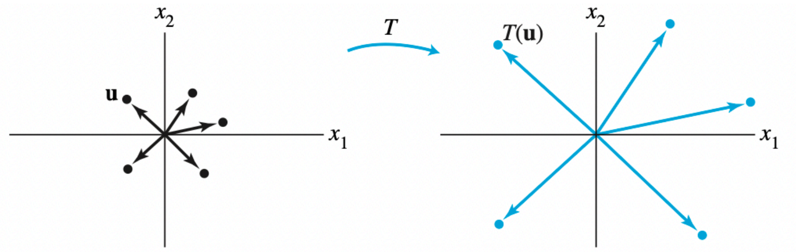


FIGURE 5 A dilation transformation.

Rotation

The transformation T : $R^2 \to R^2$ defined by

$T(u) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$

Example

Find, the image under T of $u = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathrm{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ $u + v = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$

Solution

$T(u) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$ $and\ T(v) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$

$T(u + v) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$

Note that T (u + v) is obviously equal to T (u) + T (v). It appears from Fig. 6 that T rotates u, v, and u + v counter clockwise about the origin through 90◦. In fact, T transforms the entire parallelogram determined by u and v into the one determined by T (u) and T (v).

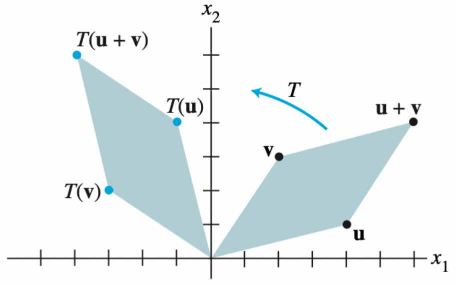


FIGURE 6 A rotation transformation.

Reflection through Origin

Example

Use a rectangular coordinate system to plot $u = \begin{bmatrix}5\\2\end{bmatrix}, v = \begin{bmatrix}-2\\4\end{bmatrix}$

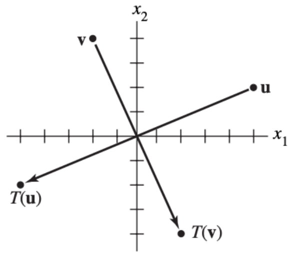
and their images under the given transformation T. Describe geometrically what T does to each vector x in R2

$$T(\mathbf{x}) = \begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

Solution

$$T(u) = \begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} = \begin{bmatrix}-5\\-2\end{bmatrix}$$

$$T(v) = \begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}\begin{bmatrix}-2\\4\end{bmatrix} = \begin{bmatrix}2\\-4\end{bmatrix}$$



A reflection through the origin.

Example

Use a rectangular coordinate system to plot $u = \begin{bmatrix}5\\2\end{bmatrix}, v = \begin{bmatrix}-2\\4\end{bmatrix}$

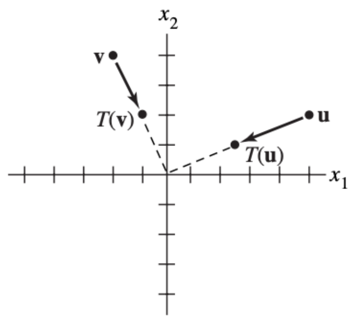
and their images under the given transformation T. Describe geometrically what T does to each vector x in R^2

$$T(\mathbf{x}) = \begin{bmatrix}.5 & 0\\0 & .5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

Solution

$$T(u) = \begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} = \begin{bmatrix}2.5\\1\end{bmatrix}$$

$$T(v) = \begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\begin{bmatrix}-2\\4\end{bmatrix} = \begin{bmatrix}-1\\2\end{bmatrix}$$



A contraction by the factor .5.

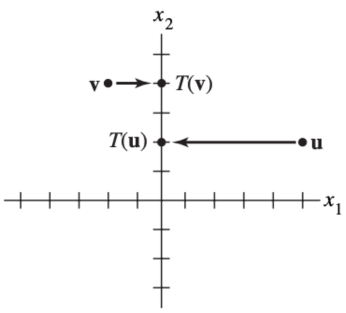
Example

Use a rectangular coordinate system to plot $u = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, v = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$
and their images under the given transformation T. Describe geometrically what T does to each vector x in R^2

$$T(\mathbf{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Solution

$$T(u) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$
$$T(v) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$



A projection onto the x_2-axis

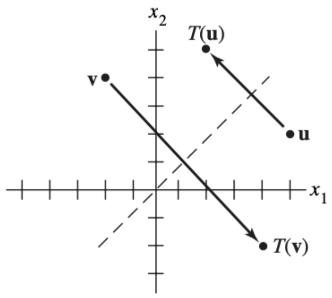
Example

Use a rectangular coordinate system to plot $u = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, v = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$
and their images under the given transformation T. Describe geometrically what T does to each vector x in R^2

$$T(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Solution

$$T(u) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$
$$T(v) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$



A reflection through the line $x_2 = x_1$.

Matrix of a Linear Transformation

Whenever a linear transformation T arises geometrically or is described in words, we usually want a “formula” for T (**x**). The discussion that follows shows that every linear transformation from R^n to R^m is actually a matrix transformation $\boldsymbol{x} \longmapsto \boldsymbol{Ax}$ and that important properties of T are intimately related to familiar properties of A. The key to finding A is to observe that T is completely determined by what it does to the columns of the n×n identity matrix I_n.

Standard Matrix of a Linear Transformation

Let T : $R^n \rightarrow R^m$ be a linear transformation. Then there exists a unique matrix A such that T(x) = Ax for all x in R^n
In fact, A is the m×n matrix whose j^{th} column is the vector T (e_j), where e_j is the j^{th} column of the identity matrix in R^n :
A=[T(e_1) ⋯ T(e_n)]
The matrix A is called the standard matrix for the linear transformation T .

We know now that every linear transformation from R^n to R^m is a matrix transformation, and vice versa.
The term linear transformation focuses on a property of a mapping, while matrix transformation describes how such a mapping is implemented

Standard Rotation Matrix

Let T : $R^2 \rightarrow R^2$ be the transformation that rotates each point in R^2 about the origin through an angle ϕ, with counter clockwise rotation for a positive angle. We could show geometrically that such a transformation is linear. Find the standard matrix A of this transformation.

$$\begin{bmatrix}1\\0\end{bmatrix} rotates\ into \begin{bmatrix}cos\phi\\sin\phi\end{bmatrix}\ and \begin{bmatrix}0\\1\end{bmatrix} rotates\ into \begin{bmatrix}-sin\phi\\cos\phi\end{bmatrix}$$

$$A = \begin{bmatrix}cos\phi & -sin\phi\\sin\phi & cos\phi\end{bmatrix}$$

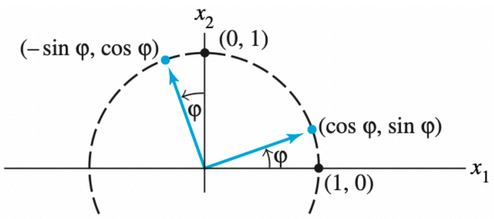
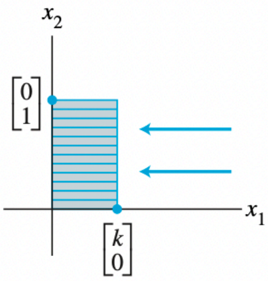
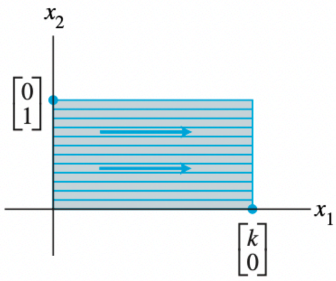
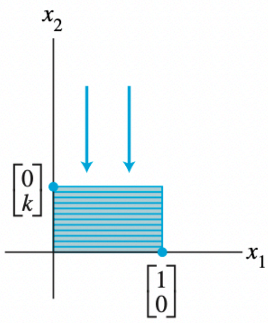
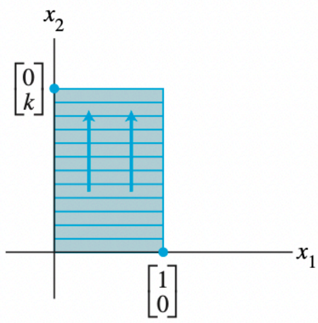
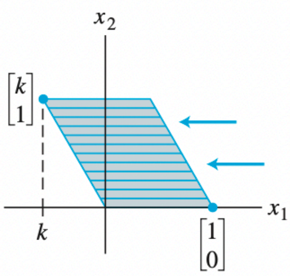
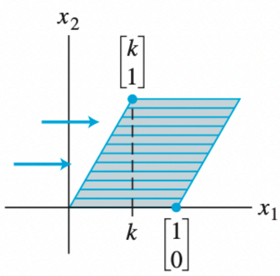
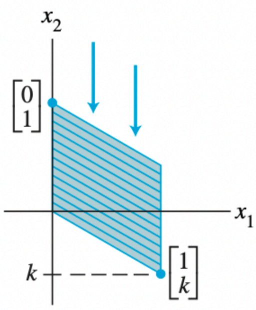
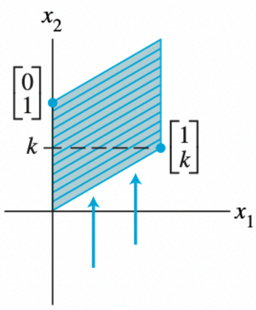
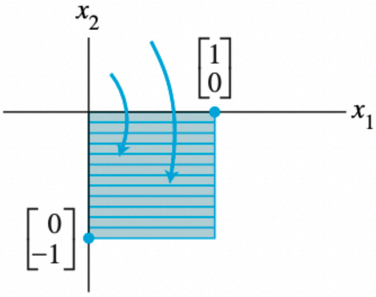
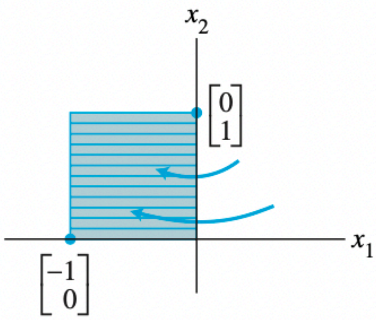
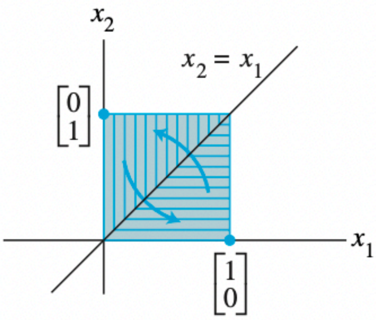
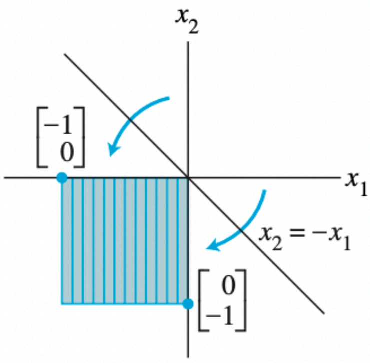
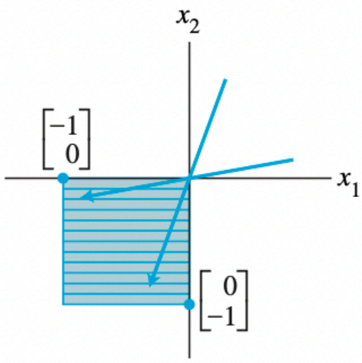
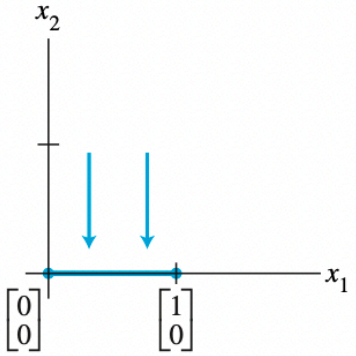
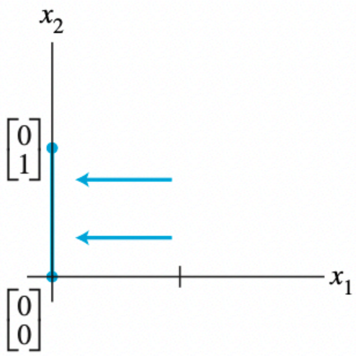


FIGURE 1 A rotation transformation.

Transformation	Image of the Unit Square		Standard Matrix
Horizontal contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix}k & 0\\0 & 1\end{bmatrix}$
Vertical contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix}1 & 0\\0 & k\end{bmatrix}$

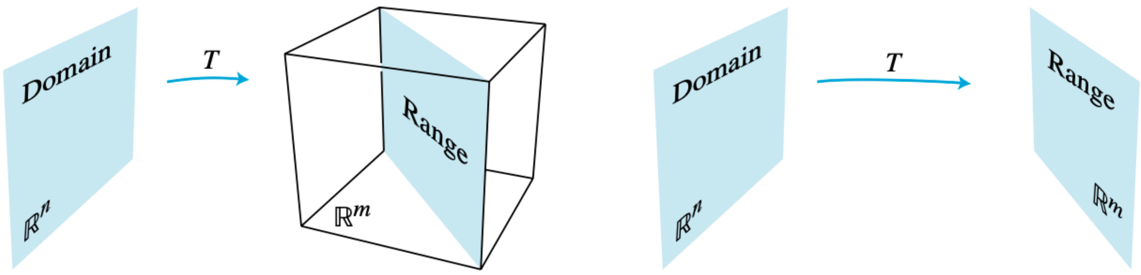
Transformation	Image of the Unit Square		Standard Matrix
Horizontal shear	 $k < 0$	 $k > 0$	$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$
Vertical shear	 $k < 0$	 $k > 0$	$\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}$

Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$
Reflection through the origin		$\begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}$

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Onto Mapping

A mapping T : $R^n \rightarrow R^m$ is said to be onto R^m if each b in R^m is the image of at least one x in R^n. The mapping T is not onto when there is some b in Rm for which the equation T (x) = b has no solution.



T is *not* onto $\mathbb{R}^m$ $\qquad$ T *is* onto $\mathbb{R}^m$

FIGURE 3 Is the range of T all of $\mathbb{R}^m$?

One-to-one Mapping

A mapping T : $R^n \rightarrow R^m$ is said to be one-to-one R^m if each b in R^m is the image of at most one x in R^n.

T is one-to-one if, for each b in R^m, the equation T (x) = b has either a unique solution or none at all.

The mapping T is not one-to-one when some b in R^m is the image of more than one vector in R^n. If there is no such b, then T is one-to-one.

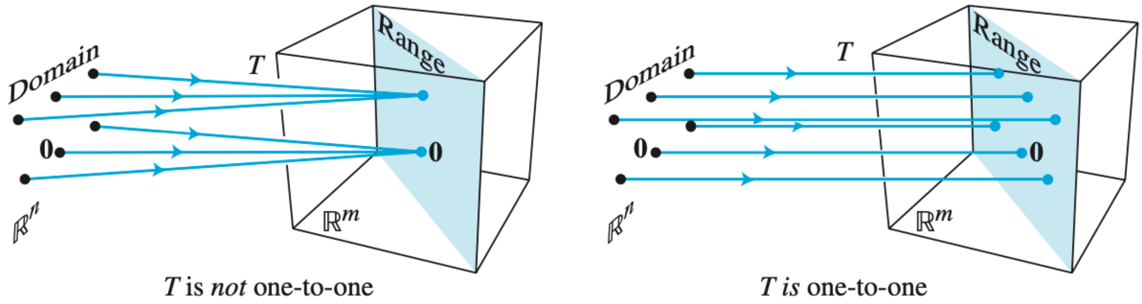


FIGURE 4 Is every **b** the image of at most one vector?

Mapping

The projection transformations are not one-to-one and do not map R^2 onto R^2.
The transformations in Reflection, Contraction, Expansion and Shear are one-to-one and do map R^2 onto R^2.

Invertible Linear Transformation

A linear transformation T : $R^n \rightarrow$ Rn is said to be **invertible** if there exists a function
S : $R^n \rightarrow R^n$ such that

- S(T(**x**))=**x** for all **x** in R^n
- T(S(**x**))=**x** for all **x** in R^n

Let T : $R^n \rightarrow R^n$ be a linear transformation and let A be the standard matrix for T . Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation S given by S(**x**) = A^{-1}**x** is the unique function satisfying

- S(T(**x**))=**x** for all **x** in R^n
- T(S(**x**))=**x** for all **x** in R^n

Example
Find the standard matrix A for the dilation transformation
T (**x**) = 3**x**, for **x** in R^2

Solution
Write

$$\mathrm{T}\,(e_1\,) = \; 3e_1 \, = \begin{bmatrix} 3 \\ 0 \end{bmatrix},\ \mathrm{T}\,(e_2\,) = 3e_2 \, = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

Example
Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ is a vertical shear transformation that maps e_1 into $e_1 - 2e_2$ but leaves the vector e_2 unchanged.

Solution
Write

$$\mathrm{T}\,(e_1\,) = e_1 - 2e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

T (e_2) = e_2 $= \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

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

Example

Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ is a horizontal shear transformation that leaves e1 unchanged and maps e_2 into $e_2 + 3e_1$.

Solution

Write

T $(e_1) = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, T $(e_2) = e_2 + 3e_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

Example

Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ rotates points (about the origin) through 3π/2 radians (counter clockwise).

Solution

T $(e_1) = -e_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, T $(e_2) = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

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

Example

Assume that T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ rotates points (about the origin) through -π/4 radians (clockwise).

Solution

T $(e_1) = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$, T (e_2) = $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

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

Example

T is a linear transformation. Find the standard matrix of T . T : $R^2 \rightarrow R^2$ first reflects points through the horizontal x_1- axis and then reflects points through the line $x_2 = x_1$.

Solution

$\mathbf{e}_1 \rightarrow \mathbf{e}_1 \rightarrow \mathbf{e}_2$ and $\mathbf{e}_2 \rightarrow -\mathbf{e}_2 \rightarrow -\mathbf{e}_1$,

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

Example
T : $R^2 \rightarrow R^2$ first performs a horizontal shear that transforms e_2 into $e_2 - 2e_1$ (leaving e_1 unchanged) and then reflects points through the line $x_2 = -x_1$.

Solution
The horizontal shear maps e1 into e1, and then the reflection in the line $x_2 = -x_1$ maps e_1 into $-e_2$.
The horizontal shear maps e_2 into e_2 into $e_2 - 2e_1$. To find the image of $e_2 - 2e_1$ when it is reflected in the line $x_2 = -x_1$, use the fact that such a reflection is a linear transformation. So, the image of $e_2 - 2e_1$ is the same linear combination of the images of e_2 and e_1, namely, $-e_1 - 2(-e_2) = -e_1 + 2e_2$.

$$\mathbf{e}_1 \rightarrow \mathbf{e}_1 \rightarrow -\mathbf{e}_2 \quad \text{and} \quad \mathbf{e}_2 \rightarrow \mathbf{e}_2 - 2\mathbf{e}_1 \rightarrow -\mathbf{e}_1 + 2\mathbf{e}_2, \quad \text{so} \quad A = \begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix}$$

To find the image of $\mathbf{e}_2 - 2\mathbf{e}_1$ when it is reflected through the vertical axis use the fact that such a reflection is a linear transformation. So, the image of $\mathbf{e}_2 - 2\mathbf{e}_1$ is the same linear combination of the images of $\mathbf{e}_2$ and $\mathbf{e}_1$, namely, $\mathbf{e}_2 + 2\mathbf{e}_1$.

Application to Computer Graphics

- Computer graphics are images displayed or animated on a computer screen. Applications of computer graphics are widespread and growing rapidly. For instance, computer-aided design (CAD) is an integral part of many engineering processes, such as the aircraft design process
- The entertainment industry has made the most spectacular use of computer graphics—from the special effects in *The Matrix* to PlayStation 2 and the Xbox.
- Most interactive computer software for business and industry makes use of computer graphics in the screen displays and for other functions, such as graphical display of data, desktop publishing, and slide production for commercial and educational presentations. Consequently, anyone studying a computer language invariably spends time learning how to use at least two-dimensional (2D) graphics.
- Basic mathematics used to manipulate and dis- play graphical images such as a wire-frame model of an airplane. Such an image (or picture) consists of a number of points, connecting lines or curves, and information about how to fill in closed regions bounded by the lines and curves. Often, curved lines are approximated by short straight-line segments, and a figure is defined mathematically by a list of points.
- Among the simplest 2D graphics symbols are letters used for labels on the screen. Some letters are stored as wire-frame objects; others that have curved portions are stored with additional mathematical formulas for the curves.

Example
The capital letter N in Fig. 1 is determined by eight points, or *vertices*. The coordinates of the points can be stored in a data matrix, D. In addition to D, it is necessary to specify which vertices are connected by lines, but we omit this detail.

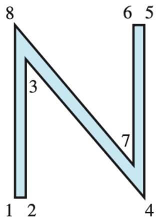


FIGURE 1
Regular N.

$$\begin{array}{ccccccccc} & \text{Vertex:} \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ x\text{-coordinate} \\ y\text{-coordinate} \end{array} \begin{bmatrix} 0 & .5 & .5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix} = D$$

Given A = $\begin{bmatrix} 1 & 0.25 \\ 0 & 1 \end{bmatrix}$, describe the effect of the shear transformation x |→Ax on the letter N

Solution
By definition of matrix multiplication, the columns of the product AD contain the images of the vertices of the letter N.

The transformed vertices are plotted in Fig. 2, along with connecting line segments that correspond to those in the original figure.
The italic N in Fig. 2 looks a bit too wide. To compensate, we can shrink the width by a scale transformation.

Example
Compute the matrix of the transformation that performs a shear transformation using A = $\begin{bmatrix} 1 & 0.25 \\ 0 & 1 \end{bmatrix}$, and then scales all x-coordinates by a factor of 0.75.

Solution
The matrix that multiplies the x-coordinate of a point by 0.75 is

$$S = \begin{bmatrix} .75 & 0 \\ 0 & 1 \end{bmatrix}$$

So the matrix of the composite transformation is

$$SA = \begin{bmatrix} .75 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & .25 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} .75 & .1875 \\ 0 & 1 \end{bmatrix}$$

The result of this composite transformation is shown in Fig. 3.

FIGURE 3
Composite transformation of N.

Homogeneous Coordinates

The mathematics of computer graphics is intimately connected with matrix multi- plication. Unfortunately, translating an object on a screen does not correspond directly to matrix multiplication because translation is not a linear transformation. The standard way to avoid this difficulty is to introduce what are called homogeneous coordinates.

Each point (x, y) in R^2 can be identified with the point (x, y, 1) on the plane in R^3 that lies one unit above the xy-plane. We say that (x, y) has *homogeneous coordinates* (x, y, 1). For instance, the point (0, 0) has homogeneous coordinates (0, 0, 1). Homogeneous coordinates for points are not added or multiplied by scalars, but they can be transformed via multiplication by 3×3 matrices.

Any linear transformation on R^2 is represented with respect to homogeneous coordinates by a partitioned matrix of the form $\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}$ where A is a 2×2 matrix. Typical examples are

$$\begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} s & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Counterclockwise rotation about the origin, angle φ; Reflection through $y = x$; Scale x by s and y by t

Translation

A translation of the form (x, y) |→ (x + h, y + k) is written in homogeneous coordinates as (x, y, 1) |→(x + h, y + k, 1). This transformation can be computed via matrix multiplication:

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x+h \\ y+k \\ 1 \end{bmatrix}$$

The translation operation is often used in machine learning for mean-centering the data, where a constant mean vector is subtracted from each row of the data set. As a result, the mean value of each column of the transformed data set becomes 0. An example of the effect of mean-centering on the scatter plot of a 2-dimensional data set is illustrated in Figure 2.1.

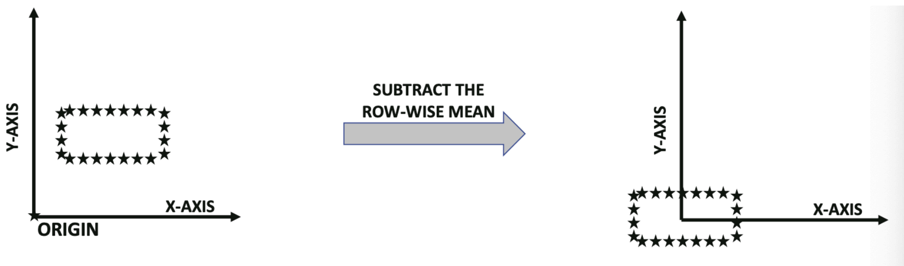


Figure 2.1: Mean-centering: a translation operation

Composite Transformations

Example

Find the 3×3 matrix that corresponds to the composite transformation of a scaling by .3, a rotation of 90◦, and finally a translation that adds (−.5, 2) to each point of a figure.

- If ϕ = π/2, then sinϕ = 1and cosϕ = 0

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \xrightarrow{\text{Scale}} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\xrightarrow{\text{Rotate}} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\xrightarrow{\text{Translate}} \begin{bmatrix} 1 & 0 & -.5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

The matrix for the composite transformation is

$$\begin{bmatrix} 1 & 0 & -.5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1 & -.5 \\ 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -.3 & -.5 \\ .3 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Homogeneous 3D Coordinates

By analogy with the 2D case, we say that (x, y, z, 1) are homogeneous coordinates for the point (x, y, z) in R^3. In general, (X, Y, Z, H) are homogeneous coordinates for (x,y,z) if H ≠ 0 and

$$x = \frac{X}{H}, \qquad y = \frac{Y}{H}, \qquad \text{and} \qquad z = \frac{Z}{H}$$

Each nonzero scalar multiple of (x, y, z, 1) gives a set of homogeneous coordinates for (x,y,z). For instance, both (10, −6, 14, 2) and (−15, 9, −21, −3) are homogeneous coordinates for (5, −3, 7).

Orthogonal Transformation

The orthogonal 2 × 2 matrices V_r and V_c that respectively rotate 2-dimensional row and column vectors by ϕ degrees in the counter-clockwise direction are as follows:

$$V_r = \begin{bmatrix} cos\phi & sin\phi \\ -sin\phi & cos\phi \end{bmatrix}, \; V_c = \begin{bmatrix} cos\phi & -sin\phi \\ sin\phi & cos\phi \end{bmatrix}$$

If we have an n×2 data matrix D, then the product D V_r will rotate each row of D using V_r whereas the product $V_c D^T$ D^T will equivalently rotate each column of D^T . One can also view a data rotation D V_r in terms of projection of the original data on a rotated axis system.

Counter-clockwise rotation of the data with a fixed axis system is the same as clockwise rotation of the axis system with fixed data.

In essence, the two columns of the transformation matrix V_r represent the mutually orthogonal unit vectors of a new axis system that is rotated clockwise by ϕ. These two new columns are shown on the left of Figure 2.2 for a counter- clockwise rotation of 30◦.

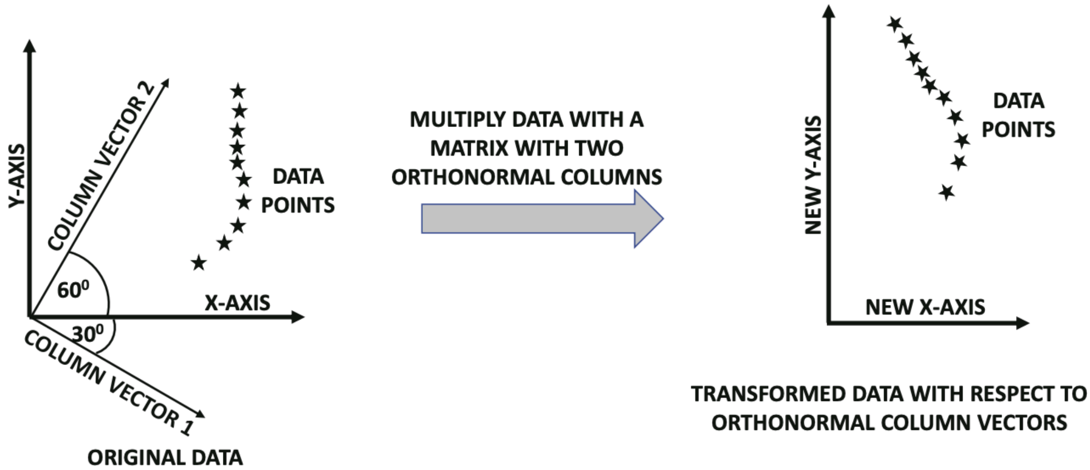


Figure 2.2: An example of counter-clockwise rotation with 30° with matrix multiplication. The two columns of the transformation matrix are shown in the figure on the left

The transformation returns the coordinates D V_r of the data points on these column vectors, because we are computing the dot product of each row of D with the (unit length) columns of V_r . In this case, the columns of V_r (orthonormal directions in new axis system) make counter-clockwise angles of −30◦ and 60◦ with the vector [1, 0].
Therefore, the corresponding matrix V_r is obtained by populating the columns with vectors of the form [cos(ϕ), sin(ϕ)]T , where ϕ is the angle each new orthonormal axis direction makes with the vector [1,0]. This results in the following matrix V_r :

$$V_r = \begin{bmatrix} \cos(30) & \sin(30) \\ -\sin(30) & \cos(30) \end{bmatrix}$$

After performing the projection of each data point on the new axes, we can reorient the figure so that the new axes are aligned with the original X- and Y -axes (as shown in the left-to-right transition of Figure 2.2). It is easy to see that the final result is a counter-clockwise rotation of the data points by 30◦ about the origin.

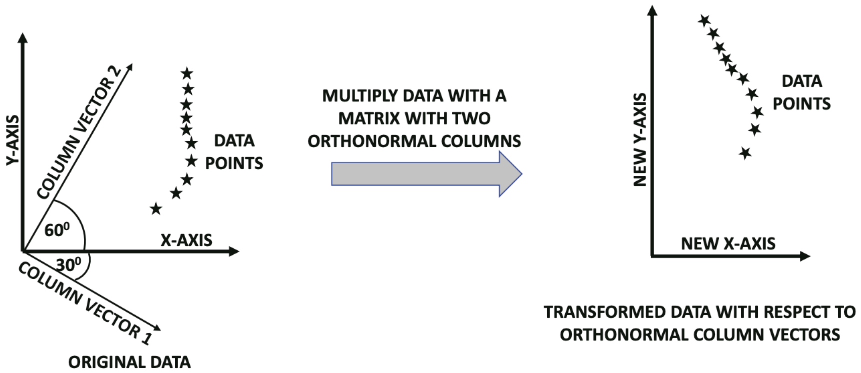


Figure 2.2: An example of counter-clockwise rotation with 30° with matrix multiplication. The two columns of the transformation matrix are shown in the figure on the left

Orthogonal matrices might include reflections. Consider the following matrix:

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

For any 2-dimensional data set contained in the n × 2 matrix D, the transformation DV of the rows of D simply flips the two coordinates in each row of D. The resulting transformation cannot be expressed purely as a rotation. This is because this transformation changes the handedness of the data — for example, if the scatter plot of the n rows of the n × 2 matrix D depicts a right hand, the scatter plot of the n × 2 matrix DV will depict a left hand. Intuitively, when you look at your reflection in the mirror, your left hand appears to be your right hand. This implies that a reflection needs to performed somewhere.

The key point is that V can be expressed as the product of a counter-clockwise rotation of 90◦, followed by a reflection across the vector [0, 1]:

$$V = \begin{bmatrix} \cos(90) & \sin(90) \\ -\sin(90) & \cos(90) \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

When a row of D is post-multiplied with V , it is first rotated counter-clockwise with 90◦ because of multiplication with the first matrix, and then its first coordinate is multiplied with −1 (i.e., reflection across the Y - axis [0, 1]) because of multiplication with the second matrix. An example of the above transformation can be elucidated by post-multiplying the 2-dimensional row vector [3, 4] with V :

$$[3,4]\,V = \underbrace{[3,4]\begin{bmatrix} \cos(90) & \sin(90) \\ -\sin(90) & \cos(90) \end{bmatrix}}_{\text{Rotate } 90^\circ \text{ counter-clockwise}} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \underbrace{[-4,3]\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{Reflect}} = [4,3]$$

Note that the intermediate result [−4, 3] is indeed a 90◦ rotation of [3, 4]. The decomposition of an orthogonal matrix into rotations and reflections is not unique. For example, if we reflected across [1,0] instead of [0,1] in the above example, then a 270◦ counter-clockwise rotation will do the same job.

An orthogonal matrix might correspond to a sequence of rotations in a space of dimensionality greater than 3. For example, if a 4-dimensional object in the xyzw-axis system is rotated once in the xy-plane with angle α and once in the zw-plane with angle β, the two independent rotations cannot be expressed by a single angle or plane of rotation. However, the resulting 4 × 4 orthogonal matrix is still called a “rotation matrix” (in spite of being a sequence of rotations). In some cases, reflections are included with rotations. When a compulsory reflection is included in the sequence, the resulting matrix is referred to as a rotreflection matrix.

Questions

1. Use matrix multiplication to find the image of the triangle with data matrix

$D = \begin{bmatrix} 5 & 2 & 4 \\ 0 & 2 & 3 \end{bmatrix}$ under the transformation that reflects points through the y-axis.

2. Find the 3×3 matrices that produce the described composite 2D transformations, using homogeneous coordinates: Translate by (−2,3), and then scale the x-coordinate by .8 and the y-coordinate by 1.2.

Answers

1.

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

2.

$$\begin{bmatrix} .8 & 0 & 0 \\ 0 & 1.2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} .8 & 0 & -1.6 \\ 0 & 1.2 & 3.6 \\ 0 & 0 & 1 \end{bmatrix}$$

10. The Invertible Matrix Theorem

Let A be a square n×n matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix.
b. A is row equivalent to the n×n identity matrix.
c. A has n pivot positions.
d. The equation A**x** = **0** has only the trivial solution.
e. The columns of A form a linearly independent set.
f. The linear transformation **x** |→ A**x** is one-to-one.
g. The equation A**x** = **b** has atleast one solution for each **b** in R^n .
h. The columns of A span R^n .
i. The linear transformation **x** |→ A**x** maps R^n onto R^n.
j. There is an n×n matrix C such that CA = I.
k. There is an n×n matrix D such that AD = I.
l. A^T is an invertible matrix.
m. The columns of A form a basis of R^n .
n. Col A = R^n
o. Dim Col A = n
p. Rank A=n
q. Nul A={0}
r. Dim Nul A=0

Difference between singular(non-invertible matrix) and non-singular(invertible matrix)

Singular or non-invertible matrix	**Non-singular or invertible matrix**
It has no inverse, A^{-1} does not exists	It has an inverse, A^{-1} exists
Its determinant is zero	The determinant is nonzero
There is no unique solution to the system ***Ax*** = b	There is a unique solution to the system ***Ax*** = b
Gaussian elimination cannot avoid a zero on the diagonal	Gaussian elimination does not encounter a zero on the diagonal
The rank is less than *n*	The rank equals *n*
Rows are linearly dependent	Rows are linearly independent
Columns are linearly dependent	Columns are linearly independent

11. Eigen Values and Eigen Vectors

Eigen Value and Eigen Vectors

An eigenvector of an n×n matrix A is a nonzero vector x such that Ax = λx for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution x of Ax = λx; such an x is called an eigenvector corresponding to λ.

Example Let A= $\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, u = $\begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and v = $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are u and v eigenvectors of A?

Solution

$Au = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4u$

$Av = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$

Thus u is an eigenvector corresponding to an eigenvalue (−4), but v is not an eigenvector of A, because Av is not a multiple of v.

Example Show that x = $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ and find the corresponding eigenvalue.

Solution

We compute

$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4x$

from which it follows that x is an eigenvector of A corresponding to the eigenvalue 4.

Eigen Space

λ is an eigenvalue of A if and only if the equation

(A − λI)x = 0

has a nontrivial solution. The set of all solutions of the above equation is just the null space of the matrix A − λI . So this set is a subspace of R^n and is called the eigenspace of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

All vectors in the nullspace of A − λ I (which we call the eigenspace) will satisfy Ax = λ x.

Example Show that 5 is an eigenvalue of $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ and determine all eigenvectors corresponding to this eigenvalue.

Solution

We must show that there is a nonzero vector x such that Ax = 5x. But this equation is equivalent to the equation (A - 5I)x = 0, so we need to compute the null space of the matrix A - 5I. We find that

$A - 5I = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} - \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} -4 & 2 \\ 4 & -2 \end{bmatrix}$

Since the columns of this matrix are clearly linearly dependent, its null space is nonzero. Thus, Ax = 5x. has a nontrivial solution, so 5 is an eigenvalue of A. We find its eigenvectors by computing the null space:

$$[A-5I \quad 0] = \begin{bmatrix} -4 & 2 & 0 \\ 4 & -2 & 0 \end{bmatrix} = \begin{bmatrix} -1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, if $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue 5, it satisfies

$-x_1 + 1/2x_2 = 0$ or

$x_1 = 1/2x_2$

The eigen vector is $x_2 \begin{bmatrix} 1 \\ 1/_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/_2 \end{bmatrix}$

The eigen space corresponding to λ = 5 consists of all multiples of $\begin{bmatrix} 1 \\ 1/_2 \end{bmatrix}$, which is the line through (1,1/2) and the origin.

Characteristic Equation

A scalar λ is an eigenvalue of an n×n matrix A if and only if λ satisfies the characteristic equation

det (A − λI) = 0

if A is an n×n matrix, then det (A − λI) is a polynomial of degree n called the characteristic polynomial of A.

In general, the (algebraic) multiplicity of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

Example Find eigen values and eigen vectors of

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

Solution

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

Det(A) = (4- λ)(-3- λ)+10 = $\lambda^2 - \lambda - 2 = 0$ is the characteristic polynomial.

(λ + 1)(λ – 2) = 0

λ = -1 or λ = 2

For $\lambda = -1$ $(A - \lambda I)x = \begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Eigen vector for $\lambda = -1$ is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For $\lambda = 2$ $(A - \lambda I)x = \begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Eigen vector for $\lambda = 2$ is $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$

The eigenspaces are the lines through x_1 = (1, 1) and x_2 = (5, 2) and the origin.

Example Find eigen values and eigen vectors of

$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

Solution

$A - \lambda I = \begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}$

Det(A) = (3- λ) -1 = $\lambda^2 - 6\lambda + 8 = 0$ is the characteristic polynomial.

(λ - 4)(λ – 2) = 0

λ = 4 or λ = 2

For $\lambda = 4$ $(A - \lambda I)x = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Eigen vector for $\lambda = 4$ is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For $\lambda = 2$ $(A - \lambda I)x = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Eigen vector for $\lambda = 2$ is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

The eigenspaces are the lines through x_1 = (1, 1) and x_2 = (-1, 1) and the origin.

Important theorems of eigenvalues and eigenvectors

1. If A is a n×n triangular matrix – upper triangular, lower triangular or diagonal, the eigenvalues of A are the diagonal entries of A .
2. λ = 0 is an eigenvalue of A if A is a singular (noninvertible) matrix.
3. A and A^T have the same eigenvalues.
4. Eigenvalues of a symmetric matrix are real.
5. Eigenvectors of a symmetric matrix are orthogonal, but only for distinct eigenvalues.
6. |det(A)| is the product of the absolute values of the eigenvalues of A .
7. Sum of all eigen values of a matrix is equal to the trace of the matrix.

The Power Method

One of the most common methods used for finding largest eigenvalues and eigenvectors is the power method. The power method applies to an n×n matrix A with a strictly dominant eigenvalue λ_1, which means that λ_1 must be larger in absolute value than all the other eigenvalues.

Note that if this largest eigenvalues is repeated, this method will not work.

Also this eigenvalue needs to be distinct. In this case, the power method produces a scalar sequence that approaches λ_1 and a vector sequence that approaches a corresponding eigenvector.

The method is as follows:

1. Select an initial vector x_0 whose largest entry is 1.
2. For k = 0, 1,...,
   a. Compute A x_k.
   b. Let μ_k be an entry in A x_k whose absolute value is as large as possible.
   c. Compute x_{k+1} = $\left(1/\mu_k\right)\mathrm{A}x_k$.
3. For almost all choices of x_0 , the sequence $\{\mu_k\}$ approaches the dominant eigenvalue, and the sequence $\{x_k\}$ approaches a corresponding eigenvector.
   If ϵ is the pre-specified percentage relative error tolerance to which you would like the answer to converge to, keep iterating until
   $\left|\frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}}\right| \times 100 \le \epsilon$ where the left hand side of the above inequality is the definition of absolute percentage relative approximate error, denoted generally by ϵ A pre-specified percentage relative tolerance of $0.5 \times 10^{2-m}$ implies at

least m significant digits are current in your answer. When the system converges, the value of λ is the largest (in absolute value) eigenvalue of A.

Example Using the power method, find the largest eigenvalue and the corresponding eigenvector of

$A = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}$

Solution

Assume $x_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Identify the largest entry μ_0 in $A\,x_0$

$Ax_0 = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \mu_0 = 5$

Scale Ax_0 by $1/\mu_0$ to get x_1, compute Ax_1, and identify the largest entry in Ax_1 :

$x_1 = (1/\mu_0)Ax_0$

$= \frac{1}{5}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.4 \end{bmatrix}$

$Ax_1 = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 8 \\ 1.8 \end{bmatrix}$

$\mu_1 = 8$

$x_2 = (1/\mu_1)Ax_1$

$= \frac{1}{8}\begin{bmatrix} 8 \\ 1.8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.225 \end{bmatrix}$

Conducting further iterations, the values of λ and the corresponding eigenvectors is given in the table below

Iteration	**λ**	**x**	$\epsilon = \left\lvert \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} \right\rvert \times 100$
1	5	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	-
2	8	$\begin{bmatrix} 8 \\ 1.8 \end{bmatrix}$	37%
3	7.125	$\begin{bmatrix} 1 \\ 0.225 \end{bmatrix}$	12%
4	7.0175	$\begin{bmatrix} 1 \\ 0.2035 \end{bmatrix}$	1.5%
5	7.0025	$\begin{bmatrix} 1 \\ 0.2005 \end{bmatrix}$	0.2%

The exact value of the eigenvalue is λ = 7

and the corresponding eigenvector is $x_1 = \begin{bmatrix} 1 \\ 0.2 \end{bmatrix}$

The Inverse Power Method

This method provides an approximation for any eigenvalue, provided a good initial estimate α of the eigenvalue λ is known. In this case, we let $\mathrm{B} = (\mathrm{A} - \alpha\mathrm{I})^{-1}$ and apply the power method to B. It can be shown that if the eigenvalues of A are $\lambda_1, \lambda_2, \ldots, \lambda_n$ then the eigenvalues of B are

$$\frac{1}{\lambda_1 - \alpha}, \frac{1}{\lambda_2 - \alpha}, \ldots\ldots, \frac{1}{\lambda_n - \alpha}$$

and the corresponding eigenvectors are the same as those for A.

Suppose, for example, that α is closer to λ_2 han to the other eigenvalues of A. Then $\frac{1}{\lambda_2-\alpha}$ will be a strictly dominant eigenvalue of B. If α is really close to λ_2, $\frac{1}{\lambda_2-\alpha}$ is much larger than the other eigenvalues of B, and the inverse power method produces a very rapid approximation to λ_2 for almost all choices of α.

The method is as follows:

- Select an initial estimate α sufficiently close to λ.
- Select an initial vector x_0 whose largest entry is 1.
- For k = 0,1,...,
    a. Solve (A-αI)y_k = x_k for y_k .

    b. Let μ_k be an entry in y_k whose absolute value is as large as possible.

    c. Compute $v_k = \alpha + ({}^1/_{\mu_k})$

    d. Compute $x_{k+1} = ({}^1/_{\mu_k})y_k$
- For almost all choices of x_0, the sequence $\{v_k\}$ approaches the eigen value λ of A, and the sequence $\{x_k\}$ approaches a corresponding eigenvector.

Notice that B, or rather $(\mathrm{A} - \alpha\mathrm{I})^{-1}$, does not appear in the algorithm.
Instead of computing $(\mathrm{A} - \alpha\mathrm{I})^{-1}x_k$ to get the next vector in the sequence, it is better to *solve* the equation (A-αI)y_k = x_k for y_k (and then scale y_k to produce x_{k+1}).

Example

Suppose 21, 3.3, and 1.9 are estimates for the eigenvalues of the matrix A below. Find the smallest eigenvalue, accurate to six decimal places.

$$A = \begin{bmatrix} 10 & -8 & -4 \\ -8 & 13 & 4 \\ -4 & 5 & 4 \end{bmatrix}$$

Solution

The two smallest eigenvalues seem close together, so we use the inverse power method for A − 1.9I .
Here x_0 was chosen arbitrarily, y_k = $(\mathrm{A} - 1.9\mathrm{I})^{-1}x_k$, μ_k is the largest entry in y_k , $v_k = 1.9 + ({}^1/_{\mu_k})$, and $x_{k+1} = ({}^1/_{\mu_k})y_k$.

k	0	1	2	3	4
$\mathbf{x}_k$	$\begin{bmatrix}1\\1\\1\end{bmatrix}$	$\begin{bmatrix}.5736\\.0646\\1\end{bmatrix}$	$\begin{bmatrix}.5054\\.0045\\1\end{bmatrix}$	$\begin{bmatrix}.5004\\.0003\\1\end{bmatrix}$	$\begin{bmatrix}.50003\\.00002\\1\end{bmatrix}$
$\mathbf{y}_k$	$\begin{bmatrix}4.45\\.50\\7.76\end{bmatrix}$	$\begin{bmatrix}5.0131\\.0442\\9.9197\end{bmatrix}$	$\begin{bmatrix}5.0012\\.0031\\9.9949\end{bmatrix}$	$\begin{bmatrix}5.0001\\.0002\\9.9996\end{bmatrix}$	$\begin{bmatrix}5.000006\\.000015\\9.999975\end{bmatrix}$
μ_k	7.76	9.9197	9.9949	9.9996	9.999975
ν_k	2.03	2.0008	2.00005	2.000004	2.0000002

As it turns out, the initial eigenvalue estimate was fairly good, and the inverse power sequence converged quickly.
The smallest eigenvalue is 2.

Questions

1. Explain eigen vectors graphically.
2. When do we use the power method?
3. When do we stop the iterations in power method?
4. Does the power method give the sign of the largest eigen value?
5. Are x_1 and x_2 eigenvectors of A? Find eigen values

$$A=\begin{bmatrix}2 & 0\\0 & -1\end{bmatrix} \qquad x_1=\begin{bmatrix}1\\0\end{bmatrix} \qquad x_2=\begin{bmatrix}0\\1\end{bmatrix}$$

6. Is $\begin{bmatrix}1\\-2\\1\end{bmatrix}$ an eigenvector of $\begin{bmatrix}3 & 6 & 7\\3 & 3 & 7\\5 & 6 & 5\end{bmatrix}$? If so, find the eigenvalue.
7. Find eigen vectors and eigen values of

$$A=\begin{bmatrix}2 & -12\\1 & -5\end{bmatrix}$$

8. One of the eigenvalues of

$A=\begin{bmatrix}5 & 6 & 2\\3 & 5 & 9\\2 & 1 & -7\end{bmatrix}$ is zero. Is A invertible?

9. Given the eigenvalues of

$$A=\begin{bmatrix}2 & -3.5 & 6\\3.5 & 5 & 2\\8 & 1 & 8.5\end{bmatrix}$$

are λ = -1.547 , 12.33, 4.711
What are the eigenvalues of B if

$$B=\begin{bmatrix}2 & 3.5 & 8\\-3.5 & 5 & 1\\6 & 2 & 8.5\end{bmatrix}$$

10. Determine the largest eigen value in magnitude and the corresponding eigen vector of the following matrices by power method. Take as $x_0=[1 \quad 1 \quad 1]^T$

a. $\begin{bmatrix} 35 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & -1 \end{bmatrix}$

b. $\begin{bmatrix} 20 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

Answers

4. Solution: No. Power method gives the largest eigen value in magnitude. If the sign of the eigen value is required, then we substitute this value in the characteristic determinant $(\mathrm{A} - \lambda\mathrm{I})$ and determine the sign of the eigen value. If $|\mathrm{A} - \lambda\mathrm{I}| = 0$ is satisfied approximately, then it is of positive sign. Otherwise, it is of negative sign.

8. λ = 0 is an eigenvalue of A, that implies A is singular and is not invertible

9. Solution: Since B = A^T, the eigenvalues of A and B are the same. Hence eigenvalues of B also are λ = -1.547 , 12.33, 4.711

10a) | λ | = 35.15, x = $[1.0 \quad 0.06220 \quad 0.02766]^T$

10b) | λ | = 20.11, x = $[1.0, \quad 0.05316, \quad 0.04759]^T$

12. Diagonalization

Similarity

If A and B are n×n matrices, then A is similar to B if there is an invertible matrix P such that $P^{-1}AP = B$,
or equivalently, $A = PBP^{-1}$
Changing A into $P^{-1}AP$ is called a similarity transformation.

If n×n matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues

Diagonalization

An n×n matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.
In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P
An n×n matrix with n distinct eigenvalues is diagonalizable.
The determinant of a diagonalizable matrix is equal to the product of its eigenvalues.
We can also write it as
$D = P^{-1}AP \;\; or \; AP = PD$

Procedure to find Matrix P and Diagonal matrix D

Let A be a given matrix

- **Step 1**: Find Characteristic equation of A
  det (A − λI)
- **Step 2**: Find Eigenvalues of A, say $\lambda_1, \lambda_2, \ldots, \lambda_n$
  Step 3: Find Eigenvectors for all eigenvalues of A, say $v_1, v_2, \ldots, v_n$ corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively.
- **Step 4**: Construct P such that columns of P are eigenvectors of A
  $P = [v_1, v_2, \ldots, v_n]$
- **Step 5**:
  As all eigenvectors of A are independent, matrix P is invertible
- **<u>Step 6</u>**:

  $A = PDP^{-1}$
  Where matrix D is called a diagonal form of matrix A

Example

a. Show that the matrix A is diagonalizable.

$$A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$$

b. Find a diagonal matrix *D* that is similar to *A*.
c. Determine the similarity transformation that diagonalizes *A*.

Solution

a. The eigenvalues and corresponding eigenvector of this matrix are

$\lambda_1 = 2, \mathbf{v}_1 = r\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ $\qquad$ $\lambda_2 = -1, \text{and } \mathbf{v}_2 = s\begin{bmatrix} -2 \\ 1 \end{bmatrix}$

Since A, a 2×2 matrix, has two linearly independent eigenvectors, it is diagonalizable.

b. A is similar to the diagonal matrix D, which has diagonal elements $\lambda_1 = 2$ and $\lambda_2 = -1$. Thus
$A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$ is similar to $D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$

c. Select two convenient linearly independent eigenvectors, say
$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$
Let these vectors be the column vectors of the diagonalizing matrix P.
$P = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}$
We get

$$P^{-1}AP = \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} = D$$

Example
Show that the following matrix A is not diagonalizable.
$$A = \begin{bmatrix} 5 & -3 \\ 3 & -1 \end{bmatrix}$$

Solution
The characteristic equation is
$$A - \lambda I_2 = \begin{bmatrix} 5-\lambda & -3 \\ 3 & -1-\lambda \end{bmatrix}$$
$$|A - \lambda I_2| = 0 \Rightarrow (5-\lambda)(-1-\lambda) + 9 = 0$$
$$\Rightarrow \lambda^2 - 4\lambda + 4 = 0 \Rightarrow (\lambda - 2)(\lambda - 2) = 0$$
There is a single eigenvalue, $\lambda = 2$. We find he corresponding eigenvectors. $(A - 2I)\,\mathbf{x} = \mathbf{0}$ gives
$$\begin{bmatrix} 3 & -3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad 3x_1 - 3x_2 = 0.$$
Thus $x_1 = r$, $x_2 = r$. The eigenvectors are nonzero vectors of the form
$$r\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
The eigenspace is a one-dimensional space. A is a 2×2 matrix, but it does not have two linearly independent eigenvectors. Thus A is not diagonalizable.

Example

The matrix A is factored in the form PDP −1 . Use the Diagonalization Theorem to find the eigenvalues of A and a basis for each eigenspace.

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

Solution

Eigenvectors form the columns of the left factor, and they correspond respectively to the eigenvalues on the diagonal of the middle factor.

$$\lambda = 5: \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \lambda = 1: \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

Example

A is a 5×5 matrix with two eigenvalues. One eigenspace is three-dimensional, and the other eigenspace is two dimensional. Is A diagonalizable? Why?

Solution

A is diagonalizable because you know that five linearly independent eigenvectors exist: three in the three-dimensional eigenspace and two in the two-dimensional eigenspace. The set of all five eigenvectors is linearly independent.

Example

A is a 3×3 matrix with two eigenvalues. Each eigenspace is one dimensional. Is A diagonalizable? Why?

Solution

No. Let v_1 and v_2 be eigenvectors that span the two one-dimensional eigenspaces. If v is any other eigenvector, then it belongs to one of the eigenspaces and hence is a multiple of either v_1 or v_2. So there cannot exist three linearly independent eigenvectors. So, A cannot be diagonalizable.

Example

A is a 7×7 matrix with three eigenvalues. One eigenspace is two-dimensional, and one of the other eigenspaces is three-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

Solution

Yes, if the third eigenspace is only one-dimensional. In this case, the sum of the dimensions of the eigenspaces will be six, whereas the matrix is 7 × 7. In that case A is not diagonalizable.

Matrices Whose Eigenvalues Are Not Distinct

It is not necessary for an n×n matrix to have n distinct eigenvalues in order to be diagonalizable.
Theorem provides a sufficient condition for a matrix to be diagonalizable.
If an n×n matrix *A* has *n* distinct eigenvalues, with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and if P = $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$
then *P* is automatically invertible because its columns are linearly independent.
When *A* is diagonalizable but has fewer than *n* distinct eigenvalues, it is still possible to build *P* in a way that makes *P* automatically invertible, as the next theorem shows.

Theorem: Let *A* be an n×n matrix whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p$.

a. For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

b. The matrix *A* is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals *n*, and this happens if and only if (*i*) the characteristic polynomial factors completely into linear factors and (*ii*) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

c. If *A* is diagonalizable and B_k is a basis for the eigenspace corresponding to B_k for each *k*, then the total collection of vectors in the sets $B_1, \ldots, B_p$ forms an eigenvector basis.

Powers of A

The eigenvalue of A^2 are exactly $\lambda_1^2, \lambda_2^2, \ldots \ldots \ldots, \lambda_n^2$ and every eigenvector of A is also an eigenvector , of A^2 .

Similarly,
The eigenvalues of A^k are $\lambda_1^k, \lambda_2^k, \ldots \ldots \ldots, \lambda_n^k$, and each eigenvector of A is still an eigenvector of A^k .

Example 1: Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k , given that $A = PDP^{-1}$, where

$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$

Solution: The standard formula for the inverse of a n×n matrix yields

$P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$

Then, by associativity of matrix multiplication,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\underbrace{(P^{-1}P)}_{I}DP^{-1} = PDDP^{-1}$$

$$= PD^2P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Again,

$$A^3 = (PDP^{-1})A^2 = (PD\underbrace{P^{-1})P}_{I}D^2P^{-1} = PDD^2P^{-1} = PD^3P^{-1}$$

- In general, for k≥ 1 ,

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2\cdot 5^k - 3^k & 5^k - 3^k \\ 2\cdot 3^k - 2\cdot 5^k & 2\cdot 3^k - 5^k \end{bmatrix}$$

Example

let $A = PDP^{-1}$ and compute A^4

$$P = \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}, D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Solution

$$P = \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}, D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$P^{-1} = \begin{bmatrix} 3 & -7 \\ -2 & 5 \end{bmatrix}, D^4 = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A^4 = \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & -7 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 226 & -525 \\ 90 & -209 \end{bmatrix}$$

Questions

1.Diagonalize the matrices if possible.

$\boldsymbol{a})\ \begin{bmatrix} 1 & 0 \\ 6 & -1 \end{bmatrix} \qquad \boldsymbol{b})\ \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$

Answers

$\boldsymbol{a})\ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \boldsymbol{b})$ Not diagonalizable

Bibliography
Linear Algebra and its Applications by David Lay
Linear Algebra A Modern Introduction by David Poole