Matrices

CS5691 : PRML - Linear Algebra – basics;

CS6015- LARP;
CS6464

Matrix Arithmetic and Operation

- **Equality:** $A = B$ provided *dimensions of A and B are equal* and $a_{ij} = b_{ij}$ for all i and j.
  Matrices of different sizes cannot be equal.
- **Addition, Subtraction:** $A_{n\times m} \pm B_{n\times m} = [a_{ij} \pm b_{ij}]$. *Matrices of different sizes cannot be added or subtracted.*
- **Scalar Multiple:** $cA = [ca_{ij}]$; c is any number.
- **Multiplication:** $A_{n\times p} * B_{p\times m} = A.B_{n\times m}$
- **Transpose:** $A = [a_{ij}]_{n\times m}$ then $A^T = [a_{ji}]_{m\times n} \forall i, j$
- **Trace:** $tr(A) = \sum_{i=1}^{n} a_{ii}$. *If A is not square then trace is not defined.*

Properties of Matrix Arithmetic and the Transpose

- $A + B = B + A$
- $A + (B + C) = (A + B) + C$
- $A(BC) = (AB)C$
- $A(B \pm C) = AB \pm AC$
- $(B \pm C)A = BA \pm CA$
- $a(B \pm C) = aB \pm aC$
- $(a \pm b)C = aC \pm bC$
- $(ab)C = a(bC)$
- $a(BC) = (aB)C = B(aC)$
- A (B) $\neq B\ (A)$, in general.

Letters in caps define matrices, while that in small denote scalars.

Properties of Matrix Arithmetic and the Transpose

- $A + 0 = 0 + A = A$
- $A - A = 0$
- $0 - A = A$
- $0A = 0 \; and \; A0 = 0$
- $A^n A^m = A^{n+m}$
- $(A^n)^m = A^{nm}$
- $(A^T)^T = A$
- $(A \pm B)^T = A^T \pm B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$

Letters in caps define matrices, while that in small denote scalars.

Theorem **Traces of AB and BA are equal.** *If AB and BA are each square, then* $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

Important properties of the inverse matrix

Suppose that A and B are invertible matrices of the same size. Then,

a) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$

b) A^{-1} is invertible and $(A^{-1})^{-1} = A$

c) For $n = 0,1,2 \ldots A^n$ is invertible and $(A^n)^{-1} = A^{-n} = (A^{-1})^n$

d) If c is any non zero scalar then cA is invertible and $(cA)^{-1} = \frac{1}{c}A^{-1}$.

e) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$

Inverse Calculation

- The matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

  will be ***invertible*** if $ad - bc \neq 0$

  and ***singular*** if $ad - bc = 0$.

- If the matrix is invertible its inverse will be,

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

The properties of these operations are (assuming that r, s are scalars and the sizes of the matrices A, B, C are chosen so that each operation is well defined):

$$A + B = B + A, \tag{1}$$
$$(A + B) + C = A + (B + C), \tag{2}$$
$$A + 0 = A, \tag{3}$$
$$r(A + B) = rA + rB, \tag{4}$$
$$(r + s)A = rA + sA, \tag{5}$$
$$r(sA) = (rs)A; \tag{6}$$
$$A(BC) = (AB)C, \tag{7}$$
$$A(B + C) = AB + AC, \tag{8}$$
$$(B + C)A = BA + CA, \tag{9}$$

$$r(AB) = (rA)B = A(rB), \tag{10}$$
$$I_m A = A = AI_n; \tag{11}$$
$$(A^T)^T = A, \tag{12}$$
$$(A + B)^T = A^T + B^T, \tag{13}$$
$$(rA)^T = rA^T, \tag{14}$$
$$(AB)^T = B^T A^T, \tag{15}$$
$$(I_n)^T = I_n; \tag{16}$$
$$AA^{-1} = A^{-1}A = I_n, \tag{17}$$
$$(rA)^{-1} = r^{-1}A^{-1}, \; r \neq 0, \tag{18}$$
$$(AB)^{-1} = B^{-1}A^{-1}, \tag{19}$$
$$(I_n)^{-1} = I_n, \tag{20}$$
$$(A^T)^{-1} = (A^{-1})^T, \tag{21}$$
$$(A^{-1})^{-1} = A. \tag{22}$$

$$\mathrm{Tr}(\mathbf{A} + \mathbf{B}) = \mathrm{Tr}(\mathbf{A}) + \mathrm{Tr}(\mathbf{B})$$
$$\mathrm{Tr}(\mathbf{E}) = n \qquad \text{(trace of identity matrix)}$$
$$\mathrm{Tr}(\mathbf{O}) = 0 \qquad \text{(trace of zero matrix)}$$
$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA})$$
$$\mathrm{Tr}(c\mathbf{A}) = c\,\mathrm{Tr}(\mathbf{A}) \qquad c \in \mathbb{C}$$
$$\mathrm{Tr}(\mathbf{A}^{\mathrm{T}}) = \mathrm{Tr}(\mathbf{A})$$

Special Matrices : Diagonal Matrix

- **Diagonal Matrix:** A square matrix is called **diagonal** if it has the following form

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & d_n \end{bmatrix}$$

- Suppose *D* is a diagonal matrix and $d_i, i = 1, .., n$ are the entries on the main diagonal.
- If one or more of the $d_i's$ are zero then the matrix is singular.

Diagonal Matrix (contd.)

- On the other hand if $d_i \neq 0, \forall i$ then the matrix is invertible and the inverse is,

$$\boldsymbol{D^{-1}} = \begin{bmatrix} \frac{1}{d_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{d_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \frac{1}{d_n} \end{bmatrix}$$

Triangular matrix

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix}_{n\times n}$$

Upper Triangular Matrix

$$L = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{bmatrix}_{n\times n}$$

Lower Triangular Matrix

- *If A is a triangular matrix with main diagonal entries* $a_{11}, a_{22}, \dots, a_{nn}$ *then if one or more of the* a_{ii}*'s are zero the matrix will be* ***singular****.*
- *On the other hand if* $a_{ii} \neq 0 \; \forall i$ *then the matrix is* ***invertible****.*

Symmetric and anti-symmetric matrices

Suppose that A is an $n \times m$ matrix, then A will be called **symmetric** if $A = A^T$.

Some properties of symmetric matrices are:

a) *For any matrix A, both AA^T and A^TA are symmetric.*

b) *If A is an invertible symmetric matrix then A^{-1} is also symmetric.*

c) *If A is invertible then AA^T and A^TA are both invertible.*

Anti-Symmetric or Skew-Symmetric:

An anti-symmetric matrix is a square matrix that satisfies the identity $\boldsymbol{A = -A^T}$.

Other Special forms of matrices:

- Toeplitz matrix
- Block Circulant Matrix
- Orthogonal (also, -skew -sym)
- PD, PSD, …
- Tri-diagonal system
- Hessian
- Jacobian
- Adjoint and Adjugate matrices
- (skew-) Hermitian (or self-adjoint) matrix
- Covariance matrix
- Periodic matrices

- Compound Matrix
- g-inv & Pseudo-inv
- GRAM matrix
- Kernel of matrix
- Schur Complement
- PERM (n)
- Skew-symmetric
- DFT Matrix
- Idempotent Matrices
- Vandermonde Matrices

Matrix Multiplication

1. The <u>order</u> makes a difference…AB is different from BA.
2. **Rule :** The number of columns in first matrix <u>must</u> equal number of rows in second matrix.

   In other words, the **inner dimensions** must be equal.
3. **Dimension of product :** **<u>The answer</u>** will be number of rows in first matrix <u>by</u> number of columns in second matrix.

   In other words, the **outer dimensions.**

$$\begin{bmatrix}4\\2\end{bmatrix}\times\begin{bmatrix}3 & 1\end{bmatrix}=\begin{bmatrix}\square & \square\\ \square & \square\end{bmatrix}$$

2 x 1 1 x 2

$$\begin{bmatrix}3 & 1\end{bmatrix}\times\begin{bmatrix}4\\2\end{bmatrix}=\begin{bmatrix}\square\end{bmatrix}$$

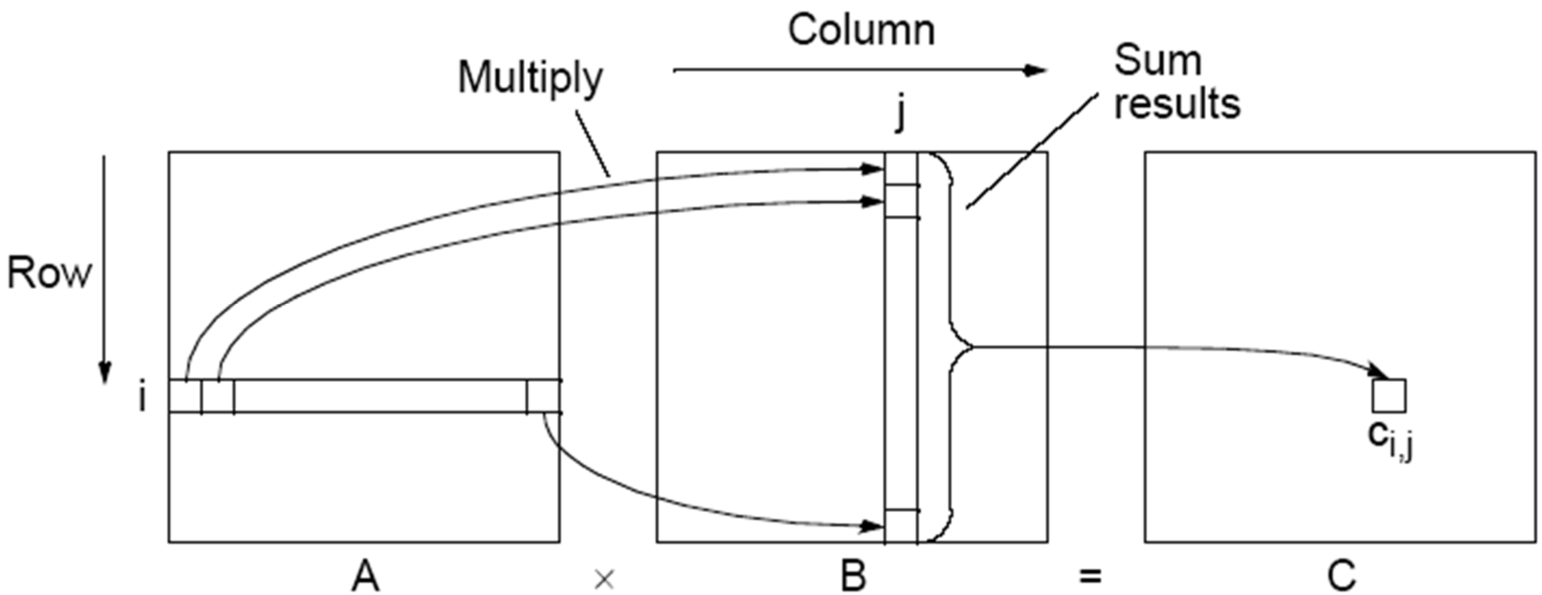
1 x 2 2 x 1

Matrix Multiplication

Multiplication of two matrices, **A** and **B**, produces the matrix **C** whose elements, $c_{i,j}$ $(0 <= i < n, 0 <= j < m)$, are computed as follows:

$$c_{i,j} = \sum_{k=0}^{l-1} a_{i,k} b_{k,j}$$

where **A** is an $n \times p$ matrix and **B** is an $p \times m$ matrix.



Matrix Notation and Matrix Multiplication

Nine co-efficients
Three unknowns
Three right-hand sides

$$\begin{aligned} 2u + v \;+\; w &= 5 \\ 4u - 6v \qquad &= -2 \\ -2u + 7v + 2w &= 9 \end{aligned}$$

$$Ax = b$$

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \qquad x = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \qquad b = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}$$

Co-efficient matrix   Solution vector   constant vector

There are two ways to multiply a matrix A and a vector x.

- One way is a row at a time, each row of A combines with x to give a component of Ax. There are three inner products when A has three rows:

$$\textbf{Ax by rows} \qquad \begin{bmatrix} 1 & 1 & 6 \\ 3 & 0 & 1 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 2 + 1\cdot 5 + 6\cdot 0 \\ 3\cdot 2 + 0\cdot 5 + 3\cdot 0 \\ 1\cdot 2 + 1\cdot 5 + 4\cdot 0 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \\ 7 \end{bmatrix}$$

- Second way, multiplication a column at a time. The product Ax is found all at once, as a combination of the three columns of A:

$$\textbf{Ax by columns} \qquad 2\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \\ 7 \end{bmatrix}$$

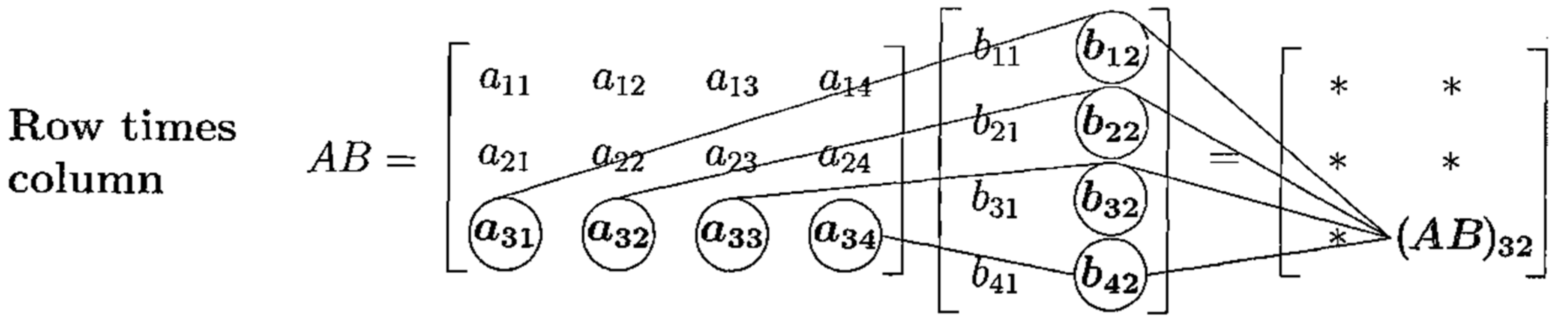
Properties of matrix multiplication

- Every product Ax can be found using whole columns. Therefore Ax is a combination of the columns of A. The coefficients are the components of x.
- The identity matrix I, with 1s on the diagonal and 0s everywhere else, leaves every vector unchanged.

Identity matrix $IA = A$ and $BI = B$.

Properties of matrix multiplication

- The i, j entry of AB is the inner product of the i-th row of A and the j-th column of B
  $$(AB)_{32} = a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} + a_{34}b_{42}$$



Properties of matrix multiplication

- Each entry of AB is the product of a row and a column:
  $(AB)_{ij} = (row\ i\ of\ A)\ times\ (column\ j\ of\ B)$
- Each column of AB is the product of a matrix and a column:
  $column\ j\ of\ AB = A\ times\ (column\ j\ of\ B)$
- Each row of AB is the product of a row and a matrix:
  $row\ i\ of\ AB = (row\ i\ of\ A)\ times\ B$

Properties of matrix multiplication

- For matrices A, B, C, D, E and F,
- Matrix multiplication is **associative**:

$$(AB)C = A(BC)$$

- Matrix operations are **distributive**:

$$A(B+C) = AB + AC \text{ and } (B+C)D = BD + CD$$

- Matrix multiplication is **not commutative**: Usually

$$FE \neq \mathrm{EF}$$

Exception :

$$E = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mathbf{1} & 0 & 1 \end{bmatrix} \qquad EF = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & 1 & 0 \\ \mathbf{1} & 0 & 1 \end{bmatrix} = FE.$$

Norms

To meter the lengths of vectors in a vector space we need the idea of a norm.

Norm is a function that maps x to a nonnegative real number

$$\|\,.\,\|:\quad F \to R^{+}$$

A Norm must satisfy following properties:

$$1-\text{Positivity}\quad \|\mathrm{x}\| > 0 \;\; ,\forall\ \mathrm{x} \neq 0$$

$$2-\text{Homogeneity}\quad \|\alpha \mathrm{x}\| = |\alpha|\,\|x\| \;,\forall\ \mathrm{x} \in \mathrm{F} \text{ and } \forall\ \alpha \in \mathrm{C}$$

$$3-\text{Triangle inequality}\quad \|\mathrm{x}+\mathrm{y}\| \leq \|x\| + \|y\| \;,\forall\ \mathrm{x},\mathrm{y} \in \mathrm{F}$$

Norm of vectors

p-norm is: $\|x\|_p = \left(\sum_i |a_i|^p \right)^{\frac{1}{p}}$ $\quad p \geq 1$

For p=1 we have 1-norm or sum norm $\|x\|_1 = \left(\sum_i |a_i| \right)$

For p=2 we have 2-norm or euclidian norm $\|x\|_2 = \left(\sum_i |a_i|^2 \right)^{1/2}$

For p=∞ we have ∞-norm or max norm $\|x\|_\infty = \max_i \{ |a_i| \}$

The l_p-Norm

The l_p- Norm for a vector x is defined as ($p \geq 1$):

$$\|x\|_{l_p} = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Examples:

- for p=2 we have the ordinary euclidian norm: $\|x\|_{l_2} = \sqrt{x^T x}$

- for p= ∞ the definition is $\|x\|_{l_\infty} = \max_{1 \leq i \leq n} |x_i|$

- a norm for matrices is induced via $\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$

- for l_2 this means :
  $||A||_2$=maximum eigenvalue of $A^T A$

Properties of Matrix Norms

- These induced matrix norms satisfy:

$$\|A\| > 0 \text{ if A} \neq 0$$

$$\|\gamma A\| = |\gamma| \cdot \|A\| \text{ for any scalar } \gamma$$

$$\|A + B\| \leq \|A\| + \|B\| \text{ (triangle inequality)}$$

$$\|AB\| \leq \|A\| \cdot \|B\|$$

$$\|Ax\| \leq \|A\| \cdot \|x\| \text{ for any vector } x$$

Condition Number

- If A is square and nonsingular, then

$$\text{cond}(A) = \|A\| \cdot \|A^{-1}\|$$

- If A is singular, then cond(A) = ∞
- If A is nearly singular, then cond(A) is large.
- The condition number measures the ratio of m maximum shrinkage:

$$\|A\| \cdot \|A^{-1}\| = \left(\max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \right) \cdot \left(\min_{x \neq 0} \frac{\|Ax\|}{\|x\|} \right)^{-1}$$

Condition Number of the Matrix

$$A = \begin{bmatrix} 100 & -200 \\ -200 & 401 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 4.01 & 2 \\ 2 & 1 \end{bmatrix}$$

Row - sum norm of A = ‖ A ‖ = Max (300, 601) = 601.

Row - sum norm of A^{-1} = ‖ A^{-1} ‖ = Max (6.01, 3) = 6.01.

Condition Number k(A) = 601 (6.01) = 3612 (large).

A is ill - conditioned.

The Gaussian Elimination Method

- The Gaussian elimination method is a technique for solving systems of linear equations of any size.
- The operations of the Gaussian elimination method are:
  1. Interchange any two equations.
  2. Replace an equation by a nonzero constant multiple of itself.
  3. Replace an equation by the sum of that equation and a constant multiple of any other equation.

Row-Reduced Form of a Matrix

- Each row consisting entirely of zeros lies below all rows having nonzero entries.
- The first nonzero entry in each nonzero row is 1 (called a leading 1).
- In any two successive (nonzero) rows, the leading 1 in the lower row lies to the right of the leading 1 in the upper row.
- If a column contains a leading 1, then the other entries in that column are zeros.

Row Operations

1. Interchange any two rows.
2. Replace any row by a nonzero constant multiple of itself.
3. Replace any row by the sum of that row and a constant multiple of any other row.

Terminology for the Gaussian Elimination Method

Unit Column

- A column in a coefficient matrix is in unit form if one of the entries in the column is a 1 and the other entries are zeros.

Pivoting

- The sequence of row operations that transforms a given column in an augmented matrix into a unit column.

Notation for Row Operations

- Letting R_i denote the i-th row of a matrix, we write

Operation 1: $R_i \leftrightarrow R_j$ to mean:
Interchange row i with row j.

Operation 2: cR_i to mean:
replace row i with c times row i.

Operation 3: $R_i + aR_j$ to mean:
Replace row i with the sum of row i and a times row j.

Example

- Pivot the matrix about the circled element

$$\begin{bmatrix} \textcircled{3} & 5 & 9 \\ 2 & 3 & 5 \end{bmatrix}$$

Solution:

$$\left[\begin{array}{cc|c} 3 & 5 & 9 \\ 2 & 3 & 5 \end{array}\right] \xrightarrow{\frac{1}{3}R_1} \left[\begin{array}{cc|c} 1 & 5/3 & 3 \\ 2 & 3 & 5 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 5/3 & 3 \\ 0 & -1/3 & -1 \end{array}\right]$$

The Gaussian Elimination Method

1. Write the augmented matrix corresponding to the linear system.
2. Interchange rows, if necessary, to obtain an augmented matrix in which the first entry in the first row is nonzero. Then pivot the matrix about this entry.
3. Interchange the second row with any row below it, if necessary, to obtain an augmented matrix in which the second entry in the second row is nonzero. Pivot the matrix about this entry.
4. Continue until the final matrix is in row-reduced form.

Augmented Matrices

- Matrices are rectangular arrays of numbers that can aid us by eliminating the need to write the variables at each step of the reduction.
- For example, the system

$$2x + 4y + 6z = 22$$
$$3x + 8y + 5z = 27$$
$$-x + y + 2z = 2$$

may be represented by the augmented matrix

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 22 \\ 3 & 8 & 5 & 27 \\ -1 & 1 & 2 & 2 \end{array}\right]$$

Augmented matrix [C|B]

Coefficient Matrix [C]

Matrices and Gaussian Elimination

- Every step in the Gaussian elimination method can be expressed with matrices, rather than systems of equations, thus simplifying the whole process:
- Steps expressed as systems of equations:

$$2x + 4y + 6z = 22$$
$$3x + 8y + 5z = 27$$
$$-x + y + 2z = 2$$

- Steps expressed as augmented matrices:

$$\left[\begin{array}{ccc|c} 2 & 4 & 6 & 22 \\ 3 & 8 & 5 & 27 \\ -1 & 1 & 2 & 2 \end{array}\right]$$

$$\begin{aligned} 2x + 4y + 6z &= 22 \\ 3x + 8y + 5z &= 27 \\ -x + y + 2z &= 2 \end{aligned} \qquad \left[\begin{array}{ccc|c} 2 & 4 & 6 & 22 \\ 3 & 8 & 5 & 27 \\ -1 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{R_1' = \frac{1}{2} R_1}$$

$$\begin{aligned} x + 2y + 3z &= 11 \\ 3x + 8y + 5z &= 27 \\ -x + y + 2z &= 2 \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 11 \\ 3 & 8 & 5 & 27 \\ -1 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{R_2' = R_2 - 3R_1}$$

$$\begin{aligned} x + 2y + 3z &= 11 \\ 2y - 4z &= -6 \\ -x + y + 2z &= 2 \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 11 \\ 0 & 2 & -4 & -6 \\ -1 & 1 & 2 & 2 \end{array}\right]$$

$$\boldsymbol{R_3' = R_3 + R_1}$$

$$\begin{aligned} x + 2y + 3z &= 11 \\ 2y - 4z &= -6 \\ 3y + 5z &= 13 \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 11 \\ 0 & 2 & -4 & -6 \\ 0 & 3 & 5 & 13 \end{array}\right]$$

$$\boldsymbol{R_2' = \frac{1}{2} R_2}$$

$$\boldsymbol{R_2' = \frac{1}{2} R_2}$$

$$\begin{aligned} x + 2y + 3z &= 11 \\ y - 2z &= -3 \\ 3y + 5z &= 13 \end{aligned} \qquad \begin{bmatrix} 1 & 2 & 3 & 11 \\ 0 & 1 & -2 & -3 \\ 0 & 3 & 5 & 13 \end{bmatrix}$$

$$\boldsymbol{R_1' = R_1 - 2R_2}$$

$$\begin{aligned} x \qquad + 7z &= 11 \\ y - 2z &= -3 \\ 3y + 5z &= 13 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 7 & 17 \\ 0 & 1 & -2 & -3 \\ 0 & 3 & 5 & 13 \end{bmatrix}$$

$$\boldsymbol{R_3' = R_3 - 3R_2}$$

$$\begin{aligned} x \qquad + 7z &= 11 \\ y - 2z &= -3 \\ 11z &= 22 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 7 & 17 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 11 & 22 \end{bmatrix}$$

$$\boldsymbol{R_3' = \frac{1}{11} R_3}$$

$$\begin{aligned} x \qquad + 7z &= 11 \\ y - 2z &= -3 \\ z &= 2 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 7 & 17 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\boldsymbol{R_1' = R_1 - 7R_3}$$

$$R_1' = R_1 - 7R_3$$

$$\begin{aligned} x \quad\quad\quad &= 3 \\ y - 2z &= -3 \\ z &= 2 \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

$$R_2' = R_2 + 2R_3$$

$$\begin{aligned} x \quad\quad &= 3 \\ y \quad &= -3 \\ z &= 2 \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Row Reduced Form of the Matrix

Thus, the solution to the system is $x = 3$, $y = 1$, and $z = 2$.

Gaussian Elimination in the case of unique solution

- With a full set of n pivots, there is only one solution.
- The system is non singular, and it is solved by forward elimination and back-substitution.

Systems with no Solution

- If there is a row in the augmented matrix containing all zeros to the left of the vertical line and a nonzero entry to the right of the line, then the system of equations has no solution.

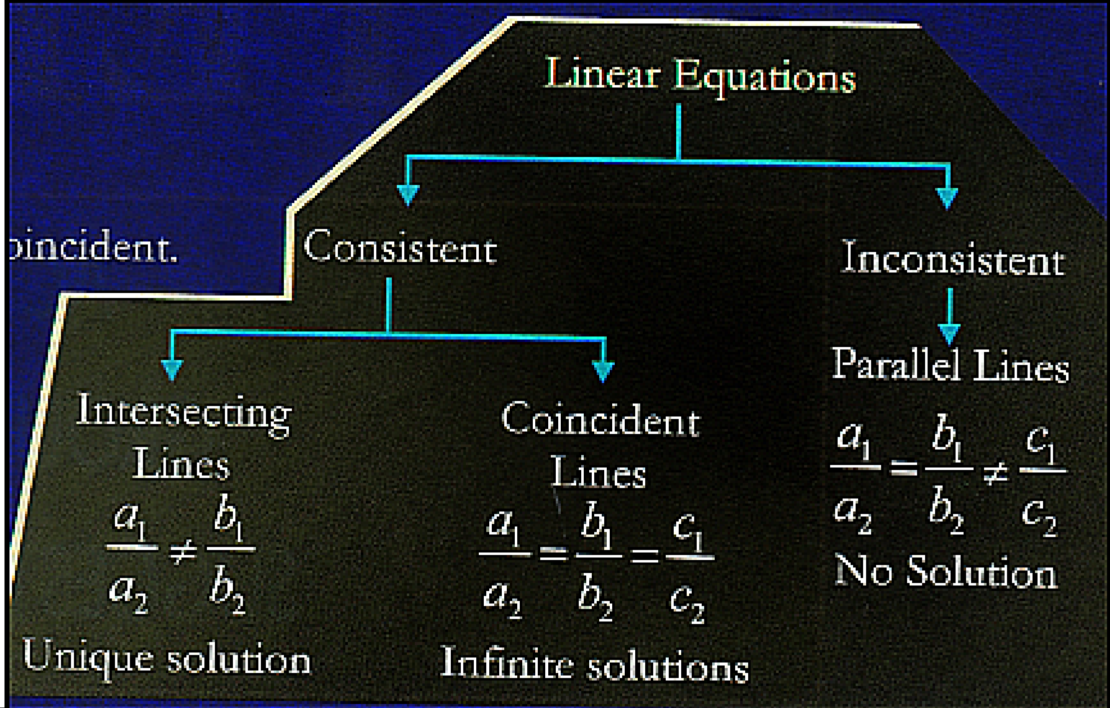
Theorem

a. If the number of equations is greater (over-determined system) than or equal to the number of variables in a linear system, then one of the following is true:
   i. The system has no solution.
   ii. The system has exactly one solution.
   iii. The system has infinitely many solutions.

b. If there are fewer equations than variables (under-determined system) in a linear system, then the system either has no solution or it has infinitely many solutions.

Linear Equation

An equation of the form ax + by + c = 0, where a, b, c are real numbers, a ≠ 0, b ≠ 0 and x, y are variables; is called a linear equation in two variables.



Example for Solving Equations with Zero Solutions

Determine whether the following equation has zero, one, or infinitely many solutions.

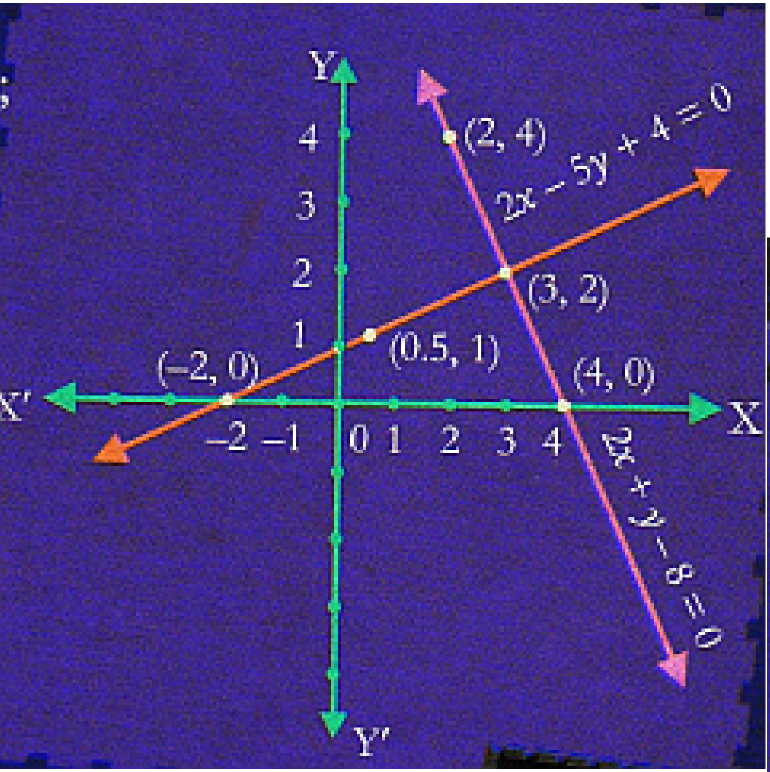
$$3x + 9x + 8 = 14x - 2x + 9$$

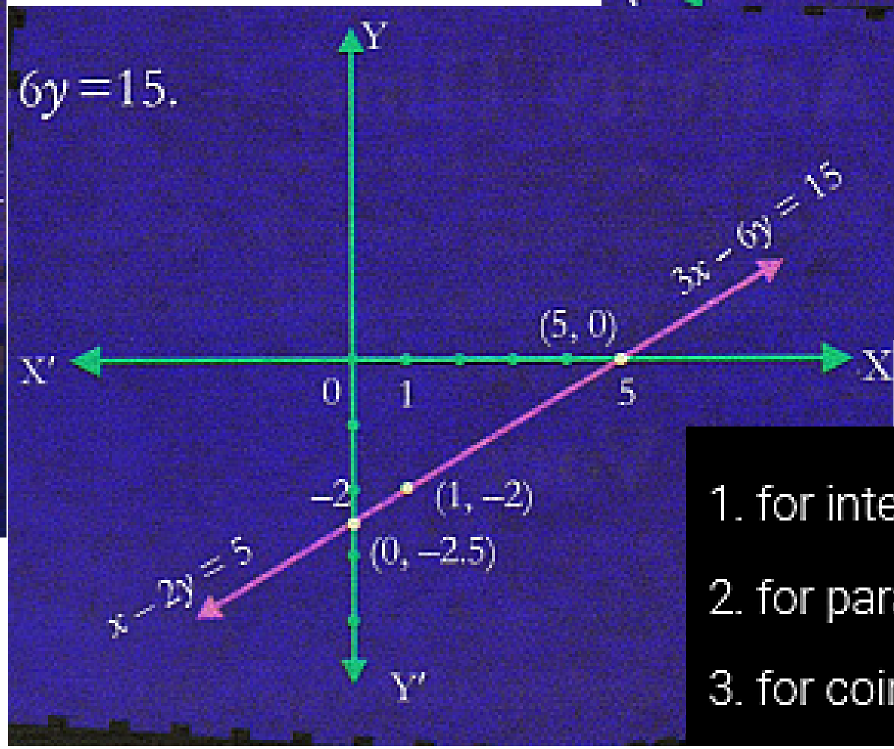
Consider the three pairs of linear equations

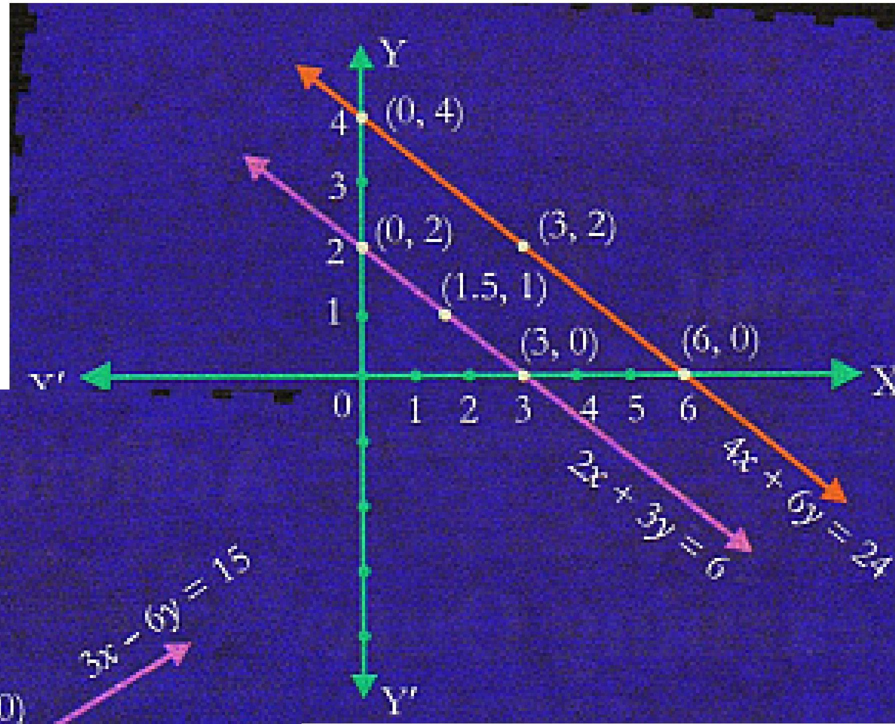
1st pair: $2x - 5y + 4 = 0$, $2x + y - 8 = 0$

2nd pair: $4x + 6y = 24$, $2x + 3y = 6$

3rd pair: $x - 2y = 5$, $3x - 6y = 15$







1. for intersecting lines, $\frac{a_1}{a_2} \neq \frac{b_1}{b_2}$
2. for parallel lines, $\frac{a_1}{a_2} = \frac{b_1}{b_2} \neq \frac{c_1}{c_2}$
3. for coincident lines, $\frac{a_1}{a_2} = \frac{b_1}{b_2} = \frac{c_1}{c_2}$

Inverse matrix

- The inverse of an n by n matrix is another n by n matrix. The inverse of A is written A^{-1} (and pronounced "A inverse").

- The fundamental property is simple: If you multiply by A and then multiply by A^{-1} , you are back where you started:

Inverse matrix If $b = Ax$ then $A^{-1}b = x$

- Thus $A^{-1}Ax = x$. The matrix A^{-1} times A is the identity matrix. ***Not all matrices have inverses. An inverse is impossible when Ax is zero and x is nonzero.*** Then A^{-1} would have to get back from $Ax = 0$ to x. No matrix can multiply that zero vector Ax and produce a nonzero vector x.

- Our goals are to define the inverse matrix and compute it and use it, when A^{-1} exists—and then to understand which matrices don't have inverses.

Properties : Inverse matrix

1K The **inverse** of A is a matrix B such that $BA = I$ and $AB = I$. There is at most one such B, and it is denoted by A^{-1}:

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I. \tag{1}$$

Note* 1. *The inverse exists if and only if elimination produces* n *pivots (row exchanges allowed). Elimination solves $Ax = b$ without explicitly finding A^{-1}.

***Note* 2.** The matrix A cannot have two different inverses, Suppose $BA = I$ and also $AC = I$. Then $B = C$, according to this "proof by parentheses":

$$B(AC) = (BA)C \quad \text{gives} \quad BI = IC \quad \text{which is} \quad B = C. \tag{2}$$

This shows that a *left-inverse* B (multiplying from the left) and a *right-inverse* C (multiplying A from the right to give $AC = I$) must be the *same matrix*.

***Note* 3.** If A is invertible, the one and only solution to $Ax = b$ is $x = A^{-1}b$:

$$\textbf{\textit{Multiply}} \quad Ax = b \quad \textbf{\textit{by}} \quad A^{-1}. \quad \textbf{\textit{Then}} \quad x = A^{-1}Ax = A^{-1}b.$$

Note* 4.** (Important) ***Suppose there is a nonzero vector* x *such that $Ax = 0$. *Then* A *cannot have an inverse*. To repeat: No matrix can bring 0 back to x.

If A is invertible, then $Ax = 0$ can only have the zero solution $x = 0$.

Properties : Inverse matrix

***Note* 5.** A 2 by 2 matrix is invertible if and only if $ad - bc$ is not zero:

$$\textbf{2 by 2 inverse} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{3}$$

This number $ad - bc$ is the *determinant* of A. A matrix is invertible if its determinant is not zero (Chapter 4). In MATLAB, the invertibility test is *to find n nonzero pivots*. Elimination produces those pivots before the determinant appears.

***Note* 6.** A diagonal matrix has an inverse provided no diagonal entries are zero:

$$\text{If} \quad A = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \quad \text{then} \quad A^{-1} = \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{bmatrix} \quad \text{and} \quad AA^{-1} = I.$$

When two matrices are involved, not much can be done about the inverse of $A + B$. The sum might or might not be invertible. Instead, it is the inverse of their *product* AB which is the key formula in matrix computations. Ordinary numbers are the same: $(a+b)^{-1}$ is hard to simplify, while $1/ab$ splits into $1/a$ times $1/b$. But for matrices *the order of multiplication must be correct*—if $ABx = y$ then $Bx = A^{-1}y$ and $x = B^{-1}A^{-1}y$. **The inverses come in reverse order**.

Properties : Inverse matrix

1L A product AB of invertible matrices is inverted by $B^{-1}A^{-1}$:

$$\textbf{Inverse of } AB \qquad (AB)^{-1} = B^{-1}A^{-1}. \tag{4}$$

Proof. To show that $B^{-1}A^{-1}$ is the inverse of AB, we multiply them and use the associative law to remove parentheses. Notice how B sits next to B^{-1}:

$$(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I$$

$$(B^{-1}A^{-1})(AB) = B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I.$$

□

A similar rule holds with three or more matrices:

$$\textbf{Inverse of } ABC \qquad (ABC)^{-1} = C^{-1}B^{-1}A^{-1}.$$

when the elimination matrices E, F, G were inverted to come back from U to A. In the forward direction, $GFEA$ was U. In the backward direction, $L = E^{-1}F^{-1}G^{-1}$ was the product of the inverses. *Since G came last, G^{-1} comes first*. Please check that A^{-1} would be $U^{-1}GFE$.

$[A|b]$

= [GFEA|GFEb];

U = GFEA;

$U^{-1} = A^{-1}(GFE)^{-1}$

$A^{-1} = U^{-1}(GFE)$

Calculation of A^{-1} : The Gauss-Jordan Method

- Given the $n \times n$ matrix A:

  1. Adjoin the $n \times n$ identity matrix I to obtain the augmented matrix $[A \mid I]$.

  2. Use a sequence of row operations to reduce $[A \mid I]$ to the form $[I \mid B]$ if possible.

- Then the matrix B is the inverse of A.

Example

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

<u>Solution</u>

- We form the augmented matrix

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

Example

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

<u>Solution</u>

- And use the Gauss-Jordan elimination method to reduce it to the form $[I \mid B]$:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\boldsymbol{R_1 - R_2}} \left[\begin{array}{ccc|ccc} -1 & -1 & 0 & 1 & -1 & 0 \\ 3 & 2 & 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 0 & 0 & 1 \end{array}\right]$$

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

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

Example

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

<u>Solution</u>

- And use the Gauss-Jordan elimination method to reduce it to the form $[I \mid B]$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & -1 & -3 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right] \xrightarrow[\boldsymbol{R_2 + R_3}]{\boldsymbol{R_1 - R_3}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -1 & -1 \\ 0 & 1 & 0 & -4 & 2 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right]$$

Previous step $\quad\quad$ $\boldsymbol{I_n}$ $\quad\quad$ $\boldsymbol{B}$

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

Finding the inverse of a square matrix using LU decomposition

The inverse [B] of a square matrix [*A*] is defined as

How can LU Decomposition be used to find the inverse?

Methods for LU-Decomp:

Doolittle decomposition*, Crout decomposition,* Cholesky decomposition, Full/partial pivots, Cormen (recursive)

Example: Inverse of a Matrix

Find the inverse of a square matrix [*A*]

$$[A]=\begin{bmatrix} 25 & 5 & 1 \\ 64 & 8 & 1 \\ 144 & 12 & 1 \end{bmatrix}$$

Using the decomposition procedure, the [*L*] and [*U*] matrices are found to be

$$[A]=[L][U]=\begin{bmatrix} 1 & 0 & 0 \\ 2.56 & 1 & 0 \\ 5.76 & 3.5 & 1 \end{bmatrix}\begin{bmatrix} 25 & 5 & 1 \\ 0 & -4.8 & -1.56 \\ 0 & 0 & 0.7 \end{bmatrix}$$

Example: Inverse of a Matrix

Solving for the each column of [*B*] requires two steps

1)Solve [*L*] [*Z*] = [*C*] for [*Z*]

2)Solve [*U*] [*X*] = [*Z*] for [*X*]

Step 1: $$[L][Z]=[C]\rightarrow\begin{bmatrix}1 & 0 & 0\\ 2.56 & 1 & 0\\ 5.76 & 3.5 & 1\end{bmatrix}\begin{bmatrix}z_1\\ z_2\\ z_3\end{bmatrix}=\begin{bmatrix}1\\ 0\\ 0\end{bmatrix}$$

This generates the equations:

$$z_1 = 1$$
$$2.56z_1 + z_2 = 0$$
$$5.76z_1 + 3.5z_2 + z_3 = 0$$

Example: Inverse of a Matrix

Solving for [*Z*]

$$z_1 \;= 1$$

$$z_2 \;=\; 0 - 2.56 z_1$$

$$= 0 - 2.56(1)$$

$$= -2.56$$

$$z_3 = 0 - 5.76 z_1 - 3.5 z_2$$

$$= 0 - 5.76(1) - 3.5(-2.56)$$

$$= 3.2$$

$$[Z] = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} =$$

Example: Inverse of a Matrix

Solving [*U*][*X*] = [*Z*] for [*X*]

$$\begin{bmatrix} 25 & 5 & 1 \\ 0 & -4.8 & -1.56 \\ 0 & 0 & 0.7 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ -2.56 \\ 3.2 \end{bmatrix}$$

$$25b_{11} + 5b_{21} + b_{31} = 1$$

$$-4.8b_{21} - 1.56b_{31} = -2.56$$

$$0.7b_{31} = 3.2$$

Example: Inverse of a Matrix

Using Backward Substitution

$$b_{31} = \frac{3.2}{0.7} = 4.571$$

$$b_{21} = \frac{-2.56 + 1.560 b_{31}}{-4.8}$$

$$= \frac{-2.56 + 1.560(4.571)}{-4.8} = -0.9524$$

$$b_{11} = \frac{1 - 5b_{21} - b_{31}}{25}$$

$$= \frac{1 - 5(-0.9524) - 4.571}{25} = 0.04762$$

So the first column of the inverse o

$$\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} =$$

Example: Inverse of a Matrix

Repeating for the second and third columns of the inverse

Second Column

$$\begin{bmatrix} 25 & 5 & 1 \\ 64 & 8 & 1 \\ 144 & 12 & 1 \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} =$$

Third Column

$$\begin{bmatrix} 25 & 5 & 1 \\ 64 & 8 & 1 \\ 144 & 12 & 1 \end{bmatrix} \begin{bmatrix} b_{13} \\ b_{23} \\ b_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} b_{13} \\ b_{23} \\ b_{33} \end{bmatrix} =$$

Example: Inverse of a Matrix

The inverse of $[A]$ is

$$[A]^{-1} = \begin{bmatrix} 0.04762 & -0.08333 & 0.03571 \\ -0.9524 & 1.417 & -0.4643 \\ 4.571 & -5.000 & 1.429 \end{bmatrix}$$

To check your work do the following operation

$$[A][A]^{-1} = [I] = [A]^{-1}[A]$$

Pseudo

If the colun ormula the pseudo
inverse:

Here A^+ is

However, if

This is a rigl

If both the pseudo inverse is
equal to th

The Moore-Penrose pseudo inverse is a generalization of the matrix inverse when the matrix may not be invertible. If A is invertible, then the Moore-Penrose pseudo inverse is equal to the matrix inverse. However, the Moore-Penrose pseudo inverse is defined even when A is not invertible.

More formally, the Moore-Penrose pseudo inverse, A^+, of an m-by-n matrix is defined by the unique n-by-m matrix satisfying the following four criteria (we are only considering the case where A consists of real numbers).

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. $(AA^+)' = AA^+$
4. $(A^+A)' = A^+A$

If A is an mxn matrix where $m > n$ and A is of full rank (= n), then

$$A^+ = (A'A)^{-1}A'$$

and the solution of $Ax = b$ is $x = A^+b$. In this case, the solution is not exact. It finds the solution that is closest in the least squares sense.

Ax

$(A^TA$

$(AA^T)^{-1} b$

$T)^{-1}] b$

$)^{-1} b$

Eigenvalues and Eigenvectors

CS6015- LARP;

CS5691 - PRML

CS6464;

Ack: Linear Algebra and Its Applications , Gilbert Strang

The Solution of $\boldsymbol{Ax = \lambda x}$

- $Ax = \lambda x$ is a nonlinear equation; λ multiplies x. If we could discover λ, then the equation for x would be **linear**.
- We could write $\lambda I x$ in place of λx, and bring this term over to the left side:

$$(A - \lambda I)x = 0$$

The vector x is in the nullspace of $A - \lambda I$.
The number λ is chosen so that $A - \lambda I$ has a nullspace.

- We want a **nonzero** eigenvector x. The vector $x = 0$ always satisfies $Ax = \lambda x$, but it is useless.
- To be of any use, the nullspace of $A - \lambda I$ must contain vectors other than zero.
- In short, $\boldsymbol{A - \lambda I}$ **must be singular**.

The Solution of $Ax = \lambda x$

5A The number λ is an eigenvalue of A if and only if $A - \lambda I$ is singular:

$$\det(A - \lambda I) = 0. \tag{10}$$

This is the characteristic equation. Each λ is associated with eigenvectors x:

$$(A - \lambda I)x = 0 \qquad \text{or} \qquad Ax = \lambda x. \tag{11}$$

The Solution of $Ax = \lambda x$

- Example:

$$A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$$

we shift A by λI to make it singular:

Subtract λI $\quad A - \lambda I = \begin{bmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{bmatrix}$

Determinant $\quad |A - \lambda I| = (4-\lambda)(-3-\lambda) + 10 \quad$ or $\quad \lambda^2 - \lambda - 2$

- This is the **characteristic polynomial.**
- Its **roots**, where the **determinant is zero**, are the **eigenvalues**.

$$\lambda^2 - \lambda - 2 = (\lambda + 1)(\lambda - 2)$$

The Solution of $Ax = \lambda x$

Eigenvalues $$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{1 \pm \sqrt{9}}{2} = -1 \text{ and } 2.$$

- There are two eigen values, because a quadratic has two roots.
- The values $\lambda = -1$ and $\lambda = 2$ lead to a solution of $Ax = \lambda x$ or $(A - \lambda I)x = 0$.

$$\lambda_1 = -1 : \qquad (A - \lambda_1 I)x = \begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The solution (the first eigenvector) is any nonzero multiple of x_1:

$$\textbf{Eigenvector for } \lambda_1 \qquad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The Solution of $Ax = \lambda x$

The solution (the first eigenvector) is any nonzero multiple of x_1:

$$\textbf{Eigenvector for } \lambda_1 \qquad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The computation for λ_2 is done separately:

$$\lambda_2 = 2: \qquad (A - \lambda_2 I)x = \begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The second eigenvector is any nonzero multiple of x_2:

$$\textbf{Eigenvector for } \lambda_2 \qquad x_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}.$$

The Solution of $\boldsymbol{Ax = \lambda x}$

- The steps in solving $Ax = \lambda x$:

1. **Compute the determinant of $\boldsymbol{A - \lambda I}$.** With λ subtracted along the diagonal, this determinant is a polynomial of degree n. It starts with $(-\lambda)^n$.
2. **Find the roots of this polynomial.** The n roots are the eigenvalues of A.
3. **For each eigenvalue solve the equation $\boldsymbol{(A - \lambda I)x = 0}$.** Since the determinant is zero, there are solutions other than $x = 0$. Those are the eigenvectors.

The Solution of $\boldsymbol{Ax = \lambda x}$ (Recap)

- The key equation was $Ax = \lambda x$.
- Most vectors x will not satisfy such an equation.
- They **change direction** when multiplied by A, so that Ax is not a multiple of x.
- ***This means that only certain special numbers are eigenvalues, and only certain special vectors $\boldsymbol{x}$ are eigenvectors***.

Example 3. The eigenvalues are on the main diagonal when A is triangular.

$$\det(A-\lambda I)=\begin{vmatrix}1-\lambda & 4 & 5\\ 0 & \frac{3}{4}-\lambda & 6\\ 0 & 0 & \frac{1}{2}-\lambda\end{vmatrix}=(1-\lambda)(\tfrac{3}{4}-\lambda)(\tfrac{1}{2}-\lambda)$$

- The determinant is just the product of the diagonal entries.
- It is zero if $\lambda = 1, \lambda = \frac{3}{4}, or\ \lambda = \frac{1}{2}$
- The eigenvalues were already sitting along the main diagonal.

5B The ***sum*** of the n eigenvalues equals the sum of the n diagonal entries:

$$\textbf{Trace of} \quad A = \lambda_1 + \cdots + \lambda_n = a_{11} + \cdots + a_{nn}. \tag{15}$$

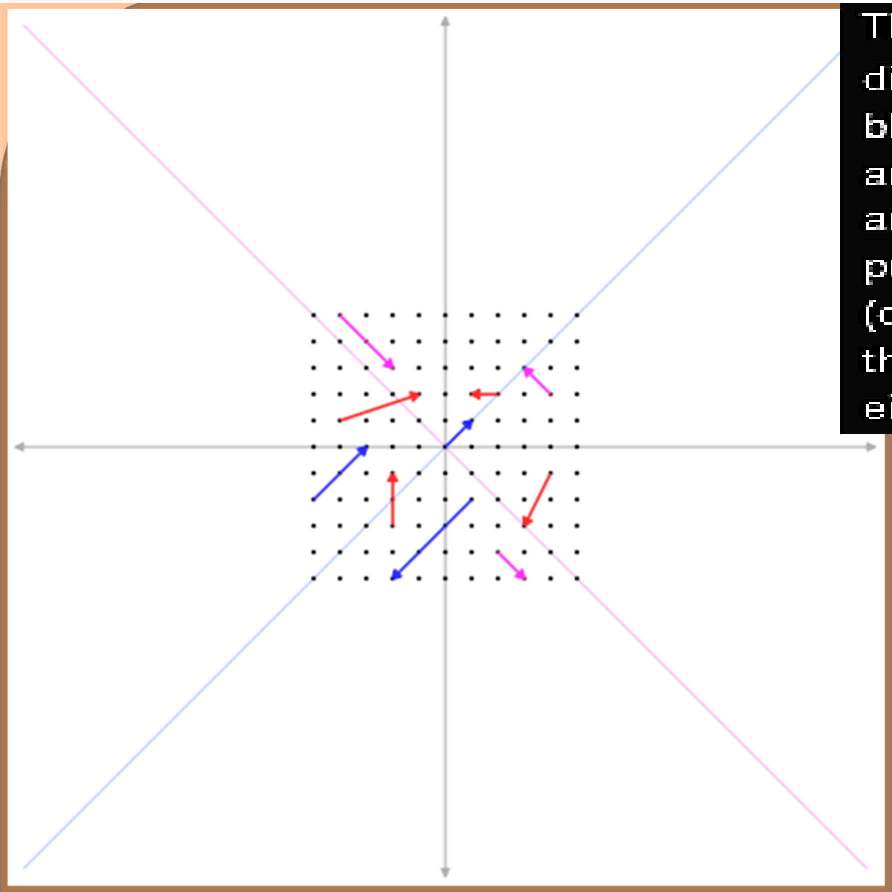
Furthermore, the ***product*** of the n eigenvalues equals the ***determinant*** of A.

For a 2 by 2 matrix, the trace and determinant tell us everything:

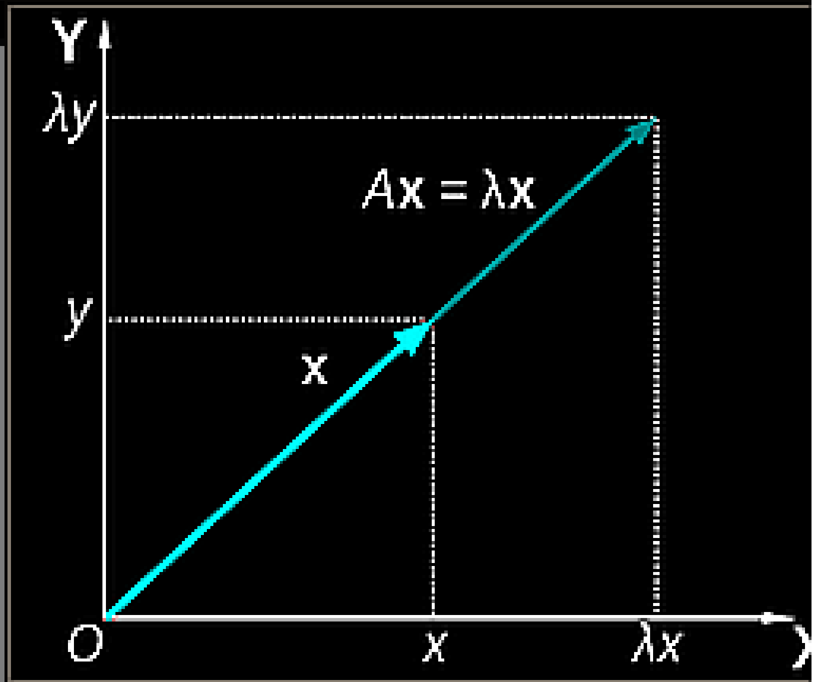
$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has trace $a + d$, and determinant $ad - bc$

$$\det(A - \lambda I) = \det \begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = \lambda^2$$

The eigenvalues are $\lambda =$



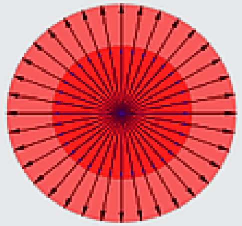
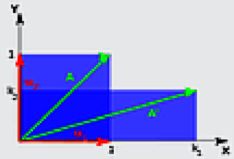
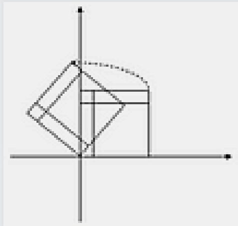
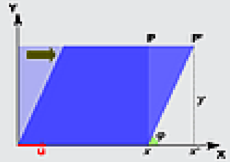
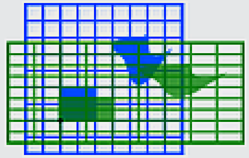
The transformation matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ preserves the direction of purple vectors parallel to $\mathbf{v}_{\lambda=1} = [1\ -1]^T$ and blue vectors parallel to $\mathbf{v}_{\lambda=3} = [1\ 1]^T$. The red vectors are not parallel to either eigenvector, so, their directions are changed by the transformation. The lengths of the purple vectors are unchanged after the transformation (due to their eigenvalue of 1), while blue vectors are three times the length of the original (due to their eigenvalue of 3).

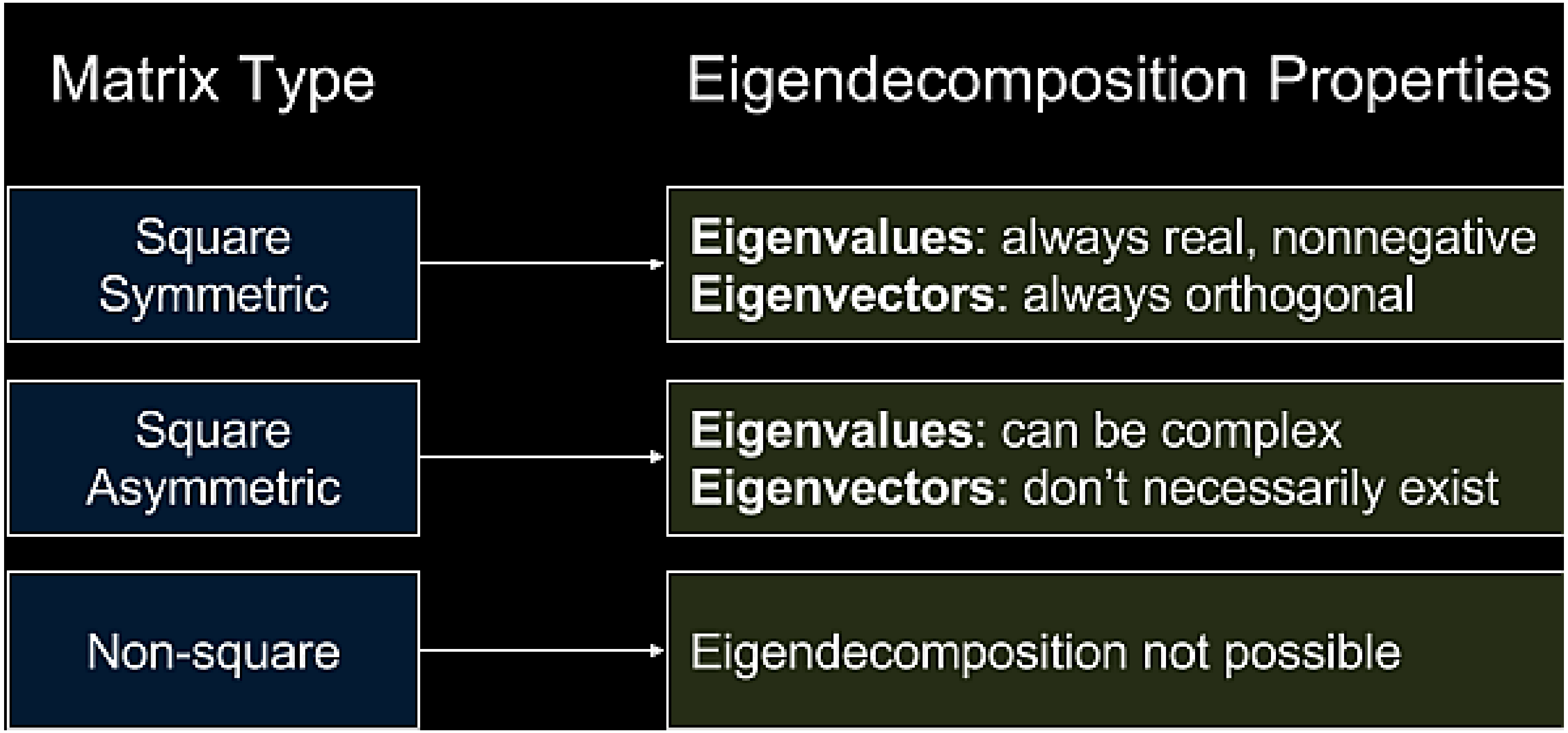


Matrix A acts by stretching the vector $\mathbf{x}$, not changing its direction, so $\mathbf{x}$ is an eigenvector of A.

https://www.geogebra.org/m/KuMAuEnd

Eigenvalues of geometric transformations

	Scaling	Unequal scaling	Rotation	Horizontal shear	Hyperbolic rotation
Illustration					
Matrix	$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$	$\begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{bmatrix}$
Characteristic polynomial	$(\lambda - k)^2$	$(\lambda - k_1)(\lambda - k_2)$	$\lambda^2 - 2\cos(\theta)\lambda + 1$	$(\lambda - 1)^2$	$\lambda^2 - 2\cosh(\varphi)\lambda + 1$
Eigenvalues, λ_i	$\lambda_1 = \lambda_2 = k$	$\lambda_1 = k_1$ $\lambda_2 = k_2$	$\lambda_1 = e^{i\theta} = \cos\theta + i\sin\theta$ $\lambda_2 = e^{-i\theta} = \cos\theta - i\sin\theta$	$\lambda_1 = \lambda_2 = 1$	$\lambda_1 = e^{\varphi} = \cosh\varphi + \sinh\varphi$ $\lambda_2 = e^{-\varphi} = \cosh\varphi - \sinh\varphi$
Algebraic mult., $\mu_i = \mu(\lambda_i)$	$\mu_1 = 2$	$\mu_1 = 1$ $\mu_2 = 1$	$\mu_1 = 1$ $\mu_2 = 1$	$\mu_1 = 2$	$\mu_1 = 1$ $\mu_2 = 1$
Geometric mult., $\gamma_i = \gamma(\lambda_i)$	$\gamma_1 = 2$	$\gamma_1 = 1$ $\gamma_2 = 1$	$\gamma_1 = 1$ $\gamma_2 = 1$	$\gamma_1 = 1$	$\gamma_1 = 1$ $\gamma_2 = 1$
Eigenvectors	All nonzero vectors	$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\mathbf{u}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ $\mathbf{u}_2 = \begin{bmatrix} 1 \\ +i \end{bmatrix}$	$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$


Matrix Type
Eigendecomposition Properties
Square Symmetric
Eigenvalues: always real, nonnegative
Eigenvectors: always orthogonal
Square Asymmetric
Eigenvalues: can be complex
Eigenvectors: don't necessarily exist
Non-square
Eigendecomposition not possible

Singular Value Decomposition

$A = U\Sigma V^T$ is known as the "**SVD**" or the ***singular value decomposition***.

The SVD is closely associated with the eigenvalue-eigenvector factorization $Q\Lambda Q^T$ of a positive definite matrix.

Any $m \times n$ matrix A can be factored into

$$A = U\Sigma V^{\mathrm{T}} = (\textbf{orthogonal})(\textbf{diagonal})(\textbf{orthogonal})$$

The columns of U ($m \times m$) are ***eigenvectors of AA^T***, and the columns of V ($n \times n$) are ***eigenvectors of A^TA***.

The r singular values on the diagonal of Σ ($m \times n$) are the ***square roots of the nonzero eigenvalues*** of both AA^T and A^TA.

See next few slides for variants →

Stack the (centered) observations into the rows of an $N \times p$ matrix $\mathbf{X}$. We construct the *singular value decomposition* of $\mathbf{X}$:

Sec. 14.5 – PP 535

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{14.54}$$

This is a standard decomposition in numerical analysis, and many algorithms exist for its computation (Golub and Van Loan, 1983, for example). Here $\mathbf{U}$ is an $N \times p$ orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$) whose columns $\mathbf{u}_j$ are called the *left singular vectors*; $\mathbf{V}$ is a $p \times p$ orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}_p$) with columns v_j called the *right singular vectors*, and $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ known as the *sin-*

Here $\mathbf{U}$ and $\mathbf{V}$ are $N \times p$ and $p \times p$ orthogonal matrices, with the columns of $\mathbf{U}$ spanning the column space of $\mathbf{X}$, and the columns of $\mathbf{V}$ spanning the row space. $\mathbf{D}$ is a $p \times p$ diagonal matrix, with diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ called the singular values of $\mathbf{X}$. If one or more values $d_j = 0$, $\mathbf{X}$ is singular.

Using the singular value decomposition we can write the least squares fitted vector as

$$\begin{aligned} \mathbf{X}\hat{\beta}^{\text{ls}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= \mathbf{U}\mathbf{U}^T\mathbf{y}, \end{aligned} \tag{3.46}$$

Strang - Sec. 6.3 – PP 367

Singular Value Decomposition: Any m by n matrix A can be factored into

$$A = U\Sigma V^{\mathrm{T}} = (\textbf{orthogonal})(\textbf{diagonal})(\textbf{orthogonal}).$$

The columns of U (m by m) are eigenvectors of AA^{T}, and the columns of V (n by n) are eigenvectors of $A^{\mathrm{T}}A$. The r singular values on the diagonal of Σ (m by n) are the square roots of the nonzero eigenvalues of both AA^{T} and $A^{\mathrm{T}}A$.

Hastie - Sec. 18.3.5 – PP 659

Given the $N \times p$ data matrix $\mathbf{X}$, let

$$\begin{aligned} \mathbf{X} &= \mathbf{UDV}^T & (18.12) \\ &= \mathbf{RV}^T & (18.13) \end{aligned}$$

be the singular-value decomposition (SVD) of $\mathbf{X}$; that is, $\mathbf{V}$ is $p \times N$ with orthonormal columns, $\mathbf{U}$ is $N \times N$ orthogonal, and $\mathbf{D}$ a diagonal matrix with elements $d_1 \geq d_2 \geq d_N \geq 0$. The matrix $\mathbf{R}$ is $N \times N$, with rows r_i^T.

Singular Value Decomposition

Remark 1.

- For positive definite matrices, Σ is Λ and $U\Sigma V^T$ is identical to $Q\Lambda Q^T$.
- For other symmetric matrices, any negative eigenvalues in Λ become positive in Σ.
- For complex matrices, Σ remains real but U and V become *unitary* (the complex version of orthogonal).

Remark 2.

U and V give orthonormal bases for all four fundamental subspaces:

first	r	columns of U:	**column space** of A
last	$m-r$	columns of U:	**left nullspace** of A
first	r	columns of V:	**row space** of A
last	$n-r$	columns of V:	**nullspace** of A

Singular Value Decomposition

Remark 3.

Eigenvectors of AA^T and A^TA must go into the columns of U and V:

$$AA^{\mathrm{T}} = (U\Sigma V^{\mathrm{T}})(V\Sigma^{\mathrm{T}}U^{\mathrm{T}}) = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}} \quad \text{and, similarly,} \quad A^{\mathrm{T}}A = V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}.$$

- U must be the eigenvector matrix for AA^T .
- The eigenvalue matrix in the middle is $\Sigma\Sigma^T$— which is $m \times m$ with $\sigma_1^2, \ldots, \sigma_r^2$ on the diagonal.
- From the $A^TA = V\Sigma^T\Sigma V^T$, the V matrix must be the eigenvector matrix for A^TA.

Singular Value Decomposition

Example 1.

This A has only one column: rank r = 1. Then Σ has only $\sigma_1 = 3$:

$$\textbf{SVD} \qquad A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = U_{3\times 3}\Sigma_{3\times 1}V^{\mathrm{T}}_{1\times 1}$$

$A^{\mathrm{T}}A$ is 1 by 1, whereas AA^{T} is 3 by 3. They both have eigenvalue 9 (whose square root is the 3 in Σ). The two zero eigenvalues of AA^{T} leave some freedom for the eigenvectors in columns 2 and 3 of U. We kept that matrix orthogonal.

Singular Value Decomposition

Example 2.

Now A has rank 2, and AA^T = $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ with $\lambda = 3$ and 1:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = U\Sigma V^{\mathrm{T}} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{matrix} /\sqrt{6} \\ /\sqrt{2} \\ /\sqrt{3} \end{matrix}$$

Notice √ 3 and √ 1. The columns of U are *left singular vectors* (unit eigenvectors of $\boldsymbol{AA^T}$).

The columns of V are *right singular vectors* (unit eigenvectors of $\boldsymbol{A^T A}$).

Let $A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$. Then

$$AA^T = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix}.$$

$$A^TA = \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix}$$

$$\begin{aligned} &= \det(xI - AA^T) = \begin{vmatrix} x-11 & -5 \\ -5 & x-11 \end{vmatrix} \\ &= (x-11)^2 - 25 \\ &= x^2 - 22x + 121 - 25 \\ &= x^2 - 22x + 96 \\ &= (x-16)(x-6). \end{aligned}$$

$$V = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & -\sqrt{2} & -1 \\ 0 & -\sqrt{2} & 2 \\ \sqrt{3} & \sqrt{2} & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix} = \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\right) \begin{bmatrix} \text{[illegible]} \end{bmatrix} \left(\frac{1}{\sqrt{17}} \begin{bmatrix} 1 & -4 \\ 4 & 1 \end{bmatrix}\right)$$

$$\begin{aligned} A &= \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right) \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix} \left(\frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ -\sqrt{2} & -\sqrt{2} & \sqrt{2} \\ -1 & 2 & 1 \end{bmatrix}\right) \end{aligned}$$

Applications of Singular Value Decomposition

Image Processing.

- Suppose a satellite takes a picture, and wants to send it to Earth.
- The picture may contain $1000 \times\ 1000$ "pixels"—a million little squares, each with a definite color.
- We can code the colors, and send back $1{,}000{,}000$ numbers.
- It is ***better to find the essential information inside the*** $\mathbf{1000} \times \mathbf{1000}$ ***matrix***, and send only that.

In SVD some σ's are significant and others are extremely small.

If we keep 20 and throw away 980, then we send only the corresponding 20 columns of U and V.

The other 980 columns are multiplied in $U\Sigma V^T$ by the small σ's that are being ignored. ***If only 20 terms are kept, we send 20 times 2000 numbers instead of a million (25 to 1 compression).***

The **conjugate transpose**, also known as the **Hermitian transpose**, of an m × n complex matrix A is an n × m matrix obtained by transposing A and applying complex conjugate on each entry (the complex conjugate of a + ib being a − i b, for real numbers a and b)

$$\boldsymbol{A}^{\mathrm{H}} = \left(\overline{\boldsymbol{A}}\right)^{\mathsf{T}} = \overline{\boldsymbol{A}^{\mathsf{T}}}$$

$$\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}} = \left(\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}}.$$

A ***matrix is full row rank*** when each of the rows of the matrix are linearly independent and full column rank when each of the columns of the matrix are linearly independent.

For a ***square matrix*** these two concepts are equivalent and we say the matrix is ***full rank if all rows and columns are linearly independen***t. A square matrix is full rank if and only if its determinant is nonzero.

For a ***non-square matrix*** with m rows and n columns, it will always be the case that either the rows or columns (whichever is larger in number) are linearly dependent. Hence when we say that a non-square matrix is full rank, we mean that the row and column rank are as high as possible, given the shape of the matrix. So, if there are ***more rows than columns (m > n), then the matrix is full rank if the matrix is full column rank.***

The ***rank of A equals the number of non-zero singular values***, which is the same as the number of non-zero diagonal elements ***in Σ*** in the singular value decomposition A = U Σ V*

If A is a matrix over the real numbers then the rank of A and the rank of its corresponding Gram matrix are equal. Thus, for real matrices

$$\operatorname{rank}(A^{\mathrm{T}}A) = \operatorname{rank}(AA^{\mathrm{T}}) = \operatorname{rank}(A) = \operatorname{rank}(A^{\mathrm{T}}).$$

suppose A is an $n \times m$ matrix and $n \neq m$. It must be that $rank(A^t) = rank(A) \leq \min(n, m) < \max(n, m)$.

Using the fact that $rank(AB) \leq rank(A)$ for any A, B for which the product is defined, we have that:

$$rank(AA^t) \leq rank(A) < \max(n, m)$$

$$rank(A^tA) \leq rank(A^t) < \max(n, m).$$

But it must be the case that the dimensions of AA^t or A^tA is $\max(n, m)$. Therefore at least one of them does not have full rank. For square matrices, not having full rank is equivalent to being singular.

Example 1

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

$$C = AA^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix}$$

$$D = A^T A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix}$$

$$\mathrm{rank(A)} = \mathrm{rank}(A^T) = \mathrm{rank(C)} = \mathrm{rank(D)} = 2$$

Example 2

$$A = \begin{bmatrix} 3 & 6 & 1 & 1 & 7 \\ 1 & 2 & 2 & 3 & 1 \\ 2 & 4 & 5 & 8 & 4 \\ 0 & 0 & 1 & 2 & 2 \end{bmatrix} \qquad \mathrm{A}^T = \begin{bmatrix} 3 & 1 & 2 & 0 \\ 6 & 2 & 4 & 0 \\ 1 & 2 & 5 & 1 \\ 1 & 3 & 8 & 2 \\ 7 & 1 & 4 & 2 \end{bmatrix}$$

$$C = AA^T = \begin{bmatrix} 96 & 27 & 71 & 17 \\ 27 & 19 & 48 & 10 \\ 71 & 48 & 125 & 29 \\ 17 & 10 & 29 & 9 \end{bmatrix} \qquad D = A^T A = \begin{bmatrix} 14 & 28 & 15 & 22 & 30 \\ 28 & 56 & 30 & 44 & 60 \\ 15 & 30 & 31 & 49 & 31 \\ 22 & 44 & 49 & 78 & 46 \\ 30 & 60 & 31 & 46 & 70 \end{bmatrix}$$

$$\mathrm{rank(A)} = \mathrm{rank}(A^T) = \mathrm{rank(C)} = \mathrm{rank(D)} = 3$$

Definition [edit]

Throughout this article, boldfaced unsubscripted $\mathbf{X}$ and $\mathbf{Y}$ are used to refer to random vectors, and unboldfaced subscripted X_i and Y_i are used to refer to scalar random variables.

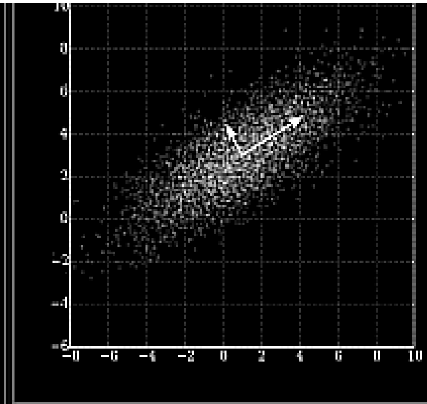
If the entries in the column vector

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$$

are random variables, each with finite variance and expected value, then the covariance matrix $\operatorname{K}_{\mathbf{X}\mathbf{X}}$ is the matrix whose (i, j) entry is the covariance[1]: p. 177

$$\operatorname{K}_{X_i X_j} = \operatorname{cov}[X_i, X_j] = \operatorname{E}[(X_i - \operatorname{E}[X_i])(X_j - \operatorname{E}[X_j])]$$

where the operator E denotes the expected value (mean) of its argument.



Sample points from a bivariate Gaussian distribution with a standarc deviation of 3 in roughly the lower left-upper right direction and of 1 in the orthogonal direction. Because the *x* and *y* components co-vary, the variances of x and y do not fully describe the distribution. A 2×2 covariance matrix is needed; the directions of the arrows correspond t the eigenvectors of this covariance matrix and their lengths to the square roots of the eigenvalues.

Conflicting nomenclatures and notations [edit]

Nomenclatures differ. Some statisticians, following the probabilist William Feller in his two-volume book *An Introduction to Probability Theory and Its Applications*,[2] call the matrix $\operatorname{K}_{\mathbf{X}\mathbf{X}}$ the **variance** of the random vector $\mathbf{X}$, because it is the natural generalization to higher dimensions of the 1-dimensional variance. Others call it the **covariance matrix**, because it is the matrix of covariances between the scalar components of the vector $\mathbf{X}$.

$$\operatorname{var}(\mathbf{X}) = \operatorname{cov}(\mathbf{X}, \mathbf{X}) = \operatorname{E}\left[(\mathbf{X} - \operatorname{E}[\mathbf{X}])(\mathbf{X} - \operatorname{E}[\mathbf{X}])^{\mathrm{T}}\right].$$

Both forms are quite standard, and there is no ambiguity between them. The matrix $\operatorname{K}_{\mathbf{X}\mathbf{X}}$ is also often called the *variance-covariance matrix*, since the diagonal terms are in fac variances.

By comparison, the notation for the cross-covariance matrix *between* two vectors is

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \operatorname{K}_{\mathbf{X}\mathbf{Y}} = \operatorname{E}\left[(\mathbf{X} - \operatorname{E}[\mathbf{X}])(\mathbf{Y} - \operatorname{E}[\mathbf{Y}])^{\mathrm{T}}\right].$$

Basic properties

For $\mathrm{K}_{\mathbf{XX}} = \operatorname{var}(\mathbf{X}) = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\mathrm{T}}\right]$ and $\mu_{\mathbf{X}} = \mathrm{E}[\mathbf{X}]$, where $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ is a n-dimensional random variable, the following basic properties apply:[4]

1. $\mathrm{K}_{\mathbf{XX}} = \mathrm{E}(\mathbf{XX}^{\mathrm{T}}) - \mu_{\mathbf{X}}\mu_{\mathbf{X}}{}^{\mathrm{T}}$
2. $\mathrm{K}_{\mathbf{XX}}$ is positive-semidefinite, i.e. $\mathbf{a}^T \mathrm{K}_{\mathbf{XX}}\, \mathbf{a} \geq 0 \quad \text{for all } \mathbf{a} \in \mathbb{R}^n$
3. $\mathrm{K}_{\mathbf{XX}}$ is symmetric, i.e. $\mathrm{K}_{\mathbf{XX}}^{\mathrm{T}} = \mathrm{K}_{\mathbf{XX}}$
4. For any constant (i.e. non-random) $m \times n$ matrix $\mathbf{A}$ and constant $m \times 1$ vector $\mathbf{a}$, one has $\operatorname{var}(\mathbf{AX} + \mathbf{a}) = \mathbf{A}\, \operatorname{var}(\mathbf{X})\, \mathbf{A}^{\mathrm{T}}$
5. If $\mathbf{Y}$ is another random vector with the same dimension as $\mathbf{X}$, then $\operatorname{var}(\mathbf{X} + \mathbf{Y}) = \operatorname{var}(\mathbf{X}) + \operatorname{cov}(\mathbf{X}, \mathbf{Y}) + \operatorname{cov}(\mathbf{Y}, \mathbf{X}) + \operatorname{var}(\mathbf{Y})$ where $\operatorname{cov}(\mathbf{X}, \mathbf{Y})$ is the cross-covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$.

For random vectors $\mathbf{X}$ and $\mathbf{Y}$, each containing random elements whose expected value and variance exist, the **cross-covariance matrix** of $\mathbf{X}$ and $\mathbf{Y}$ is defined by[1]: p.336

$$\mathrm{K}_{\mathbf{XY}} = \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \stackrel{\mathrm{def}}{=} \mathrm{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})^{\mathrm{T}}] \quad \textbf{(Eq.1)}$$

where $\mu_{\mathbf{X}} = \mathrm{E}[\mathbf{X}]$ and $\mu_{\mathbf{Y}} = \mathrm{E}[\mathbf{Y}]$ are vectors containing the expected values of $\mathbf{X}$ and $\mathbf{Y}$. The vectors $\mathbf{X}$ and $\mathbf{Y}$ need not have the same dimension, and either might be a scalar value.

The cross-covariance matrix is the matrix whose (i, j) entry is the covariance

$$\mathrm{K}_{X_i Y_j} = \operatorname{cov}[X_i, Y_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(Y_j - \mathrm{E}[Y_j])]$$

For the cross-covariance matrix, the following basic properties apply:[2]

1. $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \operatorname{E}[\mathbf{X}\mathbf{Y}^{\mathrm{T}}] - \mu_{\mathbf{X}}\mu_{\mathbf{Y}}^{\mathrm{T}}$
2. $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \operatorname{cov}(\mathbf{Y},\mathbf{X})^{\mathrm{T}}$
3. $\operatorname{cov}(\mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}) = \operatorname{cov}(\mathbf{X}_1,\mathbf{Y}) + \operatorname{cov}(\mathbf{X}_2,\mathbf{Y})$
4. $\operatorname{cov}(A\mathbf{X}+\mathbf{a}, B^{\mathrm{T}}\mathbf{Y}+\mathbf{b}) = A\,\operatorname{cov}(\mathbf{X},\mathbf{Y})\,B$
5. If $\mathbf{X}$ and $\mathbf{Y}$ are independent (or somewhat less restrictedly, if every random variable in $\mathbf{X}$ is uncorrelated with every random variable in $\mathbf{Y}$), then $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = 0_{p\times q}$

where $\mathbf{X}$, $\mathbf{X_1}$ and $\mathbf{X_2}$ are random $p \times 1$ vectors, $\mathbf{Y}$ is a random $q \times 1$ vector, $\mathbf{a}$ is a $q \times 1$ vector, $\mathbf{b}$ is a $p \times 1$ vector, A and B are $q \times p$ matrices of constants, and $0_{p\times q}$ is a $p \times q$ matrix of zeroes.

Given a sample consisting of *n* independent observations $x_1,\ldots,x_n$ of a *p*-dimensional random vector $X \in \mathbf{R}^{p\times 1}$ (a *p*×1 column-vector), an unbiased estimator of the (*p*×*p*) covariance matrix

$$\Sigma = \operatorname{E}\left[(X-\operatorname{E}[X])\,(X-\operatorname{E}[X])^{\mathrm{T}}\right]$$

is the sample covariance matrix

$$\mathbf{Q} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(x_i-\overline{x})^{\mathrm{T}},$$

where x_i is the *i*-th observation of the *p*-dimensional random vector, and the vector

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the sample mean. This is true regardless of the distribution of the random variable *X*, provided of course that the theoretical means and covariances exist. The reason

$$\mathrm{Var}[bX] = \mathrm{E}[(bX - \mathrm{E}[bX])(bX - \mathrm{E}[bX])^{\top}]$$
$$= \mathrm{E}[(bX - b\mathrm{E}[X])(bX - b\mathrm{E}[X])^{\top}]$$

Which matrices are covariance matrices?

let $\mathbf{b}$ be a $(p \times 1)$ real-valued vector, then

$$\operatorname{var}(\mathbf{b}^{\mathrm{T}}\mathbf{X}) = \mathbf{b}^{\mathrm{T}} \operatorname{var}(\mathbf{X})\mathbf{b},$$

which must always be nonnegative, since it is the variance of a real-valued random variable, so a covariance matrix is always a positive-semidefinite matrix.

The above argument can be expanded as follows:

$$w^{\mathrm{T}}\,\mathrm{E}\left[(\mathbf{X}-\mathrm{E}[\mathbf{X}])(\mathbf{X}-\mathrm{E}[\mathbf{X}])^{\mathrm{T}}\right]w = \mathrm{E}\left[w^{\mathrm{T}}(\mathbf{X}-\mathrm{E}[\mathbf{X}])(\mathbf{X}-\mathrm{E}[\mathbf{X}])^{\mathrm{T}}w\right]$$
$$= \mathrm{E}\left[\left(w^{\mathrm{T}}(\mathbf{X}-\mathrm{E}[\mathbf{X}])\right)^{2}\right] \geq 0,$$

where the last inequality follows from the observation that $w^{\mathrm{T}}(\mathbf{X}-\mathrm{E}[\mathbf{X}])$ is a scalar.

Conversely, every symmetric positive semi-definite matrix is a covariance matrix. To see this, suppose M is a $p \times p$ symmetric positive-semidefinite matrix. From the finite-dimensional case of the spectral theorem, it follows that M has a nonnegative symmetric square root, which can be denoted by $\mathrm{M}^{1/2}$. Let $\mathbf{X}$ be any $p \times 1$ column vector-valued random variable whose covariance matrix is the $p \times p$ identity matrix. Then

$$\operatorname{var}(\mathbf{M}^{1/2}\mathbf{X}) = \mathbf{M}^{1/2}\,\operatorname{var}(\mathbf{X})\,\mathbf{M}^{1/2} = \mathbf{M}.$$