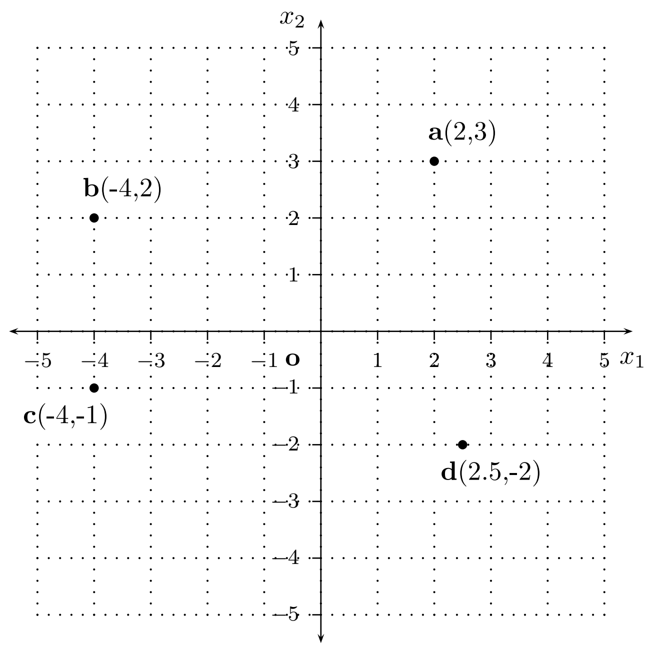
Scalers, Vectors & Matrices

A single element, value, or quantity is referred to as *scaler*. For example, number 3 is scaler, so is covariance between x_1 and x_2, denoted by $\text{Cov}(x_1, x_2)$. Finally, estimated parameter $\hat{b}_1$ is also scaler.
Two or more scalers written together in a row or a column, form a row or a column *vector*. The *order* of a row vector is $1 \times c$, where 1 indicates one row and c is the number of columns. The order of a column vector is $r \times 1$. Geometrically, when vectors are in two dimensions, it is possible to interpret vectors in physical space. Consider four points, $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ in two dimensions. Location of each point is generally specified relative to some reference point and reference axes. Following figure shows these four points when reference point is $\mathbf{o}$(0,0).



A *matrix* is defined as a rectangular array of elements (vectors or scalers) arranged in rows and columns as in

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ -4 & 2 \\ -4 & -1 \\ 2.5 & -2 \end{bmatrix}$$

and in general,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rc} \end{bmatrix}$$

There are rc elements arranged in r rows and c columns, and it is order of r by c, and it is written as $r \times c$. The element in the ith row and jth column is denoted by a_{ij}. The matrix is designated by a capital letter, **A**.

Special forms of Matrices

- **Vector**, a matrix with only a single row or of order $1 \times c$ is called row vector. Also a matrix with only a single column or of order $r \times 1$ is called column vector. Vectors will be designated by a lowercase letter, **b**. For example, a row vector

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 & \cdots & b_c \end{bmatrix}$$

while a column vector or a matrix of order $r \times 1$ is a

$$\mathbf{c} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}.$$

- **Square matrix**, matrix with same number of rows and columns, that is, $r = c$ or matrix of size $r \times r$ or $c \times c$.

- **Identity matrix** is a square matrix with ones along the diagonal and zeros everywhere else.

$$\mathbf{I}_r = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- **Scaler matrix** is similar to identity matrix but a common constant or scaler along diagonal and zero everywhere else. For example,

$$k\mathbf{I} = \begin{bmatrix} k & 0 & \cdots & 0 \\ 0 & k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k \end{bmatrix}$$

- **Diagonal matrix** is also similar to identity matrix but elements along diagonal are different and zero everywhere else. For example,

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_r \end{bmatrix}$$

- **Square Symmetric Matrix** is square matrix with $a_{ij} = a_{ji}$ for all $i \neq j$.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1c} \\ a_{12} & a_{22} & \cdots & a_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1c} & a_{2c} & \cdots & a_{cc} \end{bmatrix}$$

- **Null Matrix** is matrix with all elements are zero and denoted by $\mathbf{0}$.

Matrix Operations

- **Equality of two matrices** Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be equal when they are of the same order and $a_{ij} = b_{ij}$ for all i and j.

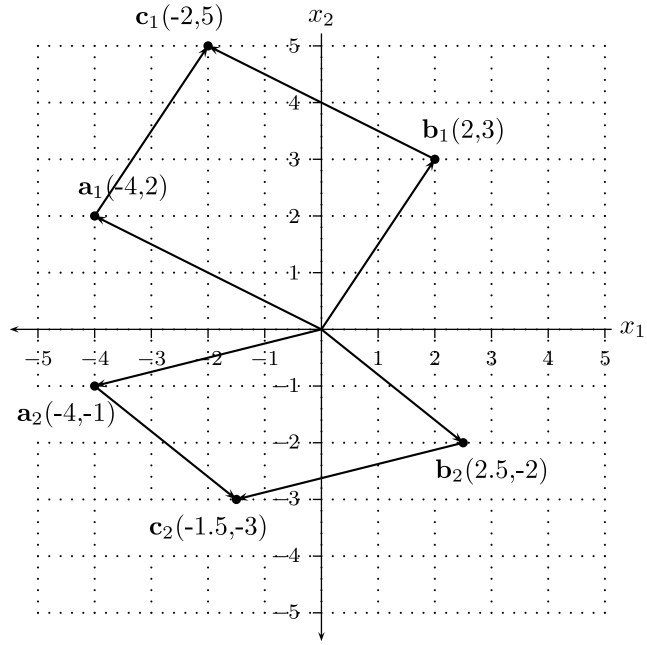
- **Addition and subtraction of two matrices** is possible if $\mathbf{A}$ and $\mathbf{B}$ are of the same order. That is, $\mathbf{C} = \mathbf{A} + \mathbf{B}$. Consider following example,

$$\mathbf{A} = \begin{bmatrix} -4 & 2 \\ -4 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 3 \\ 2.5 & -2 \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{C} = \mathbf{A} + \mathbf{B} &= \begin{bmatrix} -4+2 & 2+3 \\ -4+2.5 & -1-2 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 5 \\ -1.5 & -3 \end{bmatrix} \end{aligned}$$

Geometrically, additions of two matrices is explained as follows. Consider matrix $\mathbf{A}$ as composed of two vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$ and matrix $\mathbf{B}$ as $\mathbf{b}_1$ and $\mathbf{b}_2$. Then addition involves taking summation along same axis for $\mathbf{a}_1$ and $\mathbf{b}_1$ as well as $\mathbf{a}_2$ and $\mathbf{b}_2$. The resulting matrix then is denoted by vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ on the graph.



- **Scaler multiplication** require that each element a_{ij} be multiplied by a scaler. If k is a scaler, then $k\mathbf{A}$ is $k \times a_{ij}$. For example,

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix} \quad \text{and} \quad k = 2$$

then

$$\begin{aligned} k\mathbf{A} &= \begin{bmatrix} 2 \times 3 & 2 \times 2 \\ 2 \times 2 & 2 \times 0 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 4 \\ 4 & 0 \end{bmatrix} \end{aligned}$$

- **Matrix multiplication** is possible if two matrices are conformable. That is, the number of elements in a row of the first matrix has to be equal to the number of elements in a column of the second matrix. If $\mathbf{A}$ is of order $m \times n$ and $\mathbf{B}$ is order $n \times p$, then the product $\mathbf{AB}$ is defined to be a matrix of order $m \times p$ whose ijth element is

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

In $\mathbf{AB}$ note that $\mathbf{A}$ being postmultiplied by $\mathbf{B}$ or to $\mathbf{A}$ premultiplying $\mathbf{B}$. $\mathbf{BA}$ is usually different from $\mathbf{AB}$ and may not exist. The two products $\mathbf{AB}$ and $\mathbf{BA}$ will exist only if the matrices are of order $m \times n$ and $n \times m$. In such case, the first product will be of order $m \times m$ and the second $n \times n$.

Consider following example,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

Then $\mathbf{AB}$ is a 2×2 or

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

while $\mathbf{BA}$ is a 3×3 or

$$\mathbf{BA} = \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} & b_{11}a_{13} + b_{12}a_{23} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} & b_{21}a_{13} + b_{22}a_{23} \\ b_{31}a_{11} + b_{32}a_{21} & b_{31}a_{12} + b_{32}a_{22} & b_{31}a_{13} + b_{32}a_{23} \end{bmatrix}$$

Determine $\mathbf{AB}$ and $\mathbf{BA}$ for following matrices.

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

Answers:

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 2 \times 3 + 3 \times 1 + 5 \times 0 & 2 \times -1 + 3 \times 0 + 5 \times 2 \\ \times \;+\; \times \;+\; \times & \times \;+\; \times \;+\; \times \end{bmatrix} \\ &= \begin{bmatrix} 9 & 8 \\ 2 & 3 \end{bmatrix} \end{aligned}$$

Similarly verify that

$$\begin{aligned} \mathbf{BA} &= \begin{bmatrix} \times \;+\; \times & \times \;+\; \times & \times \;+\; \times \\ \times \;+\; \times & \times \;+\; \times & \times \;+\; \times \\ \times \;+\; \times & \times \;+\; \times & \times \;+\; \times \end{bmatrix} \\ &= \begin{bmatrix} 5 & 10 & 13 \\ 2 & 3 & 5 \\ 2 & -2 & 4 \end{bmatrix} \end{aligned}$$

- **Rules for Matrix Additions and Multiplication** are not same as those apply for scalers. Following are important rules to remember.

  1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. This is commutative law of addition.

2. $\mathbf{AB} \neq \mathbf{BA}$. In general, commutative law of multiplication does not hold. There are special square matrices for which commutative law holds. If the matrices are of order $m \times n$ and $n \times m$ then both products will exist, but they will of different orders and hence cannot be equal. If both matrices are square of same order, then both products will exist and will be of the same order but not necessarily equal.
3. $(\mathbf{A}+\mathbf{B})+\mathbf{C} = \mathbf{A}+(\mathbf{B}+\mathbf{C})$. The associative law of addition holds.
4. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. The associative law of multiplication holds.
5. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$. This is called distributive law and it holds. Similarly, distributive law of scaler multiplication also holds. That is $c(\mathbf{A}+\mathbf{B}) = c\mathbf{A}+c\mathbf{B}$ and $(c+d)\mathbf{A} = c\mathbf{A}+d\mathbf{A}$.

- **Transpose** of matrix is obtained by interchanging rows and columns. Suppose $\mathbf{A}$ is original matrix, then the first row of $\mathbf{A}$ becomes the first column of the transpose, the second row of $\mathbf{A}$ becomes the second column of the transpose. In general, the jith element in the transpose is the ijth element of the original matrix. The transposed matrix is indicated by $\mathbf{A}'$ or $\mathbf{A}^t$. Consider following example,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$
$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

Note here that $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ result in two different symmetric matrices. That is,

$$\mathbf{AA}' = \begin{bmatrix} a_{11}^2+a_{12}^2+a_{13}^2 & a_{11}a_{21}+a_{12}a_{22}+a_{13}a_{23} \\ a_{11}a_{21}+a_{12}a_{22}+a_{13}a_{23} & a_{21}^2+a_{22}^2+a_{23}^2 \end{bmatrix}$$
$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} a_{11}^2+a_{21}^2 & a_{11}a_{12}+a_{21}a_{22} & a_{11}a_{13}+a_{21}a_{23} \\ a_{11}a_{12}+a_{21}a_{22} & a_{12}^2+a_{22}^2 & a_{12}a_{13}+a_{22}a_{23} \\ a_{11}a_{13}+a_{21}a_{23} & a_{12}a_{13}+a_{22}a_{23} & a_{13}^2+a_{23}^2 \end{bmatrix}$$

Verify above results for

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 5 \\ 1 & -1 & 2 \end{bmatrix}$$

$$\mathbf{AA}' = \begin{bmatrix} \times + \times + \times & \times + \times + \times \\ \times + \times + \times & \times + \times + \times \end{bmatrix}$$
$$= \begin{bmatrix} 38 & 9 \\ 9 & 6 \end{bmatrix}$$

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} \times + \times & \times + \times & \times + \times \\ \times + \times & \times + \times & \times + \times \\ \times + \times & \times + \times & \times + \times \end{bmatrix}$$
$$= \begin{bmatrix} 5 & 5 & 12 \\ 5 & 10 & 13 \\ 12 & 13 & 29 \end{bmatrix}$$

- **Rules about transposed matrices**
  1. $(\mathbf{A}')' = \mathbf{A}$.
  2. $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.
  3. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.
  4. If $\mathbf{A}' = \mathbf{A}$, then $\mathbf{A}$ is a symmetric matrix.

If $\mathbf{x}$ is a column vector of n elements, then $\mathbf{x}'\mathbf{x}$ is sum of squares. That is because,

$$\mathbf{x}'\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
$$= x_1^2 + x_2^2 + \cdots + x_n^2.$$

Note that in above application, each element contributes equally to summation. If we want to include weighted sum, then we need to include weighting matrix, $\mathbf{W}$. That is,

$$\mathbf{x}'\mathbf{W}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} w_{11}x_1 \\ w_{22}x_2 \\ \cdots \\ w_{nn}x_n \end{bmatrix}$$
$$= w_{11}x_1^2 + w_{22}x_2^2 + \cdots + w_{nn}x_n^2$$

A very important function is obtained when we consider the general form of $\mathbf{x}'\mathbf{W}\mathbf{x}$, without the restriction that $\mathbf{W}$ is a diagonal matrix but postulating simply that $\mathbf{W}$ is symmetric. Note that for 2×2, $\mathbf{x}'\mathbf{W}\mathbf{x}$ would be

$$\mathbf{x}'\mathbf{W}\mathbf{x} = w_{11}x_1^2 + w_{22}x_2^2 + 2w_{12}x_1x_2$$

and for 3×3 case,

$$\begin{aligned}\mathbf{x}'\mathbf{W}\mathbf{x} &= w_{11}x_1^2 + w_{22}x_2^2 + w_{33}x_3^2 + \\ &\quad 2w_{12}x_1x_2 + 2w_{13}x_1x_3 + 2w_{23}x_2x_3.\end{aligned}$$

This quadratic form also has application in portfolio selection model where $\mathbf{W}$ is variance-covariance of returns while x_i reflect fraction of portfolio invested in ith investment instrument. If r_i is the expected return on ith investment instrument, then portfolio manager is expected to maximize $\mathbf{rx}$ subject to $\sum_{i=1}^{n} x_i = 1$ and $\mathbf{x}'\mathbf{W}\mathbf{x} \leq T$ where T is risk threshold, that is variation in returns that are not acceptable to portfolio manager.

- **Trace** of a square matrix is sum of the elements in the principal diagonal. That is, for $n \times n$ matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$. Note that the trace of $\mathbf{A}'\mathbf{A}$, denoted by $\mathrm{tr}(\mathbf{A}'\mathbf{A})$ is equal to $\mathrm{tr}(\mathbf{A}\mathbf{A}')$.

  For matrics $\mathbf{A}$ and $\mathbf{B}$ of order $m \times n$ and $p \times q$, respectively, following properties apply.

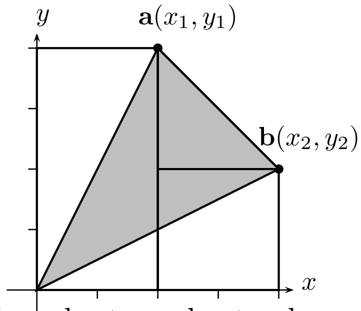
$$\begin{aligned}\mathrm{tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}) & m = n = p = q \\ \mathrm{tr}(c\mathbf{A}) &= c\ \mathrm{tr}(\mathbf{A}) & m = n \\ \mathrm{tr}(\mathbf{A}') &= \mathrm{tr}(\mathbf{A}) & m = n \\ \mathrm{tr}(\mathbf{AB}) &= \mathrm{tr}(\mathbf{BA}) & m = q, n = p\end{aligned}$$

- **Determinant** is a scaler quantity associated with any square matrix $\mathbf{A}$ and denoted by $\det(\mathbf{A})$ or $\mid \mathbf{A} \mid$. This scaler is obtained by summing various products of the elements of $\mathbf{A}$. For example, if matrix $\mathbf{A}$ 2×2 then

$$\mid \mathbf{A} \mid = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

  and if matrix is 3×3, then determinant is

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



Geometrically, determinant is a constant times area, or volume enclosed by points. For example, if there are two points $\mathbf{a}(x_1, y_1)$ and $\mathbf{b}(x_2, y_2)$ as shown in figure. Then

$$\mid \mathbf{A} \mid = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1$$

and area enclosed by[1] **a** and **b** is $\frac{1}{2} \mid x_2y_1 - x_1y_2 \mid$.

In order to understand process of obtaining the general expression, we need two more concepts, *minor* and *Co-factor*.

- **Minor** of the element a_{ij} from matrix $\mathbf{A}$ is the determinant of the matrix formed by deleting the ith row and jth column of matrix $\mathbf{A}$ and it is denoted by $\mathbf{A}_{ij}$.

- **Co-factor** of a_{ij} is the minor multiplied by $(-1)^{i+j}$ or co-factor, c_{ij} is equal $(-1)^{i+j} \mid \mathbf{A}_{ij} \mid$. These two quantities can be used to evaluate determinant of any square matrix. Thus,

  $$\mid \mathbf{A} \mid = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{in}c_{in}.$$

  For example,

  $$\begin{aligned} \mid \mathbf{A} \mid &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} \\ &= a_{11}(-1)^{(1+1)} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)^{(1+2)} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(-1)^{(1+3)} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\quad + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} \end{aligned}$$

- **Properties of Determinants**

  1. If $\mathbf{D}$ is a diagonal matrix with n rows and columns, then $\mid \mathbf{D} \mid = \prod_{i=1}^{n} d_{ii}$.
  2. If $\mathbf{I}_r$ is an identity matrix with r rows, then $\mid \mathbf{I}_r \mid = 1$.
  3. $\mid \mathbf{A}' \mid = \mid \mathbf{A} \mid$. That is, changing rows to columns does not alter value of determinant.
  4. $\mid \mathbf{AB} \mid = \mid \mathbf{A} \mid\mid \mathbf{B} \mid$.
  5. If each element in $\mathbf{A}$ is multiplied by k, the determinant of $\mathbf{A}$ is multiplied by k.
  6. If each element in $\mathbf{A}$ is multiplied by k, the determinant of $\mathbf{A}$ is multiplied by k^n where n is number of elements in matrix $\mathbf{A}$.
  7. If $\mid \mathbf{A} \mid = 0$, then $\mathbf{A}$ is said to be a singular matrix. That is, at least one row or one column is linear combination of some other row(s) or column(s).

[1]Note that area enclosed by two points and the origin is two rectangles, plus a triangle minus two triangles. That is, Area $= x_1y_1 + (x_2 - x_1)y_2 + \frac{1}{2}(x_2 - x_1)(y_1 - y_2) - \frac{1}{2}x_1y_1 - \frac{1}{2}x_2y_2$. After simplifying we get desired result.

Suppose that matrix $\mathbf{A}$ is

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

then show that $\mid \mathbf{A} \mid$ is equal to -14.

- **Inverse of matrix A** exists only if $\mathbf{A}$ is nonsingular. The inverse or reciprocal of a matrix is denoted by $\mathbf{A}^{-1}$. Note that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

That is, pre or post multiplication by inverse to original matrix results in identity matrix. To obtain inverse of matrix $\mathbf{A}$, first transpose of co-factors (adjoint of $\mathbf{A}$) is obtained. Then, adjoint of matrix $\mathbf{A}$ is divided by determinant of matrix $\mathbf{A}$. That is,

$$\mathbf{A}^{-1} = \frac{1}{\mid \mathbf{A} \mid}(\text{adj}\mathbf{A}).$$

Suppose that matrix $\mathbf{A}$ is

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

then show that inverse of $\mathbf{A}$ is

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

and verify that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_3$. Let us denote co-factor matrix by $\mathbf{C}$, then

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 3 & 3 \\ 1 & 2 \end{vmatrix} & \begin{vmatrix} 3 & 2 \\ 1 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 1 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 3 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -3 & 1 \\ -3 & 1 & 1 \\ 4 & 0 & -4 \end{bmatrix}. \end{aligned}$$

Note that $\mid \mathbf{A} \mid$ is equal to -4. Thus,

$$(\text{adj}\mathbf{A}) = \mathbf{C}' = \begin{bmatrix} 1 & -3 & 4 \\ -3 & 1 & 0 \\ 1 & 1 & -4 \end{bmatrix}$$

and hence the result.

- **Properties of Inverses**

  1. Inverse of a symmetric matrix is also symmetric.
  2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$. Inverse of a transpose is equal to transpose of inverse.
  3. $(\mathbf{AB})^{-1} = (\mathbf{B}^{-1}\mathbf{A}^{-1})$.
  4. If k is a constant, then $(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$.
  5. If $\mathbf{D}$ is a diagonal matrix, then

$$\mathbf{D}^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_r} \end{bmatrix}.$$

  6. $\mid \mathbf{A}^{-1} \mid = \frac{1}{\mid \mathbf{A} \mid}$.

- **Finding Solution to Linear Equations** can be performed using inverse and determinants. Consider following following three equations with three unknowns (x_1, x_2, and x_3).

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_1 - x_2 + x_3 &= 2 \\ 2x_1 + x_2 - x_3 &= 1 \end{aligned}$$

  This may be written in compact form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 1 \end{bmatrix}$$

$$\mathbf{Ax} = \mathbf{h}$$

  We can see that to obtain unique solution to above simultaneous equations, we must be able to compute $\mathbf{x} = \mathbf{A}^{-1}\mathbf{h}$. We know that to compute $\mathbf{A}^{-1}$, we need determinant ($\mid \mathbf{A} \mid$) to be non-zero. We also know that

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{\mid \mathbf{A} \mid}\text{adj}(\mathbf{A}) \\ &= \frac{1}{\mid \mathbf{A} \mid} \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \end{aligned}$$

where c_{ij} is transposed element of co-factor matrix. Using these results, we can write solution to simultanous equations as

$$
\begin{aligned}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \frac{1}{\mid \mathbf{A} \mid} \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \\
x_1 &= \frac{h_1 c_{11} + h_2 c_{21} + h_3 c_{31}}{\mid \mathbf{A} \mid} \\
x_2 &= \frac{h_1 c_{12} + h_2 c_{22} + h_3 c_{32}}{\mid \mathbf{A} \mid} \\
x_3 &= \frac{h_1 c_{13} + h_2 c_{23} + h_3 c_{33}}{\mid \mathbf{A} \mid}
\end{aligned}
$$

Above solution also could be written with original matrix elements. That is,

$$
\begin{aligned}
x_1 &= \frac{\begin{vmatrix} h_1 & a_{12} & a_{13} \\ h_2 & a_{22} & a_{23} \\ h_3 & a_{32} & a_{33} \end{vmatrix}}{\mid \mathbf{A} \mid} \\
x_2 &= \frac{\begin{vmatrix} a_{11} & h_1 & a_{13} \\ a_{21} & h_2 & a_{23} \\ a_{31} & h_3 & a_{33} \end{vmatrix}}{\mid \mathbf{A} \mid} \\
x_3 &= \frac{\begin{vmatrix} a_{11} & a_{12} & h_1 \\ a_{21} & a_{22} & h_2 \\ a_{31} & a_{32} & h_3 \end{vmatrix}}{\mid \mathbf{A} \mid}
\end{aligned}
$$

Numerically, note that

$$
\begin{aligned}
\mid \mathbf{A} \mid &= 1 \begin{vmatrix} -1 & 1 \\ 1 & -1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} + 1 \begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} \\
&= 0 + 3 + 3 \\
&= 6
\end{aligned}
$$

We can now compute the solution. That is,

$$
x_1 = \frac{\begin{vmatrix} 6 & 1 & 1 \\ 2 & -1 & 1 \\ 1 & 1 & -1 \end{vmatrix}}{6} = \frac{6 \times (1-1) - 1 \times (-2-1) + 1 \times (2+1)}{6} = 1
$$

$$x_2 = \frac{\begin{vmatrix} 1 & 6 & 1 \\ 1 & 2 & 1 \\ 2 & 1 & -1 \end{vmatrix}}{6} = \frac{1\times(-1-2) - 6\times(-1-2) + 1\times(1-4)}{6} = 2$$

$$x_3 = \frac{\begin{vmatrix} 1 & 1 & 6 \\ 1 & -1 & 2 \\ 2 & 1 & 1 \end{vmatrix}}{6} = \frac{1\times(-1-2) - 1\times(1-4) + 6\times(1+2)}{6} = 3$$

There are two things to note. First, we can solve a system of equations containing a large set of variables without much difficulty. Second, in most instances, we can proceed to solve these equations unless the determinant of coefficients or $|\mathbf{A}| \neq 0$. One easy way to detect that determinant is zero is by examining, if there are one or more row or column that are a linear combination of each other. For example, consider

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -2 \\ 2 & -1 & 4 \\ 2 & 6 & -4 \end{bmatrix}.$$

In this instance, the last row is twice that of the first row and $|\mathbf{A}|$ is $1\times(4-24) - 3\times(-8-8) - 2\times(12+2)$ is zero. There are, however, complex forms of situation where determinant is zero. Consider

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 3 & -2 \\ 2 & -1 & 4 \\ 1 & -11 & 14 \end{bmatrix} \\ |\mathbf{A}| &= 1\times(-14+44) - 3\times(28-4) - 2\times(-22+1) = 30 - 72 + 42 = 0. \end{aligned}$$

In such situation, matrix **A** may provide solution to not all of unknowns. We will revisit this issue when we discuss problem of finding eigenvalues and eigenvectors below.

Rank of a Matrix

Linear Dependence

Consider the set of homogeneous equations

$$\mathbf{Ax} = \mathbf{0}$$

where **A** is an $m \times n$ matrix of known constants and **x** a column vector of n unknowns. There are two types of solution to this problem. First, the trivial one, that is, $\mathbf{x} = \mathbf{0}$. That means, every element of **x** is zero. Second solution is more complicated. There are some elements of **x** are non-zero while the others are function of non-zero x's. In first case, matrix **A** is termed *linearly independent* while the second case, matrix $\mathbf{A}$ is termed *linearly dependent*. To put differently, suppose the columns of matrix $\mathbf{A}$ are denoted by $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$, if there exists a set of a scalers $x_1, x_2, \cdots, x_n$ with all $x_i \neq 0$ for all i and satisfy

$$\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_n x_n = 0$$

then the matrix $\mathbf{A}$ is linearly dependent. Note that to decide which is possibility, we need to evaluate determinant of $\mathbf{A}$ or $\mid \mathbf{A} \mid$. Consider first situation when $m = n$ or matrix $\mathbf{A}$ is a square matrix. If determinant is non-zero (positive or negative), then we have the first case and if determinant is zero, then we have the second case. To illustrate this further, Consider following set of equations:

$$\begin{bmatrix} 1 & 2 & -3 \\ 2 & 0 & 2 \\ 4 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Note that

$$\begin{aligned} \mid \mathbf{A} \mid &= 1 \begin{vmatrix} 0 & 2 \\ 1 & 3 \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 4 & 3 \end{vmatrix} - 3 \begin{vmatrix} 2 & 0 \\ 4 & 1 \end{vmatrix} \\ &= -2 + 4 - 6 \\ &= -4 \end{aligned}$$

In this case, we may conclude that the only solution is $x_1 = x_2 = x_3 = 0$. Suppose we change matrix $\mathbf{A}$ to

$$\begin{bmatrix} 1 & 2 & -3 \\ 2 & 0 & 2 \\ 4 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

In this case,

$$\begin{aligned} \mid \mathbf{A} \mid &= 1 \begin{vmatrix} 0 & 2 \\ 1 & 2 \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 4 & 2 \end{vmatrix} - 3 \begin{vmatrix} 2 & 0 \\ 4 & 1 \end{vmatrix} \\ &= -2 + 8 - 6 \\ &= 0 \end{aligned}$$

This means that it may be possible to find a set of x's, not all zero, such that

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Suppose we assumed that $x_3 = 1$ and used first two equations to solve for x_1 and x_2. Then, we would get $x_1 = -1$ and $x_2 = 2$ or any scaler multiple of these values will satisfy above equation. This leads to definition of rank and some properties of ranks.

The rank of matrix, denoted by $\rho(\mathbf{A})$ is defined as the maximum number of linearly independent columns in $\mathbf{A}$. That is, the rank of a matrix $\mathbf{A}$ is said to be r, if

- every square submatrix of order $r+1$ is singular, and
- there is at least one square submatrix of order r which is non-singular.

Theorems about ranks

1. If matrix $\mathbf{A}$ is $m \times n$, $\rho(\mathbf{A})$ is less than equal to $\min(m, n)$.
2. If there are two matrices, $\mathbf{A}$ and $\mathbf{B}$ and their product exists, then $\rho(\mathbf{AB}) = \min[\rho(\mathbf{A}), \rho(\mathbf{B})]$

Two alternatives for determining rank of a matrix are provided. First method is based on minor of a matrix while other is based on echelon (normal) form of a matrix. Both methods give identical results but method based on minor becomes far more elaborate and time consuming for a larger matrices. Recall that the rank of a matrix is the maximum number of independent columns or rows. For example, identity matrix

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

has three independent columns or rows. Thus, methods summarized below, either reduce matrix to an identity or infer structure of matrix based on characteristics of some other matrices.

Minor Based Approach

Definition of Minor: Let $\mathbf{A}$ be a $m \times n$ matrix. If we retain *any* r rows and r columns of $\mathbf{A}$, we will have a square sub-matrix of order r. The determinant of the square sub-matrix of order r is called a minor of $\mathbf{A}$ of order r. From a given matrix, we can form square sub-matrices of order $0, 1, 2, 3, \cdots, m$ if m is less than n or order $0, 1, 2, 3, \cdots, n$ if n is less than m. For example, if the matrix is 3×4, we can have square sub-matrices of order 1, 2, and 3. We cannot have square sub-matrices of order 4×4 for this example. Let

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}.$$

As you might have guessed it, there are 12 minors of order 1. That is, each element of $\mathbf{A}$ is a minor. If we retain any two rows and any two columns of $\mathbf{A}$ and then computed determinant of the square sub-matrices, then these determinants are called minors of order 2. That is,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \quad \begin{vmatrix} a_{21} & a_{24} \\ a_{31} & a_{32} \end{vmatrix}, \quad \text{etc.}$$

For a matrix of size 3×4, there would be 18 such sub-matrices and minors. If we retain any three rows and any three columns of $\mathbf{A}$ and computed the determinant of square sub-matrices, these determinant are called the minors of order 3. That is for above 3×4 matrix, we might get

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \end{vmatrix}.$$

We are now in position to relate rank and minor.

Theorem

The rank of a given matrix $\mathbf{A}$ is said to be r if

a. there is at least one minor of $\mathbf{A}$ of order r which is not equal to zero; or

b. every minor of $\mathbf{A}$ of order $r + 1$ is zero.

Note that if a minor of $\mathbf{A}$ is zero, the corresponding submatrix is singular and reverse is also true.

Illustration: Consider following 2×3 matrix A. That is,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

For this matrix, we can have rank of this matrix to be 0, 1 or 2. When all elements of matrix A are zero, we will have rank of $\mathbf{A}$ of zero. If any element of this matrix is non-zero, then we would rank of 1. Finally, there are three 2×2 sub-matrices. That is,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \begin{pmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{pmatrix}, \quad \text{and} \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}.$$

If rank of this matrix is 2, then we must have that

$$\begin{aligned} a_{11}a_{22} - a_{21}a_{12} &\neq 0 \quad \text{or} \\ a_{11}a_{23} - a_{21}a_{13} &\neq 0 \quad \text{or} \\ a_{12}a_{23} - a_{22}a_{13} &\neq 0 \end{aligned}$$

Note that to demonstrate that rank of $\mathbf{A}$ is 2, we need to show that at least one of three sub-matrices have minor that is non-zero. As one would imagine that such process for a larger matrices is tedious. For example, when matrix is 3×4, we need to compute 18 such minors to demonstrate that rank of such matrix is 2.

Theorems on Rank of Matrices

i. The rank of the transpose of a matrix is the same as that of the original matrix.

ii. The rank of a given matrix **A** remains unchanged by the operations of elementary row and column transformations.

We will use above two theorems for finding rank of a matrix using the normal form of a matrix. Every $m \times n$ matrix **A** of rank r can be reduced to any of the following forms

$$\begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \begin{pmatrix} I_r \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \mathbf{I}_r \end{pmatrix}$$

where matrix $\mathbf{I}_r$ denotes identity matrix of size r. These are called normal forms and are obtained by elementary row and column operations.

Rank of Matrix Based on Normal Form

If a matrix **A** is reduced to the form

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

by a chain of elementary row and column operations, then the rank of **A** is the order of identity matrix I_r.

Illustration Suppose

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix}.$$

Consider following steps to convert matrix A to normal forms:

Step 1. Multiply column 1 by 2 and subtract it from column 2. In addition, multiply column 1 by 3 and subtract it from column 3.

Step 2. Multiply row 1 by 2 and subtract it from row 2; and also multiply row 1 by 3 and subtract it from row 3.

Step 3. Multiply row 2 by -1.

Step 4. Multiply column 2 by 2 and subtract it from column 3.

Above four steps are summarized below.

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & -2 \\ 3 & -1 & -2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & -2 \\ 0 & -1 & -2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

That is, by using elementary row and column operations, we have reduced matrix **A** to $\begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & 0 \end{bmatrix}$. Hence, we conclude that rank of matrix **A** is 2.

Rank of Matrix Based on Echelon Form

To find the linearly independent rows of a matrix **A**, and hence to determine its rank, we may apply successive row operations in order to transform matrix **A** into a so-called echelon form. Suppose we obtain matrix B by applying such row and column operations. That is,

$$\underset{m \times n}{\mathbf{B}} = \begin{pmatrix} 1 & b_{12} & b_{13} & \cdots & b_{1r} & \cdots & b_{1n} \\ 0 & 1 & b_{23} & \cdots & b_{2r} & \cdots & b_{2n} \\ 0 & 0 & 1 & \ddots & b_{3r} & \ddots & b_{3n} \\ 0 & 0 & 0 & \cdots & 1 & \cdots & b_{rn} \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}.$$

That is, the echelon form consists of an $r \times r$ matrix in leading position in which elements below the main diagonal are identically zero and elements in main diagonal are unity. The remaining elements in the first r rows are in general nonzero, while all elements in the remaining $m - r$ rows are identically zero. *The rank of* **A** *is equal to the number of rows in the echelon form in which there is at least on nonzero element.* As an example, let us reduce

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 3 & 5 & 1 \\ 1 & 3 & 4 & 5 \end{bmatrix}.$$

Multiply row one by 2 and subtract from row two and subtract row one from row three. Then, **A** is transformed to

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

In the next step, add row two to row three, and multiply row two by -1 to obtain matrix **B**.

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

Consequently, we conclude that rank of **A** is 2. As you may noted that transforming matrix in echelon form requires fewer numerical operations compared to the approach based on minors, especially when matrices are large.

Eigenvalues and Eigenvectors

The characteristic value problem is defined as that of finding values of a scaler λ and an associated vector $\mathbf{x} \neq \mathbf{0}$ which satisfy

$$\mathbf{Ax} = \lambda \mathbf{x}$$

where $\mathbf{A}$ is some $n \times n$ matrix, λ is called a characterisitc root of $\mathbf{A}$ and $\mathbf{X}$ a characteristic vector. Alternative names are latent roots and vectors and eigenvalues and eigenvectors. Note that above equation may be written as

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

and has a nontrivial solution, $\mathbf{x} \neq \mathbf{0}$, if $(\mathbf{A} - \lambda \mathbf{I})$ is singular, that is, if $\mid (\mathbf{A} - \lambda \mathbf{I}) \mid = \mathbf{0}$ Eigenvectors are transformed covariance-variance or correlation matrices such that

1. Each vector is independent of others,
2. Each vector is of equal length, usually unit length, and
3. Eigenvalues are decreasing in magnitude where eigenvalues are associated scalars with each eigenvector.

- **Illustrative Example:** Suppose we ask five respondents two questions, attitude towards family and attitude towards church on ten point scale. Actual and the mean subtracted responses were

$$\text{Responses} = \begin{pmatrix} 6 & 7 \\ 5 & 9 \\ 8 & 6 \\ 4 & 9 \\ 7 & 9 \end{pmatrix} \quad \text{(Response - Mean)} = \begin{pmatrix} 0 & -1 \\ -1 & 1 \\ 2 & -2 \\ -2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Let us call the first column to be y_1 (with mean of 6) and the second to be y_2 (with mean of 8). We may construct covariance-variance matrix $\mathbf{S}$

$$\mathbf{S} = \begin{pmatrix} 2.5 & -1.5 \\ -1.5 & 2.0 \end{pmatrix}$$

We would like to obtain two sets of two numbers, say, x_{11}, x_{12} and x_{21}, x_{22} such that when we construct transformed variables, y_{t1} and y_{t2} transformed variables will be independent of each other. In addition, we would like each vector(s) of unit length. That means,

$\sqrt{x_{11}^2 + x_{12}^2} = 1$. Suppose λ is a scalar number and we will call it eigenvalue. Then finding eigenvector is equivalent to finding solution to following two linear equations.

$$\begin{aligned} 2.5x_{11} - 1.5x_{12} &= \lambda x_{11} \\ -1.5x_{11} + 2x_{12} &= \lambda x_{12} \end{aligned}$$

Note that using the first equation, we can solve for x_{11}. That is, we may write

$$x_{11} = -\frac{-1.5x_{12}}{2.5 - \lambda},$$

and substituting value of x_{11} in the second equation, we get

$$\begin{aligned} (-1.5) \times -\frac{-1.5x_{12}}{2.5 - \lambda} + 2x_{12} &= \lambda x_{12} \quad \text{or} \\ \frac{-(1.5)^2 x_{12}}{2.5 - \lambda} + (2 - \lambda)x_{12} &= 0 \quad \text{or} \\ -(1.5)^2 x_{12} + (2.5 - \lambda)(2 - \lambda)x_{12} &= 0 \quad \text{or} \\ \left[(2.5 - \lambda)(2 - \lambda) - (1.5)^2\right] x_{12} &= 0 \end{aligned}$$

Note a solution to above equation could be found by setting $x_{12} = 0$ and that would mean $x_{11} = 0$. This solution is not consistent with our second requirement of unit length. Another possibility would entail finding a solution for terms inside square parenthesis, or

$$(2.5 - \lambda)(2 - \lambda) - (1.5)^2 = 0$$

This is quadratic equation with λ being unknown. Two solutions can be obtained where

$$\lambda = 2.25 \pm \frac{1}{2}\sqrt{.25 + 4 \times (1.5)^2}.$$

Thus, we would get $\lambda = 3.791$ or $\lambda = 0.729$. Since these are roots of quadratic equations (in general, polynomial) these are often called *characteristic roots* or *eigenvalues.*

Let us substitute $\lambda = 3.791$ in equation relating to x_{11}. That is,

$$\begin{aligned} x_{11} &= \frac{1.5x_{12}}{2.5 - \lambda}, \\ &= \frac{1.5x_{12}}{2.5 - 3.771}, \\ &= -1.18x_{12} \end{aligned}$$

We also know that $x_{11}^2 + x_{12}^2 = 1$. Using these two equations, we may write

$$\begin{aligned}(-1.18)^2 x_{12}^2 + x_{12}^2 &= 1 \quad \text{or} \\ (1.392 + 1)x_{12}^2 &= 1 \\ x_{12} &= \pm\sqrt{\frac{1}{2.392}} \\ &= \pm 0.646\end{aligned}$$

This would mean that $x_{11} = -0.763$ when $x_{12} = 0.646$ and $x_{11} = 0.763$, if we take $x_{12} = -0.646$. If repeat computational steps x_{12} with $\lambda = 0.729$, we will get $x_{12} \pm 0.763$ and $x_{11} \pm 0.646$. Note that first vector, $[-0.763, 0.646]$ is called first eigenvector[2] and other one is $[0.646, 0.763]$. Note couple of observations. First, if we multiply our data matrix by these two eigenvectors, we would get two new transformed variables and these transformed vectors are independent of each other. To verify this, I multiplied $0 \times (-0.763) + -1 \times 0.646$ and I get -0.65. I continued this for each cell in the data matrix and obtained following transformed and mean corrected observations.

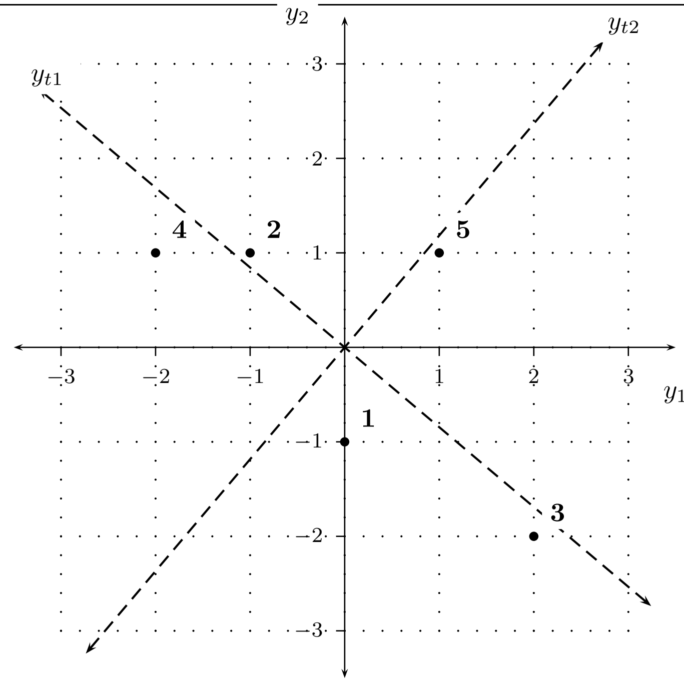
$$\left(\begin{array}{c}\text{Mean Corrected} \\ \text{Transformed Responses}\end{array}\right) = \begin{pmatrix} -0.65 & -0.76 \\ 1.41 & 0.12 \\ -2.82 & -0.24 \\ 2.17 & -0.53 \\ -0.12 & 1.41 \end{pmatrix}$$

and covariance-variance matrix is

$$\begin{pmatrix} 3.771 & 0.0 \\ 0.0 & 0.729 \end{pmatrix}.$$

Second, covariance-variance matrix of transformed matrix is nothing but eigenvalues. Graphically, determining eigenvalue is explained as a transformation of axes such that variability is maximized along the first axis.

[2] Here we have chosen one of two solutions but much of subsequent analysis does not depend on which solution is used.



- **Generalization:** Suppose we had general covariance-variance matrix such as

$$\mathbf{C} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

what are its eigenvalues? It turns out that for a simple matrix like 2×2, we can write a formula for its eigenvalues. That is,

$$\lambda = \frac{1}{2}(a_{11} + a_{22}) \pm \frac{1}{2}\sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2}.$$

Note several interesting observations.

1. When covariance is zero ($a_{12} = 0$), two eigenvalues would be a_{11} and a_{22}.
2. If we substitute $a_{11} = a_{22} = 1$, that is convert variables with variance of 1, then a_{12} would be correlation between two variables. In such instances, eigenvalues would be $\lambda = 1 \pm r_{12}$. This would imply that when two variables are perfectly correlated, one eigenvalue would be 2 and other would be zero.
3. Note that sum of variances is equal to sum of eigenvalues.

- **Properties of Eigenvalues & Eigenvectors:** There are number of interesting properties of eigenvalues. Some of these are listed below and assume that covariance-variance matrix is used for analysis.

1. The number of non-zero eigenvalues is equal rank of a matrix.

2. The product of eigenvalues is equal to determinant of a matrix.
3. The sum of eigenvalues is equal to sum of elements along the diagonal.
4. The eigenvalues of a symmetric matrix are real and orthogonal.
5. The eigenvalues of a covariance-variance matrix are positive.

Partitioned Matrices

A matrix may be divided it by means of horizontal and vertical lines and made into smaller submatrices. A partitioned matrices are often easier for computation purpose than the whole matrix. Consider a matrix $\mathbf{A}$ with dimensions 3×5. That is,

$$\mathbf{A} = \left[\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{array}\right]$$

and may partioned by means of the two lines shown to provide four submatrices,

$$\mathbf{A}_{11} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \mathbf{A}_{12} = \begin{bmatrix} a_{14} & a_{15} \\ a_{24} & a_{25} \end{bmatrix}$$

$$\mathbf{A}_{21} = \begin{bmatrix} a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \mathbf{A}_{22} = \begin{bmatrix} a_{34} & a_{35} \end{bmatrix}$$

and it could be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Note that the partitioning must extend all the way across or up and down the original matrix. The operations of addition and multiplication also apply to partitioned matrices, but matrices must be have been partitioned conformably. For example, if there is another matrix $\mathbf{B}$ also of the order 3×5 and partitioned as matrix $\mathbf{A}$. That is,

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ 2 \times 3 & 2 \times 2 \\ \mathbf{B}_{21} & \mathbf{B}_{22} \\ 1 \times 3 & 2 \times 1 \end{bmatrix}$$

then the sum $\mathbf{A} + \mathbf{B}$ can be written as

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{bmatrix}.$$

For multiplication also partitioned matrices must be conformable. Suppose matrix $\mathbf{C}$ is 5×2. That is;

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \end{bmatrix}.$$

The product $\mathbf{AC}$ then is of order 3×2. For the product may be expressed in terms of partitioned matrices, the only condition is that the partitioning of the *rows* of $\mathbf{C}$ should conform to the partitioning of the *columns* of $\mathbf{A}$. Suppose matrix $\mathbf{C}$ is partitioned as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} \\ 3 \times 2 \\ \mathbf{C}_{21} \\ 2 \times 2 \end{bmatrix}.$$

The multiplication of submatrices then is same as treating submatrices as elements. That is product

$$\begin{aligned} \mathbf{AC} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} \\ \mathbf{C}_{21} \end{bmatrix} \\ &= \begin{bmatrix} \underset{2 \times 2}{\mathbf{A}_{11}\mathbf{C}_{11}} & + & \underset{2 \times 2}{\mathbf{A}_{12}\mathbf{C}_{21}} \\ \underset{1 \times 2}{\mathbf{A}_{21}\mathbf{C}_{11}} & + & \underset{1 \times 2}{\mathbf{A}_{22}\mathbf{C}_{21}} \end{bmatrix}. \end{aligned}$$

Note that condition of conformability also would be satisfied by

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ 3 \times 1 & 3 \times 1 \\ \mathbf{C}_{21} & \mathbf{C}_{22} \\ 2 \times 1 & 2 \times 1 \end{bmatrix}.$$

and then product would appear

$$\begin{aligned} \mathbf{AC} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \underset{2 \times 1}{\mathbf{A}_{11}\mathbf{C}_{11}} & + & \underset{2 \times 1}{\mathbf{A}_{12}\mathbf{C}_{21}} & \underset{1 \times 1}{\mathbf{A}_{11}\mathbf{C}_{12}} & + & \underset{1 \times 1}{\mathbf{A}_{12}\mathbf{C}_{22}} \\ \underset{1 \times 1}{\mathbf{A}_{21}\mathbf{C}_{11}} & + & \underset{1 \times 1}{\mathbf{A}_{22}\mathbf{C}_{21}} & \underset{1 \times 1}{\mathbf{A}_{21}\mathbf{C}_{12}} & + & \underset{1 \times 1}{\mathbf{A}_{22}\mathbf{C}_{22}} \end{bmatrix}. \end{aligned}$$

The inverse of partitioned matrix sometimes provide useful starting point for analysis purpose. If $\mathbf{A}$ is a nonsingular matrix, which is partitioned

$$\mathbf{A} = \begin{bmatrix} \underset{p\times p}{\mathbf{A}_{11}} & \underset{p\times q}{\mathbf{A}_{12}} \\ \underset{q\times p}{\mathbf{A}_{21}} & \underset{q\times q}{\mathbf{A}_{22}} \end{bmatrix}$$

Using following identities

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

after some algebraic manipulation, it can be shown that

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} \underset{p\times p}{\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{E}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1})} & \underset{p\times q}{-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{E}} \\ \underset{q\times p}{-\mathbf{E}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1})} & \underset{q\times q}{\mathbf{E}} \end{bmatrix} \quad \text{or alternatively} \\ &= \begin{bmatrix} \underset{q\times q}{\mathbf{G}} & \underset{q\times p}{-\mathbf{G}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1})} \\ \underset{p\times q}{-(\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\mathbf{G}} & \underset{p\times p}{\mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{G}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1})} \end{bmatrix}. \end{aligned}$$

where $\mathbf{E} = \left[\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right]^{-1}$ and $\mathbf{G} = \left[\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right]^{-1}$ Both of these forms are useful, if $\mathbf{A}_{22}$ or $\mathbf{A}_{11}$ happens to be of simpler form. This inverse exists, if $\mathbf{A}_{11}$ and / or $\mathbf{A}_{22}$ are nonsigular. If matrix $\mathbf{A}$ is *block-diagonal* then, inverse of is of simpler form. That is,

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \quad \text{then} \\ \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix}. \end{aligned}$$

Above result holds for any number of matrices.
Determinant of partitioned matrix can be written in two different forms

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \quad \text{then} \\ \mid \mathbf{A} \mid &= \mid \mathbf{A}_{22} \mid \cdot \mid \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \mid \\ &= \mid \mathbf{A}_{11} \mid \cdot \mid \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \mid \end{aligned}$$

Note that the first result to be valid, we must have $\mathbf{A}_{22}$ to be nonsingular while the second result to be valid, we must have $\mathbf{A}_{11}$ to nonsingular. Furthermore, if any one of off-diagonal matrix is zero, that is, either $\mathbf{A}_{12}$ or $\mathbf{A}_{21}$, then determinant is equal to product of determinant of diagonal matices, or $\mid \mathbf{A}_{11} \mid \cdot \mid \mathbf{A}_{22} \mid$.

Vector and Matrix Derivatives

It is possible to write any vector or matrix in extended form and apply, element by element, rules of scaler differentiation, there is advantage in defining rules that apply to vectors and matrices. First briefly scaler rules are reviewed and then similar matrix based rules are provided.

Scaler Differentiation Rules
In following examples, u and v are real-valued differentiable functions of x and y is function u and v. Furthermore a is a real constant, then following rules apply.

- If $y = a$, then $\frac{\partial y}{\partial x} = 0$.
- If $y = u(x) + v(x)$, then
$$\frac{\partial y}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}.$$
- If $y = u(x)v(x)$, then
$$\frac{\partial y}{\partial x} = v(x)\frac{\partial u}{\partial x} + u(x)\frac{\partial v}{\partial x}.$$
- If $y = \frac{u(x)}{v(x}$, then
$$\frac{\partial y}{\partial x} = \frac{v(x)\frac{\partial u}{\partial x} - u(x)\frac{\partial v}{\partial x}}{v(x)^2}$$
provided $v(x) \neq 0$.
- If $y = u(x)^a$, then
$$\frac{\partial y}{\partial x} = au(x)^{a-1}\frac{\partial u}{\partial x}.$$
- If $y = \log[u(x)]$, then
$$\frac{\partial y}{\partial x} = \frac{1}{u(x)}\frac{\partial u}{\partial x}.$$

- If $y = \exp[u(x)]$, then
$$\frac{\partial y}{\partial x} = \exp[u(x)]\frac{\partial u}{\partial x}.$$
- If $y = a^{u(x)}$, then
$$\frac{\partial y}{\partial x} = a^{u(x)} \log[u(x)]\frac{\partial u}{\partial x},$$
provided $a > 0$.

A Matric Function with respect to a Scaler
Suppose the elements of $m \times n$ of matrix $\mathbf{U}$ be differentiable functions of a scaler variable x. Then the derivative of $\mathbf{U}$ with respect to x is defined as the $m \times n$ matrix

$$\frac{\partial \mathbf{U}}{\partial x} = \left[\frac{\partial u_{ij}}{\partial x}\right]$$

where $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$. This is easy rule to follow because element-by-element differentiation. For example, suppose that

$$\mathbf{U} = \begin{bmatrix} 2x & x^2 & 3\exp(x) \\ \log(x) & 3 & \sqrt{2x} \end{bmatrix}$$

then

$$\frac{\partial \mathbf{U}}{\partial x} = \begin{bmatrix} 2 & 2x & 3\exp(x) \\ \dfrac{1}{x} & 0 & \dfrac{2}{\sqrt{2x}} \end{bmatrix}.$$

Suppose that there is another matrix $\mathbf{V}$ of size $p \times q$ and also differentiable functions of a scaler variable x. Then following rules apply.

$$\frac{\partial(\mathbf{U} + \mathbf{V})}{\partial x} = \frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}$$

provided $m = p$ and $n = q$.

$$\frac{\partial(\mathbf{UV})}{\partial x} = \mathbf{U}\frac{\partial \mathbf{V}}{\partial x} + \mathbf{V}\frac{\partial \mathbf{U}}{\partial x}$$

provided $n = p$ and resulting matrix will be of size $m \times q$. Finally, differentiation of $\mathbf{U}^{-1}$ only exist if $m = n$ and $\mid \mathbf{U} \mid \neq 0$.

$$\frac{\partial \mathbf{U}^{-1}}{\partial x} = -\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}$$

which is complicated to remember and manipulate than scaler quotient rule. Note that

$$\frac{\partial \log \mid \mathbf{U} \mid}{\partial x} = \text{tr}(\mathbf{U}^{-1})\frac{\partial \mathbf{U}}{\partial x}$$

provided $m = n$. Moreover,

$$\frac{\partial \text{tr}(\mathbf{UA})}{\partial x} = \text{tr}(\frac{\partial \mathbf{U}}{\partial x})\mathbf{A}$$

provided $m = n$, matrix $\mathbf{A}$ has real valued constants and of size $m \times m$.

Differentiation of a Scaler Function of a Matrix with respect to Matric or Vector
Let y be scaler function of matrix $\mathbf{X}$, then the derivative of y with respect to $\mathbf{X}$ is the matrix $\frac{\partial y}{\partial \mathbf{X}} = \left[\frac{\partial y}{\partial x_{ij}}\right]$. The vector form of this differentiation appear in regression analysis while the matrix form appear with multivariate regression. Since determinants and traces are always scaler, these differentiations are summarized below. It is assumed that matrix $\mathbf{X}$ is of order $m \times n$. Note that differentiation of determinant,

$$\frac{\partial \mid \mathbf{X} \mid}{\partial \mathbf{X}} = \text{adj}(\mathbf{X}')$$

provided $m = n$ and

$$\frac{\partial \log \mid \mathbf{X} \mid}{\partial \mathbf{X}} = \frac{\text{adj}(\mathbf{X}')}{\mid \mathbf{X} \mid} = (\mathbf{X}^{-1})'$$

provided that $m = n$ and $\mid \mathbf{X} \mid \neq 0$. To illustrate, usefulness of these rules consider following example. Let

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Note that $\mid \mathbf{X} \mid = x_{11}x_{22} - x_{12}x_{21}$. Then it follows that

$$\begin{aligned}
\frac{\partial \mid \mathbf{X} \mid}{\partial x_{11}} &= x_{22} \\
\frac{\partial \mid \mathbf{X} \mid}{\partial x_{12}} &= -x_{21} \\
\frac{\partial \mid \mathbf{X} \mid}{\partial x_{21}} &= -x_{12} \\
\frac{\partial \mid \mathbf{X} \mid}{\partial x_{22}} &= x_{11} \quad \text{and, thus} \\
\frac{\partial \mid \mathbf{X} \mid}{\partial \mathbf{X}} &= \begin{bmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{bmatrix}.
\end{aligned}$$

To obtain same result using rules of differentiation, a transpose and adjoint must be calculated. That is,

$$\begin{aligned}\mathbf{X}' &= \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \\ \text{Adj}(\mathbf{X}') &= \begin{bmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{bmatrix}\end{aligned}$$

which is easier to obtain than actually calculating derivatives. Verify that for above $\mathbf{X}$

$$\frac{\partial \log \mid \mathbf{X} \mid}{\partial \mathbf{X}} = \frac{1}{x_{11}x_{22} - x_{12}x_{21}} \begin{bmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{bmatrix}$$

using both approaches.
Note that differentiation of determinant,

$$\frac{\partial \mid \mathbf{AXB} \mid}{\partial \mathbf{X}} = \mid \mathbf{AXB} \mid (\mathbf{X}')^{-1} = \mid \mathbf{AXB} \mid (\mathbf{X}^{-1})'$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices of constants and of the same size as matrix $\mathbf{X}$. Suppose

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}, \\ \mathbf{X} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad \text{and} \\ \mathbf{B} &= \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.\end{aligned}$$

Verify that $\mid \mathbf{AXB} \mid$ is equal to $2(x_{11}x_{22} - x_{21}x_{12})$ and

$$\frac{\partial \mid \mathbf{AXB} \mid}{\partial \mathbf{X}} = 2 \begin{bmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{bmatrix}.$$

Suppose $\mathbf{A}$ be an $p \times q$ matrix of constants and $\mathbf{X}$ a $m \times n$ matrix of variables as defined above. Then following derivatives of trace function are useful.

$$
\begin{aligned}
\frac{\partial \text{tr}(\mathbf{A})}{\partial \mathbf{X}} &= \mathbf{0} && p = q \\
\frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} &= \mathbf{I} && m = n \\
\frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{XA})}{\partial \mathbf{X}} &= \mathbf{A}' && p = m, q = n \\
\frac{\partial \text{tr}(\mathbf{AX}')}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}'\mathbf{A})}{\partial \mathbf{X}} &= \mathbf{A} && p = m, q = n \\
\frac{\partial \text{tr}(\mathbf{X}'\mathbf{AX})}{\partial \mathbf{X}} &= (\mathbf{A} + \mathbf{A}')\mathbf{X} && p = m, q = n. \\
\frac{\partial \text{tr}(\mathbf{X}^{-1}\mathbf{A})}{\partial \mathbf{X}} &= -\mathbf{X}^{-1}\mathbf{A}'\mathbf{X}^{-1} && p = m, q = n.
\end{aligned}
$$

Suppose

$$
\begin{aligned}
\mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \\
\mathbf{X} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.
\end{aligned}
$$

This results in

$$
\begin{aligned}
\mathbf{AX} &= \begin{bmatrix} a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22} \\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22} \end{bmatrix} \quad \text{and} \\
\text{tr}(\mathbf{AX}) &= a_{11}x_{11} + a_{12}x_{21} + a_{21}x_{12} + a_{22}x_{22}.
\end{aligned}
$$

Thus, it can be concluded that

$$
\frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \mathbf{A}'.
$$

In conclusion, it might be noted that use of matrix algebra for differentiation purposes is more compact than element-by-element differentiation.