Outline of Elementary Matrix Algebra

I. Matrices

A. A **matrix** is a rectangular array of elements arranged in m rows and n columns, of order $m \times n$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

A $1 \times n$ matrix is a **row vector** of order n

$$\boldsymbol{b}^{\mathsf{T}} = [b_1 \; b_2 \; \cdots \; b_n].$$

An $m \times 1$ matrix is a **column vector** of order m

$$\boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}.$$

B. Rules and definitions for matrix operations

1. **Equality of two matrices**

$\boldsymbol{A} = \boldsymbol{B}$ if $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the same order $m \times n$ and

$$a_{ij} = b_{ij} \text{ for all } i = 1, ..., m \text{ and } j = 1, ..., n.$$

2. **Matrix addition**

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the same order $m \times n$, then $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{C}$ such that $\boldsymbol{C}$ is also of the same order and

$$a_{ij} + b_{ij} = c_{ij} \text{ for all } i = 1, ..., m \text{ and } j = 1, ...,n.$$

3. **Scalar multiplication**

If α is a scalar, then $\alpha\boldsymbol{A} = [\alpha a_{ij}]$, i.e. each element of **A** is multiplied by α.

4. **Matrix multiplication**

If $\boldsymbol{A}$ is of order $m\times n$ and $\boldsymbol{B}$ is of order $n\times p$, i.e. the number of rows in $\boldsymbol{B}$ equals the number of columns in $\boldsymbol{A}$, then $\boldsymbol{A}$ and $\boldsymbol{B}$ are **conformable** for matrix multiplication. Given two conformable matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, the product $\boldsymbol{AB} = \boldsymbol{C}$ is defined to be the matrix of order $m\times p$ such that

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}\ ,\ i = 1,...,m \text{ and } j = 1,...,p.$$

That is, the ij^{th} element of $\boldsymbol{C}$ is formed by multiplying each element of the i^{th} **row** of $\boldsymbol{A}$ times the corresponding element of the j^{th} **column** of $\boldsymbol{B}$ and adding across $k = 1,\ldots,n$ (n equals the number of **columns** of $\boldsymbol{A}$ and the number of **rows** of $\boldsymbol{B}$). Note that $\boldsymbol{BA}$ is not defined unless $m = p$.

5. **Commutative laws** (Prove these as exercises.)

a) **Addition:** $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$

whenever either one is well-defined.

b) **Multiplication:** $\boldsymbol{AB} \neq \boldsymbol{BA}$

except in the special case of square, symmetric matrices.

6. **Associative laws**

a) **Addition:** $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$

b) **Multiplication:** $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.

7. **Distributive laws** (Prove these as exercises.)

a) **Matrix multiplication:** $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$.

b) **Scalar multiplication:** $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$

$(\alpha + \beta)\boldsymbol{A} = \alpha\boldsymbol{A} + \beta\boldsymbol{A}$.

C. **Unit, or identity matrix**

1. The **identity matrix**, $\boldsymbol{I}_n$, is a square matrix of order $n\times n$ with ones on the principal diagonal and zeros elsewhere:

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

2. A **scalar matrix**, $\sigma\boldsymbol{I}_n$, is a square $n\times n$ matrix with a common scalar element on the principal diagonal and zeros elsewhere.

Note: $\sigma\boldsymbol{A} = (\sigma\boldsymbol{I}_n)\boldsymbol{A} = \boldsymbol{A}(\sigma\boldsymbol{I}_n) = \boldsymbol{A}\sigma$.

3. A **diagonal matrix**, $\boldsymbol{\Delta}$, is a square $n\times n$ matrix with scalar elements, not necessarily equal, on the principal diagonal and zeros elsewhere.

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_n \end{bmatrix}$$

D. Transposition of a matrix

1. The **transpose** of $\boldsymbol{A}$, denoted $\boldsymbol{A}^{\mathrm{T}}$, is the matrix obtained from $\boldsymbol{A}$ by interchanging rows and columns.

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \Leftrightarrow \boldsymbol{A}^{\top}=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

2. A **symmetric** matrix, $\boldsymbol{A}=\boldsymbol{A}^{\top}$, is a **square** $n\times n$ matrix such that $a_{ij} = a_{ji}$ for all i, j = 1, 2,..., n. If $\boldsymbol{A}$ is of order $m\times n$, $m \neq n$, then both $\boldsymbol{A}\boldsymbol{A}^{\top}$ ($m\times m$) and $\boldsymbol{A}^{\top}\boldsymbol{A}$ ($n\times n$) are symmetric, though of different orders.

$$\boldsymbol{A}\boldsymbol{A}^{\top}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}=\left[\sum_{k=1}^{n} a_{ik}a_{jk}\right],$$

where i, j = 1,...,m on the far right-hand-side.

$$\boldsymbol{A}^{\top}\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}=\left[\sum_{k=1}^{m} a_{ki}a_{kj}\right],$$

where i, j = 1,...,n on the far right-hand-side.

3. **Theorems about transposed matrices** (Prove these as exercises.)

a. $(\boldsymbol{A}^{\top})^{\top}=\boldsymbol{A}$

b. $(\boldsymbol{A}+\boldsymbol{B})^{\top}=\boldsymbol{A}^{\top}+\boldsymbol{B}^{\top}$

c. $(\boldsymbol{A}\boldsymbol{B})^{\top}=\boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$, $(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})^{\top}=\boldsymbol{C}^{\top}\boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$, and so on.

E. Trace

For a square $n\times n$ matrix, $\boldsymbol{A}$, the **trace** of $\boldsymbol{A}$ is the sum of the elements on the principal main diagonal:

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Note: For all $\boldsymbol{A}$, $\text{tr}(\boldsymbol{A}^\top \boldsymbol{A}) = \text{tr}(\boldsymbol{A}\boldsymbol{A}^\top) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$. (Again, you should prove this.)

F. <u>Simple matrix operations</u>

1. A system of m **<u>simultaneous equations</u>** in n variables,

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

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

$$\vdots$$

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

may be written as $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A}$ is an $m \times n$ matrix of coefficients, $\boldsymbol{x}$ is a column vector of n elements, and $\boldsymbol{b}$ is a column vector of m elements.

2. **<u>Sum of squares</u>**

$$e_1^2 + e_2^2 + \cdots + e_n^2 = \boldsymbol{e}^\top \boldsymbol{e} ,$$

where $\boldsymbol{e}$ is a column vector of n elements.

3. **<u>Weighted sum of squares</u>**

$$a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 = \boldsymbol{x}^\top \boldsymbol{Ax} ,$$

where $\boldsymbol{A}$ is a diagonal $n \times n$ matrix and $\boldsymbol{x}$ is an $n \times 1$ column vector.

4. **<u>Quadratic forms</u>**

$$a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + a_{21}x_1x_2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 = \boldsymbol{x}^\top \boldsymbol{Ax}$$

where $\boldsymbol{A}$ is a symmetric $n \times n$ matrix and $\boldsymbol{x}$ is a column vector of n elements. We may assume that $\boldsymbol{A}$ is symmetric without loss of generality (WLOG) because

$$\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\top}\boldsymbol{A}^{\top}\boldsymbol{x} = ½\boldsymbol{x}^{\top}(\boldsymbol{A}+\boldsymbol{A}^{\top})\boldsymbol{x}$$

for all $n \times n$ matrices $\boldsymbol{A}$ and all n–vectors $\boldsymbol{x}$, due to the fact that transposing a scalar such as $\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}$ does not change the scalar.

II. Determinants

Associated with any square matrix **A** there is a scalar quantity called the **determinant** of $\boldsymbol{A}$, denoted det($\boldsymbol{A}$) or |$\boldsymbol{A}$|. The definition of the determinant of an n^{th} order matrix is

$$|\,\boldsymbol{A}\,| = \sum \pm a_{1i}a_{2j}\cdots a_{nr}$$

where

1. Each term on the right-hand side is the product of the same number of elements as the order of $\boldsymbol{A}$.
2. In each term there is one and only one element from each row and one and only one element from each column of $\boldsymbol{A}$.
3. In total there are $n! = n\cdot(n-1)\cdots 3\cdot 2\cdot 1$ different terms on the right-hand side.
4. Within each term the first subscripts are in natural order, while the second terms are permutations of the first n integers.
5. For $n > 1$, $n!$ is an even number. Half of the terms have positive signs and half have negative signs for $n > 1$. The sign of each term depends upon the

second subscripts. An **<u>inversion of natural order</u>** occurs when of two integers the larger precedes the smaller. The number of inversions in a permutation of n integers is the number of pairs of elements, ***not necessarily adjacent***, in which a larger integer precedes a smaller one. A permutation is even (carries a positive sign) when the number of inversions is even and odd (carries a negative sign) when the number of inversions is odd.

A. <u>Properties of determinants</u> (good exercises)

1. $|\boldsymbol{A}^T| = |\boldsymbol{A}|$, the determinant of the transposed matrix is equal to the determinant of the original matrix.

2. Interchanging any two columns (or rows) of $\boldsymbol{A}$ changes the sign of the determinant of $\boldsymbol{A}$.

   a. The interchange of any two elements in a permutation must always change the class of the permutation from odd to even, or vice versa;

   b. Interchanging any two columns of $\boldsymbol{A}$ thus means that the first subscripts in all terms in $|\boldsymbol{A}|$ are unchanged but the sign of each term changes, so that the sign of $|\boldsymbol{A}|$ changes.

3. The determinant of a matrix with two equal rows (or columns) is zero.

4. If every element of a row (or column) of $\boldsymbol{A}$ is multiplied by a scalar σ to give a new matrix $\boldsymbol{B}$, then $|\boldsymbol{B}| = \sigma|\boldsymbol{A}|$.

5. If every element of an n^{th} order matrix $\mathbf{A}$ is multiplied by σ,

$$|\sigma \boldsymbol{A}| = \sigma^{\boldsymbol{n}} \, |\boldsymbol{A}|.$$

6. If **A** and **B** are both ($n\times n$) matrices, then |**AB**| = |**A**||**B**|.

B. <u>Minors</u>

The **<u>minor determinant</u>** of element a_{ij} in the matrix $\boldsymbol{A}$ is the determinant of order n–1 of the matrix obtained from $\boldsymbol{A}$ by deleting the i^{th} row and the j^{th} column, and is denoted $|\boldsymbol{A}_{ij}|$.

Note: The sum of all terms in $|\boldsymbol{A}|$ involving a_{ij} as a factor is $a_{ij}(-1)^{i+j}\,|\boldsymbol{A}_{ij}|$. (Prove this as an exercise.)

C. <u>Co-factors</u>

The **<u>co-factor</u>** for the element a_{ij} in **A** is defined as

$$c_{ij} = (-1)^{i+j}\,|\boldsymbol{A}_{ij}|.$$

$|\boldsymbol{A}|$ can be expressed in terms of the elements of the i^{th} row (or j^{th} column) and their co-factors as (prove this as an exercise)

$$|\boldsymbol{A}| = \sum_{i=1}^{n} a_{ij}c_{ij} = \sum_{j=1}^{n} a_{ij}c_{ij}.$$

Consequently, a determinant is unaltered in value when a constant multiple of any row (column) is added to any other row (column). The same result holds when a linear combination of rows (columns) is added to any row (column) that is not part of the combination. (Prove these as exercises.)

D. <u>The inverse of a matrix</u>

If, for a **<u>square</u>** matrix $\boldsymbol{A}$ of order n, there exists a matrix $\boldsymbol{A}^{-1}$ such that

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$$

then $\boldsymbol{A}^{-1}$ is defined to be the **inverse** of ***A***.

Construction of $\boldsymbol{A}^{-1}$:

1. From ***A***, construct a co-factor matrix in which each element of ***A***, a_{ij}, is replaced by its co-factor, c_{ij}. The transpose of this matrix is called the **adjoint** matrix, denoted by $\boldsymbol{A}^* = \text{adj}(\boldsymbol{A}) = \text{cof}(\boldsymbol{A})^{\mathrm{T}} = [c_{ji}]$.

2. $\boldsymbol{A}^{-1}$ is then defined as $\boldsymbol{A}^{-1} = \text{adj}(\boldsymbol{A}) / |\boldsymbol{A}|$. . (Prove that this definition satisfies $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$ as an exercise.)

3. Hence, $\boldsymbol{A}^{-1}$ exists if and only if $|\boldsymbol{A}|$ does not vanish, i.e., ***A*** is **nonsingular** (If $|\boldsymbol{A}| = 0$, then ***A*** is **singular** since at least one row or column is an exact linear combination of the other rows or columns, respectively).

E. Cramer's Rule

If $\boldsymbol{Ax} = \boldsymbol{b}$, where ***A*** is $n \times n$ and nonsingular, then premultiplication of both sides by $\boldsymbol{A}^{-1} = \text{adj}(\boldsymbol{A}) / |\boldsymbol{A}|$ gives

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} = \frac{\text{adj}(\boldsymbol{A})}{|\boldsymbol{A}|}\boldsymbol{b} = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Evaluating the right-hand-side for x_i gives

$$x_i = \left(\frac{1}{|\boldsymbol{A}|}\right)\sum_{j=1}^{n}(-1)^{i+j}\,|\boldsymbol{A}_{ji}|\,b_j = \left(\frac{1}{|\boldsymbol{A}|}\right)\begin{vmatrix} a_{11} & a_{21} & \cdots & a_{n-1,1} & a_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1,i-1} & a_{2,i-1} & \cdots & a_{n-1,i-1} & a_{n,i-1} \\ b_1 & b_2 & \cdots & b_{n-1} & b_n \\ a_{1,i+1} & a_{2,i+1} & \cdots & a_{n-1,i+1} & a_{n,i+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{n-1,n} & a_{nn} \end{vmatrix}$$

$$= \left(\frac{1}{|\boldsymbol{A}|}\right)\begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & b_1 & a_{1,i+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,i-1} & b_2 & a_{2,i+1} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,i-1} & b_{n-1} & a_{n-1,i+1} & \cdots & a_{n-1,n} \\ a_{n1} & \cdots & a_{n,i-1} & b_n & a_{n,i+1} & \cdots & a_{nn} \end{vmatrix},$$

which is the ratio of two determinants. The numerator is the determinant of the matrix formed by replacing the i^{th} column of $\boldsymbol{A}$ with $\boldsymbol{b}$. The denominator is simply the determinant of $\boldsymbol{A}$.

Note: This result makes use of the two facts from above. First, $|\boldsymbol{A}| = |\boldsymbol{A}^{\top}|$ for any square matrix $\boldsymbol{A}$. Second, a determinant can be defined in terms of the expansion by the cofactors of any row or column.

III. Partitioned Matrices

Because a matrix is a rectangular array of elements (numbers), we may divide it up into parts by means of horizontal and vertical lines to get smaller arrays or submatrices such as

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

Basic operations such as addition and multiplication apply to partitioned matrices **<u>if</u>** the matrices have been partitioned conformably.

A. <u>Block-diagonal matrices</u>

Let $\boldsymbol{A}$ is nonsingular and of the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & 0 \\ 0 & \boldsymbol{A}_{22} \end{bmatrix}$$

(Prove the following 3 results as exercises.)

Then its inverse is given by

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & 0 \\ 0 & \boldsymbol{A}_{22}^{-1} \end{bmatrix}.$$

We also have

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{11}||\boldsymbol{A}_{22}|$$

as well as

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{11}||\boldsymbol{A}_{22}|.$$

B. <u>General partitioned matrices</u>

If $\boldsymbol{A}$ is partitioned as

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

and $\mathbf{A}_{11}$ is nonsingular, then

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{22}||\boldsymbol{A}_{11} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|.$$

Theorem: *If the partitioned matrix* $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}$ *is invertible (that is, nonsingular) and* **A** *is symmetric and nonsingular, then*

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}^\top(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\top)^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}^\top(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\top)^{-1} \\ -(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\top)^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & (\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\top)^{-1} \end{bmatrix}.$$

This result plays an important role in the comparative static analysis of the theory of consumer behavior (and in the theory of constrained cost minimization, with u a fixed output level), where $\boldsymbol{A}=\partial^2 u(\boldsymbol{x})/\partial\boldsymbol{x}\partial\boldsymbol{x}^\top$, $\boldsymbol{B} = -\boldsymbol{p}^\top$, and $\boldsymbol{C} = 0$, a scalar, so it is worthwhile to prove it here. Let

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{E}^\top \\ \boldsymbol{E} & \boldsymbol{F} \end{bmatrix}.$$

It follows from the definition of inverse that

$$\boldsymbol{A}\boldsymbol{D} + \boldsymbol{B}^\top\boldsymbol{E} = \boldsymbol{I}$$

$$\boldsymbol{A}\boldsymbol{E}^\top + \boldsymbol{B}^\top\boldsymbol{F} = \boldsymbol{0}$$

$$\boldsymbol{B}\boldsymbol{D} + \boldsymbol{C}\boldsymbol{E} = \boldsymbol{0}$$

$$\boldsymbol{B}\boldsymbol{E}^\top + \boldsymbol{C}\boldsymbol{F} = \boldsymbol{I}$$

First, use the invertibility of $\boldsymbol{A}$ to solve the first equation for $\boldsymbol{D}$, which gives

$$\boldsymbol{D} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}^\top\boldsymbol{E}\,.$$

Second, substitute this expression into the third equation to obtain

$$\boldsymbol{B}(\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}^\top\boldsymbol{E}) + \boldsymbol{C}\boldsymbol{E} = \boldsymbol{0}\,.$$

Third, group terms in $\boldsymbol{E}$ and solve for $\boldsymbol{E}$, which gives

$$\boldsymbol{E} = -(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.$$

Fourth, substitute the solution for $\boldsymbol{E}$ into the expression for $\boldsymbol{D}$ above to get

$$\boldsymbol{D} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.$$

Finally, substitute the solution for $\boldsymbol{E}$ into the second equation

$$\boldsymbol{A}[-\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}] + \boldsymbol{B}^{\top}\boldsymbol{F} = \boldsymbol{0}$$

$$\Rightarrow -\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{B}^{\top}\boldsymbol{F} = \boldsymbol{0},$$

and premultiply this by $-\boldsymbol{B}\boldsymbol{A}^{-1}$,

$$\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}\boldsymbol{F} = \boldsymbol{0}.$$

Now substitite the solution for $\boldsymbol{E}$ into the fourth equation

$$\boldsymbol{B}[-\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}] + \boldsymbol{C}\boldsymbol{F} = \boldsymbol{I}$$

$$\Rightarrow -\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1} + \boldsymbol{C}\boldsymbol{F} = \boldsymbol{I}.$$

Adding the two results together and grouping terms in $\boldsymbol{F}$ gives

$$\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}\boldsymbol{F} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1} + \boldsymbol{C}\boldsymbol{F} = \boldsymbol{I}$$

$$\Rightarrow (\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{F} = \boldsymbol{I}$$

$$\Rightarrow \boldsymbol{F} = (\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}.$$

Restating these results in block matrix form gives,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^{\top} \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}^{\top}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1} \\ -(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{\top})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & (\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}.$$

IV. Linear Dependence, Rank, and the Solution of Homogeneous Equations

Consider $\boldsymbol{Ax} = \boldsymbol{0}$, where $\boldsymbol{A}$ is an $m \times n$ matrix of known constants and $\boldsymbol{x}$ is an $n \times 1$ column vector of unknown variables. If the ***only*** solution is the trivial one, $\boldsymbol{x} = \boldsymbol{0}$, (every element of $\boldsymbol{x}$ is zero), then the column vectors $[\boldsymbol{a}_{\bullet 1}, \boldsymbol{a}_{\bullet 2}, \cdots, \boldsymbol{a}_{\bullet n}]$ (the first subscript "•" indicates that we are considering **columns** of $\boldsymbol{A}$) are **linearly independent**. If there **exists** some (i.e., ***any***) $\boldsymbol{x} \neq \mathbf{0}$ (that is, not all of the elements of $\boldsymbol{x}$ are zero) that is a solution, then the columns of **A** are **linearly dependent**.

A. Rank

The **rank** of $\boldsymbol{A}$, denoted rank($\boldsymbol{A}$) or $\rho(\boldsymbol{A})$, is defined as the **maximum number of linearly independent columns** of $\boldsymbol{A}$.

Note: $\rho(\boldsymbol{A}) \leq \min\{m,n\}$.

The rank of a matrix can be defined equivalently as the **maximum number of linearly independent rows**, or the **maximum order of non-zero minors**.

B. Full Rank

If $m = n$ and $\boldsymbol{A}$ is nonsingular, then $|\boldsymbol{A}| \neq 0$ and $\boldsymbol{A}^{-1}$ exists, so that the only solution to $\boldsymbol{Ax} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$ since $\boldsymbol{Ax} = \mathbf{0}$ if and only if $\boldsymbol{x} = \boldsymbol{A}^{-1}\mathbf{0} = \mathbf{0}$. **A** then has **full rank** and $\rho(\boldsymbol{A}) = n$.

C. Diagonal Matrices

The **rank** of a **diagonal matrix** is equal to the **number of nonzero elements on the main diagonal**.

Note: $\rho(\boldsymbol{I}_n) = n$.

D. Inverse Matrices

If $\boldsymbol{A}$ is nonsingular, then $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}| \neq 0$, hence, $\boldsymbol{A}^{-1}$ also has **full rank** and is **nonsingular**.

E. <u>Product Matrices</u>

1. In general, $\rho(\boldsymbol{AB}) \le \min\{\rho(\boldsymbol{A}), \rho(\boldsymbol{B})\}$, i.e., the rank of the product $\boldsymbol{AB}$ can not exceed the smaller of the ranks of $\boldsymbol{A}$ and $\boldsymbol{B}$.

2. Pre- or post-multiplication of a matrix $\boldsymbol{A}$ by a nonsingular matrix gives a product matrix whose rank is equal to the rank of $\boldsymbol{A}$.

F. <u>Dimension</u>

The **<u>dimension</u>** of a space is defined as the **<u>maximum number of linearly independent vectors</u>** in the space. If $\boldsymbol{A}$ is $m \times n$ with $\rho(\boldsymbol{A}) = r < n$, then the **<u>dimension</u>** of the **<u>solution space</u>** for $\boldsymbol{Ax} = \boldsymbol{0}$ is $n-r$. This is the **<u>null space</u>** of the linear transformation $\boldsymbol{y} = \boldsymbol{Ax}$. This function maps points in n–dimensional space ($\boldsymbol{x} \in \mathbb{R}^n$) into points in m-dimensional space ($\boldsymbol{y} \in \mathbb{R}^m$). If the dimension of the solution space of $\boldsymbol{Ax} = \boldsymbol{0}$ is $n-r$, then the dimension of the space **<u>spanned</u>** by the columns of $\boldsymbol{A}$ is r.

V. <u>Characteristic Roots and Vectors</u>

The **<u>characteristic value</u>** problem is defined as that of finding values of a scalar λ and an associated vector $\boldsymbol{x} \ne \boldsymbol{0}$ of unit length (that is, $\sqrt{\sum_{i=1}^{n} x_i^2} = 1$) which satisfy the equations $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, where $\boldsymbol{A}$ is an $n \times n$ matrix, λ is a **<u>characteristic root</u>** (also called an **<u>eigenvalue</u>** or a **<u>latent root</u>**) of $\boldsymbol{A}$, and $\boldsymbol{x}$ is a **<u>characteristic vector</u>** (**<u>eigenvector</u>** or **<u>latent vector</u>**) of $\boldsymbol{A}$.

The system of n equations $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ is equivalent to the system of n homogeneous equations $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. This latter system has a non-trivial solution (i.e., a solution with $\boldsymbol{x} \ne \boldsymbol{0}$) if and only if $\boldsymbol{A} - \lambda\boldsymbol{I}$ is singular, i.e., $|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$. This gives a polynomial of order n in the unknown λ (this can be shown by expanding the determinant explicitly), which may be solved for the possible values of λ, and then the characteristic vectors can be obtained for each of the characteristic roots.

For the most part, we will be concerned with symmetric matrices. There are two important properties of characteristic roots and vectors in the case of real, symmetric matrices:

(1) the characteristic roots will all be real (as opposed to complex numbers of the form $\lambda = \alpha + \beta i$ with $i = \sqrt{-1}$); and

(2) the characteristic vectors will be orthogonal to each other. Furthermore, if a characteristic root λ has multiplicity k (is repeated as a solution to the polynomial k times) - for example, the quadratic equation, $\lambda^2 - 2\lambda + 1 = 0$, has the root $\lambda = 1$ repeated twice as its solution - then there will be k orthogonal characteristic vectors corresponding to this root. Sometimes the set of characteristic vectors is not unique with repeated characteristic roots, particularly when $\lambda = 0$ is a repeated root.

An $n \times n$ symmetric matrix $\boldsymbol{A}$ has characteristic roots $(\lambda_1,...,\lambda_n)$, which possibly are not all distinct. Corresponding to these roots is a set of **orthogonal characteristic vectors** $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ such that $\boldsymbol{x}_i^\top \boldsymbol{x}_j = 0$ for all $i \neq j$, $i, j = 1,...,n$. The two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be **orthogonal** if and only if $\boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x} = 0$ and $\boldsymbol{x}^\top \boldsymbol{x} \neq 0, \boldsymbol{y}^\top \boldsymbol{y} \neq 0.$ We normalize the vectors so that $\boldsymbol{x}_i^\top \boldsymbol{x}_i = 1,\ i = 1, \cdots, n$. That is, each characteristic vector has unit length. Such a set of normalized vectors is called an **orthonormal set**.

If $\boldsymbol{X}$ denotes the $n \times n$ matrix whose columns are the vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, then it follows that $\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}_n$, so that $\boldsymbol{X}^\top = \boldsymbol{X}^{-1}$. That is, the transpose of $\boldsymbol{X}$ equals its inverse. We call such a matrix $\boldsymbol{X}$ an **orthogonal matrix**.

Now, if $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ for each i = 1,...,n, then it follows that $\boldsymbol{x}_j^\top \boldsymbol{A} \boldsymbol{x}_i = \lambda_i \boldsymbol{x}_j^\top \boldsymbol{x}_i = \lambda_i \delta_{ij}$, where δ_{ij} = 1 if $i = j$ and 0 otherwise (This is called the **Kronecker delta**.). Consequently, $\boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X} = \mathbf{diag}[\lambda_i]$, the diagonal matrix with the characteristic roots of $\boldsymbol{A}$ on the main diagonal and zeros everywhere else. Alternatively, by pre-multiplying $\boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X}$ by $\boldsymbol{X}$ and then post-multiplying the result by $\boldsymbol{X}^\top$ (remembering that $\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^\top = \boldsymbol{I}_n$), we can write $\boldsymbol{A} = \boldsymbol{X}\mathbf{diag}[\lambda_i]\boldsymbol{X}^\top$. Assembling the characteristic vectors of $\boldsymbol{A}$ as the columns of $\boldsymbol{X}$ and forming the product $\boldsymbol{X}^\top \boldsymbol{A} \boldsymbol{X}$ produces a diagonal matrix with the characteristic roots of $\boldsymbol{A}$ displayed on the principle diagonal, $\mathbf{diag}(\lambda_i)$.

This is a very useful result. The reason is that $\boldsymbol{X}$ is as legitimate a basis for n space as $\boldsymbol{I}_n$ is, in that $\boldsymbol{X}$ has full rank (spans n dimensions), each vector (axis) of $\boldsymbol{X}$ has unit length, and each pair of vectors (axes) are perpendicular (orthogonal) to each other. Thus $\boldsymbol{X}$ simply represents a **rotation of axes** relative to $\boldsymbol{I}_n$ in n dimensional space. The characteristic roots, or eigen values, λ_i, identify the **relative weights** (i.e., the lengths in terms of the matrix $\boldsymbol{A}$) of the new basis vectors compared to the orginal basis, $\boldsymbol{I}_n$.

VI. Quadratic Forms and Positive Definite Matrices

If $\boldsymbol{x}$ denotes an $n \times 1$ column vector and $\boldsymbol{A}$ an $n \times n$ real symmetric matrix, then $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$ defines a quadratic form in the elements of $\boldsymbol{x}$. The quadratic form and the matrix $\boldsymbol{A}$

are **positive definite** if and only if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0 \ \forall \ \boldsymbol{x} \neq \mathbf{0}$, and **positive semidefinite** if and only if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0 \ \forall \ \boldsymbol{x} \in \mathbb{R}^n$. If $\boldsymbol{A}$ is positive definite then $\boldsymbol{A}$ is nonsingular. If $\boldsymbol{A}$ is positive semidefinite and nonsingular, then $\boldsymbol{A}$ is positive definite.

A. Suppose that $\boldsymbol{A}$ is an $n \times n$ real, symmetric, and positive definite matrix. Note that $\boldsymbol{I}_n$ is positive definite.

1. If $\boldsymbol{B}$ is $n \times s$ ($s < n$) with $\rho(\boldsymbol{B}) = s$, then $\boldsymbol{B}^\top \boldsymbol{A}\boldsymbol{B}$ is positive definite. This follows because $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{y}$ is an $n \times 1$ column vector, $\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y} > 0 \ \forall \ \boldsymbol{y} \neq \mathbf{0}$ since $\boldsymbol{A}$ is positive definite, and $\boldsymbol{x}^\top \boldsymbol{B}^\top \boldsymbol{A}\boldsymbol{B}\boldsymbol{x} = \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y} > 0 \ \forall \ \boldsymbol{x} \neq \mathbf{0}$ by the full rank of $\boldsymbol{B}$.

2. $\boldsymbol{A}^{-1}$ is positive definite.

3. $\boldsymbol{A}$ is positive definite if and only if all the characteristic roots of $\boldsymbol{A}$ are positive.

4. $|\boldsymbol{A}| > 0$.

Thus, a (symmetric) positive definite matrix is nonsingular, has positive characteristic roots, and a positive determinant. All principle minors are positive (The principle minors are the determinants of the various submatrices formed by deleting corresponding rows and columns associated with the diagonal elements of the original matrix). The inverse of a positive definite matrix is positive definite.

B. Let $\boldsymbol{X}$ be the orthogonal matrix that diagonalizes $\boldsymbol{A}$, that is, $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X} = \mathbf{diag}[\lambda_i]$. Define the diagonal matrix $\boldsymbol{\Delta} = \mathbf{diag}\left(\sqrt{\lambda_i}\right)$. Then

$$\boldsymbol{\Delta}^{-1}\boldsymbol{X}^{\top}\boldsymbol{A}\boldsymbol{X}\boldsymbol{\Delta} = \mathbf{diag}\left[1\big/\sqrt{\lambda_i}\right]\mathbf{diag}[\lambda_i]\mathbf{diag}\left[1\big/\sqrt{\lambda_i}\right] = \boldsymbol{I}_n$$

This is equivalent to $\boldsymbol{Q}^{\top}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{I}_n$, where $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{\Delta}$. $\boldsymbol{Q}$ is nonsingular since $\boldsymbol{X}$ and $\boldsymbol{\Delta}$ are, so that $\boldsymbol{A} = \boldsymbol{Q}^{\top}\boldsymbol{Q}$. Thus, if $\boldsymbol{A}$ is (symmetric and) positive definite, then we can always find a nonsingular matrix $\boldsymbol{Q}$ such that $\boldsymbol{A} = \boldsymbol{Q}^{\top}\boldsymbol{Q}$.

Note: The trace of $\boldsymbol{A}$ is the sum of its characteristic roots, since

$$\sum_{i=1}^{n}\lambda_i = \mathrm{tr}\left(\mathrm{diag}(\lambda_i)\right) = \mathrm{tr}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right) = \mathrm{t}r\left(\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}'\right) = \mathrm{tr}\left(\boldsymbol{A}\right)$$

because $\boldsymbol{X}\boldsymbol{X}^{\top} = \boldsymbol{I}_n$.

C. If $\boldsymbol{A}$ is positive (semi-)definite, then $-\boldsymbol{A}$ is negative (semi-)definite.

VII. Differential Calculus in Matrix Notation

A. First, consider the simple linear function $\boldsymbol{a}'\boldsymbol{x} = \sum_{i=1}^{n} a_i x_i$ **.** If we take the partial derivative with respect to each x$_i$, we obtain

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_1} = a_1$$

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_2} = a_2$$

$$\vdots$$

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_n} = a_n$$

Arranging the n partial derivatives as an n×1 vector, we have

$$\frac{\partial(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{a}\,.$$

B. Next, consider the set of m linear functions $\boldsymbol{Ax}$, where $\boldsymbol{A}$ is an m×n matrix, so that

$$\boldsymbol{Ax} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} x_j \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{1\bullet}\boldsymbol{x} \\ \boldsymbol{a}_{2\bullet}\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_{m\bullet}\boldsymbol{x} \end{bmatrix},$$

where the "•" for the second subscript on the far right-hand-side indicates **rows** of **A**. For each i = 1,...,m (that is, each **row** of $\boldsymbol{A}$), we have from the previous case that

$$\frac{\partial(\boldsymbol{a}_{i\bullet}\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} = (\boldsymbol{a}_{i\bullet})^{\top},$$

an $n\times1$ **column** vector. Since $\boldsymbol{Ax}$ is an $m\times1$ column vector, to keep track of rows and columns properly we should have $\partial(\boldsymbol{Ax})/\partial\boldsymbol{x}^{\top}$ defined as an $m\times n$ matrix. That is, the number of **rows** of $\partial(\boldsymbol{Ax})/\partial\boldsymbol{x}^{\top}$ equals the number of **rows** of $\boldsymbol{Ax}$, while the number of **columns** of $\partial(\boldsymbol{Ax})/\partial\boldsymbol{x}^{\top}$ equals the number of elements of $\boldsymbol{x}$. We arrange the **rows** $\partial(\boldsymbol{a}_{i\bullet}\boldsymbol{x})/\partial\boldsymbol{x}^{\top} = \mathbf{a}_{i\bullet}$ vertically to obtain the definition

$$\frac{\partial(\boldsymbol{Ax})}{\partial \boldsymbol{x}^{\top}} = \begin{bmatrix} \boldsymbol{a}_{1\bullet} \\ \boldsymbol{a}_{2\bullet} \\ \vdots \\ \boldsymbol{a}_{m\bullet} \end{bmatrix} = \boldsymbol{A}.$$

Similarly, if we arrange the **columns** of $\partial(\boldsymbol{a}_{i\bullet}\boldsymbol{x})^{\top}/\partial\boldsymbol{x} = (\boldsymbol{a}_{i\bullet})^{\top}$ side by side, we have

$$\frac{\partial(\boldsymbol{Ax})^{\top}}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{x}^{\top}\boldsymbol{A}^{\top})}{\partial \boldsymbol{x}} = \left[(\boldsymbol{a}_{1\bullet})^{\top} \quad (\boldsymbol{a}_{2\bullet})^{\top} \quad \cdots \quad (\boldsymbol{a}_{m\bullet})^{\top}\right] = \boldsymbol{A}^{\top},\ n\times m.$$

To summarize, since $\boldsymbol{Ax}$ is an $m\times1$ column vector and $\boldsymbol{x}^{\top}$ is an $n\times1$ row vector, we must have $\partial(\boldsymbol{Ax})/\partial\boldsymbol{x}^{\top}$ as an $m\times n$ matrix. Also, since $\boldsymbol{x}^{\top}\boldsymbol{A}^{\top} = (\boldsymbol{Ax})^{\top}$ is a $1\times m$ row vector and $\boldsymbol{x}$ is an $n\times1$ column vector, we need $\partial(\boldsymbol{Ax})^{\top}/\partial\boldsymbol{x}$ to be an $n\times m$ matrix. This is precisely what these definitions give us.

However, note that it is essential to keep track of and account for the relative **positions** and **orientation** of the **function** (i.e, $\boldsymbol{Ax}$) and the **variables** (i.e, $\boldsymbol{x}$).

C. Consider $\boldsymbol{x}^\top\boldsymbol{A}$ where $\boldsymbol{A}$ is an n×m matrix and $\boldsymbol{x}$ is an n×1 column vector,

$$\boldsymbol{x}^\top\boldsymbol{A} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n a_{i1}x_i & \sum_{i=1}^n a_{i2}x_i & \cdots & \sum_{i=1}^n a_{im}x_i \end{bmatrix}, 1\times m.$$

For each j = 1,...,m we have $\partial(\boldsymbol{a}_{\bullet j}^\top\boldsymbol{x})/\partial\boldsymbol{x} = \boldsymbol{a}_{\bullet j}$, an n×1 column vector. Since $\boldsymbol{x}^\top\boldsymbol{A}$ is a 1×m row vector, we must have $\partial(\boldsymbol{x}^\top\boldsymbol{A})/\partial\boldsymbol{x}$ as an n×m matrix. So we arrange the columns $\boldsymbol{a}_{\bullet j}$ side by side to obtain the definition

$$\frac{\partial\left(\boldsymbol{x}^\top\boldsymbol{A}\right)}{\partial\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{a}_{\bullet 1} & \boldsymbol{a}_{\bullet 2} & \cdots & \boldsymbol{a}_{\bullet m} \end{bmatrix} = \boldsymbol{A}.$$

D. The quadratic form $\boldsymbol{x}^\top\boldsymbol{Ax}$, where $\boldsymbol{A}$ is a symmetric n×n matrix and $\boldsymbol{x}$ is an n×1 column vector, satisfies

$$\boldsymbol{x}^\top\boldsymbol{Ax} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n + a_{22}x_2^2 + 2a_{23}x_2x_3 + \cdots + a_{nn}x_n^2.$$

$$= a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n + a_{22}x_2^2 + 2a_{23}x_2x_3 + \cdots + a_{nn}x_n^2 .$$

Note that assuming symmetry in the quadratic form is without loss in generality, since $\boldsymbol{x}^\top\boldsymbol{Ax} = ½\boldsymbol{x}^\top\boldsymbol{Ax} + ½\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{x} = ½\boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{A}^\top)\boldsymbol{x}$.

Taking the partial derivative of $\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}$ with respect to the individual elements of $\boldsymbol{x}$, gives

$$\frac{\partial\left(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}\right)}{\partial x_1} = 2(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) = 2\boldsymbol{a}_{1\bullet}\boldsymbol{x}$$

$$\frac{\partial\left(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}\right)}{\partial x_2} = 2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n) = 2\boldsymbol{a}_{2\bullet}\boldsymbol{x}$$

$$\vdots$$

$$\frac{\partial\left(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}\right)}{\partial x_n} = 2(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n) = 2\boldsymbol{a}_{n\bullet}\boldsymbol{x}$$

where $a_{ij} = a_{ji}$ for all $i,j = 1,...,n$ has been used (e.g., $a_{21} = a_{12}$). Since $\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}$ is a scalar, we must have $\partial(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x})/\partial\boldsymbol{x}$ as an $n\times1$ column vector. Therefore, we arrange the elements of $\partial(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x})/\partial x_i$ vertically to obtain the definition

$$\frac{\partial\left(\boldsymbol{x}^{\top}\boldsymbol{A}\boldsymbol{x}\right)}{\partial\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}\,.$$

E. Consider $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$, where $\boldsymbol{y}$ is an $m\times1$ vector, with each element of $\boldsymbol{y}$ a function of the $n\times1$ vector $\boldsymbol{x}$, $\boldsymbol{f}:\mathbb{R}^n \to \mathbb{R}^m$, that is,

$$\begin{aligned} y_1 &= f_1(\boldsymbol{x}) \\ y_2 &= f_2(\boldsymbol{x}) \\ &\vdots \\ y_m &= f_m(\boldsymbol{x}). \end{aligned}$$

Each y_i can be partially differentiated with respect to each x_j, giving a total of mn partial derivatives, which we arrange as an $m\times n$ matrix,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\top}} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\top}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, m \times n..$$

Note that this is fully consistent with the definition in part **B** above for **linear** functions. This notation helps us keep proper track of location and orientation of vector-valued functions and matrix-valued derivatives.

F. To distinguish between minimum and maximum points, we need to look at second-order derivatives. Let $y = f(\boldsymbol{x})$ be a scalar (i.e., real-valued) function of the $n\times 1$ vector $\boldsymbol{x}$ and set the $n\times 1$ vector of first-order partial derivatives, $\partial y / \partial \boldsymbol{x} = \partial f(\boldsymbol{x})/\partial \boldsymbol{x}$, equal to $\mathbf{0}$ (the $n\times 1$ column vector of zeros), to find a stationary point for y – that is, y is neither increasing nor decreasing in any of the elements of $\boldsymbol{x}$ at the point $\boldsymbol{x}^{\circ}$, say.

The point $y^{\circ} = f(\boldsymbol{x}^{\circ})$ will be a unique local maximum (minimum) if

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 y}{\partial x_i \partial x_j} dx_i dx_j < 0 \ (> 0, \text{respectively})$$

for every vector $\boldsymbol{dx}$ such that the individual dx_i's are arbitrarily small, not all of them are zero, and all of the second-order partial derivatives, $\partial^2 y/\partial x_i \partial x_j$, are evaluated at the stationary point, $\boldsymbol{x}^{\circ}$.

In matrix notation, we define the **<u>Hessian matrix</u>** for $y = f(\boldsymbol{x})$ as

$$\frac{\partial^2 y}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} = \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}.$$

Note: ***The Hessian matrix of a twice continuously differentiable function is symmetric. This is known as Young's Theorem.***

Using matrix notation, the sufficient second-order condition for a local maximum is

$$d\boldsymbol{x}^\top \left(\frac{\partial^2 y}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} \right) d\boldsymbol{x} < 0 \ \forall \ d\boldsymbol{x} \neq \boldsymbol{0} : \|d\boldsymbol{x}\| \equiv \sqrt{\textstyle\sum_{i=1}^n dx_i^2} < \varepsilon ,$$

where $\varepsilon > 0$ is an arbitrarily small positive number. Similarly, the sufficient second-order condition for a local minimum is

$$d\boldsymbol{x}^\top \left(\frac{\partial^2 y}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} \right) d\boldsymbol{x} > 0 \ \forall \ d\boldsymbol{x} \neq \boldsymbol{0} : \|d\boldsymbol{x}\| \equiv \sqrt{\textstyle\sum_{i=1}^n dx_i^2} < \varepsilon .$$

Therefore, if $\partial^2 y / \partial \boldsymbol{x} \partial \boldsymbol{x}^\top$, is positive (negative) definite at the point $\boldsymbol{x}^\circ$, then $\boldsymbol{x}^\circ$ is a unique local minimum (maximum).

Discussion: Because the Hessian matrix is evaluated at the point $\boldsymbol{x}^\circ$, the quadratic form $Q(d\boldsymbol{x}) = d\boldsymbol{x}^\top \left(\frac{\partial^2 y}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} \right) d\boldsymbol{x} = d\boldsymbol{x}^\top \left(\frac{\partial^2 f(\boldsymbol{x}^\circ)}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} \right) d\boldsymbol{x}$ is a function of the $n \times 1$ vector $d\boldsymbol{x}$, and the Hessian matrix is an $n \times n$ matrix of constants with respect to $d\boldsymbol{x}$, for a ***given*** value of $\boldsymbol{x}^\circ$. Note that $Q(d\boldsymbol{x})$ is homogeneous of degree 2 in $d\boldsymbol{x}$; that is,

$$Q(t d\boldsymbol{x}) \equiv t^2 Q(d\boldsymbol{x}) \ \forall \ t \geq 0 .$$

As a result, if we can find a vector $\boldsymbol{dx}$ with length ε that causes Q to become positive (in the case where we are checking to see if $\boldsymbol{x}^\circ$ maximizes y), then we can also find a vector $\boldsymbol{z} \equiv t\boldsymbol{dx}$ of any length we desire that also causes Q to be positive. This is why the conditions on the Hessian (e.g., negative definiteness) can be stated in terms of ***any*** $\boldsymbol{z} \in \mathbb{R}^n$, not just for the $\boldsymbol{z}$**'**s contained in an n-dimensional ball with radius ε.

Jacobians

Let

$$\begin{aligned} y_1 &= f_1(\boldsymbol{x}) \\ y_2 &= f_2(\boldsymbol{x}) \\ &\vdots \\ y_n &= f_n(\boldsymbol{x}) \end{aligned}$$

be n functions of n variables, i.e., $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$, $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$. Suppose each f_i is twice continuously differentiable in each x_j. The ***Jacobian*** determinant (or Jacobian) is defined by

$$|\boldsymbol{J}| = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}.$$

The n functions $f_1,\ldots,f_n$ are ***functionally independent*** in the n-dimensional ball around the point $\boldsymbol{x}$ defined by $\|\boldsymbol{x} + \Delta\boldsymbol{x} - \boldsymbol{x}\| = \|\Delta\boldsymbol{x}\| = \sqrt{\sum_{j=1}^n \Delta x_j^2} < \varepsilon$ for some small positive $\varepsilon > 0$ if and only if $|\boldsymbol{J}| \neq 0$ at $\boldsymbol{x}$.

<u>Note:</u> $\forall\ \boldsymbol{x}+\Delta\boldsymbol{x}$ ***close enough*** to $\boldsymbol{x}$, the vector-valued function $\boldsymbol{f}(\boldsymbol{x}+\Delta\boldsymbol{x})$ is given by

$$f_i(\boldsymbol{x}+\Delta\boldsymbol{x}) = f_i(\boldsymbol{x}) + \sum_{j=1}^{n} \frac{\partial f_i(\boldsymbol{x})}{\partial x_j}\Delta x_j + o(\varepsilon)$$

$$= f_i(\boldsymbol{x}) + \frac{\partial f_i(\boldsymbol{x})}{\partial \boldsymbol{x}^{\top}}\Delta\boldsymbol{x} + o(\varepsilon),\ i = 1,\cdots,n,$$

where $\lim_{\varepsilon\to 0}\left[o(\varepsilon)/\varepsilon\right] = 0$, by Taylor's theorem. Therefore, for small enough ε, we can ignore the term $o(\varepsilon)$ and write

$$\Delta y_i = f_i(\boldsymbol{x}+\Delta\boldsymbol{x}) - f_i(\boldsymbol{x}) = \frac{\partial f_i(\boldsymbol{x})}{\partial \boldsymbol{x}^{\top}}\Delta\boldsymbol{x},\ i = 1,\cdots,n.$$

In matrix notation we have

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^{\top}}\Delta\boldsymbol{x} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\boldsymbol{x})}{\partial x_1} & \frac{\partial f_n(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{bmatrix} = \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_n \end{bmatrix} = \Delta\boldsymbol{y},$$

which is a ***linearized*** representation of the ***local*** transformation from ***changes*** in $\boldsymbol{x}$ to ***changes*** in $\boldsymbol{y}$. Note that the linear map associated with the $n\times n$ matrix $\partial\boldsymbol{f}(\boldsymbol{x})/\partial\boldsymbol{x}^{\top}$ is well-defined globally, i.e., $\forall\ \Delta\boldsymbol{x}\in\mathbb{R}^n$. In general, however, the linear map provides a "very close" approximation to the possibly nonlinear map $\boldsymbol{f}(\boldsymbol{x}+\Delta\boldsymbol{x})$ only for arbitrarily small changes in $\boldsymbol{x}$ in any direction, $\|\Delta\boldsymbol{x}\| < \varepsilon$.

As $\varepsilon\to 0$***, the linear approximation*** $\boldsymbol{f}(x+\Delta x) - \boldsymbol{f}(x) = \frac{\partial\boldsymbol{f}(\boldsymbol{x})}{\partial\boldsymbol{x}^{\top}}\Delta x$ ***becomes <u>exact</u>.***

<u>Functional Dependence and Independence</u>

From linear (matrix) algebra, we know that the matrix $\left[\partial f_i(\boldsymbol{x})/\partial x_j\right]$ **spans** *n* dimensions if and only if $|\boldsymbol{J}| \neq 0$. Keep in mind that $\left[\partial f_i(\boldsymbol{x})/\partial x_j\right]$ is independent of (i.e., is not a function of) $\Delta\boldsymbol{x}$.

Suppose that $f_2(\boldsymbol{x}) = g(f_1(\boldsymbol{x}))$, where g is an arbitrary continuously differentiable function of $y_1 = f_1(\boldsymbol{x})$. Without loss of generality (WLOG), we only need to consider f_1 and f_2. since we can always simply reorder the rows of $\boldsymbol{f}$ to meet our notational needs. Then

$$\frac{\partial f_2(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{dg(f_1(\boldsymbol{x}))}{dy_1} \cdot \frac{\partial f_1(\boldsymbol{x})}{\partial \boldsymbol{x}}.$$

by the chain rule. That is, the *n*–vector $\partial f_2(\boldsymbol{x})/\partial \boldsymbol{x}$ is ***proportional*** to the *n*–vector $\partial f_1(\boldsymbol{x})/\partial \boldsymbol{x}$ at each point $\boldsymbol{x}$. This implies that $|\boldsymbol{J}| = 0$ from results in linear algebra, and the Jacobian matrix is singular (the Jacobian determinant vanishes). In other words, any functinoal dependency between a pair of functions $\{f_1, f_2\}$ implies that the above linear representation can not span all *n*–dimensions, even in an arbitrarily small neighborhood of $\boldsymbol{x}$.

Note that the factor of proportionality, $dg(f_1(\boldsymbol{x}))/dy_1$ in general changes with $\boldsymbol{x}$. But remember that this factor is ***constant*** with respect to $\Delta\boldsymbol{x}$ in the above linearized representation for the changes in $\boldsymbol{y}$.

Now suppose, again, WLOG, that

$$f_n(\boldsymbol{x}) = g\left([f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_{n-1}(\boldsymbol{x})]^\top\right)$$

for some arbitrary, continuously differentiable, real-valued function

$$g([y_1, y_2, \ldots, y_{n-1}]^\top),\ g : \mathbb{R}^{n-1} \to \mathbb{R}\,.$$

Then, for each j = 1, ..., n, we have

$$\frac{\partial f_n(\boldsymbol{x})}{\partial x_j} = \sum_{i=1}^{n-1} \frac{\partial g([\,f_1(\boldsymbol{x}),\, f_2(\boldsymbol{x}), \ldots,\, f_{n-1}(\boldsymbol{x})]^\top)}{\partial y_i} \frac{\partial f_i(\boldsymbol{x})}{\partial x_j}$$

Clearly, this means that, for each $\boldsymbol{x}$, $\partial f_n(\boldsymbol{x})/\partial \boldsymbol{x}$ is an exact linear combination of the first n–1 column vectors $\{\partial f_1(\boldsymbol{x})/\partial \boldsymbol{x}, \cdots, \partial f_{n-1}(\boldsymbol{x})/\partial \boldsymbol{x}\}$, with a coefficient matrix that si defined by the n–1 vectors $\{\partial g_1(\boldsymbol{y}_{\sim n})/\partial \boldsymbol{y}_{\sim n}, \cdots, \partial g_{n-1}(\boldsymbol{y}_{\sim n})/\partial \boldsymbol{y}_{\sim n}\}$, which in general vary with $\boldsymbol{x}$, but are constant with respect to $\Delta \boldsymbol{x}$.

Summarizing the implications of these developments:

1. ***At any point x, if any*** f_i ***is a (continuously differentiable) function of any of the remaining*** f_k, $k \neq i$, ***then |J| = 0 and the Jacobian matrix*** $\left[\partial f_i/\partial x_j\right]$ ***can not span n-space at that point.***

2. ***The functions {***f_1, ..., f_n***} mapping*** $\mathbb{R}^n$ ***into*** $\mathbb{R}^n$ ***are functionally independent at the point x if and only if*** $|\boldsymbol{J}| \neq 0$ ***at x.***

3. ***If*** $|\boldsymbol{J}| \neq 0\ \forall\ \boldsymbol{x}$ ***in the domain of f, say,*** $\mathcal{X} \subset \mathbb{R}^n$***, then the map*** $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ ***is functionally independent on*** $\mathcal{X}$***.***

Functional independence is thus equivalent to the full dimension of the linear space locally spanned by a system of n functions in n variables. This concept plays a fundamental role in virtually all qualitative analyses of comparative statics and dynamics properties in economics.