Linear Algebra and Matrices

Notes from Charles A. Rohde

Fall 2003

Contents

æ

Chapter 1

Linear Vector Spaces

1.1 Definition and Examples

Definition 1.1 A linear vector space $\mathcal{V}$ is a set of points (called vectors) satisfying the following conditions:

(1) An operation $+$ exists on $\mathcal{V}$ which satisfies the following properties:

(a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(b) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$

(c) A vector $\mathbf{0}$ exists in $\mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$

(d) For every $\mathbf{x} \in \mathcal{V}$ a vector $-\mathbf{x}$ exists in $\mathcal{V}$ such that $(-\mathbf{x}) + \mathbf{x} = \mathbf{0}$

(2) An operation $\circ$ exists on $\mathcal{V}$ which satisfies the following properties:

(a) $\alpha \circ (\mathbf{x} + \mathbf{y}) = \alpha \circ \mathbf{y} + \alpha \circ \mathbf{x}$

(b) $\alpha \circ (\beta \circ \mathbf{x}) = (\alpha\beta) \circ \mathbf{x}$

(c) $(\alpha + \beta) \circ \mathbf{x} = \alpha \circ \mathbf{x} + \beta \circ \mathbf{x}$

(d) $1 \circ \mathbf{x} = \mathbf{x}$

where the scalars α and β belong to a field $\mathcal{F}$ with the identity being given by 1. For ease of notation we will eliminate the $\circ$ in scalar multiplication.

In most applications the field $\mathcal{F}$ will be the field of real numbers $\mathbf{R}$ or the field of complex numbers $\mathbf{C}$. The $\mathbf{0}$ vector will be called the **null vector** or the **origin**.

example 1: Let $\mathbf{x}$ represent a point in two dimensional space with addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} \quad \text{and} \quad \alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}$$

The origin and negatives are defined by

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad -\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix}$$

example 2: Let $\mathbf{x}$ represent a point in n dimensional space (called Euclidean space and denoted by $\mathbf{R}^n$) with addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

The origin and negatives are defined by

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad -\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$$

example 3: Let $\mathbf{x}$ represent a point in n dimensional complex space (called Unitary space and denoted by $\mathbf{C}^n$) with addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

The origin and negatives are defined by

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad - \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$$

In this case the x_i and y_i can be complex numbers as can the scalars.
example 4: Let $\mathbf{p}$ be an nth degree polynomial i.e.

$$\mathbf{p}(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$$

where the α_i are complex numbers. Define addition and scalar multiplication by

$$(\mathbf{p}_1 + \mathbf{p}_2)(x) = \sum_{i=0}^{n} (\alpha_{1i} + \alpha_{2i}) x^i \quad \text{and} \quad \alpha \mathbf{p} = \sum_{i=0}^{n} \alpha \alpha_i x^i$$

The origin is the null function and $-\mathbf{p}$ is defined by

$$-\mathbf{p} = \sum_{i=0}^{n} -\alpha_i x^i$$

example 5: Let $X_1, X_2, \ldots, X_n$ be random variables defined on a probability space and define $\mathcal{V}$ as the collection of all linear combinations of $X_1, X_2, \ldots, X_n$. Here the vectors in $\mathcal{V}$ are random variables.

1.2 Linear Independence and Bases

Definition 1.2 A finite set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is said to be a **linearly independent set** if

$$\sum_i \alpha_i \mathbf{x}_i = \mathbf{0} \Longrightarrow \alpha_i = 0 \ \text{ for each } \ i$$

If a set of vectors is not linearly independent it is said to be **linearly dependent**. If the set of vectors is empty we define $\sum_i \mathbf{x}_i = \mathbf{0}$ so that, by convention, the empty set of vectors is a linearly independent set of vectors.

Theorem 1.1 The set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is linearly dependent if and only if

$$\mathbf{x}_t = \sum_{i=1}^{t-1} \alpha_i \mathbf{x}_i \text{ for some } t \geq 2$$

Proof: Sufficiency is obvious since the equation

$$\mathbf{x}_t - \sum_{i=1}^{t-1} \alpha_i \mathbf{x}_i = \mathbf{0}$$

does not imply that all of the coefficients of the vectors are equal to 0.

Conversely, if the vectors are linearly dependent then we have

$$\sum_i \beta_i \mathbf{x}_i = \mathbf{0}$$

where at least two of the βs are non-zero (we assume that none of the $\mathbf{x}$'s are the zero vector). Thus, with suitable relabelling, we have

$$\sum_{i=1}^{t} \beta_i \mathbf{x}_i = \mathbf{0}$$

where $\beta_t \neq 0$. Thus

$$\mathbf{x}_t = -\sum_{i=1}^{t-1} \left(\frac{\beta_i}{\beta_t}\right) \mathbf{x}_i = \sum_{i=1}^{t-1} \alpha_i \mathbf{x}_i \quad \square$$

Definition 1.3 A **linear basis** or **coordinate system** in a vector space $\mathcal{V}$ is a set $\mathcal{E}$ of linearly independent vectors in $\mathcal{V}$ such that each vector in $\mathcal{V}$ can be written as a linear combination of the vectors in $\mathcal{E}$.

Since the vectors in $\mathcal{E}$ are linearly independent the representation as a linear combination is unique. If the number of vectors in $\mathcal{E}$ is finite we say that $\mathcal{V}$ is **finite dimensional**.

Definition 1.4 The **dimension** of a vector space is the number of vectors in any basis of the vector space.

If we have a set of linearly independent vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$, this set may be extended to form a basis. To see this let $\mathcal{V}$ be finite dimensional with basis $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$. Consider the set

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$$

Let $\mathbf{z}$ be the first vector in this set which is a linear combination of the preceeding ones. Since the $\mathbf{x}$'s are linearly independent we must have $\mathbf{z} = \mathbf{y}_i$ for some i. Thus the set

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_n\}$$

can be used to construct every vector in $\mathcal{V}$. If this set of vectors is linearly independent it is a basis which includes the $\mathbf{x}$'s. If not we continue removing $\mathbf{y}$'s until the remaining vectors are linearly independent.

It can be proved, using the Axiom of Choice, that every vector space has a basis. In most applications an explicit basis can be written down and the existence of a basis is a vacuous question.

Definition 1.5 A non-empty subset $\mathcal{M}$ of a vector space $\mathcal{V}$ is called a **linear manifold** or a **subspace** if $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ implies that every linear combination $\alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{M}$.

Theorem 1.2 The intersection of any collection of subspaces is a subspace.

Proof: Since $\mathbf{0}$ is in each of the subspaces it is in their intersection. Thus the intersection is non-empty. If $\mathbf{x}$ and $\mathbf{y}$ are in the intersection then $\alpha\mathbf{x} + \beta\mathbf{y}$ is in each of the subspaces and hence in their intersection. It follows that the intersection is a subspace. $\square$

Definition 1.6 If $\{\mathbf{x}_i, i \in I\}$ is a set of vectors the subspace **spanned** by $\{\mathbf{x}_i, i \in I\}$ is defined to be the intersection of all subspaces containing $\{\mathbf{x}_i, i \in I\}$.

Theorem 1.3 If $\{\mathbf{x}_i, i \in I\}$ is a set of vectors the subspace **spanned** by $\{\mathbf{x}_i, i \in I\}$, $\mathbf{sp}(\{\mathbf{x}_i, i \in I\})$, is the set of all linear combinations of the vectors in $\{\mathbf{x}_i, i \in I\}$.

Proof: The set of all linear combinations of vectors in $\{\mathbf{x}_i, i \in I\}$ is obviously a subspace which contains $\{\mathbf{x}_i, i \in I\}$. Conversely, $\mathbf{sp}(\{\mathbf{x}_i, i \in I\})$ contains all linear combinations of vectors in $\{\mathbf{x}_i, i \in I\}$ so that the two subspaces are identical. $\square$

It follows that an alternative characterization of a basis is that it is a set of linearly independent vectors which spans $\mathcal{V}$.

example 1: If $\mathcal{X}$ is the empty set then the space spanned by $\mathcal{X}$ is the $\mathbf{0}$ vector.
example 2: In $\mathbf{R}^3$ let $\mathcal{X}$ be the space spanned by $\mathbf{e}_1, \mathbf{e}_2$ where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Then $\mathbf{sp}\{\mathbf{e}_1, \mathbf{e}_2\}$ is the set of all vectors of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

example 3: If $\{X_1, X_2, \ldots, X_n\}$ is a set of random variables then the space spanned by $\{X_1, X_2, \ldots, X_n\}$ is the set of all random variables of the form $\sum_i a_i X_i$. This set is a basis if $P(X_i = X_j) = 0$ for $i \neq j$. æ

Chapter 2

Linear Transformations

Let $\mathcal{U}$ be a p dimensional vector space and let $\mathcal{V}$ be an n dimensional vector space.

2.1 Sums and Scalar Products

Definition 2.1 A linear transformation $\mathbf{L}$ from $\mathcal{U}$ to $\mathcal{V}$ is a mapping (function) from $\mathcal{U}$ to $\mathcal{V}$ such that

$$\mathbf{L}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{L}(\mathbf{x}) + \beta\mathbf{L}(\mathbf{y}) \text{ for every } \mathbf{x}, \mathbf{y} \in \mathcal{U} \text{ and all } \alpha, \beta$$

Definition 2.2 The sum of two linear transformations $\mathbf{L}_1$ and $\mathbf{L}_2$ is defined by the equation

$$\mathbf{L}(\mathbf{x}) = \mathbf{L}_1(\mathbf{x}) + \mathbf{L}_2(\mathbf{x}) \text{ for every } \mathbf{x} \in \mathcal{U}$$

Similarly $\alpha\mathbf{L}$ is defined by

$$[\alpha\mathbf{L}](\mathbf{x}) = \alpha\mathbf{L}(\mathbf{x})$$

The transformation $\mathbf{O}$ defined by $\mathbf{O}(\mathbf{x}) = \mathbf{0}$ has the properties:

$$\mathbf{L} + \mathbf{O} = \mathbf{O} + \mathbf{L} \ ; \ \mathbf{L} + (-\mathbf{L}) = (-\mathbf{L}) + \mathbf{L} = \mathbf{O}$$

Also note that $\mathbf{L}_1 + \mathbf{L}_2 = \mathbf{L}_2 + \mathbf{L}_1$. Thus linear transformations with addition and scalar multiplication as defined above constitute an additive commutative group.

2.2 Products

If $\mathcal{U}$ is p dimensional, $\mathcal{V}$ is m dimensional and $\mathcal{W}$ is n dimensional and

$$\mathbf{L}_1 : \mathcal{U} \rightarrow \mathcal{V} \text{ and } \mathbf{L}_2 : \mathcal{V} \rightarrow \mathcal{W}$$

we define the **product**, $\mathbf{L}$, of $\mathbf{L}_1$ and $\mathbf{L}_2$ by

$$\mathbf{L}(\mathbf{x}) = \mathbf{L}_2(\mathbf{L}_1(\mathbf{x}))$$

Note that the product of linear transformations is not commutative. The following are some properties of linear transformations

$$\begin{aligned} \mathbf{LO} &= \mathbf{OL} = \mathbf{O} \\ \mathbf{L}_1(\mathbf{L}_2 + \mathbf{L}_3) &= \mathbf{L}_1\mathbf{L}_2 + \mathbf{L}_1\mathbf{L}_3 \\ \mathbf{L}_1(\mathbf{L}_2\mathbf{L}_3) &= (\mathbf{L}_1\mathbf{L}_2)\mathbf{L}_3 \end{aligned}$$

If $\mathcal{U} = \mathcal{V}$, $\mathbf{L}$ is called a linear transformation on a vector space and we can define the identity transformation $\mathbf{I}$ by the equation

$$\mathbf{I}(\mathbf{x}) = \mathbf{x}$$

Then

$$\mathbf{LI} = \mathbf{IL} = \mathbf{L}$$

We can also define, in this case, $\mathbf{A}^2$, and in general $\mathbf{A}^n$ for any integer n. Note that

$$\mathbf{A}^n\mathbf{A}^m = \mathbf{A}^{n+m} \text{ and } (\mathbf{A}^n)^m = \mathbf{A}^{nm}$$

If we define $\mathbf{A}^0 = \mathbf{I}$ then we can define $p(\mathbf{A})$, where p is a polynomial by

$$p(\mathbf{A})(\mathbf{x}) = \sum_{i=0}^{n} \alpha_i \mathbf{A}^i \mathbf{x}$$

æ

Chapter 3

Matrices

3.1 Definitions

If $\mathbf{L}$ is a linear transformation from $\mathcal{U}$ (p dimensional) to $\mathcal{V}$ (n dimensional) and we have bases $\{\mathbf{x}_j : \ j \in J\}$ and $\{\mathbf{y}_i : \ i \in I\}$ respectively for $\mathcal{U}$ and $\mathcal{V}$ then for each $j \in J$ we have

$$\mathbf{L}(\mathbf{x}_j) = \sum_{i=1}^{n} \ell_{ij}\mathbf{y}_i \ \text{ for some choice of } \ \ell_{ij}$$

The $n \times p$ array

$$\begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1p} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \ell_{n1} & \ell_{n2} & \cdots & \ell_{np} \end{bmatrix}$$

is called the **matrix** of $\mathbf{L}$ relative to the bases $\{\mathbf{x}_j : \ j \in J\}$ and $\{\mathbf{y}_i : \ i \in I\}$. To determine the matrix of $\mathbf{L}$ we express the transformed jth vector of the basis in the $\mathcal{U}$ space in terms of the basis of the $\mathcal{V}$ space. The n numbers so obtained form the jth column of the matrix of $\mathbf{L}$. Note that the matrix of $\mathbf{L}$ depends on the bases chosen for $\mathcal{U}$ and $\mathcal{V}$.

In typical applications a specific basis is usually present. Thus in $\mathbf{R}^n$ we usually choose the basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where

$$\mathbf{e}_i = \begin{bmatrix} \delta_{1i} \\ \delta_{2i} \\ \vdots \\ \delta_{ni} \end{bmatrix} \quad \text{and} \quad \delta_{ji} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

In such circumstances we call $\mathbf{L}$ the matrix of the linear transformation.

Definition 3.1 A matrix $\mathbf{L}$ is an n by p array of scalars and is a carrier of a linear transformation ($\mathbf{L}$ carries the basis of $\mathcal{U}$ to the basis of $\mathcal{V}$).

The scalars making up the matrix are called the elements of the matrix. We will write

$$\mathbf{L} = \{\ell_{ij}\}$$

and call ℓ_{ij} the i, j element of $\mathbf{L}$ (ℓ_{ij} is thus the element in the ith row and jth column of the matrix $\mathbf{L}$). The zero or null matrix is the matrix each of whose elements is 0 and corresponds to the zero transformation. A one by p matrix is called a **row vector** and is written as

$$\boldsymbol{\ell}_i^T = (\ell_{i1}, \ell_{i2}, \ldots, \ell_{ip})$$

Similarly an n by one matrix is called a **column vector** and is written as

$$\boldsymbol{\ell}_j = \begin{bmatrix} \ell_{1j} \\ \ell_{2j} \\ \vdots \\ \ell_{nj} \end{bmatrix}$$

With this notation we can write the matrix $\mathbf{L}$ as

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{\ell}_1^T \\ \boldsymbol{\ell}_2^T \\ \vdots \\ \boldsymbol{\ell}_n^T \end{bmatrix}$$

Alternatively if $\boldsymbol{\ell}_j$ denotes the jth column of $\mathbf{L}$ we may write the matrix $\mathbf{L}$ as

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{\ell}_1 & \boldsymbol{\ell}_2 & \cdots & \boldsymbol{\ell}_p \end{bmatrix}$$

3.2 Sums and Products of Matrices

If $\mathbf{L}_1$ and $\mathbf{L}_2$ are linear transformations then their sum has matrix $\mathbf{L}$ defined by the equation

$$\begin{aligned}\mathbf{L}(\mathbf{x}_j) &= (\mathbf{L}_1 + \mathbf{L}_2)(\mathbf{x}_j) \\ &= \mathbf{L}_1(\mathbf{x}_j + \mathbf{L}_2(\mathbf{x}_j) \\ &= \sum_i \ell_{ij}^{(1)} \mathbf{y}_i + \sum_i \ell_{ij}^{(2)} \mathbf{y}_i \\ &= \sum_i \left(\ell_{ij}^{(1)} + \ell_{ij}^{(2)} \right) \mathbf{y}_i\end{aligned}$$

Thus we have the following

Definition 3.2 The sum of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\}$$

Note that $\mathbf{A}$ and $\mathbf{B}$ both must be of the same order for the sum to be defined.

Definition 3.3 The multiplication of a matrix by a scalar is defined by the equation

$$\lambda \mathbf{A} = \{\lambda a_{ij}\}$$

Matrix addition and scalar multiplication of a matrix have the following properties:

- $\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$
- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A}\mathbf{B}) + \mathbf{C}$
- $\mathbf{A} + (-\mathbf{A}) = \mathbf{O}$

If $\mathbf{L}_1$ and $\mathbf{L}_2$ are linear transformations then their product has matrix $\mathbf{L}$ defined by the equation

$$\mathbf{L}(\mathbf{x}_j) = \mathbf{L}_2(\mathbf{L}_1(\mathbf{x}_j))$$

$$
\begin{aligned}
&= \mathbf{L}_2\left(\sum_k \ell_{kj}^{(1)}\mathbf{y}_k\right) \\
&= \left(\sum_k \ell_{kj}^{(1)}\mathbf{L}_2(\mathbf{y}_k)\right) \\
&= \sum_k \ell_{kj}^{(1)}\left(\sum_i \ell_{ik}^{(2)}\mathbf{z}_i\right) \\
&= \sum_i \left(\sum_k \ell_{ik}^{(2)}\ell_{kj}^{(1)}\right)\mathbf{z}_i
\end{aligned}
$$

Thus we have

Definition 3.4 The product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined by the equation

$$\mathbf{AB} = \{\sum_k a_{ik}b_{kj}\}$$

Thus the matrix of the product can be found by taking the ith row of $\mathbf{A}$ times the jth column of $\mathbf{B}$ element by element and summing. Note that the product is defined only if the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$. We say that $\mathbf{A}$ premultiplies $\mathbf{B}$ or that $\mathbf{B}$ post multiplies $\mathbf{A}$ in the product $\mathbf{AB}$. Matrix multiplication is not commutative.

Provided the indicated products exist matrix multiplication has the following properties:

- $\mathbf{AO} = \mathbf{O}$ and $\mathbf{OA} = \mathbf{O}$
- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$
- $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$
- $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

If $n = p$ the matrix $\mathbf{A}$ is said to be a square matrix. In this case we can compute $\mathbf{A}^n$ for any integer n which has the following properties:

- $\mathbf{A}^n\mathbf{A}^m = \mathbf{A}^{n+m}$

- $(\mathbf{A}^n)^m = \mathbf{A}^{nm}$

The **identity** matrix $\mathbf{I}$ can be found using the equation

$$\mathbf{I}(\mathbf{x}_j) = \mathbf{x}_j$$

so that the jth column of the identity matrix consists of a one in the jth row and zeros elsewhere. The identity matrix has the property that

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

If we define $\mathbf{A}^0 = \mathbf{I}$ then if p is a polynomial we may define $p(\mathbf{A})$ by the equation

$$p(\mathbf{A}) = \sum_{i=0}^{n} \alpha_i \mathbf{A}^i$$

3.3 Conjugate Transpose and Transpose Operations

Definition 3.5 The **conjugate transpose** of $\mathbf{A}$, $\mathbf{A}^*$ is

$$\mathbf{A}^* = \{\bar{a}_{ji}\}$$

where $\bar{a}$ is the complex conjugate of a. Thus if $\mathbf{A}$ is n by p, the conjugate transpose $\mathbf{A}^*$, is p by n with i, j element equal to the complex conjugate of the j, i element of $\mathbf{A}$.

Definition 3.6 The **transpose** of $\mathbf{A}$, $\mathbf{A}^T$ is

$$\mathbf{A}^T = \{a_{ji}\}$$

Thus if $\mathbf{A}$ is n by p the transpose $\mathbf{A}^T$ is p by n with i, j element equal to the j, i element of $\mathbf{A}$.

If $\mathbf{A}$ is a matrix with real elements then $\mathbf{A}^* = \mathbf{A}^T$.

Definition 3.7 A square matrix is said to be

- **Hermitian** if $\mathbf{A}^* = \mathbf{A}$

- **symmetric** if $\mathbf{A}^T = \mathbf{A}$
- **normal** if $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$

The following are some properties of the conjugate transpose and transpose operations:

$$\begin{array}{cc}
(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^* & (\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T \\
\mathbf{I}^* = \mathbf{I}\,;\ \mathbf{O}^* = \mathbf{O}^T & \mathbf{I}^T = \mathbf{I} \\
(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^* & (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \\
(\alpha\mathbf{A})^* = \bar{\alpha}\mathbf{A}^* & (\alpha\mathbf{A})^T = \alpha\mathbf{A}^T
\end{array}$$

Definition 3.8 A square matrix $\mathbf{A}$ is said to be

- **diagonal** if $a_{ij} = 0$ for $i \neq j$
- **upper right triangular** if $a_{ij} = 0$ for $i > j$
- **lower left triangular** if $a_{ij} = 0$ for $i < j$

A **column vector** is an n by one matrix and a **row vector** is a one by n matrix. If $\mathbf{x}$ is a column vector then $\mathbf{x}^T$ is the row vector with the same elements.

3.4 Invertibility and Inverses

Lemma 3.1 $\mathbf{A} = \mathbf{B}$ if and only if $\mathbf{Ax} = \mathbf{Bx}$ for all $\mathbf{x}$.

Proof: If $\mathbf{A} = \mathbf{B}$ the assertion is obvious.
If $\mathbf{Ax} = \mathbf{Bx}$ for all $\mathbf{x}$ then we need only take $\mathbf{x} = \mathbf{e}_i$ for $i = 1, 2, \ldots, n$ where $\mathbf{e}_i$ is the column vector with 1 in the ith row and zeros elsewhere. Thus $\mathbf{A}$ and $\mathbf{B}$ are equal row by row. □

If $\mathbf{A}$ is n by p and $\mathbf{B}$ is p by m are written as

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

then the product $\mathbf{AB}$ satisfies

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

if the partitions are conformable.

If a vector is written as $\mathbf{x} = \sum_i \alpha_i \mathbf{e}_i$ where the $\mathbf{e}_i$ form a basis for $\mathcal{V}$ then the product $\mathbf{Ax}$ gives the transformed value of $\mathbf{x}$ under the linear transformation $\mathbf{A}$ which corresponds to the matrix $\mathbf{A}$ via the formula

$$\mathbf{A}(\mathbf{x}) = \sum_i \alpha_i \mathbf{A}(\mathbf{e}_i)$$

Definition 3.9 A square matrix $\mathbf{A}$ is said to be **invertible** if

(1) $\mathbf{x}_1 \neq \mathbf{x}_2 \Longrightarrow \mathbf{A}\mathbf{x}_1 \neq \mathbf{A}\mathbf{x}_2$

(2) For every vector $\mathbf{y}$ there exists at least one $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{y}$.

If $\mathbf{A}$ is invertible and maps $\mathcal{U}$ into $\mathcal{V}$, both n dimensional, then we define $\mathbf{A}^{-1}$, which maps $\mathcal{V}$ into $\mathcal{U}$, by the equation

$$\mathbf{A}^{-1}(\mathbf{y}_0) = \mathbf{x}_0$$

where $\mathbf{y}_0$ is in $\mathcal{V}$ and $\mathbf{x}_0$ is in $\mathcal{U}$ and satisfies $\mathbf{A}\mathbf{x}_0 = \mathbf{y}_0$ (such an $\mathbf{x}_0$ exists by (2) and is unique by (1)). Thus defined $\mathbf{A}^{-1}$ is a legitimate transformation from $\mathcal{V}$ to $\mathcal{U}$. That $\mathbf{A}^{-1}$ is a linear transformation from $\mathcal{V}$ to $\mathcal{U}$ can be seen by the fact that if $\alpha\mathbf{y}_1 + \beta\mathbf{y}_2 \in \mathcal{V}$ then there exist unique $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{U}$ such that $\mathbf{y}_1 = \mathbf{A}(\mathbf{x}_1)$ and $\mathbf{y}_2 = \mathbf{A}(\mathbf{x}_2)$. Since

$$\begin{aligned} \mathbf{A}(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) &= \alpha\mathbf{A}(\mathbf{x}_1) + \beta\mathbf{A}(\mathbf{x}_2) \\ &= \alpha\mathbf{y}_1 + \beta\mathbf{y}_2 \end{aligned}$$

it follows that

$$\begin{aligned} \mathbf{A}^{-1}(\alpha\mathbf{y}_1 + \beta\mathbf{y}_2) &= \alpha\mathbf{x}_1 + \beta\mathbf{x}_2 \\ &= \alpha\mathbf{A}^{-1}(\mathbf{y}_1) + \beta\mathbf{A}^{-1}(\mathbf{y}_2) \end{aligned}$$

Thus $\mathbf{A}^{-1}$ is a linear transformation and hence has a matrix representation. Finding the matrix representation of $\mathbf{A}^{-1}$ is not an easy task however.

Theorem 3.2 If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices such that

$$\mathbf{AB} = \mathbf{CA} = \mathbf{I}$$

then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1} = \mathbf{B} = \mathbf{C}$.

Proof: If $\mathbf{Ax}_1 = \mathbf{Ax}_2$ then $\mathbf{CAx}_1 = \mathbf{CAx}_2$ so that $\mathbf{x}_1 = \mathbf{x}_2$ and (1) of defintion 3.9 is satisfied. If $\mathbf{y}$ is any vector then defining $\mathbf{x} = \mathbf{By}$ implies $\mathbf{Ax} = \mathbf{ABy} = \mathbf{y}$ so that (2) of definition 3.9 is also satisfied and hence $\mathbf{A}$ is invertible. Since $\mathbf{AB} = \mathbf{I} \Longrightarrow \mathbf{B} = \mathbf{A}^{-1}$ and $\mathbf{CA} = \mathbf{I} \Longrightarrow \mathbf{C} = \mathbf{A}^{-1}$ the conclusions follow. $\square$

Theorem 3.3 A matrix $\mathbf{A}$ is invertible if and only if $\mathbf{Ax} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ or equivalently if and only if every $\mathbf{y} \in \mathcal{V}$ can be written as $\mathbf{y} = \mathbf{Ax}$.

Proof: If $\mathbf{A}$ is invertible then $\mathbf{Ax} = \mathbf{0}$ and $\mathbf{A0} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$ since otherwise we have a contradiction. If $\mathbf{A}$ is invertible then by (2) of definition 3.9 we have that $\mathbf{y} = \mathbf{Ax}$ for every $\mathbf{y} \in \mathcal{V}$.

Suppose that $\mathbf{Ax} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$. Then if $\mathbf{x}_1 \neq \mathbf{x}_2$ we have that $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) \neq \mathbf{0}$ so that $\mathbf{Ax}_1 \neq \mathbf{Ax}_2$ i.e. (1) of definition 3.9 is satisfied. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for $\mathcal{U}$ then

$$\sum_i \alpha_i \mathbf{A}(\mathbf{x}_i) = \mathbf{0} \Longrightarrow \mathbf{A}\left(\sum_i \alpha_i \mathbf{x}_i\right) = \mathbf{0} \Longrightarrow \sum_i \alpha_i \mathbf{x}_i = \mathbf{0} \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

It follows that $\{\mathbf{Ax}_1, \mathbf{Ax}_2, \ldots, \mathbf{Ax}_n\}$ is a basis for $\mathcal{V}$ so that we can write each $\mathbf{y} \in \mathcal{V}$ as $\mathbf{y} = \mathbf{Ax}$ since

$$\mathbf{y} = \sum_i \beta_i \mathbf{A}(\mathbf{x}_i) = \mathbf{A}\left(\sum_i \beta_i \mathbf{x}_i\right) = \mathbf{Ax}$$

Thus (2) of definition 3.9 is satisfied and $\mathbf{A}$ is invertible.

Suppose now that for each $\mathbf{y} \in \mathcal{V}$ we can write $\mathbf{y} = \mathbf{Ax}$ Let $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ be a basis for $\mathcal{V}$ and let $\mathbf{x}_i$ be such that $\mathbf{y}_i = \mathbf{Ax}_i$. Then if $\sum_i \alpha_i \mathbf{x}_i = \mathbf{0}$ we have

$$\mathbf{0} \;=\; \mathbf{A}\left(\sum_i \alpha_i \mathbf{x}_i\right)$$

$$
\begin{aligned}
&= \sum_i \alpha_i \mathbf{A}(\mathbf{x}_i) \\
&= \sum_i \alpha_i \mathbf{y}_i
\end{aligned}
$$

which implies that each of the α_i equal 0. Thus $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis for $\mathcal{U}$. It follows that every $\mathbf{x}$ can be written as $\sum_i \beta \mathbf{x}_i$ and hence $\mathbf{A}\mathbf{x} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$ i.e. (1) of definition 3.9 is satisfied. By assumption, (2) of definition 3.9 is satisfied so that $\mathbf{A}$ is invertible. $\square$

Theorem 3.4

(1) If $\mathbf{A}$ and $\mathbf{B}$ are invertible so is $\mathbf{AB}$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

(2) If $\mathbf{A}$ is invertible and $\alpha \neq 0$ then $\alpha\mathbf{A}$ is invertible and $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$

(3) If $\mathbf{A}$ is invertible then $\mathbf{A}^{-1}$ is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

Proof: Since

$$(\mathbf{AB})\mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

and

$$\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$$

by Theorem 3.2, $(\mathbf{AB})^{-1}$ exists and equals $\mathbf{B}^{-1}\mathbf{A}^{-1}$

Since

$$(\alpha\mathbf{A})\alpha^{-1}\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

and

$$\alpha^{-1}\mathbf{A}^{-1}(\alpha\mathbf{A}) = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

it again follows from Theorem 3.2 that $(\alpha\mathbf{A})^{-1}$ exists and equals $\alpha^{-1}\mathbf{A}^{-1}$. $\square$

Since

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

it again follows from Theorem 3.2 that $\mathbf{A}^{-1}$ exists and equals $\mathbf{A}$.

In many problems one “guesses” the inverse of $\mathbf{A}$ and then verifies that it is in fact the inverse. The following theorems help.

Theorem 3.5

(1) Let

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad \text{where } \mathbf{A} \text{ is invertible}$$

Then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Q}^{-1} \end{bmatrix}$$

where $\mathbf{Q} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$.

(2) Let

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad \text{where } \mathbf{D} \text{ is invertible}$$

Then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

Proof:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Q}^{-1} \end{bmatrix}$$

is equal to

$$\begin{bmatrix} \mathbf{A}\mathbf{A}^{-1} + \mathbf{A}\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} - \mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{A}\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} + \mathbf{B}\mathbf{Q}^{-1} \\ \mathbf{C}\mathbf{A}^{-1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} - \mathbf{D}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\mathbf{Q}^{-1} + \mathbf{D}\mathbf{Q}^{-1} \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} \mathbf{I} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} - \mathbf{B}\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{B}\mathbf{Q}^{-1} + \mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{C}\mathbf{A}^{-1} + (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})\mathbf{Q}^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})\mathbf{Q}^{-1} \end{bmatrix}$$

which is seen to be the identity matrix. □

The proof for (2) follows similarly.

Theorem 3.6 Let $\mathbf{A}$ be n by n, $\mathbf{U}$ be m by n, $\mathbf{S}$ be m by m and $\mathbf{V}$ be m by n. Then if $\mathbf{A}$, $\mathbf{A} + \mathbf{U}^T\mathbf{S}\mathbf{V}$ and $\mathbf{S} + \mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}$ are each invertible we have

$$\left[\mathbf{A} + \mathbf{U}^T\mathbf{S}\mathbf{V}\right] = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\left[\mathbf{S} + \mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\mathbf{S}\mathbf{V}\mathbf{A}^{-1}$$

Proof:

$$\begin{aligned}
\left[\mathbf{A}+\mathbf{U}^T\mathbf{S}\mathbf{V}\right]\left[\mathbf{A}+\mathbf{U}^T\mathbf{S}\mathbf{V}\right]^{-1} &= \mathbf{I}-\mathbf{U}^T\mathbf{S}\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\mathbf{S}\mathbf{V}\mathbf{A}^{-1} \\
&\quad +\mathbf{U}^T\mathbf{S}\mathbf{V}\mathbf{A}^{-1}-\mathbf{U}^T\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\mathbf{S}\mathbf{V}\mathbf{A}^{-1} \\
&= \mathbf{I}+\mathbf{U}^T\left(\mathbf{I}-\mathbf{S}\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\right)\mathbf{S}\mathbf{V}\mathbf{A}^{-1} \\
&\quad -\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\mathbf{S}\mathbf{V}\mathbf{A}^{-1} \\
&= \mathbf{I}+\mathbf{U}^T\left(\mathbf{I}-\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]\left[\mathbf{S}+\mathbf{S}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\mathbf{S}\right]^{-1}\right)\mathbf{S}\mathbf{V}\mathbf{A}^{-1} \\
&= \mathbf{I}\ \square
\end{aligned}$$

Corollary 3.7 If $\mathbf{S}^{-1}$ exists in the above theorem then

$$\left[\mathbf{A}+\mathbf{U}^T\mathbf{S}\mathbf{V}\right]^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{U}^T\left[\mathbf{S}^{-1}+\mathbf{V}\mathbf{A}^{-1}\mathbf{U}^T\right]^{-1}\mathbf{V}\mathbf{A}^{-1}$$

Corollary 3.8 If $\mathbf{S}=\mathbf{I}$ and $\mathbf{u}$ and $\mathbf{v}$ are vectors then

$$\left[\mathbf{A}+\mathbf{u}\mathbf{v}^T\right]=\mathbf{A}^{-1}-\frac{1}{(1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})}\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}$$

The last corollary is used as starting point for the development of a theory of recursuve estimation in least squares and Kalman filtering in information theory.

3.5 Direct Products and Vecs

- If $\mathbf{A}$ is a $p\times q$ matrix and $\mathbf{B}$ is an $r\times s$ matrix then their direct product denoted by $\mathbf{A}\otimes\mathbf{B}$ is the $pr\times qs$ matrix defined by

$$\mathbf{A}\otimes\mathbf{B}=(a_{ij}\mathbf{B})=\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1q}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2q}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}\mathbf{B} & a_{p2}\mathbf{B} & \dots & a_{pq}\mathbf{B} \end{bmatrix}$$

– Properties of direct products:

$$
\begin{aligned}
\mathbf{0} \otimes \mathbf{A} &= \mathbf{A} \otimes \mathbf{0} = \mathbf{0} \\
(\mathbf{A}_1 + \mathbf{A}_2) \otimes \mathbf{B} &= \mathbf{A}_1 \otimes \mathbf{B} + \mathbf{A}_2 \otimes \mathbf{B} \\
\mathbf{A} \otimes (\mathbf{B}_1 + \mathbf{B}_2) &= \mathbf{A} \otimes \mathbf{B}_1 + \mathbf{A} \otimes \mathbf{B}_2 \\
\alpha \mathbf{A} \otimes \beta \mathbf{B} &= \alpha\beta(\mathbf{A} \otimes \mathbf{B}) \\
\mathbf{A}_1\mathbf{A}_2 \otimes \mathbf{B}_1\mathbf{B}_2 &= (\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) \\
(\mathbf{A} \otimes \mathbf{B})^T &= \mathbf{A}^T \otimes \mathbf{B}^T \\
\text{rank}\,(\mathbf{A} \otimes \mathbf{B}) &= \text{rank}\,(\mathbf{A})\text{rank}\,(\mathbf{B}) \\
\text{trace}\,(\mathbf{A} \otimes \mathbf{B}) &= [\text{trace}\,(\mathbf{A})][\text{trace}[\mathbf{B}] \\
(\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\
\det(\mathbf{A} \otimes \mathbf{B}) &= [\det(\mathbf{A})]^p[\det(\mathbf{B})]^m
\end{aligned}
$$

- If $\mathbf{Y}$ is an $n \times p$ matrix:
  – $\text{vec}_R(\mathbf{Y})$ is the $np \times 1$ matrix with the rows of $\mathbf{Y}$ stacked on top of each other.
  – $\text{vec}_C\,(\mathbf{Y})$ is the $np \times 1$ matrix with the columns of $\mathbf{Y}$ stacked on top of each other.
  – The i, j element of $\mathbf{Y}$ is the $(i-1)p + j$ element of $\text{vec}_R\,(\mathbf{Y})$.
  – The i, j element of $\mathbf{Y}$ is the $(j-1)p + i$ element of $\text{vec}_C\,(\mathbf{Y})$.
  –

$$\text{vec}_C\,(\mathbf{Y}^T) = \text{vec}_R\,(\mathbf{Y})$$

$$\text{vec}_R\,(\mathbf{ABC}) = (\mathbf{A} \otimes \mathbf{C}^T)\text{vec}_R\,(\mathbf{B})$$

  where $\mathbf{A}$ is $n \times q_1$, $\mathbf{B}$ is $q_1 \times q_2$ and $\mathbf{C}$ is $q_2 \times p$.

- The following are some relationships between vecs and direct products.

$$
\begin{aligned}
\mathbf{ab}^T &= \mathbf{b}^T \otimes \mathbf{a} \\
&= \mathbf{a} \otimes \mathbf{b}^T \\
\text{vec}_C\,(\mathbf{ab}^T) &= \mathbf{b} \otimes \mathbf{a}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{vec}_R\,(\mathbf{a}\mathbf{b}^T) &= \mathbf{a}\otimes\mathbf{b} \\
\mathrm{vec}_C\,(\mathbf{A}\mathbf{B}\mathbf{C}) &= (\mathbf{C}^T\otimes\mathbf{A})\mathrm{vec}_C(\mathbf{B}) \\
\mathrm{trace}\,(\mathbf{A}^T\mathbf{B}) &= [\mathrm{vec}_C\,(\mathbf{A})]^T[\mathrm{vec}_C\,(\mathbf{B})]
\end{aligned}
$$

Chapter 4

Matrix Factorization

4.1 Rank of a Matrix

In matrix algebra it is often useful to have the matrices expressed in as simple a form as possible. In particular, if a matrix is diagonal the operations of addition, multiplication and inversion are easy to perform.

Most of these methods are based on considering the rows and columns of a matrix as vectors in a vector space of the appropriate dimension.

Definition: If $\mathbf{A}$ is an n by p matrix then

(1) The row rank of $\mathbf{A}$ is the number of linearly independent rows of the matrix considered as vectors in p dimensional space.

(2) The column rank of $\mathbf{A}$ is the number of linearly independent columns of the matrix considered as vectors in n dimensional space.

Theorem 4.1 Let $\mathbf{A}$ be an n by p matrix. The set of all vectors $\mathbf{y}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$ is a vector space with dimension equal to the column rank of $\mathbf{A}$ and is called the **range space** of $\mathbf{A}$ and is denoted by $\mathcal{R}(\mathbf{A})$.

Proof: $\mathcal{R}(\mathbf{A})$ is non empty since $\mathbf{0} \in \mathcal{R}(\mathbf{A})$ If $\mathbf{y}_1$ and $\mathbf{y}_2$ are in $\mathcal{R}(\mathbf{A})$ then

$$\begin{aligned}\mathbf{y} &= \alpha_1\mathbf{y}_1 + \alpha_2\mathbf{y}_2 \\ \mathbf{y} &= \alpha_1\mathbf{A}\mathbf{x}_1 + \alpha_2\mathbf{A}\mathbf{x}_2 \\ \mathbf{y} &= \mathbf{A}(\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2)\end{aligned}$$

It follows that $\mathbf{y} \in \mathcal{R}(\mathbf{A})$ and hence $\mathcal{R}(\mathbf{A})$ is a vector space. If $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p\}$ are the columns of $\mathbf{A}$ then for every $\mathbf{y} \in \mathcal{R}(\mathbf{A})$ we can write

$$\mathbf{y} = \mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_p\mathbf{a}_p$$

Thus the columns of $\mathbf{A}$ span $\mathcal{R}(\mathbf{A})$. It follows that the dimension of $\mathcal{R}(\mathbf{A})$ is equal to the column rank of $\mathbf{A}$. $\square$

Theorem 4.2 Let $\mathbf{A}$ be an n by p matrix. The set of all vectors $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$ is a vector space of dimension equal to $p -$ column rank of $\mathbf{A}$. This vector space is called the **null space** of $\mathbf{A}$ and is denoted by $\mathcal{N}(\mathbf{A})$.

Proof: Since $\mathbf{0} \in \mathcal{N}(\mathbf{A})$ it follows that $\mathcal{N}(\mathbf{A})$ is non empty. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are in $\mathcal{N}(\mathbf{A})$ then

$$\begin{aligned}\mathbf{A}(\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2) &= \alpha_1\mathbf{A}\mathbf{x}_1 + \alpha_2\mathbf{A}\mathbf{x}_2 \\ &= \mathbf{0}\end{aligned}$$

so that $\mathcal{N}(\mathbf{A})$ is a vector space.

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be a basis for $\mathcal{N}(\mathbf{A})$. Then there exists vectors $\{\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_p\}$ such that $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ is a basis for p dimensional space. It follows that $\{\mathbf{A}\mathbf{x}_1, \mathbf{A}\mathbf{x}_2, \ldots, \mathbf{A}\mathbf{x}_p\}$ spans the range space of $\mathbf{A}$ since

$$\begin{aligned}\mathbf{y} &= \mathbf{A}\mathbf{x} \\ &= \mathbf{A}\left(\sum_i \alpha_i\mathbf{x}_i\right) \\ &= \sum_i \alpha_i\mathbf{A}\mathbf{x}_i\end{aligned}$$

Since $\mathbf{A}\mathbf{x}_i = \mathbf{0}$ for $i = 1, 2, \ldots, k$ it follows that $\{\mathbf{A}\mathbf{x}_{k+1}, \mathbf{A}\mathbf{x}_{k+2}, \ldots, \mathbf{A}\mathbf{x}_p\}$ spans the range space of $\mathbf{A}$. Suppose now that $\alpha_{k+1}, \alpha_{k+2} \ldots, \alpha_p$ are such that

$$\alpha_{k+1}\mathbf{A}\mathbf{x}_{k+1} + \alpha_{k+2}\mathbf{A}\mathbf{x}_{k+2} + \cdots + \alpha_p\mathbf{A}\mathbf{x}_p = \mathbf{0}$$

Then

$$\mathbf{A}(\alpha_{k+1}\mathbf{x}_{k+1} + \alpha_{k+2}\mathbf{x}_{k+2} + \cdots + \alpha_p\mathbf{x}_p) = \mathbf{0}$$

It follows that

$$\alpha_{k+1}\mathbf{x}_{k+1} + \alpha_{k+2}\mathbf{x}_{k+2} + \cdots + \alpha_p\mathbf{x}_p \in \mathcal{N}(\mathbf{A})$$

Since $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ is a basis for $\mathcal{N}(\mathbf{A})$ it follows that for some $\alpha_1, \alpha_2, \ldots, \alpha_k$ we have

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_k\mathbf{x}_k = \alpha_{k+1}\mathbf{x}_{k+1} + \alpha_{k+2}\mathbf{x}_{k+2} + \cdots + \alpha_p\mathbf{x}_p$$

or

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_k\mathbf{x}_k - \alpha_{k+1}\mathbf{x}_{k+1} + \alpha_{k+2}\mathbf{x}_{k+2} + \cdots + \alpha_p\mathbf{x}_p = \mathbf{0}$$

Since $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ is a basis for p dimensional space we have that $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. It follows that the set

$$\{\mathbf{A}\mathbf{x}_{k+1}, \mathbf{A}\mathbf{x}_{k+2}, \ldots, \mathbf{A}\mathbf{x}_p\}$$

is linearly independent and spans the range space of $\mathbf{A}$. The dimension of $\mathcal{R}(\mathbf{A}$ is thus $p - k$. But by Theorem 4.1 we also have that the dimension of $\mathcal{R}(\mathbf{A})$ is equal to the column rank of $\mathbf{A}$. Thus

$$p - k = \text{column rank of } \mathbf{A}$$

so that

$$k = p - \text{column rank of } \mathbf{A} = \dim\left(\mathcal{N}(\mathbf{A})\right. \quad \square$$

Definition 4.2 The elementary row (column) operations on a matrix are defined to be

(1) The interchange of two rows (columns).

(2) The multiplication of a row (column) by a non zero scalar.

(3) The addition of one row (column) to another row (column).

Lemma 4.3 Elementary row (column) operations do not change the row (column) rank of a matrix.

Proof: The row rank of a matrix $\mathbf{A}$ is the number of linearly independent vectors in the set defined by

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

Clearly interchange of two rows does not change the rank nor does the multiplication of a row by a non zero scalar. The addition of $a\mathbf{a}_j^T$ to $\mathbf{a}_i^T$ produces the new matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_i^T + a\mathbf{a}_j^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

which has the same rank as $\mathbf{A}$. The statements on the column rank of $\mathbf{A}$ are obtained by considering the row vectors of $\mathbf{A}^T$ and using the above results. □

Lemma 4.4 Elementary row (column) operations on a matrix can be achieved by pre (post) multiplication by non-singular matrices.

Proof: Interchange of the ith and jth rows can be achieved by pre-multiplication by the matrix

$$\mathbf{E}_1 = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ \mathbf{e}_{i-1}^T \\ \mathbf{e}_j^T \\ \mathbf{e}_{i+1}^T \\ \vdots \\ \mathbf{e}_{j-1}^T \\ \mathbf{e}_i^T \\ \mathbf{e}_{j+1}^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$$

Multiplication by a non zero scalar and addition of a scalar multiple of one row to another row can be achieved by pre multiplication by the matrices $\mathbf{E}_2$ and $\mathbf{E}_3$ where

$$\mathbf{E}_2 = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ a\mathbf{e}_i^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix} \quad \text{and} \quad \mathbf{E}_3 = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ a\mathbf{e}_j^T + \mathbf{e}_i^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$$

The statements concerning column operations are obtained by using the above results on $\mathbf{A}$ and then transposing the final results. □

Lemma 4.5 The matrices which produce elementary row (column) operations are non singular.

Proof: The inverse of $\mathbf{E}_1$ is $\mathbf{E}_1^T$. The inverse of $\mathbf{E}_2$ is the matrix

$$\mathbf{E}_2 = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ \frac{1}{a}\mathbf{e}_i^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$$

Finally the inverse of $\mathbf{E}_3$ is the matrix

$$[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_j - a\mathbf{e}_i, \ldots, \mathbf{e}_n]$$

Again the results for column operations follow from the row results. □

Lemma 4.6 The column rank of a matrix is invariant under pre multiplication by a non singular matrix. Similarly the row rank of a matrix is invariant under post multiplication by a non singular matrix.

Proof: By Theorem 4.2 the dimension of $\mathcal{N}(\mathbf{A})$ is equal to $p-$column rank of $\mathbf{A}$. Since

$$\mathcal{N}(\mathbf{A} = \{\mathbf{x} : \ \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

it follows that $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ if and only $\mathbf{E}\mathbf{A}\mathbf{x} = \mathbf{0}$ where $\mathbf{E}$ is non singular. Thus the dimension of $\mathcal{N}(\mathbf{A})$ is invariant under pre multiplication by a non singular matrix and the conclusion follows. The result for post multiplication follow by considering $\mathbf{A}^T$ and transposing. □

Corollary 4.7 The column rank of a matrix is invariant under premultiplication by an elementary matrix . Similarly the row rank of a matrix is invariant under post multiplication by an elementary matrix.

By suitable choices of elementary matrices we can thus write

$$\mathbf{PA} = \mathbf{E}$$

where $\mathbf{P}$ is the product of non singular elementary matrices and hence is non singular. The matrix $\mathbf{E}$ is a **row echelon** matrix i.e. a matrix with r non zero rows in which the first k columns consist of $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_r$ and $\mathbf{0}$'s in some order. The remaining $p - r$ columns have zeros below the first r rows and arbitrary elements in the remaining positions. It thus follows that

$$r = \text{row rank of } \mathbf{A} \leq \text{column rank of } \mathbf{A}$$

and hence

$$\text{row rank of } \mathbf{A}^T \leq \text{column rank of } \mathbf{A}^T$$

But we also have that

$$\text{row rank of } \mathbf{A}^T = \text{column rank of } \mathbf{A}$$

and

$$\text{row rank of } \mathbf{A} = \text{column rank of } \mathbf{A}^T$$

It follows that

$$\text{column rank of } \mathbf{A} \leq \text{row rank of } \mathbf{A}$$

and hence we have:

Theorem 4.8 row rank of $\mathbf{A}$ = column rank of $\mathbf{A}$

Alternative Proof of Theorem 4.8 Let $\mathbf{A}$ be an $n \times p$ matrix i.e.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

We may write $\mathbf{A}$ in terms of its columns or rows i.e.

$$\mathbf{A} = \left[\mathbf{A}_1^c\ ,\ \mathbf{A}_2^c,\ \ldots\ ,\mathbf{A}_p^c\right] = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

where the jth column vector, $\mathbf{A}_j^c$, and ith row vector, $\mathbf{a}_i^T$ are given by

$$\mathbf{A}_j^c = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \quad \mathbf{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}$$

The column vectors of $\mathbf{A}$ span a space which is a subset of R^n and is called the **column space** of $\mathbf{A}$ and denoted by $\mathcal{C}(\mathbf{A})$ i.e.

$$\mathcal{C}(\mathbf{A}) = \mathbf{sp}\left(\mathbf{A}_1^c, \mathbf{A}_2^c, \ldots, \mathbf{A}_p^c\right)$$

The row vectors of $\mathbf{A}$ span a space which is a subspace of R^p and is called the **row space** of $\mathbf{A}$ and denoted by $\mathcal{R}(\mathbf{A})$ i.e.

$$\mathcal{R}(\mathbf{A}) = \mathbf{sp}\left(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\right)$$

The dimension of $\mathcal{C}(\mathbf{A})$ is called the **column rank** of $\mathbf{A}$ and is denoted by $C = c(\mathbf{A})$. Similarly the dimension of $\mathcal{R}(\mathbf{A})$ is called the **row rank** of $\mathbf{A}$ and is denoted by $R = r(\mathbf{A})$.

Result 1: Let

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^c & \mathbf{b}_2^c & \ldots & \mathbf{b}_C^c \end{bmatrix}$$

where the columns of $\mathbf{B}$ form a basis for the column space of $\mathbf{A}$. Then there is a $C \times p$ matrix $\mathbf{L}$ such that

$$\mathbf{A} = \mathbf{BL}$$

It follows that the rows of $\mathbf{A}$ are linear combinations of the rows of $\mathbf{L}$ and hence

$$\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{L})$$

Hence

$$R = \dim[\mathcal{R}(\mathbf{A})] \leq \dim[\mathcal{R}(\mathbf{L})] \leq C$$

i.e. the row rank of $\mathbf{A}$ is less than or equal to the column rank of $\mathbf{A}$.

Result 2:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \vdots \\ \mathbf{r}_R^T \end{bmatrix}$$

where the rows of $\mathbf{R}$ form a basis for the row space of $\mathbf{A}$. Then there is a $n \times R$ matrix $\mathbf{K}$ such that

$$\mathbf{A} = \mathbf{KR}$$

It follows that the columns of $\mathbf{A}$ are linear combinations of the columns of $\mathbf{K}$ and hence

$$\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{K})$$

Hence

$$C = \dim[\mathcal{C}(\mathbf{A})] \leq \dim[\mathcal{C}(\mathbf{K})] \leq R$$

i.e. the column rank of $\mathbf{A}$ is less than or equal to the row rank of $\mathbf{A}$.

Hence we have that the row rank and column rank of a matrix are equal. Th common value is called the **rank** of $\mathbf{A}$.

Reference Harville, David A. (1997) Matrix Algebra From a Statistician's Perspective. Springer (pages 36-40).

We thus define the rank of a matrix $\mathbf{A}$, $\rho(\mathbf{A})$ to be the number of linearly independent rows or the number of linearly independent columns in the matrix $\mathbf{A}$. Note that $\rho(\mathbf{A})$ is unaffected by pre or post multiplication by non singular matrices. By pre and post multiplying by suitable elementary matrices we can write any matrix as

$$\mathbf{PAQ} = \mathbf{B} = \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

where $\mathbf{O}$ denotes a matrix of zeroes of the appropriate order. Since $\mathbf{P}$ and $\mathbf{Q}$ are non singular we have

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{Q}^{-1}$$

which is an example of a **matrix factorization**.

Another factorization of considerable importance is contained in the following Lemma.

Lemma 4.9 If $\mathbf{A}$ is an n by p matrix of rank r then there are matrices $\mathbf{B}$ (n by r) and $\mathbf{C}$ (r by p) such that

$$\mathbf{A} = \mathbf{BC}$$

where $\mathbf{B}$ and $\mathbf{C}$ are both of rank r.

Proof: Let the rows of $\mathbf{C}$ form a basis for the row space of $\mathbf{A}$. Then $\mathbf{C}$ is r by p of rank r. Since the rows of $\mathbf{C}$ form a basis for the row space of $\mathbf{A}$ we can write the ith row of $\mathbf{A}$ as $\mathbf{a}_i^T = \mathbf{b}_i^T\mathbf{C}$ for some choice of $\mathbf{b}_i$. Thus $\mathbf{A} = \mathbf{BC}$. $\square$

The fact that the rank of $\mathbf{B}$ in Lemma 4.9 is r follows from the following important result

Theorem 4.10 $\text{rank}\,(\mathbf{AB}) \leq \min\{\text{rank}\,(\mathbf{A}), \text{rank}\,(\mathbf{B})\}$

Proof: Each row of $\mathbf{AB}$ is a linear combination of the rows of $\mathbf{B}$. Thus

$$\text{rank}\,(\mathbf{AB}) \leq \text{rank}\,(\mathbf{B})$$

Similarly each column of $\mathbf{AB}$ is a linear combination of the columns of $\mathbf{A}$ so that

$$\text{rank}\,(\mathbf{AB}) \leq \text{rank}\,(\mathbf{A})$$

The conclusion follows. $\square$

Theorem 4.11 An n by n matrix is non singular if and only if it is of rank n.

Proof: For any matrix we can write

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are non singular. Since the right hand side of this equation is invertible if and only if $r = n$ the conclusion follows. $\square$

Definition 4.3 If $\mathbf{A}$ is an n by p matrix then the p by p matrix $\mathbf{A}^*\mathbf{A}$ ($\mathbf{A}^T\mathbf{A}$ if $\mathbf{A}$ is real) is called the **Gram** matrix of $\mathbf{A}$.

Theorem 4.12 If $\mathbf{A}$ is an n by p matrix then

(1) $\text{rank}\,(\mathbf{A}^*\mathbf{A}) = \text{rank}\,(\mathbf{A}) = \text{rank}\,(\mathbf{AA}^*)$

(2) $\text{rank}\,(\mathbf{A}^T\mathbf{A}) = \text{rank}\,(\mathbf{A}) = \text{rank}\,(\mathbf{A}\mathbf{A}^T)$ if $\mathbf{A}$ is real.

Proof: Assume that the field of scalars is such that $\sum_i z_i^* z_i = 0$ implies that $z_1 = z_2 = \cdots = z_n = 0$. If $\mathbf{Ax} = \mathbf{0}$ then $\mathbf{A}^*\mathbf{Ax} = \mathbf{0}$. Also if $\mathbf{A}^*\mathbf{Ax} = \mathbf{0}$ then $\mathbf{x}^*\mathbf{A}^*\mathbf{Ax} = 0$ or $\mathbf{z}^*\mathbf{z} = 0$ so that $\mathbf{z} = \mathbf{0}$ i.e. $\mathbf{Ax} = \mathbf{0}$. Hence $\mathbf{A}^*\mathbf{Ax} = \mathbf{0}$ and $\mathbf{Ax} = \mathbf{0}$ are equivalent statements. It follows that

$$\mathcal{N}(\mathbf{A}^*\mathbf{A}) = \{\mathbf{x} : \ \mathbf{A}^*\mathbf{Ax} = \mathbf{0}\} = \{\mathbf{x} : \ \mathbf{Ax} = \mathbf{0}\} = \mathcal{N}(\mathbf{A})$$

Thus $\text{rank}\,(\mathbf{A}^*\mathbf{A}) = \text{rank}\,(\mathbf{A})$. The conclusion that $\text{rank}\,(\mathbf{A}^*\mathbf{A}) = \text{rank}\,(\mathbf{A})$ follows by symmetry. $\square$

Theorem 4.13 $\text{rank}\,(\mathbf{A}+\mathbf{B}) \leq \text{rank}\,(\mathbf{A}) + \text{rank}\,(\mathbf{B})$

Proof: If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r$ and $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_s$ denote linearly independent columns of $\mathbf{A}$ and $\mathbf{B}$ respectively then

$$\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_s\}$$

span the column space of $\mathbf{A}+\mathbf{B}$. Hence

$$\text{rank}\,(\mathbf{A}+\mathbf{B}) \leq r + s = \text{rank}\,(\mathbf{A}) + \text{rank}\,(\mathbf{B}) \ \ \square$$

Definition 4.4

(1) A square matrix $\mathbf{A}$ is **idempotent** if $\mathbf{A}^2 = \mathbf{A}$.

(2) A square matrix $\mathbf{A}$ is **nilpotent** if $\mathbf{A}^m = \mathbf{O}$ for some integer m greater than one.

Definition 4.5 The **trace** of a square matrix is the sum of its diagonal elements i.e.

$$\text{tr}\,(\mathbf{A}) = \sum_i a_{ii}$$

Theorem 4.14 $\text{tr}\,(\mathbf{AB}) = \text{tr}\,(\mathbf{BA})$; $\text{tr}\,(\mathbf{A}+\mathbf{B}) = \text{tr}\,(\mathbf{A}) + \text{tr}\,(\mathbf{B})$; $\text{tr}\,(\mathbf{A}^T) = \text{tr}\,(\mathbf{A})$

Theorem 4.15 If $\mathbf{A}$ is idempotent then

$$\text{rank}\,(\mathbf{A}) = \text{tr}\,(\mathbf{A})$$

Proof: We can write $\mathbf{A} = \mathbf{BC}$ where $\mathbf{B}$ is n by r of rank r, $\mathbf{C}$ is r by n of rank r and r is the rank of $\mathbf{A}$. Thus we have

$$\mathbf{A}^2 = \mathbf{BCBC} = \mathbf{A} = \mathbf{BC}$$

Hence

$$\mathbf{B}^*\mathbf{BCBCC}^* = \mathbf{B}^*\mathbf{BCC}^*$$

Since $\mathbf{B}$ and $\mathbf{C}$ are each of rank r we have that $\mathbf{CB} = \mathbf{I}$ and hence

$$\mathrm{tr}\,(\mathbf{BC}) = \mathrm{tr}\,(\mathbf{CB}) = \mathrm{tr}\,(\mathbf{I}) = r = \mathrm{rank}\,(\mathbf{A}) \quad \square$$

Chapter 5

Linear Equations

5.1 Consistency Results

Consider a set of n linear equations in p unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1p}x_p & = & y_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2p}x_p & = & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{np}x_p & = & y_n \end{array}$$

which may be compactly written in matrix notation as

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

where $\mathbf{A}$ is the n by p matrix with i, j element equal to a_{ij}, $\mathbf{x}$ is the p by one column vector with jth element equal to x_j and $\mathbf{y}$ is the n by one column vector with ith element equal to y_i.

Often such "equations" arise without knowledge of whether they are really equations, i.e. does there exist a vector $\mathbf{x}$ which satisfies the equations? If such an $\mathbf{x}$ exists the equations are said to be **consistent**, otherwise they are said to be **inconsistent**.

The equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ can be discussed from a vector space perspective since $\mathbf{A}$ is a carrier of a linear transformation from the p dimensional space spanned by the $\mathbf{x}_i$'s to the n dimensional space where $\mathbf{y}$ "lives". It follows that a solution exists if and only if $\mathbf{y}$ is in $\mathcal{R}(\mathbf{A})$, the range space of $\mathbf{A}$. Thus a solution exists if and only if $\mathbf{y}$ is in the column space of $\mathbf{A}$ defined as the set of all vectors of the form $\{\mathbf{y} : \ \mathbf{y} = \mathbf{A}\mathbf{x} \ \text{ for some } \ \mathbf{x}\}$. If we consider the augmented matrix $[\mathbf{A}, \mathbf{y}]$ we have the following theorem on consistency

Theorem 5.1 The equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ are consistent if and only if

$$\text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}(\mathbf{A})$$

Proof: Obviously

$$\text{rank}([\mathbf{A}, \mathbf{y}]) \geq \text{rank}(\mathbf{A})$$

If the equations have a solution then there exists $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{y}$ so that

$$\text{rank}(\mathbf{A}) \leq \text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}([\mathbf{A}, \mathbf{A}\mathbf{x}]) = \text{rank}(\mathbf{A}[\mathbf{I}, \mathbf{x}]) \leq \text{rank}(\mathbf{A})$$

It follows that

$$\text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}(\mathbf{A})$$

if the equations are consistent.

Conversely if

$$\text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}(\mathbf{A})$$

then $\mathbf{y}$ is a linear combination of the columns of $\mathbf{A}$ so that a solution exists. □

Theorem 5.1 gives conditions under which solutions to the equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ exist but we need to know the form of the solution in order to explicitly solve the equations. The equations $\mathbf{A}\mathbf{x} = \mathbf{0}$ are called the **homogeneous equations**. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are solutions to the homogeneous equations then $\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2$ is also a solution to the homogeneous equations. In general, if $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s\}$ span the null space of $\mathbf{A}$ then $\mathbf{x}_0 = \sum_{i=1}^{s} \alpha_i \mathbf{x}_i$ is a solution to the homogeneous equations. If $\mathbf{x}_p$ is a **particular solution** to the equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ then $\mathbf{x}_p + \mathbf{x}_0$ is also a solution to the equations $\mathbf{A}\mathbf{x} = \mathbf{y}$. If $\mathbf{x}$ is any solution then $\mathbf{x} = \mathbf{x}_p + (\mathbf{x} - \mathbf{x}_p)$ so that every solution is of this form. We thus have the following theorem

Theorem 5.2 A general solution to the consistent equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ is of the form

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_0$$

where $\mathbf{x}_p$ is any particular solution to the equations and $\mathbf{x}_0$ is any solution to the homogeneous equations $\mathbf{Ax} = \mathbf{0}$.

A final existence question is to determine when the equations $\mathbf{Ax} = \mathbf{y}$ have a unique solution. The solution is unique if and only if the only solution to the homogeneous equations is the $\mathbf{0}$ vector i.e. the null space of $\mathbf{A}$ contains only the $\mathbf{0}$ vector. If the solution $\mathbf{x}$ is unique the homogeneous equations cannot have a non zero solution $\mathbf{x}_0$ since then $\mathbf{x} + \mathbf{x}_0$ would be another solution. Conversely, if the homogeneous equations have only the $\mathbf{0}$ vector as a solution then the solution to $\mathbf{Ax} = \mathbf{y}$ is unique (if there were two different solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ then $\mathbf{x}_1 - \mathbf{x}_2$ would be a non zero solution to the homogeneous equations).

Since the null space of $\mathbf{A}$ contains only the $\mathbf{0}$ vector if and only if the columns of $\mathbf{A}$ are linearly independent we have that the solution is unique if and only if the rank of $\mathbf{A}$ is equal to p where we assume with no loss of generality that $\mathbf{A}$ is n by p with $p \leq n$. We thus have the following theorem.

Theorem 5.3 Let $\mathbf{A}$ be n by p with $p \leq n$ then

(1) The equations $\mathbf{Ax} = \mathbf{y}$ have the general solution

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_0$$

where $\mathbf{Ax}_p = \mathbf{y}$ and $\mathbf{Ax}_0 = \mathbf{0}$ if and only if

$$\text{rank}([\mathbf{A}, \mathbf{y}]) = \text{rank}(\mathbf{A})$$

(2) The solution is unique ($\mathbf{x}_0 = \mathbf{0}$) if and only if

$$\text{rank}(\mathbf{A}) = p$$

5.2 Solutions to Linear Equations

5.2.1 A is p by p of rank p

Here there is a unique solution by Theorem 5.3 with explicit form given by $\mathbf{A}^{-1}\mathbf{y}$. The proof of this result is trivial given the following Lemma

Lemma 5.4 A p by p matrix is non singular if and only if $\text{rank}(\mathbf{A}) = p$.

Proof: If $\mathbf{A}^{-1}$ exists it is the unique matrix satisfying

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

Thus $p \geq \text{rank}(\mathbf{A}) \geq p$ i.e. the rank of $\mathbf{A}$ is p if $\mathbf{A}$ is non singular.

Conversely, if $\text{rank}(\mathbf{A}) = p$ then there exists a unique solution $\mathbf{x}_i$ to the equation $\mathbf{A}\mathbf{x}_i = \mathbf{e}_i$ for $i = 1, 2, \ldots, p$. If we define $\mathbf{B}$ by

$$\mathbf{B} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p]$$

then $\mathbf{AB} = \mathbf{I}$. Similarly there exists a unique solution $\mathbf{z}_i$ to the equation $\mathbf{A}^T\mathbf{z}_i$ for $i = 1, 2, \ldots, p$ so that if we define

$$\mathbf{C}^T = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_p]$$

then $\mathbf{A}^T\mathbf{C}^T = \mathbf{I}$ or $\mathbf{CA} = \mathbf{I}$. It follows that that $\mathbf{C} = \mathbf{B}$ i.e. that $\mathbf{A}^{-1}$ exists. $\Box$

5.2.2 A is n by p of rank p

If $\mathbf{A}$ is n by p of rank p where $p \leq n$ then by theorem 5.3 a unique solution exists and we need only find its form. Since the rank of $\mathbf{A}$ is p the rank of $\mathbf{A}^T\mathbf{A}$ is p and hence $\mathbf{A}^T\mathbf{A}$ has an inverse. Thus if we define $\mathbf{x}_1 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$ we have that

$$\mathbf{A}\mathbf{x}_1 = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{y}$$

so that $\mathbf{x}_1$ is the unique solution to the equations. $\Box$

5.2.3 A is n by p of rank r where $r \leq p \leq n$

In this case a solution may not exist and even if a solution exists it may not be unique. Suppose that a matrix $\mathbf{A}^-$ can be found such that

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

Such a matrix $\mathbf{A}^-$ is called a **generalized inverse** of $\mathbf{A}$

If the equations $\mathbf{Ax} = \mathbf{y}$ have a solution then $\mathbf{x}_1 = \mathbf{Ay}$ is a solution since

$$\mathbf{Ax}_1 = \mathbf{A}(\mathbf{A}^-\mathbf{y}) = \mathbf{AA}^-(\mathbf{Ax}) = \mathbf{AA}^-\mathbf{Ax} = \mathbf{Ax} = \mathbf{y}$$

so that $\mathbf{A}^-\mathbf{x}_1$ is a solution. The general solution is thus obtained by characterizing the solutions to the homogeneous equations i.e. we need to characterize the vectors in the null space of $\mathbf{A}$. If $\mathbf{z}$ is an arbitrary vector then $(\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}$ is in the null space of $\mathbf{A}$. The dimension of the set $\mathcal{U}$ which can be written in this form is equal to the rank of $\mathbf{I} - \mathbf{A}^-\mathbf{A}$. Since

$$\begin{aligned}(\mathbf{I} - \mathbf{A}^-\mathbf{A})^2 &= \mathbf{I} - \mathbf{A}^-\mathbf{A} - \mathbf{A}^-\mathbf{A} + \mathbf{A}^-\mathbf{AA}^-\mathbf{A} \\ &= \mathbf{I} - \mathbf{A}^-\mathbf{A}\end{aligned}$$

we see that

$$\begin{aligned}\text{rank}\,(\mathbf{I} - \mathbf{A}^-\mathbf{A}) &= \text{tr}\,(\mathbf{I} - \mathbf{A}^-\mathbf{A}) \\ &= p - \text{tr}\,(\mathbf{A}^-\mathbf{A}) \\ &= p - \text{rank}\,(\mathbf{A}^-\mathbf{A})) \\ &= p - \text{rank}\,(\mathbf{A}) \\ &= p - r\end{aligned}$$

It follows that $\mathcal{U}$ is of dimension $p-r$ and hence $\mathcal{U} = \mathcal{N}(\mathbf{A})$. Thus we have the following theorem

Theorem 5.5 If $\mathbf{A}$ is n by p of rank $r \leq p \leq n$ then

(1) If a solution exists the general solution is given by

$$\mathbf{x} = \mathbf{A}^-\mathbf{y} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}$$

where $\mathbf{z}$ is an arbitrary p by one vector and $\mathbf{A}^-$ is a generalized inverse of $\mathbf{A}$.

(2) If $\mathbf{x}_0 = \mathbf{By}$ is a solution to $\mathbf{Ax} = \mathbf{y}$ for every $\mathbf{y}$ for which the equations are consistent then $\mathbf{B}$ is a generalized inverse of $\mathbf{A}$.

Proof: Since we have established (1) we prove (2) Let $\mathbf{y} = \mathbf{A}\boldsymbol{\ell}$ where $\boldsymbol{\ell}$ is arbitrary. The equations $\mathbf{Ax} = \mathbf{A}\boldsymbol{\ell}$ have a solution by assumption which is of the form $\mathbf{BA}\boldsymbol{\ell}$. It follows that $\mathbf{ABA} = \mathbf{A}\boldsymbol{\ell}$. Since $\boldsymbol{\ell}$ is arbitrary we have that $\mathbf{ABA} = \mathbf{A}$ i.e. $\mathbf{B}$ is a generalized inverse of $\mathbf{A}$. □

5.2.4 Generalized Inverses

We now establish that a generalized inverse of $\mathbf{A}$ always exists. Recall that we can find non singular matrices $\mathbf{P}$ and $\mathbf{Q}$ such that

$$\mathbf{PAQ} = \mathbf{B} \;\text{ i.e. }\; \mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{Q}^{-1}$$

where $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

Matrix multiplication shows that $\mathbf{B}\mathbf{B}^{-}\mathbf{B} = \mathbf{B}$ where

$$\mathbf{B}^{-} = \begin{bmatrix} \mathbf{I} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$$

where $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ are arbitrary. Since

$$\begin{aligned} \mathbf{A}(\mathbf{Q}\mathbf{B}^{-}\mathbf{P})\mathbf{A} &= \mathbf{P}^{-1}\mathbf{B}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{B}^{-}\mathbf{P}\mathbf{P}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ &= \mathbf{P}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ &= \mathbf{A} \end{aligned}$$

we see that $\mathbf{Q}\mathbf{B}^{-}\mathbf{P}$ is a generalized inverse of $\mathbf{A}$ establishing the existence of a generalized inverse for any matrix.

We also note that if

$$\mathbf{PA} = \begin{bmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

Then another generalized inverse of $\mathbf{A}$ is $\mathbf{B}^{-}\mathbf{P}$ where

$$\mathbf{B}^{-} = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

æ

Chapter 6

Determinants

6.1 Definition and Properties

Definition 6.1 Let $\mathbf{A}$ be square matrix with columns $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$. The **determinant** of $\mathbf{A}$ is that function $\det : \mathbf{R}^p \to \mathbf{R}$ such that

(1) $\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, c\mathbf{x}_m, \ldots, \mathbf{x}_p) = c \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)$

(2) $\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{x}_k, \ldots, \mathbf{x}_p = \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)$

(3) $\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p) = 1$ where $\mathbf{e}_i$ is the ith unit vector.

Properties of Determinants

(1) If $\mathbf{x}_i = \mathbf{0}$ then $\mathbf{x}_i = 0\mathbf{x}_i$ so that by (1) of the definition $\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p) = 0$ if $\mathbf{x}_i = \mathbf{0}$.

(2) $\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + c\mathbf{x}_k, \ldots, \mathbf{x}_p) = \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)$. If $c = 0$ this result is trivial. If $c \neq 0$ we note that

$$\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + c\mathbf{x}_k, \ldots, \mathbf{x}_p) \quad = \quad -\frac{1}{c} \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + c\mathbf{x}_k, \ldots, -c\mathbf{x}_k, \ldots, \mathbf{x}_p)$$

$$
\begin{aligned}
&= -\frac{1}{c}\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, -c\mathbf{x}_k, \ldots, \mathbf{x}_p) \\
&= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)
\end{aligned}
$$

(3) If two columns of $\mathbf{A}$ are identical then $\det = 0$. Similarly if the columns of $\mathbf{A}$ are linearly dependent then the determinant of $\mathbf{A}$ is 0.

(4) If $\mathbf{x}_k$ and $\mathbf{x}_m$ are interchanged then the determinant changes sign i.e.

$$\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_p) = -\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)$$

To prove this we note that

$$
\begin{aligned}
\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_p) &= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{x}_k, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_p) \\
&= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{x}_k, \ldots, \mathbf{x}_k - (\mathbf{x}_m + \mathbf{x}_k), \ldots, \mathbf{x}_p) \\
&= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{x}_k, \ldots, -\mathbf{x}_m, \ldots, \mathbf{x}_p) \\
&= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots, -\mathbf{x}_m, \ldots, \mathbf{x}_p) \\
&= -\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p)
\end{aligned}
$$

Thus the interchange of an even number of columns does not change the sign of the determinant while the interchange of an odd number of columns does change the sign.

(5) $\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{y}, \ldots, \mathbf{x}_p)$ is equal to

$$\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_p) + \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{y}, \ldots, \mathbf{x}_p)$$

If the vectors $\mathbf{x}_i$ for $i \neq m$ are dependent the assertion is obvious since both terms are 0. If the $\mathbf{x}_i$ are linearly independent for $i \neq m$ and $\mathbf{x}_m = \sum_i c_i \mathbf{x}_i$ the assertion follows by repeated use of (2) and the fact that the first term of the expression is always zero.

If the $\mathbf{x}_i$ are linearly independent then $\mathbf{y} = \sum_i d_i \mathbf{x}_i$ so that

$$
\begin{aligned}
\det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \mathbf{y}, \ldots, \mathbf{x}_p) &= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m + \sum_i d_i \mathbf{x}_i, \ldots, \mathbf{x}_p) \\
&= \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, (1 + d_m)\mathbf{x}_m, \ldots, \mathbf{x}_p)
\end{aligned}
$$

$$
\begin{aligned}
&= (1+d_m)\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m,\ldots,\mathbf{x}_p)\\
&= \det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p)+\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,d_m\mathbf{x}_m,\ldots,\mathbf{x}_p)\\
&= \det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p)+\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\sum_i d_i\mathbf{x}_i,\ldots,\mathbf{x}_p)\\
&= \det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p)+\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{y},\ldots,\mathbf{x}_p)
\end{aligned}
$$

Using the above results we can show that the determinant of a matrix is a well defined function of a matrix and give a method for computing the determinant. Since any column vector of $\mathbf{A}$, say $\mathbf{x}_m$ can be written as

$$\mathbf{x}_m=\sum_{i=1}^{p}a_{im}\mathbf{e}_i$$

we have that

$$
\begin{aligned}
\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p) &= \det(\sum_{i_1}a_{i_11}\mathbf{e}_{i_1},\mathbf{x}_2,\ldots,\mathbf{x}_p)\\
&= \sum_{i_1}a_{i_11}\det(\mathbf{e}_{i_1},\mathbf{x}_2,\ldots,\mathbf{x}_p)\\
&= \sum_{i_1}a_{i_11}\sum_{i_2}a_{i_22}\det(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\ldots,\mathbf{x}_p)\\
&= \sum_{i_1}\sum_{i_2}\cdots\sum_{i_p}a_{i_11}a_{i_22}\cdots a_{i_pp}\det(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\ldots,\mathbf{e}_{i_p})
\end{aligned}
$$

Now note that $\det(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\ldots,\mathbf{e}_p)$ is 0 if any of the subscripts are the same, $+1$ if $i_1,i_2,\ldots,i_p$ is an even permutation of $1,2,\ldots,p$ and -1 if $i_1,i_2,\ldots,i_p$ is an odd permutation of $1,2,\ldots,p$. Hence we can evaluate the determinant of a matrix and it is a uniquely defined function satisfying the conditions (1), (2) and (3).

6.2 Computation

The usual method for computing a determinant is to note that if $\mathbf{x}_m=\sum_i a_{im}\mathbf{x}_i$ then

$$
\begin{aligned}
\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_p) &= \sum_i a_{im}\det(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{e}_m,\ldots,\mathbf{x}_p)\\
&= \sum_i a_{im}\mathbf{A}_{im}
\end{aligned}
$$

where

$$\mathbf{A}_{im} = \det(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{e}_m, \ldots, \mathbf{x}_p)$$

is called the **cofactor** of a_{im} and is simply the determinant of $\mathbf{A}$ when the mth column of $\mathbf{A}$ is replaced by $\mathbf{e}_i$.

We now note that

$$\mathbf{A}_{im} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_p} a_{i_1} a_{i_2} \cdots a_{i_p} \det(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_{i_p})$$

Hence

$$\mathbf{A}_{im} = (-1)^{i+m} \det(\mathbf{M}_{im})$$

where $\mathbf{M}_{im}$ is the matrix formed by deleting the ith row of $\mathbf{A}$ and the jth column of $\mathbf{A}$. Thus the value of the determinant of a p by p matrix can be reduced to the calculation of p determinants of $p-1$ by $p-1$ matrices. The determinant of $\mathbf{M}_{im}$ is called the **minor** of a_{im}.

æ

Chapter 7

Inner Product Spaces

7.1 definitions and Elementary Properties

Definition 7.1 An **inner product** on a linear vector space $\mathcal{V}$ is a function $(\,,\,)$ mapping $\mathcal{V} \times \mathcal{V}$ into $\mathbf{C}$ such that

1. $(\mathbf{x}, \mathbf{x}) \geq 0$ with $(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{x}, \mathbf{y})}$
3. $(\alpha\mathbf{x} + \mathbf{y}, vbfz) = \alpha(\mathbf{x}, \mathbf{z}) + \beta(\mathbf{y}, \mathbf{z})$

A vector space with an inner product is called an **inner product space**.

Definition 7.2: The **norm** or **length** of $\mathbf{x}$, $||\mathbf{x}||$ is defined by

$$||\mathbf{x}||^2 = (\mathbf{x}, \mathbf{x})$$

Definition 7.3: The vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** if $(\mathbf{x}, \mathbf{x}) = 0$. We write $\mathbf{x} \perp \mathbf{y}$ if $\mathbf{x}$ is orthogonal to $\mathbf{y}$. More generally

(1) $\mathbf{x}$ is said to be orthogonal to the set of vectors $\mathcal{X}$ if

$$(\mathbf{x}, \mathbf{y}) = 0 \;\text{ for every }\; \mathbf{y} \in \mathcal{X}$$

(2) $\mathcal{X}$ and $\mathcal{Y}$ are said to be orthogonal if

$$(\mathbf{x}, \mathbf{y}) = 0 \text{ for every } \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{y} \in \mathcal{Y}$$

Definition 7.4: $\mathcal{X}$ is an **orthonormal** set of vectors if

$$(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{y} \\ 0 & \text{if } \mathbf{x} \neq \mathbf{y} \end{cases}$$

$\mathbf{X}$ is a **complete orthonormal** set of vectors if it is not contained in a larger set of orthonormal vectors.

Lemma 7.1 If $\mathcal{X}$ is an orthonormal set then its vectors are linearly independent.

Proof: If $\sum_i \alpha_i \mathbf{x} = \mathbf{0}$ for $\mathbf{x}_i \in \mathcal{X}$ then

$$\mathbf{0} = \left(\sum_i \alpha_i \mathbf{x}_i, \mathbf{x}_j \right) = \sum_i \alpha_i (\mathbf{x}_i, \mathbf{x}_j) = \alpha_j$$

It follws that each of the α_j are equal to 0 and hence that the $\mathbf{x}_j$s are linearly independent. $\square$

Lemma 8.2 Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be an orthonormal basis for $\mathcal{X}$. Then $\mathbf{x} \perp \mathcal{X}$ if and only if $\mathbf{x} \perp \mathbf{x}_i$ for $i = 1, 2, \ldots, p$.

Proof: If $\mathbf{x} \perp \mathcal{X}$ then by definition $\mathbf{x} \perp \mathbf{x}_i$ for $i = 1, 2, \ldots, p$.

Conversely, if $\mathbf{x} \perp \mathbf{x}_i$ for $i = 1, 2, \ldots, p$. Then if $\mathbf{y} \in \mathcal{X}$ we have $\mathbf{y} = \sum_i \alpha_i \mathbf{x}_i$. Thus

$$(\mathbf{x}, \mathbf{y}) = \sum_i \bar{\alpha}_i (\mathbf{x}, \mathbf{x}_i) = 0$$

Thus $\mathbf{x} \perp \mathcal{X}$. $\square$

Theorem 7.3 (Bessel's Inequality) Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be an orthonormal set in a linear vector space $\mathcal{V}$ with an inner product. Then

1. $\sum_i |\alpha_i|^2 \leq ||\mathbf{x}||^2$ where $\mathbf{x}$ is any vector in $\mathcal{V}$ and $\alpha_i = (\mathbf{x}, \mathbf{x}_i)$.
2. The vector $\mathbf{r} = \mathbf{x} - \sum_i \alpha_i \mathbf{x}$ is orthogonal to each $\mathbf{x}_j$ and hence to the space spanned by $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$.

Proof: We note that

$$
\begin{aligned}
||\mathbf{x} - \sum_i \alpha_i \mathbf{x}_i||^2 &= \left(\mathbf{x} - \sum_i \alpha_i \mathbf{x}_i, \mathbf{x} - \sum_i \alpha_i \mathbf{x}_i\right) \\
&= (\mathbf{x}, \mathbf{x}) - \sum_i (\alpha_i (\mathbf{x}_i, \mathbf{x}) - (\mathbf{x}, \sum_i \alpha_i \mathbf{x}_i) + (\sum_i \alpha_i \mathbf{x}_i, \sum_i \alpha_i \mathbf{x}_i) \\
&= ||\mathbf{x}||^2 - \sum_i \alpha_i \bar{\alpha}_i - \sum_i \bar{\alpha}_i (\mathbf{x}, \mathbf{x}_i) + \sum_i \alpha_i (\mathbf{x}_i, \sum_i \alpha_i \mathbf{x}_i) \\
&= ||\mathbf{x}||^2 - 2 \sum_i |\alpha_i|^2 + \sum_i \alpha_i (\mathbf{x}_i, \alpha \mathbf{x}_i) \\
&= ||\mathbf{x}||^2 - \sum_i |\alpha_i|^2
\end{aligned}
$$

Since $||\mathbf{x} - \sum_i \alpha_i \mathbf{x}_i||^2 \geq 0$ result (1) follows.

For result (2) we note that

$$\left(\mathbf{x} - \sum_i \alpha_i \mathbf{x}_i, \mathbf{x}_j\right) = (\mathbf{x}, \mathbf{x}_j) - \sum_i \alpha_i (\mathbf{x} - i, \mathbf{x}_j) = (\mathbf{x}, \mathbf{x}_j) - (\mathbf{x}, \mathbf{x}_j) = 0 \quad \square$$

Theorem 7.4 (Cauchy Schwartz Inequality) If $\mathbf{x}$ and $\mathbf{y}$ are vectors in an inner product space then

$$|(\mathbf{x}, \mathbf{y})| \leq ||\mathbf{x}|| \, ||\mathbf{y}||$$

Proof: If $\mathbf{y} = \mathbf{0}$ equality holds. If $\mathbf{y} \neq \mathbf{0}$ then $\frac{\mathbf{y}}{||\mathbf{y}||}$ is orthonormal and by Bessel's inequality we have

$$\left|\left(\mathbf{x}, \frac{\mathbf{y}}{||\mathbf{y}||}\right)\right|^2 \leq ||\mathbf{x}||^2 \text{ or } |(\mathbf{x}, \mathbf{y})| \leq ||\mathbf{x}|| \, ||\mathbf{y}|| \quad \square$$

Theorem 7.5 If $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is any finite orthonormal set in a vector space $\mathcal{V}$ then the following conditions are equivalent:

(1) $\mathcal{X}$ is complete.

(2) $(\mathbf{x}, \mathbf{x}_i) = \mathbf{0}$ for $i = 1, 2, \ldots, n$ implies that $\mathbf{x} = \mathbf{0}$.

(3) The space spanned by $\mathcal{X}$ is equal to $\mathcal{V}$.

(4) If $\mathbf{x} \in \mathcal{V}$ then $\mathbf{x} = \sum_i (\mathbf{x}, \mathbf{x}_i)\mathbf{x}_i$.

(5) If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathcal{V}$ then

$$(\mathbf{x}, \mathbf{y}) = \sum_i (\mathbf{x}, \mathbf{x}_i)(\mathbf{x}_i, \mathbf{y})$$

(6) If $\mathbf{x} \in \mathcal{V}$ then

$$||\mathbf{x}||^2 = \sum_i |(\mathbf{x}_i, \mathbf{x})|^2$$

Result (5) is called Parseval's identity.

Proof:
(1) $\Longrightarrow$ (2) If $\mathcal{X}$ is complete and $(\mathbf{x}, \mathbf{x}_i) = 0$ for each i then $\frac{\mathbf{x}}{||\mathbf{x}||}$ could be "added" to the set $\mathcal{X}$ which would be a contradiction. Thus $\mathbf{x} = \mathbf{0}$.

(2) $\Longrightarrow$ (3) If $\mathbf{x}$ is not a linear combination of the $\mathbf{x}_i$ then $\mathbf{x} - \sum_i (\mathbf{x}, \mathbf{x}_i)\mathbf{x}_i$ is not equal to $\mathbf{x}$ and is orthogonal to $\mathcal{X}$ which is a contradiction. Thus the sapace spanned by $\mathcal{X}$ is equal to $\mathcal{V}$¿

(3) $\Longrightarrow$ (4) If $\mathbf{x} \in \mathcal{V}$ we can write $\mathbf{x} = \sum_j \alpha_j \mathbf{x}_j$. It follows that

$$(\mathbf{x}, \mathbf{x}_i) = \sum_j \alpha_j (\mathbf{x}_j, \mathbf{x}_i) = \alpha_i$$

(4) $\Longrightarrow$ (5) Since

$$\mathbf{x} = \sum_i (\mathbf{x}, \mathbf{x}_i)\mathbf{x}_i \;\text{ and }\; \mathbf{y} = \sum_j (\mathbf{y}, \mathbf{x}_j)\mathbf{x}_j$$

we have

$$\begin{aligned}
(\mathbf{x}, \mathbf{y}) &= (\sum_i (\mathbf{x}, \mathbf{x}_i)\mathbf{x}_i, \sum_j (\mathbf{y}, \mathbf{x}_j)\mathbf{x}_j) \\
&= \sum_i (\mathbf{x}, \mathbf{x}_i)\overline{(\mathbf{y}, \mathbf{x}_i)} \\
&= \sum_i (\mathbf{x}, \mathbf{x}_i)(\mathbf{x}_i, \mathbf{y})
\end{aligned}$$

(5) $\Longrightarrow$ (6) In result (5) simply set $\mathbf{x} = \mathbf{y}$.

(6) $\Longrightarrow$ (1) If $\mathbf{x}_o$ is orthogonal to each $\mathbf{x}_i$ then by (6) we have

$$||\mathbf{x}_o||^2 = \sum_i |(\mathbf{x}_i, \mathbf{x}_o)|^2 = 0$$

so that $\mathbf{x}_o = \mathbf{0}$. $\square$

7.2 Gram Schmidt Process

The Gram Schmidt process can be used to construct an orthonormal basis for a finite dimensional vector space. Start with a basis for $\mathcal{V}$ as $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Form

$$\mathbf{y}_1 = \frac{\mathbf{x}_1}{||\mathbf{x}_1||}$$

Next define

$$\mathbf{z}_2 = \mathbf{x}_2 - (\mathbf{x}_2, \mathbf{y}_1)\mathbf{y}_1$$

Since the $\mathbf{x}_i$ are linearly independent $\mathbf{z}_2 \neq \mathbf{0}$ and $\mathbf{z}_2$ is orthogonal to $\mathbf{y}_1$. Hence $\mathbf{y}_2 = \frac{\mathbf{z}_2}{||\mathbf{z}_2||}$ is orthogonal to $\mathbf{y}_1$ and has unit norm. If $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r$ have been so chosen then we form

$$z_{r+1} = \mathbf{x}_{r+1} - \sum_{i=1}^{r} (\mathbf{x}_{r+1}, \mathbf{y}_i)\mathbf{y}_i$$

Since the $\mathbf{x}_i$ are linearly independent $||\mathbf{z}_{r+1}|| > 0$ and since $\mathbf{z}_{r+1} \perp \mathbf{y}_i$ for $i = 1, 2, \ldots, r$ it follows that

$$\mathbf{y}_{r+1} = \frac{\mathbf{z}_{r+1}}{||\mathbf{z}_{r+1}||}$$

may be "added" to the set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r\}$ to form a new orthonormal set. The process necessarily stops with $\mathbf{y}_n$ since there can be at most n elements in a linearly independent set. $\square$

7.3 Orthogonal Projections

Theorem 7.6 (Orthogonal Projection) Let $\mathcal{U}$ be a subspace of an inner product space $\mathcal{V}$ and let $\mathbf{y}$ be a vector in $\mathcal{V}$ which is not in $\mathcal{U}$. Then there exists a unique vector $\mathbf{y}_{\mathcal{U}} \in \mathcal{U}$ and and a unique vector $\mathbf{e} \in \mathcal{V}$ such that

$$\mathbf{y} = \mathbf{y}_{\mathcal{U}} + \mathbf{e} \ \text{ and } \ \mathbf{e} \perp \mathcal{U}$$

Proof: Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ be a basis of $\mathcal{U}$. Since $\mathbf{y} \notin \mathcal{U}$ the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r, \mathbf{y}\}$ is a linearly independent set. Use the Gram Schmidt process on this set to form a set of orthonormal vectors $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r, \mathbf{y}_{r+1}\}$. Since $\mathbf{y}$ is in the space spanned by these vectors we can write

$$\mathbf{y} = \sum_{i=1}^{r+1} \alpha_i \mathbf{y}_r = \sum_{i=1}^{r} \alpha_i \mathbf{y}_r + \alpha_{r+1} \mathbf{y}_{r+1}$$

If we define

$$\mathbf{y}_{\mathcal{U}} = \sum_{i=1}^{r} \alpha_i \mathbf{y}_i \ \text{ and } \ \mathbf{e} = \alpha_{r+1} \mathbf{y}_{r+1}$$

then $\mathbf{Y}_{\mathcal{U}} \in \mathcal{U}$ and $\mathbf{e} \perp \mathcal{U}$. To show uniqueness, suppose that we also have

$$\mathbf{y} = \mathbf{z}_1 + \mathbf{e}_1 \ \text{ where } \ \mathbf{z}_1 \in \mathcal{U} \ \text{ and } \ \mathbf{e}_1 \perp \mathcal{U}$$

Then we have

$$(\mathbf{z}_1 - \mathbf{y}_{\mathcal{U}}) + (\mathbf{e}_1 - \mathbf{e}) = \mathbf{0}$$

so that $(\mathbf{z}_1 - \mathbf{y}_{\mathcal{U}}, \mathbf{z}_1 - \mathbf{y}_{\mathcal{U}}) = 0$ and hence $\mathbf{z}_1 = \mathbf{y}_{\mathcal{U}}$ and $\mathbf{e}_1 = \mathbf{e}$. $\square$

Definition 7.4 The vector $\mathbf{e}$ in Theorem 7.6 is called the **orthogonal projection** from $\mathbf{y}$ to $\mathcal{U}$ and the vector $\mathbf{y}_{\mathcal{U}}$ is called the **orthogonal projection** of $\mathbf{y}$ on $\mathcal{U}$.

Theorem 7.7 The projection of $\mathbf{y}$ on $\mathcal{U}$ has the property that

$$||\mathbf{y} - \mathbf{y}_{\mathcal{U}}|| = \overset{\min}{\mathbf{x}} \{||\mathbf{y} - \mathbf{x}|| : \ \mathbf{x} \in \mathcal{U}\}$$

Proof: Let $\mathbf{x} \in \mathcal{U}$ Then $(\mathbf{x} - \mathbf{y}_{\mathcal{U}}, \mathbf{e}) = 0$. Hence

$$||\mathbf{y} - \mathbf{x}|| \ = \ (\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x})$$

$$
\begin{aligned}
&= (\mathbf{y} - \mathbf{y}_{\mathcal{U}} + \mathbf{y}_{\mathcal{U}} - \mathbf{x}, \mathbf{y} - \mathbf{y}_{\mathcal{U}} + \mathbf{y}_{\mathcal{U}} - \mathbf{x}) \\
&= ||\mathbf{y} - \mathbf{y}_{\mathcal{U}}|| + ||\mathbf{y}_{\mathcal{U}} - \mathbf{x}|| + (\mathbf{y} - \mathbf{y}_{\mathcal{U}}, \mathbf{y}_{\mathcal{U}} - \mathbf{x}) + (\mathbf{y}_{\mathcal{U}} - \mathbf{x}, \mathbf{y} - \mathbf{y}_{\mathcal{U}}) \\
&= ||\mathbf{y} - \mathbf{y}_{\mathcal{U}}|| + ||\mathbf{y}_{\mathcal{U}} - \mathbf{x}|| \\
&\geq ||\mathbf{y} - \mathbf{y}_{\mathcal{U}}||
\end{aligned}
$$

with the minimum occurring when $\mathbf{x} = \mathbf{y}_{\mathcal{U}}$. Thus the projection minimizes the distance from $\mathcal{U}$ to $\mathbf{x}$. $\Box$ æ

Chapter 8

Characteristic Roots and Vectors

8.1 Definitions

Definition 8.1: A scalar λ is a **characteristic root** and a non-zero vector $\mathbf{x}$ is a **characteristic vector** of the matrix $\mathbf{A}$ if

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

Other names for charateristic roots are proper value, latent root, eigenvalue and secular value with similar adjectives appying to characteristic vectors.

If λ is a characteristic root of $\mathbf{A}$ let the set $\mathcal{X}_\lambda$ be defined by

$$\mathcal{X}_\lambda = \{\mathbf{x} \,:\, \mathbf{Ax} = \lambda\mathbf{x}\}$$

The **geometric multiplicity** of λ is defined to be the dimension of $\mathcal{X}_\lambda \cup \{\mathbf{0}\}$. λ is said to be a simple characteristic root if its geometric multiplicity is 1.

Definition 8.2: The **spectrum** of $\mathbf{A}$ is the set of all characteristic roots of $\mathbf{A}$ and is denoted by $\Lambda(\mathbf{A})$

Since $\mathbf{Ax} = \lambda\mathbf{x}$ if and only if $[\mathbf{A} - \lambda]\,\mathbf{x} = \mathbf{0}$ we see that $\lambda \in \Lambda(\mathbf{A})$ is equivalent to the set of λ such that $\mathbf{A} - \lambda\mathbf{I}$. Similarly $\mathbf{x}$ is a characteristic vector corresponding to λ

if and only $\mathbf{x}$ is in the null space of $\mathbf{A} - \lambda\mathbf{I}$. Using the properties of determinants we have the following Theorem.

Theorem 8.1 $\lambda \in \Lambda(\mathbf{A})$ if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

If the field of scalars is the set of complex numbers (or any algebraically closed field) then if $\mathbf{A}$ is p by p we have

$$\det(\mathbf{A} - \lambda\mathbf{A}) = \prod_{i=1}^{q}(\lambda - \lambda_i)^{n_i}$$

where

$$\sum_{i=1}^{q} = p \text{ and } n_i \geq 1$$

The **algebraic multiplicity** of the characteristic root λ_i is the number n_i which appears in the above expression.

The algebraic and geometric multiplicity of a characteristic root need not correspond. To see this let

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

Then

$$\det\left(\mathbf{A} - \lambda\mathbf{I}\right) = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (\lambda - 1)^2$$

It follows that both of the characteristic roots of $\mathbf{A}$ are equal to 1. Hence the algebraic multiplicity of the characteristic root 1 is equal to 2. However, the equation

$$\mathbf{Ax} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

yields the equations

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

Thus the geometric multiplicity of $\mathbf{A}$ is equal to 1.

Lemma 8.2

(1) The spectrum of $\mathbf{TAT}^{-1}$ is the same as the spectrum of $\mathbf{A}$.

(2) If λ is a characteristic root of $\mathbf{A}$ then $p(\lambda)$ is a characteristic root of $p(\mathbf{A})$ where p is any polynomial.

Proof: Since $\mathbf{TAT}^{-1} - \lambda\mathbf{I} = \mathbf{T}[\mathbf{A} - \lambda\mathbf{I}]\mathbf{T}^{-1}$ it follows that $[\mathbf{A} - \lambda\mathbf{I}]\mathbf{x} = \mathbf{0}$ if and only if $[\mathbf{TAT}^{-1} - \lambda\mathbf{I}]\mathbf{Tx} = \mathbf{0}$ and the spectrums are thus the same.

If $\mathbf{Ax} = \lambda\mathbf{x}$ then $\mathbf{A}^n\mathbf{x} = \lambda^n\mathbf{x}$ for any integer n so that $p(\mathbf{Ax} = p(\lambda)\mathbf{x}$. □

8.2 Factorizations

Theorem 8.3 (Schur'sTriangular Factorization) Let $\mathbf{A}$ be any p by p matrix. Then there exists a unitary matrix $\mathbf{U}$ i.e. $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ such that

$$\mathbf{U}^*\mathbf{AU} = \mathbf{T}$$

where $\mathbf{T}$ is an upper right triangular matrix with diagonal elements equal to the characteristic roots of $\mathbf{A}$

Proof: Let $p = 2$ and let $\mathbf{u}_1$ be such that

$$\mathbf{Au}_1 = \lambda_1\mathbf{u}_1 \;\text{ and }\; \mathbf{u}_1\mathbf{u}_1 = 0$$

Define the matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$ where $\mathbf{u}_2$ is chosen so that $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. Then

$$\begin{aligned} \mathbf{U}^*\mathbf{AU} &= [\mathbf{u}_1, \mathbf{u}_2]^*\mathbf{A}[\mathbf{u}_1, \mathbf{u}] \\ &= \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \end{bmatrix} [\lambda\mathbf{u}_1, \mathbf{Au}_2] \\ &= \begin{bmatrix} \lambda & \mathbf{u}_1^*\mathbf{Au}_2 \\ 0 & \mathbf{u}_2^*\mathbf{Au}_2 \end{bmatrix} \end{aligned}$$

Note that $\mathbf{u}_2^*\mathbf{Au}_2 = \lambda_2$ is the other characteristic root of $\mathbf{A}$. Now assume that the result is true for $n = N$ and let $n = N + 1$. Let $\mathbf{Au}_1 = \lambda_1\mathbf{u}_1$ and $\mathbf{u}_1^*\mathbf{u}_1 = 1$ and let $\{vbfb_1, \mathbf{b}_2, \ldots, \mathbf{b}_N\}$ be such that

$$\mathbf{U}_1 = [\mathbf{u}_1, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N]$$

is unitary. Thus

$$\begin{aligned}
\mathbf{U}_1^*\mathbf{A}\mathbf{U}_1 &= [\mathbf{u}_1, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N]^{\mathbf{A}}[\mathbf{u}_1, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N] \\
&= [\mathbf{u}_1, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N]^*[\lambda_1\mathbf{u}_1, \mathbf{A}\mathbf{b}_1, \mathbf{A}\mathbf{b}_2, \ldots, \mathbf{A}\mathbf{b}_N] \\
&= \begin{bmatrix} \lambda_1 & \mathbf{u}_1^*\mathbf{A}\mathbf{u}_1 & \cdots & \mathbf{u}_1^*\mathbf{A}\mathbf{u}_N \\ \mathbf{0} & & \mathbf{B}_N & \end{bmatrix}
\end{aligned}$$

Since the characteristic roots of the above matrix are the solutions to the equation

$$(\lambda_1 - \lambda)\det(\mathbf{B}_N - \lambda\mathbf{I}) = 0$$

it follows that the characteristic roots of $\mathbf{B}_N$ are the remaining characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_{N+1}$ of $\mathbf{A}$. By the induction assumption there exists a unitary matrix $\mathbf{L}$ such that

$$\mathbf{L}^*\mathbf{B}_N\mathbf{L}$$

is upper right triangular with diagonal elements equal to $\lambda_1, \lambda_2, \ldots, \lambda_{N+1}$. Let

$$\mathbf{U}_{N+1} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{L} \end{bmatrix} \text{ and } \mathbf{U} = \mathbf{U}_1\mathbf{U}_{N+1}$$

Then

$$\begin{aligned}
\mathbf{U}^*\mathbf{A}\mathbf{U} &= \mathbf{U}_{N+1}^*\mathbf{U}_1\mathbf{A}\mathbf{U}_1\mathbf{U}_{N+1} \\
&= \mathbf{U}_{N+1}^*\begin{bmatrix} \lambda_1 & \mathbf{t}^* \\ \mathbf{0} & \mathbf{B}_N \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{L}^* \end{bmatrix}\begin{bmatrix} \lambda_1 & \mathbf{b}^* \\ \mathbf{0} & \mathbf{B}_N \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0}\mathbf{L} & \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & \mathbf{b}^* \\ \mathbf{0} & \mathbf{L}^*\mathbf{B}_N\mathbf{L} \end{bmatrix}
\end{aligned}$$

which is upper right triangular as claimed with the diagonal elements equal to the characteristic roots of $\mathbf{A}$. □

Theorem 8.4 If $\mathbf{A}$ is normal ($\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$) then there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{D}$$

where $\mathbf{D}$ is diagonal with elements equal to the characteristic roots of $\mathbf{A}$.

Proof: By Theorem 8.3 there is a unitary matrix $\mathbf{U}$ such that $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$ where $\mathbf{T}$ is upper right triangular. Thus we als have $\mathbf{U}^*\mathbf{A}^*\mathbf{U} = \mathbf{T}^*$ where $\mathbf{T}^*$ is lower right triangular. It follows that

$$\mathbf{U}^*\mathbf{A}\mathbf{A}^*\mathbf{U} = \mathbf{T}\mathbf{T}^* \text{ and } \mathbf{U}^*\mathbf{A}^*\mathbf{A}\mathbf{U} = \mathbf{T}^*\mathbf{T}$$

Hence the off diagonal elements of $\mathbf{T}$ vanish i.e. $\mathbf{T}$ is diagonal as claimed. $\square$

Theorem 8.5 (Cochran's Theorem) If $\mathbf{A}$ is normal then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{E}_i$$

where $\mathbf{E}_i^*\mathbf{E}_j = \mathbf{O}$ for $i \neq j$, $\mathbf{E}_i^2 = \mathbf{E}_i$ and the λ_i are the characteristic roots of $\mathbf{A}$.

Proof: By Theorem 8.3 there is a unitary matrix $\mathbf{U}$ such that $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{D}$ is diagonal with diagonal elements equal to the characteristic roots of $\mathbf{A}$. Hence

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{D}\mathbf{U}^* \\ &= \mathbf{U}\begin{bmatrix} \lambda_1 \mathbf{u}_1^* \\ \lambda_2 \mathbf{u}_2^* \\ \vdots \\ \lambda_n \mathbf{u}_n^* \end{bmatrix} \\ &= \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^* \end{aligned}$$

Defining $\mathbf{E}_i = \mathbf{u}_i\mathbf{u}_i^*$ completes the proof. $\square$

This representation of $\mathbf{A}$ is called the **spectral representation** of $\mathbf{A}$. In most applications $\mathbf{A}$ is symmetric.

The spectral representation can be rewritten as

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{E}_i = \sum_{j=1}^{q} \lambda_j \mathbf{E}_j$$

where $\lambda_1 < \lambda_2 < \cdots < \lambda_q$ are the distinct characteristic roots of $\mathbf{A}$. Now define

$$\mathbf{S}_0 = \mathbf{O} \text{ and } \mathbf{S}_i = \sum_{j=1}^{i} \mathbf{E}_j \text{ for } i = 1, 2, \ldots, q$$

Then $\mathbf{E}_i = \mathbf{S_i} - \mathbf{S_{i-1}}$ for $i = 1, 2, \ldots, q$ and hence

$$\begin{aligned}
\mathbf{A} &= \sum_{j=1}^{q} \lambda_j \mathbf{E}_j \\
&= \sum_{j=1}^{q} \lambda_j (\mathbf{S}_j - \mathbf{S}_{j-1}) \\
&= \sum_{j=1}^{q} \lambda_j \, \Delta \mathbf{S}_{\lambda_j}
\end{aligned}$$

which may be written as a Stielges integral i.e.

$$\mathbf{A} = \int \lambda d\mathbf{S}(\lambda)$$

which is also called the spectral representation of $\mathbf{A}$.

The spectral representation of $\mathbf{A}$ has the property that

$$\begin{aligned}
(\mathbf{Ax}, \mathbf{y}) &= (\int \lambda d\mathbf{S}(\lambda)\mathbf{x}, \mathbf{y}) \\
&= \left(\sum_{j=1}^{q} \lambda_j \, \Delta \mathbf{S}_{\lambda_j} \mathbf{x}, \mathbf{y} \right) \\
&= \sum_{j=1}^{q} \lambda_j \, (\Delta \mathbf{S}_{\lambda_j} \mathbf{x}, \mathbf{y}) \\
&= \sum_{j=1}^{q} \lambda_j [(\mathbf{S}_j \mathbf{x}, \mathbf{y}) - (\mathbf{S}_{j-1} \mathbf{x}, \mathbf{y})] \\
&= \int \lambda \, d(\mathbf{S}(\lambda)\mathbf{x}, \mathbf{y})
\end{aligned}$$

In more advanced treatments applicable to Hilbert spaces the property just established is taken as the defining property for $\int \lambda \, d(\mathbf{S}(\lambda)$.

8.3 Quadratic Forms

If $\mathbf{A}$ is a symmetric matrix with real elements $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is called a **quadratic form**. A quadratic form is said to be

- **positive definite** if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
- **non negative definite** if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$

If $\mathbf{A}$ is positve definite then all of its characteristic roots are positive while if $\mathbf{A}$ is non negative definite then all of its characteristic roots are non negative. To prove this note that if $\mathbf{x}_i$ is a characteristic vector of $\mathbf{A}$ corresponding to λ_i then $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$ and hence $\mathbf{x}_i^T\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i^T\mathbf{x}_i$.

If $\lambda_i \neq \lambda_j$ then the corresponding characteristic vectors are orthogonal since

$$\lambda_j\mathbf{x}_i^T\mathbf{x}_j = \mathbf{x}_i^T\mathbf{A}\mathbf{x}_j = (\mathbf{A}\mathbf{x}_i)\mathbf{x}_j = \lambda_i\mathbf{x}_i^T\mathbf{x}_j$$

which implies that $\mathbf{x}_i^T\mathbf{x}_j = 0$ i.e. that $\mathbf{x}_i$ and $\mathbf{x}_j$ are orthogonal if $\lambda_i \neq \lambda_j$.

Another way of characterizing characteristic roots is to consider the stationary values of

$$\frac{\mathbf{X}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

as $\mathbf{x}$ varies over $\mathbf{R}^p$. Since

$$\mathbf{A} = \sum_{i=1}^{p} \lambda_i\mathbf{E}_i$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ we have that

$$\mathbf{A}\mathbf{x} = \sum_{i=1}^{p} \lambda_i\mathbf{E}_i\mathbf{x}$$

It follows that

$$\mathbf{X}^T\mathbf{A}\mathbf{x} \quad = \quad \sum_{i=1}^{p} \lambda_i\mathbf{x}^T\mathbf{E}_i\mathbf{x}$$

$$
\begin{aligned}
&= \sum_{i=1}^{p} \lambda_i \mathbf{x}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{x} \\
&= \sum_{i=1}^{p} \lambda_i \|\mathbf{E}_i \mathbf{x}\|^2
\end{aligned}
$$

Since $\mathbf{E}_i = \mathbf{y}_i \mathbf{y}_i^T$, where the $\mathbf{y}_i$ form an orthonormal basis we have that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ where

$$
\mathbf{U} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p]
$$

It follows that $\mathbf{U}^{-1} = \mathbf{U}^T$ and hence that $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ Thus

$$
\begin{aligned}
\mathbf{I} &= \mathbf{U}\mathbf{U}^T \\
&= [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p] \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_p \end{bmatrix} \\
&= \sum_{i=1}^{p} \mathbf{y}_i \mathbf{y}_i^p \\
&= \sum_{i=1}^{p} \mathbf{E}_i
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbf{x}^T\mathbf{x} &= \sum_{i=1}^{p} \mathbf{x}^T \mathbf{E}_i \mathbf{x} \\
&= \sum_{i=1}^{p} \mathbf{x}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{x} \\
&= \sum_{i=1}^{p} \|\mathbf{E}_i \mathbf{x}\|^2
\end{aligned}
$$

Thus

$$
\frac{\mathbf{X}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\sum_{i=1}^{p} \lambda_i \|\mathbf{E}_i \mathbf{x}\|^2}{\sum_{i=1}^{p} \|\mathbf{E}_i \mathbf{x}\|^2}
$$

If $\mathbf{x}$ is orthogonal to $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_k$ then

$$\frac{\mathbf{X}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\sum_{i=k+1}^{p} \lambda_i \|\mathbf{E}_i\mathbf{x}\|^2}{\sum_{i=k+1}^{p} \|\mathbf{E}_i\mathbf{x}\|^2}$$

and it follows that maximum of

$$\frac{\mathbf{X}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

subject to $\mathbf{E}_i\mathbf{x} = \mathbf{0}$ for $i = 1, 2, \ldots, k$ is λ_{k+1}.

The **norm** of a matrix $\mathbf{A}$ is defined as

$$\|\mathbf{A}\| = \{\mathbf{x} : \overset{\textbf{max}}{\|\mathbf{x}\|} \neq \mathbf{0}\} \frac{\|\mathbf{A}\|}{\|\mathbf{x}\|}$$

If $\mathbf{P}$ and $\mathbf{Q}$ are orthogonal matrices of order $n \times n$ and $p \times p$ respectively then

$$\|\mathbf{PAQ}\| = \|\mathbf{A}\|$$

Lemma 8.6 The characterisitic roots of a Hermitian matrix are real and the characteristic roots of a real symmetric matrix are real.

Proof: If λ is a characteristic root of $\mathbf{A}$ then

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \text{ and } \mathbf{x}^*\mathbf{A}^* = \bar{\lambda}\mathbf{x}^*$$

Since $\mathbf{A}$ is Hermitian it follows that

$$\mathbf{x}^*\mathbf{A}^* = \mathbf{x}^*\mathbf{A} = \bar{\lambda}\mathbf{x}^*$$

and hence

$$\lambda\mathbf{x}^*\mathbf{x} = \bar{\lambda}\mathbf{x}^*\mathbf{x}$$

or $\lambda = \bar{\lambda}$. If $\mathbf{A}$ is a real symmetric matrix it is Hemitian so that the result for real symmetric matrices is true.

Lemma 8.7 A necessary and sufficient condition that $\mathbf{A} = \mathbf{0}$ is that $\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = 0$ for all $\mathbf{x}$ and $\mathbf{y}$.

Theorem 8.8 A necessary and sufficient condition that a symmetric matrix on a real inner product space satisfies $\mathbf{A} = \mathbf{0}$ is that $\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = 0$.

Proof: Necessity is obvious.

To prove sufficiency we note that

$$\begin{aligned} \langle \mathbf{A}(\mathbf{x}+\mathbf{y}), (\mathbf{x}+\mathbf{y}) \rangle &= \langle (\mathbf{A}\mathbf{x}+\mathbf{A}\mathbf{y}), (\mathbf{x}+\mathbf{y}) \rangle \\ . &= \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{A}\mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle \end{aligned}$$

It follows that

$$\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{A}\mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{A}(\mathbf{x}+\mathbf{y}), (\mathbf{x}+\mathbf{y}) \rangle - \langle \mathbf{A}(\mathbf{x}, (\mathbf{x} \rangle - \langle \mathbf{A}(\mathbf{y}, (\mathbf{y} \rangle$$

Thus the real part of $\langle \mathbf{A}\mathbf{x} \langle = 0$ since $\mathbf{A}$ is symmetric. Thus $\langle \mathbf{A}\mathbf{x} \langle = 0$ and by Lemma 8.2 the result follows.

8.4 Jordan Canonical Form

The most general canonical form for matrices is the Jordan Canonical Form. This particular canonical form is used in the theory of Markov chains.

The Jordan canonical form is a nearly diagonal canonical form for a matrix. Let $\mathbf{J}_{p_i}(\lambda_i)$ be $p_i \times p_i$ matrix of the form

$$\mathbf{J}_{p_i}(\lambda_i) = \lambda_i \mathbf{I}_{p_i} + \mathbf{N}_{p_i}$$

where

$$\mathbf{N}_{p_i} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

is a nilpotent matrix of order p_i i.e. $\mathbf{N}_{p_i}^{p_i} = \mathbf{0}$.

Theorem 8.9 (Jordan Canonical Form). Let $\lambda_1, \lambda_2, \ldots, \lambda_q$ be the q distinct characteristic roots of a $p \times p$ matrix $\mathbf{A}$. Then there exists a non singular matrix $\mathbf{V}$ such that

$$\mathbf{V}\mathbf{A}\mathbf{V}^{-1} = \text{diag}\left(\mathbf{J}_{p_1}, \mathbf{J}_{p_2}, \ldots, \mathbf{J}_{p_q}\right)$$

One use of the Jordan canonical form occurs when we want an expression for $\mathbf{A}^n$. Using the Jordan canonical form we see that

$$\mathbf{V}\mathbf{A}^n\mathbf{V}^{-1} = \text{diag}\left(\mathbf{J}_{p_1}^n, \mathbf{J}_{p_2}^n, \ldots, \mathbf{J}_{p_q}^n\right)$$

Thus if $n \geq \max p_i$ we have

$$\begin{aligned} \mathbf{J}_{p_i}^n(\lambda_i) &= (\lambda_i \mathbf{I}_{p_i} + \mathbf{N}_{p_i})^n \\ &= \sum r = 0^n \binom{n}{r} \lambda_i^{n-r} \mathbf{N}_{p_i}^r \end{aligned}$$

by the binomial expansion. Since $\mathbf{N}_{p_i}^r = \mathbf{0}$ if $r \geq p_i$ we obtain

$$\mathbf{J}_{p_i}^n(\lambda_i) = \sum_{r=0}^{p_i-1} \binom{n}{r} \lambda_i^{n-r} \mathbf{N}_{p_i}^r$$

Thus for large n we have a relatively simple expression for $\mathbf{A}^n$.

8.5 The Singular Value Decomposition (SVD) of a Matrix

Theorem 8.10 If A is any n by p matrix then there exists matrices U, V and D such that

$$A = UDV^T$$

where

- U is n by p and $U^T U = I_p$
- V is p by p and $V^T V = I_p$

- $D = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_p)$ and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$$

Proof: Let $\mathbf{x}$ be such that $||\mathbf{x}|| = 1$ and $||A\mathbf{x}|| = ||A||$. That such an $\mathbf{x}$ can be chosen follows from the fact that the norm is a continuous function on a closed and bounded subset of $\mathbf{R}^n$ and hence achieves its maximum at a point in the set. Let $\mathbf{y}_1 = A\mathbf{x}$ and define

$$\mathbf{y} = \frac{\mathbf{y}_1}{||A||}$$

Then

$$||\mathbf{y}|| = 1 \text{ and } A\mathbf{x} = ||A||\mathbf{y} = \sigma\mathbf{y}$$

where $\sigma = ||A||$
Let U_1 and V_1 be such that

$$U = [\mathbf{y}, U_1] \text{ and } [\mathbf{x}, V_1] \text{ are orthogonal}$$

where U is n by p and V is p by p. It then follows that

$$U^T AV = \begin{bmatrix} \mathbf{y}^T \\ U_1^T \end{bmatrix} A\,[\mathbf{x}, V_1] = \begin{bmatrix} \mathbf{y}^T A\mathbf{x} & \mathbf{y}^T AV_1 \\ U_1^T A\mathbf{x} & U_1^T AV_1 \end{bmatrix}$$

Thus

$$A_1 = U^T AV = \begin{bmatrix} \sigma & \mathbf{w}^T \\ \mathbf{0} & B_1 \end{bmatrix}$$

where $\mathbf{w}^T = \mathbf{y}^T AV_1,\ B_1 = U_1^T AV_1$ and $\mathbf{y}^T A\mathbf{x} = \sigma\mathbf{y}^T\mathbf{y} = \sigma$

If we define

$$\mathbf{z} = A_1 \begin{bmatrix} \sigma \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \sigma & \mathbf{w}^T \\ \mathbf{0} & B_1 \end{bmatrix} \begin{bmatrix} \sigma \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \sigma^2 + \mathbf{w}^T\mathbf{w} \\ B_1\mathbf{w} \end{bmatrix}$$

Then

$$||\mathbf{z}||^2 = [\sigma^2 + \mathbf{w}^T\mathbf{w}, \mathbf{w}^T B_1^T] \begin{bmatrix} \sigma^2 + \mathbf{w}^T\mathbf{w} \\ B_1\mathbf{w} \end{bmatrix} = (\sigma^2 + \mathbf{w}^T\mathbf{w})^2 + \mathbf{w}^T B_1^T B_1\mathbf{w}$$

Thus

$$\sigma^2 = ||A||^2 = ||U^T AV||^2 = ||A_1||^2 \geq (\sigma^2 + \mathbf{w}^T \mathbf{w})$$

and it follows that $\mathbf{w} = \mathbf{0}$. Thus

$$U^T AV = \begin{bmatrix} \sigma & \mathbf{0}^T \\ \mathbf{0} & B_1 \end{bmatrix}$$

Induction now establishes the result. The orthogonality properties of U and V then imply that

$$U^T AV = D \iff A = UDV^T$$

8.5.1 Use of the SVD in Data Reduction and Reconstruction

Let $\mathbf{Y}$ be a data matrix consisting of n observations on p response variables. That is

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}$$

The singular value decomposition of $\mathbf{Y}$ allows us to represent $\mathbf{Y}$ as

$$\mathbf{Y} = U_1 D_1 V_1^T$$

where U_1 is n by p, V_1 is p by p and D_1 is a p by p diagonal matrix with ith diagonal element d_i. We know that $d_i \geq 0$ and we may assume that

$$d_1 \geq d_2 \geq \cdots \geq d_p$$

Recall that the matrices U_1 and V_1 have the properties

$$U_1^T U_1 = I_p \;\text{ and }\; V_1^T V_1 = I_p$$

so that the columns of U_1 form an orthonormal basis for the column space of $\mathbf{Y}$ i.e. the subspace of $\mathbf{R}^n$ spanned by the columns of $\mathbf{Y}$. Similarly the columns of V_1 form an orthonormal basis for the row space of $\mathbf{Y}$ i.e. the subspace of $\mathbf{R}^p$ space spanned by the rows of $\mathbf{Y}$.

The rows of $\mathbf{Y}$ represent points in p dimensional space and the

$$y_{i1}, y_{i2}, \ldots, y_{ip}$$

represent the coordinates of the ith individual's values relative to the axes of the p original variables. Two individuals are "close" in this space if the distance (norm) between them is "small". In this case the two individuals have nearly the same values on each of the variables i.e. their coordinates with respect to the original variables are nearly equal. We refer to this subspace of $\mathbf{R}^p$ as **individual space** or **observation space**.

The columns of $\mathbf{Y}$ represent p points in $\mathbf{R}^n$ and represent the observed values of a variable on n individuals. Two points in this space are close if they have similar values over the entire collection of observed individuals i.e. the two variables are measuring the same thing. We call this space **variable space**.

The distance between individuals i and i' is given by

$$||\mathbf{y}_i - \mathbf{y}_{i'}|| = \left[\sum_{j=1}^{p}(y_{ij} - y_{i'j})^2\right]^{1/2}$$

For any data matrix $\mathbf{Y}$ there is an orthogonal matrix V_1 such that if $\mathbf{Z} = \mathbf{Y}V_1$ then the distances between individuals are preserved i.e.

$$||\mathbf{z}_i - \mathbf{z}_{i'}|| = ||\mathbf{y}_i V_1 - \mathbf{y}_{i'} V_1|| = \left[(\mathbf{y}_i - \mathbf{y}_{i'})^T V_1^T V_1 (\mathbf{y}_i - \mathbf{y}_{i'})\right]^{1/2} = ||\mathbf{y}_i - \mathbf{y}_{i'}||$$

Thus the original data is replaced by a new data set in which the coordinate axes of the individuals are now the columns of $\mathbf{Z}$.

We now note that

$$U_1 = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p]$$

where $\mathbf{u}_j$ is the jth column vector of U_1 and that

$$V_1 = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p]$$

where $\mathbf{v}_j$ is the jth column vector of V_1. Thus we can write

$$\mathbf{Y} = U_1 D V_1^T = \sum_{j=1}^{p} d_j^2 \mathbf{u}_j \mathbf{v}_j^T$$

That is we may reconstruct $\mathbf{Y}$ as a sum of constants (the d_j^2) times matrices which are the **outer product** of vectors which span variable space (the $\mathbf{u}_j$) and individual space (the $\mathbf{v}_j$). The interpretation of these vectors is of importance to understanding the nature of the representaion. We note that

$$\mathbf{Y}\mathbf{Y}^T = (U_1 D_1 V_1^T)(V_1 D_1 U_1^T) = U_1 D_1^2 U_1^T$$

Thus

$$\mathbf{Y}\mathbf{Y}^T U_1 = U_1 D_1^2$$

which shows that the columns of U_1 are the eigenvectors of $\mathbf{Y}\mathbf{Y}^T$ with eigenvalues equal to the d_j^2.

We also note that

$$\mathbf{Y}^T\mathbf{Y} = (V_1 D_1 U_1^T)(U_1 D_1 V_1^T) = V_1 D_1^2 V_1^T$$

so that

$$\mathbf{Y}^T\mathbf{Y} V_1 = V_1 D_1^2$$

Thus the columns of V_1 are the eigenvalues of $\mathbf{Y}^T\mathbf{Y}$ with eigenvalues again equal to the d_j^2.

The relationship

$$\mathbf{Y} V_1 = U_1 D_1$$

shows that the columns of V_1 transform an individual's values on the original variables to values in the new variables U_1 with relative importance given by the d_j

If some of the d_j are zero or are approximately 0 then we can approximate $\mathbf{Y}$ as

$$\mathbf{Y} \sim \sum_{j=1}^{p} d_j^2 \mathbf{u}_j \mathbf{v}_j^T$$

which tells us that an approximating subspace has the same dimension in variable space or observation space. æ