Notes on Abstract Algebra

JOHN R. SMITH
email: jrsmith@ucdavis.edu

Physics Department
University of California, Davis
Davis, CA 95616-8677, USA

April 14, 2019

1 Logic

Definition If a statement A leads logically to another statement B we say that 'if A, then B' and write $A \to B$. Another way to think about implication is that $A \to B$ can be thought of as 'if A is true, then B must be true' ('B, if A') or that B is *necessary* for A. This can also be expressed as 'A *only if* B'.

$A \to B$ implies that A being true is sufficient for B to be true. If the we have both $A \to B$ and $B \to A$, then we can say that A is *necessary and sufficient* for B – which can also be stated as $A \iff B$.

Definition n-dimensional Affine Space $\mathbb{R}_n$. An ordered n-tuple of real numbers, i.e., a system of $(x_1, x_2, ..., x_n)$ of n real numbers in a definite order, is called a point (n a positive integer). The numbers $x_1, x_2, ..., x_n$ are called the coordinates of the point; in particular x_1 is called its first, x_2 its second, ..., x_n its n-th coordinate. Two points $P = (x_1, x_2, ..., x_n)$ and $Q = (y_1, y_2, ..., y_n)$ are said to be the *same* or to coincide if and only if $x_1 = y_1, x_2 = y_2, ..., x_n = y_n$. The totality of all n-tuples of real numbers is called n-dimensional Affine Space and is denoted by $\mathbb{R}_n$.

Definition The laws of addition:

A1: The Commutative Law of addition: $x + y = y + x$ for all pairs of numbers x and y.

A2: The Associative Law of addition: $(x + y) + z = x + (y + z)$ for any three numbers x, y and z.

A3: The existence of a zero. There is a number, called the zero and denoted by 0, which has the property that

$$x + 0 = x \text{ and } 0 + x = x \text{ for all numbers } x.$$

A4: The existence of negatives: Corresponding to each number x there is a number called the negative of x and written $-x$ which satisfies

$$x + -x = 0 \text{ and } -x + x = 0.$$

Definition Distributive Law

D1: The Distributive Law:

$$\begin{aligned}(x+y)z &= xz+yz\\ x(y+z) &= xy+xz\end{aligned}$$

Definition The laws of multiplication:

M1: The Commutative Law of multiplication : $xy = yx$ for all pairs of numbers x and y.

M2: The Associative Law of multiplication: $(xy)z = x(yz)$ for any three numbers x, y and z.

M3: The existence of a unity. There is a number, called the unity and denoted by 1, which has the property that

$$x1 = x \text{ and } 1x = x \text{ for all numbers } x.$$

M4: The existence of inverses: Corresponding to each number x there is a number called the inverse of x and written x^{-1} which satisfies

$$xx^{-1} = 1 \text{ and } x^{-1}x = 1.$$

2 Mappings

Definition An injective function $f : X \to Y$ (also known as injection, or one-to-one function) is a function f that maps distinct elements of its domain to distinct elements; that is, $f(x_1) = f(x_2)$ implies $x_1 = x_2$. Equivalently, $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$ in the equivalent contrapositive statement.

Definition A surjective function $f : X \to Y$ (also known as surjection, or onto function) is a function f that every element y can be mapped from some element x so that $f(x) = y$. In other words, every element of the function's codomain is the image of at least one element of its domain. It is not required that x be unique; the function f may map one or more elements of the set X to the same element of the set Y.

Definition Homorphisms and Mappings:
Monomorphism: 1-1, Epimorphism: onto, Isomorphism: 1-1 and onto.
Homomorphisms preserve structure (i.e. multiplication and addition in Groups and Rings).
Homomorphisms that have the same object and image space are called Endomorphisms.
And Endomorphism that is also an Isomorphism is called and Automorphism.
In a Group G, the mappings of G into itself, given by $g \to x^{-1}gx$, where x is any fixed element of G, is an Automorphism, known as an Inner Automorphism.

Suppose $f : A \to B$ and $g : B \to C$. Hence $(g \circ f) : A \to C$ exists. Show
(a) If f and g are 1-1 (monomorphisms), then $g \circ f$ is 1-1:
Suppose $(g \circ f)(x) = (g \circ f)(y)$. Then $g(f(x)) = g(f(y))$. Because g is 1-1, $f(x) = f(y)$ and because f is 1-1, $x = y$. Therefore $(g \circ f)(x) = (g \circ f)(y)$, implies $x = y$; hence $g \circ f$ is 1-1.
(b) If f and g are onto (epimorphisms), then $g \circ f$ is onto.

Suppose $c \in C$. Because g is onto, there exists $b \in B$ for which $g(b) = c$. Because f is onto, there exists $a \in A$ for which $f(a) = b$. Thus $(g \circ f)(a) = g(f(a)) = g(b) = c$, Hence $g \circ f$ is onto.

(c) If $g \circ f$ is 1-1, then f is 1-1.
Prove contrapositive, Suppose f is not 1-1., Then there exists distinct elements $x, y \in A$ for which $f(x) = f(y)$. Then $(g \circ f)(x) = g(f(x)) = g(f(y)) = (g \circ f)(y)$. Hence $g \circ f$ is not 1-1.Therefore if $g \circ f$ is 1-1, then f must be 1-1.
(d) If $g \circ f$ is onto, then g is onto.
Prove contrapositive, If $a \in A$, then $(g \circ f)(a) = g(f(a)) \in g(B)$. Hence $(g \circ f)(A) \subseteq g(B)$. Suppose g is not onto. Then $g(B)$ is properly contained in C and so $(g \circ f)(A)$ is properly contained in C; thus $g \circ f$ is not onto. Therefore if $g \circ f$ is onto, then g must be onto.

3 Linear Dependence

Definition Vectors are directed segments. The vector in $\mathbb{R}_n$ whose components are $a_1, a_2, ..., a_n$ is denoted by

$$\mathfrak{a} = \{a_1, a_2, ..., a_n\} \text{ (braces)}.$$

The zero vector in $\mathbb{R}_n$ is denoted as

$$\mathfrak{o} = \{0, 0, ..., 0\},$$

with n components each 0.

Definition We call p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ **linearly dependent** if there are p numbers $\lambda_1, \lambda_2, ..., \lambda_p$ not all 0, such that

$$\lambda_1\mathfrak{a}_1 + \lambda_2\mathfrak{a}_2 + ... + \lambda_p\mathfrak{a}_p = \mathfrak{o} \quad \text{(the zero vector)}.$$

If no such p numbers exist, then the p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are called **linearly independent**.

Theorem 3.1 *It a subset of the p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ is linearly dependent, then the set of p vectors is itself linearly dependent.*

Suppose the first r vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_r, (r < p)$are linearly dependent. Then there are r numbers $\lambda_1, \lambda_2, ..., \lambda_r$ not all zero such that

$$\lambda_1\mathfrak{a}_1 + \lambda_2\mathfrak{a}_2 + ... + \lambda_r\mathfrak{a}_r = \mathfrak{o}.$$

Setting $\lambda_{r+1}, \lambda_{r+2}, ..., \lambda_p$ to zero we have

$$\lambda_1\mathfrak{a}_1 + \lambda_2\mathfrak{a}_2 + ... + \lambda_r\mathfrak{a}_r + \lambda_{r+1}\mathfrak{a}_{r+1} + ... + \lambda_p\mathfrak{a}_p = \mathfrak{o}.$$

where the numbers $\lambda_1, \lambda_2, ..., \lambda_p$ are not all zero. Therefore the p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly dependent.

Theorem 3.2 *If the p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly independent, then so is every subset of these p vectors.*

For otherwise, by Theorem 3.1, the p vectors would be linearly dependent.

Theorem 3.3 *If the p vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly dependent and if $p > 1$, then at least one of these vectors is a linear combination of the others.*

For, there are numbers λ_i, which do not all vanish such that

$$\lambda_1\mathfrak{a}_1 + \lambda_2\mathfrak{a}_2 + ... + \lambda_p\mathfrak{a}_p = \mathfrak{o}.$$

Suppose $\lambda_p \neq 0$. The we can solve for $\mathfrak{a}_p$,

$$\mathfrak{a}_p = -\frac{\lambda_1}{\lambda_p}\mathfrak{a}_1 - \frac{\lambda_2}{\lambda_p}\mathfrak{a}_2 - ... - \frac{\lambda_{p-1}}{\lambda_p}\mathfrak{a}_{p-1},$$

i.e. $\mathfrak{a}_p$ is a linear combination of the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_{p-1}$.

Theorem 3.4 *If one of the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ is a linear combination of the others, then vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly dependent.*

Suppose

$$\mathfrak{a}_1 = \lambda_2\mathfrak{a}_2 + \lambda_3\mathfrak{a}_3 + ... + \lambda_p\mathfrak{a}_p.$$

Then we have that

$$\mathfrak{a}_1 - \lambda_2\mathfrak{a}_2 - \lambda_3\mathfrak{a}_3 - ... - \lambda_p\mathfrak{a}_p = \mathfrak{o},$$

and since the coefficient of $\mathfrak{a}_1$ is not zero, the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly dependent.

Theorem 3.5 *If the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly independent, and if the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ...\mathfrak{a}_p, \mathfrak{b}$ are linearly dependent, then $\mathfrak{b}$ is a linear combination of $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$.*

We have a relation of the form

$$\lambda_1\mathfrak{a}_1 + \lambda_2\mathfrak{a}_2 + ... + \lambda_p\mathfrak{a}_p + \lambda_{p+1}\mathfrak{b} = \mathfrak{o},$$

where not all the λ_i vanish. Now λ_{p+1} cannot be 0 because the last term would drop out and then all the other λ_i would vanish because of the linear independence of the $\mathfrak{a}_i$ vectors. Therefore $\lambda_{p+1} \neq 0$ and so we have

$$\mathfrak{b} = -\frac{\lambda_1}{\lambda_{p+1}}\mathfrak{a}_1 - \frac{\lambda_2}{\lambda_{p+1}}\mathfrak{a}_2 - ... - \frac{\lambda_p}{\lambda_{p+1}}\mathfrak{a}_p.$$

Theorem 3.6 *If the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are linearly independent, and if $\mathfrak{b}$ is a not a linear combination of $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$, then the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ...\mathfrak{a}_p, \mathfrak{b}$ are linearly independent.*

Theorem 3.6 follows from Theorem 3.5.

4 Vector Spaces in $\mathbb{R}_n$

Let L be a *non-empty* set of vectors of $\mathbb{R}_n$ with the properties:

1. If $\mathfrak{a}$ is a vector of the set L, then $\lambda\mathfrak{a}$ belongs to L for every real λ.
2. If $\mathfrak{a}$ and $\mathfrak{b}$ are two (not necessarily distinct) vectors of the set L, then the vector $\mathfrak{a} + \mathfrak{b}$ also belongs to L.

Such a set of vectors, satisfying 1. and 2., is called a **vector space** in $\mathbb{R}_n$.

Theorem 4.1 Steinitz replacement theorem. *Let $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ be any p vectors, and let L be the vector space spanned by them. Let $\mathfrak{b}_1, \mathfrak{b}_2, ..., \mathfrak{b}_q$ be any q linearly independent vectors of L. Then we may replace a certain set of q vectors from among the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ by the vectors $\mathfrak{b}_1, \mathfrak{b}_2, ..., \mathfrak{b}_q$ so that the remaining vectors of $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ together with the vectors $\mathfrak{b}_1, \mathfrak{b}_2, ..., \mathfrak{b}_q$ span the entire vector space L.*

Proof Use mathematical induction.

Theorem 4.2 *Any set of more than p linear combinations of p given vectors is always linearly dependent.*

Theorem 4.3 *The maximal number of linearly independent vectors in $\mathbb{R}_n$ is n*

Definition Dimension and Basis
Let L be any vector space in $\mathbb{R}_n$. The maximal number of linearly independent vectors is $\leq n$. Let p be this maximal number of independent vectors in L. We call p the **dimension** of L. Thus,

$$0 \leq p \leq n.$$

We call any system of p linearly independent vectors of L, such as $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ a **basis** of L.

Theorem 4.4 *Every vector of L can be represented in exactly one way as a linear combination of the basis vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$.*

Theorem 4.5 *The dimension of the vector space spanned by the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ is equal to the maximal number of linearly independent vectors among $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$.*

Theorem 4.6 *Any k $(k \leq p)$ linearly independent vectors*

$$\mathfrak{b}_1, \mathfrak{b}_2, ..., \mathfrak{b}_k$$

of L can be extended to form a basis of L by suitably adjoining to them $p - k$ other vectors $\mathfrak{a}_{k+1}, \mathfrak{a}_{k+2}, ..., \mathfrak{a}_p$.

Let L' and L'' be any two vector spaces in $\mathbb{R}_n$. By the **intersection** D of L' and L'' we mean the set of all those vectors belonging to both L' and L''.

If we form the totality S of all vectors of the form $\mathfrak{a}' + \mathfrak{a}''$, where $\mathfrak{a}'$ belongs to L' and $\mathfrak{a}''$ belongs to L''. S is called the **sum** of L' and L''.

Theorem 4.7 *If the dimensions of the vector spaces L', L'', D, and S are respectively p', p'', d and s, then*

$$p' + p'' = d + s.$$

5 Linear Spaces

A linear space $\mathbb{L}_p$ of dimension p is the totality of points of $\mathbb{R}_n$ into which a fixed point P is carried by the vectors of a p-dimensional vector space L of $\mathbb{R}_n$. Suppose there are given $p+1$ points

$$P_k = (x_k^1, x_k^2, ..., x_k^n),\ \ k = 0, 1, ..., p$$

which do not lie an any linear space of dimension $< p$. Let us consider the vectors $\mathfrak{a}_k = \overrightarrow{P_0 P_k}, k = 1, 2, ..., p$. These p vectors are linearly independent. For if they were linearly dependent, then the vector space spanned by them would have dimension $< p$. By applying these vectors to this vector space to P_0, we obtain a linear space which on the one hand has dimension $< p$, and on the other contains the points $P_0, P_1, ..., P_p$, which contradicts our assumption.

All points of a linear space are equivalent to one another because any point of it can be used to obtain the entire linear space.

6 Linear Equations

Let a system of m linear equations in n unknowns $x^1, x^2, ..., x^n$ be given in the following form:

$$\begin{array}{ccccc} a_{11}x^1 & + a_{12}x^2 & + \ldots & + a_{1n}x^n = b_1 \\ a_{21}x^1 & + a_{22}x^2 & + \ldots & + a_{2n}x^n = b_2 \\ \cdots & \cdots & \cdots & \cdots\cdots \\ a_{i1}x^1 & + a_{i2}x^2 & + \ldots & + a_{in}x^n = b_i \\ \cdots & \cdots & \cdots & \cdots\cdots \\ a_{m1}x^1 & + a_{m2}x^2 & + \ldots & + a_{mn}x^n = b_m \end{array} \tag{1}$$

which can be written in the following form based on the rules of matrix multiplication

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ \cdot \\ x^i \\ \cdot \\ x^n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ b_i \\ \cdot \\ b_m \end{bmatrix} \tag{2}$$

To every column of the matrix of coefficients we can associate a vector

$$\mathfrak{a}_k = \{a_{1k}, a_{2k}, ..., a_{mk}\},\ \ k = 1, 2, ..., n.$$

The original system of equations can be expressed as a single vector equation

$$\mathfrak{a}_1 x^2 + \mathfrak{a}_2 x^2 + ... + \mathfrak{a}_n x^n = \mathfrak{b}, \tag{3}$$

where $\mathfrak{b} = \{b_1, b_2, ..., b_m\}$.

Consider the vector space $\mathbb{L}$ spanned by the vectors $\mathfrak{a}_k, k = 1, ..., n$ in $\mathbb{R}_m$. If Eq.(3) is solvable for $x^1, x^2, ..., x^n$, then $\mathfrak{b}$ must be a linear combination of the $\mathfrak{a}_k$ and therefore belongs to $\mathbb{L}$. Conversely if

$\mathfrak{b}$ belongs to $\mathbb{L}$ then Eq.(3) is solvable.

Consider the vector space $\mathbb{L}'$ spanned by the vectors $\mathfrak{a}_1, \mathfrak{a}_1, ..., \mathfrak{a}_n, \mathfrak{b}$. A vector $\mathfrak{c}$ in $\mathbb{L}'$ is of the form

$$\mathfrak{c} = \lambda_1 \mathfrak{a}_1 + \lambda_2 \mathfrak{a}_2 + ... + \lambda_n \mathfrak{a}_n + \lambda \mathfrak{b}.$$

If $\mathfrak{b}$ belongs to $\mathbb{L}$, then Eq. (3) is satisfied by certain numbers $x^1, x^2, ..., x^n$ which can be substituted for $\mathfrak{b}$ and so $\mathfrak{c}$ is a linear combination of the $\mathfrak{a}_k, k = 1, ..., n$ and therefore belongs to $\mathbb{L}$. In this case $\mathbb{L}' \subseteq \mathbb{L}$ and also $\mathbb{L} \subseteq \mathbb{L}'$ and therefore $\mathbb{L}'$ and $\mathbb{L}$ are identical.

Theorem 6.1 *The equation (3) is solvable for the x^i, if and only if, the maximal number of linearly independent vectors among $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n$ is the same as the maximal number of linearly independent vectors among $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n, \mathfrak{b}$.*

Definition The **rank** of a matrix is equal to the maximal number of independent column vectors in the matrix.

Consider the following matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} & b_i \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

obtained by appending the vector $\{b_1, b_2, ..., b_m\}$ to the final column. This matrix is called the **augmented** matrix of the system of equations (1).

Theorem 6.2 *Eqs. (1) are solvable for the x^i if and only if the rank of the coefficient matrix of the system (1) is equal to the rank of the augmented matrix*

The following system of equations is called a **homogenous** *in the* x^i

$$\begin{array}{ccccc} a_{11}x^1 & + a_{12}x^2 & + \dots & + a_{1n}x^n = 0 \\ a_{21}x^1 & + a_{22}x^2 & + \dots & + a_{2n}x^n = 0 \\ \dots & \dots & \dots & \dots\dots \\ a_{i1}x^1 & + a_{i2}x^2 & + \dots & + a_{in}x^n = 0 \\ \dots & \dots & \dots & \dots\dots \\ a_{m1}x^1 & + a_{m2}x^2 & + \dots & + a_{mn}x^n = 0 \end{array} \tag{4}$$

Any set of numbers $x^1, x^2, ..., x^n$ which satisfy Eq. (4) are now taken to be the components of an n-dimensional vector $\mathfrak{r} = \{x^1, x^2, ..., x^n\}$. We call such a vector a *vector solution* of the system (4). The totality of vector solutions of the system of equations (4) forms a vector space.

Theorem 6.3 *If the coefficient matrix of the system of equations (4) has rank r, then the set of vector solutions of this system is an $(n-r)$-dimensional vector space.*

Theorem 6.4 *The system of equations (4) has a non-trivial solution, i.e. a solution $x^1, x^2, ..., x^n$ such that not all the x^i vanish, if and only if, the rank of the matrix of (4) is less than n.*

Theorem 6.5 *If the number of equations in the system of homogenous equations (4) is less than the number of unknowns, then the system must have non-trivial solutions.*

Theorem 6.6 *All the solutions of a non-homogenous system of equations (1) are of the form $z = \mathfrak{r} + \eta$, where $\mathfrak{r}$ is a fixed solution of the non-homogenous system (1) and η runs though all solutions of the corresponding homogenous system (4).*

Let s be the maximal number of linearly independent rows of the matrix row vectors. Let us assume that the rows of the homogenous system are ordered so that the first s rows are independent to start with. This involves no loss of generality since it does not affect that rank. The i-th equation of this system is, by definition of the scalar product of two vectors, equivalent to the vector equation

$$\mathfrak{a}_i \cdot \mathfrak{r} = 0,$$

where we set $\mathfrak{r} = \{x^1, x^2, ..., x^n\}$.
Every system of n numbers $x^1, x^2, ..., x^n$ satisfying the first s equation of (the re-ordered) (4), satisfies all of the equations of (4). Because the first s-rows are linearly independent, all the rows satisfy a relation of the form

$$\mathfrak{a}_k = \lambda_1^k \mathfrak{a}_1 + \lambda_2^k \mathfrak{a}_2 + ... + \lambda_s^k \mathfrak{a}_s, \text{ for } 1 \leq k \leq m.$$

If we apply the distributive law on the right, we obtain

$$\mathfrak{a}_k \cdot \mathfrak{r} = \lambda_1^k (\mathfrak{a}_1 \cdot \mathfrak{r}) + \lambda_2^k (\mathfrak{a}_2 \cdot \mathfrak{r}) + ... + \lambda_s^k (\mathfrak{a}_s \cdot \mathfrak{r}).$$

Since $\mathfrak{a}_i \cdot \mathfrak{r} = 0$ for $i = 1, 2, ..., s$ the right hand side of this last equation becomes 0, so that

$$\mathfrak{a}_k \cdot \mathfrak{r} = 0, \quad k = 1, 2, ..., m,$$

and in particular for $k = s+1, s+2, ..., m$ as was to be proved.

The vector space of all the solutions of the first s equations of (4) is thus identical with the space of solutions of the entire system (4). It dimension is this $n - r$ since r is the rank of the matrix of coefficients in (2). It thus follows that the rank of the matrix of the first s equations, i.e of the matrix which consists only of the first s rows of (4) is equal to $n-(n-r) = r$. But the column vectors of this matrix are vectors with s components. Thus the maximal number of linearly independent column vectors of this matrix is $\leq s$ since by previous theorem any $s+1$ vectors of s-dimensional vector space are linearly dependent.

Therefore the maximal number of linearly independent column vectors of a matrix is at most as large as the maximal number of linearly independent row vectors.

Now consider the transpose of the matrix of coefficients of (4). The maximal number of independent column vectors of this matrix is s, and the maximal number of row vectors is r. Applying the previous result we also obtain $s \leq r$. Hence we have $r \leq s$ and $s \leq r$ and therefore s = r. Thus we have proved

Theorem 6.7 *The maximal number of linearly independent column vectors of a matrix is equal to the maximal number of linearly independent row vectors of that matrix.*

Theorem 6.8 *With any given vector space L of $\mathbb{R}_n$, we can always associate a system of homogenous linear equations in n unknowns such that all the vectors of L, and no others, are vector solutions of this system.*

Let $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ be a basis of L, where we set $\mathfrak{a}_i = \{\mathfrak{a}_{i1}, \mathfrak{a}_{i2}, ..., \mathfrak{a}_{in}\}$. If in addition we set

$$\mathfrak{r} = \{x^1, x^2, ..., x^n\}$$

then the equations

$$\mathfrak{a}_i \cdot \mathfrak{r} = 0, \quad i = 1, 2, ..., p, \tag{5}$$

form a system of homogenous linear equations in the unknowns $x^1, x^2, ..., x^n$ whose matrix is

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix}$$

Let p be the dimension of L.The rank of this system must be $n - p$. The totality of vectors which are orthogonal to L is a vector space L' of dimension $n - p$. It consists precisely of all the vector solutions of (5).

We now seek a those vectors which are othogonal to the vector space L' just found, We chose a basis $\mathfrak{b}_1, \mathfrak{b}_2, ..., \mathfrak{b}_{n-p}$ of L'.The $\mathfrak{b}_i$ as vectors of L' are orthogonal to L, and this in particular to $\mathfrak{a}_1, \mathfrak{a}_2, ...\mathfrak{a}_p$. Thus, for $i = 1, 2, ..., p$, we have

$$\mathfrak{b}_1 \cdot \mathfrak{a}_i = 0, \quad \mathfrak{b}_2 \cdot \mathfrak{a}_2 = 0, \quad ..., \quad \mathfrak{b}_{n-p} \cdot \mathfrak{a}_i = 0.$$

Now just as for L, the totality of all vectors orthogonal to L' consists of all the vector solutions of the equations

$$\mathfrak{b}_1 \cdot \mathfrak{r} = 0, \quad \mathfrak{b}_2 \cdot \mathfrak{r} = 0, \quad ..., \quad \mathfrak{b}_{n-p} \cdot \mathfrak{r} = 0. \tag{6}$$

Thus the $\mathfrak{a}_i, i = 1, 2, ..., p$ are orthogonal to L'. But the vectors which are orthogonal to L' form a vector space L'' of dimension $n - (n - p) = p$.Since the vectors $\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_p$ are in L'' and are linearly independent, then form a basis of L''. Therefore L'' is identical with L. Thus the vectors of L are precisely the vector solutions of equations (6), i.e, of a certain system of $n - p$ linear homogenous equations.

Theorem 6.9 *The point solutions of a solvable system of equations of the type (1) form a linear space. This space has dimensions $n - rm$ where r is the rank of the coefficient matrix of the system. Conversely, every linear space may be represented as the totality of all point solutions of some suitable system of linear equations.*

7 Equivalence Relations

Theorem 7.1 *The equivalence classes theorem*
Suppose in a set S we have a relation R defined between certain pairs of elements, xRy meaning that x and y stand in the given relation R.
Suppose further that R has the following 3 properties:

(1) It is reflexive: i.e. xRx for all $x \in S$.

(2) It is symmetric: i.e. $xRy \Rightarrow yRx$.

(3) It is transitive: i.e. xRy and $yRz \Rightarrow xRz$.

Then R is an equivalence relations: i.e. it divides S into mutually exclusive subsets so that every element of S is in one and only one subset, and so that two elements are in the same subset if and only if they stand in the relation R to one another.

The subsets are called equivalence classes defined by R.

Given any element x in S consider all the elements y such that xRy. These form a subset of S: let us call it A_x. We show first that two elements are in the same subset A_x if and only if they stand in the relation R to one another. Suppose yRz and $y \in A_x$. Then xRy and so xRz since R is transitive. Hence $z \in A_x$. Conversely, if y and z are both in A_x we have xRy and xRz, i.e. yRx by the symmetric property and xRz: thus yRz by transitivity.

For each element x we now have a subset A_x, but these will not all be distinct. We will show that two such are either mutually exclusive or else identical. Suppose A_x and A_y both have an element z, Then xRz and yRz, snf do zRy by the symmetric property; hence xRy by the transitive property. Now take any element w of A_x. Then xRw and since xRy we have yRx, giving yRw. So w is in A_y. Hence we have shown that if A_x and A_y have one element z in common, any element w of A_x is in A_y, i.e, $A_x \subset A_y$. Similarly $A_y \subset A_x$, and so the subsets A_x and A_y are identical.
Thus we have mutually exclusive subsets $A_{x_1}, A_{x_2}, \ldots\ldots$ Finally by the reflexive property any element t is in one of the subsets, viz. A_t.

8 Subgroups and Cosets

Theorem 8.1 *A necessary and sufficient condition for a non-empty subset H to be a subgroup of a group G is that $gh^{-1} \in H$ for all $g, h \in H$.*

(1) Necessary: If H is a subgroup $h^{-1} \in H$ and so $gh^{-1} \in H$.

(2) Sufficient: If g is any element of H we have $gg^{-1} \in H$, i.e, $e \in H$. Hence $eg^{-1} = g^{-1} \in H$, i.e, H contains inverses. Thus if $g, h \in H$, so is h^{-1} and hence $g(h^{-1})^{-1} \in H$, i.e. $gh \in H$.

Given a group G and a subgroup H it is possible to decompose the whole group into subsets 'parallel' to H, two elements being in the same subset if their 'difference' is in H. This concept is better adapted in the product notation where the difference between g and k is expressed as $k^{-1}g$. We therefore decompose group G into subsets such that g and k are in the same subset if $k^{-1}g \in H$. These subsets are known as cosets and we have a true decomposition into mutually exclusive subsets because $k^{-1}g \in H$ sets up an equivalence relation gRk if $k^{-1}g \in H$.

Theorem 8.2 *Equivalence Classes of G relative to H The relation defined by gRk above is an equivalence relation.*

(1) Reflexive: For any $g \in H$ we have gRg because $g^{-1}g = e \in H$ since H is a subgroup.

(2) Symmetric: If gRk then $k^{-1}g \in H$. Hence its inverse is in H, i.e. $g^{-1}k \in H$ and so kRq.

(3) Transitive: If gRk and kRl then $k^{-1}g \in H$ and $l^{-1}k \in H$. Hence $(l^{-1}k)(k^{-1}g) = l^{-1}g \in H$ and so gRl.

These mutual exclusive classes are called left cosets of G relative to H. So if gRk then kRg and $g^{-1}k \in H$ or $k = gh$ for some $h \in H$. Hence the coset containing g is precisely the complex gH, the set $\{gh\}$ for all $h \in H$.

Theorem 8.3 *Left Cosets The left cosets of G relative to H are the complexes gH, $g \in G$. Any left coset may be expressed in this form for any g in it. Any two cosets are either the same of have no element in common.*

We may similarly define *right cosets* of G relative to H as the equivalence classes under the relation given by $gk^{-1} \in H$. They are the complexes Hg.

We normally will use left cosets and therefore will just refer to them as cosets.

9 Some Results in the Theory of Numbers

Definition If a and b are two integers such that $a - b$ is divisible by n, we say that a is congruent to b modulo n, and write $a \equiv b \mod n$, or merely $a \equiv b(n)$.

Theorem 9.1 *Lagrange's theorem.*
The order of any subgroup of a finite group is a factor of the order of the group.
The order of a finite group G is a multiple of the order of every one of its subgroups.

Suppose H is a subgroup of G, where the order of G is n and that of H is m.
The whole group G may be decomposed into cosets gH relative to H. The number of elements in each coset must be m, since the coset gH is formed by taking all the elements of H and pre-multiplying them by g, and all the elements so formed are different by the Cancellation Law. But the cosets are completely disjoint. Hence if there are r cosets we must have that $n = mr$, and so m is a factor of n.

Each element a of G generates a cyclic subgroup, whose order is simple the order of a. Therefore we have:

Corollary 9.2 *(Corollary 1): Every element of a finite group G has as order a divisor of the order of G.*

Corollary 9.3 *(Corollary 2): Every group G of prime order p is cyclic.*

Proof For the cyclic subgroup A generated by any element $a \neq e$ in such a group has an order $n > 1$ dividing p. But this implies $n = p$, and so $G = A$ is cyclic. ∎

Corollary 9.4 *(Corollary 3): The only abstract groups of order four are the cyclic group of that order and the four-group (Vierergruppe).*

Proof If a group G has order 4, contains and element of order 4, it is cyclic. Otherwise, by Corollary 1, all elements of G except e must have order 2. Call them a, b, c. By the cancellation law, ab cannot be either $ac = a, eb = b$, of $aa = e$; hence $ab = e$. By symmetry, $ac = ca = b, bc = cb = a, ba = c$. But these, together with $a^2 = b^2 = c^2 = e$, and $ex = xe = x$ for all x, give the multiplication table of the four-group. ∎

Theorem 9.5 *(Fermat's Little Theorem): If p is prime, $a^p \equiv a(p)$, and if a is not a mulltiple of p, then $a^{p-1} \equiv 1(p)$*

Proof (1) The numbers $a, 2a, 3a, ..., (p-1)a$ are all different modulo p if a is not a multiple of p, and so they must be $1, 2, ..., (p-1)$ modulo p in some order. Thus

$$a.2a.3a...(p-1)a \equiv 1.2.3...(p-1).$$

Hence $a^{(p-1)} \equiv 1(p)$. Thus $a^p \equiv a(p)$ if a is not a multiple of p, and this latter result is true also if a is a multiple of p, since then both sides are 0. Hence it is true for all a. ∎

Proof (2) The multiplication group mod p (excluding zero) has $(p-1)$ elements. The order of any element a of this group is then a divisor of $(p-1)$, by Corollary 1, so that $a^{(p-1)} \equiv 1(p)$ whenever $a \neq 0(p)$. If we multiply by a or both sides, we obtain the desired congruence, except for the case $a \equiv 0(p)$, for which the conclusion is trivially true. ∎

Proof (3) For a fixed prime p, let $P(n)$ be the proposition that $n^p \equiv n(p)$. Then $P(0)$ and $P(1)$ are obvious. In the binomial expansion for $(n+1)^p$, every coefficient except the first and the last is divisible by p, hence $(n+1)^p \equiv n^p+1(p)$, whence $P(n)$ implies $(n+1)^p \equiv n+1(p)$, which is the proposition $P(n+1)$. ∎

Theorem 9.6 *(Wilson's Theorem): Let $p > 2$ be a prime number. The field $\mathbb{Z}_p$ is the splitting field of $x^p - x$.*

Proof Since $\mathbb{Z}_p$ is the splitting field of $x^p - x$, we have

$$x^p - x = x(x-1)(x-2)\dots(x-(p-1)).$$

Thus

$$x^{(p-1)} - 1 = (x-1)(x-2)\dots(x-(p-1)),$$

and substituting $x = 0$ gives

$$-1 = (-1)(-2)\dots(-(p-1)).$$

There are an even number of factos on the right hand side, so the minus signs cancel out, leaving

$$(p-1)! \equiv -1(\mod p),$$

which is Wilson's Theorem.

10 Chinese Remainder Theorem

Consider a sequence of congruence equations:

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ &\vdots \\ x &\equiv a_k \pmod{n_k}, \end{aligned}$$

where the n_i are coprime in pairs. Let $N = n_1 n_2 \dots n_k$ and define $N_i = N/n_i$ be the product of all moduli but one. As the n_i are pairwise coprime, N_i and n_i are coprime. Therefore, by the Bezout identity, there exist integers M_i and m_i such that

$$M_i N_i + m_i n_i = 1.$$

Therefore a solution to the system of congruences is

$$x = \sum_{i=1}^{k} a_i M_i N_i.$$

Definition We restate the result as: if the n_i are pairwise coprime, the map

$$x \pmod N \to (x \pmod{n_1}, \ldots, x \pmod{n_k})$$

defines a ring homomorphism

$$\mathbb{Z}/N\mathbb{Z} \cong \mathbb{Z}/n_1\mathbb{Z} \times \ldots \times \mathbb{Z}/n_k\mathbb{Z}$$

between the ring of integers modulo N and the direct product of rings of integers modulo the n_i. This means that for doing a sequence of arithmetic operations in $\mathbb{Z}/N\mathbb{Z}$, one may do the same computation independently in each $\mathbb{Z}/n_i\mathbb{Z}$ and then get the result by applying the isomorphism (from the right to the left).

$$R/(\cap A_i) \cong R/(A_1) \times \ldots \times R/(A_k)$$

We have a surjective map and know the kernel is the intersection of all the A_i's. If you take a map and quotient by the kernel, then that is isomorphic to the image – which is the whole set (everything) because the map is surjective. Therefore the LHS is isomorphic to the RHS.

11 Fields

Theorem 11.1 *(Fields have no non-trivial Ideals): If a field F has an ideal I, then $I = (0)$ or $I = (1) = F$.*

Proof Let $a \in I$ and $a \neq 0$. Then $aa^{-1} = 1 \in I$ and so $f.1 = f \in I$ for any $f \in F$ Therefore $F \subset I$. But $I \subset F$ by definition of Ideals. Therefore $I = F$. The only other possibility is $a = 0$ in which case we have the trivial Ideal $I = (0) = 0$.

If F is any field, then the smallest subfield of F that contains the identify element 1 is called the *prime subfield* of F. If F is a finite field, then its prime subfield is isomorphic to $\mathbb{Z}_p$, where $p = \text{char}(F)$ for some prime p. When we want to emphasize $\mathbb{Z}/p\mathbb{Z}$ with its field structure, we denote it by $\mathbb{F}_p$. The prime subfield of $\mathbb{Z}_p$ which is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ is $\mathbb{F}_p$. The only prime fields are the fields of residues modulo p and the field of rationals.

The prime subfield of a field F is the intersection of all subfields of F. The intersection of any two subfields is a field and similarly for all the other subfields. Also the intersection of all subfields has no proper subfields, since any such would be subfields of R and hence must contain the intersection which is impossible. Hence the intersection is a prime subfield. F cannot contain two distinct prime subfields, since their intersection would also be a subfield and hence must be identical with both.

12 Finite Fields

Theorem 12.1 *(Characteristic of a Field): In a field F, all non-zero elements have the same additive order. This is called the* characteristic *of F and is said to be zero if all non-zero elements have infinite order.*

Proof Suppose that any given non-zero element y has finite order m. Then if x is any other non-zero element, $(mx)y = m(xy) = x(my) = 0$. But $y \neq 0$ and so $mx = 0$. This is a step that could not be performed in the case of rings, which may have zero divisors. Also, suppose $rx = 0$ for some $r < m$. Then $x(ry) = (rx)y = 0$ as before and since $x \neq 0$, $ry = 0$, which is impossible since $r < m$ and m is the order of y. Hence x also has order m.

Theorem 12.2 *If F has a finite characteristic, this must be a prime.*

Proof Suppose F has as characteristic a composite number hk. If a is any non-zero element of F, a^2 is non-zero and so has order hk, so that $hka^2 = 0$, i.e., $(ha)(ka) = 0$ and so either ha or ka is 0, which is impossible, since both h and k are less than hk and a must have order hk.

Theorem 12.3 *If F has characteristic p, it contains a subfield isomorphic to the field of residues modulo p.*

Proof Consider the unity 1 and denote 1+1 by 2, 1+1+1 by 3, etc. Then 1 has order p, and so $p.1 = 0$ (i.e., the sum of the p 1's is the zero element). Now consider the subset of elements $0, 1, 2, \ldots, (p-1)$. The sum and difference of any two of these is in the set as are 0 and 1. Furthermore the product of any two is in the set, since the product is the ordinary product of residues modulo p, by the Distributive Law. For example,

$$2.3 = (1+1)(1+1+1) = 1+1+1+1+1+1 = 6.$$

Since sum and product behave as they do for residues modulo p the subset is isomorphic to the set of residues modulo p under residue sum and product and so, since p is prime, inverses exist in the subset and the subset is isomorphic to the field of residues modulo p.

Theorem 12.4 *Let F be a field of characteristic 0, it contains a subfield isomorphic to the field of rational numbers.*

Proof The unity 1 has infinite order and so the elements $0, 1, 2 \ldots$ are all distinct. Consider the subset of all elements m/n where m and n are co-prime, n is non-zero and m/n means $m.n^{-1}$. It is easily seen that this subset contains the sum, difference, product and quotient of any two of its elements (except dividing by 0, of course) and also contains 0 and 1, and so is a subfield, and is trivially isomorphic to the field of rationals.

Definition A finite field is a field which has a finite number of elements.

Theorem 12.5 *Any finite field has a characteristic p, for p a finite prime (i.e., it cannot have characteristic 0).*

For by Theorem 12.4 a field with characteristic 0 contains an infinite subfield and so must itself be infinite.

Theorem 12.6 *Let F be a finite field of characteristic p. The F has p^n elements, for some positive integer n.*

Proof Recall that if F has characteristic p, then the ring homomorphism $\phi : \mathbb{Z} \to F$ defined by $\phi(n) = n.1$ for all $n \in \mathbb{Z}$ has kernel $p\mathbb{Z}$, and thus the image of ϕ is a subfield K of F isomorphic to $\mathbb{Z}_p$. Since F is finite, it must certainly have finite dimension as a vector space over K, say $[F : K] = n$. If $v_1, v_2, \ldots, v_n$ is a basis for F over K, then each element of F has the form $a_1v_1 + a_2v_2 + \ldots + a_nv_n$ for elements $a_1, a_2, \ldots, a_n \in K$. Thus to define an elements of F there are n coefficients a_i, and for each coefficient there are p choices, since K has only p elements. Therefore the total number of ways in which an element in F can be defined is p^n. ∎

Theorem 12.7 *Let F be a finite field with p^n elements. Then F is the splitting field of the polynomial $x^{p^n} - x$ over the prime subfield of F.*

Proof Let F be a finite field of characteristic p. Then as in Theorem 12.6 the field F is an extension of degree n of its prime subfield K, which is isomorphic to $\mathbb{Z}_p$. Since F has p^n elements, the order of the multiplicative group $F^\times$ of non-zero elements of F is $p^n - 1$. Therefor $x^{p^n-1} \equiv 1$ for all $0 \neq x \in F$, and so $x^{p^n} = x$ for all $x \in F$. The polynomial $f(x) = x^{p^n} - x$ has at most p^n roots, and so its roots must be precisely the elements of F. Thus F is the splitting field of $f(x)$ over K. ∎

13 Descartes Rule of Signs

Theorem 13.1 *(Descartes Rule of Signs): The number of variatons of sign in the subsequent terms of a polynomial, s, minus the number of positive roots, p, is a non-negative even integer.*

$$s - p = 2r, \quad r \geq 0$$

a) $s - p$ is an even integer.

Let $f(x)$ be a polynomial with a positive leading coefficient .

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n, \quad \text{and} \quad a_0 > 0.$$

Then since the sequence of signs starts and ends with a + we must have an even number of variations $+-$ and $-+$ of sign in the sequence

$$+ \ldots + .$$

to get back eventually to find the final + sign so s is even.

Similarly p is even because the function must start at $y = a_0$ and can only cross the x-axis an even number of times to eventually again be in the $y > 0$ region. If there are multiple roots of the form $(x - \alpha)^k = 0$, then these are counted with their respective multiplicity. If k is even, then the function does not cross the x-axis at the multiple root. If k is odd, then the function crosses the x axis and goes into the $y < 0$ region. In either case, p is even. Hence if $a_0 > 0$, we have that $s - p$ is even.
Similar arguments hold for the case when $a_0 < 0$. In that case both s and p are odd numbers and therefore $s - p$ is still even.
b) $s - p$ is non-negative.
Use induction on the degree of the polynomial.
$n = 1$. $x + a_n = 0$. Zero variations if a_n is positive and zero positive roots so $s - p = 0$ is even. If $a_n < 0$ then there is one variation in sign and one positive root, $s = 1$ and $p = 1$ and $s - p = 0$.

Induct on $n = k$.

$$f(x) = a_0x^k + a_1x^{k-1} + a_2x^{k-2} + \ldots + a_{k-1}x + a_k.$$

Take the derivative with respect to x,

$$f'(x) = ka_0x^{k-1} + (k-1)a_1x^{k-2} + (k-2)a_2x^{k-3} + \ldots + a_{k-1}.$$

Now, since the k terms that multiply the various x powers are all positive then

$$s = \text{sign variation}[a_0, a_1, a_2, \ldots, a_{k-1}, a_k]$$

and

$$s' = \text{sign variation}[a_0, a_1, a_2, \ldots, a_{k-1}]$$

Therefore

$$s \geq s'$$

.

Since the roots p' are interspersed between the roots of $f(x)$ by Rolle's theorem, the number of positive roots can be $p-1$ which occur between the roots of $f(x)$ plus possibly one more between $x=0$ and the first root of α_1 of $f(x)$, so:

$$p' \geq p - 1.$$

. Therefore we have

$$s \geq s' \geq p' \geq p - 1$$

or

$$s - p \geq -1$$

. Part a) shows that $s-p$ must be an even number in all cases, so $s - p \geq -1 \implies s - p \geq 0$. Hence

$$s - p = 2r, \quad r \geq 0.$$

14 Determinants and Their Properties

Consider a set of vectors $V = \{\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n\}$ in $\mathbb{R}^n$. The Determinant $D(\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n)$ has the following properties

i) Invariance property: $D(\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n)$ is to remain unchanged if some $\mathfrak{a}_i$ is replaced by $\mathfrak{a}_i + \mathfrak{a}_j \quad (i \neq j)$;

ii) Homogeneity property: $D(\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n)$ is to go over into $\lambda D(\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n)$ if $\mathfrak{a}_i$ is replaced by $\lambda \mathfrak{a}_i$;

iii) the Normalization: $D(e_1, e_2, ..., e_n) = 1$.

The determinant is the unique function of vectors $\{\mathfrak{a}_1, \mathfrak{a}_2, ..., \mathfrak{a}_n\}$ that satisfy the above properties [1].

Conventions: We adopt the Range and Summation Conventions:

Definition Range Convention. *When a small Latin suffix (superscript or subscript) occurs unrepeated in a term, it is understood to take all the values* $1, 2, ..., n$*, where* n *is the number of dimensions of the space.*

Definition Summation Convention. *When a small Latin suffix is repeated in a term, summation with respect to that suffix is understood, the range of summation being* $1, 2, ..., n$.

Kronecker delta:

$$\delta_r^s = \begin{cases} 1 & \text{if } s = r \\ 0 & \text{if } r \neq s \end{cases}$$

Also, the another form of the Kronecker delta is δ_{rs} which satisfies the same cases as above.

Permutation Symbol: Consider a set σ of $\sigma_1, \sigma_2, ..., \sigma_n$ numbers taken from the set $\{1, 2, ..., n\}$. Let

$$\epsilon_{\sigma_1\sigma_2...\sigma_n} = \begin{cases} +1 & \text{if } \sigma \text{ is an even permutation of } \{1, 2, ..., n\} \\ -1 & \text{if } \sigma \text{ is an odd permutation of } \{1, 2, ..., n\} \\ 0 & \text{if } \sigma \text{ has any duplicates} \end{cases}$$

Let the S_n be the set of all permutation of the numbers $\{1, 2, ..., n\}$. In the case of n-by-n matrix A we obtain the complete development of the determinant of A as follows:

$$\det|A| = \sum_{\sigma\in S_n} \epsilon_{\sigma_1\sigma_2...\sigma_n} a_{1\sigma_1} a_{2\sigma_2} ... a_{n\sigma_n} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \tag{7}$$

If the rows are permuted by $\rho \in S_n$, we find

$$\epsilon_{\rho_1\rho_2...\rho_n} \det|A| = \sum_{\sigma\in S_n} \epsilon_{\sigma_1\sigma_2...\sigma_n} a_{\rho_1\sigma_1} a_{\rho_2\sigma_2} ... a_{\rho_n\sigma_n} \tag{8}$$

Cofactor of an element a_{rs},

$$M_{rs} = \text{cofactor}(a_{rs}) = \sum_{\sigma\in S_n,\ \sigma_r=s} \epsilon_{\sigma_1\sigma_2...\sigma_n} a_{1\sigma_1} a_{2\sigma_2} ... a_{r-1\sigma_{r-1}} a_{r+1\sigma_{r+1}} .... a_{n\sigma_n} \tag{9}$$

Relation between cofactors and minors obtained by taking the determinant obtained from A by striking out row r and column s from A. We denote such minor determinants by A_{rs} :

$$M_{rs} = (-1)^{r+s} A_{rs}$$

For every r, $1 \le r \le n$ we can write $\det|A|$ in terms of a cofactor expansion

$$\det|A| = \sum_{s=1}^{n} a_{rs}\, M_{rs} = \sum_{s=1}^{n} (-1)^{r+s}\, a_{rs}\, A_{rs} \tag{10}$$

Also, if there is a mismatch between the row indices between the matrix element and the cofactor or minor, i.e.,

$$\sum_{s=1}^{n} a_{ks}\, M_{rs} = \sum_{s=1}^{n} (-1)^{r+s}\, a_{ks}\, A_{rs} = 0, \text{ for } \quad k \neq r \tag{11}$$

The case $k \neq r$ corresponds to a determinant with two identical rows and hence vanishes because the rows are linearly dependent. Putting both of the above equations together we find

$$\sum_{s=1}^{n} a_{ks}\, M_{rs} = \sum_{s=1}^{n} (-1)^{r+s}\, a_{ks}\, A_{rs} = \delta_{kr} \det|A| \tag{12}$$

Schur's complement: Suppose p, q are nonnegative integers, and suppose A, B, C, D are respectively p-by-p, p-by-q, q-by-p, and q-by-q matrices of complex numbers. Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{13}$$

so that M is a $(p+q)$-by-$(p+q)$ matrix.

If D is invertible, then the Schur complement of the block D of the matrix M is the p-by-p matrix defined by

$$M/D := A - BD^{-1}C. \tag{14}$$

and

$$\det M = \det D \det(A - BD^{-1}C).$$

If A is invertible, the Schur complement of the block A of the matrix M is the q-by-q matrix defined by

$$M/A := D - CA^{-1}B. \tag{15}$$

and

$$\det M = \det A \det(D - CA^{-1}B).$$

In the case that A or D is singular, substituting a generalized inverse for the inverses on M/A and M/D yields the generalized Schur complement.

Laplace's General Expansion Theorem: Divide the set of integers $1, 2, ..., n$ into two complementary sets $\{\alpha_1, \alpha_2, ..., \alpha_p\}$ and $\{\beta_1, \beta_2, ..., \beta_q\}$ where $n = p + q$ and

$$\alpha_1 < \alpha_2 < ... < \alpha_p \text{ and } \beta_1 < \beta_2 < ... < \beta_q.$$

These two sets remained fixed throughout the following:

$$\det A = \sum_{\substack{\rho_1,\rho_2,...,\rho_p \\ \rho_1<\rho_2<...<\rho_p}} (-1)^{(\alpha_1+\alpha_2+...+\alpha_p+\rho_1+\rho_2...+\rho_p)} \begin{vmatrix} a_{\alpha_1\rho_1} & a_{\alpha_1\rho_2} & \dots & a_{\alpha_1\rho_p} \\ a_{\alpha_2\rho_1} & a_{\alpha_2\rho_2} & \dots & a_{\alpha_2\rho_p} \\ \dots & \dots & \dots & \dots \\ a_{\alpha_i\rho_1} & a_{\alpha_i\rho_2} & \dots & a_{\alpha_i\rho_p} \\ \dots & \dots & \dots & \dots \\ a_{\alpha_p\rho_1} & a_{\alpha_p\rho_2} & \dots & a_{\alpha_p\rho_p} \end{vmatrix} \begin{vmatrix} a_{\beta_1\sigma_1} & a_{\beta_1\sigma_2} & \dots & a_{\beta_1\sigma_q} \\ a_{\beta_2\sigma_1} & a_{\beta_2\sigma_2} & \dots & a_{\beta_2\sigma_q} \\ \dots & \dots & \dots & \dots \\ a_{\beta_i\sigma_1} & a_{\beta_i\sigma_2} & \dots & a_{\beta_i\sigma_q} \\ \dots & \dots & \dots & \dots \\ a_{\beta_q\sigma_1} & a_{\beta_q\sigma_2} & \dots & a_{\beta_q\sigma_q} \end{vmatrix} \tag{16}$$

The sets $\{\alpha_1, \alpha_2, ..., \alpha_p\}$ and $\{\beta_1, \beta_2, ..., \beta_q\}$ were defined to be fixed. The summation is taken over all p-tuples $(\rho_1, \rho_2, ..., \rho_p)$ from among the numbers $1, 2, ..., n$ for which $\rho_1 < \rho_2 < ... < \rho_p$ and where

$$\sigma_1 < \sigma_2 < ... < \sigma_q$$

are the remaining q integers.

15 Cauchy-Binet Formula

Application 1: **Example Partitioned Matrix and the Cauchy-Binet Formula**

$$M = \left[\begin{array}{cccc|cccc} 0 & 0 & \dots & 0 & b_{11} & b_{12} & \dots & b_{1n} \\ 0 & 0 & \dots & 0 & b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & b_{m1} & b_{m2} & \dots & b_{mn} \\ \hline c_{11} & c_{12} & \dots & c_{1m} & 1 & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & c_{2m} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ c_{n1} & c_{n2} & \dots & c_{nm} & 0 & 0 & \dots & 1 \end{array}\right] \tag{17}$$

In the above M is an $(m+n)$-by-$(m+n)$ matrix which is partitioned into: A an m-by-m matrix of 0's, B an m-by-n matrix, C an n-by-m matrix and D is the n-by-n Identity matrix. Then using the determinant in the Schur's Complement form we have

$$\det M = \det D \det(A - BD^{-1}C) = \det(-BC) = (-1)^m \det(BC).$$

We can obtain a similar result by row operations. Denote the $(m+i)$ row by $\mathfrak{a}_i$. Adding a multiple of row i to row j does not change the value of $\det M$. We successively subtract the $\sum_{i=1}^{n} b_{k\,i}\mathfrak{a}_i$ from the k-th row of M, for $k = 1, 2, ..., m$ changing M into a different matrix with the same determinant, viz.,

$$\tilde{M} = \left[\begin{array}{cccc|cccc} -d_{11} & -d_{12} & \dots & -d_{1m} & 0 & 0 & \dots & 0 \\ -d_{21} & -d_{22} & \dots & -d_{2m} & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -d_{m1} & -d_{m2} & \dots & -d_{mm} & 0 & 0 & \dots & 0 \\ \hline c_{11} & c_{12} & \dots & c_{1m} & 1 & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & c_{2m} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ c_{n1} & c_{n2} & \dots & c_{nm} & 0 & 0 & \dots & 1 \end{array}\right], \tag{18}$$

where $d_{kj} = \sum_{i=1}^{n} b_{ki}c_{ij}$, for $k, j = 1, ..., m$. Since $\det M = \det \tilde{M}$ we have another demonstration of

$$\det M = (-1)^m \det(BC) \tag{19}$$

where B and C are the matrices corresponding to b_{kj} and c_{ij} respectively.

Another representation of $\det M$ can be obtained by applying Laplace's Generalized Theorem successively. First swap the bottom n rows with the top m rows (by moving each of the bottom n rows up m positions and changing the sign of $\det M$ accordingly

$$\det M = (-1)^{m^2} \det \left[\begin{array}{cccc|cccc} c_{11} & c_{12} & \dots & c_{1m} & 1 & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & c_{2m} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ c_{n1} & c_{n2} & \dots & c_{nm} & 0 & 0 & \dots & 1 \\ \hline 0 & 0 & \dots & 0 & b_{11} & b_{12} & \dots & b_{1n} \\ 0 & 0 & \dots & 0 & b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & b_{m1} & b_{m2} & \dots & b_{mn} \end{array}\right] \tag{20}$$

Using Laplace's Generalized Theorem the first time across the bottom m rows and taking advantage of the m-by-m block of zeros in the lower left hand corner to drop some of the minors in the expansion we have only a sum over the m-by-m minors in the b_{ij} elements starting in columns $m+1$ through $m+n$ of M. After forming any one of these minors we have to select m columns to remove to form the algebraic complement to use in the Laplace Theorem. The n-by-n complementary minor consists of the n-by-m c_{ij} matrix elements with an additional $n-m$ columns which remain in from the former Identity matrix in the upper right hand side of the working matrix.

We now apply Laplace's Generalized Theorem a second time, now expanding over the first m columns of the complementary n-by$n$ minor. We must again sum over the choice of m rows from the first n rows. The complementary

minor to this second set of minors is obtained from the original n-by-n Identity matrix, I_n, with m columns removed by forming the first set of minors in the b_{ij} and removing the m rows to form the second set of minors. We are left with an $n-m$-by-$n-m$ matrix from I_n. If the rows that were struck out do not match exactly the columns that were struck out then there will appear in the the reduced matrix rows and columns with all zero entries and the corresponding determinant vanishes. So in the second application of Laplace's Generalized theorem there is only one term surviving in the sum, for any given choice of m numbers $1 \leq \rho_1 \leq \rho_2 \leq ... \leq \rho_m \leq n$, the only term that survive is

$$\begin{vmatrix} c_{1\rho_1} & c_{1\rho_2} & \dots & c_{1\rho_m} \\ c_{2\rho_1} & c_{2\rho_2} & \dots & c_{2\rho_m} \\ . & . & . & . \\ c_{m\rho_1} & c_{m\rho_2} & \dots & c_{m\rho_m} \end{vmatrix} \begin{vmatrix} b_{1\rho_1} & b_{1\rho_2} & \dots & b_{1\rho_m} \\ b_{2\rho_1} & b_{2\rho_2} & \dots & b_{2\rho_m} \\ . & . & . & . \\ b_{m\rho_1} & b_{m\rho_2} & \dots & b_{m\rho_m} \end{vmatrix} \tag{21}$$

Putting both results together and using the fact that both methods give an expression for $\det M$ we obtain the Cauchy-Binet Formula for matrices B (n-by-m) and C (m-by-n) with $m \leq n$:

$$\det(BC) = \sum_{\substack{\rho_1,\rho_2,...,\rho_m=1 \\ \rho_1<\rho_2<...<\rho_m}}^{n} \begin{vmatrix} c_{1\rho_1} & c_{1\rho_2} & \dots & c_{1\rho_m} \\ c_{2\rho_1} & c_{2\rho_2} & \dots & c_{2\rho_m} \\ . & . & . & . \\ c_{m\rho_1} & c_{m\rho_2} & \dots & c_{m\rho_m} \end{vmatrix} \begin{vmatrix} b_{1\rho_1} & b_{1\rho_2} & \dots & b_{1\rho_m} \\ b_{2\rho_1} & b_{2\rho_2} & \dots & b_{2\rho_m} \\ . & . & . & . \\ b_{m\rho_1} & b_{m\rho_2} & \dots & b_{m\rho_m} \end{vmatrix} \tag{22}$$

Application 2: **Alternate Derivation of the Cauchy-Binet Formula**

Let A be an m-by-n matrix and let B be an n-by-m matrix. Let $1 \leq j_1, j_2, ..., j_m \leq n$. Let $A_{j_1 j_2 ... j_m}$ denote the m-by-m matrix consisting of the m columns $j_1, j_2, ..., j_m$ of A. Let $B_{j_1 j_2 ... j_m}$ denote the m-by-m matrix consisting of the m rows $j_1, j_2, ..., j_m$ of B. Let $\{k_1, k_2, ..., k_m\}$ be an ordered m-tuple of integers and let $\{j_1, j_2, ..., j_m\}$ be the same set of integers but arranged into non-decreasing order:

$$\mathrm{J}1 \leq j_2 \leq ... \leq j_m.$$

Then we have the relationships that

$$\begin{aligned} \det\left(B_{k_1 k_2 ... k_m}\right) &= \epsilon_{k_1 k_2 ... k_m} \det\left(B_{j_1 j_2 ... j_m}\right), \\ \det\left(B_{j_1 j_2 ... j_m}\right) &= \epsilon_{k_1 k_2 ... k_m} \det\left(B_{k_1 k_2 ... k_m}\right) \text{ no implied sum on } k_i's. \end{aligned}$$

Now from the definition of determinant for the m-by-m matrix AB we have

$$\begin{aligned} \det(AB) &= \sum_{1 \leq l_1, ... l_m \leq m} \epsilon_{l_1 l_2 ... l_m} \left(\sum_{k=1}^{n} a_{1k} b_{k l_1}\right) ... \left(\sum_{k=1}^{n} a_{mk} b_{k l_m}\right) \\ &= \sum_{1 \leq k_1, ... k_m \leq n} (a_{1k_1} ... a_{mk_m}) \sum_{1 \leq l_1, ... l_m \leq m} \epsilon_{l_1 l_2 ... l_m} b_{k_1 l_1} ... b_{k_m l_m} \\ &= \sum_{1 \leq k_1, ... k_m \leq n} (a_{1k_1} ... a_{mk_m}) \det\left(B_{k_1 ... k_m}\right) \\ &= \sum_{1 \leq k_1, ... k_m \leq n} (a_{1k_1} ... a_{mk_m} \epsilon_{k_1 ... k_m}) \det\left(B_{j_1 ... j_m}\right) \\ &= \sum_{1 \leq k_1, ... k_m \leq n} (\epsilon_{k_1 ... k_m} a_{1k_1} ... a_{mk_m}) \det\left(B_{j_1 ... j_m}\right) \\ &= \sum_{1 \leq j_1 \leq j_2 \leq ... \leq j_m \leq n} \det\left(A_{j_1 ... j_m}\right) \det\left(B_{j_1 ... j_m}\right) \end{aligned}$$

If any two j's are equal then $\det(A_{j_1 \ldots j_m}) = 0$

In the case of A and $B = A^T$ we immediately have

$$\det(AA^T) = \sum_{1 \le j_1 \le j_2 \le \ldots \le j_m \le n} [\det(A_{j_1 \ldots j_m})]^2$$

‘

16 Vandermonde Determinant

The Vandermonde Determinant for n variables $x = [x_1, x_2, \ldots, x_n]$ is given by

$$V_n(x) = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & x_2^3 & \ldots & x_2^n \\ 1 & x_3 & x_3^2 & x_3^3 & \ldots & x_2^n \\ \vdots & & & & & \\ 1 & x_n & x_n^2 & x_n^3 & \ldots & x_n^n \end{bmatrix}$$

Except for the first column, subtract x_1 times the previous column to the rest of the columns. Since a multiple of any column added to any other column does not change the determinant, we can write

$$\det V_n(x) = \det \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & x_2^2(x_2 - x_1) & \ldots & x_2^{n-1}(x_2 - x_1) \\ 1 & x_3 - x_1 & x_3(x_3 - x_1) & x_3^2(x_3 - x_1) & \ldots & x_3^{n-1}(x_3 - x_1) \\ \vdots & & & & & \\ 1 & x_n - x_1 & x_n(x_n - x_1) & x_n^2(x_n - x_1) & \ldots & x_n^{n-1}(x_n - x_1) \end{bmatrix}$$

Expanding the determinant using the elements in the first row we have

$$\det V_n(x) = \det \begin{bmatrix} x_2 - x_1 & x_2(x_2 - x_1) & x_2^2(x_2 - x_1) & \ldots & x_2^{n-1}(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) & x_3^2(x_3 - x_1) & \ldots & x_3^{n-1}(x_3 - x_1) \\ \quad \vdots & & & & \\ x_n - x_1 & x_n(x_n - x_1) & x_n^2(x_n - x_1) & \ldots & x_n^{n-1}(x_n - x_1) \end{bmatrix}$$

Factoring out $x_2 - x_1$ from the first row, $x_3 - x_1$ from the second row, and so on until the last row we have

$$\det V_n(x_1, x_2, \ldots, x_n) = [\prod_{i=2}^{n}(x_i - x_1)] \det \begin{bmatrix} 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \ldots & x_3^{n-1} \\ \vdots & & & & \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{bmatrix} = \prod_{i=2}^{n}(x_i - x_1) \ \det V_{n-1}(x_2, x_3, \ldots, x_n)$$

Proceeding in a similar way by subtracting x_2 times the previous column except for the first column we can continue

$$\det V_n(x_1, x_2, \ldots, x_n) = [\prod_{i=2}^{n}(x_i - x_1)] \, [\prod_{i=3}^{n}(x_i - x_2)] \ \det V_{n-2}(x_3, x_4, \ldots, x_n)$$

To arrive at

$$\det V_n(x_1, x_2, \ldots, x_n) = \prod_{\substack{1<j \\ i<j}}^{n}(x_j - x_i) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

17 Cayley-Hamilton Theorem

The characteristic equation for eigenvalues $0 = \det|A - \lambda I|$ can be expanded as a polynomial in λ

$$\phi(\lambda) = \det|A - \lambda I| = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0, \quad a_n = (-1)^n.$$

And the characteristic equation can be written in the form $\phi(\lambda) = 0$.

If in any polynomial $f(\lambda) = \sum c_i\lambda^i$ we substitute the matrix A for λ and multiply the constant term c_0 by I, we obtain a matrix denoted by $f(A)$.

Theorem 17.1 *(Cayley-Hamilton) Any square matrix satisfies its characteristic equation.*

Proof Let $\phi(\lambda)$ be the characteristic determinant of A. We want to show that $\phi(A) = 0$. Since the elements of $A - \lambda I$ are linear functions of λ, and the elements of its adjoint [the transpose of the matrix of signed minors of $A - \lambda I$ – denoted by Adj_$(A - \lambda I)$] are (n-1)-rowed determinants, they are polynomials in λ of degree $\leq n - 1$. Let C stand for the Adj-$(A - \lambda I)$. If the element in the i-th row and j-th column of C is $\sum_k c_{ijk}\lambda^k$, then

$$C = \sum_{k=0}^{n-1} C_k\lambda^k, \quad C_k = (c_{ijk}) \quad (i, j = 1, 2, \ldots, n)$$

Using the relationship between the adjoint and the determinant, $\phi(\lambda) = \det|A - \lambda I|$, we can write

$$(A - \lambda I)C = \phi(\lambda)I.$$

This expression can be expanded as

$$A\sum_{k=0}^{n-1} C_k\lambda^k - \lambda\sum_{k=0}^{n-1} C_k\lambda^k = \sum_{k=0}^{n} a_k\lambda^k I$$

Equating the terms free of λ and the coefficients of $\lambda, \lambda^1, ..., \lambda^{n-1}, \lambda^n$ we get

$$\begin{aligned} AC_0 &= a_0 I, \\ AC_1 - C_0 &= a_1 I, \\ AC_2 - C_1 &= a_2 I, \\ \ldots &\quad \ldots \\ AC_{n-1} - C_{n-2} &= a_{n-1} I, \\ -C_{n-1} &= a_n I. \end{aligned}$$

Multiply these equations on the left by $I, A, A^2, ..., A^{n-1}, A^n$ respectively and add; we get

$$0 = a_0 I + a_1 A + a_2 A^2 + \ldots + a_{n-1}A^{n-1} + a_n A^n = \phi(A),$$

and we see that the matrix A satisfies its own characteristic equation. ∎

18 Sylvesters Criterion for Positive Definite Hermitian Matrices

Theorem 18.1 *(Sylvester's Criterion) An n-x-n Hermitian matrix M is positive-definite if and only if all the leading principal minors are positive.*

Proof Let M be an n x n Hermitian positive definite matrix. Then $0 \leq x^\dagger Mx$ if x is not the zero vector. Let M_k be the k-th leading principal minor (the first k-rows and k-columns in the upper left hand corner of M). Select a non-zero vector x as follows

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \vec{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then we have that $0 \leq x^\dagger Mx = \vec{x}^\dagger M_k \vec{x}$. Therefore, since $\vec{x} \neq \vec{0}$ is otherwise arbitrary, then all the eigenvalues of M_k must be positive and hence $\det |M_k| > 0$
To prove the reverse implication, we use induction. The general form of an $(n+1)$-x-$(n+1)$ Hermitian matrix is

$$M_{n+1} = \begin{bmatrix} M_n & \vec{v} \\ \vec{v}^\dagger & d \end{bmatrix}$$

Denote the $(n+1)$ dimensional vector x as

$$x = \begin{bmatrix} \vec{x} \\ x_{n+1} \end{bmatrix}$$

Then

$$x^\dagger M_{n+1} x = \vec{x}^\dagger M_n \vec{x} + x_{n+1}\vec{x}^\dagger \vec{v} + \overline{x}_{n+1}\vec{v}^\dagger \vec{x} + d|x_{n+1}|^2$$

By completing the square, this last expression is equal to

$$(\vec{x}^\dagger + \vec{v}^\dagger M_n^{-1}\overline{x}_{n+1}) M_n (\vec{x} + x_{n+1} M_n^{-1}\vec{v}) - |x_{n+1}|^2 \vec{v}^\dagger M_n^{-1}\vec{v} + d|x_{n+1}|^2$$

$$= (\vec{x} + \vec{c})^\dagger M_n (\vec{x} + \vec{c}) + |x_{n+1}|^2 (d - \vec{v}^\dagger M_n^{-1}\vec{v}),$$

where $\vec{c} = x_{n+1} M_n^{-1}\vec{v}$. M_n^{-1} exists because the eigenvalues of M_n are all positive. The first term is positive by the induction hypothesis. Using this useful determinant expansion

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det |A| \det |D - CA^{-1}B|$$

We have that

$$\det |M_{n+1}| = \det |M_n| (d - \vec{v}^\dagger M_n^{-1}\vec{v}) > 0,$$

which implies $(d - \vec{v}^\dagger M_n^{-1}\vec{v}) > 0$. Hence $x^\dagger M_{n+1} x > 0$ (is also positive definite). ∎

19 Resultant of Two Polynomials

20 Resultant Expressed as a Sylvester Determinant

Consider two polynomials

$$f(x) = \sum_{i=0}^{n} f_i\, x^{n-i} = f_0 x^n + f_1 x^{n-1} + f_2 x^{n-2} + \ldots + f_{n-1} x + f_n$$
$$g(x) = \sum_{j=0}^{m} g_j\, x^{m-j} = g_0 x^m + g_1 x^{m-1} + g_2 x^{m-2} + \ldots + g_{m-1} x + g_m$$

Let use assume that $f(x)$ and $g(x)$ have a common factor $(x - \alpha)$. Then

$$f(x) = (x-\alpha) f_1(x) \quad \text{and}$$
$$g(x) = (x-\alpha) g_1(x).$$

Then

$$f(x) g_1(x) = g(x) f_1(x) = \frac{f(x) g(x)}{x - \alpha}.$$

Assuming that

$$f_1(x) = -\sum_{i=0}^{n-1} z_i\, x^{n-1-i} = -(z_0 x^{n-1} + z_1 x^{n-2} + z_2 x^{n-3} + \ldots + z_{n-2} x + z_{n-1})$$
$$g_1(x) = \sum_{j=0}^{m-1} y_j\, x^{m-1-i} = y_0 x^{m-1} + y_1 x^{m-2} + y_2 x^{m-3} + \ldots + y_{m-2} x + y_{m-1}$$

By equating like powers of x, we can write out a linear system corresponding to the equation

$$f(x) g_1(x) - g(x) f_1(x) = 0$$

we find

$$\begin{aligned}
-g(x) f_1(x) = \; & g_0 z_0 x^{n+m-1} + g_0 z_1 x^{n+m-2} + g_0 z_2 x^{n+m-3} + \ldots + g_0 z_{n-2} x^{m+1} + g_0 z_{n-1} x^m + \\
& g_1 z_0 x^{n+m-2} + g_1 z_1 x^{n+m-3} + \ldots + g_1 z_{n-3} x^{m+1} + g_1 z_{n-2} x^m + g_1 z_{n-1} x^{m-1} + \\
& g_2 z_0 x^{n+m-3} + g_2 z_1 x^{n+m-4} + \ldots + g_2 z_{n-3} x^m + g_2 z_{n-2} x^{m-1} + g_2 z_{n-1} x^{m-2} + \\
& \vdots \quad + \\
& \ldots + (g_{m-2} z_{n-1} + g_{m-1} z_{n-2} + g_m z_{n-3}) x^2 + (g_{m-1} z_{n-1} + g_m z_{n-2}) x + g_m z_{n-1}
\end{aligned}$$

$$
\begin{aligned}
g(x)g_1(x) = \; & f_0y_0x^{n+m-1} + f_0y_1x^{n+m-2} + f_0y_2x^{n+m-3} + \ldots + f_0y_{m-2}x^{n+1} + f_0y_{n-1}x^n + \\
& f_1y_0x^{n+m-2} + f_1y_1x^{n+m-3} + \ldots + f_1y_{n-3}x^{n+1} + f_1y_{m-2}x^n + f_1y_{n-1}x^{n-1} + \\
& f_2y_0x^{n+m-3} + f_2y_1x^{n+m-4} + \ldots + f_2y_{n-3}x^n + f_2y_{m-2}x^{n-1} + f_2y_{n-1}x^{n-2} + \\
& \vdots \; + \\
& \ldots + (f_{n-2}y_{m-1} + f_{n-1}y_{m-2} + f_ny_{m-3})x^2 + (f_{n-1}y_{m-1} + f_ny_{m-2})x + f_ny_{m-1}
\end{aligned}
$$

In order to satisfy $f(x)g_1(x) - g(x)f_1(x) = 0$, all the terms multiplying like powers of x each must vanish. This can be arranged as a linear system in the vector $[z_0, z_1, ..., z_{n-1}, y_0, y_2, ...y_{m-1}]$. The coefficient matrix of this linear system is equal to the transpose of the Sylvester Resultant matrix which is defined as

$$
R_{m,n}(g,f) = \underbrace{\left[\begin{array}{ccccc}
g_0 & \cdots & g_m & & \\
& \ddots & \cdots & \ddots & \\
& & g_0 & \cdots & g_m \\
f_0 & \cdots & & f_n & \\
& \ddots & \cdots & \ddots & \\
& & f_0 & \cdots & f_n
\end{array}\right]}_{m+n}
\begin{array}{l}
\left.\begin{array}{l}\\\\\end{array}\right\} n \text{ lines} \\
\left.\begin{array}{l}\\\\\end{array}\right\} m \text{ lines}
\end{array}
$$

The corresponding linear system that sets all the coefficients of the same powers of x to 0 in $f(x)g_1(x) - g(x)f_1(x) = 0$ is

$$
\left[\begin{array}{ccccc}
g_0 & \cdots & g_m & & \\
& \ddots & \cdots & \ddots & \\
& & g_0 & \cdots & g_m \\
f_0 & \cdots & & f_n & \\
& \ddots & \cdots & \ddots & \\
& & f_0 & \cdots & f_n
\end{array}\right]^T
\left[\begin{array}{c} z_0 \\ \vdots \\ z_{n-1} \\ y_0 \\ \vdots \\ y_{m-1} \end{array}\right] =
\left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{array}\right]
$$

The criterion for a non-trivial solution to the above linear system is

$$
\det \left[\begin{array}{ccccc}
g_0 & \cdots & g_m & & \\
& \ddots & \cdots & \ddots & \\
& & g_0 & \cdots & g_m \\
f_0 & \cdots & & f_n & \\
& \ddots & \cdots & \ddots & \\
& & f_0 & \cdots & f_n
\end{array}\right] = 0
$$

The rank of the matrix $R_{m,n}(g,f)$ also gives the multiplicity of the common root of $f(x)$ and $g(x)$. The leading term of this determinant is $g_0^n f_n^m$. The terms of in the expansion an $(n+m) \times (n+m)$ matrix M of the form $\sum_\sigma \epsilon_{r_1,r_2,r_3,\ldots r_{n+m}} m_{1r_1} m_{2r_2} m_{3r_3} \ldots m_{(n+m)r_{n+m}}$ where the sum is over all permutations σ of the numbers $[1, 2, \ldots, n+m]$ and $sgn(\sigma) = \epsilon_{r_1,r_2,r_3,\ldots r_{n+m}}$ is the sign of the permutation. In the case of the Sylvester Resultant the terms take the form

$$
sgn(\sigma) g_0^{\nu_0} g_1^{\nu_1} \ldots g_m^{\nu_m} f_0^{\mu_0} f_1^{\mu_1} \ldots f_n^{\mu_n},
$$

where

$$n = \nu_0 + \nu_1 + \nu_2 + \ldots + \nu_m$$
$$m = \mu_0 + \mu_1 + \mu_2 + \ldots + \mu_n$$

Factoring out $g_0^n f_0^m$ the generic term in the expansion of the determinant becomes

$$sgn(\sigma) g_0^n f_0^m (\frac{g_1}{g_0})^{\nu_1} (\frac{g_2}{g_0})^{\nu_2} \ldots (\frac{g_m}{g_0})^{\nu_m} (\frac{f_1}{f_0})^{\mu_1} (\frac{f_2}{f_0})^{\mu_2} \ldots (\frac{f_n}{f_0})^{\mu_n}$$

Hence the Sylvester Resultant has the form

$$R_{m,n}(g,f) = g_0^n f_0^m \phi(\alpha, \beta)$$

where ϕ is an integral function of the roots α_i and β_i. As the determinant vanishes for any common roots $\alpha_i = \beta_k$, the determinant is divisible by $(\alpha_i - \beta_k)$ and the determinant is divisible by

$$g_0^n f_0^m \prod_{i,k} (\alpha_i - \beta_k)$$

Theorem 20.1 *$R_{m,n}(g,f)$ is a homogenous polynomial with integer coefficients g_i, f_j.*

(i) $R_{m,n}(g,f)$ is homogenous of degree n in $g_m, ..., g_0$ and degree m in $f_n, ..., f_0$.

(ii)If g_i and f_i are regarded as having degree i. then $R_{m,n}(g,f)$ is homogenous of degree mn.

Proof It is obvious that $R_{m,n}(g,f)$ is a homogenous polynomial with integer coefficients in the coefficients g_i, f_j, of total degree $m+n$. Moreover, to replace g_i by tg_i and f_j by uf_j in the Sylvester matrix means that we multiply each of the first m rows by t and each of the last n rows by u, and thus the determinant $R_{m,n}(g,f)$ by $t^m u^n$, which shows part (i). It is here best to treat g_i, f_j, t and u as different inderminants a do the calculations in $F(g_0, ..., g_m, f_0, ..., f_n, t, u)$. Similarly, (ii) follows because to replace each g_i by $t^i g_i$ and each f_j by $t^j f_j$ in $R_{m,n}(g,f)$ yields the same result as multiplying the ith row by t^{m+i} for $i = 1, 2, ..., m$ and by t^i for $i = m+1, ..., m+n$, and the jth column by t^{-j}. The overall effect it to multiply the determiant $R_{m,n}(g,f)$ by t^K where $K = mn + \sum_{i=1}^{m+n} i - \sum_{j=1}^{m+n} j = mn$. ∎

21 Notes on Riemannian Hypersurfaces

Let us summarize some facts about Riemannian Geometry:

Tangent Vectors $\mathfrak{r}_j = \partial \mathfrak{r} / \partial x^j$, $\mathfrak{a} = a^i \mathfrak{r}_i$, second derivatives: $\mathfrak{r}_{ij} = \frac{\partial^2 \mathfrak{r}}{\partial x^i \partial x^j}$
Metric: $g_{ij} = \mathfrak{r}_i \cdot \mathfrak{r}_j = g_{ji}$ (symmetric).
Dot product: $\mathfrak{a} \cdot \mathfrak{b} = g_{ij} a^i b^j$.

Consider the determinant

$$g = \det |g_{ij}| = \begin{vmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ g_{i1} & g_{i2} & \cdots & g_{in} \\ \cdots & \cdots & \cdots & \cdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{vmatrix} \tag{23}$$

We suppose, here and throughout, that g is not zero. Let Δ^{ij} be the cofactor of g_{ij} in this determinant, so that

$$g_{mr}\Delta^{ms} = g_{rm}\Delta^{sm} = \delta^s_r\, g,$$

which follows from the ordinary rules for developing a determinant.

Let us construct new quantities g^{ij}. The values of the components of g^{ij} are equal to the cofactor of the g_{ij}, divided by the full determinant g,

$$g^{kl} = g^{-1}\Delta^{kl}. \tag{24}$$

Alternatively the cofactor of g_{ij} can be expressed in terms of g^{ij} as folllows

$$\Delta^{ij} = g\, g^{ij}. \tag{25}$$

g^{ij} satisfies the following equations:

$$g_{ik}g^{kl} = \delta^l_i \quad \text{inverse} \tag{26}$$

Now g_{ij} is symmetric and we can also show that g^{ij} is symmetric. Multiply Eq. (26) by $g_{ls}g^{ir}$. The left hand side becomes

$$g_{ik}g^{kl}g_{ls}g^{ir} = g_{ki}g^{ir}g^{kl}g_{sl} = \delta^r_k g^{kl}g_{sl} = g_{sl}g^{rl},$$

while the right hand side becomes

$$\delta^l_i g_{ls}g^{ir} = g_{is}g^{ir} = g_{si}g^{ir} = \delta^r_s.$$

Therefore we obtain

$$g_{sl}g^{rl} = \delta^r_s \quad \text{inverse} \tag{27}$$

Comparing Eqs (26) and (27), we find that

$$g^{kl} = g^{lk}$$

Since g_{ij} is symmetric, it is obvious that g^{ij} should be symmetric also.

Let us examine the implications of the above applied to surfaces embedded in $\mathbb{R}^n$. A *curve* is defined as the totality of points given by the equations

$$x^r = f^r(u) \quad (r = 1, 2, ..., n). \tag{28}$$

Here u is a parameter and $f^r(u)$ are n functions.

Next consider the totality of points given by

$$x^r = f^r(t^1, t^2, ..., t^m) \quad (r = 1, 2, ..., n), \tag{29}$$

where the t's are parameters and $m < n$. This set of points forms V_m, a subspace of V_n. According to the Implicit Function Theorem it is possible to eliminate the parameters from (29). As long as the system is differentiable enough and the the Jacobian of the first m equations does not vanish, we can find m functions $g^i(x^1, ..., x^m)$ such that

$$t^i = g^i(x^1, ..., x^m) \text{ for } i = 1, ..., m \tag{30}$$

which implies that the remaining functions $x^r, r = m+1, ..., n$ can be written in terms of the first m coordinates $x^1, ..., x^m$ as follows:

$$x^r = f^r(x^1, ..., x^m) \text{ for } r = m+1, ..., n \tag{31}$$

In the case of $m = n - 1$ there is only one such relation and so, elimination gives just one equation

$$\phi(x^1, x^2, ..., x^n) = x^n - f^n(x^1, ..., x^m) = 0. \tag{32}$$

In this case, V_{n-1} divides portions of V_n into two parts for which respectively ϕ is positive and negative. V_{n-1} is commonly referred to as a *hypersurface* in V_n.

Let V_n be a region of $\mathbb{R}^n$, bounded by a smooth (i.e. sufficiently differentiable) surface B_{n-1}, which is an $n-1$ dimensional closed manifold. We take B_{n-1} as an example of a hypersurface V_{n-1}. We has suppose that B_{n-1} is parameterized by $n-1$ independent parameters $t^\alpha (\alpha = 1, 2, ..., n-1)$. Let the hypersurface B_{n-1} be defined implicitly by the equation (32).

On the surface of B_{n-1}, $x^i = x^i(t^\alpha)$, and therefore

$$\frac{\partial \phi}{\partial x^j} \frac{\partial x^j}{\partial t^\alpha} = 0. \tag{33}$$

Equation (33) may be regarded as $n-1$ linear conditions upon the n quantities $\partial\phi/\partial x^j$. Since the matrix

$$\left[\frac{\partial x^j}{\partial t^a}\right]$$

is assumed to have its full rank value $n-1$ on B_{n-1}, the mutual ratios of the partials $\partial\phi/\partial x^j$ are fully determined by (33).

For future reference let us define a set of determinants with respect to the change of variables between the x^i and the t^α

$$D_k = (-1)^{k-1} \det \left[\frac{\partial(x^1, ..., x^{k-1}, x^{k+1}, ..., x^n)}{\partial(t^1, ..., t^{n-1})}\right], \quad k = 1, 2, ..., n \tag{34}$$

The n determinants D_k satisfy the $n-1$ linear conditions

$$D_k \frac{\partial x^k}{\partial t^\alpha} = 0, \tag{35}$$

which can be established by noting that the left hand side can be written as an n-by-n determinant which has two rows the same and therefore vanishes. In view of (33, conditions (35) imply that the $\partial\phi/\partial x^k$ and D_k are proportional

$$\frac{\partial \phi}{\partial x^k} = \alpha D_k,$$

for some scalar factor of proportionality α. Hence the $\partial\phi/\partial x^k$ determine the unit normal to B_{n-1}, namely

$$n_k = \frac{1}{\sqrt{(\nabla\phi)^2}} \frac{\partial \phi}{\partial x^k} \tag{36}$$

$$n_k = \frac{D_k}{\sqrt{g^{ij} D_i D_j}} = \frac{D_k}{D}, \tag{37}$$

where

$$D^2 = g^{ij} D_i D_j = g^{ij} \frac{\partial \phi}{\partial x^i} \frac{\partial \phi}{\partial x^j} \alpha^{-2}.$$

22 Hypersurfaces and Concepts of Area

Consider an $n-1$ dimensional hypersurface, B_{n-1}, embedded in a Riemannian manifold M_n (which we assume to be sufficiently differentiable). Let $x^1, ..., x^n$ be local coordinates in M_n and suppose that B_{n-1} is represented locally in the form

$$x^j = x^j(t^1, ..., t^{n-1}), \quad j = 1, ..., n$$

The metric tensor of M_n is denoted by g_{ij}. and the metric tensor on B_{n-1} by $\overline{g}_{\alpha\beta}$. These two tensors are related by

$$\overline{g}_{\alpha\beta} = g_{ij} \frac{\partial x^i}{\partial t^\alpha} \frac{\partial x^j}{\partial t^\beta} = \frac{\partial x^i}{\partial t^\alpha} g_{ij} \frac{\partial x^j}{\partial t^\beta} \tag{38}$$

The element of area (or volume) in M_n is

$$dV = \sqrt{g}\, dx^1 ... dx^n, \tag{39}$$

where g is the determinant of the metric tensor in M_n. Similarly the element of area (volume) in the hypersurface B_{n-1} is given by

$$dS = \sqrt{\overline{g}}\, dt^1 ... dt^{n-1}. \tag{40}$$

We will use (38) to define the element of area in B_{n-1} based on expansion of the determinant of $\overline{g}$ in terms of products of the various minors in a manner similar to the Cauchy-Binet formula or the Laplace expansion of a determinant. Let S_n stand for the set of permutations of the numbers $\{1, 2...., n\}$ and let P represent one of the members of S_n.

$\overline{g}_{\alpha\beta}$ is an (n-1)-by-(n-1) matrix and we can consider its determinant.

$$\overline{g} = \det |\overline{g}| = \sum_{P \in S_{n-1}} \epsilon_{p_1 p_2 ... p_{n-1}} \overline{g}_{1p_1} \overline{g}_{2p_2} ... \overline{g}_{n-1 p_{n-1}} \tag{41}$$

We can think of the right hand side of (38) as representing the matrix product of three matrices.

$$\begin{aligned} J &= \left[\frac{\partial x^i}{\partial t^\alpha}\right] \text{ is } n\text{-by-}(n-1), \\ G &= \Big[g\Big] \text{ is } n\text{-by-}n \end{aligned}$$

and (38) can be written in the form:

$$\begin{aligned} \overline{g}_{\alpha\beta} &= \sum_{i,j=1}^{n} \left[\frac{\partial x^i}{\partial t^\alpha}\right]^T_{\alpha i} [g]_{ij} \left[\frac{\partial x^i}{\partial t^\beta}\right]_{j\beta} = \frac{\partial x^i}{\partial t^\alpha} g_{ij} \frac{\partial x^j}{\partial t^\beta} \text{ summation convention on } i \text{ and } j \\ \left[\overline{g}_{\alpha\beta}\right] &= J^T G J \end{aligned}$$

Inserting this computation into (41) we have $\overline{g}_{\alpha\beta}$ is an (n-1)-by-(n-1) matrix and we can consider its determinant. Let $\overline{g} = \det |\overline{g}_{\alpha\beta}|$.

$$
\begin{aligned}
\overline{g} &= \sum_{P\in S_{n-1}} \epsilon_{p_1 p_2 \ldots p_{n-1}} \sum_{\substack{i1,..,i_{n-1}=1 \\ k_1,\ldots,k_{n-1}=1}}^{n} (J^T_{1i_1} G_{i_1 k_1} J_{k_1 p_1})(J^T_{2i_2} G_{i_2 k_2} J_{k_2 p_2}) \ldots (J^T_{n-1 i_{n-1}} G_{i_{n-1} k_{n-1}} J_{k_{n-1} p_{n-1}}) \\
&= \sum_{P\in S_{n-1}} \epsilon_{p_1 p_2 \ldots p_{n-1}} \sum_{\substack{i1,..,i_{n-1}=1 \\ k_1,\ldots,k_{n-1}=1}}^{n} (J^T_{1i_1} \ldots J^T_{n-1 i_{n-1}})(G_{i_1 k_1} \ldots G_{i_{n-1} k_{n-1}})(J_{k1 p_1} \ldots J_{k_{n-1} p_{n-1}}) \\
&= \sum_{\substack{i1,..,i_{n-1}=1 \\ k_1,\ldots,k_{n-1}=1}}^{n} (J^T_{1i_1} \ldots J^T_{n-1 i_{n-1}})(G_{i_1 k_1} \ldots G_{i_{n-1} k_{n-1}}) \sum_{P\in S_{n-1}} \epsilon_{p_1 p_2 \ldots p_{n-1}} (J_{k1 p_1} \ldots J_{k_{n-1} p_{n-1}})
\end{aligned}
$$

Examining the last term in the sum on the right hand side we find a sum over the $n-1$ set $\{p1, p2, ..., p_{n-1}\}$ for a particular choice of index values: $\{k1, ...k_{n-1}\}$

$$
\sum_{P\in S_{n-1}} \epsilon_{p_1 p_2 \ldots p_{n-1}} (J_{k1 p_1} \ldots J_{k_{n-1} p_{n-1}}) = \epsilon_{k_1 k_2 \ldots k_{n-1}} |J_k| = \epsilon_{k_1 k_2 \ldots k_{n-1}} (-1)^{1+k} D_k \tag{42}
$$

where the index k is not included in $\{k1, ...k_{n-1}\}$ and final result involves the respective determinant D_k.

At this point notice that the sum over the $n-1$ variables $k1, k2, ..., k_{n-1}$ involves summing over all values from 1 to n for each k_i, but that these values must be distinct. If $k_i = k_j$ for some pair i, j, then the sum is zero because of the presence of the permutation symbol $\epsilon_{k_1 k_2 \ldots k_{n-1}}$. This means that there are $n-1$ distinct k values chosen from the full set of n values, therefore one of the k_i values is to be omitted in this sum. Let k stand for this missing value and J_k stand for the matrix J with the k-th row deleted. The set of k_i values in any set is also propagated on to the G_{ik} matrices and the corresponding k-th column of G will not contribute this particular set of choices. There will be n such cases and therefore we insert a sum over $k = 1, 2, ..., n$ to account for each case in the following. The sum over the $i_1, ..., i_{n-1}$ indices follows that same way.

$$
\begin{aligned}
\overline{g} &= \sum_{k=1}^{n} \sum_{\substack{i1,..,i_{n-1}=1 \\ k_1,\ldots,k_{n-1}=1,\neq k}}^{n} (J^T_{1i_1} \ldots J^T_{n-1 i_{n-1}})(G_{i_1 k_1} \ldots G_{i_{n-1} k_{n-1}}) \epsilon_{k_1 k_2 \ldots k_{n-1}} (-1)^{1+k} D_k \\
&= \sum_{i,k=1}^{n} \sum_{i1,..,i_{n-1}=1,\neq i}^{n} (J^T_{1i_1} \ldots J^T_{n-1 i_{n-1}}) \epsilon_{i_1 i_2 \ldots i_{n-1}} (-1)^{i+k} \text{cofactor}(g_{ik}) (-1)^{1+k} D_k \\
&= \sum_{i,k=1}^{n} \sum_{i1,..,i_{n-1}=1,\neq i}^{n} \epsilon_{i_1 i_2 \ldots i_{n-1}} (J^T_{1i_1} \ldots J^T_{n-1 i_{n-1}}) (-1)^{i+k} \Delta^{ik} (-1)^{1+k} D_k \\
&= \sum_{i,k=1}^{n} (-1)^{1+i} D_i (-1)^{i+k} \Delta^{ik} (-1)^{1+k} D_k \\
&= \sum_{i,k=1}^{n} D_i \, g \, g^{ik} \, D_k = g \sum_{i=1}^{n} D^k \, D_k = g \, D^2,
\end{aligned}
$$

where $g = |g_{ij}|$ as before and the sum is over the n values of i and k which correspond to the rows or columns which are crossed out to form the n minors each of dimension (n-1)-by-(n-1). Hence we find that

$$
\overline{g} = g \, D^2.
$$

We finally derive the relationship between the element of surface are on the hyperplane H and the metric of the n-dimensional Riemann M_n space that H is embedded inside.

$$dS = \sqrt{\overline{g}}\, dt^1 ... dt^{n-1} = \sqrt{g\, D^2}\, dt^1 ... dt^{n-1} \tag{43}$$

If the n-1 dimensional hypersurface H which is embedded in Cartesian $\mathbb{R}^n$ is given in parametric form as follows:

$$\mathfrak{r}(x) = \{x, f(x)\}$$

Then $g_{ij} = I_n$ (n-by-n identity matrix) and J is an n-by-$n-1$ matrix:

$$J = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ \frac{\partial f}{\partial x^1} & \frac{\partial f}{\partial x^2} & \dots & \frac{\partial f}{\partial x^{n-1}} \end{bmatrix} \tag{44}$$

$$J^T = \begin{bmatrix} 1 & 0 & \dots & 0 & \frac{\partial f}{\partial x^1} \\ 0 & 1 & \dots & 0 & \frac{\partial f}{\partial x^2} \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 1 & \frac{\partial f}{\partial x^n} \end{bmatrix} \tag{45}$$

Working out the minors, we find $D_n = 1$ (cross out the last row of J), and crossing out any of the other rows $i < n$ inserts a 0 in the i-th diagonal and shifts all the other rows $i+1, ..., n$ up one row. The resulting determinant is equal to $D_k = (-1)^{1+k} \frac{\partial f}{\partial x_k}$
Therefore we obtain

$$\overline{g} = \sum_{k=1}^{n} D_k = (\frac{\partial f}{\partial x^1})^2 + (\frac{\partial f}{\partial x^2})^2 + ... + (\frac{\partial f}{\partial x^n})^2 + 1 \tag{46}$$

And the resulting magnitude, given that $g_{ij} = I_n$ is $\sqrt{D^2} = \sqrt{1 + |\nabla f|^2}$ and

$$dS = \sqrt{1 + |\nabla f|^2}\ dt^1 ... dt^{n-1}. \tag{47}$$

23 Application to the Green Theorem

In Riemannian Geometry the divergence of a vector $\mathfrak{b}$ is given by

$$\nabla \cdot \mathfrak{b} = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} (\sqrt{g}\, b^i) \text{ summation convention}$$

Let us form the integral of the divergence of $\mathfrak{b}$ over a region R in V_n bounded by the closed surface B_{n-1}

$$\int_R \nabla \cdot \mathfrak{b}\, dV = \int_R \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} (\sqrt{g}\, b^i)\, dV = \int_R \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} (\sqrt{g}\, b^i)\, \sqrt{g}\, dx^1 ... dx^n \tag{48}$$

Now let us integrate the differentiated terms over the range of the appropriate variable for each one. We obtain an integral over the boundary surface B_{n-1}: for the k term this is (k not summed)

$$\int_B \sqrt{g}\, b^k dx^1 ... dx^{k-1} dx^{k+1} ... dx^n = \int_B \sqrt{g}\, b^k D_k\, dt^1 ... dt^{n-1}, \tag{49}$$

where D_k is given as in (34) represents the transformation of the integration variables from the x^i to the coordinates as given on B_{n-1}.

From the above set of equations we see that

$$\begin{aligned}\int_{R_n} \nabla \cdot \mathfrak{b}\, dV_n &= \int_{B_{n-1}} b^k D_k \sqrt{g}\, dt^1 ... dt^{n-1} \\ &= \int_{B_{n-1}} b^k n_k \sqrt{g}\, D\ dt^1 ... dt^{n-1} \\ &= \int_{B_{n-1}} \mathfrak{b} \cdot \mathfrak{n}\, dS,\end{aligned}$$

where dS is given by (43).

24 Extension and Stokes' Theorem

In a space of n dimensions, the **Generalized Kronecker delta** $\delta^{k_1\, k_2\, \cdots\, k_m}_{s_1\, s_2\, \ldots\, s_m}$ for any positive integer m is defined as

$\delta^{k_1\, k_2\, ...k_m}_{s_1\, s_2\, ...s_m} = +1$ if $k_1\, k_2\, ...k_m$ are distinct integers selected from the range $1, 2, ..., n$, and if $s_1, s_2, ..., s_m$ is an *even* permutation of $k_1\, k_2\, ...k_m$.

$\delta^{k_1\, k_2\, ...k_m}_{s_1\, s_2\, ...s_m} = -1$ if $k_1\, k_2\, ...k_m$ are distinct integers selected from the range $1, 2, ..., n$, and if $s_1, s_2, ..., s_m$ is an *odd* permutation of $k_1\, k_2\, ...k_m$.

$\delta^{k_1\, k_2\, ...k_m}_{s_1\, s_2\, ...s_m} = 0$ if any two $k_1\, k_2\, ...k_m$ are, equal, or if any two $s_1, s_2, ..., s_m$ are equal, or if the set of numbers $\{k_1, k_2, ..., k_m\}$ differs, apart from order, from the set $\{s_1, s_2, ..., s_m\}$.

In a space of n dimensions, in which no metric is assigned, consider a subspace V_m of m dimensions $m \leq n$ defined by the parametric equations

$$x^k = x^k(y^1, y^2, ..., y^m),\ \ k = 1, 2, ..., n.$$

In this V_m consider a certain region of R_m. Let us divide R_m into cells by m families of surfaces

$$f^\alpha(y) = c^\alpha,\ \ \alpha = 1, 2, ..., m.$$

where c^α are constants, each taking on a number of discrete values and so forming a family of surfaces.

It is easy to attach a meaning to the statement that a point P of V_m lies in a certain cell, because P may be said to lie between two surfaces of the same family, say family α, if the expression $f^\alpha(y) - c^\alpha$ changes sign when we change c^α from the value belonging to one of these surfaces to the value belonging to the other. Here y refers to the parameter values at the point P.

Now pick one of the cells and let A be the point corresponding to one of the corners in the cell. Through A there passes one surface of each family. Let the corresponding constants be $c^1, c^2, ..., c^m$. Passing along the edge for which c^1 alone changes, we arrive at another corner B_1 for which the constants have values $c^1 + \Delta c^1, c^2, ..., c^m$. In passing from A to B_1 the parameters change from y^α at A to $y^\alpha + \Delta_1 y^\alpha$ at B_1 and the coordinates from x^k to $x^k + \Delta_1 x^k$.

Similarly the parameter values and the coordinates of all corners $B_1, B_2, ..., B_m$ of the cell reached by traveling from A along an edge of the cell can be written as

$$y^\alpha + \Delta_\beta y^\alpha,$$
$$x^\alpha + \Delta_\beta x^k.$$

and we form the determinants

$$\Delta^{k_1 \dots k_m} = \begin{vmatrix} \Delta_1 x^{k_1} & \Delta_1 x^{k_2} & \dots & \Delta_1 x^{k_m} \\ \Delta_2 x^{k_1} & \Delta_2 x^{k_2} & \dots & \Delta_2 x^{k_m} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \Delta_m x^{k_1} & \Delta_m x^{k_2} & \dots & \Delta_m x^{k_m} \end{vmatrix} \tag{50}$$

Introducing the generalized Kronecker delta $\delta^{k_1\, k_2 \dots k_m}_{s_1\, s_2 \dots s_m}$, these determinants can be written compactly using the Summation Convention

$$\Delta^{k_1 \dots k_m} = \delta^{k_1\, k_2 \dots k_m}_{s_1\, s_2 \dots s_m}\, \Delta_1 x^{s_1}\, \Delta_2 x^{s_2} \dots \Delta_m x^{s_m} \tag{51}$$

Now if the edges had been taken in a different order, then the various determinants might have opposite signs, but apart from this the determinants are independent of the order taken. Also if we have used a different corner to start from the determinants formed similarly would have different values. But since we are going to be working with infinitesimal, the differences will be of higher order than the determinants in (51). Using differential displacements, $d_\beta x^i$ (where β indexes the β-th edge of the cell) we have

$$d\tau_m^{k_1 \dots k_m} = \delta^{k_1\, k_2 \dots k_m}_{s_1\, s_2 \dots s_m}\, d_1 x^{s_1}\, d_2 x^{s_2} \dots d_m x^{s_m} \tag{52}$$

The set of quantities $d\tau_m^{k_1 \dots k_m}$ is called the *extension* of the infinitesimal m-cell. No metrical concepts have entered into the definition of extension. Extension is the nearest concept closest to the idea of elementary volume one can define in a non-metrical space.

The edges of the infinitesimal m-cell can be written as

$$d_\beta x^k = \frac{\partial x^k}{\partial y^\alpha}\, d_\beta y^\alpha \tag{53}$$

25 Applications In Mechanics

Definition Lie Algebras. Here $F = \mathbb{R}$ or $\mathbb{C}$. A Lie Algebra of F is a pair $(\mathfrak{g}, [\cdot, \cdot])$, where $\mathfrak{g}$ is a vector space over F and

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$

is an F-blinear map satisfying the following properties

$$X, Y] = -[Y, X]$$
$$[X, [Y, Z]]] + [Z, [X, Y]]] + [Y, [Z, X]]] = 0,$$

The later is the Jacobi identity.

Definition Structure Constants: Given the elements X_a, X_b and X_c in a Lie Algebra we define the Structure Constraints c_{ab}^c to be the coefficients in the expansion of the Lie Bracket:

$$[X_a, X_b] = \sum_c c_{ab}^c \, X_c.$$

References

[1] O. Schreier and F. Sperner,"Introduction to Modern Algebra and Matrix Theory", Second Edition, Chelsea Publishing, (1959).