1. Some Basic Mathematics

1.1. The Space R^n and Its Topology

The space R^n is the set of all n-tuples of real numbers $(x_1, \cdots, x_n)$. It is the prototype of an n-dimensional real vector space.

The **distance function** between any pair of points **x** and **y** is any mapping $d : R^n \times R^n \to R$ such that

(a) $d(x, y) \geq 0$. [positivity]

(b) $d(x, y) = 0$ iff $x = y$.

(c) $d(x, y) = d(y, x)$. [symmetry]

(d) $d(x, z) \leq d(x, y) + d(y, z)$. [Triangular inequality]

For example, the (Euclidean) **distance function** between any pair of points $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$ in R^n is defined as

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} \qquad (1.1)$$

A **neighborhood** of radius r of the point **x** in R^n is the (open) set of points

$$N_r(\mathbf{x}) = \{ \mathbf{y} \mid d(\mathbf{x}, \mathbf{y}) < r \}$$

A set of points in R^n is **discrete** if each point has a neighborhood which contains no other points of the set.

A set of points S of R^n is **open** if every point **x** in S has a neighborhood entirely contained in S.
Discrete sets are therefore not open.

A simple example of an open set in R is the open interval $S = (a, b) = \{x \mid a < x < b\}$.

Note that $[a, b) = \{x \mid a \leq x < b\}$ is not an open set.

The R^n space is continuous since a line joining any 2 points in it can be infinitely

subdivided. Equivalently, we say that any 2 distinct points in R^n have neighborhoods that do not intersect. This is called the **Hausdorff property** of R^n.
The distance function $d(\mathbf{x},\mathbf{y})$ induced a **topology** on R^n since the open sets O_i it defines satisfy the axioms of a **topological space**,

(T.i.) $\bigcap_{i=1}^{N} O_i$, with N finite, is open.

(T.ii.) $\bigcup_{i=1}^{\infty} O_i$ is open.

Note that, by definition, the empty set Φ and R^n are open.

It can be shown that all distance functions induce the same topology on R^n. This is called the **natural** topology of R^n.

The concept of open set is more general than the distance function. Thus, one can define neighborhood in terms of open sets for spaces in which the concept of distance is not defined.

1.2. Mappings

A map *f* from a space *M* to a space *N* is denoted as

$$f : M \to N \qquad \text{with} \qquad x \mapsto f(x)$$

It associates an element $x \in M$ with exactly one element $f(x) \in N$.

It is permissible to have many points in *M* mapping to the same point in *N*, i.e., $f(x_1) = f(x_2) = \cdots$, the map is then called **many-to-one**.

For any subset $S \subseteq M$, the set $T = f(S) \subseteq N$ is called the **image** of *S* under *f*.

Conversely, the set $S = f^{-1}(T)$ is called the **inverse image** of *T*. Note that $f^{-1}(T)$ should in general be read as a single symbol since the map f^{-1} may not exist. If every point in the map has a distinct image in *N*, the mapping is called **one-to-one** (1-1). In which case, f^{-1} exists and is called the **inverse** of *f*.

Consider 2 maps

$$f : M \to N \qquad \text{and} \qquad g : N \to P$$

with $\quad x \mapsto y = f(x) \qquad \text{and} \qquad y \mapsto z = g(y)$.

The **composite** of *f* and *g* is the mapping

$$g \circ f : M \to P$$

with $\quad x \mapsto z = (g \circ f)(x) = g\left[f(x) \right]$

If a map is defined for all points of *M*, then *f* maps *M* **into** *N*. If every point in *N* has at least one inverse image, then *f* maps *M* **onto** *N*. If *f* is 1-1 onto, it is **bijective**.

Assuming *M*, *N* to be topological spaces, the map $f : M \to N$ is **continuous** at a point $x \in M$ if any open set of *N* containing *f*(*x*) contains the image of an open set of *M*.

Let *f* be a 1-1 map of an open set $M \in R^n$ onto another open set $N \in R^n$. Thus,

$$f : M \to N \qquad \text{with} \qquad \mathbf{x} \mapsto \mathbf{y} = \mathbf{f}(\mathbf{x})$$

where $y_i = f_i\left(x_1, \cdots, x_n\right)$. If the functions f_i are all C^k-differentiable, the map is said to be C^k-differentiable. The **Jacobian matrix** of a C^1 map is defined as the matrix with the (*i*, *j*) element equal to $\frac{\partial f_i}{\partial x_j}$. The determinant of this matrix is called the **Jacobian**

$$J = \frac{\partial\left(f_1, \cdots, f_n\right)}{\partial\left(x_1, \cdots, x_n\right)} = \det\left|\frac{\partial f_i}{\partial x_j}\right| \qquad (1.5)$$

If $J \neq 0$ at **x**, the **inverse function theorem** then says *f* is 1-1 onto in some neighborhood of **x**.

If a function $g\left(\mathbf{x}\right) = g\left(x_1, \cdots, x_n\right)$ is mapped into a function $g_*\left(\mathbf{y}\right) = g_*\left(y_1, \cdots, y_n\right)$ by the rule

$$g_*\left[\mathbf{f}\left(\mathbf{x}\right)\right] = g_*\left[f_1\left(x_1, \cdots, x_n\right), \cdots, f_n\left(x_1, \cdots, x_n\right)\right] = g\left(x_1, \cdots, x_n\right) = g\left(\mathbf{x}\right)$$

then the integral of *g* over *M* is equal to the integral of *g***J* over *N*, i.e.,

$$\int_M dx_1 \cdots dx_n \; g\left(x_1, \cdots, x_n\right) = \int_N dy_1 \cdots dy_n \; J \; g_*\left(y_1, \cdots, y_n\right) \qquad (1.6)$$

1.3. Real Analysis

1.3.1. Analyticity

A real function $f(x)$ is **analytic** at $x = x_0$ if it has a Taylor expansion in some neighborhood of x_0, i.e.,

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n f}{dx^n} \right|_{x=x_0} (x - x_0)^n \tag{1.7}$$

Note that infinite differentiability ($f \in C^{\infty}$) is a necessary, but not sufficient, condition for analyticity ($f \in C^{\omega}$). Thus, $f \in C^{\omega} \to f \in C^{\infty}$ but the converse is not necessarily true. For example, all orders of derivatives of $\exp\left(-\frac{1}{x^2}\right)$ exist and equal to 0 at $x = 0$. Hence, the Taylor series in (1.7) with $x_0 = 0$ is identically zero for arbitrary *x*. On the other hand, $\exp\left(-\frac{1}{x^2}\right) \neq 0$ for $x \neq 0$. Therefore, $\exp\left(-\frac{1}{x^2}\right)$ is infinitely differentiable but non-analytic at $x = 0$. One consequence of this is that the complex function $\exp\left(-\frac{1}{z^2}\right)$ has an essential singularity at $z = 0$ even though it is well behaved on the real line.

A real- valued function $g(\mathbf{x}) = g(x_1, \cdots, x_n)$ defined on an open region $S \in R^n$ is said to be **square- integrable** if the integral

$$\int_S d^n x \left[g(\mathbf{x}) \right]^2 = \int_S dx_1 \cdots dx_n \left[g(x_1, \cdots, x_n) \right]^2 \tag{1.8}$$

exists. A basic theorem in functional analysis says that any square- integrable function *g* can be approximated by another analytic function *g'* so that $\int_S d^n x (g - g')^2$ is arbitrarily small. For this reason, most functions encountered in physical problems can be assumed to be analytic without incurring serious errors.

1.3.2. Operators

An **operator** *A* on functions $f(\mathbf{x})$ defined on R^n maps each function *f* into another function $A(f)$. For example, one can define a multiplicative operator by

$$A(f) = g\ f \qquad \text{so that} \qquad [A(f)](\mathbf{x}) = g(\mathbf{x})\ f(\mathbf{x})$$

where *g* is some function defined on R^n. Some differential operators on $f(x)$ defined on *R* are

$$D(f) = \frac{df}{dx} \qquad \text{so that} \qquad [D(f)](x) = \frac{df(x)}{dx}$$

$$E(f) = f^2 + \frac{d^2 f}{dx^2} \qquad \text{so that} \qquad [E(f)](x) = [f(x)]^2 + \frac{d^2 f(x)}{dx^2}$$

Given a fixed kernel function *g*, one can define an integral operator $G(f)$ by the convolution

$$[G(f)](x) = \int_{x_0}^{x} dy\ g(x,y)\ f(y)$$

Note that the full definition of an operator should include a specification of its **domain**. For example, the domain of *D* is the set $f \in C^1$. That of *E* is $f \in C^2$. That of *G* is all *f* such that the integrals $\int_{x_0}^{x} dy\ g(x,y)\ f(y)$ are finite.

1.3.3. Commutators

The **commutator** of 2 operators *A* and *B* is another operator defined by

$$[A,B] = AB - BA$$

so that

$$[A,B](f) = (AB - BA)(f) \quad = A\left[B(f)\right] - B\left[A(f)\right] \qquad (1.9)$$

for arbitrary functions $f(\mathbf{x})$ defined on R^n. *A* and *B* are said to **commute** if $[A,B] = 0$. Note that the domain of $[A,B]$ is in general different from those of *A* and *B*. For example, if $A = \frac{d}{dx}$ and $B = x\frac{d}{dx}$, then the domains of *A* and *B* are $f \in C^1$ but that of $[A,B]$ is $f \in C^2$. On the other hand,

$$\left[\frac{d}{dx}, x\frac{d}{dx}\right](f) = \frac{d}{dx}\left(x\frac{df}{dx}\right) - x\frac{d}{dx}\left(\frac{df}{dx}\right) = \frac{df}{dx}$$

or $$\left[\frac{d}{dx}, x\frac{d}{dx}\right] = \frac{d}{dx}$$

so that the domain of $[A,B]$ is effectively $f \in C^1$.

1.4. Group Theory

1.4.1. Definition

A collection of elements G together with a binary operation, $\cdot$, is called a **group** $(G, \cdot)$ if for all $x, y, z \in G$, we have

(G1) **Closure** so that $x \cdot y \in G$.

(G2) **Associativity** so that $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.

(G3) A **right identity** $e \in G$ so that $x \cdot e = x$.

(G4) **Right inverse** $x^{-1} \in G$ so that $x \cdot x^{-1} = e$.

Note that it is common practice to refer to the group $(G, \cdot)$ simply as G and to write the binary composition $x \cdot y$ simply as $x\, y$.

The group is **abelian** if, furthermore, we have

(G5) **Commutativity** so that $x \cdot y = y \cdot x$.

In which case, the binary operation is usually denoted by +.

1.4.2. Example

One example of a **finite group** (group with finite number of elements) is the group of all permutations of n objects. Thus, the binary operation denotes the successive application of 2 permutations. The **order**, or number of elements, of the group is $n!$. The identity element is the permutation that leaves all objects unchanged.

1.4.3. Corollaries

Some direct consequences of the axioms (G1-4) that can be easily proved are

1. *e* is unique.
2. The right inverse is also a left inverse.
3. The inverse x^{-1} of a given element x is unique.

1.4.4. Lie Groups

Roughly speaking, a **Lie group** is a **continuous group** such that any open set of its elements has a 1-1 map onto an open set of R^n for some n. (A precise definition will be given in Chapter 2).

One example of a Lie group is the **translation group** of R^n. Each point $\mathbf{a} \in R^n$ corresponds to an element of the group $T_{\mathbf{a}}$ with $T_{\mathbf{a}}\mathbf{x} = \mathbf{x} + \mathbf{a} \quad \forall\, \mathbf{x} \in R^n$. Hence, the group has a 1-1 map onto R^n. The binary composition is simply vector addition:

$T_{\mathbf{a}} \cdot T_{\mathbf{b}} = T_{\mathbf{a}+\mathbf{b}}$ with $T_{\mathbf{a}} \cdot T_{\mathbf{b}}\mathbf{x} = T_{\mathbf{a}}\left(\mathbf{x} + \mathbf{b}\right) \quad = \mathbf{x} + \mathbf{a} + \mathbf{b} \quad = T_{\mathbf{a}+\mathbf{b}}\mathbf{x}$.

1.4.5. Subgroups

A **subgroup** *S* of a group *G* is a subset of *G* that is itself a group under the same binary operation. Note that all subgroups of a group must contain the same identity element *e*.

Using the group of permutations of *n* objects as an example, the set of all permutations that leave a particular element unchanged forms a subgroup. Indeed, this subgroup is the same as, or **isomorphic** to, the permutation group of $n - 1$ objects. The set of all even permutations also form a subgroup but that of the odd permutations does not.

1.4.6. Isomorphisms and Homomorphisms

Two groups $(G_1, \cdot\ )$ and $(G_2, *\ )$ are **isomorphic** if there exists a 1-1 map

$$f : G_1 \to G_2 \qquad \text{with} \qquad x \mapsto f(x)$$

that preserves the group operation, i.e.,

$$f(x \cdot y) = f(x) * f(y) \qquad (1.10)$$

If *f* is many-to-1, the relation is called a **homomorphism**.

As an example, let G_1 be the set of all positive real numbers and · the multiplication ×. Let G_2 be the set of all real numbers and * the addition +. These groups are isomorphic with *f* being the logarithmic function, i.e.,

$$\ln(x \times y) = \ln x + \ln y \qquad \forall\ x, y > 0$$

A trivial homomorphism of any group is the mapping of every element into the identity *e*. Less trivial is the mapping of a permutation group into the group $(\{1, -1\}, \times\ )$ so that all even (odd) permutations are mapped into 1 (–1).

1.5. Linear Algebra

1.5.1. Vector Space

A set V is a **vector space** over the field $\mathbb{K}$ if

1. $(V,+)$ is an abelian group, i.e., for all $\overline{x}, \overline{y}, \overline{z} \in V$,
   a) $\overline{x} + \overline{y} \in V$ (closure)

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

   c) $\exists\, \overline{0} \in V$ such that $\overline{0} + \overline{x} = \overline{x} + \overline{0} = \overline{x}$ (identity / zero)

   d) $\overline{x} + \overline{y} = \overline{y} + \overline{x}$ (commutativity)
2. Scalar multiplication by $a, b \in \mathbb{K}$ satisfies
   a) $a\,\overline{x} \in V$ (closure)

   b) $a(\overline{x} + \overline{y}) = a\overline{x} + b\overline{y}$ (distributivity)

   $(a + b)\overline{x} = a\overline{x} + b\overline{x}$

   c) $(ab)\overline{x} = a(b\overline{x}) = ab\overline{x}$ (associativity)

   d) $1\,\overline{x} = \overline{x}$ where 1 is the identity of $\mathbb{K}$.

We shall always assume $\mathbb{K}$ is either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

Also, the zero vector $\overline{0}$ is often written simply as 0.

1.5.2. Examples

Some examples of vector spaces are

a) The set of all $n \times n$ matrices, with matrix addition as + and multiplication by numbers as scalar multiplication.

b) The set of all real continuous function $f(x)$ defined on the interval $a \le x \le b$ with ordinary addition and multiplication of real numbers as vector addition and scalar multiplication, respectively.

1.5.3. Linear Independence and Basis

An expression like

$$a\bar{x} + b\bar{y} + c\bar{z} \tag{1.11}$$

is called a **linear combination** of the vectors $\bar{x}, \bar{y}$ and $\bar{z}$. A set *S* of elements $\{\bar{x}_1, \cdots, \bar{x}_m\}$ of *V* is **linearly independent** if

$$\sum_{j=1}^{m} a_j \bar{x}_j = 0 \tag{1.12}$$

only if $a_j = 0$ for all *j*. Otherwise, *S* is said to be **linearly dependent**. *S* is called a **maximal set** if the inclusion of any vector makes it linearly dependent. Thus, any element in *V* can be expressed as a linear combination of the vectors in a maximal set, which is therefore a **basis** of *V*. Obviously, all bases have the same number of elements *n*, if *n* is finite. We call *n* the **dimension** of *V*. Given a basis $\{\bar{x}_1, \cdots, \bar{x}_n\}$, any vector $\bar{y} \in V$ can be written as

$$\bar{y} = \sum_{i=1}^{n} a_i \bar{x}_i \tag{1.13}$$

where the numbers $\{a_i\}$ are called the **components** of $\bar{y}$ on this basis.

1.5.4. Subspace

A **subspace** of *V* is a subset *S* of *V* that is also a vector space. Since it must be closed with respect to addition, *S* always includes the zero vector. Any set of vectors $\{\bar{y}_1, \cdots, \bar{y}_m\}$ is said to **generate** *S* is any vector $\bar{y} \in S$ can be written as

$$\bar{y} = \sum_{i=1}^{m} a_i \bar{y}_i$$

If $m < n$, *S* is necessarily a **proper subspace**. In any case, the dimension of *S* is the maximal number of independent vectors in its generator.

1.5.5. Norms

The magnitude of a vector $\overline{x}$ is given by its **norm** $n(\overline{x})$. Here, *n* is a mapping

$$n : V \to \mathbb{R}$$

which

a) is positive definite: $n(\overline{x}) \geq 0$ with $n(\overline{x}) = 0$ iff $\overline{x} = \overline{0}$.

b) is linearly scaled: $n(a\overline{x}) = |a|\, n(\overline{x})$ for all $a \in \mathbb{K}$.

c) satisfies the **triangular inequality**: $n(\overline{x} + \overline{y}) \leq n(\overline{x}) + n(\overline{y})$.

A vector space equipped with a norm is called a **normed vector space**. A commonly used notation is $n(\overline{x}) = \|\overline{x}\|$.

1.5.6. Examples of Norms

$\mathbb{R}^n$ is the prototype of a vector space. Here, vector addition is defined as

$$\mathbf{x}+\mathbf{y}=\left(x_1+y_1,\cdots,x_n+y_n\right) \tag{1.14}$$

and multiplication by scalar as

$$a\mathbf{x}=\left(ax_1,\cdots,ax_n\right) \tag{1.15}$$

Given a distance function $d\left(\mathbf{x},\mathbf{y}\right)$, we can define a norm as the distance from the origin, i.e., $n\left(\mathbf{x}\right)=d\left(\mathbf{x},\mathbf{0}\right)$. Of the four distance functions discussed in §1.1, three can be used for such purposes. Thus, corresponding to *d*, *d'* and *d'''*, we have

$$n\left(\mathbf{x}\right)=\sqrt{\left(x_1\right)^2+\cdots+\left(x_n\right)^2} \tag{1.16}$$

$$n'\left(\mathbf{x}\right)=\sqrt{4\left(x_1\right)^2+0.1\left(x_2\right)^2+\cdots+\left(x_n\right)^2} \tag{1.17}$$

$$n'''\left(\mathbf{x}\right)=\max\left(\left|x_1\right|,\cdots,\left|x_n\right|\right) \tag{1.18}$$

However,

$$d''\left(\mathbf{x},\mathbf{y}\right)=\exp\left[d\left(\mathbf{x},\mathbf{y}\right)\right]-1$$

cannot define a norm since it does not scale linearly, i.e.,

$$d''\left(a\mathbf{x},\mathbf{0}\right)\neq\left|a\right|d''\left(\mathbf{x},\mathbf{0}\right)$$

1.5.7. Inner Products

The **inner product** on a vector space *V* is a mapping

$$V \times V \to \mathbb{K} \qquad \text{by} \qquad (\bar{x},\bar{y}) \mapsto \langle \bar{x} | \bar{y} \rangle$$

that is **sesquilinear**, i.e.,

1. $\langle \bar{y} | \bar{x} \rangle = \langle \bar{x} | \bar{y} \rangle^{*}$. (1.22)

2. $\langle a\bar{x} + b\bar{y} | \bar{z} \rangle = a^{*} \langle \bar{x} | \bar{z} \rangle + b^{*} \langle \bar{y} | \bar{z} \rangle$ for all $a, b \in \mathbb{K}$. (1.21)

where * is the complex conjugate. For a real vector space, we usually write $\langle \bar{x} | \bar{y} \rangle = \bar{x} \cdot \bar{y}$. A vector space equipped with an inner product is called an **inner product space**.

If the inner product is also positive definite, i.e.,

$$\langle \bar{x} | \bar{x} \rangle \geq 0 \quad \text{with} \quad \langle \bar{x} | \bar{x} \rangle = 0 \quad \text{iff} \quad \bar{x} = \bar{0}. \qquad (1.23)$$

it can induce a distance function and a norm.

For example, an inner product induces on $\mathbb{R}^n$ a distance function

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

and a norm

$$n(\mathbf{x}) = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

called the **Euclidean norm**. Such a space is called an **Euclidean space** and denoted $\mathbb{E}^n$. Note that an Euclidean norm satisfies the **parallelogram rule**:

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right)$$

An inner product that is not positive definite can induce a **pseudo-norm** $n(\mathbf{x}) = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ that is not positive definite and viloates the triangular inequality.

Note that some author refer to the inner product defined here as a **sequilinear mapping** and reserves the term **inner product** to the contraction of the metric tensor with 2 vectors.

1.6. The Algebra of Square Matrices

1.6.1. Linear Transformations

A **linear transformation** *T* on a vector space *V* over $\mathbb{K}$ is a map

$$T : V \to V \qquad \text{by} \qquad \overline{x} \mapsto T\left(\overline{x}\right)$$

such that for all $a, b \in \mathbb{K}$ and $\overline{x}, \overline{y} \in V$,

$$T\left(a\overline{x} + b\overline{y}\right) = aT\left(\overline{x}\right) + bT\left(\overline{y}\right) \tag{1.24}$$

Given a basis $\left\{\overline{e}_i \, ; i = 1, \cdots, n\right\}$ for *V*, we have

$$\overline{x} = \sum_{i=1}^{n} \overline{e}_i x_i \tag{1.25}$$

$$T\left(\overline{x}\right) = T\left(\sum_{i=1}^{n} \overline{e}_i x_i\right) = \sum_{i=1}^{n} T\left(\overline{e}_i\right) x_i$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{e}_j T_{ji} x_i \tag{1.26$'$}$$

where we've used the fact that $T\left(\overline{e}_i\right)$ is a vector so that by (1.25), we can write

$$T\left(\overline{e}_i\right) = \sum_{j=1}^{n} \overline{e}_j T_{ji} \tag{1.26a$'$}$$

Here T_{ij} are called the **components** of *T* and can be represented as an *n*×*n* matrix **T** with elements $\left(\mathbf{T}\right)_{ij} = T_{ij}$. Note that **T** defined here is the transpose of that used by Schutz. This difference is marked by a ′ on the equation number. By (1.26′), the effects of *T* on *V* are contained entirely in **T**. Treating $\overline{x}$ as a column matrix **x**, (1.26′) can be written as a matrix equation

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

where **y** is a column matrix with elements

$$y_i = \left(\mathbf{T}\mathbf{x}\right)_i = \sum_{j=1}^{n} T_{ij} x_j \tag{1.26b$'$}$$

1.6.2. Group of Transformations

Consider 2 successive linear transformations S and T on a vector $\bar{x} \in V$:

$$ST(\bar{x}) = S\sum_{i,j} \bar{e}_i T_{ij} x_j \quad = \sum_{i,j,k} \bar{e}_k S_{ki} T_{ij} x_j$$

$$= \sum_{k,j} \bar{e}_k \left(\sum_i S_{ki} T_{ij} \right) x_j \qquad (1.29')$$

Thus, ST can be treated as a single transformation U with components

$$U_{kj} = (ST)_{kj} \quad = \sum_i S_{ki} T_{ij} \qquad (1.30')$$

which can be written as a matrix equation

$$\mathbf{U} = \mathbf{ST}$$

It is straightforward to show that matrix products are associative, i.e.,

$$(\mathbf{RS})\mathbf{T} = \mathbf{R}(\mathbf{ST}) = \mathbf{RST}$$

Thus, the set of all non-singular $n \times n$ matrices is a group using matrix product as group multiplication. Here, the identity group element is the unit matrix $\mathbf{I} = \{\delta_{ij}\}$. The inverse of group element $\mathbf{T}$ is its matrix inverse $\mathbf{T}^{-1}$ so that $\mathbf{TT}^{-1} = \mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$. The non-singular requirement is to ensure the existence of these inverses. For matrices with elements in $\mathbb{K}$, this is a Lie group called the **General Linear** group of dimension n over $\mathbb{K}$ and denoted $GL(n, \mathbb{K})$. Owing to the isomorphism (existence of 1-1 onto map, i.e., bijection) between linear transformations and matrices, this is also the group of the transformations.

1.6.3. Determinants

The **transpose** $\mathbf{A}^T$ of a matrix **A** has elements

$$\left(\mathbf{A}^T\right)_{ij} = \left(\mathbf{A}\right)_{ji} = A_{ji} \qquad (1.33)$$

If **A** is complex, the **adjoint** $\mathbf{A}^+$ of **A** has elements

$$\left(\mathbf{A}^+\right)_{ij} = \left(\mathbf{A}\right)^*_{ji} = A^*_{ji} \qquad (1.33a)$$

where * denotes the complex conjugation.

The **determinant** det(**A**) of **A** can be defined recursively by the **rule of cofactors**. To begin, let $\mathbf{M}(i, j)$ be the $(n-1)\times(n-1)$ matrix obtained by striking out the *i*th row and *j*th column of the $n\times n$ matrix **A**. The **minor** of an element a_{ij} of **A** is defined as $\det\left(\mathbf{M}(i, j)\right)$. The **cofactor** A^{ij} of a_{ij} is the signed minor, namely,

$$A^{ij} = (-)^{i+j} \det\left(\mathbf{M}(i, j)\right)$$

The determinant of **A** is then defined as the sum of the cofactors of any row of **A**,

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} A^{ij} \qquad \text{for any } i \qquad (1.38)$$

$$= \sum_{j=1}^{n} (-)^{i+j} a_{ij} \det\left(\mathbf{M}(i, j)\right)$$

Eq(1.38) is called a **row expansion** for $\det(\mathbf{A})$. It replaces the calculation of the determinant of an $n\times n$ matrix by those of $(n-1)\times(n-1)$ matrices. With the understanding that a 1×1 matrix is a number, (1.38) provides a recursive formula for $\det(\mathbf{A})$. For example expansion of the 1st row gives

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\det(d) - b\det(c) \quad = ad - bc$$

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a\det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b\det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c\det\begin{pmatrix} d & e \\ g & h \end{pmatrix}$$

$$= a(ei - fh) - b(di - fg) + c(dh - eg)$$

A better definition of determinants based on its total antisymmetry and relation to the volume element will be given in §4.12. We shall list without proof the following useful properties, (see Apostol, "Linear Algebra", for proof),

1. $\det(\mathbf{A}) = \det(\mathbf{A}^T)$.

2. $\sum_{j=1}^{n} a_{ij} A^{kj} = \det(\mathbf{A})\, \delta_{ik}$.

1.6.4. Linear Independence

Some properties of det(**A**) are

1. If any 2 rows are interchanged, det(**A**) changes sign.
2. If the elements of a row are all multiplied by the same factor λ, so is det(**A**).
3. If a row is replaced by the sum of itself and any multiple of another row, det(**A**) remains unchanged.
4. Properties (1) and (3) imply $\det\left(\mathbf{A}\right)=0$ if any row is a linear combination of the other rows.
5. The above properties remain valid if all rows are replaced by columns.

Each row (column) of an $n\times n$ matrix **A** can be taken as an n-D vector. From properties (4) and (5), we see that $\det\left(\mathbf{A}\right)\neq 0$ iff the row (column) vectors of **A** are linearly independent.

1.6.5. Inverse

The elements of the inverse of **A** can be written as

$$\left(\mathbf{A}^{-1}\right)_{ij} = \frac{1}{\det\left(\mathbf{A}\right)} A^{ji} \qquad (1.39)$$

so that

$$\left(\mathbf{A}\mathbf{A}^{-1}\right)_{ij} = \sum_k \left(\mathbf{A}\right)_{ik} \left(\mathbf{A}^{-1}\right)_{kj} \quad = \frac{1}{\det\left(\mathbf{A}\right)} \sum_k a_{ik} A^{jk} \quad = \delta_{ij}$$

1.6.6. Similarity Transforms

A **similarity transformation** of **A** by a nonsingular matrix **B** is a map

$$\mathbf{A} \mapsto \mathbf{B}^{-1}\mathbf{AB}$$

Using

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

we have

$$\det\left(\mathbf{B}^{-1}\mathbf{AB}\right) = \det\left(\mathbf{B}^{-1}\right)\det(\mathbf{A})\det(\mathbf{B}) \quad = \det\left(\mathbf{B}^{-1}\mathbf{B}\right)\det(\mathbf{A})$$

$$= \det(\mathbf{A}) \qquad (1.45)$$

The **trace** of **A** is the sum of its diagonal elements, i.e.,

$$tr(\mathbf{A}) = \sum_i a_{ii} \qquad (1.40)$$

Using

$$tr(\mathbf{AB}) = \sum_{i\,j} a_{ij} b_{ji} = tr(\mathbf{BA})$$

$$tr(\mathbf{ABC}) = \sum_{i\,j\,k} a_{ij} b_{jk} c_{ki} = tr(\mathbf{CAB}) = tr(\mathbf{BCA})$$

we have

$$tr\left(\mathbf{B}^{-1}\mathbf{AB}\right) = tr\left(\mathbf{BB}^{-1}\mathbf{A}\right) = tr(\mathbf{A}) \qquad (1.44)$$

Thus, both determinant and trace are invariant under a similarity transform.

1.6.7. Eigen Problem

The transformation equation,

$$A\left(\overline{V}\right) = \lambda \overline{V} \qquad (1.47)$$

with λ being a constant, or its matrix counterpart

$$\mathbf{Av} = \lambda \mathbf{v} \qquad (1.47a)$$

is called an **eigen-equation**. The constant λ is called an **eigenvalue** and $\overline{V}$ or **v** the corresponding **eigenvector**. In components, (1.47) becomes

$$\sum_j a_{ij} V_j = \lambda V_i \qquad \Rightarrow \qquad \sum_j \left(a_{ij} - \lambda \delta_{ij}\right) V_j = 0 \qquad (1.48)$$

which admits a non-zero eigenvector iff

$$\det\left(\mathbf{A} - \lambda \mathbf{I}\right) = 0 \qquad (1.49)$$

Eq(1.49) is known as the **secular equation**. If **A** is an $n \times n$ matrix, (1.49) is an nth degree algebraic equation of λ. By the fundamental theorem of algebra, there is always n complex solutions which we denote by $\left\{\lambda_1, \cdots, \lambda_n\right\}$. It is easily shown that

1. Eigenvalues of **A** and $\mathbf{A}^T$ are identical.
2. $\det\left(\mathbf{A}\right) = \lambda_1 \cdots \lambda_n$.
3. $tr\left(\mathbf{A}\right) = \lambda_1 + \cdots + \lambda_n$.

1.7. Bibliography