Linear Algebra

Course 211

David Simms

School of Mathematics

Trinity College

Dublin

Original typesetting

Brendan Griffin.

Alterations

Geoff Bradley.
Table of Contents

1 Vector Spaces 1–1
  2.1 The Definition 2–1
  2.2 Basic Properties of Linear Operators 2–3
  2.3 Examples 2–4
  2.4 Properties Continued 2–8
  2.5 Operator Algebra 2–10
  2.6 Isomorphisms of $\mathcal{L}(M, N)$ with $K^{m \times n}$ 2–11
2 Linear Operators 2–1
  2.1 The Definition 2–1
  2.2 Basic Properties of Linear Operators 2–3
  2.3 Examples 2–4
  2.4 Properties Continued 2–8
  2.5 Operator Algebra 2–10
3 Changing Basis and Einstein Convention 3–1
4 Linear Forms and Duality 4–1
  4.1 Linear Forms 4–1
  4.2 Duality 4–3
  4.3 Systems of Linear Equations 4–4
5 Tensors 5–1
  5.1 The Definition 5–1
  5.2 Contraction 5–3
  5.3 Examples 5–4
  5.4 Bases of Tensor Spaces 5–5
6 Vector Fields 6–1
  6.1 The Definition 6–1
  6.2 Velocity Vectors 6–3
  6.3 Differentials 6–4
  6.4 Transformation Law 6–7
7 Scalar Products 7–1
  7.1 The Definition 7–1
  7.2 Properties of Scalar Products 7–1
  7.3 Raising and Lowering Indices 7–5
  7.4 Orthogonality and Diagonal Matrix (including Sylvester’s Law of Inertia) 7–7
  7.5 Special Spaces (including Jacobi’s Theorem) 7–12
8 Linear Operators 2
  8.1 Adjoint and Isometries 8–1
  8.2 Eigenvalues and Eigenvectors 8–4
  8.3 Spectral Theorem and Applications 8–7
9 Skew-Symmetric Tensors and Wedge Product 9–1
  9.1 Skew-Symmetric Tensors 9–1
  9.2 Wedge Product 9–5
10 Classification of Linear Operators 10–1
  10.1 Hamilton-Cayley and Primary Decomposition 10–1
  10.2 Diagonalisable Operators 10–4
  10.3 Conjugacy Classes 10–6
  10.4 Jordan Forms 10–9
  10.5 Determinants 10–18
11 Orientation 11–1
  11.1 Orientation of Vector Spaces 11–1
  11.2 Orientation of Coordinate Systems 11–6
12 Manifolds and (n)-dimensional Vector Analysis 12–1
  12.1 Gradient 12–1
  12.2 3-dimensional Vector Analysis 12–3
  12.3 Results 12–4
  12.4 Closed and Exact Forms 12–7
  12.5 Contraction and Results (including Poincaré Lemma) 12–9
  12.6 Surface in $\mathbb{R}^3$ 12–14
  12.7 Integration on a Manifold (Sketch) 12–16
  12.8 Stokes Theorem and Applications 12–18
Recall that in course 131 you studied the notion of a linear vector space. In that course the scalars were real numbers. We will study the more general case, where the set of scalars is any field $K$. For example $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/(p)$.

**Definition.** Let $K$ be a field. A set $M$ is called a vector space over the field $K$ (or a $K$-vector space) if

(i) an operation

$$M \times M \rightarrow M$$

$$(x, y) \mapsto x + y$$

is given, called *addition of vectors*, which makes $M$ into a commutative group;

(ii) an operation

$$K \times M \rightarrow M$$

$$(\lambda, x) \mapsto \lambda x$$

is given, called *multiplication of a vector by a scalar*, which satisfies:

(a) $\lambda(x + y) = \lambda x + \lambda y$,
(b) $(\lambda + \mu)x = \lambda x + \mu x$,
(c) $\lambda(\mu x) = (\lambda\mu)x$,
(d) $1x = x$

for all $\lambda, \mu \in K$, $x, y \in M$, where 1 is the unit element of the field $K$. 

1–1
The elements of $M$ are then called the *vectors*, and the elements of $K$ are called the *scalars* of the given $K$-vector space $M$.

**Examples:**

1. The set of 3-dimensional geometrical vectors (as in 131) is a real vector space ($\mathbb{R}$-vector space).
2. The set $\mathbb{R}^n$ (as in 131) is a real vector space.
3. If $K$ is any field then the following are $K$-vector spaces:
   
   (a) $K^n = \{(\alpha_1, \ldots, \alpha_n) : \alpha_1, \ldots, \alpha_n \in K\}$, with vector addition:
   
   $$(\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n) = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n),$$
   
   and scalar multiplication:
   
   $$\lambda(\alpha_1, \ldots, \alpha_n) = (\lambda \alpha_1, \ldots, \lambda \alpha_n).$$
   
   (b) The set $K^{m\times n}$ of $m \times n$ matrices ($m$ rows and $n$ columns) with entries in $K$ ($m, n$ fixed integers $\geq 1$), with vector addition:
   
   $$
   \begin{pmatrix}
   \alpha_{11} & \cdots & \alpha_{1n} \\
   \vdots & & \vdots \\
   \alpha_{m1} & \cdots & \alpha_{mn}
   \end{pmatrix} + 
   \begin{pmatrix}
   \beta_{11} & \cdots & \beta_{1n} \\
   \vdots & & \vdots \\
   \beta_{m1} & \cdots & \beta_{mn}
   \end{pmatrix} = 
   \begin{pmatrix}
   \alpha_{11} + \beta_{11} & \cdots & \alpha_{1n} + \beta_{1n} \\
   \vdots & & \vdots \\
   \alpha_{m1} + \beta_{m1} & \cdots & \alpha_{mn} + \beta_{mn}
   \end{pmatrix},
   $$
   
   and scalar multiplication:
   
   $$\lambda
   \begin{pmatrix}
   \alpha_{11} & \cdots & \alpha_{1n} \\
   \vdots & & \vdots \\
   \alpha_{m1} & \cdots & \alpha_{mn}
   \end{pmatrix} = 
   \begin{pmatrix}
   \lambda \alpha_{11} & \cdots & \lambda \alpha_{1n} \\
   \vdots & & \vdots \\
   \lambda \alpha_{m1} & \cdots & \lambda \alpha_{mn}
   \end{pmatrix}.$$
   
   (c) The set $K^X$ of all maps from $X$ to $K$ ($X$ a fixed non-empty set), with vector addition:
   
   $$(f + g)(x) = f(x) + g(x),$$
   
   and scalar multiplication:
   
   $$(\lambda f)(x) = \lambda f(x)$$
   
   for all $x \in X$, $f, g \in K^X$, $\lambda \in K$.  

1-2
Definition. Let $N \subseteq M$, and let $M$ be a $K$-vector space. Then $N$ is called a $K$-vector subspace of $M$ if $N$ is non-empty, and

(i) $x, y \in N \Rightarrow x + y \in N$ closed under addition;

(ii) $\lambda \in K, x \in N \Rightarrow \lambda x \in N$ closed under scalar multiplication.

Thus $N$ is itself a $K$-vector space.

Examples:

1. $\{(\alpha, \beta, \gamma) : 3\alpha + \beta - 2\gamma = 0; \; \alpha, \beta, \gamma \in \mathbb{R}\}$ is a vector subspace of $\mathbb{R}^3$.

2. $\{v : v \cdot n = 0\}$, $n$ fixed, is a vector subspace of the space of 3-dimensional geometric vectors (see Figure 1.1).

3. The set $C^0(\mathbb{R})$ of continuous functions is a real vector subspace of the set $\mathbb{R}^\mathbb{R}$ of all maps $\mathbb{R} \to \mathbb{R}$.

4. Let $V$ be an open subset of $\mathbb{R}$. We denote by

\[ C^0(V) \text{ the space of all continuous real valued functions on } V, \]
\[ C^r(V) \text{ the space of all real valued functions on } V \text{ having continuous } r\text{th derivative,} \]
\[ C^\infty(V) \text{ the space of all real valued functions on } V \text{ having derivatives of all } r. \]

Then

\[ C^\infty(V) \subset \cdots \subset C^{r+1}(V) \subset C^r(V) \subset \cdots \subset C^0(V) \subset \mathbb{R}^V \]

is a sequence of real vector subspaces.
5. The space of solutions of the differential equation
\[ \frac{d^2u}{dx^2} + w^2u = 0 \]
is a real vector subspace of \( C^\infty(\mathbb{R}) \).

**Definition.** Let \( u_1, \ldots, u_r \) be vectors in a \( K \)-vector space \( M \), and let \( \alpha_1, \ldots, \alpha_r \) be scalars. Then the vector \( \alpha_1 u_1 + \cdots + \alpha_r u_r \) is called a **linear combination** of \( u_1, \ldots, u_r \). We write
\[ S(u_1, \ldots, u_r) = \{ \alpha_1 u_1 + \cdots + \alpha_r u_r : \alpha_1, \ldots, \alpha_r \in K \} \]
to denote the set of all linear combinations of \( u_1, \ldots, u_r \). \( S(u_1, \ldots, u_r) \) is a \( K \)-vector subspace of \( M \), and is called the **subspace generated by** \( u_1, \ldots, u_r \).

If \( S(u_1, \ldots, u_r) = M \), we say that \( u_1, \ldots, u_r \) **generate** \( M \) (i.e. for each \( x \in M \) there exists \( \alpha_1, \ldots, \alpha_r \in K \) such that \( x = \alpha_1 u_1 + \cdots + \alpha_r u_r \)).

**Examples:**
1. The vectors \((1, 2), (-1, 1)\) generate \( \mathbb{R}^2 \) (see Figure 1.2), since
\[ (\alpha, \beta) = \frac{\alpha + \beta}{3} (1, 2) + \frac{\beta - 2\alpha}{3} (-1, 1). \]

2. The functions \( \cos \omega x, \sin \omega x \) generate the space of solutions of the differential equation:
\[ \frac{d^2u}{dx^2} + w^2u = 0. \]
**Figure 1.2**

**Definition.** Let \( u_1, \ldots, u_r \) be vectors in a \( K \)-vector space \( M \). Then

1. \( u_1, \ldots, u_r \) are **linearly dependent** if there exist \( \alpha_1, \ldots, \alpha_r \in K \) not all zero such that
   \[
   \alpha_1 u_1 + \cdots + \alpha_r u_r = 0;
   \]
2. \( u_1, \ldots, u_r \) are **linearly independent** if
   \[
   \alpha_1 u_1 + \cdots + \alpha_r u_r = 0
   \]
   implies that \( \alpha_1, \ldots, \alpha_r \) are all zero.

**Example:** \( \cos \omega x, \sin \omega x \; (\omega \neq 0) \) are linearly independent functions in \( C^\infty(\mathbb{R}) \).

**Proof of This**

Let

\[
\alpha \cos \omega x + \beta \sin \omega x = 0; \quad \alpha, \beta \in \mathbb{R}
\]

be the zero function. Put \( x = 0 : \alpha = 0 \); put \( x = \frac{\pi}{2\omega} : \beta = 0. \)<

**Note.** If \( u_1, \ldots, u_r \) are linearly dependent, with

\[
\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r = 0,
\]

and \( \alpha_1 \; (\text{say}) \neq 0 \) then

\[
u_1 = - (\alpha_1^{-1} \alpha_2 u_2 + \cdots + \alpha_1^{-1} \alpha_r u_r).
\]

Thus \( u_1, \ldots, u_r \) linearly dependent iff one of them is a linear combination of the others.

**Definition.** A sequence of vectors \( u_1, \ldots, u_n \) in a \( K \)-vector space \( M \) is called a **basis** for \( M \) if

1. \( u_1, \ldots, u_n \) are linearly independent;
2. \( u_1, \ldots, u_n \) generate \( M \).

**Definition.** If \( u_1, \ldots, u_n \) is a basis for a vector space \( M \) then for each \( x \in M \) we have:

\[
x = \alpha^1 u_1 + \cdots + \alpha^n u_n
\]

for a sequence of scalars:

\[
(\alpha^1, \ldots, \alpha^n),
\]

which are called the **coordinates of** \( x \) **with respect to the basis** \( u_1, \ldots, u_n \).
The coordinates of \( x \) are uniquely determined once the basis is chosen because:
\[
x = \alpha^1 u_1 + \cdots + \alpha^n u_n = \beta^1 u_1 + \cdots + \beta^n u_n
\]
implies:
\[
(\alpha^1 - \beta^1)u_1 + \cdots + (\alpha^n - \beta^n)u_n = 0,
\]
and hence
\[
\alpha^1 - \beta^1 = 0, \ldots, \alpha^n - \beta^n = 0,
\]
by the linear independence of \( u_1, \ldots, u_n \). So
\[
\alpha^1 = \beta^1, \ldots, \alpha^n = \beta^n.
\]
A choice of basis therefore gives a well-defined bijective map:
\[
M \rightarrow K^n
\]
\[
x \mapsto \text{coordinates of } x,
\]
called the coordinate map wrt the given basis.

The following theorem (our first) implies that any two bases for \( M \) must have the same number of elements.

**Theorem 1.1.** Let \( M \) be a \( K \)-vector space, \( u_1, \ldots, u_n \) be linearly independent in \( M \), and \( y_1, \ldots, y_r \) generate \( M \). Then \( n \leq r \).

**Proof**
\[
u_1 = \alpha_1 y_1 + \cdots + \alpha_r y_r
\]
(say), since \( y_1, \ldots, y_r \) generate \( M \). \( \alpha_1, \ldots, \alpha_r \) are not all zero, since \( u_1 \neq 0 \).
Therefore \( \alpha_1 \neq 0 \) (say). Therefore \( y_1 \) is a linear combination of \( u_1, y_2, y_3, \ldots, y_r \).
Therefore \( u_1, y_2, y_3, \ldots, y_r \) generate \( M \).

Therefore
\[
u_2 = \beta_1 u_1 + \beta_2 y_2 + \beta_3 y_3 + \cdots + \beta_r y_r
\]
(say). \( \beta_2, \ldots, \beta_r \) are not all zero, since \( u_1, u_2 \) are linearly independent. Therefore \( \beta_2 \neq 0 \) (say). Therefore \( y_2 \) is a linear combination of \( u_1, u_2, y_3, \ldots, y_r \).

Continuing in this way, if \( n > r \) we get \( u_1, \ldots, u_r \) generate \( M \), and hence \( u_n \) is a linear combination of \( u_1, \ldots, u_r \), which contradicts the linear independence of \( u_1, \ldots, u_n \). Therefore \( n \leq r \).

**Note.** If \( u_1, \ldots, u_n \) and \( y_1, \ldots, y_r \) are two bases for \( M \) then \( n = r \).

**Definition.** A vector space \( M \) is called finite-dimensional if it has a finite basis. The number of elements in a basis is then called the dimension of \( M \), denoted by \( \dim M \).
Examples:

1. The $n$ vectors:

\[ e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 0, 1) \]

form a basis for $K^n$ as a vector-space, called the usual basis for $K^n$.

Proof of This \( \Rightarrow \) We have

\[ \alpha_1 e_1 + \cdots + \alpha_n e_n = \alpha_1 (1, 0, \ldots, 0) + \cdots + \alpha_n (0, \ldots, 0, 1) \]

\[ = (\alpha_1, \alpha_2, \ldots, \alpha_n). \]

Therefore

(a) \( e_1, \ldots, e_n \) generate \( K^n \);

(b) \( \alpha_1 e_1 + \cdots + \alpha_n e_n = 0 \Rightarrow \omega_1 = 0, \ldots, \omega_n = 0. \)

Therefore \( \alpha_1, \ldots, \alpha_n \) are linearly independent. \(<\)

2. The \( mn \) matrices:

\[
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix},
\begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix},
\ldots,
\begin{pmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 1
\end{pmatrix}
\]

form a basis for \( K^{m \times n} \) as a \( K \)-vector space.

3. The functions \( \cos \omega x, \sin \omega x \) form a basis for the solutions of the equation

\[ \frac{d^2 u}{dx^2} + \omega^2 u = 0 \quad (\omega \neq 0). \]

4. The functions

\[ 1, x, x^2, \ldots, x^n \]

form a basis for the subspace of \( C^\infty(\mathbb{R}) \) consisting of polynomial functions of degree \( \leq n \).

5. \( \dim K^n = n; \ dim K^{m \times n} = mn \). We have:

\[ \dim \mathbb{C}^{m \times n} = \begin{cases} 
mn & \text{as a complex vector space;} \\
2mn & \text{as a real vector space.} 
\end{cases} \]
Given any linearly independent set of vectors we can add extra ones to form a basis. Given any generating set of vectors we can discard some to form a basis. More generally:

**Theorem 1.2.** Let $M$ be a vector space with a finite generating set (or a vector subspace of such a space). Let $Z$ be a generating set, and let $X$ be a linearly independent subset of $Z$. Then $M$ has a finite basis $Y$ such that

$$X \subset Y \subset Z.$$ 

**Proof** Among all the linearly independent subsets of $Z$ which contain $X$ there is one at least

$$Y = \{u_1, \ldots, u_n\},$$

with a maximal number of elements, $n$ (say).

Now if $z \in Z$ then $z, u_1, \ldots, u_n$ are linearly dependent. Therefore there exist scalars $\lambda, \alpha_1, \ldots, \alpha_n$ not all zero such that

$$\lambda z + \alpha_1 u_1 + \cdots + \alpha_n u_n = 0.$$ 

$\lambda \neq 0$, since $u_1, \ldots, u_n$ are linearly independent. Therefore $z$ is a linear combination of $u_1, \ldots, u_n$.

But $Z$ generates $M$. Therefore $u_1, \ldots, u_n$ generate $M$. Therefore $u_1, \ldots, u_n$ form a basis for $M$. $\Box$
Chapter 2

Linear Operators 1

2.1 The Definition

Definition. Let $M, N$ be $K$-vector spaces. A map

$$M \xrightarrow{T} N$$

is called a linear operator (or linear map or linear function or linear transformation or linear homomorphism) if

(i) $T(x + y) = Tx + Ty$ (group homomorphism);

(ii) $T\alpha x = \alpha Tx$ for all $x, y \in M, \alpha \in K$.

A linear operator is called a (linear) isomorphism if $T$ is bijective. We say that $M$ is isomorphic to $N$ if there exists a linear isomorphism

$$M \rightarrow N.$$

Note. Geometrically:

(i) means that $T$ preserves parallelograms (see Figure 2.1);

(ii) means that $T$ preserves collinearity (see Figure 2.2).
Examples:

1. If

\[ A = (\alpha_j^i) = \begin{pmatrix} \alpha_1^1 & \ldots & \alpha_1^n \\ \vdots & \ddots & \vdots \\ \alpha_m^1 & \ldots & \alpha_m^n \end{pmatrix} \in K^{m \times n}, \]

we denote by

\[ K^n \xrightarrow{A} K^m \]
the linear operator given by matrix multiplication by $A$ acting on elements of $K^n$ written as $n \times 1$ columns. Since

$$A(x + y) = Ax + Ay,$$

$$A\alpha x = \alpha Ax$$

for matrix multiplication, it follows that $A$ is a linear operator.

E.g.

$$A = \begin{pmatrix} 3 & 7 & 2 \\ -2 & 5 & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 3}$$

Now:

$$\mathbb{R}^3 \to \mathbb{R}^2 : \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \mapsto \begin{pmatrix} 3\alpha + 7\beta + 2\gamma \\ -2\alpha + 5\beta + \gamma \end{pmatrix}.$$  

2. Take

$$\frac{d}{dt} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}).$$

Now:

$$\frac{d}{dt} [x(t) + y(t)] = \frac{d}{dt} x(t) + \frac{d}{dt} y(t),$$

$$\frac{d}{dt} cx(t) = c \frac{d}{dt} x(t)$$

for all $c \in \mathbb{R}$. Therefore $\frac{d}{dt}$ is a linear operator.

3. The Laplacian

$$\Delta = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2} : C^\infty(\mathbb{R}^3) \to C^\infty(\mathbb{R}^3)$$

is a linear operator.

### 2.2 Basic Properties of Linear Operators

1. If $M \xrightarrow{T} N$ is a linear operator and $u_1, \ldots, u_r \in M$; $\alpha_1, \ldots, \alpha_r \in K$ then

$$T(\alpha_1 u_1 + \cdots + \alpha_r u_r) = \alpha_1 Tu_1 + \cdots + \alpha_r Tu_r,$$

i.e.

$$T \sum_{i=1}^r \alpha_i u_i = \sum_{i=1}^r \alpha_i Tu_i,$$

i.e. $T$ preserves linear combinations, i.e. $T$ can be moved across summations and scalars.
2. If \( M \rightarrow N \) are linear operators, if \( u_1, \ldots, u_m \) generate \( M \), and if \( Su_i = Tu_i \) \((i = 1, \ldots, m)\) then \( S = T \).

\[ \text{Proof of This} \]

Let \( x \in M \). Then \( x = \sum_{i=1}^{m} \alpha_i u_i \) (say). Therefore

\[
Sx = S \sum_{i=1}^{m} \alpha_i u_i = \sum_{i=1}^{m} \alpha_i Su_i = \sum_{i=1}^{m} \alpha_i Tu_i = T \sum_{i=1}^{m} \alpha_i u_i = Tx.
\]

\[ \Box \]

Thus two linear operators which agree on a generating set must be equal.

3. Let \( u_1, \ldots, u_n \) be a basis for \( M \), and \( w_1, \ldots, w_n \) be arbitrary vectors in \( N \). Then we can define a linear operator \( M \rightarrow N \) by

\[
T(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_1 w_1 + \cdots + \alpha_n w_n.
\]

Thus \( T \) is the unique linear operator such that

\[
Tu_i = w_i \quad (i = 1, \ldots, m).
\]

We say that \( T \) is defined by \( Tu_i = w_i \), and extended to \( M \) by linearity.

**Definition.** Let \( M \rightarrow N \) be a linear operator. Then

\[
\ker T = \{ x \in M : Tx = 0 \}
\]

is a vector subspace of \( M \), called the kernel of \( M \), and

\[
\im T = \{ Tx : x \in M \}
\]

is a vector subspace of \( N \), called the image of \( T \). The dimension of \( \im T \) is called the rank of \( T \),

\[ \text{rank } T = \dim \im T. \]
2.3 Examples

1. Consider the matrix operator

\[ K^n \xrightarrow{A} K^m, \]

where \( A \in K^{m \times n}, \)

\[ A = \begin{pmatrix}
\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\
\vdots & \vdots & & \vdots \\
\alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn}
\end{pmatrix} \]

(say).

\[ \ker T = \{ x = (x_1, \ldots, x_n) : Ax = 0 \} \]

is the space of solutions of

\[ \begin{pmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \vdots & \vdots \\
\alpha_{m1} & \cdots & \alpha_{mn}
\end{pmatrix} \begin{pmatrix}
x_1 \\
\vdots \\
x_n
\end{pmatrix} = \begin{pmatrix}
0 \\
\vdots \\
0
\end{pmatrix}, \]

i.e. The space of solutions of the \( m \) homogeneous linear equations in \( n \) unknowns, whose coefficients are the rows of \( A \):

\[ \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n = 0 \]

\[ \vdots \]

\[ \alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + \alpha_{in}x_n = 0 \]

\[ \vdots \]

\[ \alpha_{m1}x_1 + \alpha_{m2}x_2 + \cdots + \alpha_{mn}x_n = 0 \]

Number of equations = \( m = \) number of rows of \( A = \dim K^m \).

Number of unknowns = \( n = \) number of columns of \( A = \dim K^n \).

We see that \((x_1, x_2, \ldots, x_n) \in \ker A\) iff the dot product:

\[ (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in}) \cdot (x_1, x_2, \ldots, x_n) \quad (i = 1, \ldots, m) \]

with each row of \( A \) is zero. Therefore

\[ \ker A = (\text{row } A)^\perp, \]

where row \( A \) is the vector subspace of \( K^n \) generated by the \( m \) rows of \( A \) (see Figure 2.3).

Now row \( A \) is unchanged by the following elementary row operations:
(i) multiplying a row by a non-zero echelon;
(ii) interchanging rows;
(iii) adding to one row a scalar multiple of another row.

So \( \text{ker } A \) is also unchanged by these operations.

To obtain a basis for row \( \text{ker } A \), and from this a basis for \( \text{ker } A \), carry out elementary row operations in order to bring the matrix to row echelon form (i.e. so that each row begins with more zeros than the previous row).

Example: Let

\[
A = \begin{pmatrix} 2 & 1 & -1 & 3 \\ -1 & 1 & 2 & 1 \\ 4 & 0 & -1 & 2 \end{pmatrix} : \mathbb{R}^4 \to \mathbb{R}^3.
\]

Now

\[
A \sim \begin{pmatrix} 2 & 1 & -1 & 3 \\ 0 & 3 & 3 & 5 \\ 0 & -2 & 1 & -4 \end{pmatrix} \begin{array}{c} 2\text{ row } 2 + \text{ row } 1 \\ \text{ row } 3 - 2 \text{ row } 1 \end{array}
\]

\[
\sim \begin{pmatrix} 2 & 1 & -1 & 3 \\ 0 & 3 & 3 & 5 \\ 0 & 0 & 9 & -2 \end{pmatrix} \begin{array}{c} 3\text{ row } 3 + 2\text{ row } 2. \end{array}
\]

Since the new rows are in row echelon form they are linearly independent. Therefore row \( \text{ker } A \) is 3-dimensional, with basis \((2, 1, -1, 3), (0, 3, 3, 5), (0, 9, -2)\). Therefore

\[
(\alpha, \beta, \gamma, \delta) \in \text{ker } A \iff 2\alpha + \beta - \gamma + 3\delta = 0
\]
\[
3\beta + 3\gamma + 5\delta = 0
\]
\[
9\gamma - 2\delta = 0
\]

\[
\iff \gamma = \frac{2}{9}\delta
\]
\[
3\beta = -3\gamma - 5\delta = -\frac{2}{3}\delta - 5\delta = -\frac{47}{3}\delta
\]
\[
2\alpha = -\beta + \gamma - 3\delta = -\frac{47}{9}\delta + \frac{2}{9}\delta - 3\delta = -\frac{8}{9}\delta
\]

\[
\iff (\alpha, \beta, \gamma, \delta) = (-\frac{4}{9}\delta, -\frac{47}{9}\delta, \frac{2}{9}\delta, \delta) = \delta(-4, -17, 2, 9)
\]

Therefore \( \text{ker } A \) is 1-dimensional, with basis \((-4, -17, 2, 9)\).
If \[ A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \\ \alpha_{21} & \cdots & \alpha_{2j} & \cdots & \alpha_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mj} & \cdots & \alpha_{mn} \end{pmatrix} \in K^{m \times n} \]

then

\[ Ae_j = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \\ \alpha_{21} & \cdots & \alpha_{2j} & \cdots & \alpha_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mj} & \cdots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j^{th} \text{ slot} \]

\[ = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix} = j^{th} \text{ column of } A. \]

Therefore

\[ \text{im } A = \{ Ax : x \in K^n \} = \{ A(\alpha_1 e_1 + \cdots + \alpha_n e_n) : \alpha_1, \ldots, \alpha_n \in K \} = \{ \alpha_1 Ae_1 + \cdots + \alpha_n Ae_n : \alpha_1, \ldots, \alpha_n \in K \} = \text{col } A, \]

where \( \text{col } A \) is the vector subspace of \( K^m \) generated by the \( n \) columns of \( A \).

To find a basis for \( \text{im } A = \text{col } A \) we carry out elementary column operations on \( A \).

Example: If

\[ A = \begin{pmatrix} 2 & 1 & -1 & 3 \\ -1 & 1 & 2 & 1 \\ 4 & 0 & -1 & 2 \end{pmatrix} \]
then

\[
A \sim \begin{pmatrix} 2 & 0 & 0 & 0 \\ -1 & 3 & 3 & 5 \\ 4 & -4 & 2 & -8 \\ 2 & 0 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ 4 & -4 & 6 & -4 \end{pmatrix} \sim \begin{pmatrix} \text{2 col 2} - \text{col 1} \\ \text{2 col 3} + \text{col 1} \\ \text{2 col 4} - 3 \text{ col 1} \\ \text{col 3} - 2 \text{ col 2} \\ 3 \text{ col 4} - 5 \text{ col 2} \end{pmatrix}.
\]

Therefore \( \text{im } A = \text{col } A \) has basis \((2, -1, 4), (0, 3, -4), (0, 0, 6)\). Therefore \( \text{rank } A = \text{dim } \text{im } A = 3 \).

2. Let

\[
D = \frac{d}{dt} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}) \quad (D x(t) = \frac{d}{dt} x(t)).
\]

(i) Let \( \lambda \in \mathbb{R} \) and \( D - \lambda \) be the operator

\[
(D - \lambda) = \frac{d}{dt} x(t) - \lambda x(t).
\]

Then

\[
x \in \text{ker}(D - \lambda) \iff (D - \lambda)x = 0 \iff \frac{dx}{dt} = \lambda x \iff x(t) = ce^{\lambda t}.
\]

Therefore \( \text{ker}(D - \lambda) \) is 1-dimensional, with basis \( e^{\lambda t} \).

(ii) To determine \( \text{ker}(D - \lambda)^k \) we must solve:

\[
(D - \lambda)^k x = 0.
\]

Put \( x(t) = e^{\lambda t} y(t) \). Then

\[
(D - \lambda)x = Dx(t) - \lambda x(t)
= \lambda e^{\lambda t} y(t) + e^\lambda Dy(t) - \lambda e^{\lambda t} y(t)
= e^\lambda Dy(t).
\]

Therefore

\[
(D - \lambda)^2 x = e^\lambda D^2 y(t)
\]

\[
:\quad (D - \lambda)^k x = e^\lambda D^k y(t).
\]

2–8
Therefore

\[(D - \lambda)^k x = 0 \iff e^{\lambda t} D^k y(t) = 0\]

\[\iff D^k y(t) = 0\]

\[\iff y(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_{k-1} t^{k-1}\]

\[\iff x(t) = (c_0 + c_1 t + \cdots + c_{k-1} t^{k-1}) e^{\lambda t}.\]

Therefore ker\((D - \lambda)^k\) is \(k\)-dimensional, with basis \(e^{\lambda t}, te^{\lambda t}, t^2 e^{\lambda t}, \ldots, t^{k-1} e^{\lambda t}\).

## 2.4 Properties Continued

**Theorem 2.1.** Let \(M \xrightarrow{T} N\) be a linear operator, where \(M\) is finite dimensional. Let \(u_1, \ldots, u_k\) be a basis for ker\(T\), and let \(T w_1, \ldots, T w_r\) be a basis for im\(T\). Then

\[u_1, \ldots, u_k, w_1, \ldots, w_r\]

is a basis for \(M\).

**Proof** We have two things to show:

(i) **Linear independence:** Let

\[\sum \alpha_i u_i + \sum \beta_j w_j = 0\]

Apply \(T:\)

\[0 + \sum \beta_j T w_j = 0.\]

Therefore \(\beta_j = 0\) for all \(j\). Therefore \(\alpha_i = 0\) for all \(i\).

Therefore \(u_1, \ldots, u_k, w_1, \ldots, w_r\) are linearly independent.

(ii) **Generate:** Let \(x \in M\). Then

\[Tx = \sum \beta_j T w_j \quad (\text{say}).\]

Therefore

\[Tx = T \sum \beta_j w_j.\]

Therefore

\[T [x - \sum \beta_j w_j] = 0.\]

Therefore

\[x - \sum \beta_j w_j \in \text{ker} \, T.\]

2-9
Therefore
\[
x - \sum \beta_j w_j = \sum \alpha_i u_i \quad \text{(say)}.
\]

Therefore
\[
x = \sum \alpha_i u_i + \sum \beta_j w_j.
\]

Therefore \( u_1, \ldots, u_k, w_1, \ldots, w_r \) generate \( M \).

**Corollary 2.1.** \( \dim \ker T + \dim \im T = \dim M \).

**Corollary 2.2.** If \( \dim M = \dim N \) then
\[
T \text{ is injective} \iff \ker T = \{0\} \iff \dim \im T = \dim N \iff T \text{ is surjective}.
\]

### 2.5 Operator Algebra

If \( M, N \) are \( K \)-vector spaces, we denote by
\[
\mathcal{L}(M, N)
\]

the set of all linear operators \( M \to N \), and we denote by
\[
\mathcal{L}(M)
\]

the set of all linear operators \( M \to M \).

**Theorem 2.2.** We have:

(i) \( \mathcal{L}(M, N) \) is a \( K \)-vector space, with
\[
(S + T)x = Sx + Tx,
\]
\[
(\alpha T)x = \alpha(Tx)
\]

for all \( S, T \in \mathcal{L}(M, N) \), \( x \in M \), \( \alpha \in K \).

(ii) Composition of operators gives a multiplication
\[
\mathcal{L}(L, M) \times \mathcal{L}(M, N) \to \mathcal{L}(L, N) \quad \left\{ \begin{array}{c}
L \xrightarrow{T} M \xrightarrow{S} N,
\end{array} \right.
\]

with
\[
(ST)x = S(Tx) \quad \text{for all } x \in L,
\]

which satisfies

(a) \( (RS)T = R(ST) \),
\[(b)\] \(R(S + T) = RS + RT,\)
\[(c)\] \((R + S)T = RT + ST,\)
\[(d)\] \((\alpha S)T = \alpha(ST) = S(\alpha T),\)
provided each is well-defined.

Proof Straight forward verification.

**Corollary 2.3.** \(L(M)\) is

(i) a \(K\)-vector space: \(S + T, \alpha S;\)

(ii) a ring: \(S + T, ST;\)

(iii) \((\alpha S)T = \alpha(ST) = S(\alpha T): \alpha S, ST,\)
i.e. \(L(M)\) is a \(K\)-algebra.

### 2.6 Isomorphisms of \(L(M, N)\) with \(K^{m \times n}\)

**Definition.** Let \(u_1, \ldots, u_n\) be a basis for \(M\), and let \(w_1, \ldots, w_m\) be a basis for \(N\). Let \(M \xrightarrow{T} N\). Put Then we have:

\[
Tu_1 = \alpha_1^1 w_1 + \alpha_1^2 w_2 + \cdots + \alpha_1^i w_i + \cdots + \alpha_1^m w_m, \\
\vdots \\
Tu_j = \alpha_j^1 w_1 + \alpha_j^2 w_2 + \cdots + \alpha_j^i w_i + \cdots + \alpha_j^m w_m, \\
\vdots \\
Tu_n = \alpha_n^1 w_1 + \alpha_n^2 w_2 + \cdots + \alpha_n^i w_i + \cdots + \alpha_n^m w_m,
\]
(say) where:

\[
A = (\alpha_j^i) = \begin{pmatrix}
\alpha_1^1 & \alpha_2^1 & \cdots & \alpha_j^1 & \cdots & \alpha_n^1 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\alpha_1^i & \cdots & \alpha_j^i & \cdots & \alpha_n^i \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\alpha_1^m & \cdots & \alpha_j^m & \cdots & \alpha_n^m
\end{pmatrix} \in K^{m \times n}. 
\]

**Note.** The coordinates of \(Tu_j\) form the \(j^{th}\) column of \(A\) - NOTE THE TRANSPOSE! We call \(A\) the **matrix of** \(T\) wrt the bases \(u_1, \ldots, u_n\) for \(M\) and \(w_1, \ldots, w_m\) for \(N\),

\[
Tu_j = \sum_{i=1}^m \alpha_j^i \omega_i.
\]
Theorem 2.3. \( \mathcal{L}(M, N) \to K^{m \times n} \) is a linear isomorphism where \( T \to \) matrix of \( T \) w.r.t. basis \( u_1, \ldots, u_n; \omega_1, \cdots, \omega_m \).

Proof Let \( T \) have matrix \( A = (\alpha_j^i) \), and let \( S \) have matrix \( B = (\beta_j^i) \). Then
\[
(T + S)u_j = Tu_j + Su_j = \sum_{i=1}^{m} \alpha_j^i w_i + \sum_{i=1}^{m} \beta_j^i w_i = \sum_{i=1}^{m} (\alpha_j^i + \beta_j^i) w_i.
\]
Therefore \( T + S \) has matrix \( (\alpha_j^i + \beta_j^i) = A + B \). Also
\[
(\lambda T)u_j = \lambda (Tu_j) = \lambda \sum_{i=1}^{m} \alpha_j^i w_i = \sum_{i=1}^{m} \lambda \alpha_j^i w_i.
\]
Therefore \( \lambda T \) has matrix \( (\lambda \alpha_j^i) = \lambda A \).

Corollary 2.4. \( \dim \mathcal{L}(M, N) = \dim M \cdot \dim N \).

Theorem 2.4. If \( L \xrightarrow{T} M \) has matrix \( A = (\alpha_j^i) \) wrt basis \( v_1, \ldots, v_p, u_1, \ldots, u_n \), and \( M \xrightarrow{S} N \) has matrix \( B = (\beta_j^i) \) wrt basis \( u_1, \ldots, u_n, w_1, \ldots, w_m \) then \( L \xrightarrow{ST} N \) has basis
\[
BA = \left( \sum_{k=1}^{n} \beta_k^i \alpha_j^k \right) = (\gamma_j^i)
\]
(say), wrt basis \( v_1, \ldots, v_p, w_1, \ldots, w_m \).

Proof
\[
(ST)v_j = S(Tv_j) = S \left( \sum_{k=1}^{n} \alpha_j^k u_k \right) = \sum_{k=1}^{n} \alpha_j^k Su_k
\]
\[
= \sum_{k=1}^{n} \alpha_j^k \sum_{i=1}^{m} \beta_k^i w_i = \sum_{i=1}^{m} \left( \sum_{k=1}^{n} \beta_k^i \alpha_j^k \right) w_i = \sum_{i=1}^{m} \gamma_j^i w_i.
\]

Corollary 2.5. If \( \dim M = n \) then each choice of basis \( u_1, \ldots, u_n \) of \( M \) defines an isomorphism of \( K \)-algebras:
\[
\mathcal{L}(M) \to K^n : T \mapsto \text{matrix of } T \text{ wrt } u_1, \ldots, u_n.
\]

Note. If \( M \xrightarrow{T} M \) has matrix \( A = (\alpha_j^i) \) wrt basis \( u_1, \ldots, u_n \) then
\[
(i) \ Tu_j = \sum_{i=1}^{n} \alpha_j^i u_i, \text{ by definition;}
\]

2–12
(ii) the elements of the $j^{th}$ column of $A$ are the coordinates of $Tu_j$;

(iii) $\lambda_0 I + \lambda_1 T + \lambda_2 T^2 + \cdots + \lambda_r T^r$ has matrix $\alpha_0 I + \alpha_1 A + \cdots + \alpha_r A^r$;

(iv) $T^{-1}$ has matrix $A^{-1}$,

since we have an algebra isomorphism.

**Theorem 2.5.** Let $M \xrightarrow{T} N$ have matrix $A = (\alpha_j^i)$ wrt bases $u_1, \ldots, u_n$ for $M$ and $w_1, \ldots, w_m$ for $N$. Let $x$ have coordinates

$$X = (\xi^i) = \begin{pmatrix} \xi^1 \\ \vdots \\ \xi^n \end{pmatrix}$$

wrt $u_1, \ldots, u_n$. Then $Tx$ has coordinates

$$AX = \left( \sum_{i=1}^m \alpha_j^i \xi^j \right)$$

wrt $w_1, \ldots, w_m$.

**Proof**

$$Tx = T\left( \sum_{j=1}^n \xi^j u_j \right) = \sum_{j=1}^n \xi^j Tu_j = \sum_{j=1}^n \xi^j \sum_{i=1}^m \alpha_j^i w_i = \sum_{i=1}^m \left( \sum_{j=1}^n \alpha_j^i \xi^j \right) w_i,$$

as required. ◀

**Note.** We have thus a commutative diagram:

$$
\begin{array}{ccc}
M & \xrightarrow{T} & N \\
\downarrow & & \downarrow \\
K^n & \xrightarrow{A} & K^m
\end{array}
\quad
\begin{array}{ccc}
x & \xrightarrow{T} & Tx \\
\downarrow & & \downarrow \\
x \text{ coord.} & \xrightarrow{A} & Tx \text{ coord}
\end{array}
$$
Chapter 3

Changing Basis and Einstein Convention

Definition. If \( u_1, \ldots, u_n \) and \( w_1, \ldots, w_n \) are two bases for \( M \) then we have:

\[
\begin{align*}
  u_1 &= p^1_1 w_1 + p^2_1 w_2 + \cdots + p^n_1 w_n \\
  &\quad \vdots \\
  u_j &= p^1_j w_1 + p^2_j w_2 + \cdots + p^n_j w_n \\
  &\quad \vdots \\
  u_n &= p^1_n w_1 + p^2_n w_2 + \cdots + p^n_n w_n
\end{align*}
\]

(say). Put

\[
P = (p^i_j) = \begin{pmatrix}
  p^1_1 & \cdots & p^1_n \\
  p^2_1 & \cdots & p^2_n \\
  \vdots & \ddots & \vdots \\
  p^n_1 & \cdots & p^n_n
\end{pmatrix}
\]

Note. The new coordinates of the old basis vector \( u_j \) form the \( j^{th} \) column of \( P \) - NOTE THE TRANSPOSE! We call \( P \) the transition matrix from the (old) basis \( u_1, \ldots, u_n \) to the (new) basis \( w_1, \ldots, w_n \):

\[
u_j = \sum_{i=1}^n p^i_j w_i.
\]

Theorem 3.1. If \( x \) has old coordinates

\[
X = (\xi^i) = \begin{pmatrix}
  \xi^1 \\
  \vdots \\
  \xi^n
\end{pmatrix}
\]

3-1
then \( x \) has new coordinates

\[ PX = \sum_{j=1}^{n} (p_j^i \xi^j) = (\eta^i) \]

(say).

\textbf{Proof ▶}

\[ x = \sum_{j=1}^{n} \xi^j u_j = \sum_{j=1}^{n} \xi^j \sum_{i=1}^{n} p_j^i w_i = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} p_j^i \xi^j \right) w_i = \sum_{i=1}^{n} \eta^i w_i. \]

\textbf{ Armstrong ▶

We shall often use the \textit{Einstein summation convention (s.c.)} when dealing with basis and coordinates in a fixed \( n \)-dimensional vector space \( M \). Repeated indices (one up, one down) are summed from 1 to \( n \) (\textit{contraction} of repeated indices). Non-repeated indices may take each value 1 to \( n \).

\textbf{Example:}

- \( \alpha^i \) denotes
  \[
  \begin{pmatrix}
  \alpha^1 \\
  \vdots \\
  \alpha^n
  \end{pmatrix}
  \] (column matrix; upper index labels the row).

- \( \alpha_i \) denotes
  \[(\alpha_1, \ldots, \alpha_n)\] (row matrix; lower index labels the column).

- \( \alpha_j^i \) denotes
  \[
  \begin{pmatrix}
  \alpha_1^1 & \cdots & \alpha_n^1 \\
  \vdots & \ddots & \vdots \\
  \alpha_1^n & \cdots & \alpha_n^n
  \end{pmatrix}
  \] (square matrix).

- \( u_i \) denotes \( u_1, \ldots, u_n \) (basis).

- \( \alpha^i u_i \) denotes \( \alpha^1 u_1 + \cdots + \alpha^n u_n \).

- \( \alpha^i \beta_i \) denotes \( \alpha^1 \beta_1 + \cdots + \alpha^n \beta_n \) (dot product).
\begin{itemize}
    \item $\alpha_i^j \beta_j^k$ denotes $AB$ (matrix product).
\end{itemize}

Also
\begin{align*}
    Tu_j &= \alpha_j^i u_i \quad (\alpha_j^i \text{ matrix of operator } T) \\
    u_j &= p_j^i w_i \quad (p_j^i \text{ transition matrix from } u_i \text{ to } w_i).
\end{align*}

If $x$ has components $\xi^i$ wrt $u_i$ then $Tx$ has components $\alpha_j^i \xi^i$ wrt $u_i$. If $x$ has components $\xi^j$ wrt $u_i$ then $x$ has components $p_j^i \xi^i$ wrt $w_i$.

\begin{itemize}
    \item $\delta_j^i$ denotes the unit matrix
      \[
      I = \begin{pmatrix}
      1 & 0 & \ldots & 0 \\
      0 & 1 & \ldots & 0 \\
      \vdots & \ddots & \ddots & \vdots \\
      0 & \ldots & 0 & 1
      \end{pmatrix}.
      \]
    \item If $Q = P^{-1}$ then $(q_j^i)$ denotes $Q$ (inverse matrix) and
      \[
      q_k^i p_j^k = \delta_j^i = p_k^i q_j^k.
      \]
\end{itemize}

**Theorem 3.2.** Let $M \rightarrow N$ have matrix $A$ wrt basis $u_1, \ldots, u_n$. Let $P$ be the transition matrix to (new) basis $w_1, \ldots, w_n$. Then $T$ has (new) matrix
\[
PAP^{-1}
\]
wrt $w_1, \ldots, w_n$.

**Proof** Let $P = (p_j^i)$, $A = (\alpha_j^i)$, $P^{-1} = Q = (q_j^i)$. Then
\[
Tu_j = \alpha_j^i u_i; \quad u_j = p_j^i w_i; \quad w_j = q_j^i u_i.
\]
Therefore
\[
Tw_j = T q_j^i u_i = q_j^i T u_i = q_j^i \alpha_i^k u_k = q_j^i \alpha_i^k p_k^i w_i = \underbrace{p_k^i \alpha_i^k q_j^i}_{PAP^{-1}} w_i,
\]
as required. \(\blacksquare\)
Chapter 4
Linear Forms and Duality

4.1 Linear Forms

Definition. Fix $M$ a $K$-vector space. A scalar valued linear function
\[ f : M \to K \]
is called a linear form on $M$.

If $f$ is a linear form on $M$, and $x$ is a vector in $M$, we write
\[ \langle f, x \rangle \]
to denote the value of $f$ on $x$. This notation has the advantage of treating $f$ and $x$ in a symmetrised way:

(i) $\langle f, x + y \rangle = \langle f, x \rangle + \langle f, y \rangle$,
(ii) $\langle f + g, x \rangle = \langle f, x \rangle + \langle g, x \rangle$,
(iii) $\langle \alpha f, x \rangle = \alpha \langle f, x \rangle = \langle f, \alpha x \rangle$,
(iv) $\left\langle \sum_{i=1}^{r} \alpha_i f^i, \sum_{j=1}^{s} \beta^j x_j \right\rangle = \sum_{i=1}^{r} \sum_{j=1}^{s} \alpha_i \beta^j \langle f^i, x_j \rangle$.

If $M$ is finite dimensional, with basis $u_1, \ldots, u_n$, then each $x \in M$ can be written uniquely as
\[ x = \alpha^1 u_1 + \cdots + \alpha^n u_n = \sum_{i=1}^{n} \alpha^i u_i = \alpha^i u_i. \]

We write
\[ \langle u^i, x \rangle = \alpha^i \]

4–1
to denote the $i^{th}$ coordinate of $x$ wrt basis $u_1, \ldots, u_n$. We have:

$$\langle u^i, x + y \rangle = \langle u^i, x \rangle + \langle u^i, y \rangle,$$

$$\langle u^i, \alpha x \rangle = \alpha \langle u^i, x \rangle.$$ 

Thus $u^i$ is a linear form on $M$, called the $i^{th}$ coordinate function wrt basis $u_1, \ldots, u_n$. We have:

1. $\langle u^i, u_j \rangle = \left\{ \begin{array}{ll}
1 & \text{if } i = j \\
0 & \text{if } i \neq j
\end{array} \right.$ \quad (Kronecker delta);
2. $x = \sum_{i=1}^{n} \langle u^i, x \rangle u_i$ for all $x \in M$;
3. $\langle \alpha_1 u^1 + \cdots + \alpha_n u^n, \beta_1 u_1 + \cdots + \beta^n u_n \rangle = \alpha_1 \beta^1 + \cdots + \alpha_n \beta^n = \alpha_i \beta^i$ \quad (dot product).

**Theorem 4.1.** If $u_1, \ldots, u_n$ is a basis for $M$ then the coordinate functions $u^1, \ldots, u^n$ form a basis for the space $M^*$ of linear forms on $M$ (called the dual space of $M$), called the dual basis, and

$$f = \sum_{i=1}^{n} \langle f, u_i \rangle u^i \quad \text{for each } f \in M^*.$$ 

**Proof** We have to show that $u^1, \ldots, u^n$ generate $M$, and are linearly independent.

(i) **Generate:** Let $f \in M^*$; $\langle f, u_j \rangle = \beta_j$ (say). Then

$$\left\langle \sum_{i=1}^{n} \beta_i u^i, u_j \right\rangle = \sum_{i=1}^{n} \beta_i \langle u^i, u_j \rangle = \sum_{i=1}^{n} \beta_i \delta^i_j = \beta_j = \langle f, u_j \rangle.$$ 

Therefore $\sum_{i=1}^{n} \beta_i u^i$ and $f$ are linear forms on $M$ which agree on the basis vectors $u_1, \ldots, u_n$. Therefore

$$f = \sum_{i=1}^{n} \beta_i u^i = \sum_{i=1}^{n} \langle f, u_i \rangle u^i.$$ 

(ii) **Linear independence:** Let $\sum_{i=1}^{n} \beta_i u^i = 0$. Then

$$\left\langle \sum_{i=1}^{n} \beta_i u^i, u_j \right\rangle = 0$$

4–2
for all \( j = 1, \ldots, n \). Therefore
\[
\sum_{i=1}^{n} \beta_i \delta^i_j = 0
\]
for all \( j = 1, \ldots, n \). Therefore \( \beta_j = 0 \) for all \( j = 1, \ldots, n \). Therefore \( u^1, \ldots, u^n \) are linearly independent.

**Corollary 4.1.** \( \dim M^* = \dim M \).

**Note.** We denote by \( x, y, z \) the coordinate function on \( K^3 \) wrt basis \( e_1, e_2, e_3 \), and we denote by \( x^1, \ldots, x^n \) the coordinate function on \( K^n \) wrt basis \( e_1, \ldots, e_n \). These coordinates are called the **usual coordinates**.

### 4.2 Duality

Let \( M \) be finite dimensional, with dual space \( M^* \). If \( x \in M \) and \( f \in M^* \) then

(i) \( f \) is a linear form on \( M \) whose value on \( x \) is \( \langle f, x \rangle \);

(ii) we identify \( x \) with the linear form on \( M^* \) whose value on \( f \) is \( \langle f, x \rangle \):

\[
\begin{align*}
\langle f, \cdot \rangle, \\
x = \langle \cdot, x \rangle.
\end{align*}
\]

What we are doing is identifying \( M \) with the dual of \( M^* \), by means of the linear isomorphism:

\[
\begin{align*}
M & \to M^{**} \\
x & \mapsto \langle \cdot, x \rangle.
\end{align*}
\]

This is a linear map, and is bijective because:

(i) \( \dim M^{**} = \dim M^* = \dim M \),

(ii) \( \langle \cdot, x \rangle = 0 \Rightarrow \langle u^i, x \rangle = 0 \) for all \( x \Rightarrow x = 0 \). So the map is injective (kernel = \{0\}), and hence by (i) surjective.

If \( u_1, \ldots, u_n \) is a basis for \( M \), and \( u^1, \ldots, u^n \) the dual basis for \( M^* \) then

\[
\langle u^i, u_j \rangle = \delta^i_j
\]

shows that \( u_1, \ldots, u_n \) is the basis dual to \( u^1, \ldots, u^n \).

The identification of vectors \( x \in M \) as linear forms on \( M^* \) is called **duality**. A basis \( u^1, \ldots, u^n \) for \( M^* \) is called a **linear coordinate system** on \( M \), and consists of coordinate functions wrt its dual basis \( u_1, \ldots, u_n \).
4.3 Systems of Linear Equations

**Definition.** If $f^1, \ldots, f^k$ are linear forms on $M$ then we consider the vector subspace of $M$ on which

\[ f^1 = 0, \ldots, f^k = 0 \quad (\ast). \]

Any vector in this subspace is called a *solution* of the equations $(\ast)$. Thus $x \in M$ is a solution iff

\[ (f^1, x) = 0, \ldots, (f^k, x) = 0. \]

The set of solutions is called the *solution space* of the system of $k$ homogeneous equations $(\ast)$. The dimension of the space $\mathcal{S}(f^1, \ldots, f^k)$ generated by $f^1, \ldots, f^k$ is called the *rank* (number of linearly independent equations) of the system of equations.

In particular, if $u^1, \ldots, u^n$ is a linear coordinate system on $M$ then we can write the equations as:

\[
\begin{align*}
  f^1 &\equiv \beta^1_1 u^1 + \cdots + \beta^1_n u^n = 0 \\
  \vdots \\
  f^k &\equiv \beta^k_1 u^1 + \cdots + \beta^k_n u^n = 0
\end{align*}
\]

The coordinate map $M^* \to K^n$ maps

\[
\begin{align*}
  f^1 &\mapsto (\beta^1_1, \ldots, \beta^1_n) \\
  \vdots \\
  f^k &\mapsto (\beta^k_1, \ldots, \beta^k_n).
\end{align*}
\]

Thus it maps $\mathcal{S}(f^1, \ldots, f^k)$ isomorphically onto the row space of $B = (\beta^i_j)$. Therefore

\[
\text{rank of system} = \text{dimension of row space of } B = \dim \text{ row } B.
\]

**Example:** The equations

\[
\begin{align*}
  3x - 4y + 2z &= 0, \\
  2x + 7y + 3z &= 0,
\end{align*}
\]

where $x, y, z$ are the usual coordinates on $\mathbb{R}^3$, have

\[
\text{rank} = \dim \text{ row } \begin{pmatrix} 3 & -4 & 2 \\ 2 & 7 & 3 \end{pmatrix} = 2.
\]
Theorem 4.2. A system of $k$ homogeneous linear equations of rank $r$ on an $n$-dimensional vector space $M$ has a solution space of dimension $n - r$.

Proof. Let

$$f^1 = 0, \ldots, f^k = 0$$

be the system of equations. Let $u^1, \ldots, u^r$ be a basis for $\mathcal{S}(f^1, \ldots, f^k)$. Extend to a basis $u^1, \ldots, u^r, u^{r+1}, \ldots, u^n$ for $M^*$. Let $u_1, \ldots, u_r, u_{r+1}, \ldots, u_n$ be the dual basis of $M$. Then

$$x = \alpha^1 u_1 + \cdots + \alpha^r u_r + \alpha^{r+1} u_{r+1} + \cdots + \alpha^n u_n \in \text{solution space}$$

$$\iff \alpha^1 = \langle u^1, x \rangle = 0, \ldots, \alpha^r = \langle u^r, x \rangle = 0$$

$$\iff x = \alpha^{r+1} u_{r+1} + \cdots + \alpha^n u_n.$$ 

Therefore $u_{r+1}, \ldots, u_n$ is a basis for the solution space. Therefore solution space has dimension $n - r$. □

Theorem 4.3. Let $B \in K^{k \times n}$, where $K$ is a field. Then

$$\dim \text{row } B = \dim \text{col } B \ (= \text{rank } B).$$

Proof. Consider the $k$ homogeneous linear equations on $K^n$ with coefficients $B = (\beta^i_j)$:

$$\beta_1^1 x^1 + \cdots + \beta_1^n x^n = 0$$

$$\vdots$$

$$\beta_k^1 x^1 + \cdots + \beta_k^n x^n = 0.$$ 

Now

$$n - \dim \text{row } B = n - \text{rank of equations}$$

$$= \text{dimension of solution space}$$

$$= \dim \ker B$$

$$= n - \dim \text{im } B$$

$$= n - \dim \text{col } B.$$ 

Therefore $\dim \text{col } B = \dim \text{row } B$. □
Chapter 5

Tensors

5.1 The Definition

**Definition.** Let $M$ be a finite dimensional vector space over a field $K$, let $M^*$ be the dual space, and let $\dim M = n$. A *tensor* over $M$ is a function of the form

$$T : M_1 \times M_2 \times \cdots \times M_k \to K,$$

where each $M_i = M$ or $M^*$ ($i = 1, \ldots, k$), and which is linear in each variable (*multilinear*).

Two tensors $S, T$ are said to be of the *same type* if they are defined on the same set $M_1 \times \cdots \times M_k$.

*Example:* A tensor of type

$$T : M \times M^* \times M \to K$$

is a scalar valued function $T(x, f, y)$ of three variables ($x$ a vector, $f$ a linear form, $y$ a vector) such that

$$T(\alpha x + \beta y, f, z) = \alpha T(x, f, z) + \beta T(y, f, z) \quad \text{linear in 1st variable},$$

$$T(x, \alpha f + \beta g, z) = \alpha T(x, f, z) + \beta T(x, g, z) \quad \text{linear in 2nd variable},$$

$$T(x, f, \alpha y + \beta z) = \alpha T(x, f, z) + \beta T(x, f, z) \quad \text{linear in 3rd variable}.$$

If $u_i$ is a basis for $M$, and $u^i$ is the dual basis for $M^*$ then the array of $n^3$ scalars

$$\alpha^i_j_k = T(u_i, u^j, u_k)$$

are called the *components* of $T$. 

5-1
If \( x, f, y \) have components \( \xi^i, \eta_j, \rho^k \) respectively then
\[
T(x, f, y) = T(\xi^i u_i, \eta_j w^j, \rho^k u_k) = \xi^i \eta_j \rho^k \alpha^i_j k
\]
(using summation notation), i.e. the components of \( T \) contracted by the components of \( x, f, y \).

The set of all tensors over \( M \) of a given type form a \( K \)-vector space if we define
\[
(S + T)(x_1, \ldots, x_k) = S(x_1, \ldots, x_k) + T(x_1, \ldots, x_k),
\]
\[
(\lambda T)(x_1, \ldots, x_k) = \lambda(T(x_1, \ldots, x_k)).
\]

The vector space of all tensors of type \( M \times M^* \times M \to K \) (say) has dimension \( n^3 \), since \( T \mapsto T(u_i, w^j, u_k) \) (components of \( T \) maps it isomorphically onto \( K^{n^3} \).

**Definition.** If \( S : M_1 \times \cdots \times M_k \to K \) and \( T : M_{k+1} \times \cdots \times M_l \to K \) are tensors over \( M \) then we define their **tensor product** \( S \otimes T \) to be the tensor:
\[
S \otimes T : M_1 \times \cdots \times M_k \times M_{k+1} \times \cdots \times M_l \to K,
\]
where
\[
S \otimes T(x_1, \ldots, x_l) = S(x_1, \ldots, x_k)T(x_{k+1}, \ldots, x_l).
\]

**Example:** If \( S \) has components \( \alpha^i_j k \), and \( T \) has components \( \beta^r s \) then \( S \otimes T \) has components \( \alpha^i_j k \beta^r_s \), because
\[
S \otimes T(u_i, w^j, u_k, u^r, u^s) = S(u_i, w^j, u_k)T(u^r, u^s).
\]

Tensors satisfy algebraic laws such as:
(i) \( R \otimes (S + T) = R \otimes S + R \otimes T \),
(ii) \( (\lambda R) \otimes S = \lambda (R \otimes S) = R \otimes (\lambda S) \),
(iii) \( (R \otimes S) \otimes T = R \otimes (S \otimes T) \).
But
\[
S \otimes T \neq T \otimes S
\]
in general. To prove those we look at components wrt a basis, and note that
\[
\alpha^i_j k (\beta^r_s + \gamma^r_s) = \alpha^i_j k \beta^r_s + \alpha^i_j k \gamma^r_s,
\]
for example, but
\[
\alpha^i \beta^j \neq \beta^j \alpha^i
\]
in general.
5.2 Contraction

**Definition.** Let \( T : M_1 \times \cdots \times M_r \times \cdots \times M_s \times \cdots \times M_k \to K \) be a tensor, with
\[
M_r = M^* , \quad M_s = M
\]
(say). Then we can *contract* the \( r^{th} \) index of \( T \) with the \( s^{th} \) index to get a new tensor
\[
S : M_1 \times \cdots \times M_r \times \cdots \times M_s \times \cdots \times M_k \to K
\]
defined by
\[
S(x_1, x_2, \ldots, x_{k-2}) = T(x_1, \ldots, u^i_r \text{th slot}, \ldots, u^i_s \text{th slot}, \ldots, x_{k-2}),
\]
where \( u_i \) is a basis for \( M \).

To show that \( S \) is well-defined we need:

**Theorem 5.1.** The definition of contraction is independent of the choice of basis.

**Proof** ▶ Put
\[
R(f, x) = T(x_1, x_2, \ldots, f, \ldots, x, \ldots, x_{k-2}).
\]
Then if \( u_i, w_i \) are bases:
\[
R(w^i, u_i) = R(p^{i}_k u^k, q^{i}_l u_l) = p^{i}_k q^{i}_l R(u^k, u_l) = \delta^{i}_k R(u^k, u_l) = R(u^k, u_k),
\]
as required. ◀

**Example:** If \( T \) has components \( \alpha^{i j k} \) wrt basis \( u_i \) then contraction of the \( 2^{nd} \) and \( 4^{th} \) indices gives a tensor with components
\[
\beta^{i k m} = T(u^i, u_j, u_k, u^m) = \alpha^{i k j} m.
\]

Thus when we contract we eliminate one upper (contravariant) index and one lower (covariant) index.

5.3 Examples

A vector \( x \in M \) is a tensor:
\[
x : M^* \to K
\]
with components $\alpha^i = \langle w^i, x \rangle$ (one contravariant index).

A linear form $f \in M^*$ is a tensor:

$$f : M \to K$$

with components $\alpha_i = \langle f, u_i \rangle$ (one covariant index).

A tensor with two covariant indices:

$$T : M \times M \to K,$$

with $T(u_i, u_j) = \alpha_{ij}$, is called a bilinear form or scalar product.

**Example:** The dot product

$$K^n \times K^n \to K$$

$$((\alpha^1, \ldots, \alpha^n), (\beta^1, \ldots, \beta^n)) \mapsto \alpha^1 \beta^1 + \cdots + \alpha^n \beta^n$$

is a bilinear form on $K^n$.

If $M \to M$ is a linear operator, we shall identify it with the tensor:

$$T : M^* \times M \to K$$

by

$$T(f, x) = \langle f, Tx \rangle.$$ 

This tensor has components

$$\alpha^i_j = T(u^i, u_j) = \langle u^i, Tu_j \rangle = \text{matrix of linear operator } T$$

(one contravariant index, one covariant index).

**Note (The Transformation Law).** Let $p^j_i$ be the transition matrix from basis $u_i$ to basis $w_i$, with inverse matrix $q^i_j$. Let $T$ be a tensor $M \times M^* \times M \to K$ (say). Then

$$\overbrace{T(w_i, w^j, w_k)}^{\text{new comps.}} = T(q^r_i u_r, p^j_s u^s, q^l_k u_l) = q^r_i p^j_s q^l_k \overbrace{T(u_r, u^s, u_l)}^{\text{old comps.}},$$

i.e. Upper indices contract with $p$, lower indices contract with $q$. 

5-4
5.4 Bases of Tensor Spaces

Let \( M \times M^* \times M \to K \) (say) be a tensor with components \( \alpha_{i,j}^k \) wrt basis \( u_i \). Then the tensor:

\[
\alpha_{i,j}^k u^i \otimes u_j \otimes u^k
\]

is of the same type as \( T \), and has components

\[
\alpha_{i,j}^k u^i \otimes u_j \otimes u^k \left[ u_r, u^s, u_t \right] = \alpha_{i,j}^k \langle u^i, u_r \rangle \langle u^s, u_j \rangle \langle u^k, u_t \rangle \\
= \alpha_{i,j}^k \delta_r^i \delta_j^s \delta_t^k \\
= \alpha_{r,s}^t.
\]

Therefore (**) has the same components as \( T \). Therefore

\[
T = \alpha_{i,j}^k u^i \otimes u_j \otimes u^k.
\]

Therefore \( u^i \otimes u_j \otimes u^k \) is a basis for the \( n^3 \)-dimensional space of all tensors of type (\( * \)).
Chapter 6

Vector Fields

6.1 The Definition

Let $V$ be an open subset of $\mathbb{R}^n$. Let $x^1, \ldots, x^n$ be the usual coordinate functions on $\mathbb{R}^n$. Let $V \subseteq \mathbb{R}$. If $a = (a_1, \ldots, a_n) \in V$ then we define the partial derivative of $f$ wrt $i^{th}$ variable at $a$:

$$\frac{\partial f}{\partial x^i}(a) = \lim_{t \to 0} \frac{f(a_1, \ldots, a_i + t, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{t}$$

$$= \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t}$$

$$= \frac{d}{dt} f(a + te_i)|_{t=0}$$

(see Figure 6.1). If it exists for each $a \in V$ then we have:

$$\frac{\partial f}{\partial x^i} : V \to \mathbb{R}.$$
Note that \( \frac{\partial x^i}{\partial x^j} = \delta^i_j \).

If all repeated partial derivatives of all orders:

\[
\frac{\partial^r f}{\partial x^{i_1} \cdots \partial x^{i_r}} = \frac{\partial}{\partial x^{i_1}} \cdots \frac{\partial}{\partial x^{i_r}} f : V \to \mathbb{R}
\]

exist we call \( f \) \( C^\infty \). We denote by \( C^\infty(V) \) the space of all \( C^\infty \) functions \( V \to \mathbb{R} \). \( C^\infty(V) \) is an \( \mathbb{R} \)-algebra:

(i) \( (f + g)(x) = f(x) + g(x) \),

(ii) \( (fg)(x) = f(x)g(x) \),

(iii) \( (\alpha f)(x) = \alpha(f(x)) \).

Each sequence \( \alpha^1, \ldots, \alpha^n \) of elements of \( C^\infty(V) \) defines a linear operator

\[
v = \alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n}
\]
on $C^\infty(V)$, where

$$(vf)(x) = \alpha^1(x) \frac{\partial f}{\partial x^1}(x) + \cdots + \alpha^n(x) \frac{\partial f}{\partial x^n}(x).$$

Such an operator

$$v : C^\infty(V) \to C^\infty(V)$$

is called a (contravariant) vector field on $V$.

Now for each fixed $a$ we denote by

$$\frac{\partial}{\partial x^i}$$

the operator given by:

$$\frac{\partial}{\partial x^i} f = \frac{\partial f}{\partial x^i}(a).$$

Thus $\frac{\partial}{\partial x^i}$ acts on any function $f$ which is defined and $C^1$ on an open set containing $a$. We define the linear combination $\sum_{i=1}^n \alpha^i \frac{\partial}{\partial x^i}$ by

$$\left(\alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n}\right) f = \alpha^1 \frac{\partial f}{\partial x^1}(a) + \cdots + \alpha^n \frac{\partial f}{\partial x^n}(a).$$

The set of linear combinations

$$\left\{ \alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n} : \alpha^1, \ldots, \alpha^n \in \mathbb{R} \right\}$$

is called the tangent space to $\mathbb{R}^n$ at $a$, denoted $T_a \mathbb{R}^n$. Thus $T_a \mathbb{R}^n$ is a real $n$-dimensional vector space, with basis

$$\frac{\partial}{\partial x^1_a}, \ldots, \frac{\partial}{\partial x^n_a}.$$

The operators $\frac{\partial}{\partial x^i_a}$ are linearly independent, since

$$\alpha^1 \frac{\partial}{\partial x^1_a} + \cdots + \alpha^n \frac{\partial}{\partial x^n_a} = 0 \Rightarrow \left(\alpha^1 \frac{\partial}{\partial x^1_a} + \cdots + \alpha^n \frac{\partial}{\partial x^n_a}\right) x^i = 0 \Rightarrow \alpha^i = 0,$$

since $\frac{\partial x^j}{\partial x^i}(a) = \delta^j_i$.

If $v = \alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n}$ ($\alpha^i \in C^\infty(V)$) is a vector field on $V$ then we have (see Figure 6.2), for each $x \in V$ a tangent vector

$$v_x = \alpha^1(x) \frac{\partial}{\partial x^1_x} + \cdots + \alpha^n(x) \frac{\partial}{\partial x^n_x} \in T_x \mathbb{R}^n.$$
We call \( v_x \) the value of \( v \) at \( x \), and note that

\[
v_x f = \left( \alpha^1(x) \frac{\partial}{\partial x^1} + \cdots + \alpha^n(x) \frac{\partial}{\partial x^n} \right) f
\]

\[
= \alpha^1(x) \frac{\partial f}{\partial x^1}(x) + \cdots + \alpha^n(x) \frac{\partial f}{\partial x^n}(x)
\]

\[
= (vf)(x)
\]

for all \( x \in V \). Thus \( v \) is determined by its values \( \{v_x : x \in V\} \), and vice versa. Thus a contravariant vector field is a function on \( V \)

\[
x \mapsto v_x,
\]

which maps to each point \( x \in V \) a tangent vector \( v_x \in T_x \mathbb{R}^n \).

### 6.2 Velocity Vectors
Let $\beta(t) = (\beta^1(t), \ldots, \beta^n(t))$ be a sequence of real valued $C^\infty$ functions defined on an open subset of $\mathbb{R}$. Thus $\beta = (\beta^1, \ldots, \beta^n)$ is a curve in $\mathbb{R}^n$ (see Figure 6.3). If $f$ is a $C^\infty$ real-valued function on an open set in $\mathbb{R}^n$ containing $\beta(t)$ then the rate of change of $f$ along the curve $\beta$ at parameter $t$ is

$$
\frac{d}{dt} f(\beta(t)) = \frac{d}{dt} f(\beta^1(t), \ldots, \beta^n(t))
$$

$$
= \frac{\partial f}{\partial x^1}(\beta(t)) \frac{d}{dt} \beta^1(t) + \cdots + \frac{\partial f}{\partial x^n}(\beta(t)) \frac{d}{dt} \beta^n(t) \quad \text{(by the chain rule)}
$$

$$
= \left[ \frac{d}{dt} \beta^1(t) \frac{\partial}{\partial x^1_{\beta(t)}} + \cdots + \frac{d}{dt} \beta^n(t) \frac{\partial}{\partial x^n_{\beta(t)}} \right] f
$$

$$
= \dot{\beta}(t)f,
$$

where

$$
\dot{\beta}(t) = \frac{d}{dt} \beta^1(t) \frac{\partial}{\partial x^1_{\beta(t)}} + \cdots + \frac{d}{dt} \beta^n(t) \frac{\partial}{\partial x^n_{\beta(t)}} \in T_{\beta(t)} \mathbb{R}^n
$$

is called the velocity vector of $\beta$ at $t$.

We note that if $\beta(t)$ has coordinates

$$
\beta^i(t) = x^i(\beta(t))
$$

then $\dot{\beta}(t)$ has components

$$
\frac{d}{dt} \beta^i(t) = \frac{d}{dt} x^i(\beta(t))
$$

= rate of change of $x^i$ along $\beta$ at $t$ wrt basis $\frac{\partial}{\partial x^1_{\beta(t)}}, \ldots, \frac{\partial}{\partial x^n_{\beta(t)}}$.

In particular, if $\alpha = (\alpha^1, \ldots, \alpha^n) \in \mathbb{R}^n$ and $a = (a^1, \ldots, a^n) \in \mathbb{R}^n$ then the straight line through $a$ (see Figure 6.4) in the direction of $\alpha$:

$$
(a^1 + ta^1, \ldots, a^n + ta^n)
$$

Figure 6.3

Figure 6.4
has velocity vector at \( t = 0 \):
\[
\alpha^1 \frac{\partial}{\partial x^1_a} + \cdots + \alpha^n \frac{\partial}{\partial x^n_a} \in T_a \mathbb{R}^n.
\]
Thus each tangent vector is a velocity vector.

### 6.3 Differentials

**Definition.** If \( a \in \mathbb{R}^n \), and \( f \) is a \( C^\infty \) function on an open neighbourhood of \( a \) then the *differential of \( f \) at \( a \), denoted*

\[
df_a,
\]

is the linear form on \( T_a \mathbb{R}^n \) defined by

\[
\langle df_a, \dot{\beta}(t) \rangle = \frac{d}{dt} f(\beta(t)) = \dot{\beta}(t) f
\]

for any velocity vector \( \dot{\beta}(t) \), such that \( \beta(t) = a \).

Thus

(i) \( \langle df_{\beta(t)}, \dot{\beta}(t) \rangle \) = rate of change of \( f \) along \( \beta \) at \( t \) (see Figure 6.5),

(ii) \( \langle df_a, v \rangle = vf \) (for all \( v \in T_a \mathbb{R}^n \)) = rate of change of \( f \) along \( v \).

**Theorem 6.1.** \( dx^i_a, \ldots, dx^n_a \) is the basis of \( T_a \mathbb{R}^n^* \) dual to the basis \( \frac{\partial}{\partial x^1_a}, \ldots, \frac{\partial}{\partial x^n_a} \) for \( T_a \mathbb{R}^n \).

**Proof**

\[
\left\langle dx^i_a, \frac{\partial}{\partial x^j_a} \right\rangle = \left. \frac{\partial x^i}{\partial x^j}(a) \right| = \delta^i_j,
\]

as required.
**Definition.** If $V$ is open in $\mathbb{R}^n$ then a **covariant vector field** $\omega$ on $V$ is a function on $V$:

$$\omega : x \mapsto \omega_x \in T_x^*\mathbb{R}^n.$$ 

The covariant vector fields on $V$ can be added:

$$(\omega + \eta)_x = \omega_x + \eta_x,$$

and multiplied by elements of $C^\infty(V)$:

$$(f\omega)_x = f(x)\omega_x.$$ 

Each covariant vector field $\omega$ on $V$ can be written uniquely as

$$\omega_x = \beta_1(x)dx_1 + \cdots + \beta_n(x)dx_n.$$ 

Thus

$$\omega = \beta_1dx^1 + \cdots + \beta_ndx^n$$

(we confine ourselves to $\beta_i \in C^\infty(V)$).

If $f \in C^\infty(V)$ then the covariant vector field

$$df : x \mapsto df_x$$

is called the **differential** of $f$. Thus we have:

- **contravariant vector fields**:
  
  $$v = \alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n}, \quad \alpha^i \in C^\infty(V);$$

- **covariant vector fields**:
  
  $$\omega = \beta_1dx^1 + \cdots + \beta_ndx^n, \quad \beta \in C^\infty(V);$$

and more general **tensor fields**, e.g.

$$S = \alpha_{i}^{j}k dx^{i} \otimes \frac{\partial}{\partial x^{j}} \otimes dx^{k}, \quad \alpha_{i}^{j}k \in C^\infty(V),$$

a function on $V$ whose value at $x$ is

$$S_x = \alpha_{i}^{j}k(x)dx^{i}_x \otimes \frac{\partial}{\partial x^{j}_x} \otimes dx^{k}_x,$$

a tensor over $T_x\mathbb{R}^n$.

We can add, multiply and contract tensor fields pointwise (carrying out the operation at each point $x \in V$). For example:
(i) \((R + S)_x = R_x + S_x\),
(ii) \((R \otimes S)_x = R_x \otimes S_x\),
(iii) \((\text{contracted } S)_x = \text{contracted } (S_x)\),
(iv) \((fS)_x = f(x)S_x \quad f \in C^\infty(V)\).

Contracting the covariant vector field \(\omega = \beta_1 dx^1 + \cdots + \beta_n dx^n\) with the contravariant vector field \(v = \alpha^1 \frac{\partial}{\partial x^1} + \cdots + \alpha^n \frac{\partial}{\partial x^n}\) gives the scalar field
\[
\langle \omega, v \rangle = \beta_1 \alpha^1 + \cdots + \beta_n \alpha^n.
\]
In particular, if \(f \in C^\infty(V)\) has differential \(df\) then the scalar field
\[
\langle df, v \rangle = vf
\]
is the rate of change of \(f\) along \(v\).

If \(\omega = \beta_1 dx^1 + \cdots + \beta_n dx^n\) then
\[
\beta_i = \text{\(i^{th}\) component of } \omega = \left\langle \omega, \frac{\partial}{\partial x^i} \right\rangle.
\]
In particular:
\[
\text{\(i^{th}\) component of } df = \left\langle df, \frac{\partial}{\partial x^i} \right\rangle = \frac{\partial f}{\partial x^i}.
\]
Therefore
\[
df = \frac{\partial f}{\partial x^1} dx^1 + \cdots + \frac{\partial f}{\partial x^n} dx^n \quad \text{Chain Rule},
\]
rate of change of \(f = \frac{\partial f}{\partial x^1}\) rate of change of \(x^1 + \cdots + \frac{\partial f}{\partial x^n}\) rate of change of \(x^n\).

### 6.4 Transformation Law

A sequence
\[
y = (y^1, \ldots, y^n) \quad (y^i \in C^\infty(V))
\]
is called a \((C^\infty)\) coordinate system on \(V\) if
\[
V \rightarrow W \quad x \mapsto y(x) = (y^1(x), \ldots, y^n(x))
\]
maps $V$ homeomorphically onto an open set $W$ in $\mathbb{R}^n$, and if

$$x^i = F^i(y^1, \ldots, y^n),$$

where $F^i \in C^\infty(W)$.

![Diagram of homeomorphism](image)

**Figure 6.6**

*Example:* $(r, \theta)$ is a $C^\infty$ coordinate system on $\{(x, y) : y \text{ or } x > 0\}$ (see Figure 6.6), where $r = \sqrt{x^2 + y^2}$, $\theta$ unique solution of $x = r \cos \theta, y = r \sin \theta$ ($-\pi < \theta < \pi$).

If $a \in V$, and $\beta$ is the parametrised curve – the curve along which all $y^j (j \neq i)$ are constant, and $y^i$ varies by $t$ – such that

$$y(\beta(t)) = y(a) + te_i$$

![Diagram of parametrised curve](image)
(see Figure 6.7) then the velocity vector of $\beta$ at $t = 0$ is denoted:

$$\frac{\partial}{\partial y'_a}$$

Thus if $f$ is $C^\infty$ in a neighbourhood of $a$ then

$$\frac{\partial f}{\partial y'^i}(a) = \frac{d}{dt} f(\beta(t))|_{t=0} = \text{rate of change of } f \text{ along the curve } \beta.$$ 

If we write $f$ as a function of $y^1, \ldots, y^n$:

$$f = F(y^1, \ldots, y^n)$$

(say), then

$$\frac{\partial f}{\partial y'^i}(a) = \frac{d}{dt} f(\beta(t))|_{t=0} = \frac{d}{dt} F(y(a+t\epsilon_i))|_{t=0} = \frac{\partial F}{\partial x^i}(y(a)),$$

i.e. to calculate $\frac{\partial f}{\partial y'^i}(a)$ write $f$ as a function $F$ of $y^1, \ldots, y^n$, and calculate $\frac{\partial F}{\partial x^i}$ (partial derivative of $F$ wrt $i^{th}$ slot):

$$\frac{\partial f}{\partial y'^i} = \frac{\partial F}{\partial x^i}(y^1, \ldots, y^n).$$

Now if $\beta$ is any parametrised curve at $a$, with $\beta(t) = a$ (see Figure 6.8), then

$$\langle df_a, \beta(t) \rangle = \frac{d}{dt} f(\beta(t))$$

$$= \frac{d}{dt} F(y^1(\beta(t)), \ldots, y^n(\beta(t)))$$

$$= \sum_{i=1}^n \frac{\partial F}{\partial x^i}(y^1(\beta(t)), \ldots, y^n(\beta(t))) \frac{d}{dt} y^i(\beta(t))$$

$$= \sum_{i=1}^n \frac{\partial f}{\partial y^i}(\beta(t)) \langle dy'_a, \dot{\beta}(t) \rangle$$
Therefore
\[ df_a = \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}(a) dy_i^a. \]

Therefore
\[ df = \sum_{i=1}^{n} \frac{\partial f}{\partial y^i} dy^i. \]

The operators
\[ \frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_n} \]
are linearly independent, since \( \frac{\partial}{\partial y^i} y^j = \delta^i_j \). Therefore these operators form a basis for \( T_a \mathbb{R}^n \), with dual basis
\[ dy_1^a, \ldots, dy_n^a, \]

since \( \langle dy_i^a, \frac{\partial}{\partial y^i} \rangle = \frac{\partial y_i^a}{\partial y^j}(a) = \delta^i_j \).

If \( z^1, \ldots, z^n \) is a \( C^\infty \) coordinate system on \( W \) then on \( V \cap W \):
\[ dz_i = \sum_{i=1}^{n} \frac{\partial z_i}{\partial y^j} dy^j. \]

Therefore \( \frac{\partial z_i}{\partial y^j} \) is the transition matrix from basis \( \frac{\partial}{\partial y^i} \) to basis \( \frac{\partial}{\partial z^i} \). Therefore
\[ \frac{\partial}{\partial y^j} = \sum_{i=1}^{n} \frac{\partial z_i}{\partial y^j} \frac{\partial}{\partial z^i} \]
on \( V \cap W \).

If (say) \( g = g_{ij} dy^i \otimes dy^j \) is a tensor field on \( V \), with component \( g_{ij} \) wrt coordinates \( y^i \), then
\[ g = g_{ij} \left( \frac{\partial y^i}{\partial z^k} dz^k \right) \otimes \left( \frac{\partial y^j}{\partial z^l} dz^l \right) = \frac{\partial y^i}{\partial z^k} \frac{\partial y^j}{\partial z^l} g_{ij} dz^k \otimes dz^l, \]
6–11
using s.c., and therefore \( g \) has component

\[
\frac{\partial y^j}{\partial z^k} \frac{\partial y^i}{\partial z^l} g_{ij}
\]

wrt coordinates \( z^i \).

**Example:** On \( \mathbb{R}^n \):

(i) usual coordinates \( x, y \);

(ii) polar coordinates \( r, \theta \).

\[
x = r \cos \theta, \quad y = r \sin \theta.
\]

So

\[
dx = \frac{\partial x}{\partial r} dr + \frac{\partial x}{\partial \theta} d\theta = \cos \theta \, dr - r \sin \theta \, d\theta,
\]

\[
dy = \frac{\partial y}{\partial r} dr + \frac{\partial y}{\partial \theta} d\theta = \sin \theta \, dr + r \cos \theta \, d\theta.
\]

The matrix

\[
\begin{pmatrix}
\cos \theta & -r \sin \theta \\
\sin \theta & r \cos \theta
\end{pmatrix}
\]

is the transition matrix from \( r, \theta \) to \( x, y \):

\[
\frac{\partial}{\partial r} = \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y} = \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y},
\]

\[
\frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial}{\partial y} = -r \sin \theta \frac{\partial}{\partial x} + r \cos \theta \frac{\partial}{\partial y}.
\]
Chapter 7

Scalar Products

7.1 The Definition

Definition. A tensor of type $M \times M \to K$ is called a scalar product or (bilinear form) (i.e. two lower indices).

Example: The dot product $K^n \times K^n \to K$. Writing $X, Y$ as $n \times 1$ columns:

\[
((\alpha_1, \ldots, \alpha_n), (\beta_1, \ldots, \beta_n)) \mapsto \alpha_1\beta_1 + \cdots + \alpha_n\beta_n
\]

\[
(X, Y) \mapsto X^tY.
\]

7.2 Properties of Scalar Products

1. If ($\cdot, \cdot$) is a scalar product on $M$ with components $G = (g_{ij})$ wrt basis $u_i$, if $x$ has components $X = (\phi^i)$ and $y$ has components $Y = (\nu^j)$ ($g_{ij} = (u_i|u_j)$ and ($\cdot, \cdot$) = $g_{ij}\nu^i\otimes \nu^j$) then

\[
(x|y) = (\phi^i u_i|\nu^j u_j)
\]

\[
= \phi^i \nu^j (u_i|u_j)
\]

\[
= g_{ij} \phi^i \nu^j
\]

\[
= \left( \begin{array}{c} \phi^1 \\ \vdots \\ \phi^n \end{array} \right) \left( \begin{array}{c} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{array} \right) \left( \begin{array}{c} \nu^1 \\ \vdots \\ \nu^n \end{array} \right)
\]

\[
= X^t G Y.
\]

Note. The dot product has matrix $I$ wrt $e_i$, since $e_i.e_j = \delta^i_j$.  

7–1
2. If \( P = (p_{ij}) \) is the transition matrix to new basis \( w_i \) then new matrix of \( (\cdot|\cdot) \) is \( Q^tGQ \), where \( Q = P^{-1} \).

**Proof of This**

As a tensor with two lower indices, new components of \( (\cdot|\cdot) \) are:

\[
q^k_i q^l_j g_{kl} = q^k_i g_{ki} q^l_j = Q^tGQ.
\]

Check:

\[
(PX)^tQ^tGQ(Y) = X^tP^tQ^tGQY = X^tGY.
\]

\( \triangleleft \)

3. \( (\cdot|\cdot) \) is called **symmetric** if

\[
(x|y) = (y|x)
\]

for all \( x, y \). This is equivalent to \( G \) being a symmetric matrix \( G^t = G \):

\[
g_{ij} = (u_i|u_j) = (u_j|u_i) = g_{ji}.
\]

A symmetric scalar product defines an **associated quadratic form**

\[
F : M \to K
\]

by

\[
F(x) = (x|x) = X^tGX = \begin{pmatrix} \xi^1 & \cdots & \xi^n \end{pmatrix} \begin{pmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{pmatrix} \begin{pmatrix} \xi^1 \\ \vdots \\ \xi^n \end{pmatrix} = g_{ij} \xi^i \xi^j,
\]

i.e.

\[
F = \begin{pmatrix} u^1 & \cdots & u^n \end{pmatrix} \begin{pmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{pmatrix} \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix} = g_{ij} u^i u^j.
\]

\( u^i u^j \) is a product of linear forms, and is a function:

\[
(u^i u^j)(x) = u^i(x) u^j(x).
\]
Example: If \( x, y, z \) are coordinate functions on \( M \) then
\[
F = \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 3 & 2 & 3 \\ 2 & -7 & -1 \\ 3 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}
\]
\[
= 3x^2 - 7y^2 + 2z^2 + 4xy + 6xz - 2yz.
\]
(Thus quadratic form \( \equiv \) homogeneous \( 2^{nd} \) degree polynomial).

The quadratic form \( F \) determines the symmetric scalar product \( (\cdot|\cdot) \) uniquely because:
\[
(x + y|x + y) = (x|x) + (x|y) + (y|x) + (y|y),
\]
\[
2(x|y) = F(x + y) - F(x) - F(y) \quad \text{(if} 1 + 1 \neq 0),
\]
and \( g_{ij} = (u_i|u_j) \) are called the components of \( F \) wrt \( u_i \).

**Definition.** \( (\cdot|\cdot) \) is called *non-singular* if
\[
(x|y) = 0 \text{ for all } y \in M \Rightarrow x = 0,
\]
i.e.
\[
X^tGY = 0 \text{ for all } Y \in K^n \Rightarrow X = 0,
\]
i.e.
\[
X^tG = 0 \Rightarrow X = 0,
\]
i.e.
\[
\det G \neq 0.
\]

**Definition.** A tensor field \( (\cdot|\cdot) \) with two lower indices on an open set \( V \subset \mathbb{R}^n \):
\[
(\cdot|\cdot) = g_{ij}dy^i \otimes dy^j
\]
(say), \( y^i \) coordinates on \( V \), is called a *metric tensor* if
\[
(\cdot|\cdot)_x
\]
is a symmetric non-singular scalar product on \( T_x\mathbb{R}^n \) for each \( x \in V \), i.e.
\[
g_{ij} = g_{ji} \quad \text{and} \quad \det g_{ij} \text{ nowhere zero}.
\]
The associated field \( ds^2 \) of quadratic forms:
\[
ds^2 = g_{ij}dy^i dy^j
\]
is called the line-element associated with the metric tensor.
Example: On \( \mathbb{R}^n \) the usual metric tensor
\[
dx \otimes \dx + \dy \otimes \dy,
\]
with line element \( ds^2 = (dx)^2 + (dy)^2 \), has components
\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\]
wrt coordinates \( x, y \).

If
\[
v = v^1 \frac{\partial}{\partial x} + v^2 \frac{\partial}{\partial y}, \quad w = w^1 \frac{\partial}{\partial x} + w^2 \frac{\partial}{\partial y}
\]
then
\[
(v|w) = \begin{pmatrix} v^1 & v^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\
0 & 1 \end{pmatrix} \begin{pmatrix} w^1 \\
w^2 \end{pmatrix} = v^1 w^1 + v^2 w^2 \quad \text{(dot product)}
\]
\[
ds^2[v] = (v|v) = (v^1)^2 + (v^2)^2 = \|v\|^2 \quad \text{(Euclidean norm)}.
\]

If \( r, \theta \) are polar coordinates:
\[
x = r \cos \theta, \quad y = r \sin \theta,
\]
then
\[
dx = \cos \theta \, dr - r \sin \theta \, d\theta,
\]
\[
dy = \sin \theta \, dr + r \cos \theta \, d\theta
\]
and
\[
ds^2 = (dx)^2 + (dy)^2
\]
\[
= (\cos \theta \, dr - r \sin \theta \, d\theta)^2 + (\sin \theta \, dr + r \cos \theta \, d\theta)^2
\]
\[
= (dr)^2 + r^2 (d\theta)^2
\]
has components
\[
\begin{pmatrix}
1 & 0 \\
0 & r^2
\end{pmatrix}
\]
wrt coordinates \( r, \theta \).

If
\[
v = \alpha^1 \frac{\partial}{\partial r} + \alpha^2 \frac{\partial}{\partial \theta}, \quad w = \beta^1 \frac{\partial}{\partial r} + \beta^2 \frac{\partial}{\partial \theta}
\]
then
\[
(v|w) = \alpha^1 \beta^1 + r^2 \alpha^2 \beta^2,
\]
\[
\|v\|^2 = (\alpha^1)^2 + r^2 (\alpha^2)^2.
\]

7–4
7.3 Raising and Lowering Indices

**Definition.** Let $M$ be a finite dimensional vector space with a fixed non-singular symmetric scalar product $(\cdot|\cdot)$. If $x \in M$ is a vector (one upper index), we associate with it 

$$\tilde{x} \in M^*,$$

a linear form (one lower index) defined by:

$$\langle \tilde{x}, y \rangle = (x|y) \quad \text{for all } y \in M.$$

We call the operation

$$M \rightarrow M^*$$

$$x \mapsto \tilde{x}$$

*lowering the index.* Thus

$$\tilde{x} \equiv (x|\cdot) \equiv \text{‘take scalar product with } x\text{‘.}$$

If $x = \alpha^i u_i$ has components $\alpha^i$ then $\tilde{x}$ has components

$$\alpha_j = \langle \tilde{x}, u_j \rangle = (x|u_j) = (\alpha^i u_i|u_j) = \alpha^i (u_i|u_j) = \alpha^i g^i j.$$

Since $(\cdot|\cdot)$ is non-singular, $g_{ij}$ is invertible, with inverse $g^{ij}$ (say), and we have

$$\alpha^j = \alpha_i g^{ij}.$$

Thus

$$M \rightarrow M^*$$

$$x \mapsto \tilde{x}$$

is a linear isomorphism, with inverse

$$f \mapsto \tilde{f}$$

(say), called *raising the index.* So

$$x = \alpha^i u_i = \sim,$$

$$\tilde{x} = \alpha_i u^i = f$$

and

$$\langle x|y \rangle = \langle f|y \rangle = \langle f, y \rangle = \langle \tilde{x}, y \rangle.$$
To lower: contract with $g_{ij} \ (\alpha_j = \alpha^i g_{ij})$.

To raise: contract with $g^{ij} \ (\alpha^j = \alpha_i g^{ij})$.

Let $M \overset{T}{\to} M$ be a linear operator and $(\cdot | \cdot)$ be symmetric. The matrix of $T$ is:

$$\alpha^i_j = (u^i, Tu_j),$$

one up, one down mixed components of $T$.

$$\alpha_{ij} = (u_i | Tu_j),$$

two down covariant components of $T$.

$$\alpha_{ij} = (u_i | \alpha_j^k u_k) = (u_i | u_k) \alpha^k_j = g_{ik} \alpha^k_j$$

(lower by contraction with $g_{ij}$). Therefore

$$\alpha^i_j = g^{ik} \alpha_k j$$

(raise by contraction with $g^{ij}$).

If we take the covariant components $\alpha_{ij}$, and raise the second index we get

$$\alpha^i_j = \alpha_{ik} g^{kj}.$$

$\alpha_{ij}$ are the components of the tensor $B$ (two lower indices) defined by:

$$B(x, y) = (x | Ty),$$

since

$$B(u_i, u_j) = (u_i | Tu_j) = \alpha_{ij}.$$

$\alpha^i_j$ are the components of an operator $T^*$ (one upper index, one lower index) defined by:

$$(T^* x | y) = (x | Ty),$$

since $T^*$ has components

$$\gamma_{ij} = (u_i | T^* u_j) = (T^* u_j | u_i) = (u_j | Tu_i) = \alpha_{ji},$$

and therefore $T^*$ has mixed components:

$$\gamma^i_j = g^{ik} \gamma_{kj} = \alpha_{jk} g^{ki} = \alpha_j^i.$$

$T^*$ is called the adjoint of operator $T$.  

7–6
7.4 Orthogonality and Diagonal Matrix

**Definition.** If $(\cdot|\cdot)$ is a scalar product on $M$ and

$$(x|y) = 0,$$

we say that $x$ is **orthogonal to** $y$ wrt $(\cdot|\cdot)$.

If $N$ is a vector subspace of $M$, we write

$$N^\perp = \{x \in M : (x|y) = 0 \text{ for all } y \in N\},$$

and call it the **orthogonal complement of** $N$ wrt $(\cdot|\cdot)$ (see Figure 7.1).

![Figure 7.1](image)

We denote by $(\cdot|\cdot)_N$ the scalar product on $N$ defined by

$$(x|y)_N = (x|y) \text{ for all } x, y \in N,$$

and call it the **restriction of** $(\cdot|\cdot)$ **to** $N$.

**Definition.** Let $N_1, \ldots, N_k$ be vector subspaces of a vector space $M$. Then we write

$$N_1 + \cdots + N_k = \{x_1 + \cdots + x_k : x_1 \in N_1, \ldots, x_k \in N_k\},$$

and call it the **sum of** $N_1, \ldots, N_k$. Thus $M = N_1 + \cdots + N_k$ iff each $x \in M$ can be written as a sum

$$x = x_1 + \cdots + x_k, \quad x_i \in N_i.$$
We call $M$ a direct sum of $N_1, \ldots, N_k$, and write

$$M = N_1 \oplus \cdots \oplus N_k$$

if for each $x \in M$ there exists unique $(x_1, \ldots, x_k)$ (for example, see Figure 7.2) such that

$$x = x_1 + \cdots + x_k \quad \text{and} \quad x_i \in N_i.$$

![Figure 7.2](Image)

**Theorem 7.1.** Let $(\cdot, \cdot)$ be a scalar product on $M$. Let $N$ be a finite-dimensional vector subspace such that $(\cdot, \cdot)_N$ is non-singular. Then

$$M = N \oplus N^\perp.$$

**Proof** Let $x \in M$ (see Figure 7.3). Define $f \in N^*$ by

$$\langle f, y \rangle = (x | y)$$
for all $y \in N$.

Since $(\cdot|\cdot)_N$ is non-singular we can raise the index of $f$, and get a unique vector $z \in N$ such that

$$
\langle f, y \rangle = (z|y)
$$

for all $y \in N$, i.e.

$$
(x|y) = (z|y)
$$

for all $y \in N$, i.e.

$$
(x - z|y) = 0
$$

for all $y \in N$, i.e.

$$
x - z \in N^\perp,
$$

Figure 7.3

i.e.

$$
x = z + (x - z)\quad\forall z \in N^\perp
$$

uniquely, as required. ▶

**Lemma 7.1.** Let $(\cdot|\cdot)$ be a symmetric scalar product, not identically zero on a vector space $M$ over a field $K$ of characteristic $\neq 2$. (i.e. $1 + 1 \neq 0$). Then there exists $x \in M$ such that

$$
(x|x) \neq 0.
$$

**Proof** ▶ Choose $x, y \in M$ such that $(x|y) \neq 0$. Then

$$
(x + y|x + y) = (x|x) + (x|y) + (y|x) + (y|y).
$$

Hence $(x + y|x + y), (x|x), (y|y)$ are not all zero. Hence result. ▶

7–9
Theorem 7.2. Let $(\cdot|\cdot)$ be a symmetric scalar product on a finite-dimensional vector space $M$. Then $M$ has a basis of mutually orthogonal vectors:

$$(u_i|u_j) = 0 \quad \text{if } i \neq j,$$

i.e. the scalar product has a diagonal matrix

$$
\begin{pmatrix}
\alpha_1 & 0 & \cdots & 0 \\
0 & \alpha_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \alpha_n
\end{pmatrix},
$$

where $\alpha_i = (u_i|u_i)$.

**Proof** Theorem holds if $(x|y) = 0$ for all $x, y \in M$. So suppose $(\cdot|\cdot)$ is not identically zero.

Now we use induction on $\dim M$. Theorem holds if $\dim M = 1$. So assume $\dim M = n > 1$, and that the theorem holds for all spaces of dimension less than $n$.

Choose $u_1 \in M$ such that

$$(u_1|u_1) = \alpha_1 \neq 0.$$

Let $N$ be the subspace generated by $u_1$. $(\cdot|\cdot)_N$ has $1 \times 1$ matrix $(\alpha_1)$, and therefore is non-singular. Therefore

$$M = N \oplus N^\perp$$

$$\dim : n = 1 + n - 1.$$

By the induction hypothesis $N^\perp$ has basis

$$u_2, \ldots, u_n$$

(say) of mutually orthogonal vectors. Therefore $u_1, u_2, \ldots, u_n$ is a basis for $M$ of mutually orthogonal vectors, as required. \hfill \blacksquare

If $M$ is a complex vector space, we can put

$$w_i = \frac{u_i}{\sqrt{\alpha_i}}.$$

7–10
for each $\alpha_i > 0$. Then $(w_i | w_i) = 1$ or 0, and rearranging we have a basis wrt which $(\cdot | \cdot)$ has matrix

$$
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
1 & 0 & \cdots & 0
\end{pmatrix}
$$

($r \times r$ diagonal block top left), and the associated quadratic form is a sum of squares:

$$(w^1)^2 + \cdots + (w^r)^2.$$

If $M$ is a real vector space, we can put

$$w_i = \begin{cases} u_i / \sqrt{\alpha_i} & \alpha_i > 0; \\ u_i / \sqrt{-\alpha_i} & \alpha_i < 0; \\ u_i & \alpha_i = 0. \end{cases}$$

Then $(w_i | w_i) = \pm 1$ or 0, and rearranging we have a basis wrt which $(\cdot | \cdot)$ has matrix

$$
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
& 0 & \cdots & 0 \\
& & \ddots & \vdots \\
& & & 0
\end{pmatrix}
$$

and the associated quadratic form is a sum and difference of squares:

$$(w^1)^2 + \cdots + (w^r)^2 - (w^{r+1})^2 - \cdots - (w^{r+s})^2.$$
Example: Let $(\cdot, \cdot)$ be a scalar product on a 3-dimensional space $M$ which has matrix

$$A = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 0 & -1 \\ 2 & -1 & -3 \end{pmatrix}$$

wrt a basis with coordinate functions $x, y, z$.

To find new coordinate functions wrt which $(\cdot, \cdot)$ has a diagonal matrix.  

Method: Take the associated quadratic form

$$F = 4x^2 - 3z^2 + 4xy + 4xz - 2yz,$$

and write it as a sum and difference of squares, by ‘completing squares’. We have:

$$F = 4(x^2 + xy + xz) - 3z^2 - 2yz$$
$$= 4(x + \frac{1}{2}y + \frac{1}{2}z)^2 - y^2 - z^2 - 2yz - 3z^2 - 2yz$$
$$= 4(x + \frac{1}{2}y + \frac{1}{2}z)^2 - (y^2 + 4yz + 4z^2)$$
$$= 4(x + \frac{1}{2}y + \frac{1}{2}z)^2 - (y + 2z)^2 + 0z^2$$
$$= 4u^2 - v^2 + 0w.$$

Therefore $(\cdot, \cdot)$ has diagonal matrix

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

wrt to coordinate functions

$$u = x + \frac{1}{2}y + \frac{1}{2}z,$$
$$v = y + 2z,$$
$$w = z.$$ 

The transition matrix is

$$P = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$ 

Check: $P^tDP = A$?

$$\begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 0 & -1 \\ 2 & -1 & -3 \end{pmatrix}.$$
For a symmetric scalar product on a real vector space the number of + signs, and the number of − signs, when the matrix is diagonalised, is independent of the coordinates chosen:

**Theorem 7.3 (Sylvester’s Law of Inertia).** Let \(u_1, \ldots, u_n\) and \(w_1, \ldots, w_n\) be bases for a real vector space, and let

\[
F = (u^1)^2 + \cdots + (u^r)^2 - (u^{r+1})^2 - \cdots - (u^{r+s})^2
\]
\[
= (w^1)^2 + \cdots + (w^t)^2 - (w^{t+1})^2 - \cdots - (w^{t+k})^2
\]

be a quadratic form diagonalised by each of the two bases. Then \(r = t\) and \(s = k\).

**Proof** Suppose \(r \neq t, r > t\) (say). The space of solutions of the \(n - r + t\) homogeneous linear equations

\[
u^{r+1} = 0, \ldots, u^n = 0, w^1 = 0, \ldots, w^t = 0
\]

has dimension at least

\[
n - (n - r + t) = r - t > 0.
\]

Therefore there exists a non-zero solution \(x\) so

\[
F(x) = (u^1(x))^2 + \cdots + (u^r(x))^2 > 0
\]
\[
= -(w^{t+1}(x))^2 - \cdots - (w^{t+k}(x))^2 \leq 0,
\]

which is clearly a contradiction. Therefore \(r = t\), and similarly \(s = k\).

**7.5 Special Spaces**

**Definition.** A real vector space \(M\) with a symmetric scalar product \((\cdot | \cdot)\) is called a *Euclidean space* if the associated quadratic form is positive definite, i.e.

\[
F(x) = (x|x) > 0 \quad \text{for all } x \neq 0,
\]

i.e. there exists basis \(u_1, \ldots, u_n\) such that \((\cdot | \cdot)\) has matrix

\[
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1
\end{pmatrix}
\]

(all + signs).

7–13
\[ F = (u^1)^2 + \cdots + (u^n)^2, \]
\[ (u_i|u_j) = \delta^i_j, \]
i.e. \( u_1, \ldots, u_n \) is orthonormal.

We write
\[ \|x\| = \sqrt{(x|x)} \quad (x \in M), \]
and call it the norm of \( x \). We have
\[ \|x + y\| \leq \|x\| + \|y\| \quad (\text{Triangle Inequality}). \]
Thus \( M \) is a normed vector space, and therefore a metric space, and therefore a topological space.

The scalar product also satisfies:
\[ \|(x|y)\| \leq \|x\| \|y\| \quad (\text{Schwarz Inequality}). \]

We define the angle \( \theta \) between two non-zero vectors \( x, y \) by:
\[ \frac{(x|y)}{\|x\|\|y\|} = \cos \theta \quad (0 \leq \theta \leq \pi) \]
(see Figure 7.4).

![Figure 7.4](image)

If \( M \) is an \( n \)-dimensional vector space with scalar product having an orthonormal basis (e.g. a complex vector space or a Euclidean vector space) then the transition matrix \( P \) from one orthonormal basis to another satisfies:
\[ P_{\text{new}}^t P_{\text{old}} = I, \]
i.e.

\[ P^t P = I \]

i.e. \( P \) is an **orthogonal matrix**, i.e.

\[
\begin{pmatrix}
\cdots & i^{th} \text{ col of } P & \cdots \\
\vdots & \mathbf{j}^{th} \text{ col of } P & \vdots \\
\end{pmatrix}
\begin{pmatrix}
\vdots \\
0 \\
\vdots \\
0 \\
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 & 1 \\
\end{pmatrix},
\]

i.e.

\[(i^{th} \text{ col of } P).(j^{th} \text{ col of } P) = \delta_{ij},\]

i.e. the columns of \( P \) form an orthonormal basis of \( K^n \).

Also

\[ P \text{ orthonormal } \iff P^t = P^{-1} \]

\[ \iff PP^t = I \]

\[ \iff \text{ the rows of } P \text{ form an orthonormal basis of } K^n. \]

**Definition.** A real 4-dimensional vector space \( M \) with scalar product \((\cdot,\cdot)\) of type ++ -- is called a **Minkowski space**. A basis \( u_1, u_2, u_3, u_4 \) is called a **Lorentz basis** if wrt \( u_i \) the scalar product has matrix

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 \\
\end{pmatrix},
\]

i.e.

\[ F = (u^1)^2 + (u^2)^2 + (u^3)^2 - (u^4)^2. \]

The transition matrix \( P \) from one Lorentz basis to another satisfies:

\[
P^t \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 \\
\end{pmatrix} P = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 \\
\end{pmatrix}.
\]

Such a matrix \( P \) is called a **Lorentz matrix**.

**Example:** On \( \mathbb{C}^n \) we define the **hermitian dot product** \((x|y)\) of vectors

\[ x = (\alpha_1, \ldots, \alpha_n), \quad y = (\beta_1, \ldots, \beta_n) \]

\[ 7-15 \]
to be
\[(x|y) = \alpha_1\overline{\beta_1} + \cdots + \alpha_n\overline{\beta_n}.\]

This has the property of being positive definite, since:
\[(x|x) = \alpha_1\overline{\alpha_1} + \cdots + \alpha_n\overline{\alpha_n} = \|\alpha_1\|^2 + \cdots + \|\alpha_n\|^2 > 0 \quad \text{if } x \neq 0.\]

More generally:

**Definition.** If \(M\) is a complex vector space then a **hermitian scalar product** \((\cdot|\cdot)\) on \(M\) is a function
\[M \times M \to \mathbb{C}\]
such that
\[
\begin{align*}
(i) \quad (x + y|z) &= (x|z) + (y|z), \\
(ii) \quad (\alpha x|z) &= \alpha (x|z), \\
(iii) \quad (x|y + z) &= (x|y) + (x|z), \\
(iv) \quad (x|\alpha y) &= \overline{\alpha}(x|y), \\
(v) \quad (x|y) &= (y|x).
\end{align*}
\]
(i) and (ii) imply linear in the first variable, (iii) and (iv) imply conjugate-linear in the second variable, (v) implies conjugate-symmetric.

If, in addition, \((x|x) > 0\) for all \(x \neq 0\) then we call \((\cdot|\cdot)\) a **positive definite** hermitian scalar product.

**Definition.** A complex vector space \(M\) with a positive definite hermitian scalar product \((\cdot|\cdot)\) is called a **Hilbert space**.

**Note.** For a finite dimensional complex space \(M\) with an hermitian form \((\cdot|\cdot)\) we can prove (in exactly the same way as for a real space with symmetric scalar product):
1. There exists basis wrt which \((\cdot|\cdot)\) has matrix

\[
\begin{pmatrix}
1 & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 1 &\vdots \\
0 & \cdots & 0 & 0
\end{pmatrix}.
\]

2. The number of + signs and the number of \(-\) signs are each uniquely determined by \((\cdot|\cdot)\).

3. \(M\) is a Hilbert space iff all the signs are +.

Thus \(M\) is a Hilbert space iff \(M\) has an orthonormal basis. The transition matrix \(P\) from one orthonormal basis to another satisfies:

\[P_{\text{new}}^t P_{\text{old}} = I,\]

i.e.

\[P_{\text{new}}^t = P_{\text{old}} = I.\]

Such a matrix is called a unitary matrix.

A Hilbert space \(M\) is a normed space, hence a metric space, hence a topological space if we define:

\[\|x\| = \sqrt{(x|x)}.\]

To test how many +, − signs a quadratic form has we can use determinants:

**Example:**

\[F = ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}y\right)^2 + \frac{ac - b^2}{a}y^2\]
on a 2-dimensional space, with coordinate functions \(x, y\) and matrix \(\begin{pmatrix} a & b \\ b & c \end{pmatrix}\).

Therefore

\[
++ \iff a > 0, \quad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \\
-- \iff a < 0, \quad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \\
+- \iff a b & c \\ b & c \end{vmatrix} < 0.
\]

More generally:

**Theorem 7.4 (Jacobi’s Theorem).** Let \(F\) be a quadratic form on a real vector space \(M\), with symmetric matrix \(g_{ij}\) wrt basis \(u_i\). Suppose each of the determinants

\[
\Delta_i = \begin{vmatrix} g_{11} & \cdots & g_{1i} \\ \vdots & \ddots & \vdots \\ g_{i1} & \cdots & g_{ii} \end{vmatrix}
\]

is non-zero \((i = 1, \ldots, n)\). Then there exists a basis \(w_i\) such that \(F\) has matrix

\[
\begin{pmatrix}
\frac{1}{\Delta_1} \\
\frac{\Delta_1}{\Delta_2} \\
\ddots \\
\frac{\Delta_{n-1}}{\Delta_n}
\end{pmatrix}
\]

i.e.

\[
F = \frac{1}{\Delta_1} (w^1)^2 + \frac{\Delta_1}{\Delta_2} (w^2)^2 + \cdots + \frac{\Delta_{n-1}}{\Delta_n} (w^n)^2.
\]

Thus

\[
F \text{ is } +\text{ve definite } \iff \Delta_1, \Delta_2, \ldots, \Delta_n \text{ all positive}, \\
F \text{ is } -\text{ve definite } \iff \Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots.
\]

**Proof** \(F(x) = (x|x)\), where \((\cdot | \cdot)\) is a symmetric scalar product, \((u_i|u_j) = g_{ij}\).

Let

\[
N_i = S(u_1, \ldots, u_i).
\]

\((\cdot | \cdot)_{N_i}\) is non-singular, since \(\Delta_i \neq 0\) for \(i = 1, \ldots, n\).

Now

\[
\{0\} \subset N_1 \subset N_2 \subset \cdots \subset N_{i-1} \subset N_i \subset \cdots \subset N_n = M.
\]

7–18
Therefore

\[ N_i = N_{i-1} \oplus (N_i \cap N_{i-1}^\perp) \]

\[ \dim : i = (i - 1) + 1. \]

Choose non-zero \( w_i \in N_i \cap N_{i-1}^\perp \). Then

\[ w_1, \ldots, w_{i-1}, w_i, \ldots, w_n \]

are mutually orthogonal, and \( w_i \) is orthogonal to \( u_1, \ldots, u_{i-1} \). Therefore \( w_i \) is \textit{not} orthogonal to \( u_i \), since \(( \cdot | \cdot )\) is non-singular. Therefore we can choose \( w_i \) such that \(( u_i | w_i ) = 1\).

It remains to show that

\[ (w_i | w_i) = \frac{\Delta_{i-1}}{\Delta_i}. \]

To do this we write

\[ \lambda_1 u_1 + \cdots + \lambda_{i-1} u_{i-1} + \lambda_i u_i = w_i. \]

Taking scalar product with \( w_i, u_1, u_2, \ldots, u_i \) we get:

\[ 0 + \cdots + 0 + \lambda_i = (w_i | w_i) \]

\[ \lambda_1 g_{11} + \cdots + \lambda_{i-1} g_{1,i-1} + \lambda_i g_{1i} = 0 \]

\[ \lambda_1 g_{21} + \cdots + \lambda_{i-1} g_{2,i-1} + \lambda_i g_{2i} = 0 \]

\[ \vdots \]

\[ \lambda_1 g_{i-1,1} + \cdots + \lambda_{i-1} g_{i-1,i-1} + \lambda_i g_{i-1,i} = 0 \]

\[ \lambda_1 g_{i1} + \cdots + \lambda_{i-1} g_{i,i-1} + \lambda_i g_{ii} = 1 \]

Therefore

\[ (w_i | w_i) = \lambda_i = \begin{vmatrix} g_{11} & \cdots & g_{1,i-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ g_{i-1,1} & \cdots & g_{i-1,i-1} & 0 \\ g_{i1} & \cdots & g_{i,i-1} & 1 \end{vmatrix} = \frac{\Delta_{i-1}}{\Delta_i}, \]

as required. \( \blacksquare \)

This has an application in Calculus:
Theorem 7.5 (Criteria for local maxima or minima). Let \( f \) be a scalar field on a manifold \( X \) such that \( df_X = 0 \), and let \( y^i \) be coordinates on \( X \) at \( a \). Put

\[
\Delta_i = \begin{vmatrix}
\frac{\partial^2 f}{\partial y^1 y^i} & \cdots & \frac{\partial^2 f}{\partial y^n y^i} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial y^n y^i} & \cdots & \frac{\partial^2 f}{\partial y^n y^n}
\end{vmatrix}.
\]

Then

1. If \( \Delta_i(a) > 0 \) for \( i = 1, \ldots, n \) then there exists open nbd \( V \) of \( a \) such that

\[ f(x) > f(a) \quad \text{for all } x \in V, \ x \neq a, \]

i.e. \( a \) is a local minima of \( f \);

2. If \( \Delta_1(a) < 0, \Delta_2(a) > 0, \Delta_3(a) < 0, \ldots \) then there exists open nbd \( V \) of \( a \) such that

\[ f(x) < f(a) \quad \text{for all } x \in V, \ x \neq a, \]

i.e. \( a \) is a local maxima of \( f \).

To make sure that \( \|x\| = \sqrt{(x|x)} \) is a norm on a Euclidean or a Hilbert space we need to show that the triangle inequality holds.

**Theorem 7.6.** Let \( M \) be a Euclidean or a Hilbert space. Then

(i) \( (x|y) \leq \|x\| \|y\| \)  \( \quad \text{Schwarz} \),

(ii) \( \|x + y\| \leq \|x\| + \|y\| \)  \( \quad \text{Triangle} \).

**Proof**

(i) Let \( x, y \in M \). Then

\[ (x|y) = |(x|y)|e^{i\theta}, \quad (y|x) = |(x|y)|e^{-i\theta} \]

(say). So for all \( \lambda \in \mathbb{R} \) we have:

\[
0 \leq (\lambda e^{-i\theta} x + y|\lambda e^{-i\theta} x + y) \\
= \|x\|^2 \lambda^2 + \lambda e^{-i\theta} (x|y) + \lambda e^{i\theta} (y|x) + \|y\|^2 \\
= \|x\|^2 + 2\lambda |(x|y)| + \|y\|^2.
\]

Therefore

\[ |(x|y)|^2 \leq \|x\|^2 \|y\|^2 \quad (\lambda^2 \leq 4ac). \]

Therefore

\[ |(x|y)| \leq \|x\| \|y\|. \]
(ii)

\[ \|x + y\|^2 = (x + y|x + y) \]
\[ = \|x\|^2 + (x|y) + (y|x) + \|y\|^2 \]
\[ \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \]
\[ = (\|x\| + \|y\|)^2. \]

Therefore

\[ \|x + y\| \leq \|x\| + \|y\|. \]
Chapter 8

Linear Operators 2

8.1 Adjoint and Isometries

Let $M$ be a finite dimensional vector space with a fixed non-singular symmetric or hermitian scalar product $(\cdot | \cdot)$. Recall that if

$$M \xrightarrow{T} M$$

is a linear operator then the adjoint of $T$ is the operator

$$M \xrightarrow{T^*} M,$$

which satisfies

$$(x | Ty) = (T^*x | y)$$

for all $x, y \in M$.

If $(\cdot | \cdot)$ has matrix $G$ wrt basis $u_i$ and $T$ has matrix $A$ then $T^*$ has matrix

$$A^* = G^{-1}A^tG \quad (G^{-1}A^tG \text{ in hermitian case})$$

because

$$X^tGAY = X^tGAG^{-1}GY = [G^{-1}A^tGX]^tGY,$$

and similarly

$$X^tG\overline{A}Y = X^tG\overline{A}G^{-1}G\overline{Y} = [G^{-1}A^tGX]^tG\overline{Y}.$$

8-1
Examples:

1. $M$ Euclidean, basis orthonormal:
   
   \[ A^* = A^t. \]

2. $M$ Hilbert, basis orthonormal:

   \[ A^* = \overline{A}^t. \]

3. $M$ Minkowski, basis Lorentz:

   \[
   A^* = \begin{pmatrix}
   1 & 0 & 0 & 0 \\
   0 & 1 & 0 & 0 \\
   0 & 0 & 1 & 0 \\
   0 & 0 & 0 & -1
   \end{pmatrix}
   \begin{pmatrix}
   1 & 0 & 0 & 0 \\
   0 & 1 & 0 & 0 \\
   0 & 0 & 1 & 0 \\
   0 & 0 & 0 & -1
   \end{pmatrix}.
   

Definition. An operator $M \xrightarrow{T} M$ is called an isometry if

   \[(T x | Ty) = (x | y) \quad \text{for all } x, y \in M,\]

   i.e. $T$ preserves $( \cdot | \cdot )$, i.e.

   \[(T^* T x | y) = (x | y),\]

   i.e.

   \[T^* T = 1,\]

   i.e.

   \[T^* = T^{-1}.\]

Examples:

1. $M$ Euclidean, basis orthonormal, $A$ matrix of $T$:

   \[ T \text{ is an isometry } \iff A^t A = I, \]

   i.e. $A$ is an orthogonal matrix.

2. $M$ Hilbert, basis orthonormal, $A$ matrix of $T$:

   \[ T \text{ is an isometry } \iff \overline{A}^t A = I, \]

   i.e. $A$ is a unitary matrix.
3. M Minkowski, basis Lorentz, A matrix of T:

\[ T \text{ is an isometry } \iff GA^tGA = I \iff A^tGA = G, \]

i.e. \( A \) is a Lorentz matrix.

**Definition.** An isometry of a Euclidean space is called an *orthogonal transformation*. An isometry of a Hilbert space is called a *unitary transformation*. An isometry of a Minkowski space is called a *Lorentz transformation*.

**Definition.** An operator \( M \xrightarrow{T} M \) is called *self-adjoint* if

\[ T^* = T, \]

i.e.

\[ (Tx|y) = (x|Ty) \quad \text{for all } x, y \in M, \]

i.e.

\[ (u_i|Tu_j) = (Tu_i|u_j) = (u_j|Tu_i), \]

i.e. covariant components of \( T \) are symmetric.

(In quantum mechanics physical quantities are always represented by self-adjoint operators).

**Examples:**

1. \( M \) Euclidean, basis orthonormal, A matrix of \( T \):

\( T \) is self-adjoint \( \iff A^t = A, \)

i.e. \( A \) is symmetric.

2. \( M \) Hilbert, basis orthonormal, A matrix of \( T \):

\( T \) is self-adjoint \( \iff \overline{A} = A, \)

i.e. \( A \) is hermitian.

3. \( M \) Minkowski, basis Lorentz, A matrix of \( T \):

\( T \) is self-adjoint \( \iff GA^tG = A. \)

**Summary:** Let \( M \xrightarrow{T} M \) have matrix \( A \) wrt orthonormal or Lorentz basis. Then:

<table>
<thead>
<tr>
<th>Space:</th>
<th>Euclidean</th>
<th>Hilbert</th>
<th>Minkowski</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix of ( T^* ):</td>
<td>( A^t )</td>
<td>( \overline{A} )</td>
<td>( GA^tG )</td>
</tr>
<tr>
<td>( T ) self-adjoint:</td>
<td>( A^t = A )</td>
<td>( \overline{A} = A )</td>
<td>( GA^tG = A )</td>
</tr>
<tr>
<td>( A ) symmetric</td>
<td>( A ) Hermitian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T ) an isometry:</td>
<td>( A^tA = I )</td>
<td>( \overline{A}A = I )</td>
<td>( A^tGA = G )</td>
</tr>
<tr>
<td>( A ) orthogonal</td>
<td>( A ) unitary</td>
<td>( A ) Lorentz</td>
<td></td>
</tr>
</tbody>
</table>
8.2 Eigenvalues and Eigenvectors

Definition. A vector space \( N \subset M \) is called invariant under a linear operator \( M \xrightarrow{T} M \) if

\[ T(N) \subset N, \]

i.e. \( x \in N \Rightarrow Tx \in N \).

A non-zero vector in a 1-dimensional invariant subspace under \( T \) is called an eigenvector of \( T \):

(i) \( x \in M \) is called an eigenvector of \( T \), with eigenvalue \( \lambda \) if

(a) \( x \neq 0 \),

(b) \( Tx = \lambda x \), where \( \lambda \) is a scalar (see Figure 8.1);

(ii) \( \lambda \in K \) is called an eigenvalue of \( T \) if there exists \( x \neq 0 \) such that

\[ Tx = \lambda x, \]

i.e.

\[ (T - \lambda I)x = 0, \]

i.e.

\[ \ker(T - \lambda I) \neq \{0\}. \]

\[ \ker(T - \lambda I) = \{x \in M : Tx = \lambda x\} \]

is called the \( \lambda \)-eigenspace of \( T \). It is the vector subspace consisting of all eigenvectors of \( T \) having eigenvalue \( \lambda \), together with the zero vector.
**Definition.** If $M \xrightarrow{T} M$ is a linear operator on a vector space of finite dimension $n$, with matrix $A = (\alpha_{ij})$ wrt basis $u_i$, then the polynomial of degree $n$ with coefficients in $K$:

$$\operatorname{char} T = \det \begin{vmatrix} \alpha_1^1 - X & \alpha_1^2 & \ldots & \alpha_1^n \\ \alpha_2^1 & \alpha_2^2 - X & \ldots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_n^1 & \ldots & \alpha_n^n - X \end{vmatrix} = \det (A - XI)$$

is called the **characteristic polynomial** of $T$.

$\operatorname{char} T$ is well-defined, independent of choice of basis $u_i$, since if $B$ is the matrix of $T$ wrt another basis then

$$B = PAP^{-1}.$$  

Therefore

$$\det (B - XI) = \det (PAP^{-1} - XI)$$

$$= \det P (A - XI) P^{-1}$$

$$= \det P \det (A - XI) \det P^{-1}$$

$$= \det (A - XI),$$

since $\det P \det P^{-1} = \det PP^{-1} = \det I = 1$.

**Theorem 8.1.** If $M \xrightarrow{T} M$ is a linear operator and $\dim M < \infty$ and $\lambda \in K$ then

$\lambda$ is an eigenvalue of $T$ $\iff$ $\lambda$ is a zero of $\operatorname{char} T$.

**Proof ▶** Let $T$ have matrix $A$ wrt basis $u_i$. Then

$\lambda$ is an eigenvalue of $T$ $\iff$ there exists $y \in M$ such that $(T - \lambda I)y = 0$

$\iff$ there exists $Y \in K^n$ such that $(A - \lambda I)Y = 0$

$\iff \det (A - \lambda I) = 0$

$\iff \lambda$ is a zero of $\det (A - XI)$.

**Corollary 8.1.** If $T$ is a linear operator on a finite dimensional complex space then $T$ has an eigenvalue, and therefore eigenvectors.
Theorem 8.2. Let $M \xrightarrow{T} M$ be a linear operator on a finite dimensional vector space $M$. Then $T$ has a diagonal matrix

$$
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix}
$$

wrt a basis $u_1, \ldots, u_n$ iff $u_i$ is an eigenvector of $T$, with eigenvalue $\lambda_i$, for $i = 1, \ldots, n$.

Proof

$$
Tu_1 = \lambda_1 u_1 + 0u_2 + \cdots + 0u_n \\
Tu_2 = 0u_1 + \lambda_2 u_2 + \cdots + 0u_n \\
\vdots \\
Tu_n = 0u_1 + 0u_2 + \cdots + \lambda_n u_n,
$$

hence result.

Theorem 8.3. Let $M \xrightarrow{T} M$ be a self-adjoint operator on a Hilbert space $M$. Then all the eigenvalues of $T$ are real.

Proof

Let $Tx = \lambda x, \quad x \neq 0, \quad \lambda \in \mathbb{C}$. Then

$$
\lambda(x|x) = (\lambda x|x) = (Tx|x) = (x|Tx) = (x|\lambda x) = \overline{\lambda}(x|x).
$$

$(x|x) \neq 0$. Therefore $\lambda = \overline{\lambda}$. Therefore $\lambda$ is real.

Corollary 8.2. Let $A$ be a hermitian matrix. Then $\mathbb{C}^n \xrightarrow{A} \mathbb{C}^n$ is a self-adjoint operator wrt hermitian dot product. Therefore all the roots of the equation

$$
\det(A - XI) = 0
$$

are real.

Corollary 8.3. Let $T$ be a self-adjoint operator on a finite dimensional Euclidean space. Then $T$ has an eigenvector.

Proof

Wrt an orthonormal basis $T$ has a real symmetric matrix $A$:

$$
\overline{A}^t = A^t = A.
$$

Therefore $A$ is hermitian. Therefore $\det(A - XI) = 0$ has real roots. Therefore $T$ has an eigenvalue. Therefore $T$ has eigenvectors.
Theorem 8.4. Let \( N \) be invariant under a linear operator \( M \to T M \). Then \( N^\perp \) is invariant under \( T^* \).

Proof ▶ Let \( x \in N^\perp \). Then for all \( y \in N \) we have:

\[
(T^*x|y) = (x|Ty) = 0.
\]

Therefore \( T^*x \in N^\perp \). ▶

Definition. \( M \to T M \) is a normal operator if

\[
T^*T = TT^*,
\]
i.e. \( T \) commutes with \( T^* \).

Examples:

(i) \( T \) self-adjoint ⇒ \( T \) normal.

(ii) \( T \) an isometry ⇒ \( T \) normal.

Theorem 8.5. Let \( S, T \) be commuting linear operators \( M \to M \) \((ST = TS)\). Then each eigenspace of \( S \) is invariant under \( T \).

Proof ▶

\[
Sx = \lambda x \Rightarrow S(Tx) = T(Sx) = T(\lambda x) = \lambda(Tx),
\]
i.e. \( x \in \lambda \)-eigenspace of \( S \) ⇒ \( Tx \in \lambda \)-eigenspace of \( S \). ▶

8.3 Spectral Theorem and Applications

Theorem 8.6 (Spectral theorem). Let \( M \to T M \) be either a self-adjoint operator on a finite dimensional Euclidean space or a normal operator on a finite dimensional Hilbert space. Then \( M \) has an orthonormal basis of eigenvectors of \( T \).

Proof ▶ (By induction on \( \dim M \)). True for \( \dim M = 1 \). Let \( \dim M = n \), and assume the theorem holds for spaces of dimension \( \leq n - 1 \).

Let \( \lambda \) be an eigenvalue of \( T \), \( M_\lambda \) the \( \lambda \)-eigenspace. \( \langle \cdot | \cdot \rangle_{M_\lambda} \) is non-singular, since \( \langle \cdot | \cdot \rangle \) is positive definite. Therefore

\[
M = M_\lambda \oplus M_\lambda^\perp.
\]
$M_\lambda$ is $T$-invariant. Therefore $M_\lambda^\perp$ is $T^*$-invariant. $T^*$ commutes with $T$. Therefore $M_\lambda$ is $T^*$-invariant. Therefore $M_\lambda^\perp$ is $T$-invariant.

Now

$$(T^* x | y) = (x | Ty)$$

for all $x, y \in M_\lambda^\perp$. Therefore $(T^*)_M^\perp$ is the adjoint of $T_{M_\lambda^\perp}$. Therefore

$T$ self-adjoint $\Rightarrow T_{M_\lambda^\perp}$ is self-adjoint,

and

$T$ normal $\Rightarrow T_{M_\lambda^\perp}$ is normal.

But $\dim M_\lambda^\perp \leq n - 1$. Therefore, by induction hypothesis $M_\lambda^\perp$ has an orthonormal basis of eigenvectors of $T$. Therefore $M = M_\lambda \oplus M_\lambda^\perp$ has an orthonormal basis of eigenvectors of $T$. \hfill \blacksquare

Applications:

1. Let $A$ be a real symmetric $n \times n$ matrix. Then

   (i) $\mathbb{R}^n$ has an orthonormal basis of eigenvectors $u_1, \ldots, u_n$ of $A$, with eigenvalues $\lambda_1, \ldots, \lambda_n$ (say),

   (ii) if $P$ is the matrix having $u_1, \ldots, u_n$ as rows then $P$ is an orthogonal matrix and

\[
    PAP^{-1} = \begin{pmatrix}
    \lambda_1 & 0 & \cdots & 0 \\
    0 & \lambda_2 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \vdots \\
    0 & \cdots & 0 & \lambda_n
\end{pmatrix} = Q^tAQ,
\]

   where $Q = P^{-1}$.

Proof $\blacktriangleleft$ $\mathbb{R}^n$ is a Euclidean space wrt the dot product, $e_1, \ldots, e_n$ is an orthonormal basis. Operator $\mathbb{R} \stackrel{A}{\rightarrow} \mathbb{R}^n$ has symmetric matrix $A$ wrt orthonormal basis $e_1, \ldots, e_n$. Therefore $A$ is self-adjoint. Therefore $\mathbb{R}^n$ has an orthonormal basis $u_1, \ldots, u_n$, with eigenvalues $\lambda_1, \ldots, \lambda_n$.

Let $P$ be the transition matrix from orthonormal $e_1, \ldots, e_n$ to orthonormal $u_1, \ldots, u_n$, with inverse matrix $Q$. $P$ is an orthogonal matrix, and therefore

\[
    Q = P^{-1} = P^t.
\]

8–8
Q is the transition matrix from $u_i$ to $e_i$. Therefore

\[ u_j = q^1_j e_1 + \cdots + q^n_j e_n = (q^1_j, \ldots, q^n_j) = j^{th} \text{ column of } Q = j^{th} \text{ row of } P. \]

Matrix of operator $A$ wrt basis $u_i$ is:

\[
PAP^{-1} = PAP^t = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.
\]

\[ \text{Note. If } F \text{ has matrix } A = (\alpha_{ij}), \text{ and } (\cdot|\cdot) \text{ has matrix } G = (g_{ij}) \text{ then } T \text{ has matrix } (\alpha^i_j) = (g^{ik}\alpha_{kj}) = G^{-1}A. \]

2. **Principal axes theorem** Let $F$ be a quadratic form on a finite dimensional Euclidean space $M$. Then $M$ has an orthonormal basis $u_1, \ldots, u_n$ which diagonalises $F$:

\[ F = \lambda_1(u^1)^2 + \cdots + \lambda_n(u^n)^2. \]

Such a basis is called a set of principal axes for $F$.

**Proof** Let $F(x) = B(x, x)$, where $B$ is a symmetric bilinear form. Raising an index of $B$ gives a self-adjoint operator $T$:

\[ (x|Ty) = B(x, y) = (Tx|y). \]

Let $u_1, \ldots, u_n$ be an orthonormal basis of $M$ of eigenvectors of $T$, with eigenvalues $\lambda_1, \ldots, \lambda_n$ (say). Then wrt $u_i$ the quadratic form $F$ has matrix:

\[ B(u_i, u_j) = (u_i|Tu_j) = (u_i|\lambda_j u_j) = \lambda_j \delta^i_j, \]

i.e.

\[
\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix},
\]

as required. \[ \text{Note. If } F \text{ has matrix } A = (\alpha_{ij}), \text{ and } (\cdot|\cdot) \text{ has matrix } G = (g_{ij}) \text{ then } T \text{ has matrix } (\alpha^i_j) = (g^{ik}\alpha_{kj}) = G^{-1}A. \]

8-9
Therefore $\lambda_1, \ldots, \lambda_n$ are the roots of

$$\det(G^{-1}A - XI) = 0,$$

i.e.

$$\det(A - XG) = 0.$$

3. Consider the surface

$$ax^2 + by^2 + cz^2 + 2hxy + 2gxz + 2fyz = k \quad (k > 0)$$

in $\mathbb{R}^3$. The LHS is a quadratic form with matrix

$$A = \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}$$

wrt usual coordinate functions $x, y, z$. By the principal axes theorem we can choose new orthonormal coordinates $X, Y, Z$ such that equation becomes:

$$\lambda_1X^2 + \lambda_2Y^2 + \lambda_3Z^2 = k,$$

where $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues of $A$.

The surface is:

an **ellipsoid** if $\lambda_1, \lambda_2, \lambda_3$ are all $> 0$, i.e. if the quadratic form is positive definite, i.e.

$$a > 0, \quad ab - h^2 > 0, \quad \det A > 0 \quad \text{by Jacobi;}$$

a **hyperboloid of 1-sheet** (see Figure 8.2) if the quadratic form is of type $++-$ (e.g. $X^2 + Y^2 = Z^2 + 1$), i.e.

$$a > 0, \quad ab - h^2 > 0, \quad \det A < 0$$

or

$$a > 0, \quad ab - h^2 < 0, \quad \det A < 0$$

or

$$a < 0, \quad ab - h^2 < 0, \quad \det A < 0;$$
a hyperboloid of 2-sheets (see Figure 8.3) if the quadratic form is of type $+--$ (e.g. $X^2 + Y^2 = Z^2 - 1$), i.e.

- $a > 0$, $ab - h^2 < 0$, $\det A > 0$
- or $a < 0$, $ab - h^2 < 0$, $\det A > 0$
- or $a < 0$, $ab - h^2 > 0$, $\det A > 0$. 

Figure 8.2
Figure 8.3
Chapter 9

Skew-Symmetric Tensors and Wedge Product

9.1 Skew-Symmetric Tensors

Definition. A bijective map

\[ \sigma : \{1, 2, \ldots, r\} \rightarrow \{1, 2, \ldots, r\} \]

is called a permutation of degree \( r \). The group \( S_r \) of all permutations of degree \( r \) is called the symmetric group of degree \( r \). Thus \( S_r \) is a group of order \( r! \).

Let \( T^rM \) denote the space of all tensors over \( M \) of type

\[ M \times M \times \cdots \times M \rightarrow K. \]

Thus \( T^rM \) consists of all tensors \( T \) with components

\[ T(u_{i_1}, \ldots, u_{i_r}) = \alpha_{i_1 \ldots i_r}, \quad (r \text{ lower indices}), \]

\[ T = \alpha_{i_1 \ldots i_r} u^{i_1} \otimes \cdots \otimes u^{i_r}. \]

\( u^{i_1} \otimes \cdots \otimes u^{i_r} \) is a basis for \( T^rM \).

For each \( \sigma \in S_r \), and each \( T \in T^rM \) we define \( \sigma T \in T^rM \) by:

\[ (\sigma T)(x_1, \ldots, x_r) = T(x_{\sigma(1)}, \ldots, x_{\sigma(r)}). \]

If \( T \) has components \( \alpha_{i_1 \ldots i_r} \) then \( \sigma T \) has components \( \beta_{i_1 \ldots i_r} \), where

\[ \beta_{i_1 \ldots i_r} = (\sigma T)(u_{i_1}, \ldots, u_{i_r}) = T(u_{\sigma(i_1)}, \ldots, u_{\sigma(i_r)}) = \alpha_{i_{\sigma(1)} \ldots i_{\sigma(r)}}. \]
Theorem 9.1. The group $S_r$ acts on $T^r M$ by linear transformations, i.e.

(i) $\sigma.(\alpha T + \beta S) = \alpha(\sigma.T) + \beta(\sigma.S)$,

(ii) $\sigma.(\tau.T) = (\sigma\tau).T$,

(iii) $1.T = T$

for all $\alpha, \beta \in K$, $\sigma, \tau \in S_r$, $S, T \in T^r M$.

Proof ▶ e.g. (ii)

$$[\sigma.(\tau.T)](x_1, \ldots, x_r) = (\tau.T)[x_{\sigma(1)}, \ldots, x_{\sigma(r)}]$$

$$= T(x_{\sigma(\tau(1))}, \ldots, x_{\sigma(\tau(r))})$$

$$= [(\sigma\tau).T](x_1, \ldots, x_r).$$

Therefore $\sigma.(\tau.T) = (\sigma\tau).T$. ◀

Note. If $\sigma \in S_r$, we put

$$\epsilon^\sigma = \begin{cases} +1 & \text{if } \sigma \text{ is an even permutation;} \\ -1 & \text{if } \sigma \text{ is an odd permutation} \end{cases} = \text{sign of } \sigma.$$

We have:

(i) $\epsilon^{\sigma\tau} = \epsilon^\sigma \epsilon^\tau$,

(ii) $\epsilon^1 = 1$,

(iii) $\epsilon^\sigma^{-1} = \epsilon^\sigma$.

Definition. $T \in T^r M$ is skew-symmetric if

$$\sigma.T = \epsilon^\sigma T \quad \text{for all } \sigma \in S_r,$$

i.e.

$$T(x_{\sigma(1)}, \ldots, x_{\sigma(r)}) = \epsilon^\sigma T(x_1, \ldots, x_r)$$

for all $\sigma \in S_r$, $x_1, \ldots, x_r \in M$, i.e. the components $\alpha_{i_1 \ldots i_r}$ of $T$ satisfy:

$$\alpha_{i_{\sigma(1)} \ldots i_{\sigma(r)}} = \epsilon^\sigma \alpha_{i_1 \ldots i_r}.$$

Example: $T \in T^3 M$, with components $\alpha_{ijk}$ is skew-symmetric iff

$$\alpha_{ijk} = -\alpha_{jik} = \alpha_{kji} = -\alpha_{kij} = \alpha_{kji} = -\alpha_{ikj}.$$  

It follows that if $T$ is skew-symmetric, with components $\alpha_{i_1 \ldots i_r}$ (from now on assume $K$ has characteristic zero, i.e. $\alpha \neq 0 \Rightarrow \alpha + \alpha + \cdots + \alpha \neq 0$) then
1. \( \alpha_{i_1 \ldots i_r} = 0 \) if \( i_1, \ldots, i_r \) are not all distinct;

2. if we know \( \alpha_{i_1 \ldots i_r} \) for all increasing sequences \( i_1 < \cdots < i_r \) then we know \( \alpha_{i_1 \ldots i_r} \) for all sequences \( i_1, \ldots, i_r \);

3. if \( T \) is skew-symmetric, with components \( \alpha_{i_1 \ldots i_r} \) and \( S \) is skew-symmetric, with components \( \beta_{i_1 \ldots i_r} \), and if \( \alpha_{i_1 \ldots i_r} = \beta_{i_1 \ldots i_r} \) for all increasing sequences \( i_1 < \cdots < i_r \) then \( T = S \).

**Theorem 9.2.** Let \( T \in T^r M \). Then

\[
\sum_{\sigma \in S_r} e^\sigma \sigma.T
\]

is skew-symmetric.

**Proof** Let \( \tau \in S_r \). Then

\[
\tau \left( \sum_{\sigma \in S_r} e^\sigma \sigma.T \right) = e^\tau \sum_{\sigma \in S_r} e^{\tau \sigma} (\tau \sigma).T = e^\tau \sum_{\sigma \in S_r} \sigma e^\sigma (\sigma.T),
\]

as required. ▶

**Definition.** The linear operator

\[ A : T^r M \rightarrow T^r M \]

defined by

\[ AT = \frac{1}{r!} \sum_{\sigma \in S_r} e^\sigma \sigma.T \]

is called the skew-symmetriser.

**Example:** Let \( T \in T^3 M \) have components \( \alpha_{ijk} \). Then

\[
AT(x, y, z) =
\frac{1}{6} [T(x, y, z) - T(y, x, z) + T(y, z, x) - T(x, z, y) + T(z, x, y) - T(z, y, x)],
\]

and \( AT \) has components

\[
\beta_{ijk} = \frac{1}{6} (\alpha_{ijk} - \alpha_{ikj} + \alpha_{jki} - \alpha_{jik} + \alpha_{kij} - \alpha_{kji}).
\]
Theorem 9.3. Let \( S \in T^*M, \ T \in T'M \). Then

(i) \( \mathcal{A}[(\mathcal{A}S) \otimes T] = \mathcal{A}[S \otimes T] = \mathcal{A}[S \otimes \mathcal{A}T], \)

(ii) \( \mathcal{A}(S \otimes T) = (-1)^s \mathcal{A}(T \otimes S). \)

Proof

(i) We first note that if \( \tau \in \mathcal{S}_s \) then

\[
[(\tau.S) \otimes T](x_1, \ldots, x_s, x_{s+1}, \ldots, x_{s+t}) = (\tau.S)(x_1, \ldots, x_s)T(x_{s+1}, \ldots, x_{s+t})
\]

\[
= S(x_{\tau(1)}, \ldots, x_{\tau(s)})T(x_{s+1}, \ldots, x_{s+t})
\]

\[
= S(x_{\tau'(1)}, \ldots, x_{\tau'(s)})T(x_{\tau'(s+1)}, \ldots, x_{\tau'(s+t)})
\]

\[
= [\tau'.(S \otimes T)](x_1, \ldots, x_s, x_{s+1}, \ldots, x_{s+t}),
\]

where

\[
\tau' = \begin{pmatrix}
1 & \ldots & s & s + 1 & \ldots & s + t \\
\tau(1) & \ldots & \tau(s) & s + 1 & \ldots & s + t
\end{pmatrix}.
\]

Thus

\[
(\tau.S) \otimes T = \tau'.(S \otimes T)
\]

and \( e^{\tau'} = e^\tau. \)

Now

\[
\mathcal{A}[(\mathcal{A}S) \otimes T] = \frac{1}{(s+t)!} \sum_{\sigma \in \mathcal{S}_{s+t}} e^\sigma \left[ \left( \frac{1}{s!} \sum_{\tau \in \mathcal{S}_s} e^{\tau.S} \right) \otimes T \right]
\]

\[
= \frac{1}{s!} \sum_{\tau \in \mathcal{S}_s} \frac{1}{(s+t)!} \sum_{\sigma \in \mathcal{S}_{s+t}} e^{\sigma \tau'} (\sigma \tau').(S \otimes T)
\]

\[
= \frac{1}{s!} \sum_{\tau \in \mathcal{S}_s} \mathcal{A}(S \otimes T)
\]

\[
= \mathcal{A}(S \otimes T).
\]

(ii) Let

\[
\tau = \begin{pmatrix}
1 & \ldots & s & s + 1 & \ldots & s + t \\
t + 1 & \ldots & t + s & 1 & \ldots & t
\end{pmatrix}
\]

so that \( e^\tau = (-1)^t \). Then

\[
[\tau.(S \otimes T)](x_1, \ldots, x_s, x_{s+1}, \ldots, x_{s+t}) = S \otimes T[x_{t+1}, \ldots, x_{t+s}, x_1, \ldots, x_t]
\]

\[
= T(x_1, \ldots, x_t)S[x_{t+1}, \ldots, x_{t+s}]
\]

\[
= (T \otimes S)[x_1, \ldots, x_{t+s}].
\]
Therefore
\( \tau.(S \otimes T) = T \otimes S. \)

Therefore
\[
A(S \otimes T) = \frac{1}{(s + t)!} \sum_{\sigma \in S_{s+t}} \epsilon^{\sigma} \sigma \tau.(T \otimes S)
\]
\[
= \epsilon^{\tau} \frac{1}{(s + t)!} \sum_{\sigma \in S_{s+t}} \epsilon^{\sigma}.(T \otimes S)
\]
\[
= (-1)^{st} A(T \otimes S),
\]
as required.  \( \blacksquare \)

### 9.2 Wedge Product

**Definition.** If \( S \in T^s M \) and \( T \in T^t M \), we define their wedge product (also called exterior product) by

\[
S \wedge T = \frac{1}{s!t!} \sum_{\sigma \in S_{s+t}} \epsilon^{\sigma}(S \otimes T) = \frac{(s + t)!}{s!t!} A(S \otimes T).
\]

**Example:** Let \( S, T \in M^* \) have components \( \alpha_i, \beta_i \) wrt \( u_i \). Then

\[
S \wedge T = S \otimes T - T \otimes S.
\]

Therefore

\[
S \wedge T[x, y] = S(x)T(y) - T(x)S(y),
\]
and \( S \wedge T \) has components

\[
\gamma_{ij} = S \wedge T[u_i, u_j] = S(u_i)T(u_j) - T(u_i)S(u_j) = \alpha_i \beta_j - \beta_i \alpha_j.
\]

**Theorem 9.4.** The wedge product has the following properties:

1. \( (R + S) \wedge T = R \wedge T + S \wedge T, \)
2. \( R \wedge (S + T) = R \wedge S + R \wedge T, \)
3. \( (\lambda R) \wedge S = \lambda (R \wedge S) = R \wedge (\lambda S), \)
4. \( R \wedge (S \wedge T) = (R \wedge S) \wedge T, \)

9-5
5. $S \wedge T = (-1)^s T \wedge S$, 
6. $R_1 \wedge \cdots \wedge R_k = \frac{(r_1 + \cdots + r_k)!}{r_1! \cdots r_k!} \mathcal{A}(R_1 \otimes \cdots \otimes R_k)$.

(i), (ii) and (iii) imply bilinear; (iv) implies associative; (v) implies graded commutative.

**Proof** e.g. 4.

\[
(R \wedge S) \wedge T = \frac{(r+s)!}{(r+s)!} \mathcal{A} \left[ \frac{(r+s)!}{r!s!} (\mathcal{A}(R \otimes S) \otimes T) \right] = \frac{(r+s)!}{r!s!t!} \mathcal{A}(R \otimes S \otimes T) \varepsilon \sim R \wedge (S \wedge T).
\]

5. 

\[
S \wedge T = \frac{(s+t)!}{s!t!} \mathcal{A}(S \otimes T) = (-1)^s \frac{(t+s)!}{t!s!} \mathcal{A}(T \otimes S) = (-1)^s T \wedge S.
\]

6. By induction on $k$: true for $k = 1$, assume true for $k - 1$. Then:

\[
(R_1 \wedge \cdots \wedge R_{k-1}) \wedge R_k = \frac{(r_1 + \cdots + r_{k-1} + r_k)!}{(r_1 + \cdots + r_{k-1})! r_k!} \mathcal{A} \left[ \frac{r_1 + \cdots + r_{k-1}!}{r_1! \cdots r_{k-1}!} \mathcal{A}((R_1 \otimes \cdots \otimes R_{k-1}) \otimes R_k) \right] = \frac{(r_1 + \cdots + r_k)!}{r_1! \cdots r_k!} \mathcal{A}[R_1 \otimes \cdots \otimes R_k].
\]

**Note.** For each integer $r > 0$ we write 

$M^{(r)}$ for the space of all skew-symmetric tensors of type 

\[
M \times \cdots \times M \rightarrow K;
\]

$M^{(r)}$ for the space of all skew-symmetric tensors of type 

\[
M^* \times \cdots \times M^* \rightarrow K;
\]

$M^{(0)} = K = M^{(0)}$. 

9–6
If \( S \in M(s) \) or \( S \in M(s) \), we say that \( S \) has degree \( s \), and we have
\[
S \wedge T = (-1)^s T \wedge S \quad \text{if} \quad s = \deg S, \ t = \deg T.
\]

Thus

1. \( S \wedge T = T \wedge S \) if either \( s \) or \( T \) has even degree;
2. \( S \wedge T = -T \wedge S \) if both \( S \) and \( T \) have odd degree;
3. \( S \wedge S = 0 \) if \( S \) has odd degree, since \( S \wedge S = -S \wedge S \);
4. \( T_1 \wedge T_2 \wedge \cdots \wedge S \wedge \cdots \wedge T_k = 0 \) if \( S \) has odd degree;
5. If \( x_1, \ldots, x_r \in M \) and \( i_1, \ldots, i_r \) are selected from \( \{1, 2, \ldots, r\} \) then
\[
x_{i_1} \wedge \cdots \wedge x_{i_r} = \epsilon_{i_1 \ldots i_r} x_1 \wedge x_2 \wedge \cdots \wedge x_r,
\]
where
\[
\epsilon_{i_1 \ldots i_r} = \begin{cases} 
1 & \text{if } i_1, \ldots, i_r \text{ is an even permutation of } 1, \ldots, r; \\
-1 & \text{if } i_1, \ldots, i_r \text{ is an odd permutation of } 1, \ldots, r; \\
0 & \text{otherwise}
\end{cases} = \epsilon^{i_1 \ldots i_r}
\]
is called a permutation symbol;
6. If \( x_j = \alpha_j^i y_i, \ \alpha_j^i = A \in K^{r \times r} \) then
\[
x_1 \wedge \cdots \wedge x_r = (\alpha_1^i y_i) \wedge \cdots \wedge (\alpha_r^i y_i) \\
= \alpha_1^i \cdots \alpha_r^i y_{i_1} \wedge \cdots \wedge y_{i_r} \\
= \epsilon_{i_1 \ldots i_r} \alpha_1^{i_1} \cdots \alpha_r^{i_r} y_1 \wedge \cdots \wedge y_r \\
= (\det A) y_{i_1} \wedge \cdots \wedge y_{i_r}.
\]

**Theorem 9.5.** Let \( M \) be \( n \)-dimensional, with basis \( u_1, \ldots, u_n \). Then

(i) if \( r > n \) then \( M^{(r)} = \{0\} \),

(ii) if \( 1 \leq r \leq n \) then
\[
\dim M^{(r)} = \frac{n!}{r!(n-r)!},
\]
and \( \{u^{i_1} \wedge \cdots \wedge u^{i_r}\}_{i_1 < \cdots < i_r} \) is a basis for \( M^{(r)} \).

**Proof**
(i) If \( r > n \) and \( T \in M^{(r)} \) has components \( \alpha_{i_1 \ldots i_r} \) then the indices \( i_1, \ldots, i_r \) cannot be distinct. Therefore \( T = 0 \).

(ii) We have

\[
\langle u^i, u_j \rangle = \begin{cases} 
1 & i = j; \\
0 & i \neq j
\end{cases} = \delta_j^i.
\]

More generally:

\[
u^{i_1} \wedge \cdots \wedge u^{i_r}[u_{j_1}, \ldots, u_{j_r}] = \sum_{\sigma \in S_r} \epsilon^\sigma u^{i_1} \otimes \cdots \otimes u^{i_r}[u_{j_{\sigma(1)}}, \ldots, u_{j_{\sigma(r)}}] = \sum_{\sigma \in S_r} \epsilon^\sigma \delta_{j_{\sigma(1)}}^{i_1} \cdots \delta_{j_{\sigma(r)}}^{i_r}
\]

\[
= \begin{cases} 
1 & \text{if } i_1, \ldots, i_r \text{ are distinct and an even permutation of } j_1, \ldots, j_r; \\
-1 & \text{if } i_1, \ldots, i_r \text{ are distinct and an odd permutation of } j_1, \ldots, j_r; \\
0 & \text{otherwise}
\end{cases}
\]

\[= \delta_{i_1 \ldots i_r}^{j_1 \ldots j_r} \text{ (general Kronecker delta).}\]

It follows that if \( 1 \leq r \leq n \), and if \( T \in M^{(r)} \) has components \( \alpha_{i_1 \ldots i_r} \) then the tensor:

\[
(*) \sum_{i_1 < \cdots < i_r} \alpha_{i_1 \ldots i_r} u^{i_1} \wedge \cdots \wedge u^{i_r}
\]

has components

\[
\sum_{i_1 < \cdots < i_r} \alpha_{i_1 \ldots i_r} u^{i_1} \wedge \cdots \wedge u^{i_r}[u_{j_1}, \ldots, u_{j_r}] = \sum_{i_1 < \cdots < i_r} \alpha_{i_1 \ldots i_r} \delta_{i_1 \ldots i_r}^{j_1 \ldots j_r}
\]

provided \( j_1 < \cdots < j_r \). Therefore (*) has the same components as \( T \). Therefore

\[
(**) \{u^{i_1} \wedge \cdots \wedge u^{i_r}\}_{i_1 < \cdots < i_r}
\]

generate \( M^{(r)} \). Also

\[(*) = 0 \Rightarrow \alpha_{j_1 \ldots j_r} = 0.
\]

Therefore (**) are linearly independent. Therefore (**) form a basis for \( M^{(r)} \). □
Chapter 10

Classification of Linear Operators

10.1 Hamilton-Cayley and Primary Decomposition

Let $M \xrightarrow{T} M$ be a linear operator on a vector space $M$ over a field $K$. Let $K[X]$ be the ring of polynomials in $X$ with coefficients in $K$. If $p = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \cdots + \alpha_r X^r$, write

$$p(T) = \alpha_0 1 + \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_r T^r \in \mathcal{L}(M).$$

**Theorem 10.1 (Hamilton-Cayley).** Let $T \in \mathcal{L}(M)$ have characteristic polynomial $p$, and let $\dim M < \infty$. Then $p(T) = 0$.

**Proof** Let $T$ have matrix $\alpha^i_j$ wrt basis $u_i$. Put $P = (p^i_j)$, where $p^i_j = \alpha^i_j - X \delta^i_j$. Then $P$ is an $n \times n$ matrix of polynomials, and $\det P = p$ is the characteristic polynomial.

Let $q^i_j = (-1)^{i+j}$ times the determinant of the matrix got from $P$ by removing the $i^{th}$ column and $j^{th}$ row. Then $Q = (q^i_j)$ is also an $n \times n$ matrix of polynomials, and

$$PQ = (\det P) I,$$

i.e.

$$p^i_k q^k_j = p^i_j \delta^i_j$$

Therefore

$$p(T) u_j = p(T) \delta^i_j u_i = p^i_k(T) q^k_j(T) u_i$$

$$= q^k_j(T) [\alpha^i_k - T \delta^i_k] u_i = q^k_j(T) [\alpha^i_k u_i - Tu_k] = 0$$

10-1
Therefore $p(T) = 0$, as required.

**Example:** \[
\begin{pmatrix}
\alpha & \beta \\
\gamma & \delta
\end{pmatrix}: \mathbb{R}^2 \to \mathbb{R}^2.
\]

\[
p = \begin{vmatrix}
\alpha - X & \beta \\
\gamma & \delta - X
\end{vmatrix} = X^2 - (\alpha + \delta)X + \alpha\delta - \beta\gamma.
\]

Therefore
\[
p \begin{pmatrix}
\alpha & \beta \\
\gamma & \delta
\end{pmatrix} = \left(\begin{pmatrix}
\alpha & \beta \\
\gamma & \delta
\end{pmatrix} \right)^2 - (\alpha + \delta) \begin{pmatrix}
\alpha & \beta \\
\gamma & \delta
\end{pmatrix} + (\alpha\delta - \beta\gamma) \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
= \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix}.
\]

**Theorem 10.2 (Primary Decomposition Theorem).** Let $T \in \mathcal{L}(M)$, and let
\[
(T - \lambda_1)^{r_1} \ldots (T - \lambda_k)^{r_k} = 0,
\]
where $\lambda_1, \ldots, \lambda_k$ are distinct scalars, and $r_1, \ldots, r_k$ are positive integers. Then
\[
M = M_1 \oplus \cdots \oplus M_k,
\]
where $M_i = \ker(T - \lambda_i)^{r_i}$ for $i = 1, \ldots, k$.

**Proof** Let
\[
f = (X - \lambda_1)^{r_1} \ldots (X - \lambda_k)^{r_k},
\]
\[
g_i = (X - \lambda_i)^{r_i},
\]
\[
f = g_i h_i
\]
(say), so $f(T) = 0$ and $M_i = \ker g_i(T)$.

Now $h_1, \ldots, h_k$ have hcf 1. Therefore there exist
\[
\Theta_1, \ldots, \Theta_k \in K[X]
\]
such that
\[
\Theta_1 h_1 + \cdots + \Theta_k h_k = 1.
\]

Put $P_i = \Theta_i(T)h_i(T)$. Then

(i) $P_1 + \cdots + P_k = 1$. Also:
(ii) for each $x \in M$,
\begin{equation}
g_i(T)P_i x = g_i(T)\Theta_i(T)h_i(T)x = f(T)\Theta_i(T)x = 0.
\end{equation}

Therefore $P_i x \in M_i$,

(iii) if $x_i \in M_i$ and $j \neq i$ then
\begin{equation}
P_j x_i = \Theta_j(T)h_j(T)x_i = 0,
\end{equation}
since $g_i$ is a factor of $h_j$, and $g_i(T)x_i = 0$.

Thus

1. for each $x \in M$ we have
\begin{equation}
x = 1x = P_1 x + \cdots + P_k x, \quad P_i x \in M_i;
\end{equation}

2. if $x = x_1 + \cdots + x_k$, with $x_i \in M_i$, then (for example, see Figure 10.1)
\begin{equation}
x_i = (P_1 + \cdots + P_k)x_i = P_i x_i = P_i (x_1 + \cdots + x_k) = P_i x.
\end{equation}

Therefore $x_i$ is uniquely determined by $x$. Therefore
\begin{equation}
M = M_1 \oplus \cdots \oplus M_k,
\end{equation}
Note. Each subspace $M_i$ is invariant under $T$, since
\begin{align*}
x \in M_1 & \Rightarrow g_i(T)x = 0 \\
& \Rightarrow Tg_i(T)x = 0 \\
& \Rightarrow g_i(T)Tx = 0 \\
& \Rightarrow Tx \in M_i.
\end{align*}

Therefore, if we take bases for $M_1, \ldots, M_k$, and put them together to get a basis for $M$, then wrt this basis $T$ has matrix
\[
\begin{pmatrix}
A_1 & 0 \\
A_2 & \ddots \\
0 & A_k
\end{pmatrix},
\]
where $A_i$ is the matrix of $T_{M_i}$, the restriction of $T$ to $M_i$.

Note also that
\[
(T_{M_i} - \lambda_i 1)^{r_i} = 0.
\]
Example: Let $M \xrightarrow{T} M$ and $T^2 = T$ ($T$ a projection operator). Then

$$T(T - 1) = 0.$$ 

Therefore, by primary decomposition,

$$M = \ker(T - 1) \oplus \ker T = 1\text{-eigenspace} \oplus 0\text{-eigenspace}.$$ 

If $T$ has rank $r$, and $u_1, \ldots, u_r$ is basis of 1-eigenspace; $u_{r+1}, \ldots, u_n$ is basis of 0-eigenspace then, wrt $u_1, \ldots, u_n$ $T$ has matrix

$$
\begin{pmatrix}
1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
0 \\
\vdots \\
0 & \ldots & 0
\end{pmatrix}
= \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.
$$

10.2 Diagonalisable Operators

Let $M \xrightarrow{T} M; \dim M < \infty$. Then:

$(T - \lambda_1) \ldots (T - \lambda_k) = 0$ ($\lambda_1, \ldots, \lambda_k$ distinct)

$\Rightarrow M = \ker(T - \lambda_1) \oplus \cdots \oplus (T - \lambda_k) 1$ by Primary Decomposition

$\Rightarrow M = (\lambda_1\text{-eigenspace}) \oplus \cdots \oplus (\lambda_k\text{-eigenspace})$

$\Rightarrow T$ has a diagonal matrix wrt some basis of $M$

$\Rightarrow M$ has basis consisting of eigenvectors of $T$; and $(T - \lambda_1) \ldots (T - \lambda_k)u = 0$

for each eigenvector $u$, where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of $T$

$\Rightarrow (T - \lambda_1) \ldots (T - \lambda_k) = 0$ ($\lambda_1, \ldots, \lambda_k$ distinct).

Definition. A linear operator $T$ with any one (and hence all) of the above properties is called diagonalisable.

Example: The operator

$$
\begin{pmatrix}
1 & 0 \\
1 & 1
\end{pmatrix}: \mathbb{R}^2 \to \mathbb{R}^2
$$

10-5
is not diagonalisable.

**Proof of This**

The characteristic polynomial is

\[
\begin{vmatrix}
1 - X & 0 \\
1 & 1 - X
\end{vmatrix} = (1 - X)^2.
\]

Therefore 1 is the only eigenvalue.

Also

\[
(\alpha, \beta) \in 1\text{-eigenspace} \iff \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}
\]

\[
\iff \begin{pmatrix} \alpha \\ \alpha + \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}
\]

\[
\iff \alpha = 0.
\]

Therefore 1-eigenspace = \{\beta(0,1) : \beta \in \mathbb{R}\} is 1-dimensional. Therefore \(\mathbb{R}^2\) does not have a basis of eigenvectors of \(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\). \(\triangleright\)

**Theorem 10.3.** Let \(S, T \in \mathcal{L}(M)\) be diagonalisable (\(\dim M < \infty\)). Then there exists a basis wrt which both \(S\) and \(T\) have diagonal matrices (\(S, T\) simultaneously diagonalisable) iff \(ST = TS\) (\(S, T\) commute).

**Proof**

(i) Let \(M\) have a basis wrt which \(S\) has diagonal matrix

\[
A = \begin{pmatrix} 
\lambda_1 & & \\
& \ddots & \\
& & \lambda_n
\end{pmatrix}
\]

and \(T\) has diagonal matrix

\[
B = \begin{pmatrix} 
\mu_1 & & \\
& \ddots & \\
& & \mu_n
\end{pmatrix}
\]

Then \(AB = BA\). Therefore \(ST = TS\).

(ii) Let \(ST = TS\). Since \(S\) is diagonalisable we have:

\[
M = M_1 \oplus \cdots \oplus M_i \oplus \cdots \oplus M_k,
\]

10–6
distinct sum of eigenspaces of $S$. Since $S$ and $T$ commute, $T$ leaves each $M_i$ invariant:

$$T_{M_i} : M_i \to M_i.$$ 

Since $T$ is diagonalisable we have:

$$(T - \mu_1 1) \ldots (T - \mu_l 1) = 0,$$

distinct $\mu_1, \ldots, \mu_l$. Therefore

$$(T_{M_i} - \mu_1 1_{M_i}) \ldots (T_{M_i} - \mu_l 1_{M_i}) = 0.$$ 

Therefore $T_{M_i}$ is diagonalisable. Therefore $M_i$ has a basis of eigenvectors of $T$. Therefore $M$ has a basis of eigenvectors of $S$, and of $T$. 

\section{10.3 Conjugacy Classes}

\textit{Problem}: Given two linear operators

$$S, T : M \to M,$$

to determine whether they are equivalent up to an isomorphism of $M$, i.e. is there a linear isomorphism $M \xrightarrow{R} M$ so that the diagram

$$\begin{array}{ccc}
M & \xrightarrow{S} & M \\
R \downarrow & & \downarrow R \\
M & \xrightarrow{T} & M
\end{array}$$

commutes, i.e.

$$RS = TR,$$

i.e.

$$RSR^{-1} = T.$$ 

\textbf{Definition}. $S$ is \textit{conjugate to} $T$ if there exists a linear isomorphism $R$ such that

$$RSR^{-1} = T.$$ 

Conjugacy is an equivalence relation on $\mathcal{L}(M)$; the equivalence classes are called \textit{conjugacy classes}. 

10–7
If $T \in \mathcal{L}(M)$ has matrix $A \in K^{n \times n}$ wrt some basis of $M$ then the set of all matrices which can represent $T$ is:

$$\{PAP^{-1} : P \in K^{n \times n} \text{ is invertible}\},$$

which is a conjugacy class in $K^{n \times n}$.

Conversely, the set of all linear operators on $M$ which can be represented by $A$ is:

$$\{RTR^{-1} : R \text{ is a linear isomorphism of } M\},$$

which is a conjugacy class in $\mathcal{L}(M)$.

Hence we have a bijective map from the set of conjugacy classes in $\mathcal{L}(M)$ to the set of conjugacy classes in $K^{n \times n}$ (see Figure 10.2).

The problem of determining which conjugacy class $T$ belongs to is thus equivalent to determining which conjugacy class $A$ belongs to.

A simple way of distinguishing conjugacy classes is to use properties such as: rank, trace, determinant, eigenvalues, characteristic polynomial, which are the same for all elements of a conjugacy class.

Examples:
1. Let
\[
J = \begin{pmatrix}
\lambda & 0 & 0 & 0 \\
1 & \lambda & 0 & 0 \\
0 & 1 & \lambda & 0 \\
0 & 0 & 1 & \lambda
\end{pmatrix}, \quad \text{a Jordan } \lambda\text{-block of size } 4.
\]
(4 × 4, λ on diagonal, 1 just below diagonal, zero elsewhere).

\[
J - \lambda I = \begin{pmatrix}
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}.
\]

Therefore
\[
(J - \lambda I)e_1 = e_2, \quad (J - \lambda I)e_2 = e_3, \quad (J - \lambda I)e_3 = e_4, \quad (J - \lambda I)e_4 = 0.
\]

Thus
\[
K^4 \quad e_1 \quad e_2 \quad e_3 \quad e_4
\]
\[
\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow
\]
\[
\text{im}(J - \lambda I) \quad e_2 \quad e_3 \quad e_4 \quad 0
\]
\[
\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow
\]
\[
\text{im}(J - \lambda I)^2 \quad e_3 \quad e_4 \quad 0
\]
\[
\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow
\]
\[
\text{im}(J - \lambda I)^3 \quad e_4 \quad 0
\]
\[
\downarrow \quad \downarrow
\]
\[
\{0\} \quad 0
\]

Thus
\[
\text{im}(J - \lambda I) \text{ has basis } e_2, e_3, e_4, \quad \text{rank}(J - \lambda I) = 3,
\]
\[
\text{im}(J - \lambda I)^2 \text{ has basis } e_3, e_4, \quad \text{rank}(J - \lambda I)^2 = 2,
\]
\[
\text{im}(J - \lambda I)^3 \text{ has basis } e_4, \quad \text{rank}(J - \lambda I)^3 = 1,
\]
\[
(J - \lambda I)^4 = 0.
\]

\[
\text{char } J = \begin{vmatrix}
\lambda - X & 0 & 0 & 0 \\
1 & \lambda - X & 0 & 0 \\
0 & 1 & \lambda - X & 0 \\
0 & 0 & 1 & \lambda - X
\end{vmatrix} = (\lambda - X)^4.
\]

Therefore \(\lambda\) is the only eigenvalue of \(J\), and the \(\lambda\)-eigenspace = \(\ker(J - \lambda I)\) has basis \(e_4\).
2. Let

\[ J = \begin{pmatrix}
\lambda & 0 & 0 \\
1 & \lambda & 0 \\
0 & 1 & \lambda \\
0 & 0 & \lambda \\
0 & 0 & 1 \\
\lambda & 0 & 0
\end{pmatrix} \]

(Jordan $\lambda$-blocks on diagonal: $3 \times 3, 2 \times 2, 2 \times 2$).

\[ K^7 \quad e_1 \quad e_2 e_4 e_6 \quad e_3 e_5 e_7 \]
\[ \downarrow \quad \text{s}_1 \quad \text{s}_2 \quad \text{s}_3 \]
\[ \text{im}(J - \lambda I) \quad e_2 \quad e_3 e_5 e_7 \]
\[ \downarrow \quad \text{s}_1 \quad \text{s}_2 \]
\[ \text{im}(J - \lambda I)^2 \quad e_3 \]
\[ \downarrow \quad \text{s}_1 \]
\[ \{0\} \]

where $s_1$, $s_2$ and $s_3$ are the dimensions of the kernel of $(J - \lambda I)$ restricted to im$(J - \lambda I)^2$, im$(J - \lambda I)$ and $K^7$ respectively.

\[ \text{char } J = (\lambda - X)^7. \lambda \text{ is the only eigenvector; dim } \lambda-\text{eigenspace} = 3 = \text{number of Jordan blocks}. \]

\[ (J - \lambda)e_1 = e_2 \]
\[ (J - \lambda)e_2 = e_3 \]
\[ (J - \lambda)e_3 = 0 \quad \text{eigenvector} \]

\[ (J - \lambda)e_4 = e_5 \]
\[ (J - \lambda)e_5 = 0 \quad \text{eigenvector} \]

\[ (J - \lambda)e_6 = e_7 \]
\[ (J - \lambda)e_7 = 0 \quad \text{eigenvector}. \]
10.4 Jordan Forms

**Definition.** A square matrix $J \in K^{n \times n}$ is called a *Jordan matrix* if it is of the form

$$J = \begin{pmatrix}
J_1 & & \\
& J_2 & \\
& & \ddots \\
& & & J_l
\end{pmatrix},$$

where each $J_i$ is a Jordan block.

**Example:**

$$J = \begin{pmatrix}
\lambda & 0 & 0 \\
1 & \lambda & 0 \\
0 & 1 & \lambda \\
\lambda & 0 & \\
1 & \lambda & \\
\mu & 0 & \\
1 & \mu & \\
\mu & 0 & \\
1 & \mu & \\
\end{pmatrix},$$

(where $\lambda \neq \mu$ (say)) is a $9 \times 9$ Jordan matrix.

Note that

(i) $\text{char } J = (\lambda - X)^5(\mu - X)^4$; eigenvalue $\lambda$, with *algebraic multiplicity* 5; eigenvalue $\mu$, with algebraic multiplicity 4.

(ii) $\text{dimension of } \lambda\text{-eigenspace} = \text{number of } \lambda\text{-blocks} = \text{geometric multiplicity of eigenvalue } \lambda = 2$; geometric multiplicity of eigenvalue $\mu = 2$. 


(iii) \[
J - \lambda I = \begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
0 & 0 \\
1 & 0 \\
\mu - \lambda & 0 \\
1 & \mu - \lambda \\
\end{pmatrix}
\quad \leftarrow \text{non-sing.}
\]

\[
(J - \lambda I)^2 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
1 & 0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
0 & 0 \\
0 & 0 \\
(\mu - \lambda)^2 & 0 \\
1 & (\mu - \lambda)^2 \\
\end{pmatrix}
\quad \leftarrow \text{non-sing.}
\]

\[
(J - \lambda)^3 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
0 & 0 \\
0 & 0 \\
\text{N.S.} & \text{N.S.} \\
\end{pmatrix}
\]

Therefore
\[
\text{rank}(J - \lambda I) = 2 + 1 + 4, \\
\text{rank}(J - \lambda I)^2 = 1 + 0 + 4, \\
\text{rank}(J - \lambda I)^3 = 0 + 0 + 4.
\]
More generally, if $J$ is a Jordan $n \times n$ matrix, with

- $b_1 \lambda$-blocks of size 1,
- $b_2 \lambda$-blocks of size 2,
- \[ \vdots \]
- $b_k \lambda$-blocks of size $k$,

and if $\lambda$ has algebraic multiplicity $m$ then

$$b_1 + 2b_2 + 3b_3 + \cdots = m,$$

rank$(J - \lambda I)$ has rank:

$$0b_1 + 1b_2 + 2b_3 + 3b_4 + \cdots + (n - m),$$

rank$(J - \lambda I)^2$ has rank:

$$0b_1 + 0b_2 + 1b_3 + 2b_4 + \cdots + (n - m),$$

rank$(J - \lambda I)^3$ has rank:

$$0b_1 + 0b_2 + 0b_3 + 1b_4 + 2b_5 + \cdots + (n - m),$$

and so on. Hence the number $b_k$ of $\lambda$-blocks of size $k$ in $J$ is uniquely determined by the conjugacy class of $J$.

**Theorem 10.4.** Let $T \in \mathcal{L}(M)$ be a linear operator on a finite dimensional vector space over a field $K$ which is algebraically closed. Then $T$ can be represented by a Jordan matrix $J$. The matrix $J$, which by the preceding is uniquely determined, apart from the arrangement of the blocks on the diagonal, is called the **Jordan form** of $T$.

**Proof** Since $K$ is algebraically closed, the characteristic polynomial is a product of linear factors; so, by Hamilton-Cayley we have

$$(T - \lambda_1)^{r_1} \cdots (T - \lambda_k)^{r_k} = 0$$

(say), with $\lambda_1, \ldots, \lambda_k$ distinct factors.

By primary decomposition:

$$M = M_1 \oplus \cdots \oplus M_k,$$
where \( M_i = \ker(T - \lambda_i 1)^r \). We will show that \( M_i \) has a basis wrt which \( T_{M_i} \) has a Jordan matrix with \( \lambda_i \) on the diagonal.

Put \( S = T_{M_i} - \lambda_i 1_{M_i} \). Then
\[
M_i \xrightarrow{S} M_i
\]
and \( S^r = 0 \), i.e. \( S \) is a nilpotent operator. Suppose \( S^r = 0 \) but \( S^{r-1} \neq 0 \), and consider:
\[
\begin{array}{c}
M_i \\
\downarrow \text{im } S \\
\downarrow \text{im } S^2 \\
\vdots \\
\downarrow \text{im } S^{r-3} \\
\downarrow \text{im } S^{r-2} \\
\downarrow \text{im } S^{r-1} \\
\{0\}
\end{array}
\]

Choose a basis \( z_1, \ldots, z_{s_1} \) for \( \text{im } S^{r-1} \). Choose \( y_1, \ldots, y_{s_1} \in \text{im } S^{r-2} \) such that \( Sy_j = z_j \). Extend to a basis \( z_1, \ldots, z_{s_1}, \ldots, z_{s_2} \) for the kernel of \( \text{im } S^{r-2} \to \text{im } S^{r-1} \). Thus \( y_1, \ldots, y_{s_1}, z_1, \ldots, z_{s_2} \) is a basis for \( \text{im } S^{r-2} \).

Now repeat the construction: choose \( x_1, \ldots, x_{s_1}, y_1, \ldots, y_{s_2} \in \text{im } S^{r-3} \) such that \( Sx_j = y_j \), \( Sy_i = z_i \). Extend to a basis \( z_1, \ldots, z_{s_2}, \ldots, z_{s_3} \) for the kernel of \( \text{im } S^{r-3} \to \text{im } S^{r-2} \). Thus \( x_1, \ldots, x_{s_1}, y_1, \ldots, y_{s_2}, z_1, \ldots, z_{s_3} \) is a basis for \( \text{im } S^{r-3} \).

Continue in this way until we get a basis
\[
a_1, \ldots, a_{s_1}, b_1, \ldots, b_{s_2}, \ldots, y_1, \ldots, y_{s_{r-1}}, z_1, \ldots, z_{s_r},
\]
(say), for \( M_i \), with
\[
Sa_j = b_j, Sb_j = c_j, \ldots, Sy_j = z_j, Sz_j = 0.
\]

Now write the basis elements in order, by going down each column of (*) in turn, starting at the left most column (and leaving out any column whose elements have already been written down).
Relative to this basis for \( M_i \) the matrix of \( S \) is a Jordan matrix with zeros on the diagonal. Therefore the matrix of \( T_{M_i} = S + \lambda_i 1_{M_i} \) is a Jordan matrix with \( \lambda_i \) on the diagonal.

Putting together these bases for \( M_1, \ldots, M_k \) we get a basis for \( M \) wrt which \( T \) has a Jordan matrix, as required.

*Example:* To find a Jordan matrix \( J \) conjugate to the matrix

\[
A = \begin{pmatrix}
  5 & 4 & 3 \\
  -1 & 0 & -3 \\
  1 & -2 & 1
\end{pmatrix}.
\]

The characteristic polynomial is

\[
p = \begin{vmatrix}
  5 - X & 4 & 3 \\
  -1 & -X & -3 \\
  1 & -2 & 1 - X
\end{vmatrix}
= (5 - X)[-X(1 - X) - 6] - 4[-(1 - X) + 3] + 3[2 + X]
= (5 - X)[X^2 - X - 6] - 4[X + 2] + 3[2 + X]
= 5X^2 - 5X - 30 - X^3 + X^2 + 6X - 4X - 8 + 3X + 6
= -X^3 + 6X^2 - 32
= (X + 2)(-X^2 + 8X - 16)
= -(X + 2)(X - 4)^2.
\]

Therefore, by Hamilton-Cayley the operator \( \mathbb{R}^3 \xrightarrow{A} \mathbb{R}^3 \) satisfies:

\[(A + 2I)(A - 4I)^2 = 0.\]

Therefore, by primary decomposition:

\[\mathbb{R}^3 = \ker(A + 2I) \oplus \ker(A - 4I)^2.\]

Now

\[
\ker(A + 2I) = \ker \begin{pmatrix}
  7 & 4 & 3 \\
  -1 & 2 & -3 \\
  1 & -2 & 3
\end{pmatrix}
= \ker \begin{pmatrix}
  7 & 4 & 3 \\
  0 & 18 & -18 \\
  0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
  \text{row 3 + row 2} \\
  \text{7 row 2 + row 1}
\end{pmatrix}
\]

10–15
Therefore

\[(\alpha, \beta, \gamma) \in \ker(A + 2I) \iff 7\alpha + 4\beta + 3\gamma = 0\]
\[\beta - \gamma = 0\]
\[\iff 7\alpha + 7\gamma = 0\]
\[\beta - \gamma = 0\]
\[\iff (\alpha, \beta, \gamma) = (\alpha, -\alpha, -\alpha) = \alpha(1, -1, -1).\]

Therefore \(\ker(A + 2I)\) has basis \(u_1 = (1, -1, -1)\).

\[
\ker(A - 4I)^2 = \ker \begin{pmatrix}
1 & 4 & 3 \\
-1 & -4 & -3 \\
1 & -2 & -3
\end{pmatrix}^2
= \ker \begin{pmatrix}
0 & -18 & -18 \\
0 & 18 & 18 \\
0 & 18 & 18
\end{pmatrix}
= \ker \begin{pmatrix}
0 & 1 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

Therefore

\[(\alpha, \beta, \gamma) \in \ker(A - 4I)^2 \iff \beta + \gamma = 0\]
\[\iff (\alpha, \beta, \gamma) = (\alpha, \beta, -\beta) = \alpha(1, 0, 0) + \beta(0, 1, -1).\]

Therefore \((1, 0, 0), (0, 1, -1)\) is a basis for \(\ker(A - 4I)^2\).

Put \(u_2 = (1, 0, 0), \quad u_3 = (A - 4I)u_2 = (1, -1, 1)\).

So \(u_1, u_2, u_3\) is a basis for \(\mathbb{R}^3\) such that

\[(A + 2I)u_1 = 0,\]
\[(A - 4I)u_2 = u_3,\]
\[(A - 4I)u_3 = 0,\]

i.e.

\[Au_1 = -2u_1,\]
\[Au_2 = 4u_2 + u_3,\]
\[Au_3 = 4u_3.\]
Therefore wrt basis

\[ u_1 = (1, -1, -1) = e_1 - e_2 - e_3, \]
\[ u_2 = (1, 0, 0) = e_1, \]
\[ u_3 = (1, -1, 1) = e_1 - e_2 + e_3 \]

the operator \( A \) has matrix

\[
J = \begin{pmatrix}
-2 & 0 & 0 \\
0 & 4 & 0 \\
0 & 1 & 4
\end{pmatrix}
\]

Let \( P \) be the transition matrix from \((e_1, e_2, e_3)\) to \((u_1, u_2, u_3)\).

\[
P = \begin{pmatrix}
1 & 1 & 1 \\
-1 & 0 & -1 \\
-1 & 0 & 1
\end{pmatrix}^{-1} = \frac{1}{2} \begin{pmatrix}
0 & -1 & -1 \\
2 & 2 & 0 \\
0 & -1 & 1
\end{pmatrix} = \begin{pmatrix}
0 & -\frac{1}{2} & -\frac{1}{2} \\
1 & 1 & 0 \\
0 & -\frac{1}{2} & -\frac{1}{2}
\end{pmatrix}.
\]

Therefore

\[
PAP^{-1} = \begin{pmatrix}
0 & -\frac{1}{2} & -\frac{1}{2} \\
1 & 1 & 0 \\
0 & -\frac{1}{2} & \frac{1}{2}
\end{pmatrix} \begin{pmatrix}
5 & 4 & 3 \\
-1 & 0 & -3 \\
1 & -2 & 1
\end{pmatrix} \begin{pmatrix}
1 & 1 & 1 \\
-1 & 0 & -1 \\
-1 & 0 & 1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
4 & 4 & 0 \\
1 & -1 & 2 \\
-1 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 1 & 1 \\
-1 & 0 & -1 \\
-1 & 0 & 1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
-2 & 0 & 0 \\
0 & 4 & 0 \\
0 & 1 & 4
\end{pmatrix} = J,
\]

as required.

**Note.** (i)

\[
u_1 = e_1 - e_2 - e_3 \quad \Rightarrow \quad e_1 = u_2 \\
u_2 = e_1 \quad \Rightarrow \quad u_1 + u_3 = 2e_1 - 2e_2 \\
u_3 = e_1 - e_2 + e_3 \quad \Rightarrow \quad u_1 - u_3 = -2e_3
\]

\[
e_1 = u_2 \quad \Rightarrow \quad e_2 = -\frac{1}{2}u_1 + u_2 - \frac{1}{2}u_3 \\
e_3 = -\frac{1}{2}u_1 + \frac{1}{2}u_3 \quad \Rightarrow \quad P = \begin{pmatrix}
0 & -\frac{1}{2} & -\frac{1}{2} \\
1 & 1 & 0 \\
0 & -\frac{1}{2} & \frac{1}{2}
\end{pmatrix}.
\]

10–17
(ii)
\[
\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
-1 & 0 & -1 & 0 & 1 & 0 \\
-1 & 0 & 1 & 0 & 0 & 1 \\
\end{pmatrix}
\sim
\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 \\
0 & 1 & 2 & 1 & 1 & 0 \\
\end{pmatrix}
\sim
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 \\
0 & 0 & 2 & 0 & -1 \\
\end{pmatrix}
\sim
\begin{pmatrix}
2 & 2 & 0 & 2 & 1 & -1 \\
0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 2 & 0 & -1 & 1 \\
\end{pmatrix}
\sim
\begin{pmatrix}
2 & 0 & 0 & 0 & -1 & -1 \\
0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 2 & 0 & -1 & 1 \\
\end{pmatrix}
\]

Example: To find the Jordan form of
\[
A = \begin{pmatrix}
1 & 1 & 3 \\
5 & 2 & 6 \\
-2 & -1 & -3 \\
\end{pmatrix}
\]

\[
\text{char} A = \begin{vmatrix}
1-X & 1 & 3 \\
5 & 2-X & 6 \\
-2 & -1 & -3-X \\
\end{vmatrix}
= (1-X)\{(2-X)(-3-X)+6\} - 5(-3-X) + 3[5+2(2-X)]
= (1-X)[X^2+X] - 5X - 3 - 2X
= X^2 + X - X^3 - X^2 + 5X + 3 - 6X - 3
= -X^3.
\]

Therefore operator \(\mathbb{R}^3 \xrightarrow{A} \mathbb{R}^3\) satisfies
\[
A^3 = 0.
\]

Now
\[
A^2 = \begin{pmatrix}
1 & 1 & 3 \\
5 & 2 & 6 \\
-2 & -1 & -3 \\
\end{pmatrix}
\begin{pmatrix}
1 & 1 & 3 \\
5 & 2 & 6 \\
-2 & -1 & -3 \\
\end{pmatrix}
= \begin{pmatrix}
0 & 0 & 0 \\
3 & 3 & 9 \\
-1 & -1 & -3 \\
\end{pmatrix}
\]

So put
\[
\begin{align*}
u_1 &= e_1 = (1, 0, 0) = e_1, \\
u_2 &= Ae_1 = (1, 5, -2) = e_1 + 5e_2 - 2e_3, \\
u_3 &= A^2e_1 = (0, 3, -1) = 3e_2 - e_3.
\end{align*}
\]

So wrt new basis \(u_1, u_2, u_3\) operator \(A\) has matrix
\[
J = \begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
\end{pmatrix}
\]

10–18
Also
\[ u_2 - 2u_3 = e_1 - e_2. \]
So
\[
\begin{align*}
e_2 &= u_1 - u_2 + 2u_3, \\
e_3 &= 3e_2 - u_3 = 3u_1 - 3u_2 + 5u_3.
\end{align*}
\]
Thus
\[
\begin{align*}
e_1 &= u_1, \\
e_2 &= u_1 - u_2 + 2u_3, \\
e_3 &= 3u_1 - 3u_2 + 5u_3.
\end{align*}
\]
Therefore \( PAP^{-1} = J \), where
\[
P = \begin{pmatrix} 1 & 1 & 3 \\ 0 & -1 & -3 \\ 0 & 2 & 5 \end{pmatrix}.
\]
Check:
\[
PA = JP \iff \begin{pmatrix} 1 & 1 & 3 \\ 0 & -1 & -3 \\ 0 & 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 \\ 0 & -1 & -3 \\ 0 & 2 & 5 \end{pmatrix} \iff \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 3 \\ 0 & -1 & -3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 3 \\ 0 & -1 & -3 \end{pmatrix}.
\]

### 10.5 Determinants

Let \( M \) be a vector space of finite dimension \( n \), and \( M \overset{T}{\rightarrow} N \) be a linear operator. The \textit{pull-back} (or \textit{transpose}) of \( T \) is the operator \( M \overset{T^*}{\leftarrow} N^* \) defined by
\[
\langle T^* f, x \rangle = \langle f, Tx \rangle.
\]
The transpose of \( T^* \) is \( T \) itself (by duality) written \( T_* \). So:
\[
\begin{align*}
M &\overset{T_*}{\rightarrow} N \ (T_* = T), \\
M^* &\overset{T^*_*}{\leftarrow} N^*, \\
\langle T^* f, x \rangle &= \langle f, T_* x \rangle.
\end{align*}
\]
If \( M = N \) and \( T \) has matrix \( A = (\alpha^j_i) \) wrt basis \( u_i \) then
\[
\langle T^* w^j, u_i \rangle = \langle u^j, T_* u_i \rangle = \langle w^j, \alpha^k_i u_k \rangle = \alpha^j_i.
\]
Therefore \( T^* \) has matrix \( A^t \) wrt basis \( u^i \).

More generally we have the pull-back
\[
M^{(r)} \xrightarrow{T^*} N^{(r)},
\]
and push-forward
\[
M_{(r)} \xrightarrow{T_*} N_{(r)}
\]
defined by
\[
(T^* S)(x_1, \ldots, x_r) = S(T_* x_1, \ldots, T_* x_r),
\]
\[
(T_* S)(f^1, \ldots, f^r) = S(T^* f^1, \ldots, T^* f^r).
\]
These maps \( T^*, T_* \) are linear, and preserve the wedge-product. In particular the spaces \( M^{(n)} \) and \( M_{(n)} \) are 1-dimensional. Therefore for \( M \xrightarrow{T} M \) the push-forward:
\[
M_{(n)} \xrightarrow{T_*} M_{(n)};
\]
and the pull-back
\[
M^{(n)} \xrightarrow{T^*} M^{(n)}
\]
must each be multiplication by a scalar (called \( \det T, \det T^* \) respectively).

To see what these scalars are let \( T \) have matrix \( A = (\alpha^j_i) \) wrt basis \( u_i \). Then
\[
T_* (u_1 \wedge \cdots \wedge u_n) = Tu_1 \wedge \cdots \wedge Tu_n
\]
\[
\quad = (\alpha^1_{i_1} u_{i_1}) \wedge \cdots \wedge (\alpha^n_{i_n} u_{i_n})
\]
\[
\quad = \det A u_1 \wedge \cdots \wedge u_n.
\]
Therefore \( M_{(n)} \xrightarrow{T_*} M_{(n)} \) is multiplication by \( \det A \). Similarly \( M^{(n)} \xrightarrow{T^*} M^{(n)} \) is multiplication by \( \det A^t = \det A \). Therefore
\[
\det T = \det T^* = \det A,
\]
independent of choice of basis.

**Example:** Let \( \dim M = 4 \), and \( M \xrightarrow{T} M \) have matrix \( A = (\alpha^j_i) \) wrt basis \( u_i \). Then
\[
\dim M(2) = \frac{4!}{2!2!} = 6,
\]

10–20
and wrt basis $u_1 \wedge u_2, u_1 \wedge u_3, u_1 \wedge u_4, u_2 \wedge u_3, u_2 \wedge u_4, u_3 \wedge u_4$

$$T_* : M(2) \rightarrow M(2)$$

satisfies

$$T_*(u_1 \wedge u_2) = Tu_1 \wedge Tu_2$$

$$= (\alpha_1^1 u_1 + \alpha_1^2 u_2 + \alpha_1^3 u_3 + \alpha_1^4 u_4) \wedge (\alpha_2^1 u_1 + \alpha_2^2 u_2 + \alpha_2^3 u_3 + \alpha_2^4 u_4)$$

$$= (\alpha_1^1 \alpha_2^2 - \alpha_1^2 \alpha_2^1) u_1 \wedge u_2 + \ldots$$

Therefore matrix of $T_*$ is a $6 \times 6$ matrix whose entries are $2 \times 2$ subdeterminants of $A$.

**Theorem 10.5.** If $M \xrightarrow{T} M$ has rank $r$ then

(i) $M(r) \xrightarrow{T_*} M(r)$ is non-zero,

(ii) $M(r+1) \xrightarrow{T_*} M(r+1)$ is zero.

**Proof**

(i) Let $y_1, \ldots, y_r$ be a basis for $\text{im} T$, and let $y_i = Tx_i$. Then

$$T_* x_1 \wedge \cdots \wedge x_r = Tx_1 \wedge \cdots \wedge Tx_r = y_1 \wedge \cdots \wedge y_r \neq 0.$$

(ii) Let $u_i$ be a basis for $M$. Then

$$T_* u_1 \wedge \cdots \wedge u_{r+1} = Tu_1 \wedge \cdots \wedge Tu_{r+1} = 0,$$

since $Tu_1, \ldots, Tu_{r+1} \in \text{im} T$, which has dimension $r$. Therefore linearly independent.

**Corollary 10.1.** If $T$ has matrix $A$ then $\text{rank} T = r$ if and only if all $(r + 1) \times (r + 1)$ subdeterminants are zero, and there exists at least one non-zero $r \times r$ subdeterminant.
Chapter 11
Orientation

11.1 Orientation of Vector Spaces

Let $M$ be a finite dimensional real vector space. Let $P = (p_{ij}^j)$ be the transition matrix from basis $u_1, \ldots, u_n$ to basis $w_1, \ldots, w_n$:

$$u_j = p_{ij}^j w_i.$$ 

Then

$$u_1 \wedge \cdots \wedge u_n = \det P w_1 \wedge \cdots \wedge w_n.$$ 

**Definition.** $u_1, \ldots, u_n$ has *same orientation as* $w_1, \ldots, w_n$ if $u_1 \wedge \cdots \wedge u_n$ is a positive multiple of $w_1 \wedge \cdots \wedge w_n$, i.e. $\det P > 0$. Otherwise *opposite orientation as*, i.e. $\det P < 0$.

‘Same orientation as’ is an equivalence relation on the set of all bases for $M$. There are just two equivalence classes. We call $M$ an *oriented* vector space if one of these classes has been designated as *positively oriented bases* and the other as *negatively oriented bases*. We call this *choosing an orientation for $M$*.

For $\mathbb{R}^n$ we may designate the equivalence class of the usual basis $e_1, \ldots, e_n$ as being positively oriented bases. This is called the *usual orientation of $\mathbb{R}^n$.*

**Example:** In $\mathbb{R}^3$, with usual orientation. $e_1, e_2, e_3$ (see Figure 11.1) is positively oriented (by definition).

$$e_2 \wedge e_1 \wedge e_3 = -e_1 \wedge e_2 \wedge e_3,$$

$$e_2 \wedge e_3 \wedge e_1 = e_1 \wedge e_2 \wedge e_3.$$
Therefore $e_2, e_1, e_3$ is negatively oriented and $e_2, e_3, e_1$ is positively oriented.

**Definition.** Let $M$ be a real vector space of finite dimension $n$ with a non-singular symmetric scalar product $(\cdot | \cdot)$. We call $u_1, \ldots, u_n$ a *standard basis* if

$$
(u_i | u_j) = \begin{pmatrix}
\pm 1 & \cdots & \pm 1 \\
\vdots & \ddots & \vdots \\
\pm 1 & \cdots & \pm 1
\end{pmatrix}.
$$

Recall that such bases for $M$ exist, and the numbers of $-$ signs is uniquely determined.

**Theorem 11.1.** Let $M$ be oriented. Then the $n$-form

$$
\text{vol} = u_1 \wedge u_2 \wedge \cdots \wedge u_n
$$

is independent of the choice of positively oriented standard basis for $M$. It is called the *volume form on* $M$.

If $v_1, \ldots, v_n$ is any positively oriented basis for $M$ then

$$
\text{vol} = \sqrt{(-1)^s \det(v_i | v_j)} v_1 \wedge \cdots \wedge v^n.
$$

*Proof* ▶
1. Let $w_1, \ldots, w_n$ be another positively oriented standard basis for $M$: $w^i = p^i_j w^j$ (say). Therefore

$$w^1 \land \cdots \land w^n = \det P u^1 \land \cdots \land u^n.$$ 

But $\det P > 0$ and

$$P^t \begin{pmatrix} \pm 1 \\ \ldots \\ \pm 1 \end{pmatrix} P = \begin{pmatrix} \pm 1 \\ \ldots \\ \pm 1 \end{pmatrix}.$$ 

Therefore

$$(-1)^s (\det P)^2 = (-1)^s.$$ 

Therefore

$$(\det P)^2 = 1.$$ 

Therefore $\det P = 1$. Therefore

$$w^1 \land \cdots \land w^n = u^1 \land \cdots \land u^n = \text{vol},$$

as required.

2. Let $u^i = p^i_j v^j$ (say), $\det P > 0$. Then

$$P^t \begin{pmatrix} \pm 1 \\ \ldots \\ \pm 1 \end{pmatrix} P = G,$$

where $g_{ij} = (v_i | v_j)$. Therefore

$$(-1)^s (\det P)^2 = \det G.$$ 

Therefore

$$\det P = \sqrt{(-1)^s \det G}.$$ 

Therefore

$$\text{vol} = u^1 \land \cdots \land u^n = \det P v^1 \land \cdots \land v^n = \sqrt{(-1)^s \det G} v^1 \land \cdots \land v^n,$$

as required. ◀

**Corollary 11.1.** $\text{vol}$ has components $\sqrt{\det g_{ij}} \epsilon_{i_1 \ldots i_n}$ wrt any positively oriented basis.
**Example:** Take $\mathbb{R}^n$ with usual orientation and dot product. Let

$$D = \{t^1 v_1 + \cdots + t^n v_n : 0 \leq t^i \leq 1\}$$

be the parallelogram spanned by vectors $v_1, \ldots, v_n$. Let $A$ be the matrix having $v_1, \ldots, v_n$ as columns.

$$\mathbb{R}^n \xrightarrow{A} \mathbb{R}^n, \quad v_i = Ae_i.$$

![Figure 11.2](image-url)

(For example, see Figure 11.2).

$$\text{vol}(v_1, \ldots, v_n) = \text{vol}(Ae_1, \ldots, Ae_n)$$

$$= \det A \text{vol}(e_1, \ldots, e_n)$$

$$= \det A$$

$$= \pm |\det A|$$

$$= \pm \text{Lebesgue measure of } D,$$

and Lebesgue measure of $D = \sqrt{|\det(v_i|v_j)|}$.

We continue to consider a real oriented vector space $M$ of finite dimension $n$ with a non-singular symmetric scalar product $(\cdot, \cdot)$ with $s$ signs.

$M^{(r)}$ denotes the vector space of skew-symmetric tensors of type

$$M \times \cdots \times M \rightarrow K.$$

**Theorem 11.2.** There exists a unique linear operator

$$M^{(r)} \xrightarrow{\Psi} M^{(n-r)},$$

11–4
called the Hodge star operator, with the property that for each positively oriented standard basis \( u_1, \ldots, u_n \) we have

\[
* (u^1 \wedge \cdots \wedge u^r) = s_{r+1} \cdots s_n u^{r+1} \wedge \cdots \wedge u^n \quad \text{(no summation here),}
\]

where

\[
g_{ij} = \begin{pmatrix}
s_1 & 0 \\
s_2 & \ddots \\
0 & \ddots & s_n
\end{pmatrix} = g^{ij}, \quad s_1 = \pm 1.
\]

**Example:** If \( M \) is 3-dimensional oriented Euclidean, and \( u_1, u_2, u_3 \) is any positively oriented orthonormal then

\[
M^{(1)} \to M^{(2)}, \quad M^{(2)} \to M^{(1)},
\]

with

\[
* (\alpha_1 u^1 + \alpha_2 u^2 + \alpha_3 u^3) = \alpha_1 u^2 \wedge u^3 + \alpha_2 u^3 \wedge u^1 + \alpha_3 u^1 \wedge u^2,
\]

\[
* (\alpha_1 u^2 \wedge u^3 + \alpha_2 u^3 \wedge u^1 + \alpha_3 u^1 \wedge u^2) = \alpha_1 u_1 + \alpha_2 u^2 + \alpha_3 u^3.
\]

Thus, if \( v \) has components \( \alpha_i \) wrt \( u^i \), and \( w \) has components \( \beta_i \) wrt \( u^i \) then

\[
* (v \wedge w) \text{ has components } \epsilon^{ijk} \alpha_j \beta_k \text{ wrt } u^i \text{ for any positively oriented orthonormal basis } u_i.
\]

We write \( v \times w = *(v \wedge w) \), and call it the **vector product** of \( v \) and \( w \) because:

\[
v \times w = *(v \wedge w)
\]

\[
= *[(\alpha_1 u^1 + \alpha_2 u^2 + \alpha_3 u^3) \wedge (\beta_1 u^1 + \beta_2 u^2 + \beta_3 u^3)]
\]

\[
= *[(\alpha_2 \beta_3 - \alpha_3 \beta_2) u^2 \wedge u^3 + \ldots]
\]

\[
= (\alpha_2 \beta_3 - \alpha_3 \beta_2) u^1 + \ldots,
\]

as required.

**Proof** (of theorem) If \( * \) exists then it must be unique, from the definition. Thus it is sufficient to define one such operator \( * \). For any positively oriented basis we define \( * \omega \) by contraction as:

\[
(*\omega)_{i_1 \cdots i_r} = \frac{(-1)^r}{r!} g^{i_1 j_1} \cdots g^{i_r j_r} \omega_{j_1 \cdots j_r} \sqrt{|\det g_{ij}|} \epsilon_{i_1 \cdots i_r i_{r+1} \cdots i_n}.
\]

1-5
\(*\omega\) is then well-defined independent of choice of basis, since contraction is independent of a choice of basis. Thus wrt a positively oriented standard basis \(u_1, \ldots, u_n\),

\[
g^{ij} = \begin{cases} s_i & i = j \\ 0 & i \neq j \end{cases}\quad \text{and} \quad \omega = u^1 \wedge \cdots \wedge u^r.
\]

\(\omega_{i_1, r} = 1\), other components of \(\omega\) by skew-symmetry, otherwise zero. Therefore

\[
(\ast \omega)_{r+1, \ldots, n} = \frac{(-1)^{s}}{r!} s_1 s_2 \ldots s_r \omega_{1, r} r! 1 \cdot \ldots \cdot 1 = s_{r+1} \cdots s_n,
\]
as required as other components of \(\ast \omega\) by skew-symmetry are zero.

**Theorem 11.3.** There is a unique scalar product \((\cdot | \cdot)\) on each \(M(r)\) such that

\[
\omega \land \eta = (\ast \omega | \eta) \text{ vol}
\]
for each \(\omega \in M(r), \eta \in M^{(n-r)}\). The scalar product is non-singular and symmetric for a standard basis

\[
(u^1 \wedge \cdots \wedge u^r | u^1 \wedge \cdots \wedge u^r) = (u^1 | u^1) \cdots (u^r | u^r) = \pm 1,
\]
and \(u^1 \wedge \ldots u^r\) is orthogonal to the other basis element of \(\{u^1 \wedge \cdots \wedge u^r\}_{i_1 < \cdots < i_r}\).

**Proof** Define \((\cdot | \cdot)\) by \(\omega \land \eta = (\ast \omega | \eta) \text{ vol}\). Then \((\cdot | \cdot)\) is a bilinear form on \(M^{(n-r)}\).

If \(u_1, \ldots, u_n\) is a positively oriented standard basis for \(M\) then

\[
\text{vol} = \underbrace{u^1 \wedge \cdots \wedge u^r \wedge u^{r+1} \wedge \cdots \wedge u^n}_{\omega} \quad \underbrace{\omega}_{\eta} = (s_{r+1} \ldots s_n u^{r+1} \wedge \cdots \wedge u^n | u^{r+1} \wedge \cdots \wedge u^n) \text{ vol}.
\]

Therefore

\[
(u^{r+1} \wedge \cdots \wedge u^n | u^{r+1} \wedge \cdots \wedge u^n) = s_{r+1} \cdots s_n = (u^{r+1} | u^{r+1}) \cdots (u^n | u^n),
\]
as required. Similarly other scalar products give zero.

Similarly other scalar products give zero.

**Example:** If \(u_1, u_2, u_3\) is an orthonormal basis for \(M\) then \(u^2 \wedge u^3, u^3 \wedge u^1, u^1 \wedge u^3\) is an orthonormal basis for \(M^{(2)}\), since

\[
(u^2 \wedge u^3 | u^2 \wedge u^3) = (u^2 | u^2)(u^3 | u^3) = 1.1 = 1,
(u^2 \wedge u^3 | u^3 \wedge u^1) = (u^2 | u^3)(u^3 | u^1) = 0.0 = 0.
\]

11–6
Orientation of Coordinate Systems

**Definition.** Let $X$ be an $n$-dimensional manifold. Two coordinate systems on $X$: $y^1, \ldots, y^n$ with domain $V$, and $z^1, \ldots, z^n$ with domain $W$ have the same orientation if

$$\frac{\partial(y^1, \ldots, y^n)}{\partial(z^1, \ldots, z^n)} > 0$$

on $V \cap W$. We call $X$ oriented if a family of coordinate systems is given on $X$ whose domains cover $X$, and such that any two have the same orientation. We then call these coordinate systems positively oriented.

**Note.** On $V \cap W$,

$$dy^i = \frac{\partial y^i}{\partial z^j} dz^j.$$

Therefore

$$dy^1 \wedge \cdots \wedge dy^n = \det \left( \frac{\partial y^i}{\partial z^j} \right) dz^1 \wedge \cdots \wedge dz^n = \frac{\partial(y^1, \ldots, y^n)}{\partial(z^1, \ldots, z^n)} dz^1 \wedge \cdots \wedge dz^n.$$

Therefore for each $a \in V \cap W$, $\frac{\partial y^1}{\partial z^1}, \ldots, \frac{\partial y^n}{\partial z^n}$ has same orientation as $\frac{\partial z^1}{\partial y^1}, \ldots, \frac{\partial z^n}{\partial y^n}$. Thus each tangent space $T_a X$ is an $n$-dimensional oriented vector space.

If $X$ has a metric tensor $(\cdot \cdot)$ then $T_a X$ has a non-singular symmetric scalar product $(\cdot \cdot)_a$ for each $a \in X$. Therefore we can define a differential $n$-form vol on $X$, called the volume form on $X$ by:

$$(\text{vol})_a = \text{the volume form on } T_a X.$$

Also, if $\omega$ is a differential $r$-form on $X$ then we can define a differential $(n - r)$-form on $X$, $\ast \omega$, called the Hodge star of $\omega$ by

$$(\ast \omega)_a = \ast(\omega_a) \quad \text{for each } a \in X.$$

If $u_1, \ldots, u_n$ are positively oriented vector fields on $X$ with domain $V$, and

$$ds^2 = \pm(u^1)^2 \pm \cdots \pm (u^n)^2$$

then

$$\text{vol} = u^1 \wedge \cdots \wedge u^n$$
on $V$. 

11-7
If $y^1, \ldots, y^n$ is a positively oriented coordinate system on $X$ with domain $V$ then

$$\text{vol} = \sqrt{|\det g_{ij}|} \, dy^1 \wedge \cdots \wedge dy^n$$
on $V$, where

$$g_{ij} = \left( \frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right).$$

Examples:

1. $\mathbb{R}^2$, with usual coordinates: $x, y$, and polar coordinates: $r, \theta$ (see Figure 11.3). Take $x, y$ as positively oriented:

$$x = r \cos \theta, \quad y = r \sin \theta.$$

Therefore

$$dx \wedge dy = (\cos \theta \, dr - r \sin \theta \, d\theta) \wedge (\sin \theta \, dr + r \cos \theta \, d\theta) = r \, dr \wedge d\theta.$$

![Figure 11.3](image)

$r > 0$. Therefore $r, \theta$ is positively oriented.

area element $= dx \wedge dy = r \, dr \wedge d\theta,$

$$ds^2 = (dx)^2 + (dy)^2 = (dr)^2 + (r \, d\theta)^2.$$  

Therefore

$$\ast dx = dy, \quad \ast dy = -dx,$$

$$\ast dr = r \, d\theta, \quad \ast (r \, d\theta) = -dr.$$
2. Unit sphere \( S^2 \) in \( \mathbb{R}^3 : x^2 + y^2 + z^2 = 1 \). On \( S^2 \) we have:

\[
x \, dx + y \, dy + z \, dz = 0.
\]

Therefore (wedge with \( dx \)):

\[
y \, dx \wedge dy + z \, dx \wedge dz = 0.
\]

Therefore

\[
dx \wedge dy = \frac{z}{y} \, dz \wedge dx.
\]

Therefore the coordinate system \( x, y \) on \( z > 0 \) has the same orientation as the coordinate system \( z, x \) on \( y > 0 \).

We orient \( S^2 \) so that these coordinates are positively oriented. Now

\[
ds^2 = (dx)^2 + (dy)^2 + (dz)^2
\]

\[
= (dx)^2 + (dy)^2 + \left( -\frac{x}{z} \, dx - \frac{y}{z} \, dy \right)^2 \quad \text{(on } z > 0 \text{)}
\]

\[
= \left( 1 + \frac{x^2}{z^2} \right) (dx)^2 + 2 \frac{xy}{z^2} \, dx \, dy + \left( 1 + \frac{y^2}{z^2} \right) (dy)^2.
\]

Therefore wrt coordinates \( x, y \),

\[
g_{ij} = \begin{pmatrix}
1 + \frac{x^2}{z^2} & \frac{xy}{z^2} \\
\frac{xy}{z^2} & 1 + \frac{y^2}{z^2}
\end{pmatrix}.
\]

Therefore

\[
\det g_{ij} = 1 + \frac{x^2}{z^2} + \frac{y^2}{z^2} = \frac{x^2 + y^2 + z^2}{z^2} = \frac{1}{z^2}.
\]

Therefore

\[
\text{area element } = \frac{1}{|z|} \, dx \wedge dy.
\]
Chapter 12

Manifolds and \((n)\)-dimensional Vector Analysis

This chapter could be considered a continuation of Chapter 11.

12.1 Gradient

**Definition.** If \(f\) is a scalar field on a manifold \(X\), with non-singular metric tensor \((\cdot|\cdot)\), then we define the *gradient* of \(f\) to be the vector field \(\text{grad} \, f\) such that

\[
(\text{grad} \, f | v) = < df | v > = v f = \text{rate of change of } f \text{ along } v
\]

for each vector field \(v\).

Thus \(\text{grad} \, f\) is raising the index of \(df\).

\[
df = \frac{\partial f}{\partial y^i} dy^i
\]

has components

\[
\frac{\partial f}{\partial y^i},
\]

and

\[
\text{grad} \, f = g^{ij} \frac{\partial f}{\partial y^j} \frac{\partial}{\partial y^i}
\]

has components

\[
g^{-1} \frac{\partial f}{\partial y^i}.
\]

**Theorem 12.1.** *If the metric tensor is positive definite then*
(i) \( \text{grad } f \) is in the direction of fastest increase of \( f \);

(ii) the rate of change of \( f \) in the direction of fastest increase is \( \| \text{grad } f \| \);

(iii) \( \text{grad } f \) is orthogonal to the level surfaces of \( f \) (see Figure 12.1).

\[
\text{grad } f
\]

\[
\text{Figure 12.1}
\]

**Proof**

(i) The rate of change of \( f \) along any unit vector field \( v \) has absolute value

\[
|vf| = |(\text{grad } f|v)| \leq \| \text{grad } f \| \|v\| = \| \text{grad } f \|
\]

by Cauchy-Schwarz.

(ii) For

\[
v = \frac{\text{grad } f}{\| \text{grad } f \|}
\]

the maximum is attained:

\[
|vf| = \left| \left( \text{grad } f \left| \frac{\text{grad } f}{\| \text{grad } f \|} \right. \right) \right| = \| \text{grad } f \|.
\]

(iii) If \( v \) is tangented to the level surface \( f = c \) then

\[
(\text{grad } f|v) = vf = 0.
\]

**Definition.** If \( f \) is a scalar field on a \( 2n \)-dimensional manifold \( X \), with coordinates

\[
x^i = (q^1 \ldots q^n \ p_1 \ldots p_n)
\]
and skew-symmetric tensor

\[ \langle \cdot, \cdot \rangle = \sum_{i=1}^{n} dp_i \wedge dq^i \]

then along a curve \( \alpha \) whose velocity vector is \( \text{grad} \, f \) we have:

\[ \frac{dx^i}{dt} = g^{ij} \frac{\partial}{\partial x^j}, \]

i.e.

\[
\begin{pmatrix}
\frac{dq^1}{dt} \\
\vdots \\
\frac{dq^n}{dt}
\end{pmatrix} = \begin{pmatrix}
0 & 1 & \cdots & 0 \\
-1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 \\
\end{pmatrix}
\begin{pmatrix}
\frac{\partial f}{\partial q^1} \\
\vdots \\
\frac{\partial f}{\partial q^n}
\end{pmatrix}
\]

i.e.

\[
\frac{dq^i}{dt} = \frac{\partial f}{\partial p_i}, \\
\frac{dp_i}{dt} = -\frac{\partial f}{\partial q^i}
\]

(\textit{Hamiltonian Equations of Motion}).

\textbf{Note.}

\[
\frac{d}{dt} f(\alpha(t)) = \frac{\partial f}{\partial x^i} (\alpha(t)) \frac{dx^i}{dt}(\alpha(t)) = g^{ij}(\alpha(t)) \frac{\partial f}{\partial x^j}(\alpha(t)) \frac{\partial f}{\partial x^i}(\alpha(t)) = 0,
\]

since \( g^{ij} \) is skew-symmetric.

i.e.

rate of change of \( f \) along \( \text{grad} \, f \) = \( \langle df, \text{grad} \, f \rangle = (\text{grad} \, f \mid \text{grad} \, f) = 0, \)

since \( \langle \cdot, \cdot \rangle \) is skew-symmetric.
12.2 3-dimensional Vector Analysis

Given $X$ a 3-dimensional manifold, $x, y, z$ positively oriented orthonormal coordinates, i.e. metric tensor with line element $(dx)^2 + (dy)^2 + (dz)^2$. Write

$$\nabla = \left( \frac{\partial}{\partial x} \frac{\partial}{\partial y} \frac{\partial}{\partial z} \right), \quad dr = (dx \ dy \ dz),$$

$$dS = (dy \wedge dz, \ dz \wedge dx, \ dx \wedge dy), \quad dV = dx \wedge dy \wedge dz,$$

$$F = (F^1 \ F^2 \ F^3) = (F_1 \ F_2 \ F_3), \quad \nabla f = \left( \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \frac{\partial f}{\partial z} \right) \text{ for a scalar field } f,$$

$$\nabla \times F = \left( \frac{\partial F^3}{\partial y} - \frac{\partial F^2}{\partial z}, \right), \quad \nabla \cdot F = \frac{\partial F^1}{\partial x} + \frac{\partial F^2}{\partial y} + \frac{\partial F^3}{\partial z}.$$

The vector field

$$\vec{F} = F \nabla = F^1 \frac{\partial}{\partial x} + F^2 \frac{\partial}{\partial y} + F^3 \frac{\partial}{\partial z}$$

components $F$, corresponds, lowering the index, to the 1-form (components $F$)

$$F.dr = F_1 dx + F_2 dy + F_3 dz,$$

$$df = (\nabla f).dr,$$

$$d[F.dr] = (\nabla \times F).dS,$$

$$d[F.dS] = (\nabla . F)dV,$$

$$*1 = dV,$$

$$*dr = dS,$$

$$*dS = dr,$$

$$*V = 1.$$

Now

<table>
<thead>
<tr>
<th>1-forms</th>
<th>2-forms</th>
<th>3-froms</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega(X)$</td>
<td>$\Omega^1(X)$</td>
<td>$\Omega^2(X)$</td>
</tr>
</tbody>
</table>
| $F.dr$ | $\nabla \times F).dS$ | $\nabla F.dr$ | $F.dS$ | $\nabla . F)dV$

$\vec{F}$, grad $f$, curl $\vec{F}$
12.3 Results

<table>
<thead>
<tr>
<th>Field</th>
<th>Components</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>grad $f$</td>
<td>$\nabla f$</td>
<td>vector</td>
</tr>
<tr>
<td>curl $\vec{F}$</td>
<td>$\nabla \times F$</td>
<td>vector</td>
</tr>
<tr>
<td>div $\vec{F}$</td>
<td>$\nabla \cdot F$</td>
<td>scalar</td>
</tr>
</tbody>
</table>

Theorem 12.2. Let $v$ be a vector field on a manifold, with non-singular symmetric metric tensor. Let $\omega$ be the 1-form given by lowering the index of $v$. Then the scalar field $\text{div} v$ defined by:

$$d * \omega = (\text{div} v) \text{vol}$$

is called the divergence of $v$. If $v$ has components $v^i$ wrt coordinates $y^i$ then

$$\text{div} v = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y^i} (\sqrt{g} v^i),$$

where $g = |\det g_{ij}|$.

Proof

$$(*\omega)_{j_1 \cdots j_{n-1}} = g^{i_j} v_j \sqrt{g} \epsilon_{i_1 \cdots j_{n-1}} = \sqrt{g} v^i \epsilon_{i_1 \cdots j_{n-1}}.$$ Therefore

$$*\omega = \sqrt{g} v^1 dy^2 \wedge dy^3 \wedge \cdots \wedge dy^n - \sqrt{g} v^2 dy^1 \wedge dy^3 \wedge \cdots \wedge dy^n + \ldots$$

Therefore

$$d * \omega = \frac{\partial}{\partial y^i} (\sqrt{g} v^i) dy^1 \wedge dy^2 \wedge \cdots \wedge dy^n = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y^i} (\sqrt{g} v^i) \text{vol},$$
as required.

Theorem 12.3. Let $f$ be a scalar field on a manifold, with non-singular symmetric metric tensor. Then the scalar field

$$\Delta f = \text{div} \text{grad} f = *d * df$$
is called the Laplacian of $f$. Wrt coordinates $y^i$ we have

$$\Delta f = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y^i} \left( \sqrt{g} g^{i_j} \frac{\partial f}{\partial y^j} \right).$$

Thus

$$\Delta f = \frac{1}{\sqrt{g}} \left( \frac{\partial}{\partial y^1} \cdots \frac{\partial}{\partial y^n} \right) \sqrt{g} \begin{pmatrix} g^{i_1} & \cdots & g^{i_n} \\ \vdots & \ddots & \vdots \\ g^{n_1} & \cdots & g^{n_n} \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial y^1} \\ \vdots \\ \frac{\partial f}{\partial y^n} \end{pmatrix}.$$
Examples:

(i) $\mathbb{R}^n$, with $(ds)^2 = (dx^1)^2 + \cdots + (dx^n)^2$, $(g^{ij}) = I$, $g = 1$. Therefore

$$\Delta f = \left( \frac{\partial}{\partial x^1} \ldots \frac{\partial}{\partial x^n} \right) \left( \begin{array}{c} \frac{\partial f}{\partial x^1} \\ \vdots \\ \frac{\partial f}{\partial x^n} \end{array} \right) = \frac{\partial^2 f}{\partial x^1^2} + \cdots + \frac{\partial^2 f}{\partial x^n^2}.$$ 

Therefore

$$\Delta = \frac{\partial^2}{\partial x^1^2} + \cdots + \frac{\partial^2}{\partial x^n^2} \text{ usual Laplacian.}$$

(ii) $\mathbb{R}^4$, with $(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2 - (dt)^2$: Minkowski.

$$\Delta f = \left( \frac{\partial}{\partial x} \frac{\partial}{\partial y} \frac{\partial}{\partial z} \frac{\partial}{\partial t} \right) \left( \begin{array}{cccc} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{array} \right) \left( \begin{array}{c} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial t} \end{array} \right)$$

$$= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} - \frac{\partial^2 f}{\partial t^2}.$$ 

Therefore

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial t^2} \text{ wave operator.}$$

(iii) $S^2$, with $(ds)^2 = (d\theta)^2 + (\sin \theta d\varphi)^2$, $g_{ij} = \left( \begin{array}{cc} 1 & 0 \\ 0 & \sin^2 \theta \end{array} \right)$, $g = \sin^2 \theta$.

$$\Delta f = \frac{1}{\sin \theta} \left( \frac{\partial}{\partial \theta} \frac{\partial}{\partial \varphi} \right) \sin \theta \left( \begin{array}{cc} 1 & 0 \\ 0 & \frac{1}{\sin \theta} \end{array} \right) \left( \begin{array}{c} \frac{\partial f}{\partial \theta} \\ \frac{\partial f}{\partial \varphi} \end{array} \right)$$

$$= \frac{1}{\sin \theta} \left( \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{\partial}{\partial \varphi} \left( \frac{1}{\sin \theta} \frac{\partial f}{\partial \varphi} \right) \right)$$

Definition. Let $X$ be a 3-dimensional oriented manifold with non-singular symmetric metric tensor. Let $v$ be a vector field corresponding to the 1-form $\omega$. Then $\text{curl } v$ is the vector field corresponding to the 1-form $*d\omega$. Wrt positively oriented coordinates $y^i$, curl $v$ has components:

$$\epsilon^{ijk} \frac{1}{\sqrt{g}} \left( \frac{\partial v_k}{\partial y^j} - \frac{\partial v_j}{\partial y^k} \right),$$

where $g = | \det g_{ij} |$. 

12–6
12.4  Closed and Exact Forms

Definition. A differential form $\omega \in \Omega^r(X)$ is called

(i) closed if $d\omega = 0$;

(ii) exact if $\omega = d\eta$ for some $\eta \in \Omega^{r-1}(X)$.

We note that

1. $\omega$ exact $\Rightarrow$ $\omega$ closed ($d\eta = 0$),

2. $\omega$ an exact 1-form $\Rightarrow$ $\omega = df$ (say)

$$\Rightarrow \int_{\alpha} \omega = \int_{\alpha} df = 0$$

for each closed curve $\alpha$.

Examples:

1. If $\omega = P \, dx + Q \, dy$ is a 1-form on an open $V \subset \mathbb{R}^2$ then

$$\omega = dP \wedge dx + dQ \wedge dy$$

$$= \left( \frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy \right) \wedge dx + \left( \frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy \right) \wedge dy$$

$$= \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy.$$ 

Therefore

$$\omega$$ is closed $\iff$ $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$ on $V$, 

$$\omega$$ is exact $\iff$ $P = \frac{\partial f}{\partial x}, Q = \frac{\partial f}{\partial y}$

for some scalar field $f$ on $V$.

2. The 1-form

$$\omega = \frac{x \, dy - y \, dx}{x^2 + y^2}$$

on $\mathbb{R}^2 - \{0\}$ is called the angle-form about 0. We have:

$$\frac{\partial}{\partial x} \left( \frac{x}{x^2 + y^2} \right) = + \frac{1}{x^2 + y^2} - \frac{2x^2}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

$$\frac{\partial}{\partial y} \left( \frac{-y}{x^2 + y^2} \right) = - \frac{1}{x^2 + y^2} + \frac{2y^2}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$
Therefore $\omega$ is closed.

$\omega$ is not exact because if $\alpha(t) = (\cos t, \sin t)$ is the unit circle about 0 $(0 \leq t \leq 2\pi)$ then

$$\int_\alpha \omega = \int_0^{2\pi} \frac{\cos t \sin t - \sin t(-\sin t)}{\cos^2 t + \sin^2 t} dt = \int_0^{2\pi} 1 dt = 2\pi \neq 0.$$  

However, on $\mathbb{R}^2 - \{\text{negative or zero } x\text{-axis}\}$ we have

$$x = r \cos \theta, \quad y = r \sin \theta,$$

where $\theta$ is a scalar field, with $-\pi \leq \theta \leq \pi$ and

$$\omega = r \cos \theta.(-r \cos \theta) - r \sin \theta.(-r \sin \theta) \frac{d\theta}{r^2 \cos^2 \theta + r^2 \sin^2 \theta} = d\theta.$$  

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure12.2.png}
\caption{The angle change along the curve.}
\end{figure}

Therefore, if $\alpha$ is a curve from $a$ to $b$ (see Figure 12.2),

$$\int_\alpha \omega = \int_\alpha d\theta = \theta(b) - \theta(a) = \text{change in angle along } \alpha.$$  

Note.

$$\frac{dz}{z} \frac{z}{\overline{z}} = \frac{(x - iy)(dx + i dy)}{x^2 + y^2} = \frac{x dx + y dy}{x^2 + y^2} + i \frac{y dx - x dy}{x^2 + y^2}.$$  

Therefore $\omega = \text{im} \frac{dz}{\overline{z}}.$
3. If
\[ \vec{F} = F_1 \frac{\partial}{\partial x} + F_2 \frac{\partial}{\partial y} + F_3 \frac{\partial}{\partial z} \]
is a force field in \( \mathbb{R}^3 \) then
\[ F \cdot dr = F_1 dx + F_2 dy + F_3 dz, \quad F_i = F_i \]
is called the \textit{work element}.
\[ \int F \cdot dr = \text{work done by the force } \vec{F} \text{ along the curve } \alpha. \]
\( \vec{F} \) is \textit{conservative} if \( F \cdot dr \) is exact, i.e.
\[ F \cdot dr = dV, \]
where \( V \) is a scalar field. \( V \) is called a \textit{potential function} for \( \vec{F} \).
Work done by \( \vec{F} \) along \( \alpha \) from \( a \) to \( b \) (see Figure 12.3) is
\[ \int_a F \cdot dr = \int_a dV = V(b) - V(a) = \text{potential difference}. \]

![Figure 12.3](image)

A necessary condition that \( \vec{F} \) be conservative is that \( F \cdot dr \) be closed, i.e.
\[ \nabla \times F = 0. \]

### 12.5 Contraction and Results

**Definition.** An open set \( V \subset \mathbb{R}^n \) is called \textit{contractible} to \( a \in V \) if there exists a \( C^\infty \) map
\[ V \times [0,1] \xrightarrow{\varphi} V \]
such that

\[ \varphi(x, 1) = x, \]
\[ \varphi(x, 0) = a \]

for all \( x \in V \).

_Example:_ \( V \) star-shaped \( \Rightarrow \) \( V \) contractible. \( \varphi(x, t) = tx + (1 - t)a \) (see Figure 12.4).

\[ \text{Figure 12.4} \]

**Theorem 12.4 (Poincaré Lemma).** Let \( \omega \in \Omega^r(V) \), where \( V \) is a contractible open subset of \( \mathbb{R}^n \), and \( r \geq 1 \). Then \( \omega \) is exact iff \( \omega \) is closed.

**Proof** Let \( I = [0, 1] \) be the unit interval \( 0 \leq t \leq 1 \), and define a linear operator (‘homotopy’)

\[ \Omega^r(V \times I) \xrightarrow{H} \Omega^{r-1}(V) \]

for each \( r \geq 1 \) by

\[ H[f dt \wedge dx^J] = \left( \int_0^1 f dt \right) dx^J, \]
\[ H[f dx^J] = 0. \]

Now calculate the operator \( dH + Hd \):

(i) if \( \eta = f dt \wedge dx^J \) then

\[ dH \eta + H \, d\eta = d \left[ \left( \int_0^1 f \, dt \right) dx^J \right] + H \left[ -\frac{\partial f}{\partial x^i} dt \wedge dx^i \wedge dx^J \right] \]
\[ = \left( \int_0^1 \frac{\partial f}{\partial x^i} dt \right) dx^i \wedge dx^J - \left( \int_0^1 \frac{\partial f}{\partial x^i} dt \right) dx^i \wedge dx^J \]
\[ = 0; \]
(ii) if $\eta = f \, dx^J$ then

$$dH_\eta + H_\eta \, d\eta = 0 + H \left[ \frac{\partial f}{\partial t} \, dt \wedge dx^J + \frac{\partial f}{\partial x^i} \, dx^i \wedge dx^J \right]$$

$$= \left( \int_0^1 \frac{\partial f}{\partial t} \, dt \right) \, dx^J + 0$$

$$= [f(x, t)]_{t=0}^{t=1} \, dx^J.$$

(iii) Now let $V$ be contractible, with

$$V \times I \overset{\varphi}{\to} V$$

Figure 12.5

a $C^\infty$ map such that (see Figure 12.5)

$$\varphi(x, 1) = x,$$

$$\varphi(x, 0) = a.$$

So

$$\varphi^i(x, 1) = x^i,$$

$$\varphi^i(x, 0) = a^i.$$

Therefore

$$\frac{\partial \varphi^i}{\partial x^j} = \begin{cases} \delta_j^i \text{ at } t = 1; \\ 0 \text{ at } t = 0. \end{cases}$$

Let $\omega \in \Omega^r(V)$, say $\omega = g(x)dx^{i_1} \wedge \cdots \wedge dx^{i_r}$. Apply $\varphi^*$:

$$\varphi^* \omega = g(\varphi(x, t))(d\varphi^{i_1}) \wedge \cdots \wedge (d\varphi^{i_r})$$

$$= g(\varphi(x, t)) \left[ \frac{\partial \varphi^{i_1}}{\partial x^{j_1}} dx^{j_1} + \frac{\partial \varphi^{i_1}}{\partial t} dt \right] \wedge \cdots \wedge \left[ \frac{\partial \varphi^{i_r}}{\partial x^{j_r}} dx^{j_r} + \frac{\partial \varphi^{i_r}}{\partial t} dt \right]$$

$$= g(\varphi(x, t)) \left[ \frac{\partial \varphi^{i_1}}{\partial x^{j_1}} \cdots \frac{\partial \varphi^{i_r}}{\partial x^{j_r}} dx^{j_1} \wedge \cdots \wedge dx^{j_r} + \left( \frac{\partial \varphi^{i_1}}{\partial t} \cdots \frac{\partial \varphi^{i_r}}{\partial t} dt \wedge \cdots \wedge dt \right) \wedge dt \right].$$
Apply $dH + Hd$:

$$(dH + Hd)\varphi^*\omega = \left[ g(\varphi(x,t)) \frac{\partial \varphi^i}{\partial x^i} \cdots \frac{\partial \varphi^r}{\partial x^r} \right]_{t=0}^1 dx^1 \wedge \cdots \wedge dx^r + 0$$

$$= g(x) \delta^i_1 \cdots \delta^r_i dx_i \wedge \cdots \wedge dx_r$$

$$= g(x) dx^i \wedge \cdots \wedge dx^r$$

$$= \omega.$$

Hence:

$$(dH + Hd)\varphi^*\omega = \omega$$

for all $\omega \in \Omega^r(V)$.

(iv) Let $\omega$ be closed. Then

$$d\varphi^*\omega = \varphi^* d\omega = 0.$$

Therefore

$$dH \varphi^*\omega = \omega.$$

Therefore $\omega$ is exact. ▲

**Theorem 12.5.** Let $\omega$ be a closed $r$-form, with domain $V$ open in manifold $X$. Let $a \in V$. Then there exists an open neighbourhood $W$ of $a$ such that $\omega$ is exact on $W$.

**Proof** ▲ Let $y$ be a coordinate system on $X$ at $a$, domain $U \subset V$, say. Let $W \subset U$ be an open neighbourhood of $a$ such that $y(W)$ is an open ball (see Figure 12.8).

![Figure 12.6](image_url)
Consider
\[ W \xrightarrow{\varphi} y(W), \quad W \xleftarrow{\varphi} y(W) \]
(open ball), where \( \varphi \) is the inverse map.

\[ d\varphi^*\omega = \varphi^*d\omega = 0, \text{ since } \omega \text{ is closed. Therefore } \varphi^*\omega \text{ is closed. Therefore } \varphi^*\omega = d\eta \text{ on } y(W), \text{ by Poincaré. Therefore } \]
\[ dy^*\eta = y^*d\eta = y^*\varphi^*\omega = \omega \]
on \( W \). Therefore \( \omega \) is exact on \( W \).

**Theorem 12.6.** Let \( X \) be a 2-dimensional oriented manifold with a positive definite metric tensor. Let \( u_1, u_2 \) be positively oriented orthonormal vector fields on \( X \), with domain \( V \) (moving frame). Then there exists a unique 1-form \( \omega \) on \( V \) such that
\[
(*) \quad \begin{pmatrix} du^1 \\ du^2 \end{pmatrix} = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} \wedge \begin{pmatrix} u^1 \\ u^2 \end{pmatrix}
\]
on \( V \). \( \omega \) is called the connection form (gauge field) wrt moving frame \( u_1, u_2 \).

**Proof** Any 1-form \( \omega \) on \( V \) can be written uniquely as:
\[ \omega = \alpha u^1 + \beta u^2, \quad \alpha, \beta \text{ scalar fields.} \]
\( \omega \) satisfies \((*)\) if and only if
\[
\begin{align*}
du^1 &= -\omega \wedge u^2, \\
du^2 &= \omega \wedge u^1,
\end{align*}
\]
\( \Leftrightarrow \)
\[
\begin{align*}
du^1 &= -(\alpha u^1 + \beta u^2) \wedge u^2 = -\alpha u^1 \wedge u^2, \\
du^2 &= (\alpha u^1 + \beta u^2) \wedge u^1 = -\beta u^1 \wedge u^2.
\end{align*}
\]
Thus \( \alpha, \beta \) are uniquely determined.

**Theorem 12.7.** Let \( X \) be a 2-dimensional oriented manifold with positive definite metric tensor. Let \( u_1, u_2 \) be a moving frame with domain \( V \), with connection form \( \omega \) and
\[ d\omega = Ku^1 \wedge u^2 = K \text{ area element.} \]
Then the scalar field \( K \) is independent of the choice of moving frame, and is called the Gaussian curvature of \( X \).
Proof ▶ Let $w_1, w_2$ be another moving frame with domain $V$:
\[
\begin{pmatrix}
w_1 \\
w_2
\end{pmatrix} = \begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2
\end{pmatrix}
\]  
(say). Write this in matrix form as:
\[
w = Pu.
\]
Also
\[
\begin{pmatrix}
du_1 \\
du_2
\end{pmatrix} = \begin{pmatrix}
0 & -\omega \\
\omega & 0
\end{pmatrix} \wedge \begin{pmatrix}
u_1 \\
u_2
\end{pmatrix}.
\]
Write this in matrix form as:
\[
du = \Omega \wedge u.
\]
Then
\[
dw = (dP) \wedge u + P du
\]
\[
= (dP) \wedge u + P\Omega \wedge u
\]
\[
= (dP + P\Omega) \wedge u
\]
\[
= (dP + P\Omega) \wedge P^{-1}w
\]
\[
= [(dP)P^{-1} + \Omega P^{-1}] \wedge w
\]
\[
= \left[\begin{pmatrix}
-\sin \theta d\theta & -\cos \theta d\theta \\
\cos \theta d\theta & -\sin \theta d\theta
\end{pmatrix} \begin{pmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{pmatrix}
+ \begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix} \begin{pmatrix}
0 & -\omega \\
\omega & 0
\end{pmatrix} \begin{pmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{pmatrix}\right] \wedge w
\]
\[
= \left[\begin{pmatrix}
0 & -d\theta \\
d\theta & 0
\end{pmatrix} + \begin{pmatrix}
0 & -\omega \\
\omega & 0
\end{pmatrix}\right] \wedge w
\]
\[
= \begin{pmatrix}
0 & -(\omega + d\theta) \\
\omega + d\theta & 0
\end{pmatrix} \wedge w.
\]
Therefore $\omega + d\theta$ is the connection form wrt moving frame $w_1, w_2$ and
\[
d[\omega + d\theta] = d\omega + d d\theta = d\omega,
\]
as required. ▶

Example: On $S^2$, with angle coordinates $\theta, \varphi$,
\[
ds^2 = (d\theta)^2 + (\sin \theta d\varphi)^2, \quad u^1 = d\theta, \quad u^2 = \sin \theta d\varphi.
\]
Recall

\[ \omega = -\cos \theta \, d\varphi, \]
\[ d\omega = \sin \theta \, d\theta \wedge d\varphi = u^1 \wedge u^2. \]

Therefore Gaussian curvature is constant function 1.

**Theorem 12.8.** Let \( X \) be an oriented 2-dimensional manifold with positive definite metric tensor. Then the Gaussian curvature of \( X \) is zero iff for each \( a \in X \) there exists local coordinates \( x, y \) such that

\[ ds^2 = (dx)^2 + (dy)^2, \]

i.e.

\[ g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \]

**Proof** Let \( u_1, u_2 \) be a moving frame with connection form \( \omega \) on an open neighbourhood \( V \) of \( a \) on which the Poincaré lemma holds. Then, on \( V \):

Gaussian curvature is zero \( \iff \) \( d\omega = 0 \)

\[ \iff \omega = -d\theta \quad \text{(say), by Poincaré} \]
\[ \iff \omega + d\theta = 0 \]
\[ \iff u_1, u_2 \text{ can be rotated to a new frame } w_1, w_2 \]

having connection form 0

\[ \iff dw^1 = 0, dw^2 = 0 \]
\[ \iff w^1 = dx, w^2 = dy \quad \text{(say), by Poincaré} \]
\[ \iff ds^2 = (dx)^2 + (dy)^2. \]

\( \tilde{N} \)

### 12.6 Surface in \( \mathbb{R}^3 \)

Let \( X \) be a 2-dimensional submanifold of \( \mathbb{R}^3 \). Denote all vectors by their usual components in \( \mathbb{R}^3 \). Let \( \tilde{N} \) be a field of unit vectors normal to \( X \), \( t^1, t^2 \) be coordinates on \( X \), and let \( \vec{r} = (x, y, z) \). Let \( \frac{\partial \vec{r}}{\partial t^1}, \frac{\partial \vec{r}}{\partial t^2} \) be a basis for vectors tangent to \( X \) (see Figure 12.9).

\[ \frac{\partial \vec{r}}{\partial t^i} \cdot \tilde{N} = 0. \]
Therefore

\[
\frac{\partial^2 \mathbf{r}}{\partial t \partial v} \mathbf{N} + \frac{\partial \mathbf{r}}{\partial t} \frac{\partial \mathbf{N}}{\partial v} = 0.
\]

Figure 12.7
If $\mathbf{u}$ is a vector field on $X$ tangent to $X$ then
\[
\mathbf{N} \cdot \mathbf{N} = 1
\]
along $\mathbf{u}$. Therefore
\[
(\nabla_\mathbf{u} \mathbf{N}) \cdot \mathbf{N} + \mathbf{N} \cdot \nabla_\mathbf{u} \mathbf{N} = 0.
\]
Therefore $\nabla_\mathbf{u} \mathbf{N}$ is tangential to $X$.

So define tensor field $S$ on $X$
\[
S \mathbf{u} = -\nabla_\mathbf{u} \mathbf{N}.
\]

$S$ is called the \textit{shape operator}, and measures the amount of curvature of $X$ in $\mathbb{R}^3$. With coordinates $t^1, t^2$ $S$ has covariant components
\[
S_{ij} = \left( \frac{\partial}{\partial t^i} S \frac{\partial}{\partial t^j} \right) = \frac{\partial \mathbf{N}}{\partial t^i} - \frac{\partial \mathbf{N}}{\partial t^j} = \frac{\partial^2 x}{\partial t^i \partial t^j} \mathbf{N} \quad \text{(symmetric)}.
\]

Therefore $S_e$ is a self-adjoint operator on $T_a X$ for each $a$. Therefore it has real eigenvalues $K_1, K_2$ and orthonormal eigenvectors $u_1, u_2$ (see Figure 12.8) exist, (say) $K_1 \geq K_2$.

![Figure 12.8](image)

If we intersect $X$ by a plane normal to $X$ containing the vector
\[
\cos \theta \, u_1 + \sin \theta \, u_2
\]
at $a$, we have a curve $\alpha$ of intersection along which the unit tangent vector $\mathbf{t}$ satisfies:
\[
\mathbf{t} \cdot \mathbf{N} = 0.
\]
Therefore
\[
(\nabla_\mathbf{t} \mathbf{N}) \cdot \mathbf{N} + \mathbf{t} \cdot (\nabla_\mathbf{t} \mathbf{N}) = 0.
\]
Therefore
\[
\kappa \mathbf{N} \cdot \mathbf{N} = \mathbf{t} \cdot S \mathbf{t} = 0
\]
12–17
at $a$, where $\kappa$ is the curvature of $\alpha$ at $a$. Therefore 
\[
k = \mathbf{t} \cdot \mathbf{S} \mathbf{t} = (\cos \theta u_1 + \sin \theta u_2) \cdot S(\cos \theta u_1 + \sin \theta u_2) \quad \text{(at } a) \\
= \kappa_1 \cos^2 \theta + \kappa_2 \sin^2 \theta.
\]
Therefore $u_1$ is the direction of maximum curvature $K_1$, and $u_2$ is the direction of minimum curvature $K_2$.

Put $\mathbf{N} = u_3$, with $u_1, u_2, u_3$ a moving frame.
\[
\nabla u_3 = -\omega_3^1 \otimes u_1 - \omega_3^2 \otimes u_2 - \omega_3^3 \otimes u_3.
\]
Therefore
\[
Su_1 = -\nabla_{u_1} u_3 = \langle \omega_3^1, u_1 \rangle u_1 + \langle \omega_3^2, u_1 \rangle u_2,
\]
\[
Su_2 = -\nabla_{u_2} u_3 = \langle \omega_3^1, u_2 \rangle u_1 + \langle \omega_3^2, u_2 \rangle u_2.
\]
Therefore
\[
\kappa_1 \kappa_2 = \det S \\
= \langle \omega_3^1, u_1 \rangle \langle \omega_3^2, u_2 \rangle - \langle \omega_3^1, u_2 \rangle \langle \omega_3^2, u_1 \rangle \\
= \omega_3^1 \wedge \omega_3^2(\langle u_1, u_2 \rangle) \\
= d\omega_2^1(\langle u_1, u_2 \rangle) \\
= \kappa u_1 \wedge u_2(\langle u_1, u_2 \rangle) \\
= \kappa
\]
(since $d\Omega = -\Omega \wedge \Omega$, so $d\omega_2^1 = -\omega_1^1 \wedge \omega_2^2 = -\omega_1^1 \wedge \omega_2^3 = \omega_1^1 \wedge \omega_2^3$).

\section{12.7 Integration on a Manifold (Sketch)}

Let $\omega$ be a differential $n$-form on an oriented $n$-dimensional manifold $X$. We want to define
\[
\int_X \omega,
\]
the integral of $\omega$ over $X$.

To justify in detail the construction which follows $X$ must satisfy some conditions. It is sufficient, for instance, that $X$ be a submanifold of some $\mathbb{R}^N$.

(i) Suppose
\[
\omega = f(y^1, \ldots, y^n)dy^1 \wedge \cdots \wedge dy^n \\
= y^* f(x^1, \ldots, x^n)dx^1 \wedge \cdots \wedge dx^n \\
= y^* \omega_1
\]

12–18
on the domain $V$ of a positively oriented coordinate system $y^i$ (see Figure 12.9), and that $\omega$ is zero outside $V$, and that

$$\text{supp } f = \text{ closure of } \{x \in y(V) : f(x) \neq 0\}$$

is a bounded set contained in $y(V)$. Then we define

$$\int_X \omega = \int_{y(V)} \omega_1,$$

i.e.

$$\int_X f(y^1, \ldots, y^n)dy^1 \wedge \cdots \wedge dy^n = \int_{y(V)} f(x_1, \ldots, x_n)dx_1 \cdots dx_n$$

(Lebesgue integral). The definition of $\int_X \omega$ does not depend on the choice of coordinates, since if $z^i$ with domain $W$ is another such coordinate system,

$$\omega = z^*\omega_2$$

(say), then

$$\varphi^*\omega_2 = \omega_1,$$

where $\varphi = z \circ y^{-1}$ (see Figure 12.10). Therefore

$$\int_{y(V)} \omega_1 = \int_{y(V \cap W)} \omega_1 = \int_{y(V \cap W)} \varphi^*\omega_2 = \int_{z(V \cap W)} \omega_2 = \int_{z(W)} \omega_2.$$
(ii) For a general $n$-form $\omega$ on $X$ we write

$$\omega = \omega_1 + \cdots + \omega_r,$$

where each $\omega_i$ is an $n$-form satisfying the conditions of (i), and define

$$\int_X \omega = \int_X \omega_1 + \cdots + \int_X \omega_r,$$

and check that the result is independent of the choice of $\omega_1, \ldots, \omega_r$.

**Definition.** If $X$ has a metric tensor then

$$volume \ of \ X = \int_X volume \ form.$$

**Example:** If

$$\vec{v} = v^1 \frac{\partial}{\partial x} + v^2 \frac{\partial}{\partial y} + v^3 \frac{\partial}{\partial z} = \vec{v} \cdot \nabla$$
is a vector field in $\mathbb{R}^3$,
\[ v = v_1 \, dx + v_2 \, dy + v_3 \, dz = \mathbf{\tilde{v}} \cdot d\mathbf{r} \]
is the corresponding 1-form ($v_i = v^i$), and
\[ *v = v_1 \, dy \wedge dz + v_2 \, dz \wedge dx + v_3 \, dx \wedge dy = \mathbf{\tilde{v}} \cdot d\mathbf{s} \]
is the corresponding 2-form then if $u_i$ is a moving frame, with $u_3 = \mathbf{\tilde{N}}$ normal to surface $S$ (see Figure 12.11), then
\[ \mathbf{\tilde{v}} = \alpha^i u_i, \]
\[ v = \alpha_i u^i, \]
\[ *v = \alpha_1 u^2 \wedge u^3 + \alpha_2 u^3 \wedge u^1 + \alpha_3 u^1 \wedge u^2, \]
where $\alpha_i = \alpha^i$. Therefore pull-back of $*v$ to $X$ is
\[ \alpha_3 u^1 \wedge u^2 = (\mathbf{\tilde{v}} \cdot \mathbf{\tilde{N}}) d\mathbf{S}, \]
where $d\mathbf{S} = u^1 \wedge u^2$ (area form). Therefore
\[ \int v \cdot d\mathbf{S} = \int (\mathbf{\tilde{v}} \cdot \mathbf{\tilde{N}}) dS = \text{flux of} \mathbf{\tilde{v}} \text{ across } S \]
\[ \int \mathbf{\tilde{N}} \cdot d\mathbf{S} = \int dS = \text{area of } S. \]

Note. $\mathbf{\tilde{v}} \cdot d\mathbf{r}$ is work element, $\mathbf{\tilde{v}} \cdot d\mathbf{s}$ is flux element, $\mathbf{\tilde{N}} \cdot d\mathbf{s}$ is area element of vector field $\mathbf{\tilde{v}}$. 

![Figure 12.11](image-url)
Theorem 12.9 (Stokes). (George Gabriel Stokes 1819 - 1903, Skreen Co. Sligo) Let \( \omega \) be an \((n - 1)\)-form on an \(n\)-dimensional manifold \( X \) with an \((n - 1)\)-dimensional boundary \( \partial X \) (see Figure 12.12). Then

\[
\int_X d\omega = \int_{\partial X} i^* \omega,
\]

where \( \partial X \xrightarrow{i} X \) is the inclusion map.

Proof (Sketch) We write

\[
\omega = \omega_1 + \cdots + \omega_r,
\]

where each \( \omega_i \) satisfies the conditions of either (i), (ii) or (iii) below, and we prove it for each \( \omega_i \). It then follows for \( \omega \).

\( \omega \) zero outside \( V \), and \( \text{supp} \ f \) a closed bounded subset of \( y(V) \) (see Figure 12.13).

\[
\int_{\partial X} i^* \omega = 0,
\]

since \( \omega \) is zero on \( \partial X \).

\[
d\omega = \frac{\partial f}{\partial y^1} dy^1 \wedge dy^2 \wedge \cdots \wedge dy^n.
\]
Therefore
\[
\int_X d\omega = \int_{y(V)} \frac{\partial f}{\partial x_1} dx_1 dx_2 \ldots dx_n \\
= \int_{-1}^{1} \int_{-1}^{1} [\int_{-1}^{1} \frac{\partial f}{\partial x_1} dx_1] dx_2 \ldots dx_n \\
= \int_{-1}^{1} \int_{-1}^{1} [f(1, x_2, \ldots, x_n) - f(-1, x_2, \ldots, x_n)] dx_2 \ldots dx_n \\
= 0.
\]
Therefore
\[
\int_X d\omega = 0 = \int_{\partial X} i^* \omega.
\]
For (ii) and (iii), let \( \omega \) be zero outside the domain \( W \) of functions \( y_1, y_2, \ldots, y_n \), with \( y \) a homeomorphism:
\[
y(W) = (-1, 0] \times (-1, 1) \times \cdots \times (-1, 1)
\]
(see Figure 12.14), where $y^1, y^2, \ldots, y^n$ are positively oriented coordinates on $X$ with domain $W - (W \cap \partial X)$, $y^1 = 0$ on $W \cap \partial X$, and $y^2, \ldots, y^n$ are positively oriented coordinates on $\partial X$ with domain $W \cap \partial X$. Then

(ii) if

$$\omega = f(y^1, y^2, \ldots, y^n)dy^2 \wedge \cdots \wedge dy^n \quad \text{(say)},$$

(supp $f$ closed bounded subset of $y(W))$ then

$$i^*\omega = f(0, y^2, \ldots, y^n)dy^2 \wedge \cdots \wedge dy^n,$$

$$d\omega = \frac{\partial f}{\partial y^1}dy^1 \wedge dy^2 \wedge \cdots \wedge dy^n.$$
Therefore
\[
\int_X d\omega = \int_{\partial f(W)} \frac{\partial f}{\partial x_1} dx_1 dx_2 \ldots dx_n
\]
\[
= \int_{\partial f(W \cap \partial X)} f(0, x_2, \ldots, x_n) dx_2 \ldots dx_n
\]
\[
= \int_{\partial X} i^* \omega;
\]

(iii) if
\[
\omega = f(y^1, \ldots, y^n)dy^1 \wedge dy^3 \wedge \cdots \wedge dy^n
\]
(say), (supp \( f \) closed bounded subset of \( y(W) \)) then
\[
i^* \omega = 0,
\]

since \( y^1 = 0 \) on \( W \cap \partial X \). Also
\[
d\omega = -\frac{\partial f}{\partial y^2} dy^1 \wedge dy^2 \wedge \cdots \wedge dy^n.
\]

Therefore
\[
\int_X d\omega = -\int_{-1}^0 \ldots \int_{-1}^0 \left[ \int_{-1}^0 \frac{\partial f}{\partial x_2} dx_2 \right] dx_1 \ldots dx_n = 0.
\]

Therefore
\[
\int_X d\omega = 0 = \int_{\partial X} i^* \omega.
\]
Applications of Stokes Theorem:

1. In $\mathbb{R}^2$: $\omega$ 1-form (see Figure 12.17).

\[
\int_{\partial D} (P \, dx + Q \, dy) = \int_D \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dx \wedge dy \quad \text{Green’s Theorem.}
\]

In particular:

(a) \[
\int_{\partial D} x \, dy = - \int_{\partial D} y \, dx = \frac{1}{2i} \int_{\partial D} z \, dz = \int_D dx \wedge dy = \text{area of } D.
\]

(b) If $f = u + iv$, $\omega = f \, dz$ then

\[
d\omega = \left[ -\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x} + i \left( \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) \right] \, dx \wedge dy.
\]

Now \[d\omega = 0 \iff f \text{ holomorphic , by Cauchy-Riemann. Therefore}
\[
\int_{\partial D} f(z) \, dz = 0 \text{ if } f \text{ holomorphic } \quad \text{(Cauchy)}. \]

2. In $\mathbb{R}^3$: $X$ surface (see Figure 12.18).

\[
\int_{\partial X} F \cdot dr = \int_X (\nabla \times F) \cdot dS = \int_X (\nabla \times F) \cdot \hat{N} \, dS,
\]

i.e.

work of $F$ around loop $\partial X = \text{flux of } \nabla \times F \text{ across surface } X$

$= \text{flux of } \nabla \times F \text{ through loop } \partial X$.

3. In $\mathbb{R}^3$:

\[
\int_{\partial D} F \cdot \hat{N} \, dS = \int_{\partial D} F \cdot dS = \int_D (\nabla \cdot F) \, dV,
\]

i.e. (see Figure 12.19)

flux of $F$ out of region $D = \text{integral of } \nabla \cdot F \text{ over interior of } D$.
If $\nabla \cdot F > 0$ at $a$ then $a$ is a **source** for $F$.

if $\nabla \cdot F < 0$ at $a$ then $a$ is a **sink** for $F$.

if $\nabla \cdot F = 0$ then $F$ is **source-free** (see Figure 12.20).

4. If $X$ is $n$-dimensional and $\omega \wedge \eta$ an $(n-1)$-form then

$$\int_{\partial X} \omega \wedge \eta = \int_X d(\omega \wedge \eta) \quad \text{(by Stokes)}$$

$$= \int_X (d\omega) \wedge \eta + (-1)^{r+1} \int_X \omega \wedge d\eta,$$

$\omega$ an $r$-form, by Leibnsing. Therefore

$$\int_X (d\omega) \wedge \eta = \int_{\partial X} \omega \wedge \eta + (-1)^{r+1} \int_X \omega \wedge d\eta$$

(by integration by parts) (see Figure 12.21).

**Example:** $X$ connected, 3-dimensional in $\mathbb{R}^3$.

$$\int_{\partial X} f \nabla f \cdot n dS = \int_X f(\nabla f, dS)$$

$$= \int_X [\nabla f, \nabla f + f \nabla (\nabla f)]$$

$$= \int_X [||\nabla f||^2 + f \nabla^2 f] dV.$$

Therefore

$$\nabla^2 f = 0; \ f \ or \ \nabla f \cdot n = 0 \ on \ \partial X \Rightarrow \nabla f = 0 \ on \ X \Rightarrow f = 0 \ on \ X$$

(Dirichlet Neumann).

If $X$ has a metric tensor and no boundary then

$$\int_X (*d\omega|\eta) \text{ vol} = (-1)^{r+1} \int_X (*\omega|d\eta) \text{ vol}.$$

If (say) the metric is positive definite and $n$ odd then $** = 1$, so putting $*\omega$ in place of $\omega$:

$$\int_X ((-1)^r * d * \omega|\eta) \text{ vol} = \int_X (\omega|d\eta) \text{ vol}.$$
Therefore \((-1)^r \ast d^*\) is the adjoint of the operator \(d\). Hence \(\delta = \pm \ast d^*\) is the operator adjoint to \(d\).

\[
\Delta = d\delta + \delta d
\]

is self-adjoint, and is called the Laplacian on \(f\).

5. If \(X\) is the unit ball in \(\mathbb{R}^n\) (see Figure 12.22), with \(n \geq 2\):

\[
X = \{x \in \mathbb{R}^n : \sum x_i^2 \leq 1\}
\]

then \(\partial X\) is the \((n - 1)\)-dimensional sphere

\[
\partial X = \{x \in \mathbb{R}^n : \sum x_i^2 = 1\}.
\]

**Theorem 12.10.** There is no \(C^\infty\) map

\[
X \xrightarrow{\varphi} \partial X
\]

which leaves each point of the sphere \(\partial X\) fixed.

**Proof** Suppose \(\varphi\) exists. Then we have a commutative diagram:

\[
\begin{array}{ccc}
\omega & \xrightarrow{\varphi} & \partial X \\
i & \uparrow i & \\
\alpha & \xrightarrow{1} & \partial X
\end{array}
\]

where \(i\) is the inclusion map and \(1\) the identity map.

Let

\[
\omega = x^1 dx^2 \wedge \cdots \wedge dx^n, \quad \alpha = i^* \omega.
\]

So

\[
d\omega = dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n \quad \text{(volume form on } X),
\]

\[
d\alpha = 0,
\]

since \(d\alpha\) is an \(n\)-form on \((n - 1)\)-dimensional \(\partial X\). Therefore

\[
i^* \omega = \alpha = i^* \varphi^* \alpha.
\]

Therefore

\[
\int_{\partial X} i^* \omega = \int_{\partial X} i^* \varphi^* \alpha.
\]

Therefore volume of \(X\) is

\[
\int_X d\omega = \int_X d\varphi^* \alpha = \int_X \varphi^* d\alpha \int_X \varphi^* 0 = 0.
\]

This is a contradiction so the result follows. \(\blacksquare\)
6. In $\mathbb{R}^4$, Minkowski: the electromagnetic field is a 2-form:

$$F = (E.dr) \wedge dt + B.dS,$$

where $E$ is components of electric field, and $B$ are components of magnetic field.

One of Maxwell’s equations is:

$$dF = 0$$

(the other is $d \ast F = J$ charge-current), i.e.

$$d[(E.dr) \wedge dt + B.dS] = 0,$$

i.e.

$$(\nabla \times E).dS \wedge dt + (\nabla.B)dV + \frac{\partial B}{\partial t} dS \wedge dt = 0,$$

i.e.

$$\nabla \times E = -\frac{\partial B}{\partial t},$$

$$\nabla.B = 0,$$

i.e. magnetic field is source free.

Therefore electromotive force (EMF) around loop $\partial X$ (see Figure 12.23) is

$$\int_{\partial X} E.dr \quad \text{Stokes} = \int_X (\nabla \times E).dS \quad \text{Maxwell} = \int_X \frac{d}{dt} B.dS$$

= rate of decrease of magnetic flux through loop.