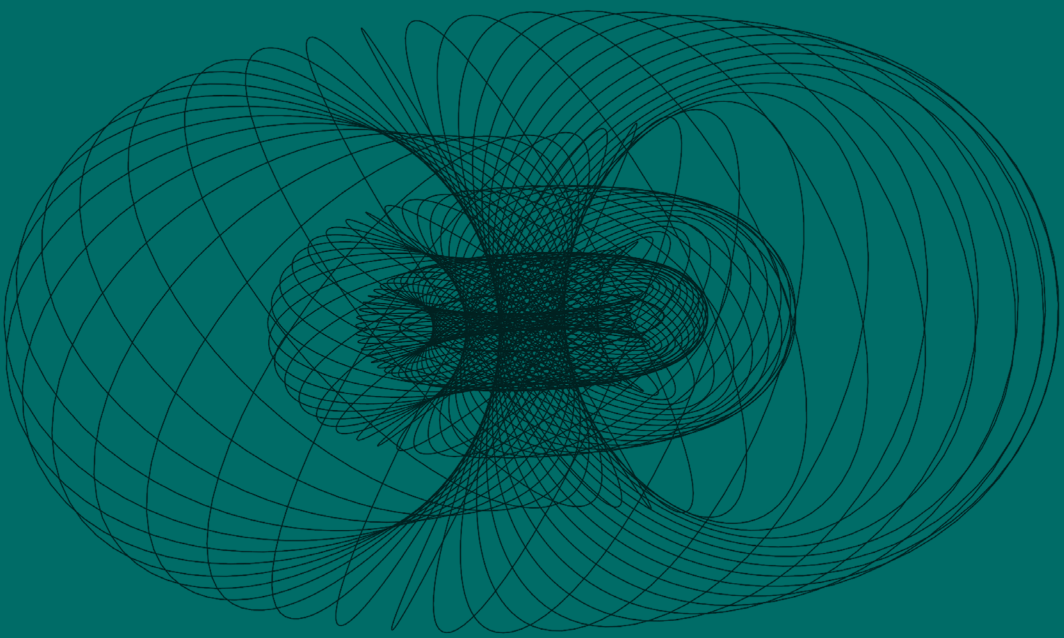
Gabriel Lugo

Differential Geometry in Physics



Differential Geometry in Physics

Gabriel Lugo

UNCW
University of North Carolina Wilmington
William Madison Randall Library

Suggested citation: Lugo, Gabriel. Differential Geometry in Physics. Wilmington. University of North Carolina Wilmington William Madison Randall Library, 2021.
doi: https: //doi.org/10.5149/9781469669267_Lugo

ISBN 978-1-4696-6924-3 (cloth: alk. paper)
ISBN 978-1-4696-6925-0 (paperback: alk. paper)
ISBN 978-1-4696-6926-7 (open access ebook)

Cover illustration: The Hopf fibration.

Published by the UNC Wilmington William Madison Randall Library

Distributed by UNC Press
www.uncpress.org

Publication of this book was supported by a grant from the Thomas W. Ross Fund from the University of North Carolina Press.

Escher detail on following page from "Belvedere" plate 230, [1]. All other figures and illustrations were produced by the author or taken from photographs taken by the author and family members. Exceptions are included with appropriate attribution.

This book is dedicated to my family, for without their love and support, this work would have not been possible. Most importantly, let us not forget that our greatest gift are our children and our children's children, who stand as a reminder, that curiosity and imagination is what drives our intellectual pursuits, detaching us from the mundane and confinement to Earthly values, and bringing us closer to the divine and the infinite.

G. Lugo (2021)

Contents

Preface

These notes were developed as part of a course on differential geometry which the author has taught for many years at UNCW. The first five chapters plus chapter six, constitute the foundation of the three-hour course. The course is cross-listed at the level of seniors and first year graduate students. In addition to applied mathematics majors, the class usually attracts a good cohort of double majors in mathematics and physics. Material from other chapters have inspired a number of honors and master level theses. This book should be accessible to students who have completed traditional training in advanced calculus, linear algebra, and differential equations. Students who master the entirety of this material will have gained insight on very powerful tools in mathematical physics at the graduate level.

There are many excellent texts in differential geometry but very few have an early introduction to differential forms and their applications to physics. It is the purpose of these notes to:

1. Provide a bridge between the very practical formulation of classical differential geometry created by early masters of the late 1800's, and the more elegant but less intuitive modern formulation in terms of manifolds, bundles and differential forms. In particular, the central topic of curvature is presented in three different but equivalent formalisms.

2. Present the subject of differential geometry with an emphasis on making the material readable to physicists who may have encountered some of the concepts in the context of classical or quantum mechanics, but wish to strengthen the rigor of the mathematics. A source of inspiration for this goal is rooted in the shock to this author as a graduate student in the 70's at Berkeley, at observing the gasping failure of communications between the particle physicists working on gauge theories and differential geometers working on connection on fiber bundles. They seemed to be completely unaware at the time, that they were working on the same subject.

3. Make the material as readable as possible for those who stand at the boundary between theoretical physics and applied mathematics. For this reason, it will be occasionally necessary to sacrifice some mathematical rigor or depth of physics, in favor of ease of comprehension.

4. Provide the formal geometrical background for the mathematical theory of general relativity.

5. Introduce examples of other applications of differential geometry to physics that might not appear in traditional texts used in courses for mathematics students. For example, several students at UNCW have written masters' theses in the theory of solitons, but usually they have followed the path of Lie symmetries in the style of Olver. We hope that the elegance of Bäcklund transforms will attract students to a geometric approach to the subject. The book is also a stepping stone to other interconnected areas of mathematics such as representation theory, complex variables and algebraic topology.

G. Lugo (2021)

Chapter 1

Vectors and Curves

1.1 Tangent Vectors

1.1.1 Definition Euclidean n-space $\mathbf{R}^n$ is defined as the set of ordered n-tuples $p(p^1, \dots, p^n)$, where $p^i \in \mathbf{R}$, for each $i = 1, \dots, n$. We may associate a position vector $\mathbf{p} = (p^1, \dots, p^n)$ with any given point a point p in n-space. Given any two n-tuples $\mathbf{p} = (p^1, \dots, p^n)$, $\mathbf{q} = (q^1, \dots, q^n)$ and any real number c, we define two operations:

$$\begin{aligned} \mathbf{p} + \mathbf{q} &= (p^1 + q^1, \dots, p^n + q^n), \\ c\,\mathbf{p} &= (c\,p^1, \dots, c\,p^n). \end{aligned} \tag{1.1}$$

These two operations of vector sum and multiplication by a scalar satisfy all the 8 properties needed to give the set $V = \mathbf{R}^n$ a natural structure of a vector space. It is common to use the same notation $\mathbf{R}^n$ for the space of n-tuples and for the vector space of position vectors. Technically, we should write $p \in \mathbf{R}^n$ when we think of $\mathbf{R}^n$ as a metric space and $\mathbf{p} \in \mathbf{R}^n$ when we think of it as vector space, but as most authors, we will freely abuse the notation. [1]

1.1.2 Definition Let x^i be the real valued functions in $\mathbf{R}^n$ such that

$$x^i(\mathbf{p}) = p^i$$

for any point $\mathbf{p} = (p^1, \dots, p^n)$. The functions x^i are then called the natural *coordinate functions*. When convenient, we revert to the usual names for the coordinates, $x^1 = x$, $x^2 = y$ and $x^3 = z$ in $\mathbf{R}^3$. A small awkwardness might

[1] In these notes we will use the following index conventions:

- In $\mathbf{R}^n$, indices such as i, j, k, l, m, n, run from 1 to n.
- In space-time, indices such as μ, ν, ρ, σ, run from 0 to 3.
- On surfaces in $\mathbf{R}^3$, indices such as $\alpha, \beta, \gamma, \delta$, run from 1 to 2.
- Spinor indices such as $A, B, \dot{A}, \dot{B}$ run from 1 to 2.

occur in the transition to modern notation. In classical vector calculus, a point in $\mathbf{R}^n$ is often denoted by $\mathbf{x}$, in which case, we pick up the coordinates with the slot projection functions $u^i : \mathbf{R}^n \to \mathbf{R}$ defined by

$$u^i(\mathbf{x}) = x^i.$$

1.1.3 Definition A real valued function in $\mathbf{R}^n$ is of class C^r if all the partial derivatives of the function up to order r exist and are continuous. The space of infinitely differentiable (smooth) functions will be denoted by $C^\infty(\mathbf{R}^n)$ or $\mathscr{F}(\mathbf{R}^n)$.

1.1.4 Definition Let V and V' be finite dimensional vector spaces such as $V = \mathbf{R}^k$ and $V' = \mathbf{R}^n$, and let $\mathbf{L}(V, V')$ be the space of linear transformations from V to V'. The set of linear functionals $\mathbf{L}(V, \mathbf{R})$ is called the *dual vector space* V^*. This space has the same dimension as V.

In calculus, vectors are usually regarded as arrows characterized by a direction and a length. Thus, vectors are considered as independent of their location in space. Because of physical and mathematical reasons, it is advantageous to introduce a notion of vectors that does depend on location. For example, if the vector is to represent a force acting on a rigid body, then the resulting equations of motion will obviously depend on the point at which the force is applied. In later chapters, we will consider vectors on curved spaces; in these cases, the positions of the vectors are crucial. For instance, a unit vector pointing north at the earth's equator is not at all the same as a unit vector pointing north at the tropic of Capricorn. This example should help motivate the following definition.

1.1.5 Definition A *tangent vector* X_p in $\mathbf{R}^n$, is an ordered pair $\{\mathbf{x}, \mathbf{p}\}$. We may regard $\mathbf{x}$ as an ordinary advanced calculus "arrow-vector" and $\mathbf{p}$ is the position vector of the foot of the arrow.

The collection of all tangent vectors at a point $\mathbf{p} \in \mathbf{R}^n$ is called the *tangent space* at $\mathbf{p}$ and will be denoted by $T_p(\mathbf{R}^n)$. Given two tangent vectors X_p, Y_p and a constant c, we can define new tangent vectors at $\mathbf{p}$ by $(X+Y)_p = X_p + Y_p$ and $(cX)_p = cX_p$. With this definition, it is clear that for each point $\mathbf{p}$, the corresponding tangent space $T_p(\mathbf{R}^n)$ at that point has the structure of a vector space. On the other hand, there is no natural way to add two tangent vectors at different points.

The set $T(\mathbf{R}^n)$ (or simply $T\mathbf{R}^n$) consisting of the union of all tangent spaces at all points in $\mathbf{R}^n$ is called the *tangent bundle*. This object is not a vector space, but as we will see later it has much more structure than just a set.

1.1.6 Definition A *vector field* X in $U \subset \mathbf{R}^n$ is a *section of the tangent bundle*, that is, a smooth function from U to $T(U)$. The space of sections $\Gamma(T(U)$ is also denoted by $\mathscr{X}(U)$.

The difference between a tangent vector and a vector field is that in the latter case, the coefficients v^i of $\mathbf{x}$ are smooth functions of x^i. Since in general

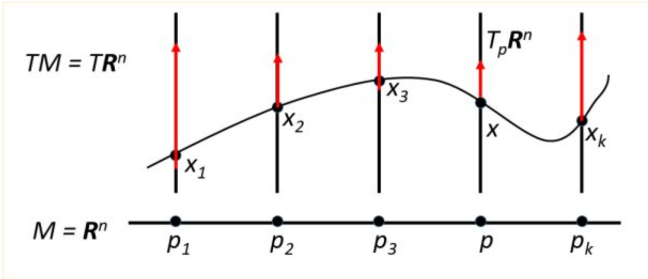


Fig. 1.1: Tangent Bundle

there are not enough dimensions to depict a tangent bundle and vector fields as sections thereof, we use abstract diagrams such as shown Figure 1.1. In such a picture, the base space M (in this case $M = \mathbf{R}^n$) is compressed into the continuum at the bottom of the picture in which several points $\mathbf{p}_1, \ldots, \mathbf{p}_k$ are shown. To each such point one attaches a tangent space. Here, the tangent spaces are just copies of $\mathbf{R}^n$ shown as vertical "fibers" in the diagram. The vector component $\mathbf{x}_p$ of a tangent vector at the point $\mathbf{p}$ is depicted as an arrow embedded in the fiber. The union of all such fibers constitutes the tangent bundle $TM = T\mathbf{R}^n$. A section of the bundle amounts to assigning a tangent vector to every point in the base. It is required that such assignment of vectors is done in a smooth way so that there are no major "changes" of the vector field between nearby points.

Given any two vector fields X and Y and any smooth function f, we can define new vector fields $X + Y$ and fX by

$$\begin{aligned} (X+Y)_p &= X_p + Y_p \\ (fX)_p &= fX_p, \end{aligned} \tag{1.2}$$

so that $\mathscr{X}(U)$ has the structure of a vector space over $\mathbf{R}$. The subscript notation X_p indicating the location of a tangent vector is sometimes cumbersome, but necessary to distinguish them from vector fields.

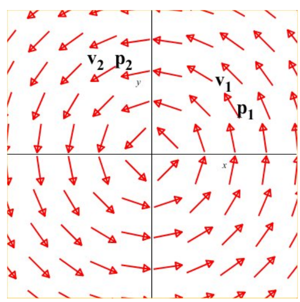


Fig. 1.2: Vector Field

Vector fields are essential objects in physical applications. If we consider the flow of a fluid in a region, the velocity vector field represents the speed and direction of the flow of the fluid at that point. Other examples of vector fields in classical physics are the electric, magnetic, and gravitational fields. The vector field in figure 1.2 represents a magnetic field around an electrical wire pointing out of the page.

1.1.7 Definition Let $X_p = \{\mathbf{x}, \mathbf{p}\}$ be a tangent vector in an open neighborhood U of a point $\mathbf{p} \in \mathbf{R}^n$ and let f be a C^∞ function in U. The directional derivative of f at the point $\mathbf{p}$, in the direction of $\mathbf{x}$, is defined by

$$X_p(f) = \boldsymbol{\nabla} f(p) \cdot \mathbf{x}, \tag{1.3}$$

where $\boldsymbol{\nabla} f(p)$ is the gradient of the function f at the point $\mathbf{p}$. The notation

$$X_p(f) \equiv \nabla_{X_p} f,$$

is also commonly used. This notation emphasizes that, in differential geometry, we may think of a tangent vector at a point as an operator on the space of smooth functions in a neighborhood of the point. The operator assigns to a function f, the directional derivative of that function in the direction of the vector. Here we need not assume as in calculus that the direction vectors have unit length.

It is easy to generalize the notion of directional derivatives to vector fields by defining

$$X(f) \equiv \nabla_X f = \boldsymbol{\nabla} f \cdot \mathbf{x}, \tag{1.4}$$

where the function f and the components of $\mathbf{x}$ depend smoothly on the points of $\mathbf{R}^n$.

The tangent space at a point p in $\mathbf{R}^n$ can be envisioned as another copy of $\mathbf{R}^n$ superimposed at the point p. Thus, at a point p in $\mathbf{R}^2$, the tangent space consist of the point p and a copy of the vector space $\mathbf{R}^2$ attached as a "tangent plane" at the point p. Since the base space is a flat 2-dimensional continuum, the tangent plane for each point appears indistinguishable from the base space as in figure 1.2.

Later we will define the tangent space for a curved continuum such as a surface in $\mathbf{R}^3$ as shown in figure 1.3. In this case, the tangent space at a point p consists of the vector space of all vectors actually tangent to the surface at the given point.

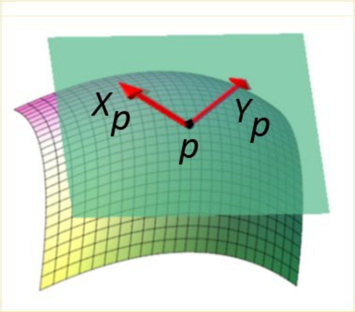


Fig. 1.3: Tangent vectors X_p, Y_p on a surface in $\mathbb{R}^3$.

1.1.8 Proposition If $f, g \in \mathscr{F}(\mathbf{R}^n)$, $a, b \in \mathbf{R}$, and $X \in \mathscr{X}(\mathbf{R}^n)$ is a vector field, then

$$\begin{aligned} X(af + bg) &= aX(f) + bX(g), \\ X(fg) &= fX(g) + gX(f). \end{aligned} \tag{1.5}$$

1.1.9 Remark The space of smooth functions is a ring, ignoring a small technicality with domains. An operator such as a vector field with the properties above, is called a *linear derivation* on $\mathscr{F}(\mathbf{R}^n)$.

Proof First, let us develop an mathematical expression for tangent vectors and vector fields that will facilitate computation.

Let $\mathbf{p} \in U$ be a point and let x^i be the coordinate functions in U. Suppose that $X_p = \{\mathbf{x}, \mathbf{p}\}$, where the components of the Euclidean vector $\mathbf{x}$ are $(v^1, \ldots, v^n)$. Then, for any function f, the tangent vector X_p operates on f according to the formula

$$X_p(f) = \sum_{i=1}^{n} v^i \left(\frac{\partial f}{\partial x^i}\right)(p). \tag{1.6}$$

It is therefore natural to identify the tangent vector X_p with the differential operator

$$\begin{aligned} X_p &= \sum_{i=1}^{n} v^i \left(\frac{\partial}{\partial x^i}\right)_p \qquad (1.7) \\ X_p &= v^1 \left(\frac{\partial}{\partial x^1}\right)_p + \cdots + v^n \left(\frac{\partial}{\partial x^n}\right)_p . \end{aligned}$$

Notation: We will be using Einstein's convention to suppress the summation symbol whenever an expression contains a repeated index. Thus, for example, the equation above could be simply written as

$$X_p = v^i \left(\frac{\partial}{\partial x^i}\right)_p . \tag{1.8}$$

This equation implies that the action of the vector X_p on the coordinate functions x^i yields the components v^i of the vector. In elementary treatments, vectors are often identified with the components of the vector, and this may cause some confusion.

The operators

$$\{e_1, \ldots, e_k\}|_p = \left\{\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p\right\}$$

form a basis for the tangent space $T_p(\mathbf{R}^n)$ at the point $\mathbf{p}$, and any tangent vector can be written as a linear combination of these basis vectors. The quantities v^i are called the **contravariant** components of the tangent vector. Thus, for example, the Euclidean vector in $\mathbf{R}^3$

$$\mathbf{x} = 3\mathbf{i} + 4\mathbf{j} - 3\mathbf{k}$$

located at a point $\mathbf{p}$, would correspond to the tangent vector

$$X_p = 3\left(\frac{\partial}{\partial x}\right)_p + 4\left(\frac{\partial}{\partial y}\right)_p - 3\left(\frac{\partial}{\partial z}\right)_p .$$

Let $X = v^i \dfrac{\partial}{\partial x^i}$ be an arbitrary vector field and let f and g be real-valued functions. Then

$$\begin{aligned} X(af+bg) &= v^i \frac{\partial}{\partial x^i}(af+bg) \\ &= v^i \frac{\partial}{\partial x^i}(af) + v^i \frac{\partial}{\partial x^i}(bg) \\ &= av^i \frac{\partial f}{\partial x^i} + bv^i \frac{\partial g}{\partial x^i} \\ &= aX(f) + bX(g). \end{aligned}$$

Similarly,

$$\begin{aligned} X(fg) &= v^i \frac{\partial}{\partial x^i}(fg) \\ &= v^i f \frac{\partial}{\partial x^i}(g) + v^i g \frac{\partial}{\partial x^i}(f) \\ &= fv^i \frac{\partial g}{\partial x^i} + gv^i \frac{\partial f}{\partial x^i} \\ &= fX(g) + gX(f). \end{aligned}$$

To re-emphasize, any quantity in Euclidean space which satisfies relations 1.5 is a called a linear derivation on the space of smooth functions. The word *linear* here is used in the usual sense of a linear operator in linear algebra, and the word derivation means that the operator satisfies Leibnitz' rule.

The proof of the following proposition is slightly beyond the scope of this course, but the proposition is important because it characterizes vector fields in a coordinate-independent manner.

1.1.10 Proposition Any linear derivation on $\mathscr{F}(\mathbf{R}^n)$ is a vector field.

This result allows us to identify vector fields with linear derivations. This step is a big departure from the usual concept of a "calculus" vector. To a differential geometer, a vector is a linear operator whose inputs are functions and whose output are functions that at each point represent the directional derivative in the direction of the Euclidean vector.

1.1.11 Example Given the point $p(1,1)$, the Euclidean vector $\mathbf{x} = (3,4)$, and the function $f(x,y) = x^2 + y^2$, we associate $\mathbf{x}$ with the tangent vector

$$X_p = 3\frac{\partial}{\partial x} + 4\frac{\partial}{\partial y}.$$

Then,

$$\begin{aligned} X_p(f) &= 3\left(\frac{\partial f}{\partial x}\right)_p + 4\left(\frac{\partial f}{\partial y}\right)_p, \\ &= 3(2x)|_p + 4(2y)|_p, \\ &= 3(2) + 4(2) = 14. \end{aligned}$$

1.1.12 Example Let $f(x, y, z) = xy^2z^3$ and $\mathbf{x} = (3x, 2y, z)$. Then

$$\begin{aligned} X(f) &= 3x\left(\frac{\partial f}{\partial x}\right) + 2y\left(\frac{\partial f}{\partial y}\right) + z\left(\frac{\partial f}{\partial z}\right) \\ &= 3x(y^2z^3) + 2y(2xyz^3) + z(3xy^2z^2), \\ &= 3xy^2z^3 + 4xy^2z^3 + 3xy^2z^3 = 10xy^2z^3. \end{aligned}$$

1.1.13 Definition Let X be a vector field in $\mathbf{R}^n$ and p be a point. A curve $\alpha(t)$ with $\alpha(0) = p$ is called an *integral curve* of X if $\alpha'(0) = X_p$, and, whenever $\alpha(t)$ is the domain of the vector field, $\alpha'(t) = X_{\alpha(t)}$.

In elementary calculus and differential equations, the families of integral curves of a vector field are called the streamlines, suggesting the trajectories of a fluid with velocity vector X. In figure 1.2, the integral curves would be circles that fit neatly along the *flow* of the vector field. In local coordinates, the expression defining integral curves of X constitutes a system of first order differential equations, so the existence and uniqueness of solutions apply locally. We will treat this in more detail in subsection 7.1.1

1.2 Differentiable Maps

1.2.1 Definition Let $F : \mathbf{R}^n \to \mathbf{R}^m$ be a vector function defined by coordinate entries $F(\mathbf{p}) = (f^1(\mathbf{p}), f^2(\mathbf{p}), \ldots f^m(\mathbf{p}))$. The vector function is called a *mapping* if the coordinate functions are all differentiable. If the coordinate functions are C^∞, F is called a smooth mapping. If $(x^1, x^2, \ldots x^n)$ are local coordinates in $\mathbf{R}^n$ and $(y^1, y^2, \ldots y^m)$ local coordinates in $\mathbf{R}^m$, a map $\mathbf{y} = F(\mathbf{x})$ is represented in advanced calculus by m functions of n variables

$$y^j = f^j(x^i), \quad i = 1 \ldots n,\ j = 1 \ldots m. \tag{1.9}$$

A map $F : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at a point $\mathbf{p} \in \mathbf{R}^n$ if there exists a linear transformation $DF(\mathbf{p}) : \mathbf{R}^n \to \mathbf{R}^m$ such that

$$\lim_{\mathbf{h}\to 0} \frac{|F(\mathbf{p}+\mathbf{h}) - F(\mathbf{p}) - DF(\mathbf{p})(\mathbf{h})|}{|\mathbf{h}|} = 0 \tag{1.10}$$

The linear transformation $DF(\mathbf{p})$is called the *Jacobian*. A differentiable map that is invertible and the inverse is differentiable, is called a *diffeomorphism*.

Remarks

1. A differentiable mapping $F : I \in \mathbf{R} \to \mathbf{R}^n$ is what we called a curve. If $t \in I = [a, b]$, the mapping gives a parametrization $\mathbf{x}(t)$, as we discussed in the previous section.

2. A differentiable mapping $F : R \in \mathbf{R}^n \to \mathbf{R}^n$ is called a coordinate transformation. Thus, for example, the mapping $F : (u, v) \in \mathbf{R}^2 \to (x, y) \in$

$\mathbf{R}^2$, given by functions $x = x(u, v),\ y = y(u, v)$, would constitute a change of coordinates from (u, v) to (x, y). The most familiar case is the polar coordinates transformation $x = r\cos\theta,\ \ y = r\sin\theta$.

3. A differentiable mapping $F : R \in \mathbf{R}^2 \to \mathbf{R}^3$ is what in calculus we called a parametric surface. Typically, one assumes that R is a simple closed region, such as a rectangle. If one denotes the coordinates in $\mathbf{R}^2$ by $(u, v) \in R$, and $\mathbf{x} \in \mathbf{R}^3$, the parametrization is written as $\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v))$. The treatment of surfaces in $\mathbf{R}^3$ is presented in chapter 4. If $\mathbf{R}^3$ is replaced by $\mathbf{R}^n$, the mapping locally represents a 2-dimensional surface in a space of n dimensions.

For each point $\mathbf{p} \in \mathbf{R}^n$, we say that the Jacobian induces a linear transformation F_* from the tangent space $T_p\mathbf{R}^n$ to the tangent space $T_{F(p)}\mathbf{R}^m$. In differential geometry we this Jacobian map is also called the *push-forward.* If we let X be a tangent vector in $\mathbf{R}^n$, then the tangent vector F_*X in $\mathbf{R}^m$ is defined by

$$F_*X(f) = X(f \circ F), \tag{1.11}$$

where $f \in \mathscr{F}(\mathbf{R}^m)$. (See figure 1.4)

$$\begin{tikzcd}
X \in T_p\mathbf{R}^n \arrow[r, "F_*"] \arrow[d] & T_{F(p)}\mathbf{R}^m \ni F_*X \arrow[d] & \\
\mathbf{R}^n \arrow[r, "F"] & \mathbf{R}^m \arrow[r, "f"] & \mathbf{R}
\end{tikzcd}$$

Fig. 1.4: Jacobian Map.

As shown in the diagram, $F_*X(f)$ is evaluated at $F(p)$ whereas X is evaluated at p. So, to be precise, equation 1.11 should really be written as

$$F_*X(f)(F(p)) = X(f \circ F)(p), \tag{1.12}$$
$$F_*X(f) \circ F = X(f \circ F), \tag{1.13}$$

As we have learned from linear algebra, to find a matrix representation of a linear map in a particular basis, one applies the map to the basis vectors. If we denote by $\{\frac{\partial}{\partial x^i}\}$ the basis for the tangent space at a point $p \in \mathbf{R}^n$ and by $\{\frac{\partial}{\partial y^j}\}$ the basis for the tangent space at the corresponding point $F(p) \in \mathbf{R}^m$ with coordinates given by $y^j = f^j(x^i)$, the push-forward definition reads,

$$\begin{aligned}
F_*(\frac{\partial}{\partial x^i})(f) &= \frac{\partial}{\partial x^i}(f \circ F), \\
&= \frac{\partial f}{\partial y^j}\frac{\partial y^j}{\partial x^i}, \\
F_*(\frac{\partial}{\partial x^i}) &= \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j}.
\end{aligned}$$

In other words, the matrix representation of F_* in standard basis is in fact the Jacobian matrix. In classical notation, we simply write the Jacobian map in the familiar form,

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j}. \tag{1.14}$$

1.2.2 Theorem If $F : \mathbf{R}^n \to \mathbf{R}^m$ and $G : \mathbf{R}^m \to \mathbf{R}^p$ are mappings, then $(G \circ F)_* = G_* \circ F_*$.
Proof Let $X \in T_p(\mathbf{R})^n$, and f be a smooth function $f : \mathbf{R}^p \to \mathbf{R}$. Then,

$$\begin{aligned}
(G \circ F)_*(X)(f) &= X(f \circ (G \circ F), \\
&= X((f \circ G) \circ F), \\
&= F_*(X)(f \circ G), \\
&= G_*(F_*(X)(f)), \\
&= (G_* \circ F_*)(X)(f).
\end{aligned}$$

1.2.3 Inverse Function Theorem. When $m = n$, mappings are called change of coordinates. In the terminology of tangent spaces, the classical inverse function theorem states that if the Jacobian map F_* is a vector space isomorphism at a point, then there exists a neighborhood of the point in which F is a diffeomorphism.

1.2.4 Remarks

1. Equation 1.14 shows that under change of coordinates, basis tangent vectors and by linearity all tangent vectors transform by multiplication by the matrix representation of the Jacobian. This is the source of the almost tautological definition in physics, that a contravariant tensor of rank one, is one that transforms like a contravariant tensor of rank one.

2. Many authors use the notation dF to denote the push-forward map F_*.

3. If $F : \mathbf{R}^n \to \mathbf{R}^m$ and $G : \mathbf{R}^m \to \mathbf{R}^p$ are mappings, we leave it as an exercise for the reader to verify that the formula $(G \circ F)_* = G_* \circ F_*$ for the composition of linear transformations corresponds to the classical chain rule.

4. As we will see later, the concept of the push-forward extends to manifold mappings $F : M \to N$.

1.3 Curves in $\mathbf{R}^3$

1.3.1 Parametric Curves

1.3.1 Definition A *curve* $\alpha(t)$ in $\mathbf{R}^3$ is a C^∞ map from an interval $I \subset \mathbf{R}$ into $\mathbf{R}^3$. The curve assigns to each value of a parameter $t \in \mathbf{R}$, a point $(\alpha^1(t), \alpha^2(t), \alpha^3(t)) \in \mathbf{R}^3$.

$$\begin{array}{rcl} I \subset \mathbf{R} & \xmapsto{\alpha} & \mathbf{R}^3 \\ t & \longmapsto & \alpha(t) = (\alpha^1(t), \alpha^2(t), \alpha^3(t)) \end{array}$$

One may think of the parameter t as representing time, and the curve α as representing the trajectory of a moving point particle as a function of time. When convenient, we also use classical notation for the position vector

$$\mathbf{x}(t) = (x^1(t), x^2(t), x^3(t)), \tag{1.15}$$

which is more prevalent in vector calculus and elementary physics textbooks. Of course, what this notation really means is

$$x^i(t) = (u^i \circ \alpha)(t), \tag{1.16}$$

where u^i are the coordinate slot functions in an open set in $\mathbf{R}^3$

1.3.2 Example Let

$$\alpha(t) = (a_1 t + b_1, a_2 t + b_2, a_3 t + b_3). \tag{1.17}$$

This equation represents a straight line passing through the point $\mathbf{p} = (b_1, b_2, b_3)$, in the direction of the vector $\mathbf{v} = (a_1, a_2, a_3)$.

1.3.3 Example Let

$$\alpha(t) = (a \cos \omega t, a \sin \omega t, bt). \tag{1.18}$$

This curve is called a circular helix. Geometrically, we may view the curve as the path described by the hypotenuse of a triangle with slope b, which is wrapped around a circular cylinder of radius a. The projection of the helix onto the xy-plane is a circle and the curve rises at a constant rate in the z-direction (See Figure 1.5a). Similarly, the equation $\alpha(t) = (a \cosh \omega t, a \sinh \omega t, bt)$ is called a hyperbolic "helix." It represents the graph of curve that wraps around a hyperbolic cylinder rising at a constant rate.

1.3.4 Example Let

$$\alpha(t) = (a(1 + \cos t), a \sin t, 2a \sin(t/2)). \tag{1.19}$$

This curve is called the *Temple of Viviani*. Geometrically, this is the curve of intersection of a sphere $x^2 + y^2 + z^2 = 4a^2$ of radius $2a$, and the cylinder

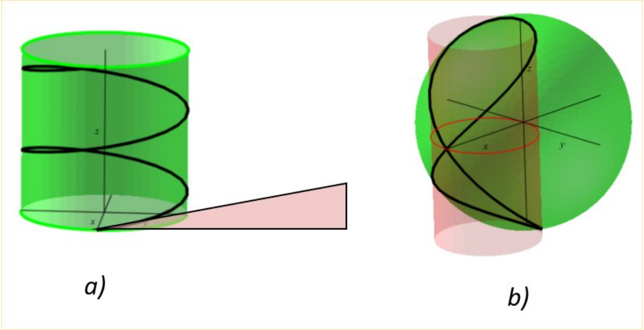
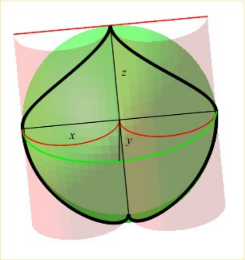


Fig. 1.5: a) Circular Helix. b) Temple of Viviani

$x^2 + y^2 = 2ax$ of radius a, with a generator tangent to the diameter of the sphere along the z-axis (See Figure 1.5b).

The Temple of Viviani is of historical interest in the development of calculus. The problem was posed anonymously by Viviani to Leibnitz, to determine on the surface of a semi-sphere, four identical windows, in such a way that the remaining surface be equivalent to a square. It appears as if Viviani was challenging the effectiveness of the new methods of calculus against the power of traditional geometry.



It is said that Leibnitz understood the nature of the challenge and solved the problem in one day. Not knowing the proposer of the enigma, he sent the solution to his Serenity Ferdinando, as he guessed that the challenge must have originated from prominent Italian mathematicians. Upon receipt of the solution by Leibnitz, Viviani posted a mechanical solution without proof. He described it as using a boring device to remove from a semisphere, the surface area cut by two cylinders with half the radius, and which are tangential to a diameter of the base. Upon realizing this could not physically be rendered as a temple since the roof surface would rest on only four points, Viviani no longer spoke of a temple but referred to the shape as a "sail."

1.3.5 Definition Let $\alpha : I \to \mathbf{R}^3$ be a curve in $\mathbf{R}^3$ given in components as above $\alpha = (\alpha^1, \alpha^2, \alpha^3)$. For each point $t \in I$ we define the *velocity* or *tangent vector* of the curve by

$$\alpha'(t) = \left(\frac{d\alpha^1}{dt}, \frac{d\alpha^2}{dt}, \frac{d\alpha^3}{dt} \right)_{\alpha(t)}. \tag{1.20}$$

At each point of the curve, the velocity vector is tangent to the curve and thus the velocity constitutes a vector field representing the velocity flow along that curve. In a similar manner the second derivative $\alpha''(t)$ is a vector field called the *acceleration* along the curve. The length $v = \|\alpha'(t)\|$ of the velocity vector is called the speed of the curve. The classical components of the velocity vector are simply given by

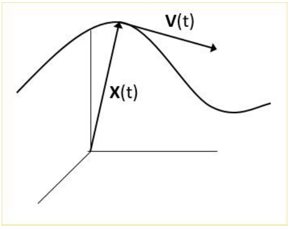
$$\mathbf{v}(t) = \dot{\mathbf{x}} \equiv \frac{d\mathbf{x}}{dt} = \left(\frac{dx^1}{dt}, \frac{dx^2}{dt}, \frac{dx^3}{dt}\right), \tag{1.21}$$

and the speed is

$$v = \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2}. \tag{1.22}$$

The notation $T(t)$ or $T_\alpha(t)$ is also used for the tangent vector $\alpha'(t)$, but for now, we reserve $T(t)$ for the unit tangent vector to be introduced in section 1.3.3 on Frenet frames.

As is well known, the vector form of the equation of the line 1.17 can be written as $\mathbf{x}(t) = \mathbf{p} + t\mathbf{v}$, which is consistent with the Euclidean axiom stating that given a point and a direction, there is only one line passing through that point in that direction. In this case, the velocity $\dot{\mathbf{x}} = \mathbf{v}$ is constant and hence the acceleration $\ddot{\mathbf{x}} = 0$. This is as one would expect from Newton's law of inertia.



The differential $d\mathbf{x}$ of the position vector given by

$$d\mathbf{x} = (dx^1, dx^2, dx^3) = \left(\frac{dx^1}{dt}, \frac{dx^2}{dt}, \frac{dx^3}{dt}\right) dt \tag{1.23}$$

which appears in line integrals in advanced calculus is some sort of an *infinitesimal tangent vector*. The norm $\|d\mathbf{x}\|$ of this infinitesimal tangent vector is called the differential of arc length ds. Clearly, we have

$$ds = \|d\mathbf{x}\| = v\, dt. \tag{1.24}$$

If one identifies the parameter t as time in some given units, what this says is that for a particle moving along a curve, the speed is the rate of change of the arc length with respect to time. This is intuitively exactly what one would expect.

The notion of infinitesimal objects needs to be treated in a more rigorous mathematical setting. At the same time, we must not discard the great intuitive value of this notion as envisioned by the masters who invented calculus, even at the risk of some possible confusion! Thus, whereas in the more strict sense of modern differential geometry, the velocity is a tangent vector and hence it is a differential operator on the space of functions, the quantity $d\mathbf{x}$ can be viewed as a traditional vector which, at the infinitesimal level, represents a linear approximation to the curve and points tangentially in the direction of $\mathbf{v}$.

1.3.2 Velocity

For any smooth function $f : \mathbf{R}^3 \to \mathbf{R}$, we formally define the action of the velocity vector field $\alpha'(t)$ as a linear derivation by the formula

$$\alpha'(t)(f) \mid_{\alpha(t)} = \frac{d}{dt}(f \circ \alpha) \mid_t . \tag{1.25}$$

The modern notation is more precise, since it takes into account that the velocity has a vector part as well as point of application. Given a point on the curve, the velocity of the curve acting on a function, yields the directional derivative of that function in the direction tangential to the curve at the point in question. The diagram in figure 1.6 below provides a more visual interpretation of the velocity vector formula 1.25, as a linear mapping between tangent spaces.

$$\begin{tikzcd}
\frac{d}{dt} \in T_t\mathbf{R} \arrow[r, "\alpha_*"] \arrow[d] & T_{\alpha(t)}\mathbf{R}^3 \ni \alpha'(t) \arrow[d] & \\
\mathbf{R} \arrow[r, "\alpha"] & \mathbf{R}^3 \arrow[r, "f"] & \mathbf{R}
\end{tikzcd}$$

Fig. 1.6: Velocity Vector Operator

The map $\alpha(t)$ from $\mathbf{R}$ to $\mathbf{R}^3$ induces a push-forward map α_* from the tangent space of $\mathbf{R}$ to the tangent space of $\mathbf{R}^3$. The image $\alpha_*(\frac{d}{dt})$ in $T\mathbf{R}^3$ of the tangent vector $\frac{d}{dt}$ is what we call $\alpha'(t)$.

$$\alpha_*(d/dt) = \alpha'(t).$$

Since $\alpha'(t)$ is a tangent vector in $\mathbf{R}^3$, it acts on functions in $\mathbf{R}^3$. The action of $\alpha'(t)$ on a function f on $\mathbf{R}^3$ is the same as the action of d/dt on the composition $(f \circ \alpha)$. In particular, if we apply $\alpha'(t)$ to the coordinate functions x^i, we get the components of the tangent vector

$$\alpha'(t)(x^i) \mid_{\alpha(t)} = \frac{d}{dt}(x^i \circ \alpha)|_t. \tag{1.26}$$

To unpack the above discussion in the simplest possible terms, we associate with the classical velocity vector $\mathbf{v} = \dot{\mathbf{x}}$ a linear derivation $\alpha'(t)$ given by

$$\begin{aligned}
\alpha'(t) &= \frac{d}{dt}(x^i \circ \alpha)_t(\partial/\partial x^i)_{\alpha(t)}, \\
&= \frac{dx^1}{dt}\frac{\partial}{\partial x^1} + \frac{dx^2}{dt}\frac{\partial}{\partial x^2} + \frac{dx^3}{dt}\frac{\partial}{\partial x^3}.
\end{aligned} \tag{1.27}$$

So, given a real valued function f in $\mathbf{R}^3$, the action of the velocity vector is given by the chain rule

$$\alpha'(t)(f) = \frac{\partial f}{\partial x^1}\frac{dx^1}{dt} + \frac{\partial f}{\partial x^2}\frac{dx^2}{dt} + \frac{\partial f}{\partial x^3}\frac{dx^3}{dt} = \boldsymbol{\nabla} f \cdot \mathbf{v}.$$

If $\alpha(t)$ is a curve in $\mathbf{R}^n$ with tangent vector $X = \alpha'(t)$, and $F : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable map, then $F_* X$ is a tangent vector to the curve $F \circ \alpha$ in $\mathbf{R}^m$. That is, F_* maps tangent vectors of α to tangent vectors of $F \circ \alpha$.

1.3.6 Definition If $t = t(s)$ is a smooth, real valued function and $\alpha(t)$ is a curve in $\mathbf{R}^3$, we say that the curve $\beta(s) = \alpha(t(s))$ is a *reparametrization* of α.

A common reparametrization of curve is obtained by using the arc length as the parameter. Using this reparametrization is quite natural, since we know from basic physics that the rate of change of the arc length is what we call speed

$$v = \frac{ds}{dt} = \|\alpha'(t)\|. \tag{1.28}$$

The arc length is obtained by integrating the above formula

$$s = \int \|\alpha'(t)\|\, dt = \int \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2}\, dt \tag{1.29}$$

In practice, it is typically difficult to find an explicit arc length parametrization of a curve since not only does one have to calculate the integral, but also one needs to be able to find the inverse function t in terms of s. On the other hand, from a theoretical point of view, arc length parameterizations are ideal, since any curve so parametrized has unit speed. The proof of this fact is a simple application of the chain rule and the inverse function theorem.

$$\begin{aligned}
\beta'(s) &= [\alpha(t(s))]' \\
&= \alpha'(t(s))t'(s) \\
&= \alpha'(t(s))\frac{1}{s'(t(s))} \\
&= \frac{\alpha'(t(s))}{\|\alpha'(t(s))\|},
\end{aligned}$$

and any vector divided by its length is a unit vector. Leibnitz notation makes this even more self-evident

$$\begin{aligned}
\frac{d\mathbf{x}}{ds} &= \frac{d\mathbf{x}}{dt}\frac{dt}{ds} = \frac{\frac{d\mathbf{x}}{dt}}{\frac{ds}{dt}} \\
&= \frac{\frac{d\mathbf{x}}{dt}}{\|\frac{d\mathbf{x}}{dt}\|}
\end{aligned}$$

1.3.7 Example Let $\alpha(t) = (a\cos\omega t, a\sin\omega t, bt)$. Then

$$\mathbf{v}(t) = (-a\omega\sin\omega t, a\omega\cos\omega t, b),$$

$$\begin{aligned} s(t) &= \int_0^t \sqrt{(-a\omega \sin \omega u)^2 + (a\omega \cos \omega u)^2 + b^2}\, du \\ &= \int_0^t \sqrt{a^2\omega^2 + b^2}\, du \\ &= ct, \quad \text{where,} \quad c = \sqrt{a^2\omega^2 + b^2}. \end{aligned}$$

The helix of unit speed is then given by

$$\beta(s) = (a \cos \frac{\omega s}{c}, a \sin \frac{\omega s}{c}, b\frac{\omega s}{c}).$$

1.3.3 Frenet Frames

Let $\beta(s)$ be a curve parametrized by arc length and let $T(s)$ be the vector

$$T(s) = \beta'(s). \tag{1.30}$$

The vector $T(s)$ is tangential to the curve and it has unit length. Hereafter, we will call T the *unit tangent* vector. Differentiating the relation

$$T \cdot T = 1, \tag{1.31}$$

we get

$$2\, T \cdot T' = 0, \tag{1.32}$$

so we conclude that the vector T' is orthogonal to T. Let N be a unit vector orthogonal to T, and let κ be the scalar such that

$$T'(s) = \kappa N(s). \tag{1.33}$$

We call N the *unit normal* to the curve, and κ the *curvature*. Taking the length of both sides of last equation, and recalling that N has unit length, we deduce that

$$\kappa = \|T'(s)\|. \tag{1.34}$$

It makes sense to call κ the curvature because, if T is a unit vector, then $T'(s)$ is not zero only if the direction of T is changing. The rate of change of the direction of the tangent vector is precisely what one would expect to measure how much a curve is curving. We now introduce a third vector

$$B = T \times N, \tag{1.35}$$

which we will call the *binormal* vector. The triplet of vectors (T, N, B) forms an orthonormal set; that is,

$$\begin{aligned} T \cdot T = N \cdot N = B \cdot B = 1, \\ T \cdot N = T \cdot B = N \cdot B = 0. \end{aligned} \tag{1.36}$$

If we differentiate the relation $B \cdot B = 1$, we find that $B \cdot B' = 0$, hence B' is orthogonal to B. Furthermore, differentiating the equation $T \cdot B = 0$, we get

$$B' \cdot T + B \cdot T' = 0.$$

rewriting the last equation

$$B' \cdot T = -T' \cdot B = -\kappa N \cdot B = 0,$$

we also conclude that B' must also be orthogonal to T. This can only happen if B' is orthogonal to the TB-plane, so B' must be proportional to N. In other words, we must have

$$B'(s) = -\tau N(s), \tag{1.37}$$

for some quantity τ, which we will call the *torsion*. The torsion is similar to the curvature in the sense that it measures the rate of change of the binormal. Since the binormal also has unit length, the only way one can have a non-zero derivative is if B is changing directions. This means that if in addition B did not change directions, the vector would truly be a constant vector, so the curve would be a flat curve embedded into the TN-plane.

The quantity B' then measures the rate of change in the up and down direction of an observer moving with the curve always facing forward in the direction of the tangent vector. The binormal B is something like the flag in the back of sand dune buggy.

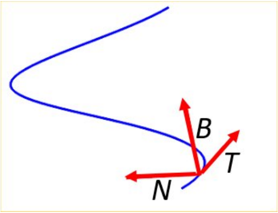


Fig. 1.7: Frenet Frame.

The set of basis vectors $\{T, N, B\}$ is called the *Frenet frame* or the *repère mobile* (moving frame). The advantage of this basis over the fixed $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ basis is that the Frenet frame is naturally adapted to the curve. It propagates along the curve with the tangent vector always pointing in the direction of motion, and the normal and binormal vectors pointing in the directions in which the curve is tending to curve. In particular, a complete description of how the curve is curving can be obtained by calculating the rate of change of the frame in terms of the frame itself.

1.3.8 Theorem Let $\beta(s)$ be a unit speed curve with curvature κ and torsion τ. Then

$$\begin{array}{rcrcrcr} T' &=& & & \kappa N & & \\ N' &=& -\kappa T & & & & \tau B \\ B' &=& & & -\tau N & & \end{array} . \tag{1.38}$$

Proof We need only establish the equation for N'. Differentiating the equation $N \cdot N = 1$, we get $2N \cdot N' = 0$, so N' is orthogonal to N. Hence, N' must be a linear combination of T and B.

$$N' = aT + bB.$$

Taking the dot product of last equation with T and B respectively, we see that

$$a = N' \cdot T, \text{ and } \quad b = N' \cdot B.$$

On the other hand, differentiating the equations $N \cdot T = 0$, and $N \cdot B = 0$, we find that

$$\begin{aligned} N' \cdot T &= -N \cdot T' = -N \cdot (\kappa N) = -\kappa \\ N' \cdot B &= -N \cdot B' = -N \cdot (-\tau N) = \tau. \end{aligned}$$

We conclude that $a = -\kappa$, $b = \tau$, and thus

$$N' = -\kappa T + \tau B.$$

The Frenet frame equations (1.38) can also be written in matrix form as shown below.

$$\begin{bmatrix} T \\ N \\ B \end{bmatrix}' = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix}. \tag{1.39}$$

The group-theoretic significance of this matrix formulation is quite important and we will come back to this later when we talk about general orthonormal frames. Presently, perhaps it suffices to point out that the appearance of an antisymmetric matrix in the Frenet equations is not at all coincidental.

The following theorem provides a computational method to calculate the curvature and torsion directly from the equation of a given unit speed curve.

1.3.9 Proposition Let $\beta(s)$ be a unit speed curve with curvature $\kappa > 0$ and torsion τ. Then

$$\begin{aligned} \kappa &= \|\beta''(s)\| \\ \tau &= \frac{\beta' \cdot [\beta'' \times \beta''']}{\beta'' \cdot \beta''} \end{aligned} \tag{1.40}$$

Proof If $\beta(s)$ is a unit speed curve, we have $\beta'(s) = T$. Then

$$\begin{aligned} T' = \beta''(s) &= \kappa N, \\ \beta'' \cdot \beta'' &= (\kappa N) \cdot (\kappa N), \\ \beta'' \cdot \beta'' &= \kappa^2 \\ \kappa^2 &= \|\beta''\|^2 \end{aligned}$$

$$\begin{aligned} \beta'''(s) &= \kappa' N + \kappa N' \\ &= \kappa' N + \kappa(-\kappa T + \tau B) \\ &= \kappa' N + -\kappa^2 T + \kappa\tau B. \end{aligned}$$

$$\begin{aligned}
\beta' \cdot [\beta'' \times \beta'''] &= T \cdot [\kappa N \times (\kappa' N + -\kappa^2 T + \kappa\tau B)] \\
&= T \cdot [\kappa^3 B + \kappa^2 \tau T] \\
&= \kappa^2 \tau \\
\tau &= \frac{\beta' \cdot [\beta'' \times \beta''']}{\kappa^2} \\
&= \frac{\beta' \cdot [\beta'' \times \beta''']}{\beta'' \cdot \beta''}
\end{aligned}$$

1.3.10 Example Consider a circle of radius r whose equation is given by

$$\alpha(t) = (r\cos t, r\sin t, 0).$$

Then,

$$\begin{aligned}
\alpha'(t) &= (-r\sin t, r\cos t, 0) \\
\|\alpha'(t)\| &= \sqrt{(-r\sin t)^2 + (r\cos t)^2 + 0^2} \\
&= \sqrt{r^2(\sin^2 t + \cos^2 t)} \\
&= r.
\end{aligned}$$

Therefore, $ds/dt = r$ and $s = rt$, which we recognize as the formula for the length of an arc of circle of radius r, subtended by a central angle whose measure is t radians. We conclude that

$$\beta(s) = (r\cos\frac{s}{r}, r\sin\frac{s}{r}, 0)$$

is a unit speed reparametrization. The curvature of the circle can now be easily computed

$$\begin{aligned}
T &= \beta'(s) = (-\sin\frac{s}{r}, \cos\frac{s}{r}, 0), \\
T' &= (-\frac{1}{r}\cos\frac{s}{r}, -\frac{1}{r}\sin\frac{s}{r}, 0), \\
\kappa &= \|\beta''\| = \|T'\|, \\
&= \sqrt{\frac{1}{r^2}\cos^2\frac{s}{r} + \frac{1}{r^2}\sin^2\frac{s}{r} + 0^2}, \\
&= \sqrt{\frac{1}{r^2}(\cos^2\frac{s}{r} + \sin^2\frac{s}{r})}, \\
&= \frac{1}{r}.
\end{aligned}$$

This is a very simple but important example. The fact that for a circle of radius r the curvature is $\kappa = 1/r$ could not be more intuitive. A small circle has large curvature and a large circle has small curvature. As the radius of the circle approaches infinity, the circle locally looks more and more like a straight line, and the curvature approaches 0. If one were walking along a great circle on a very large sphere (like the earth) one would be perceive the space to be locally flat.

1.3.11 Proposition Let $\alpha(t)$ be a curve of velocity $\mathbf{v}$, acceleration $\mathbf{a}$, speed v and curvature κ, then

$$\begin{aligned} \mathbf{v} &= vT, \\ \mathbf{a} &= \frac{dv}{dt}T + v^2\kappa N. \end{aligned} \tag{1.41}$$

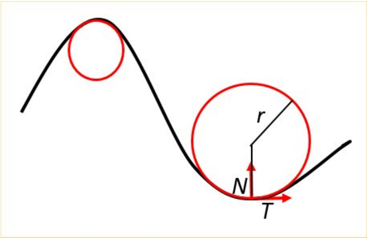


Fig. 1.8: Osculating Circle

Proof Let $s(t)$ be the arc length and let $\beta(s)$ be a unit speed reparametrization. Then $\alpha(t) = \beta(s(t))$ and by the chain rule

$$\begin{aligned} \mathbf{v} &= \alpha'(t), \\ &= \beta'(s(t))s'(t), \\ &= vT. \end{aligned}$$

$$\begin{aligned} \mathbf{a} &= \alpha''(t), \\ &= \frac{dv}{dt}T + vT'(s(t))s'(t), \\ &= \frac{dv}{dt}T + v(\kappa N)v, \\ &= \frac{dv}{dt}T + v^2\kappa N. \end{aligned}$$

Equation 1.41 is important in physics. The equation states that a particle moving along a curve in space feels a component of acceleration along the direction of motion whenever there is a change of speed, and a centripetal acceleration in the direction of the normal whenever it changes direction. The *centripetal Acceleration* and any point is

$$a = v^2\kappa = \frac{v^2}{r}$$

where r is the radius of a circle called the *osculating circle.*

The osculating circle has maximal tangential contact with the curve at the point in question. This is called contact of order 2, in the sense that the circle passes through two nearby in the curve. The osculating circle can be envisioned by a limiting process similar to that of the tangent to a curve in differential calculus. Let p be point on the curve, and let q_1 and q_2 be two nearby points. If the three points are not collinear, they uniquely determine a circle. The center of this circle is located at the intersection of the perpendicular bisectors of the segments joining two consecutive points. This circle is a "secant" approximation to the tangent circle. As the points q_1 and q_2 approach the point p, the "secant" circle approaches the osculating circle. The osculating circle, as shown in figure 1.8, always lies in the TN-plane, which by analogy is called the *osculating plane.* If $T' = 0$, then $\kappa = 0$ and the osculating circle degenerates into a circle of infinite radius, that is, a straight line. The physics interpretation of equation 1.41 is that as a particle moves along a curve, in some sense at an infinitesimal level, it is moving tangential to a circle, and hence, the centripetal acceleration at each point coincides with the centripetal acceleration along the osculating circle. As the points move along, the osculating circles move along with them, changing their radii appropriately.

1.3.12 Example (Helix)

$$
\begin{aligned}
\beta(s) &= (a\cos\frac{\omega s}{c}, a\sin\frac{\omega s}{c}, \frac{bs}{c}), \quad \text{where } c=\sqrt{a^2\omega^2+b^2},\\
\beta'(s) &= (-\frac{a\omega}{c}\sin\frac{\omega s}{c}, \frac{a\omega}{c}\cos\frac{\omega s}{c}, \frac{b}{c}),\\
\beta''(s) &= (-\frac{a\omega^2}{c^2}\cos\frac{\omega s}{c}, -\frac{a\omega^2}{c^2}\sin\frac{\omega s}{c}, 0),\\
\beta'''(s) &= (\frac{a\omega^3}{c^3}\sin\frac{\omega s}{c}, -\frac{a\omega^3}{c^3}\cos\frac{\omega s}{c}, 0),\\
\kappa^2 &= \beta''\cdot\beta'',\\
&= \frac{a^2\omega^4}{c^4},\\
\kappa &= \pm\frac{a\omega^2}{c^2}.\\
\tau &= \frac{(\beta'\beta''\beta''')}{\beta''\cdot\beta''},\\
&= \frac{b}{c}\begin{bmatrix} -\frac{a\omega^2}{c^2}\cos\frac{\omega s}{c} & -\frac{a\omega^2}{c^2}\sin\frac{\omega s}{c}, \\ \frac{a\omega^3}{c^2}\sin\frac{\omega s}{c} & -\frac{a\omega^3}{c^2}\cos\frac{\omega s}{c} \end{bmatrix}\frac{c^4}{a^2\omega^4},\\
&= \frac{b}{c}\frac{a^2\omega^5}{c^5}\frac{c^4}{a^2\omega^4}.
\end{aligned}
$$

Simplifying the last expression and substituting the value of c, we get

$$
\begin{aligned}
\tau &= \frac{b\omega}{a^2\omega^2+b^2},\\
\kappa &= \pm\frac{a\omega^2}{a^2\omega^2+b^2}.
\end{aligned}
$$

Notice that if $b=0$, the helix collapses to a circle in the xy-plane. In this case, the formulas above reduce to $\kappa = 1/a$ and $\tau = 0$. The ratio $\kappa/\tau = a\omega/b$ is particularly simple. Any curve for which $\kappa/\tau =$constant, is called a helix; the circular helix is a special case.

1.3.13 Example (Plane curves) Let $\alpha(t) = (x(t), y(t), 0)$. Then

$$\begin{aligned}
\alpha' &= (x', y', 0), \\
\alpha'' &= (x'', y'', 0), \\
\alpha''' &= (x''', y''', 0), \\
\kappa &= \frac{\|\alpha' \times \alpha''\|}{\|\alpha'\|^3}, \\
&= \frac{|\, x'y'' - y'x'' \,|}{(x'^2 + y'^2)^{3/2}}. \\
\tau &= 0.
\end{aligned}$$

1.3.14 Example Let $\beta(s) = (x(s), y(s), 0)$, where

$$\begin{aligned}
x(s) &= \int_0^s \cos\frac{t^2}{2c^2}\, dt, \\
y(s) &= \int_0^s \sin\frac{t^2}{2c^2}\, dt. \qquad (1.42)
\end{aligned}$$

Then, using the fundamental theorem of calculus, we have

$$\beta'(s) = (\cos\frac{s^2}{2c^2}, \sin\frac{s^2}{2c^2}, 0),$$

Since $\|\beta'\| = v = 1$, the curve is of unit speed, and s is indeed the arc length. The curvature is given by

$$\begin{aligned}
\kappa &= \|x'y'' - y'x''\| = (\beta' \cdot \beta')^{1/2}, \\
&= \| -\frac{s}{c^2}\sin\frac{s^2}{2c^2}, \frac{s}{c^2}\cos\frac{s^2}{2c^2}, 0\|, \\
&= \frac{s}{c^2}.
\end{aligned}$$

The functions (1.42) are the classical *Fresnel integrals* which we will discuss in more detail in the next section.

In cases where the given curve $\alpha(t)$ is not of unit speed, the following proposition provides formulas to compute the curvature and torsion in terms of α.

1.3.15 Proposition If $\alpha(t)$ is a regular curve in $\mathbf{R}^3$, then

$$\kappa^2 = \frac{\|\alpha' \times \alpha''\|^2}{\|\alpha'\|^6}, \qquad (1.43)$$

$$\tau = \frac{(\alpha'\alpha''\alpha''')}{\|\alpha' \times \alpha''\|^2}, \qquad (1.44)$$

where $(\alpha'\alpha''\alpha''')$ is the triple vector product $[\alpha' \times \alpha''] \cdot \alpha'''$.

Proof

$$\begin{aligned}
\alpha' &= vT, \\
\alpha'' &= v'T + v^2\kappa N, \\
\alpha''' &= (v^2\kappa)N' + \dots, \\
&= v^3\kappa N' + \dots, \\
&= v^3\kappa\tau B + \dots.
\end{aligned}$$

As the computation below shows, the other terms in α''' are unimportant here because $\alpha' \times \alpha''$ is proportional to B, so all we need is the B component to solve for τ.

$$\begin{aligned}
\alpha' \times \alpha'' &= v^3\kappa(T \times N) = v^3\kappa B, \\
\|\alpha' \times \alpha''\| &= v^3\kappa, \\
\kappa &= \frac{\|\alpha' \times \alpha''\|}{v^3}. \\
(\alpha' \times \alpha'') \cdot \alpha''' &= v^6\kappa^2\tau, \\
\tau &= \frac{(\alpha'\alpha''\alpha''')}{v^6\kappa^2}, \\
&= \frac{(\alpha'\alpha''\alpha''')}{\|\alpha' \times \alpha''\|^2}.
\end{aligned}$$

1.4 Fundamental Theorem of Curves

The fundamental theorem of curves basically states that prescribing a curvature and torsion as functions of some parameter s, completely determines up to position and orientation, a curve $\beta(s)$ with that given curvature and torsion. Some geometrical insight into the significance of the curvature and torsion can be gained by considering the Taylor series expansion of an arbitrary unit speed curve $\beta(s)$ about $s = 0$.

$$\beta(s) = \beta(0) + \beta'(0)s + \frac{\beta''(0)}{2!}s^2 + \frac{\beta'''(0)}{3!}s^3 + \dots \tag{1.45}$$

Since we are assuming that s is an arc length parameter,

$$\begin{aligned}
\beta'(0) &= T(0) = T_0 \\
\beta''(0) &= (\kappa N)(0) = \kappa_0 N_0 \\
\beta'''(0) &= (-\kappa^2 T + \kappa' N + \kappa\tau B)(0) = -\kappa_0^2 T_0 + \kappa_0' N_0 + \kappa_0\tau_0 B_0
\end{aligned}$$

Keeping only the lowest terms in the components of T, N, and B, we get the first order Frenet approximation to the curve

$$\beta(s) \doteq \beta(0) + T_0 s + \frac{1}{2}\kappa_0 N_0 s^2 + \frac{1}{6}\kappa_0\tau_0 B_0 s^3. \tag{1.46}$$

The first two terms represent the linear approximation to the curve. The first three terms approximate the curve by a parabola which lies in the osculating plane (TN-plane). If $\kappa_0 = 0$, then locally the curve looks like a straight line. If $\tau_0 = 0$, then locally the curve is a plane curve contained on the osculating plane. In this sense, the curvature measures the deviation of the curve from a straight line and the torsion (also called the second curvature) measures the deviation of the curve from a plane curve. As shown in figure 1.9 a non-planar space curve locally looks like a wire that has first been bent into a parabolic shape in the TN and twisted into a cubic along the B axis. So suppose that p

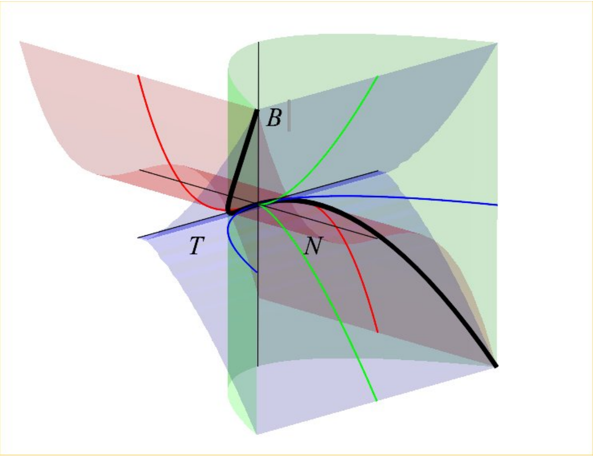


Fig. 1.9: Cubic Approximation to a Curve

is an arbitrary point on a curve $\beta(s)$ parametrized by arc length. We position the curve so that p is at the origin so that $\beta(0) = 0$ coincides with the point p. We chose the orthonormal basis vectors $\{e_1, e_2, e_3\}$ in $\mathbf{R}^3$ to coincide with the Frenet Frame T_0, N_0, B_0 at that point. then, the equation (1.46) provides a canonical representation of the curve near that point. This then constitutes a proof of the fundamental theorem of curves under the assumption the curve, curvature and torsion are analytic. (One could also treat the Frenet formulas as a system of differential equations and apply the conditions of existence and uniqueness of solutions for such systems.)

1.4.1 Proposition A curve with $\kappa = 0$ is part of a straight line.
If $\kappa = 0$ then $\beta(s) = \beta(0) + sT_0$.

1.4.2 Proposition A curve $\alpha(t)$ with $\tau = 0$ is a plane curve.
Proof If $\tau = 0$, then $(\alpha'\alpha''\alpha''') = 0$. This means that the three vectors α', α'', and α''' are linearly dependent and hence there exist functions $a_1(s)$,$a_2(s)$ and $a_3(s)$ such that

$$a_3\alpha''' + a_2\alpha'' + a_1\alpha' = 0.$$

This linear homogeneous equation will have a solution of the form

$$\alpha = \mathbf{c}_1\alpha_1 + \mathbf{c}_2\alpha_2 + \mathbf{c}_3, \quad c_i = \text{constant vectors.}$$

This curve lies in the plane

$$(\mathbf{x} - \mathbf{c}_3) \cdot \mathbf{n} = 0, \quad \text{where} \quad \mathbf{n} = \mathbf{c}_1 \times \mathbf{c}_2$$

A consequence of the Frenet Equations is that given two curves in space C and C^* such that $\kappa(s) = \kappa^*(s)$ and $\tau(s) = \tau * (s)$, the two curves are the same up to their position in space. To clarify what we mean by their "position" we need to review some basic concepts of linear algebra leading to the notion of isometries.

1.4.1 Isometries

1.4.3 Definition Let $\mathbf{x}$ and $\mathbf{y}$ be two column vectors in $\mathbf{R}^n$ and let $\mathbf{x}^T$ represent the transposed row vector. To keep track on whether a vector is a row vector or a column vector, hereafter we write the components $\{x^i\}$ of a column vector with the indices up and the components $\{x_i\}$ of a row vector with the indices down. Similarly, if A is an $n \times n$ matrix, we write its components as $A = (a^i{}_j)$. The standard *inner product* is given by matrix multiplication of the row and column vectors

$$\begin{aligned} < \mathbf{x}, \mathbf{y} > &= \mathbf{x}^T\mathbf{y}, && (1.47)\\ &=< \mathbf{y}, \mathbf{x} > . && (1.48)\end{aligned}$$

The inner product gives $\mathbf{R}^n$ the structure of a normed space by defining $\|\mathbf{x}\| =< \mathbf{x}, \mathbf{x} >^{1/2}$ and the structure of a metric space in which $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$. The real inner product is bilinear (linear in each slot), from which it follows that

$$\|\mathbf{x} \pm \mathbf{y}\|^2 = \|\mathbf{x}\|^2 \pm 2 < \mathbf{x}, \mathbf{y} > + \|\mathbf{y}\|^2. \tag{1.49}$$

Thus, we have the *polarization identity*

$$< \mathbf{x}, \mathbf{y} >= \tfrac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \tfrac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2. \tag{1.50}$$

The Euclidean inner product satisfies the relation

$$< \mathbf{x}, \mathbf{y} >= \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos\theta, \tag{1.51}$$

where θ is the angle subtended by the two vectors.

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are called *orthogonal* if $< \mathbf{x}, \mathbf{y} >= 0$, and a set of basis vectors $\mathscr{B} = \{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ is called an *orthonormal* basis if $< \mathbf{e}_i, \mathbf{e}_j >= \delta_{ij}$. Given an orthonormal basis, the dual basis is the set of linear functionals $\{\alpha^{\mathbf{i}}\}$ such that $\alpha^i(\mathbf{e}_j) = \delta^i_j$. In terms of basis components, column vectors are given by $\mathbf{x} = x^i \mathbf{e}_i$, row vectors by $\mathbf{x}^T = x_j \alpha^j$, and the inner product

$$
\begin{aligned}
< \mathbf{x}, \mathbf{y} > &= \mathbf{x}^T \mathbf{y}, \\
&= (x_i \alpha^i)(y^j \mathbf{e}_j), \\
&= (x_i y^j) \alpha^i(\mathbf{e}_j) = (x_i y^j) \delta^i_j. \\
&= x_i y^i, \\
&= \begin{bmatrix} x_1 & x_2 \ldots & x_n \end{bmatrix} \begin{bmatrix} y^1 \\ y^2 \\ \cdots \\ y^n \end{bmatrix}
\end{aligned}
$$

Since $|\cos\theta| \leq 1$, it follows from equation 1.51, a special case of the *Schwarz inequality*

$$| < \mathbf{x}, \mathbf{y} > | \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|. \tag{1.52}$$

Let F be a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$ and $\mathscr{B} = \{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ be an orthonormal basis. Then, there exists a matrix $A = [F]_{\mathscr{B}}$ given by

$$A = (a^i_j) = \alpha^i(F(\mathbf{e}_j)), \tag{1.53}$$

or in terms of the inner product,

$$A = (a_{ij}) =< \mathbf{e}_i, F(\mathbf{e}_j) > . \tag{1.54}$$

On the other hand, if A is a fixed $n \times n$ matrix, the map F defined by $F(\mathbf{x}) = A\mathbf{x}$ is a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$ whose matrix representation in the standard basis is the matrix A itself. It follows that given a linear transformation represented by a matrix A, we have

$$
\begin{aligned}
< \mathbf{x}, A\mathbf{y} > &= \mathbf{x}^T A \mathbf{y}, && (1.55) \\
&= (A^T \mathbf{x})^{\mathbf{T}} \mathbf{y}, \\
&=< A^T \mathbf{x}, \mathbf{y} > . && (1.56)
\end{aligned}
$$

1.4.4 Definition A real $n \times n$ matrix A is called *orthogonal* if $A^T A = AA^T = I$. The linear transformation represented by A is called an *orthogonal transformation.* Equivalently, the transformation represented by A is orthogonal if

$$< \mathbf{x}, A\mathbf{y} >=< A^{-1}\mathbf{x}, \mathbf{y} > . \tag{1.57}$$

Thus, real orthogonal transformations are represented by symmetric matrices (Hermitian in the complex case) and the condition $A^T A = I$ implies that $\det(A) = \pm 1$.

1.4.5 Theorem If A is an orthogonal matrix, then the transformation determined by A preserves the inner product and the norm.

Proof

$$\begin{aligned} < A\mathbf{x}, A\mathbf{y} > &=< A^T A\mathbf{x}, \mathbf{y} >, \\ &=< \mathbf{x}, \mathbf{y} > . \end{aligned}$$

Furthermore, setting $\mathbf{y} = \mathbf{x}$:

$$\begin{aligned} < A\mathbf{x}, A\mathbf{x} > &=< \mathbf{x}, \mathbf{x} >, \\ \|A\mathbf{x}\|^2 &= \|\mathbf{x}\|^2, \\ \|A\mathbf{x}\| &= \|\mathbf{x}\|. \end{aligned}$$

As a corollary, if $\{\mathbf{e}_i\}$ is an orthonormal basis, then so is $\{\mathbf{f}_i = A\mathbf{e}_i\}$. That is, an orthogonal transformation represents a rotation if $\det A = 1$ and a rotation with a reflection if $\det A = -1$.

1.4.6 Definition A mapping $F : \mathbf{R}^n \to \mathbf{R}^n$ called an *isometry* if it preserves distances. That is, if for all $\mathbf{x}, \mathbf{y}$

$$d(F(\mathbf{x}), F(\mathbf{y})) = d(\mathbf{x}, \mathbf{y}). \tag{1.58}$$

1.4.7 Example (Translations) Let $\mathbf{q}$ be fixed vector. The map $F(\mathbf{x}) = \mathbf{x} + \mathbf{q}$ is called a *translation.* It is clearly an isometry since $\|F(\mathbf{x}) - F(\mathbf{y})\| = \|\mathbf{x} + \mathbf{p} - (\mathbf{y} + \mathbf{p})\| = \|\mathbf{x} - \mathbf{y}\|$.

1.4.8 Theorem An orthogonal transformation is an isometry.

Proof Let F be an isometry represented by an orthogonal matrix A. Then, since the transformation is linear and preserves norms, we have:

$$\begin{aligned} d(F(\mathbf{x}), F(\mathbf{x})) &= \|A\mathbf{x} - A\mathbf{y}\|, \\ &= \|A(\mathbf{x} - \mathbf{y})\|, \\ &= \|\mathbf{x} - \mathbf{y}\| \end{aligned}$$

The composition of two isometries is also an isometry. The inverse of a translation by $\mathbf{q}$ is a translation by $-\mathbf{q}$. The inverse of an orthogonal transformation represented by A is an orthogonal transformation represented by A^{-1}. Consequently, the set of isometries consisting of translations and orthogonal transformations constitutes a group. Given a general isometry, we can use a translation to insure that $F(\mathbf{0}) = \mathbf{0}$. We now prove the following theorem.

1.4.9 Theorem If F is an isometry such that $F(\mathbf{0}) = \mathbf{0}$, then F is an orthogonal transformation.

Proof We need to prove that F preserves the inner product and that it is linear. We first show that F preserves norms. In fact

$$\begin{aligned}\|F(\mathbf{x})\| &= d(F(\mathbf{x}),\mathbf{0}),\\ &= d(F(\mathbf{x}),F(\mathbf{0}),\\ &= d(\mathbf{x},\mathbf{0}),\\ &= \|\mathbf{x}-\mathbf{0}\|,\\ &= \|\mathbf{x}\|.\end{aligned}$$

Now, using 1.49 and the norm preserving property above, we have:

$$\begin{aligned}d(F(\mathbf{x}),F(\mathbf{y})) &= d(\mathbf{x},\mathbf{y}),\\ \|F(\mathbf{x})-F(\mathbf{y})\|^2 &= \|\mathbf{x}-\mathbf{y}\|^2,\\ \|F(\mathbf{x})\|^2 - 2 < F(\mathbf{x}),F(\mathbf{y}) > + \|F(\mathbf{y})\|^2 &= \|\mathbf{x}\|^2 - 2 < \mathbf{x},\mathbf{y} > + \|\mathbf{y}\|^2.\\ < F(\mathbf{x}),F(\mathbf{y}) > &= < \mathbf{x},\mathbf{y} > .\end{aligned}$$

To show F is linear, let $\mathbf{e}_i$ be an orthonormal basis, which implies that $\mathbf{f}_i = F(\mathbf{e}_i)$ is also an orthonormal basis. Then

$$\begin{aligned}F(a\mathbf{x}+b\mathbf{y}) &= \sum_{i=1}^{n} < F(a\mathbf{x}+b\mathbf{y},\mathbf{f}_i > \mathbf{f}_i,\\ &= \sum_{i=1}^{n} < F(a\mathbf{x}+b\mathbf{y}),F(\mathbf{e}_i) > \mathbf{f}_i,\\ &= \sum_{i=1}^{n} < (a\mathbf{x}+b\mathbf{y}),\mathbf{e}_i > \mathbf{f}_i,\\ &= a\sum_{i=1}^{n} < \mathbf{x},\mathbf{e}_i > \mathbf{f}_i + b\sum_{i=1}^{n} < \mathbf{y},\mathbf{e}_i > \mathbf{f}_i,\\ &= a\sum_{i=1}^{n} < F(\mathbf{x}),\mathbf{f}_i > \mathbf{f}_i + b\sum_{i=1}^{n} < F(\mathbf{y}),\mathbf{f}_i > \mathbf{f}_i,\\ &= aF(\mathbf{x})+bF(\mathbf{y}).\end{aligned}$$

1.4.10 Theorem If $F:\mathbf{R}^n \to \mathbf{R}^n$ is an isometry then

$$F(\mathbf{x}) = A\mathbf{x}+\mathbf{q}, \tag{1.59}$$

where A is orthogonal.

Proof If $F(\mathbf{0} = \mathbf{q}$, then $\widetilde{F} = F - \mathbf{q}$ is an isometry with $\widetilde{F}(\mathbf{0}) = \mathbf{0}$ and hence by the previous theorem $\widetilde{F}$ is an orthogonal transformation represented by an orthogonal matrix $\widetilde{F}\mathbf{x} = A\mathbf{x}$. It follows that $F(\mathbf{x}) = A\mathbf{x}+\mathbf{q}$.

We have just shown that any isometry is the composition of translation and an orthogonal transformation. The latter is the linear part of the isometry. The orthogonal transformation preserves the inner product, lengths, and maps orthonormal bases to orthonormal bases.

1.4.11 Theorem If α is a curve in $\mathbf{R}^n$ and β is the image of α under a mapping F, then vectors tangent to α get mapped to tangent vectors to β.
Proof Let $\beta = F \circ \alpha$. The proof follows trivially from the properties of the Jacobian map $\beta_* = (F \circ \alpha)_* = F_* \alpha_*$ that takes tangent vectors to tangent vectors. If in addition F is an isometry, then F_* maps the Frenet frame of α to the Frenet frame of β.

We now have all the ingredients to prove the following:

1.4.12 Theorem (Fundamental theorem of curves) If C and $\tilde{C}$ are space curves such that $\kappa(s) = \tilde{\kappa}(s)$, and $\tau(s) = \tilde{\tau}(s)$ for all s, the curves are isometric.
Proof Given two such curves, we can perform a translation so that, for some $s = s_0$, the corresponding points on C and $\tilde{C}$ are made to coincide. Without loss of generality, we can make this point be the origin. Now we perform an orthogonal transformation to make the Frenet frame $\{T_0, N_0, B_0\}$ of C coincide with the Frenet frame $\{\tilde{T}_0, \tilde{N}_0, \tilde{B}_0\}$ of $\tilde{C}$. By Schwarz inequality, the inner product of two unit vectors is also a unit vector, if and only if the vectors are equal. With this in mind, let

$$L = T \cdot \tilde{T} + N \cdot \tilde{N} + B \cdot \tilde{B}.$$

A simple computation using the Frenet equations shows that $L' = 0$, so $L =$ constant. But at $s = 0$ the Frenet frames of the two curves coincide, so the constant is 3 and this can only happen if for all s, $T = \tilde{T}$, $N = \tilde{N}$, $B = \tilde{B}$. Finally, since $T = \tilde{T}$, we have $\beta'(s) = \tilde{\beta}'(s)$, so $\beta(s) = \tilde{\beta}(s) +$ constant. But since $\beta(0) = \tilde{\beta}(0)$, the constant is 0 and $\beta(s) = \tilde{\beta}(s)$ for all s.

1.4.2 Natural Equations

The fundamental theorem of curves states that up to an isometry, that is up to location and orientation, a curve is completely determined by the curvature and torsion. However, the formulas for computing κ and τ are sufficiently complicated that solving the Frenet system of differential equations could be a daunting task indeed. With the invention of modern computers, obtaining and plotting numerical solutions is a *routine* matter. There is a plethora of differential equations solvers available nowadays, including the solvers built-in into Maple, Mathematica, and Matlab. For plane curves, which are characterized

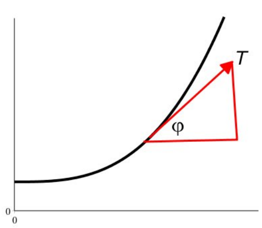


Fig. 1.10: Tangent

by $\tau = 0$, it is possible to find an integral formula for the curve coordinates in terms of the curvature. Given a curve parametrized by arc length, consider an arbitrary point with position vector $\mathbf{x} = (x, y)$ on the curve, and let φ be the angle that the tangent vector T makes with the horizontal, as shown in figure 1.10. Then, the Euclidean vector components of the unit tangent vector are given by

$$\frac{d\mathbf{x}}{ds} = \mathbf{T} = (\cos\varphi, \sin\varphi).$$

This means that

$$\frac{dx}{ds} = \cos\varphi, \quad \text{and} \quad \frac{dy}{ds} = \sin\varphi.$$

From the first Frenet equation we also have

$$\frac{d\mathbf{T}}{ds} = (-\sin\varphi\frac{d\varphi}{ds}, \cos\varphi\frac{d\varphi}{ds}) = \kappa\mathbf{N},$$

so that,

$$\left\|\frac{d\mathbf{T}}{ds}\right\| = \frac{d\varphi}{ds} = \kappa.$$

We conclude that

$$x(s) = \int \cos\varphi\, ds, \quad y(s) = \int \sin\varphi\, ds, \text{where,} \quad \varphi = \int \kappa\, ds. \tag{1.60}$$

Equations 1.60 are called the *natural equations* of a plane curve. Given the curvature κ, the equation of the curve can be obtained by "quadratures," the classical term for integrals.

1.4.13 Example Circle: $\kappa = 1/R$
The simplest natural equation is one where the curvature is constant. For obvious geometrical reasons we choose this constant to be $1/R$. Then, $\varphi = s/R$ and

$$\mathbf{x} = (R\sin\frac{s}{R}, -R\cos\frac{s}{R}),$$

which is the equation of a unit speed circle of radius R.

1.4.14 Example Cornu spiral: $\kappa = \pi s$
This is the most basic linear natural equation, except for the scaling factor of π which is inserted for historical conventions. Then $\varphi = \frac{1}{2}\pi s^2$, and

$$x(s) = C(s) = \int \cos(\tfrac{1}{2}\pi s^2)\, ds; \qquad y(s) = S(s) = \int \sin(\tfrac{1}{2}\pi s^2)\, ds. \tag{1.61}$$

The functions $C(s)$ and $S(s)$ are called *Fresnel Integrals.* In the standard classical function libraries of Maple and Mathematica, they are listed as $FresnelC$ and $FresnelS$ respectively. The fast-increasing frequency of oscillations of the integrands here make the computation prohibitive without the use of high-speed computers. Graphing calculators are inadequate to render the rapid oscillations for s ranging from 0 to 15, for example, and simple computer programs for the

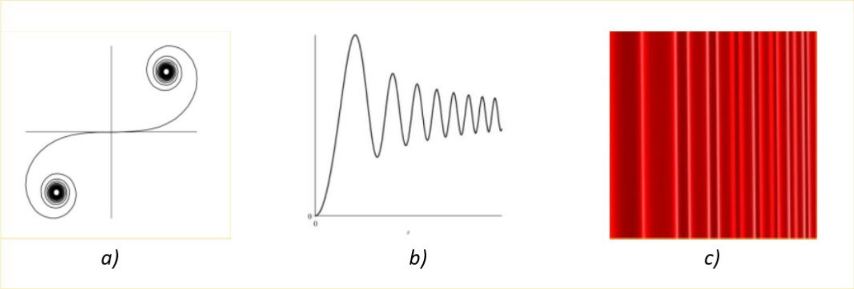


Fig. 1.11: Fresnel Diffraction

trapezoidal rule as taught in typical calculus courses, completely fall apart in this range. The *Cornu spiral* is the curve $\mathbf{x}(s) = (x(s), y(s))$ parametrized by Fresnel integrals (See figure 1.11a). It is a tribute to the mathematicians of the 1800's that not only were they able to compute the values of the Fresnel integrals to 4 or 5 decimal places, but they did it for the range of s from 0 to 15 as mentioned above, producing remarkably accurate renditions of the spiral. Fresnel integrals appear in the study of diffraction. If a coherent beam of light such as a laser beam, hits a sharp straight edge and a screen is placed behind, there will appear on the screen a pattern of diffraction fringes. The amplitude and intensity of the diffraction pattern can be obtained by a geometrical construction involving the Fresnel integrals. First consider the function $\Psi(s) = \|\mathbf{x}\|$ that measures the distance from the origin to the points in the Cornu spiral in the first quadrant. The square of this function is then proportional to the intensity of the diffraction pattern, The graph of $|\Psi(s)|^2$ is shown in figure 1.11b. Translating this curve along an axis coinciding with that of the straight edge, generates a three dimensional surface as shown from "above" in figure 1.11c. A color scheme was used here to depict a model of the Fresnel diffraction by the straight edge.

1.4.15 Example Logarithmic Spiral $\kappa = 1/(as + b)$
A logarithmic spiral is a curve in which the position vector $\mathbf{x}$ makes a constant angle with the tangent vector, as shown in figure 1.12. A formula for the curve can be found easily if one uses the calculus formula in polar coordinates

$$\tan\psi = \frac{r}{dr/d\theta}. \tag{1.62}$$

Here, ψ is the angle between the polar direction and the tangent. If ψ is constant, then one can immediately integrate the equation to get the exponential function below, in which k is the constant of integration

$$r(\theta) = ke^{(\cot\psi)\theta} \tag{1.63}$$

Derivation of formula 1.62 has fallen through the cracks in standard fat calculus textbooks, at best relegated to an advanced exercise which most students

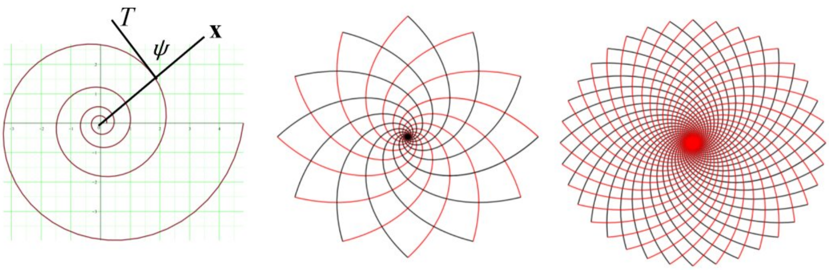


Fig. 1.12: Logarithmic Spiral

will not do. Perhaps the reason is that the section on polar coordinates is typically covered in Calculus II, so students have not yet been exposed to the tools of vector calculus that facilitate the otherwise messy computation. To fill-in this gap, we present a short derivation of this neat formula. For a plane curve in parametric polar coordinates, we have

$$\begin{aligned}\mathbf{x}(t) &= (r(t)\cos\theta(t), r(t)\sin\theta(t)),\\ \dot{\mathbf{x}} &= (\dot{r}\cos\theta - r\sin\theta\,\dot{\theta}, \dot{r}\sin\theta + r\cos\theta\,\dot{\theta}).\end{aligned}$$

A direct computation of the dot product gives,

$$|<\mathbf{x},\dot{\mathbf{x}}>|^2 = (r\dot{r})^2.$$

On the other hand,

$$\begin{aligned}|<\mathbf{x},\dot{\mathbf{x}}>|^2 &= \|\mathbf{x}\|^2\,\|\dot{\mathbf{x}}\|^2\ \cos^2\psi,\\ &= r^2(\dot{r}^2 + r^2\dot{\theta}^2)\cos^2\psi.\end{aligned}$$

Equating the two, we find,

$$\begin{aligned}\dot{r}^2 &= (\dot{r}^2 + r^2\dot{\theta}^2)\cos^2\psi,\\ (\sin^2\psi)\dot{r}^2 &= r^2\dot{\theta}^2\cos^2\psi,\\ (\sin\psi)\,dr &= r\cos\psi\ d\theta,\\ \tan\psi &= \frac{r}{dr/d\theta}.\end{aligned}$$

We leave it to the reader to do a direct computation of the curvature. Instead, we prove that if $\kappa = 1/(as+b)$, where a and b are constant, then the curve is a logarithmic spiral. From the natural equations, we have,

$$\begin{aligned}
\frac{d\theta}{ds} &= \kappa = \frac{1}{as+b},\\
\theta &= \frac{1}{a}\ln(as+b) + C, \quad C = \text{const},\\
e^{a\theta} &= B(as+b), \quad B = e^{aC} = 1/A,\\
\frac{1}{\kappa} &= Ae^{a\theta} = \frac{ds}{d\theta},\\
ds &= Ae^{as}\, d\theta.
\end{aligned}$$

Back to the natural equations, the x and y coordinates are obtained by integrating,

$$\begin{aligned}
x &= \int Ae^{a\theta}\cos\theta\, d\theta,\\
y &= \int Ae^{a\theta}\sin\theta\, d\theta.
\end{aligned}$$

We can avoid the integrations by parts by letting $z = x + iy = re^{i\theta}$. We get

$$\begin{aligned}
z &= A\int e^{a\theta}e^{i\theta}\, d\theta,\\
&= A\int e^{(a+i)\theta}\, d\theta,\\
&= \frac{A}{a+i}e^{(a+i)\theta},\\
&= \frac{A}{a+i}e^{a\theta}e^{i\theta}.
\end{aligned}$$

Extracting the real part $\|z\| = r$, we get

$$r = \frac{A}{\sqrt{a^2+1}}e^{a\theta}, \tag{1.64}$$

which is the equation of a logarithmic spiral with $a = \cot\psi$. As shown in figure 1.12, families of concentric logarithmic spirals are ubiquitous in nature as in flowers and pine cones, in architectural designs. The projection of a conical helix as in figure 4.8 onto the plane through the origin, is a logarithmic spiral. The analog of a logarithmic spiral on a sphere is called a loxodrome as depicted in figure 4.2.

1.4.16 Example Meandering Curves: $\kappa = \sin s$

A whole family of meandering curves are obtained by letting $\kappa = A\sin ks$. The meandering graph shown in picture 1.13 was obtained by numerical integration for $A = 2$ and "wave number" $k = 1$. The larger the value of A the larger the curvature of the "throats." If A is large enough, the "throats" will overlap.

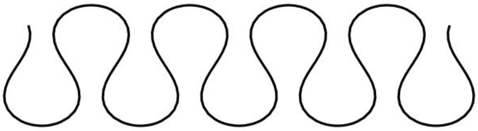


Fig. 1.13: Meandering Curve

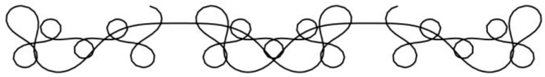


Fig. 1.14: Bimodal Meander

Using superpositions of sine functions gives rise to a beautiful family of "multi-frequency" meanders with graphs that would challenge the most skillful calligraphists of the 1800's. Figure 1.14 shows a rendition with two sine functions with equal amplitude $A = 1.8$, and with $k_1 = 1,\ k_2 = 1.2$.

Chapter 2

Differential Forms

2.1 One-Forms

The concept of the differential of a function is one of the most puzzling ideas in elementary calculus. In the usual definition, the differential of a dependent variable $y = f(x)$ is given in terms of the differential of the independent variable by $dy = f'(x)dx$. The problem is with the quantity dx. What does "dx" mean? What is the difference between Δx and dx? How much "smaller" than Δx does dx have to be? There is no trivial resolution to this question. Most introductory calculus texts evade the issue by treating dx as an arbitrarily small quantity (lacking mathematical rigor) or by simply referring to dx as an infinitesimal (a term introduced by Newton for an idea that could not otherwise be clearly defined at the time.)

In this section we introduce linear algebraic tools that will allow us to interpret the differential in terms of a linear operator.

2.1.1 Definition Let $\mathbf{p} \in \mathbf{R}^n$, and let $T_p(\mathbf{R}^n)$ be the tangent space at $\mathbf{p}$. A *1-form* at $\mathbf{p}$ is a linear map ϕ from $T_p(\mathbf{R}^n)$ into $\mathbf{R}$, in other words, a linear functional. We recall that such a map must satisfy the following properties:

$$\begin{aligned} &\text{a)} \qquad \phi(X_p) \in \mathbf{R}, \qquad && \forall X_p \in \mathbf{R}^n \\ &\text{b)} \qquad \phi(aX_p + bY_p) = a\phi(X_p) + b\phi(Y_p), \qquad && \forall a, b \in \mathbf{R},\ X_p, Y_p \in T_p(\mathbf{R}^n) \end{aligned} \tag{2.1}$$

A *1-form* is a smooth assignment of a linear map ϕ as above for each point in the space.

2.1.2 Definition Let $f : \mathbf{R}^n \to \mathbf{R}$ be a real-valued C^∞ function. We define the differential df of the function as the 1-form such that

$$df(X) = X(f), \tag{2.2}$$

for every vector field in X in $\mathbf{R}^n$. In other words, at any point $\mathbf{p}$, the differential df of a function is an operator that assigns to a tangent vector X_p the directional derivative of the function in the direction of that vector.

$$df(X)(p) = X_p(f) = \nabla f(p) \cdot \mathbf{X}(p). \tag{2.3}$$

In particular, if we apply the differential of the coordinate functions x^i to the basis vector fields, we get

$$dx^i(\frac{\partial}{\partial x^j}) = \frac{\partial x^i}{\partial x^j} = \delta^i_j. \tag{2.4}$$

The set of all linear functionals on a vector space is called the *dual* of the vector space. It is a standard theorem in linear algebra that the dual of a finite dimensional vector space is also a vector space of the same dimension. Thus, the space $T^\star_p(\mathbf{R}^n)$ of all 1-forms at $\mathbf{p}$ is a vector space which is the dual of the tangent space $T_p(\mathbf{R}^n)$. The space $T^\star_p(\mathbf{R}^n)$ is called the *cotangent space* of $\mathbf{R}^n$ at the point $\mathbf{p}$. Equation (2.4) indicates that the set of differential forms $\{(dx^1)_p, \ldots, (dx^n)_p\}$ constitutes the basis of the cotangent space which is dual to the standard basis $\{(\frac{\partial}{\partial x^1})_p, \ldots (\frac{\partial}{\partial x^n})_p\}$ of the tangent space. The union of all the cotangent spaces as $\mathbf{p}$ ranges over all points in $\mathbf{R}^n$ is called the *cotangent bundle* $T^*(\mathbf{R}^n)$.

2.1.3 Proposition Let f be a smooth function in $\mathbf{R}^n$ and let $\{x^1, \ldots x^n\}$ be coordinate functions in a neighborhood U of a point $\mathbf{p}$. Then, the differential df is given locally by the expression

$$\begin{aligned} df &= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i \qquad (2.5) \\ &= \frac{\partial f}{\partial x^i} dx^i \end{aligned}$$

Proof The differential df is by definition a 1-form, so, at each point, it must be expressible as a linear combination of the basis $\{(dx^1)_p, \ldots, (dx^n)_p\}$. Therefore, to prove the proposition, it suffices to show that the expression 2.5 applied to an arbitrary tangent vector coincides with definition 2.2. To see this, consider a tangent vector $X_p = v^j(\frac{\partial}{\partial x^j})_p$ and apply the expression above as follows:

$$\begin{aligned} (\frac{\partial f}{\partial x^i} dx^i)_p(X_p) &= (\frac{\partial f}{\partial x^i} dx^i)(v^j \frac{\partial}{\partial x^j})(p) \qquad (2.6) \\ &= v^j(\frac{\partial f}{\partial x^i} dx^i)(\frac{\partial}{\partial x^j})(p) \\ &= v^j(\frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial x^j})(p) \\ &= v^j(\frac{\partial f}{\partial x^i}\delta^i_j)(p) \\ &= (\frac{\partial f}{\partial x^i} v^i)(p) \\ &= \nabla f(p) \cdot \mathbf{x}(p) \\ &= df(X)(p) \end{aligned}$$

The definition of differentials as linear functionals on the space of vector fields is much more satisfactory than the notion of infinitesimals, since the new definition is based on the rigorous machinery of linear algebra. If α is an arbitrary 1-form, then locally

$$\alpha = a_1(\mathbf{x})dx^1+, \ldots + a_n(\mathbf{x})dx^n, \tag{2.7}$$

where the coefficients a_i are C^∞ functions. Thus, a 1-form is a smooth section of the cotangent bundle and we refer to it as a *covariant tensor* of rank 1, or simply a *covector*. The collection of all 1-forms is denoted by $\Omega^1(\mathbf{R}^n) = \mathscr{T}_1^{\,0}(\mathbf{R}^n)$. The coefficients $(a_1, \ldots, a_n)$ are called the *covariant* components of the covector. We will adopt the convention to always write the covariant components of a covector with the indices down. Physicists often refer to the covariant components of a 1-form as a covariant vector and this causes some confusion about the position of the indices. We emphasize that not all one forms are obtained by taking the differential of a function. If there exists a function f, such that $\alpha = df$, then the one form α is called *exact*. In vector calculus and elementary physics, exact forms are important in understanding the path independence of line integrals of conservative vector fields.

As we have already noted, the cotangent space $T_p^*(\mathbf{R}^n)$ of 1-forms at a point $\mathbf{p}$ has a natural vector space structure. We can easily extend the operations of addition and scalar multiplication to the space of all 1-forms by defining

$$\begin{aligned} (\alpha + \beta)(X) &= \alpha(X) + \beta(X) \\ (f\alpha)(X) &= f\alpha(X) \end{aligned} \tag{2.8}$$

for all vector fields X and all smooth functions f.

2.2 Tensors

As we mentioned at the beginning of this chapter, the notion of the differential dx is not made precise in elementary treatments of calculus, so consequently, the differential of area $dxdy$ in $\mathbf{R}^2$, as well as the differential of surface area in $\mathbf{R}^3$ also need to be revisited in a more rigorous setting. For this purpose, we introduce a new type of multiplication between forms that not only captures the essence of differentials of area and volume, but also provides a rich algebraic and geometric structure generalizing cross products (which make sense only in $\mathbf{R}^3$) to Euclidean space of any dimension.

2.2.1 Definition A map $\phi : \mathscr{X}(\mathbf{R}^n) \times \mathscr{X}(\mathbf{R}^n) \longrightarrow \mathbf{R}$ is called a *bilinear* map of vector fields, if it is linear on each slot. That is, $\forall X_i, Y_i \in \mathscr{X}(\mathbf{R}^n), \;\; f^i \in \mathscr{F}(\mathbf{R}^n)$, we have

$$\begin{aligned} \phi(f^1X_1 + f^2X_2, Y_1) &= f^1\phi(X_1, Y_1) + f^2\phi(X_2, Y_1) \\ \phi(X_1, f^1Y_1 + f^2Y_2) &= f^1\phi(X_1, Y_1) + f^2\phi(X_1, Y_2). \end{aligned}$$

2.2.1 Tensor Products

2.2.2 Definition Let α and β be 1-forms. The *tensor product* of α and β is defined as the bilinear map $\alpha \otimes \beta$ such that

$$(\alpha \otimes \beta)(X, Y) = \alpha(X)\beta(Y) \tag{2.9}$$

for all vector fields X and Y.
Thus, for example, if $\alpha = a_i dx^i$ and $\beta = b_j dx^j$, then,

$$\begin{aligned}
(\alpha \otimes \beta)(\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l}) &= \alpha(\frac{\partial}{\partial x^k})\beta(\frac{\partial}{\partial x^l}) \\
&= (a_i dx^i)(\frac{\partial}{\partial x^k})(b_j dx^j)(\frac{\partial}{\partial x^l}) \\
&= a_i \delta^i_k b_j \delta^j_l \\
&= a_k b_l.
\end{aligned}$$

A quantity of the form $T = T_{ij} dx^i \otimes dx^j$ is called a *covariant tensor of rank 2,* and we may think of the set $\{dx^i \otimes dx^j\}$ as a basis for all such tensors. The space of covariant tensor fields of rank 2 is denoted $\mathscr{T}_2^0(\mathbf{R}^n)$. We must caution the reader again that there is possible confusion about the location of the indices, since physicists often refer to the components T_{ij} as a covariant tensor of rank two, as long is it satisfies some transformation laws.

In a similar fashion, one can define the tensor product of vectors X and Y as the bilinear map $X \otimes Y$ such that

$$(X \otimes Y)(f, g) = X(f)Y(g) \tag{2.10}$$

for any pair of arbitrary functions f and g.

If $X = a^i \frac{\partial}{\partial x^i}$ and $Y = b^j \frac{\partial}{\partial x^j}$, then the components of $X \otimes Y$ in the basis $\frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}$ are simply given by $a^i b^j$. Any bilinear map of the form

$$T = T^{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \tag{2.11}$$

is called a contravariant tensor of rank 2 in $\mathbf{R}^n$. The notion of tensor products can easily be generalized to higher rank, and in fact one can have tensors of mixed ranks. For example, a tensor of contravariant rank 2 and covariant rank 1 in $\mathbf{R}^n$ is represented in local coordinates by an expression of the form

$$T = T^{ij}{}_k \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \otimes dx^k.$$

This object is also called a tensor of type $\binom{2}{1}$. Thus, we may think of a tensor of type $\binom{2}{1}$ as a map with three input slots. The map expects two functions in the first two slots and a vector in the third one. The action of the map is bilinear on the two functions and linear on the vector. The output is a real number.

A tensor of type $\binom{r}{s}$ is written in local coordinates as

$$T = T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \ldots dx^{j_s} \tag{2.12}$$

The tensor components are given by

$$T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s} = T(dx^{i_1}, \ldots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \ldots, \frac{\partial}{\partial x^{j_s}}). \tag{2.13}$$

The set $T^r_s|_p(\mathbf{R}^n)$ of all tensors of type T^r_s at a point p has a vector space structure. The union of all such vector spaces is called the *tensor bundle*, and smooth sections of the bundle are called *tensor fields* $\mathscr{T}^r_s(\mathbf{R}^n)$; that is, a tensor field is a smooth assignment of a tensor to each point in $\mathbf{R}^n$.

2.2.2 Inner Product

Let $X = a^i \frac{\partial}{\partial x^i}$ and $Y = b^j \frac{\partial}{\partial x^j}$ be two vector fields and let

$$g(X, Y) = \delta_{ij} a^i b^j. \tag{2.14}$$

The quantity $g(X, Y)$ is an example of a bilinear map that the reader will recognize as the usual dot product.

2.2.3 Definition A bilinear map $g(X, Y) \equiv < X, Y >$ on vectors is called a real *inner product* if

1. $g(X, Y) = g(Y, X)$,
2. $g(X, X) \geq 0, \ \ \forall X$,
3. $g(X, X) = 0$ iff $X = 0$.

Since we assume $g(X, Y)$ to be bilinear, an inner product is completely specified by its action on ordered pairs of basis vectors. The components g_{ij} of the inner product are thus given by

$$g(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) = g_{ij}, \tag{2.15}$$

where g_{ij} is a symmetric $n \times n$ matrix which we assume to be non-singular. By linearity, it is easy to see that if $X = a^i \frac{\partial}{\partial x^i}$ and $Y = b^j \frac{\partial}{\partial x^j}$ are two arbitrary vectors, then

$$< X, Y >= g(X, Y) = g_{ij} a^i b^j.$$

In this sense, an inner product can be viewed as a generalization of the dot product. The standard Euclidean inner product is obtained if we take $g_{ij} = \delta_{ij}$. In this case, the quantity $g(X, X) = \| X \|^2$ gives the square of the length of the vector. For this reason, g_{ij} is called a *metric* and g is called a *metric tensor*.

Another interpretation of the dot product can be seen if instead one considers a vector $X = a^i \frac{\partial}{\partial x^i}$ and a 1-form $\alpha = b_j dx^j$. The action of the 1-form on the vector gives

$$\begin{aligned}
\alpha(X) &= (b_j dx^j)(a^i \tfrac{\partial}{\partial x^i}) \\
&= b_j a^i (dx^j)(\tfrac{\partial}{\partial x^i}) \\
&= b_j a^i \delta^j_i \\
&= a^i b_i.
\end{aligned}$$

If we now define

$$b_i = g_{ij} b^j, \tag{2.16}$$

we see that the equation above can be rewritten as

$$a^i b_i = g_{ij} a^i b^j,$$

and we recover the expression for the inner product.

Equation (2.16) shows that the metric can be used as a mechanism to lower indices, thus transforming the contravariant components of a vector to covariant ones. If we let g^{ij} be the inverse of the matrix g_{ij}, that is

$$g^{ik} g_{kj} = \delta^i_j, \tag{2.17}$$

we can also raise covariant indices by the equation

$$b^i = g^{ij} b_j. \tag{2.18}$$

We have mentioned that the tangent and cotangent spaces of Euclidean space at a particular point p are isomorphic. In view of the above discussion, we see that the metric g can be interpreted on one hand as a bilinear pairing of two vectors

$$g : T_p(\mathbf{R}^n) \times T_p(\mathbf{R}^n) \longrightarrow \mathbf{R},$$

and on the other, as inducing a linear isomorphism

$$G_\flat : T_p(\mathbf{R}^n) \longrightarrow T^\star_p(\mathbf{R}^n)$$

defined by

$$G_\flat X(Y) = g(X, Y), \tag{2.19}$$

that maps vectors to covectors. To verify this definition is consistent with the action of lowering indices, let $X = a^i \frac{\partial}{\partial x^i}$ and $Y = b^j \frac{\partial}{\partial x^j}$. We show that that $G_\flat X = a_i \, dx^i$. In fact,

$$\begin{aligned}
G_\flat X(Y) &= (a_i \, dx^i)(b^j \tfrac{\partial}{\partial x^j}), \\
&= a_i b^j dx^i (\tfrac{\partial}{\partial x^j}), \\
&= a_i b^j \delta^i{}_j, \\
&= a_i b^i = g_{ij} a^j b^i, \\
&= g(X, Y).
\end{aligned}$$

The inverse map $G^\sharp : T_p^\star(\mathbf{R}^n) \longrightarrow T_p(\mathbf{R}^n)$ is defined by

$$< G^\sharp \alpha, X >= \alpha(X), \tag{2.20}$$

for any 1-form α and tangent vector X. In quantum mechanics, it is common to use Dirac's notation, in which a linear functional α on a vector space $\mathcal{V}$ is called a *bra-vector* denoted by $\langle\alpha|$, and a vector $X \in \mathcal{V}$ is called a *ket-vector*, denoted by $|X\rangle$. The, action of a bra-vector on a ket-vector is defined by the *bracket*,

$$\langle\alpha|X\rangle = \alpha(X). \tag{2.21}$$

Thus, if the vector space has an inner product as above, we have

$$\langle\alpha|X\rangle =< G^\sharp \alpha, X >= \alpha(X). \tag{2.22}$$

The mapping $C : T_p^*(\mathbf{R}^n) \to \mathbf{R}$ given by $(\alpha, X) \mapsto \langle\alpha|X\rangle = \alpha(X)$ is called a *contraction*. In passing, we introduce a related concept called the *interior product*, or contraction of a vector and a form. If α is a $(k+1)$-form and X a vector, we define

$$i_X \alpha(X_1, \ldots, X_k) = \alpha(X, X_1, \ldots, X_k). \tag{2.23}$$

In particular, for a one form, we have

$$i_X \alpha = \langle\alpha|X\rangle = \alpha(X).$$

If T is a type $\binom{1}{1}$ tensor, that is,

$$T = T^i{}_j dx^j \otimes \frac{\partial}{\partial x^i},$$

The contraction of the tensor is given by

$$\begin{aligned} C(T) &= T^i{}_j \langle dx^j | \tfrac{\partial}{\partial x^i}\rangle, \\ &= T^i{}_j \; dx^j(\tfrac{\partial}{\partial x^i}), \\ &= T^i{}_j \; \delta^j_i, \\ &= T^i{}_i. \end{aligned}$$

In other words, the contraction of the tensor is the trace of the $n \times n$ array that represents the tensor in the given basis. The notion of raising and lowering indices as well as contractions can be extended to tensors of all types. Thus, for example, we have

$$g^{ij} T_{iklm} = T^i{}_{klm}.$$

A contraction between the indices i and l in the tensor above could be denoted by the notation

$$C^1_2(T^i{}_{klm}) = T^i{}_{kim} = T_{km}.$$

This is a very simple concept, but the notation for a general contraction is a bit awkward because one needs to keep track of the positions of the indices contracted. Let T be a tensor of type $\binom{r}{s}$. A contraction C^k_l yields a tensor of type $\binom{r-1}{s-1}$. Let T be given in the form 2.12. Then,

$$C^l_k(T) = T^{i_1\dots i_{l-1},m,i_{i+1}\dots i_r}_{j_1\dots j_{k-1},m,j_{k+1}\dots j_s}\frac{\partial}{\partial x^{i_1}}\otimes\dots\otimes\widehat{\frac{\partial}{\partial x^{i_l}}}\otimes\dots\otimes\frac{\partial}{\partial x^{i_r}}\otimes dx^{j_1}\otimes\dots\otimes\widehat{dx^{j_k}}\otimes\dots dx^{j_s},$$

where the "hat" means that these are excluded. Here is a very neat and most useful result. If S is a 2-tensor with symmetric components $T_{ij} = T_{ji}$ and A is a 2-tensor with antisymmetric components $A^{ij} = -A^{ji}$, then the contraction

$$S_{ij}A^{ij} = 0 \tag{2.24}$$

The short proof uses the fact that summation indices are dummy indices and they can be relabeled at will by any other index that is not already used in an expression. We have

$$S_{ij}A^{ij} = S_{ji}A^{ij} = -S_{ji}A^{ji} = -S_{kl}A^{kl} = -S_{ji}A^{ij} = 0,$$

since the quantity is the negative of itself.

In terms of the vector space isomorphism between the tangent and cotangent space induced by the metric, the gradient of a function f, viewed as a differential geometry vector field, is given by

$$\text{Grad}\, f = G^\sharp df, \tag{2.25}$$

or in components

$$(\nabla f)^i \equiv \nabla^i f = g^{ij} f_{,j}\,, \tag{2.26}$$

where $f_{,j}$ is the commonly used abbreviation for the partial derivative with respect to x^j.

In elementary treatments of calculus, authors often ignore the subtleties of differential 1-forms and tensor products and define the differential of arc length as

$$ds^2 \equiv g_{ij}dx^i dx^j,$$

although what is really meant by such an expression is

$$ds^2 \equiv g_{ij}dx^i \otimes dx^j. \tag{2.27}$$

2.2.4 Example In cylindrical coordinates, the differential of arc length is

$$ds^2 = dr^2 + r^2 d\theta^2 + dz^2. \tag{2.28}$$

In this case, the metric tensor has components

$$g_{ij} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.29}$$

2.2.5 Example In spherical coordinates,

$$\begin{aligned} x &= r\sin\theta\cos\phi, \\ y &= r\sin\theta\sin\phi, \\ z &= r\cos\theta, \end{aligned} \tag{2.30}$$

and the differential of arc length is given by

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2. \tag{2.31}$$

In this case the metric tensor has components

$$g_{ij} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin\theta^2 \end{bmatrix}. \tag{2.32}$$

2.2.3 Minkowski Space

An important object in mathematical physics is the so-called *Minkowski space* which is defined as the pair $(M_{(1,3)}, \eta)$, where

$$M_{(1,3)} = \{(t, x^1, x^2, x^3) |\ t, x^i \in \mathbf{R}\} \tag{2.33}$$

and η is the bilinear map such that

$$\eta(X, X) = t^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \tag{2.34}$$

The matrix representing Minkowski's metric η is given by

$$\eta = \mathrm{diag}(1, -1, -1, -1),$$

in which case, the differential of arc length is given by

$$\begin{aligned} ds^2 &= \eta_{\mu\nu} dx^\mu \otimes dx^\nu \\ &= dt \otimes dt - dx^1 \otimes dx^1 - dx^2 \otimes dx^2 - dx^3 \otimes dx^3 \\ &= dt^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \end{aligned} \tag{2.35}$$

Note: Technically speaking, Minkowski's metric is not really a metric since $\eta(X, X) = 0$ does not imply that $X = 0$. Non-zero vectors with zero length are called *light-like* vectors and they are associated with particles that travel at the speed of light (which we have set equal to 1 in our system of units.)

The Minkowski metric $\eta_{\mu\nu}$ and its matrix inverse $\eta^{\mu\nu}$ are also used to raise and lower indices in the space in a manner completely analogous to $\mathbf{R}^n$. Thus, for example, if A is a covariant vector with components

$$A_\mu = (\rho, A_1, A_2, A_3),$$

then the contravariant components of A are

$$\begin{aligned} A^\mu &= \eta^{\mu\nu} A_\nu \\ &= (\rho, -A_1, -A_2, -A_3). \end{aligned}$$

2.2.4 Wedge Products and 2-Forms

2.2.6 Definition A map $\phi : T(\mathbf{R}^n) \times T(\mathbf{R}^n) \longrightarrow \mathbf{R}$ is called *alternating* if

$$\phi(X, Y) = -\phi(Y, X).$$

The alternating property is reminiscent of determinants of square matrices that change sign if any two column vectors are switched. In fact, the determinant function is a model of an alternating bilinear map on the space $M_{2\times2}$ of two by two matrices. Of course, for the definition above to apply, one has to view $M_{2\times2}$ as the space of column vectors.

2.2.7 Definition A *2-form* ϕ is a map $\phi : T(\mathbf{R}^n) \times T(\mathbf{R}^n) \longrightarrow \mathbf{R}$ which is alternating and bilinear.

2.2.8 Definition Let α and β be 1-forms in $\mathbf{R}^n$ and let X and Y be any two vector fields. The *wedge product* of the two 1-forms is the map $\alpha \wedge \beta : T(\mathbf{R}^n) \times T(\mathbf{R}^n) \longrightarrow \mathbf{R}$, given by the equation

$$\begin{aligned}(\alpha \wedge \beta)(X, Y) &= \alpha(X)\beta(Y) - \alpha(Y)\beta(X),\\ &= \begin{bmatrix} \alpha(X) & \alpha(Y) \\ \beta(X) & \beta(Y) \end{bmatrix}\end{aligned} \tag{2.36}$$

2.2.9 Theorem If α and β are 1-forms, then $\alpha \wedge \beta$ is a 2-form.
Proof Let α and β be 1-forms in $\mathbf{R}^n$ and let X and Y be any two vector fields. Then

$$\begin{aligned}(\alpha \wedge \beta)(X, Y) &= \alpha(X)\beta(Y) - \alpha(Y)\beta(X)\\ &= -(\alpha(Y)\beta(X) - \alpha(X)\beta(Y))\\ &= -(\alpha \wedge \beta)(Y, X).\end{aligned}$$

Thus, the wedge product of two 1-forms is alternating.

To show that the wedge product of two 1-forms is bilinear, consider 1-forms, α, β, vector fields X_1, X_2, Y and functions f^1, f^2. Then, since the 1-forms are linear functionals, we get

$$\begin{aligned}(\alpha \wedge \beta)(f^1X_1 + f^2X_2, Y) &= \alpha(f^1X_1 + f^2X_2)\beta(Y) - \alpha(Y)\beta(f^1X_1 + f^2X_2)\\ &= [f^1\alpha(X_1) + f^2\alpha(X_2)]\beta(Y) - \alpha(Y)[f^1\beta(X_1) + f^2\beta(X_2)]\\ &= f^1\alpha(X_1)\beta(Y) + f^2\alpha(X_2)\beta(Y) - f^1\alpha(Y)\beta(X_1) - f^2\alpha(Y)\beta(X_2)\\ &= f^1[\alpha(X_1)\beta(Y) - \alpha(Y)\beta(X_1)] + f^2[\alpha(X_2)\beta(Y) - \alpha(Y)\beta(X_2)]\\ &= f^1(\alpha \wedge \beta)(X_1, Y) + f^2(\alpha \wedge \beta)(X_2, Y).\end{aligned}$$

The proof of linearity on the second slot is quite similar and is left to the reader.

The wedge product of two 1-forms has characteristics similar to cross products of vectors in the sense that both of these products anti-commute. This means that we need to be careful to introduce a minus sign every time we interchange the order of the operation. Thus, for example, we have

$$dx^i \wedge dx^j = -dx^j \wedge dx^i$$

if $i \neq j$, whereas

$$dx^i \wedge dx^i = -dx^i \wedge dx^i = 0$$

since any quantity that equals the negative of itself must vanish.

2.2.10 Example Consider the case of $\mathbf{R}^2$. Let

$$\begin{aligned} \alpha &= a\,dx + b\,dy, \\ \beta &= c\,dx + d\,dy. \end{aligned}$$

since $dx \wedge dx = dy \wedge dy = 0$, and $dx \wedge dy = -dy \wedge dx$, we get,

$$\begin{aligned} \alpha \wedge \beta &= ad\,dx \wedge dy + bc\,dy \wedge dx, \\ &= ad\,dx \wedge dy - bc\,dx \wedge dy, \\ &= \begin{vmatrix} a & b \\ c & d \end{vmatrix}\, dx \wedge dy. \end{aligned}$$

The similarity between wedge products is even more striking in the next example, but we emphasize again that wedge products are much more powerful than cross products, because wedge products can be computed in any dimension.

2.2.11 Example For combinatoric reasons, it is convenient to label the coordinates as $\{x^1, x^2, x^3\}$. Let

$$\begin{aligned} \alpha &= a_1\,dx^1 + a_2\,dx^2 + a_3\,dx^3, \\ \beta &= b_1\,dx^1 + b_2\,dx^2 + b_3\,dx^3, \end{aligned}$$

There are only three independent basis 2-forms, namely

$$\begin{aligned} dy \wedge dz &= dx^2 \wedge dx^3, \\ dx \wedge dz &= -dx^1 \wedge dx^3, \\ dx \wedge dy &= dx^1 \wedge dx^2. \end{aligned}$$

Computing the wedge products in pairs, we get

$$\alpha \wedge \beta = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}\, dx^2 \wedge dx^3 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}\, dx^1 \wedge dx^3 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\, dx^1 \wedge dx^2.$$

If we consider vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, we see that the result above can be written as

$$\alpha \wedge \beta = (\mathbf{a} \times \mathbf{b})_1\, dx^2 \wedge dx^3 - (\mathbf{a} \times \mathbf{b})_2\, dx^1 \wedge dx^3 + (\mathbf{a} \times \mathbf{b})_3\, dx^1 \wedge dx^2 \quad (2.37)$$

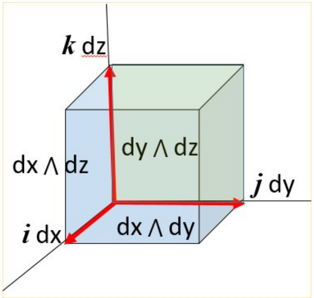


Fig. 2.1: Area Forms

It is worthwhile noticing that if one thinks of the indices in the formula above as permutations of the integers $\{1, 2, 3\}$, the signs of the three terms correspond to the signature of the permutation. In particular, the middle term indices constitute an odd permutation, so the signature is minus one. One can get a good sense of the geometrical significance and the motivation for the creation of wedge products by considering a classical analogy in the language of vector calculus. As shown in figure 2.1, let us consider infinitesimal arc length vectors $\mathbf{i}\ dx$, $\mathbf{j}\ dy$ and $\mathbf{k}\ dz$ pointing along the coordinate axes. Recall from the definition, that the cross product of two vectors is a new vector whose magnitude is the area of the parallelogram subtended by the two vectors and which points in the direction of a unit vector perpendicular to the plane containing the two vectors, oriented according to the right hand rule. Since $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are mutually orthogonal vectors, the cross product of any pair is again a unit vector pointed in the direction of the third or the negative thereof. Thus, for example, in the xy-plane the differential of area is really an oriented quantity that can computed by the cross product $(\mathbf{i}\ dx \times \mathbf{j}\ dy) = dx\ dy\ \mathbf{k}$. A similar computation yields the differential of areas in the other two coordinate planes, except that in the xz-plane, the cross product needs to be taken in the reverse order. In terms of wedge products, the differential of area in the xy-plane is $(dx \wedge dy)$, so that the oriented nature of the surface element is built-in. Technically, when reversing the order of variables in a double integral one should introduce a minus sign. This is typically ignored in basic calculus computations of double and triple integrals, but it cannot be ignored in vector calculus in the context of flux of a vector field through a surface.

2.2.12 Example One could of course compute wedge products by just using the linearity properties. It would not be as efficient as grouping into pairs, but it would yield the same result. For example, let
$\alpha = x^2 dx - y^2 dy$ and $\beta = dx + dy - 2xydz$. Then,

$$
\begin{aligned}
\alpha \wedge \beta &= (x^2 dx - y^2 dy) \wedge (dx + dy - 2xydz) \\
&= x^2\, dx \wedge dx + x^2\, dx \wedge dy - 2x^3 y\, dx \wedge dz - y^2\, dy \wedge dx \\
&\quad -y^2\, dy \wedge dy + 2xy^3\, dy \wedge dz \\
&= x^2\, dx \wedge dy - 2x^3 y\, dx \wedge dz - y^2\, dy \wedge dx + 2xy^3\, dy \wedge dz \\
&= (x^2 + y^2)\, dx \wedge dy - 2x^3 y\, dx \wedge dz + 2xy^3\, dy \wedge dz.
\end{aligned}
$$

In local coordinates, a 2-form can always be written in components as

$$\phi = F_{ij}\ dx^i \wedge dx^j \tag{2.38}$$

If we think of F as a matrix with components F_{ij}, we know from linear algebra that we can write F uniquely as a sum of a symmetric and an antisymmetric matrix, namely,

$$
\begin{aligned}
F &= S + A, \\
&= \frac{1}{2}(F + F^T) + \frac{1}{2}(F - F^T), \\
F_{ij} &= F_{(ij)} + F_{[ij]},
\end{aligned}
$$

where,

$$
\begin{aligned}
F_{(ij)} &= \frac{1}{2}(F_{ij} + F_{ji}), \\
F_{[ij]} &= \frac{1}{2}(F_{ij} - F_{ji}),
\end{aligned}
$$

are the completely symmetric and antisymmetric components. Since $dx^i \wedge dx^j$ is antisymmetric, and the contraction of a symmetric tensor with an antisymmetric tensor is zero, one may assume that the components of the 2-form in equation 2.38 are antisymmetric as well. With this mind, we can easily find a formula using wedges that generalizes the cross product to any dimension.

Let $\alpha = a_i dx^i$ and $\beta = b_i dx^i$ be any two 1-forms in $\mathbf{R}^n$, and Let X and Y be arbitrary vector fields. Then

$$
\begin{aligned}
(\alpha \wedge \beta)(X, Y) &= (a_i dx^i)(X)(b_j dx^j)(Y) - (a_i dx^i)(Y)(b_j dx^j)(X) \\
&= (a_i b_j)[dx^i(X) dx^j(Y) - dx^i(Y) dx^j(X)] \\
&= (a_i b_j)(dx^i \wedge dx^j)(X, Y).
\end{aligned}
$$

Because of the antisymmetry of the wedge product, the last of the above equations can be written as

$$
\begin{aligned}
\alpha \wedge \beta &= \sum_{i=1}^{n} \sum_{j<i}^{n} (a_i b_j - a_j b_i)(dx^i \wedge dx^j), \\
&= \frac{1}{2}(a_i b_j - a_j b_i)(dx^i \wedge dx^j).
\end{aligned}
$$

In particular, if $n = 3$, the reader will recognize the coefficients of the wedge product as the components of the cross product of $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, as shown earlier.

Remark Quantities such as $dx\, dy$ and $dy\, dz$ which often appear in calculus II, are not really well defined. What is meant by them are actually wedge products of 1-forms, but in reversing the order of integration, the antisymmetry of the wedge product is ignored. In performing surface integrals, however, the surfaces must be considered oriented surfaces and one has to insert a negative sign in the differential of surface area component in the xz-plane as shown later in equation 2.83.

2.2.5 Determinants

The properties of n-forms are closely related to determinants, so it might be helpful to digress a bit and review the fundamentals of determinants, as found in any standard linear algebra textbook such as [16]. Let $A \in M_n$ be an $n \times n$ matrix with column vectors

$$A = [\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n]$$

2.2.13 Definition A function $f : M_n \to \mathbf{R}$ is called multilinear if it is linear on each slot; that is,

$$f[\mathbf{v}_1, \dots, a_1\mathbf{v}_i + a_2\mathbf{v}_j, \dots, \mathbf{v}_n] = a_1 f[\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n] + a_2 f[\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n].$$

2.2.14 Definition A function $f : M_n \to R$ is called alternating if it changes sign whenever any two columns are switched; that is,

$$f[\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n] = -f[\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots \mathbf{v}_n]$$

2.2.15 Definition A determinant function is a map $D : M_n \to \mathbf{R}$ that is
a) Multilinear,
b) Alternating,
c) $D(I) = 1$.

One can then prove that this defines the function uniquely. In particular, if $A = (a^i{}_j)$, the determinant can be expressed as

$$\det(A) = \sum_{\pi} \operatorname{sgn}(\pi)\, a^1_{\pi(1)} a^2_{\pi(2)} \dots a^n_{\pi(n)}, \tag{2.39}$$

where the sum is over all the permutations of $\{1, 2 \dots, n\}$. The determinant can also be calculated by the cofactor expansion formula of Laplace. Thus, for example, the cofactor expansion along the entries on the first row $(a^1{}_j)$, is given by

$$\det(A) = \sum_k a^1{}_k \Delta^k{}_1, \tag{2.40}$$

where Δ is the cofactor matrix.
At this point it is convenient to introduce the totally antisymmetric *Levi-Civita* permutation symbol defined as follows:

$$\epsilon_{i_1 i_2 \dots i_k} = \begin{cases} +1 & \text{if } (i_1, i_2, \dots i_k) \text{ is an even permutation of } (1, 2, \dots, k) \\ -1 & \text{if } (i_1, i_2, \dots, i_k) \text{ is an odd permutation of } (1, 2, \dots, k) \\ 0 & \text{otherwise} \end{cases} \tag{2.41}$$

In dimension 3, there are only 6 ($3! = 6$) non-vanishing components of ϵ_{ijk}, namely,

$$\begin{aligned} \epsilon_{123} = \epsilon_{231} = \epsilon_{312} &= 1 \\ \epsilon_{132} = \epsilon_{213} = \epsilon_{321} &= -1 \end{aligned} \tag{2.42}$$

We set the Levi-Civita symbol with some or all the indices up, numerically equal to the permutation symbol will all the indices down. The permutation symbols are useful in the theory of determinants. In fact, if $A = (a^i{}_j)$ is an $n \times n$ matrix, then, equation (2.39) can be written as,

$$\det A = |A| = \epsilon^{i_1 i_2 \dots i_n} a^1{}_{i_1} a^2{}_{i_2} \dots a^n{}_{i_n}. \tag{2.43}$$

Thus, for example, for a 2×2 matrix,

$$\begin{aligned} A &= \begin{vmatrix} a^1{}_1 & a^1{}_2 \\ a^2{}_1 & a^2{}_2 \end{vmatrix}, \\ \det(A) &= \epsilon^{ij} a^1{}_i a^2{}_j, \\ &= \epsilon^{12} a^1{}_1 a^2{}_2 + \epsilon^{21} a^1{}_2 a^2{}_1, \\ &= a^1{}_1 a^2{}_2 - a^1{}_2 a^2{}_1. \end{aligned}$$

We also introduce the generalized Kronecker delta symbol

$$\delta^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} = \begin{cases} +1 & \text{if } (i_1, i_2, \dots, i_k) \text{ is an even permutation of } (j_1, j_2, \dots, j_k) \\ -1 & \text{if } (i_1, i_2, \dots, i_k) \text{ is an odd permutation of } (j_1, j_2, \dots, j_k) \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

If one views the indices i_k as labelling rows and j_k as labelling columns of a matrix, we can represent the completely antisymmetric symbol by the determinant,

$$\delta^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k} = \begin{vmatrix} \delta^{i_1}_{j_1} & \delta^{i_1}_{j_2} & \dots & \delta^{i_1}_{j_k} \\ \delta^{i_2}_{j_1} & \delta^{i_2}_{j_2} & \dots & \delta^{i_2}_{j_k} \\ \dots & \dots & \dots & \dots \\ \delta^{i_k}_{j_1} & \delta^{i_k}_{j_2} & \dots & \delta^{i_k}_{j_k} \end{vmatrix} \tag{2.45}$$

Not surprisingly, the generalized Kronecker delta is related to a product of Levi-Civita symbols by the equation

$$\epsilon^{i_1 i_2 \dots i_k} \epsilon_{j_1 j_2 \dots j_k} = \delta^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_k}, \tag{2.46}$$

which is evident since both sides are completely antisymmetric. In dimension 3, the only non-zero components of δ^{ij}_{kl} are,

$$\begin{aligned} \delta^{12}_{12} = \delta^{13}_{13} = \delta^{23}_{23} = 1 \qquad \delta^{12}_{21} = \delta^{13}_{31} = \delta^{23}_{32} = -1 \\ \delta^{21}_{21} = \delta^{31}_{31} = \delta^{32}_{32} = 1 \qquad \delta^{21}_{12} = \delta^{31}_{13} = \delta^{32}_{23} = -1. \end{aligned}$$

2.2.16 Proposition In dimension 3 the following identities hold

a) $\epsilon^{ijk}_{imn} = \delta^{jk}_{mn} = \delta^j_m \delta^k_n - \delta^j_n \delta^k_m$,

b) $\epsilon^{ijk}_{ijn} = 2\delta^k_n$,

c) $\epsilon^{ijk}_{ijk} = 3!$

Proof For part (a), we compute the determinant by cofactor expansion on the first row

$$\begin{aligned}
\epsilon^{ijk}_{imn} &= \begin{vmatrix} \delta^i_i & \delta^i_m & \delta^i_n \\ \delta^j_i & \delta^j_m & \delta^j_n \\ \delta^k_i & \delta^k_m & \delta^k_n \end{vmatrix} \\
&= \delta^i_i \begin{vmatrix} \delta^j_m & \delta^j_n \\ \delta^k_m & \delta^k_n \end{vmatrix} - \delta^i_m \begin{vmatrix} \delta^j_i & \delta^j_n \\ \delta^k_i & \delta^k_n \end{vmatrix} + \delta^i_n \begin{vmatrix} \delta^j_i & \delta^j_m \\ \delta^k_i & \delta^k_m \end{vmatrix} \\
&= 3 \begin{vmatrix} \delta^j_m & \delta^j_n \\ \delta^k_m & \delta^k_n \end{vmatrix} - \begin{vmatrix} \delta^j_m & \delta^j_n \\ \delta^k_m & \delta^k_n \end{vmatrix} + \begin{vmatrix} \delta^j_n & \delta^j_m \\ \delta^k_n & \delta^k_m \end{vmatrix} \\
&= (3-1-1) \begin{vmatrix} \delta^j_m & \delta^j_n \\ \delta^k_m & \delta^k_n \end{vmatrix} = \begin{vmatrix} \delta^j_m & \delta^j_n \\ \delta^k_m & \delta^k_n \end{vmatrix}
\end{aligned}$$

Here we used the fact that the contraction δ^i_i is just the trace of the identity matrix and the observation that we had to transpose columns in the last determinant in the next to last line. for part (b) follows easily for part(a), namely,

$$\begin{aligned}
\epsilon^{ijk}_{inj} &= \delta^{jk}_{jn}, \\
&= \delta^j_j \delta^k_n - \delta^j_n \delta^k_j, \\
&= 3\delta^k_n - \delta^k_n, \\
&= 2\delta^k_n.
\end{aligned}$$

From this, part (c) is obvious. With considerably more effort, but inductively following the same scheme, one can establish the general formula,

$$\epsilon^{i_1 \ldots i_k, i_{k+1} \ldots i_n} \epsilon_{i_1 \ldots i_k, j_{k+1} \ldots j_n} = k! \delta^{i_{k+1} \ldots i_n}_{j_{k+1} \ldots j_n}. \tag{2.47}$$

2.2.6 Vector Identities

The permutation symbols are very useful in establishing and manipulating classical vector formulas. We present here a number of examples. For this purpose, let,

$$\begin{aligned}
\mathbf{a} &= a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}, \\
\mathbf{b} &= b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}, \\
\mathbf{c} &= c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}, \\
\mathbf{d} &= d_1\mathbf{i} + d_2\mathbf{j} + d_3\mathbf{k},
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
\alpha &= a_1\, dx^1 + a_2\, dx^2 + a_3\, dx^3, \\
\beta &= b_1\, dx^1 + b_2\, dx^2 + b_3\, dx^3, \\
\gamma &= c_1\, dx^1 + c_2\, dx^2 + c_3\, dx^3, \\
\delta &= d_1\, dx^1 + d_2\, dx^2 + d_3\, dx^3,
\end{aligned}$$

1. Dot product and cross product

$$\mathbf{a} \cdot \mathbf{b} = \delta^{ij} a_i b_j = a_i b^i, \qquad (\mathbf{a} \times \mathbf{b})_k = \epsilon_k{}^{ij} a_i b_j \tag{2.48}$$

2. Wedge product

$$\alpha \wedge \beta = \epsilon^k{}_{ij} (\mathbf{a} \times \mathbf{b})_k \, dx^i \wedge dx^j. \tag{2.49}$$

3. Triple product

$$\begin{aligned}
\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= \delta_{ij}a^i(\mathbf{b}\times\mathbf{c})^l,\\
&= \delta_{ij}a^i\epsilon^j{}_{kl}b^kc^l,\\
&= \epsilon_{ikl}a^ib^kc^l,\\
\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= \det([\mathbf{abc}]), && (2.50)\\
&= (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} && (2.51)
\end{aligned}$$

4. Triple cross product: bac-cab identity

$$\begin{aligned}
[\mathbf{a}\times(\mathbf{b}\times\mathbf{c})]_l &= \epsilon_l{}^{mn}a_m(\mathbf{b}\times\mathbf{c})_n\\
&= \epsilon_l{}^{mn}a_m(\epsilon_n{}^{jk}b_jc_k)\\
&= \epsilon_l{}^{mn}\epsilon_n{}^{jk}a_mb_jc_k)\\
&= \epsilon_{mnl}\epsilon^{jkn}a^mb_jc_k)\\
&= (\delta^k_m\delta^j_l - \delta^j_l\delta^k_m)a^mb_jc_k\\
&= b_la^mc_m - c_la^mb_m.
\end{aligned}$$

Rewriting in vector form

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b}). \tag{2.52}$$

5. Dot product of cross products

$$\begin{aligned}
(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) &= \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}\times\mathbf{d}),\\
&= \mathbf{a}\cdot[\mathbf{c}(\mathbf{b}\cdot\mathbf{d}) - \mathbf{d}(\mathbf{b}\cdot\mathbf{c})]\\
&= (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}),
\end{aligned}$$

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) = \begin{vmatrix}\mathbf{a}\cdot\mathbf{c} & \mathbf{a}\cdot\mathbf{d}\\ \mathbf{b}\cdot\mathbf{c} & \mathbf{b}\cdot\mathbf{d}\end{vmatrix} \tag{2.53}$$

6. Norm of cross-product

$$\begin{aligned}
\|\mathbf{a}\times\mathbf{b}\|^2 &= (\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{b}),\\
&= \begin{vmatrix}\mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b}\\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b},\end{vmatrix},\\
&= \|\mathbf{a}||^2\|\mathbf{b}\| - (\mathbf{a}\cdot\mathbf{b})^2 && (2.54)
\end{aligned}$$

7. More wedge products. Let $C = c^k\frac{\partial}{\partial x^k}$, $D = d^m\frac{\partial}{\partial x^m}$. Then,

$$\begin{aligned}
(\alpha\wedge\beta)(C,D) &= \begin{vmatrix}\alpha(C) & \alpha(D)\\ \beta(C) & \beta(D)\end{vmatrix},\\
&= \begin{vmatrix}\mathbf{a}\cdot\mathbf{c} & \mathbf{a}\cdot\mathbf{d}\\ \mathbf{b}\cdot\mathbf{c} & \mathbf{b}\cdot\mathbf{d}\end{vmatrix} && (2.55)
\end{aligned}$$

8. Grad, Curl, Div in $\mathbf{R}^3$

   Let $\nabla_i = \frac{\partial}{\partial x^i}$, $\nabla^i = \delta^{ij}\nabla_j$, $\mathbf{A} = \mathbf{a}$ and define

   ◇ $(\nabla f)_i = \nabla_i f$

   ◇ $(\nabla \times \mathbf{A})_i = \epsilon_i{}^{jk}\nabla_j a_k$

   ◇ $\nabla \cdot \mathbf{A} = \delta^{ij}\nabla_i a_j = \nabla^j a_j$

   ◇ $\nabla \cdot \nabla(f) \equiv \nabla^2 f = \nabla^i \nabla_i f$

   (a)

$$\begin{aligned}(\nabla \times \nabla f)_i &= \epsilon_i{}^{jk}\nabla_j \nabla_f = 0,\\ \nabla \times \nabla f &= 0\end{aligned} \tag{2.56}$$

   (b)

$$\begin{aligned}\nabla \cdot (\nabla \times \mathbf{A}) &= \delta^{ij}\nabla_i(\nabla \times \mathbf{a})_j,\\ &= \delta^{ij}\nabla_i \epsilon_j{}^{kl}\nabla_k a_l,\\ &= \epsilon^{jkl}\nabla_i \nabla_j a_k,\\ \nabla \cdot (\nabla \times \mathbf{A}) &= 0\end{aligned} \tag{2.57}$$

   where in the last step in the two items above we use the fact that a contraction of two symmetric with two antisymmetric indices is always 0.

   (c) The same steps as in the bac-cab identity give

$$\begin{aligned}[\nabla \times (\nabla \times \mathbf{A})]_l &= \nabla_l(\nabla^m a_m) - \nabla^m \nabla_m a_l,\\ \nabla \times (\nabla \times \mathbf{A}) &= \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A},\end{aligned}$$

   where $\nabla^2\mathbf{A}$ means the Laplacian of each component of $\mathbf{A}$.

   This last equation is crucial in the derivation of the wave equation for light from Maxwell's equations for the electromagnetic field.

2.2.7 n-Forms

2.2.17 Definition Let $\alpha^1, \alpha^2, \alpha^2$, be one forms, and $X_1, X_2, X_3 \in \mathscr{X}$. Let π be the set of permutations of $\{1, 2, 3\}$. Then

$$\begin{aligned}(\alpha^1 \wedge \alpha^2 \wedge \alpha^3)(X_1, X_2, X_3) &= \sum_\pi \operatorname{sign}(\pi)\alpha^1(X_{\pi(1)})\alpha^2(X_{\pi(2)})\alpha^3(X_{\pi(3)}),\\ &= \epsilon^{ijk}\alpha^1(X_i)\alpha^2(X_j)\alpha^3(X_k).\end{aligned}$$

This trilinear map is an example of a alternating covariant 3-tensor.

2.2.18 Definition A 3-form ϕ in $\mathbf{R}^n$ is an alternating, covariant 3-tensor. In local coordinates, a 3-from can be written as an object of the following type

$$\phi = A_{ijk} dx^i \wedge dx^j \wedge dx^k \tag{2.58}$$

where we assume that the wedge product of three 1-forms is associative but alternating in the sense that if one switches any two differentials, then the entire expression changes by a minus sign. There is nothing really wrong with using definition (2.58). This definition however, is coordinate-dependent and differential geometers prefer coordinate-free definitions, theorems and proofs.
We can easily extend the concepts above to higher order forms.

2.2.19 Definition Let $T^0_k(\mathbf{R}^n)$ be the set multilinear maps

$$t : \underbrace{T(\mathbf{R}) \times ... \times T(\mathbf{R})}_{\text{k times}} \to \mathbf{R}$$

from k copies of $T(\mathbf{R})$ to $\mathbf{R}$. The map t is called *skew-symmetric* if

$$t(e_1, \ldots, e_k) = \text{sign}(\pi) t(e_{\pi(1)}, \ldots, e_{\pi(k)}), \tag{2.59}$$

where π is the set of permutations of $\{1, \ldots, k\}$. A skew-symmetry covariant tensor of rank k at p, is called a *k-form* at p. denote by $\Lambda^k_{(p)}(\mathbf{R}^n)$ the space of k-forms at $p \in \mathbf{R}^n$. This vector space has dimension

$$\dim \Lambda^k_p(\mathbf{R}^n) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

for $k \leq n$ and dimension 0 for $k > n$. We identify $\Lambda^0_{(p)}(\mathbf{R}^n)$ with the space of C^∞ functions at p. The union of all $\Lambda^k_{(p)}(\mathbf{R}^n)$ as p ranges through all points in $\mathbf{R}^n$ is called the bundle of k-forms and will be denoted by

$$\Lambda^k(\mathbf{R}^n) = \bigcup_p \Lambda^k_p(\mathbf{R}^n).$$

Sections of the bundle are called *k-forms* and the space of all sections is denoted by

$$\Omega^k(\mathbf{R}^n) = \Gamma(\Lambda^k(\mathbf{R}^n)).$$

A section $\alpha \in \Omega^k$ of the bundle technically should be called k-form field, but the consensus in the literature is to call such a section simply a k-form. In local coordinates, a k-form can be written as

$$\alpha = A_{i_1,\ldots i_k}(x) dx^{i_1} \wedge \ldots dx^{i_k}. \tag{2.60}$$

2.2.20 Definition The *alternation map* $A : T^0_k(\mathbf{R}^n) \to T^0_k(\mathbf{R}^n)$ is defined by

$$At(e_1, \ldots, e_k) = \frac{1}{k!} \sum_\pi (\text{sign}\pi) t(e_{\pi(1)}, \ldots, e_{\pi(k)}).$$

2.2.21 Definition If $\alpha \in \Omega^k(\mathbf{R}^n)$ and $\beta \in \Omega^l(\mathbf{R}^n)$, then

$$\alpha \wedge \beta = \frac{(k+l)!}{k!l!} A(\alpha \otimes \beta) \tag{2.61}$$

If α is a k-form and β an l-form, we have

$$\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha. \tag{2.62}$$

Now, for a little combinatorics. Factorials are unavoidable due to the permutation attributes of the wedge product. The convention here follows Marsden [20] and Spivak [34], which reduces proliferation of factorials later. Let us count the number of linearly independent differential forms in Euclidean space. More specifically, we want to find a basis for the vector space of k-forms in $\mathbf{R}^3$. As stated above, we will think of 0-forms as being ordinary functions. Since functions are the "scalars", the space of 0-forms as a vector space has dimension 1.

$\mathbf{R}^2$	Forms	Dim
0-forms	f	1
1-forms	$f dx^1, g dx^2$	2
2-forms	$f dx^1 \wedge dx^2$	1

$\mathbf{R}^3$	Forms	Dim
0-forms	f	1
1-forms	$f_1 dx^1, f_2 dx^2, f_3 dx^3$	3
2-forms	$f_1 dx^2 \wedge dx^3,\ f_2 dx^3 \wedge dx^1,\ f_3 dx^1 \wedge dx^2$	3
3-forms	$f_1 dx^1 \wedge dx^2 \wedge dx^3$	1

The binomial coefficient pattern should be evident to the reader.
It is possible define *tensor-valued differential forms*. Let $E = T^r_s(\mathbf{R^n})$ be the tensor bundle. A tensor-valued p-form is defined as a section

$$T \in \Omega^p(\mathbf{R}^n, E) = \Gamma(E \otimes \Lambda^p(\mathbf{R}^n)).$$

In local coordinates, a tensor-valued k-form is a $\binom{r}{s+p}$ tensor

$$T = T^{i_1,\ldots i_r}{}_{j_1,\ldots j_s,k_1,\ldots,k_p} \frac{\partial}{\partial x^{i_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s} \wedge dx^{k_1} \wedge \ldots \wedge dx^{k_p}. \tag{2.63}$$

Thus, for example, the quantity

$$\Omega^i{}_j = \tfrac{1}{2} R^i{}_{jkl}\, dx^k \wedge dx^l$$

would be called the components of a $\binom{1}{1}$-valued 2-form

$$\Omega = \Omega^i{}_j \frac{\partial}{\partial x^i} \otimes dx^j.$$

The notion of the wedge product can be extended to tensor-valued forms using tensor products on the tensorial indices and wedge products on the differential form indices.

2.3 Exterior Derivatives

In this section we introduce a differential operator that generalizes the classical gradient, curl and divergence operators.

2.3.1 Definition Let α be a one form in $\mathbf{R}^n$. The differential $d\alpha$ is the two-form defined by

$$d\alpha(X,Y) = X(\alpha(Y)) - Y(\alpha(X)), \tag{2.64}$$

for any pair of vector fields X and Y.

To explore the meaning of this definition in local coordinates, let $\alpha = f_i dx^i$ and let $X = \frac{\partial}{\partial x^j}$, $Y = \frac{\partial}{\partial x^k}$, then

$$\begin{aligned} d\alpha(X,Y) &= \frac{\partial}{\partial x^j}\left[f_i dx^i\left(\frac{\partial}{\partial x^k}\right)\right] - \frac{\partial}{\partial x^k}\left[f_i dx^i\left(\frac{\partial}{\partial x^j}\right)\right], \\ &= \frac{\partial}{\partial x^j}(f_i\delta^i_k) - \frac{\partial}{\partial x^k}(f_i\delta^i_j), \\ d\alpha\left(\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^j}\right) &= \frac{\partial f_k}{\partial x^j} - \frac{\partial f_j}{\partial x^k} \end{aligned}$$

Therefore, taking into account the antisymmetry of wedge products, we have.

$$\begin{aligned} d\alpha &= \frac{1}{2}\left(\frac{\partial f_k}{\partial x^j} - \frac{\partial f_j}{\partial x^k}\right) dx^j \wedge dx^k, \\ &= \frac{\partial f_k}{\partial x^j} dx^j \wedge dx^k, \\ &= df_k \wedge dx^k. \end{aligned}$$

The definition 2.64 of a differential of a 1-form can be refined to provide a coordinate-free definition in general manifolds (see 6.28,) and it can be extended to differentials of m-forms. For now, the computation immediately above suffices to motivate the following coordinate dependent definition (for a coordinate-free definition for general manifolds, see (7.17):

2.3.2 Definition Let α be an m-form, given in coordinates as in equation (2.60). The *exterior derivative* of α is the $(m+1)$-form $d\alpha$ given by

$$\begin{aligned} d\alpha &= dA_{i_1,\ldots i_m} \wedge dx^{i_1} \ldots dx^{i_m} \\ &= \frac{\partial A_{i_1,\ldots i_m}}{\partial x^{i_0}}(x) dx^{i_0} \wedge dx^{i_1} \ldots dx^{i_m}. \end{aligned} \tag{2.65}$$

In the special case where α is a 0-form, that is, a function, we write

$$df = \frac{\partial f}{\partial x^i} dx^i.$$

2.3.3 Theorem

$$
\begin{array}{ll}
\text{a)} & d : \Omega^m \longrightarrow \Omega^{m+1} \\
\text{b)} & d^2 = d \circ d = 0 \\
\text{c)} & d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta \qquad \forall \alpha \in \Omega^p, \beta \in \Omega^q
\end{array}
\tag{2.66}
$$

Proof

a) Obvious from equation (2.65).

b) First, we prove the proposition for $\alpha = f \in \Omega^0$. We have

$$
\begin{aligned}
d(d\alpha) &= d(\frac{\partial f}{\partial x^i}) \\
&= \frac{\partial^2 f}{\partial x^j \partial x^i} dx^j \wedge dx^i \\
&= \frac{1}{2}[\frac{\partial^2 f}{\partial x^j \partial x^i} - \frac{\partial^2 f}{\partial x^i \partial x^j}] dx^j \wedge dx^i \\
&= 0.
\end{aligned}
$$

Now, suppose that α is represented locally as in equation (2.60). It follows from equation 2.65, that

$$
d(d\alpha) = d(dA_{i_1,\dots i_m}) \wedge dx^{i_0} \wedge dx^{i_1} \dots dx^{i_m} = 0.
$$

c) Let $\alpha \in \Omega^p, \beta \in \Omega^q$. Then, we can write

$$
\begin{aligned}
\alpha &= A_{i_1,\dots i_p}(x) dx^{i_1} \wedge \dots dx^{i_p} \\
\beta &= B_{j_1,\dots j_q}(x) dx^{j_1} \wedge \dots dx^{j_q}.
\end{aligned}
\tag{2.67}
$$

By definition,

$$
\alpha \wedge \beta = A_{i_1 \dots i_p} B_{j_1 \dots j_q} (dx^{i_1} \wedge \dots \wedge dx^{i_p}) \wedge (dx^{j_1} \wedge \dots \wedge dx^{j_q}).
$$

Now, we take the exterior derivative of the last equation, taking into account that $d(fg) = fdg + gdf$ for any functions f and g. We get

$$
\begin{aligned}
d(\alpha \wedge \beta) &= [d(A_{i_1 \dots i_p}) B_{j_1 \dots j_q} + (A_{i_1 \dots i_p}) d(B_{j_1 \dots j_q})] \\
&\quad (dx^{i_1} \wedge \dots \wedge dx^{i_p}) \wedge (dx^{j_1} \wedge \dots \wedge dx^{j_q}) \\
&= [dA_{i_1 \dots i_p} \wedge (dx^{i_1} \wedge \dots \wedge dx^{i_p})] \wedge [B_{j_1 \dots j_q} \wedge (dx^{j_1} \wedge \dots \wedge dx^{j_q})] + \\
&\quad [A_{i_1 \dots i_p} \wedge (dx^{i_1} \wedge \dots \wedge dx^{i_p})] \wedge (-1)^p [dB_{j_1 \dots j_q} \wedge (dx^{j_1} \wedge \dots \wedge dx^{j_q})] \\
&= d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta.
\end{aligned}
$$

The $(-1)^p$ factor comes into play since in order to pass the term $dB_{j_i \dots j_p}$ through p number of 1-forms of type dx^i, one must perform p transpositions.

2.3.1 Pull-back

2.3.4 Definition Let $F : \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable mapping and let α be a k-form in $\mathbf{R}^m$. Then, at each point $y \in \mathbf{R}^m$ with $y = F(x)$, the mapping F induces a map called the *pull-back* $F^* : \Omega^k_{(F(x))} \to \Omega^k_{(x)}$ defined by

$$(F^*\alpha)_x(X_1, \dots X_k) = \alpha_{F(x)}(F_*X_1, \dots F_*X_k), \tag{2.68}$$

for any tangent vectors $\{X_1, \dots X_k\}$ in $\mathbf{R}^n$.
If g is a 0-form, namely a function, $F^*(g) = g \circ F$. We have the following theorem.

2.3.5 Theorem

$$\begin{array}{lrcl} \text{a)} & F^*(g\alpha_1) & = & (g \circ F)\, F^*\alpha, \\ \text{b)} & F^*(\alpha_1 + \alpha_2) & = & F * \alpha_1 + F^*\alpha_2, \\ \text{c)} & F^*(\alpha \wedge \beta) & = & F^*\alpha \wedge F^*\beta, \\ \text{d)} & F^*(d\alpha) & = & d(F^*\alpha.) \end{array} \tag{2.69}$$

Part (d) is encapsulated in the commuting diagram in figure 2.2.

$$\begin{tikzcd} \Omega^k(\mathbf{R}^n) \arrow[d, "d"] & \Omega^k(\mathbf{R}^m) \arrow[l, "F^*"'] \arrow[d, "d"] \\ \Omega^{k+1}(\mathbf{R}^n) & \Omega^{k+1}(\mathbf{R}^m) \arrow[l, "F^*"'] \end{tikzcd}$$

Fig. 2.2: $d\, F^* = F^*\, d$

Proof Part (a) is basically the definition for the case of 0-forms and part (b) is clear from the linearity of the push-forward. We leave part (c) as an exercise and prove part (d). In the case of a 0-form, let g, be a function and X a vector field in $\mathbf{R}^m$. By a simple computation that amounts to recycling definitions, we have:

$$\begin{aligned} d(F^*g) &= d(g \circ F), \\ (F^*dg)(X) &= dg(F_*X) = (F_*X)(g), \\ &= X(g \circ F) = d(g \circ F)(X), \\ F^*dg &= d(g \circ F), \end{aligned}$$

so, $F^*(dg) = d(F^*g)$ is true by the composite mapping theorem. Let α be a k-form

$$\alpha = A_{i_1,\dots i_k}\, dy^{i_1} \wedge \dots dy^{i_k},$$

so that

$$d\alpha = (dA_{i_1,\dots i_k}) \wedge dy^{i_1} \dots dy^{i_k}.$$

Then, by part (c),

$$\begin{aligned} F^*\alpha &= (F^*A_{i_1,\ldots i_k})\, F^*dy^{i_1} \wedge \ldots F^*dy^{i_k}, \\ d(F^*\alpha) &= dF^*(A_{i_1,\ldots i_k}) \wedge F^*dy^{i_1} \wedge \ldots F^*dy^{i_k}, \\ &= F^*(dA_{i_1,\ldots i_k}) \wedge F^*dy^{i_1} \wedge \ldots F^*dy^{i_k}, \\ &= F^*(d\alpha). \end{aligned}$$

So again, the result rests on the chain rule.

To connect with advanced calculus, suppose that locally the mapping F is given by $y^k = f^k(x^i)$. Then the pullback of the form dg given the formula above $F^*dg = d(g \circ F)$ is given in local coordinates by the chain rule

$$F^*dg = \frac{\partial g}{\partial x^j}dx^j.$$

In particular, the pull-back of local coordinate functions is given by

$$F^*(dy^i) = \frac{\partial y^i}{\partial x^j}dx^j. \tag{2.70}$$

Thus, pullback for the basis 1-forms dy^k is yet another manifestation of the differential as a linear map represented by the Jacobian

$$dy^k = \frac{\partial y^k}{\partial x^i}\, dx^i. \tag{2.71}$$

In particular, if $m = n$,

$$\begin{aligned} d\Omega &= dy^1 \wedge dy^2 \wedge \ldots \wedge dy^n, \\ &= \frac{\partial y^1}{\partial x^{i_1}} \frac{\partial y^2}{\partial x^{i_2}} \cdots \frac{\partial y^n}{\partial x^{i_n}}\, dx^{i_1} \wedge dx^{i_2} \wedge \ldots dx^{i_n}, \\ &= \epsilon^{i_1 i_2 \ldots i_n} \frac{\partial y^1}{\partial x^{i_1}} \frac{\partial y^2}{\partial x^{i_2}} \cdots \frac{\partial y^n}{\partial x^{i_n}}\, dx^1 \wedge dx^2 \wedge \ldots dx^n, \\ &= |J| \wedge dx^1 \wedge \ldots dx^n. \end{aligned} \tag{2.72}$$

So, the pull-back of the volume form,

$$F^*d\,\Omega = |J|\; dx^1 \wedge \ldots \wedge dx^n,$$

gives rise to the integrand that appears in the change of variables theorem for integration. More explicitly, let $R \in \mathbf{R}^n$ be a simply connected region, F be a mapping $F : R \in \mathbf{R}^n \to \mathbf{R}^m$, with $m \geq n$. If ω is a k- form in $\mathbf{R}^m$, then

$$\int_{F(R)} \omega = \int_R F^*\omega \tag{2.73}$$

We refer to this formulation of the change of variables theorem as integration by pull-back.

If $F : \mathbf{R}^n \to \mathbf{R}^n$ is a diffeomorphism, one can push-forward forms with the inverse of the pull-back $F_* = (F^{-1})^*$.

2.3.6 Example Line Integrals
Let $\omega = f_i\, dx^i$ be a one form in $\mathbf{R}^3$ and let C be the curve given by the mapping $\phi : I = t \in [a,b] \to \mathbf{x}(t) \in \mathbf{R}^3$. We can write $\omega = \mathbf{F} \cdot d\mathbf{x}$, where $\mathbf{F} = (f_1, f_2, f_3)$ is a vector field. Then the integration by pull-back equation 2.73 reads,

$$\begin{aligned}
\int_C \mathbf{F} \cdot d\mathbf{x} &= \int_C \omega, \\
&= \int_{\phi(I)} \omega, \\
&= \int_I \phi^* \omega, \\
&= \int_I f^i(\mathbf{x}(t)) \frac{dx^i}{dt}\, dt, \\
&= \int_I \mathbf{F}(\mathbf{x}(\mathbf{t})) \frac{d\mathbf{x}}{dt}\, dt
\end{aligned}$$

This coincides with the definition of line integrals as introduced in calculus.

2.3.7 Example Polar Coordinates
Let $x = r\cos\theta$ and $y = r\sin\theta$ and $f = f(x,y)$. Then

$$\begin{aligned}
dx \wedge dy &= (-r\sin\theta d\theta + \cos\theta dr) \wedge (r\cos\theta d\theta + \sin\theta dr), \\
&= -r\sin^2\theta d\theta \wedge dr + r\cos^2\theta dr \wedge d\theta, \\
&= (r\cos^2\theta + r\sin^2\theta)(dr \wedge d\theta), \\
&= r(dr \wedge d\theta). \\
\int\int f(x,y) dx \wedge dy &= \int\int f(x(r,\theta), y(r,\theta))\; r(dr \wedge d\theta). \qquad (2.74)
\end{aligned}$$

In this case, the element of arc length is diagonal

$$ds^2 = dr^2 + r^2 d\theta^2,$$

as it should be for an orthogonal change of variables. The differential of area is

$$\begin{aligned}
dA &= \sqrt{\det g}\; dr \wedge d\theta, \\
&= r(dr \wedge d\theta)
\end{aligned}$$

If the polar coordinates map is denoted by $F : \mathbf{R}^2 \to \mathbf{R}^2$, then equation 2.74 is just the explicit expression for the pullback of $F^*(\, f\; dA)$.

2.3.8 Example Polar coordinates are just a special example of the general transformation in $\mathbf{R}^2$ given by,

$$
\begin{aligned}
x &= x(u,v), \qquad & dx &= \frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial u}\,dv,\\
y &= y(u,v), \qquad & dy &= \frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial u}\,dv,
\end{aligned}
$$

for which

$$
\phi * (dx \wedge dy) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \, du \wedge dv \tag{2.75}
$$

2.3.9 Example Surface Integrals

Let $R \in \mathbf{R}^2$ be a simply connected region with boundary δR and let the mapping

$$
\phi : (u,v) \in R \longrightarrow \mathbf{x}(u^\alpha) \in \mathbf{R}^2
$$

describe a surface S with boundary $C = \phi(\delta R)$. Here, $\alpha = 1,2$, with $u = u^1,\ v = u^2$. Given a vector field $\mathbf{F} = (f_1, f_2, f_3)$, we assign to it the 2-form

$$
\begin{aligned}
\omega &= \mathbf{F}\cdot dS,\\
&= f_1\,dx^2 \wedge dx^3 - f_2\,dx^1 \wedge dx^3 + f_3\,dx^1 \wedge dx^2,\\
&= \epsilon^i_{jk} f_i\, dx^j \wedge dx^k.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\int\!\!\int_S \mathbf{F}\cdot d\mathbf{S} &= \int\!\!\int_S \omega,\\
&= \int\!\!\int_R \phi^*\omega,\\
&= \int\!\!\int_R \epsilon^i{}_{jk} f_i \frac{\partial x^j}{\partial u^\alpha} du^\alpha \wedge \frac{\partial x^k}{\partial u^\beta} du^\beta,\\
&= \int\!\!\int_R \mathbf{F}\cdot (\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial u})\, du \wedge dv
\end{aligned}
$$

We elaborate a bit on this slick computation, for the benefit of those readers who may have gotten got lost in the index manipulation.

$$
\begin{aligned}
\int\!\!\int_S \mathbf{F}\cdot d\mathbf{S} &= \int\!\!\int_S \omega,\\
&= \int\!\!\int_R \phi^*\omega,\\
&= \int\!\!\int_R \left[f_1\ \phi*(dx^2 \wedge dx^3) - f_2\ \phi*(dx^1 \wedge dx^3) + f_3\ \phi*(dx^1 \wedge dx^2)\right],\\
&= \int\!\!\int_R \left[f_1 \begin{vmatrix} \frac{\partial x^2}{\partial u} & \frac{\partial x^2}{\partial v} \\ \frac{\partial x^3}{\partial u} & \frac{\partial x^3}{\partial v} \end{vmatrix} - f_2 \begin{vmatrix} \frac{\partial x^1}{\partial u} & \frac{\partial x^1}{\partial v} \\ \frac{\partial x^3}{\partial u} & \frac{\partial x^3}{\partial v} \end{vmatrix} + f_3 \begin{vmatrix} \frac{\partial x^1}{\partial u} & \frac{\partial x^1}{\partial v} \\ \frac{\partial x^2}{\partial u} & \frac{\partial x^2}{\partial v} \end{vmatrix} \right] du \wedge dv\\
&= \int\!\!\int_R \mathbf{F}\cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial u}\right) du \wedge dv
\end{aligned}
$$

This pull-back formula for surface integrals is how most students are introduced to this subject in the third semester of calculus.

2.3.10 Remark

1. The differential of area in polar coordinates is of course a special example of the change of coordinate theorem for multiple integrals as indicated above.

2. As shown in equation 2.32 the metric in spherical coordinates is given by
$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2,$$
so the differential of volume is
$$\begin{aligned} dV &= \sqrt{\det g}\; dr \wedge d\theta \wedge d\phi, \\ &= r^2 \sin\theta\; dr \wedge d\theta \wedge d\phi. \end{aligned}$$

2.3.2 Stokes' Theorem in $\mathbf{R}^n$

Let $\alpha = P(x,y)\,dx + Q(x,y)\,dy$. Then,

$$\begin{aligned} d\alpha &= (\tfrac{\partial P}{\partial x} dx + \tfrac{\partial P}{\partial y}) \wedge dx + (\tfrac{\partial Q}{\partial x} dx + \tfrac{\partial Q}{\partial y}) \wedge dy \\ &= \tfrac{\partial P}{\partial y}\, dy \wedge dx + \tfrac{\partial Q}{\partial x}\, dx \wedge dy \\ &= (\tfrac{\partial Q}{\partial x} - \tfrac{\partial P}{\partial y})\, dx \wedge dy. \end{aligned} \tag{2.76}$$

This example is related to Green's theorem in $\mathbf{R}^2$. For convenience, we include here a proof of Green's Theorem in a special case. We say that a region D

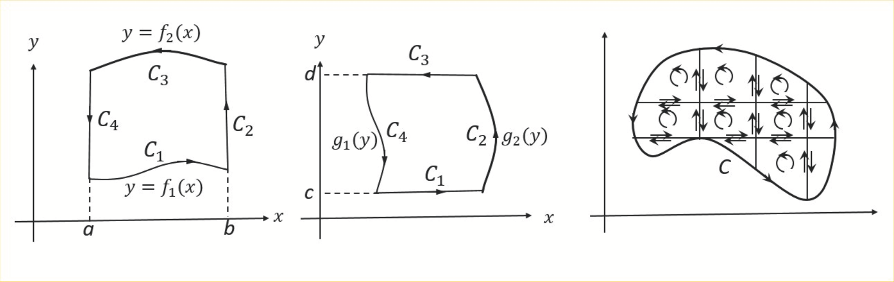


Fig. 2.3: Simple closed curve.

in the plane is of *type I* if it is enclosed between the graphs of two continuous functions of x. The region inside the simple closed curve in figure 2.3 bounded by $f_1(x)$ and $f_2(x)$, between a and b, is a region of type I. A region in the plane is of *type II* if it lies between two continuous functions of y. The region in 2.3 bounded between $c \le y \le d$, would be a region of type II.

2.3.11 Green's theorem

Let C be a simple closed curve in the xy-plane and let $\partial P/\partial x$ and $\partial Q/\partial y$ be continuous functions of (x, y) inside and on C. Let R be the region inside the closed curve so that the boundary $\delta R = C$. Then

$$\oint_{\delta R} P\, dx + Q\, dy = \int\int_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA. \tag{2.77}$$

We first prove that for a type I region such as the one bounded between a and b shown in 2.3, we have

$$\oint_C P\, dx = -\int\int_D \frac{\partial P}{\partial y}\, dA \tag{2.78}$$

Where C comprises the curves C_1, C_2, C_3 and C_4. By the fundamental theorem of calculus, we have on the right,

$$\begin{aligned}\int\int_D \frac{\partial P}{\partial y}\, dA &= \int_a^b \int_{f_1(x)}^{f_2(x)} \frac{\partial P}{\partial y}\, dydx,\\ &= \int_a^b [P(x, f_2(x)) - P(x, f_1(x))]\, dx.\end{aligned}$$

On the left, the integrals along C_2 and C_4 vanish, since there is no variation on x. The integral along C_3 is traversed in opposite direction of C_1, so we have,

$$\begin{aligned}\oint_C P(x, y)\, dx &= \int_{C_1} + \int_{C_2} + \int_{C_3} + \int_{C_4} P(x, y)\, dx,\\ &= \int_{C_1} P(x, y)\, dx - \int_{C_3} P(x, y)\, dx,\\ &= \int_a^b P(x, f_1(x))\, dx - \int_a^b P(x, f_2(x))\, dx\end{aligned}$$

This establishes the veracity of equation 2.78 for type I regions. By a completely analogous process on type II regions, we find that

$$\oint_C Q\, dy = \int\int_D \frac{\partial Q}{\partial x}\, dA. \tag{2.79}$$

The theorem follows by subdividing R into a grid of regions of both types, all oriented in the same direction as shown on the right in figure 2.3. Then one applies equations 2.78 or 2.79, as appropriate, for each of the subdomains. All contributions from internal boundaries cancel since each is traversed twice, each in opposite directions. All that remains of the line integrals is the contribution along the boundary δR.

Let $\alpha = P\, dx + Q\, dy$. Comparing with equation 2.76, we can write Green's theorem in the form

$$\int_C \alpha = \int\int_D d\alpha. \tag{2.80}$$

It is possible to extend Green's Theorem to more complicated regions that are not simple connected. Green's theorem is a special case in dimension of two of Stoke's theorem.

2.3.12 Stokes' theorem

If ω is a C^1 one form in $\mathbf{R}^n$ and S is C^2 surface with boundary $\delta S = C$, then

$$\int_{\delta S} \omega = \int\int_S d\omega. \tag{2.81}$$

Proof The proof can be done by pulling back to the uv-plane and using the chain rule, thus allowing us to use Green's theorem. Let $\omega = f_i\, dx^i$ and S be parametrized by $x^i = x^i(u^\alpha)$, where $(u^1, u^2) \in R \subset \mathbf{R}^2$. We assume that the boundary of R is a simple closed curve. Then

$$\begin{aligned}
\int_C \omega &= \int_{\delta S} f_i\, dx^i, \\
&= \int_{\delta R} f_i \frac{\partial x^i}{\partial u^\alpha}\, du^\alpha, \\
&= \int\int_R \frac{\partial}{\partial u^\beta}(f_i \frac{\partial x^i}{\partial u^\alpha})\, du^\beta \wedge du^\alpha, \\
&= \int\int_R \left[\frac{\partial f_i}{\partial x^k}\frac{\partial x^k}{\partial u^\beta}\frac{\partial x^i}{\partial u^\alpha} + f_i \frac{\partial^2 x^i}{\partial u^\beta \partial u^\alpha}\right] du^\beta \wedge du^\alpha, \\
&= \int\int_R \left[\frac{\partial f_i}{\partial x^k}\frac{\partial x^k}{\partial u^\beta}\frac{\partial x^i}{\partial u^\alpha}\right] du^\beta \wedge du^\alpha, \\
&= \int\int_R \left[\frac{\partial f_i}{\partial x^k}\frac{\partial x^k}{\partial u^\beta}\right] du^\beta \wedge \left[\frac{\partial x^i}{\partial u^\alpha}\right] du^\alpha \\
&= \int\int_S \frac{\partial f_i}{\partial x^k}\, dx^k \wedge dx^i = \int\int_S df_i \wedge dx^i \\
&= \int\int_S d\omega.
\end{aligned}$$

We present a less intuitive but far more elegant proof. The idea is formally the same, namely, we pull-back to the plane by formula 2.73, apply Green's theorem in the form given in equation 2.80, and then use the fact that the pull-back commutes with the differential as in theorem 2.69.

Let $\phi : R \subset \mathbf{R}^2 \to S$ denote the surface parametrization map. Assume that $\phi^{-1}(\delta S) = \delta(\phi^{-1} S)$, that is, the inverse of the boundary of S is the boundary of the domain R. Then,

$$\begin{aligned}
\int_{\delta S} \omega &= \int_{\phi^{-1}(\delta S)} \phi^* \omega = \int_{\delta(\phi^{-1} S)} \phi^* \omega, \\
&= \int \int_{\phi^{-1} S} d(\phi^* \omega), \\
&= \int \int_{\phi^{-1} S} \phi^* (d\omega), \\
&= \int_S d\omega.
\end{aligned}$$

The proof of Stokes' theorem presented here is one of those cases mentioned in the preface, where we have simplified the mathematics for the sake of clarity. Among other things, a rigorous proof requires one to quantify what is meant by the boundary (δS) of a region. The process involves either introducing *simplices* (generalized segments, triangles, tetrahedra...) or *singular cubes* (generalized segments, rectangles, cubes...). The former are preferred in the treatment of homology in algebraic topology, but the latter are more natural to use in the context of integration on manifolds with boundary. A singular n-cube in $\mathbf{R}^n$ is the image under a continuous map,

$$I^n : [0,1]^n \to \mathbf{R}^n,$$

of the Cartesian product of n copies of the unit interval $[0,1]$. The idea is to divide the region S into formal finite sums of singular cubes, called *chains*. One then introduces a boundary operator δ, that maps a singular n-cube and hence n-chain, into an $(n-1)$-singular cube or $(n-1)$-chain. Thus, in $\mathbf{R}^3$ for example, the boundary of a cube, is the sum $\sum c_i F_i$ of the six faces with a judicious choice of coefficients $c_i \in \{-1, 1\}$. With an appropriate scheme to label faces of singular cube and a corresponding definition of the boundary map, one proves that $\delta \circ \delta = 0$. For a thorough treatment, see the beautiful book *Calculus on Manifolds* by M. Spivak [33].

Closed and Exact forms

2.3.13 Example Let $\alpha = M(x,y)dx + N(x,y)dy$, and suppose that $d\alpha = 0$. Then, by the previous example,

$$d\alpha = (\tfrac{\partial N}{\partial x} - \tfrac{\partial M}{\partial y})\, dx \wedge dy.$$

Thus, $d\alpha = 0$ iff $N_x = M_y$, which implies that $N = f_y$ and M_x for some function $f(x,y)$. Hence,

$$\alpha = f_x\, dx + f_y\, dy = df.$$

The reader should also be familiar with this example in the context of exact differential equations of first order and conservative force fields.

2.3.14 Definition A differential form α is called *closed* if $d\alpha = 0$.

2.3.15 Definition A differential form α is called *exact* if there exists a form β such that $\alpha = d\beta$.

Since $d \circ d = 0$, it is clear that an exact form is also closed. The converse need not be true. The standard counterexample is the form,

$$\omega = \frac{-y\,dx + x\,dy}{x^2 + y^2} \tag{2.82}$$

A short computation shows that $d\omega = 0$, so ω is closed. Let $\theta = \tan^{-1}(y/x)$ be the angle in polar coordinates. One can recognize that $\omega = d\theta$, but this is only true in $\mathbf{R}^2 - L$, where L is the non-negative x-axis, $L = \{(x,0) \in \mathbf{R}^2 | x \geq 0\}$. If one computes the line integral from $(-1,0)$ to $(1,0)$ along the top half of the unit circle, the result is π. But the line integral along the bottom half of the unit circle gives $-\pi$. The integral is therefore not path independent, so $\omega \neq d\theta$ on any region that contains the origin. If one tries to find another C^1 function f such that $\omega = df$, one can easily show that $f = \theta + \text{const}$, which is not possible along L.

On the other hand, if one imposes the topological condition that the space is contractible, then the statement is true. A *contractible* space is one that can be deformed continuously to an interior point. We have the following,

2.3.16 Poincaré Lemma. In a contractible space (such as $\mathbf{R}^n$), if a differential form is closed, then it is exact.

To prove this lemma we need much more machinery than we have available at this point. We present the proof in 7.1.17.

2.4 The Hodge $\star$ Operator

2.4.1 Dual Forms

An important lesson students learn in linear algebra, is that all vector spaces of finite dimension n are isomorphic to each other. Thus, for instance, the space P_3 of all real polynomials in x of degree 3, and the space $\mathcal{M}_{2\times 2}$ of real 2 by 2 matrices are, in terms of their vector space properties, basically no different from the Euclidean vector space $\mathbf{R}^4$. As a good example of this, consider the tangent space $T_p\mathbf{R}^3$. The process of replacing $\frac{\partial}{\partial x}$ by $\mathbf{i}$, $\frac{\partial}{\partial y}$ by $\mathbf{j}$ and $\frac{\partial}{\partial z}$ by $\mathbf{k}$ is a linear, 1-1 and onto map that sends the "vector" part of a tangent vector $a^1\frac{\partial}{\partial x} + a^2\frac{\partial}{\partial y} + a^3\frac{\partial}{\partial z}$ to a regular Euclidean vector (a^1, a^2, a^3).

We have also observed that the tangent space $T_p\mathbf{R}^n$ is isomorphic to the cotangent space $T_p^\star\mathbf{R}^n$. In this case, the vector space isomorphism maps the standard basis vectors $\{\frac{\partial}{\partial x^i}\}$ to their duals $\{dx^i\}$. This isomorphism then transforms a contravariant vector to a covariant vector. In terms of components, the isomorphism is provided by the Euclidean metric that maps the components of a contravariant vector with indices up to a covariant vector with indices down.

Another interesting example is provided by the spaces $\Lambda_p^1(\mathbf{R}^3)$ and $\Lambda_p^2(\mathbf{R}^3)$, both of which have dimension 3. It follows that these two spaces must be isomorphic. In this case the isomorphism is given as follows:

$$\begin{array}{ccc} dx & \longmapsto & dy \wedge dz \\ dy & \longmapsto & -dx \wedge dz \\ dz & \longmapsto & dx \wedge dy \end{array} \tag{2.83}$$

More generally, we have seen that the dimension of the space of k-forms in $\mathbf{R}^n$ is given by the binomial coefficient $\binom{n}{k}$. Since

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!},$$

it must be true that

$$\Lambda^k_p(\mathbf{R}^n) \cong \Lambda^{n-k}_p(\mathbf{R}^n). \tag{2.84}$$

To describe the isomorphism between these two spaces, we introduce the following generalization of determinants,

2.4.1 Definition . Let $\phi : \mathbf{R}^n \to \mathbf{R}^n$ be a linear map. The unique constant $\det \phi$ such that,

$$\phi* : \Lambda^n(\mathbf{R}^n) \to \Lambda^n(\mathbf{R}^n)$$

satisfies,

$$\phi^* \omega = (\det \phi)\, \omega, \tag{2.85}$$

for all n-forms, is called the *determinant* of ϕ. This is congruent with the standard linear algebra formula 2.43, since in a particular basis, the Jacobian of a linear map is the same as the matrix the represents the linear map in that basis. Let, $g(X, Y)$ be an inner product and $\{e_1, \ldots, e_n\}$ be an orthonormal basis with dual forms $\{\theta^1, \ldots \theta^n\}$. The element of arc length is, the bilinear symmetric tensor

$$ds^2 = g_{ij}\, \theta^i \otimes \theta^j.$$

The metric then induces an n-form

$$d\Omega = \theta^1 \wedge \theta^2 \ldots \wedge \theta^n,$$

called the *volume* element. With this choice of form, the reader will recognize equation 2.85 as the integrand in the change of variables theorem for multiple integration, as in example 2.74. More generally, if $\{f_1, \ldots f_n\}$ is a positively oriented basis with dual basis $\{\phi^1, \ldots \phi^n\}$, then,

$$d\Omega = \sqrt{\det g}\; \phi^1 \wedge \ldots \wedge \phi^n. \tag{2.86}$$

2.4.2 Definition Let g be the matrix representing the components of the metric in $\mathbf{R}^n$. The *Hodge* $\star$ *operator* is the linear isomorphism $\star : \Lambda^n_p(\mathbf{R}^n) \longrightarrow \Lambda^{n-k}_p(\mathbf{R}^n)$ defined in standard local coordinates by the equation,

$$\star\,(dx^{i_1} \wedge \ldots \wedge dx^{i_k}) = \frac{\sqrt{\det g}}{(n-k)!} \epsilon^{i_1 \ldots i_k}{}_{i_{k+1} \ldots i_n}\, dx^{i_{k+1}} \wedge \ldots \wedge dx^{i_n}, \tag{2.87}$$

For flat Euclidean space $\sqrt{\det g} = 1$, so the factor in the definition may appear superfluous. However, when we consider more general Riemannian manifolds, we will have to be more careful with raising and lowering indices with the metric, and take into account that the Levi-Civita symbol is not a tensor but something slightly more complicated called a tensor density. Including the $\sqrt{\det g}$ is done in anticipation of this more general setting later. Since the forms $dx^{i_1} \wedge \ldots \wedge dx^{i_k}$ constitute a basis of the vector space $\Lambda^k_p(\mathbf{R}^n)$ and the $\star$ operator is assumed to be a linear map, equation (2.87) completely specifies the map for all k-forms. In particular, if the components of a dual of a form are equal to the components of the form, the tensor is called *self-dual.* Of course, this can only happen if the tensor and its dual are of the same rank.

A metric g on $\mathbf{R}^n$ induces an inner product on $\Lambda^k(\mathbf{R}^n)$ as follows. Let $\{e_1, , \ldots e_n\}$ by an orthonormal basis with dual basis $\theta^1, \ldots, \theta^n$. If $\alpha, \beta \in \Lambda^k(\mathbf{R}^n)$, we can write

$$\begin{aligned} \alpha &= a_{i_1 \ldots i_k}\ \theta^{i_1} \wedge, \ldots \theta^{i_k}, \\ \beta &= b_{j_1 \ldots j_k}\ \theta^{j_1} \wedge, \ldots \theta^{j_k} \end{aligned}$$

The induced inner product is defined by

$$< \alpha, \beta >^{(k)} = \frac{1}{k!} a_{i_1 \ldots i_k} b^{i_1 \ldots i_k}. \tag{2.88}$$

If $\alpha, \beta \in \Lambda^k(\mathbf{R}^n)$, then $\star\beta \in \Lambda^{n-k}(\mathbf{R}^n)$, so $\alpha \wedge \star\beta$ must be a multiple of the volume form. The Hodge $\star$ operator is the unique isomorphism such that

$$\alpha \wedge \star\beta = < \alpha, \beta >^{(k)}\ d\,\Omega. \tag{2.89}$$

Clearly,

$$\alpha \wedge \star\beta = \star\alpha \wedge \beta$$

When it is evident that the inner product is the induced inner product on $\Lambda^k(\mathbf{R}^n)$ the indicator (k) is often suppressed. An equivalent definition of the induced inner product of two k-forms is given by

$$< \alpha, \beta > = \int (\alpha \wedge \star\beta)\ d\,\Omega. \tag{2.90}$$

If α is a k-form and β is a $(k-1)$-form, one can define the adjoint or *co-differential* by

$$< \delta\alpha, \beta > = < \alpha, d\beta > . \tag{2.91}$$

The adjoint is given by

$$\delta = (-1)^{nk+n+1} \star d \star . \tag{2.92}$$

In particular,

$$\delta = \begin{cases} \quad - \star d \star & \text{if } n \text{ is even} \\ (-1)^k \star\ d \star & \text{if } n \text{ is odd} \end{cases} \tag{2.93}$$

The differential maps $(k-1)$-forms to k-forms, and the co-differential maps k-forms to $(k-1)$-forms. It is also the case that $\delta \circ \delta = 0$ The combination,

$$\Delta = (d + \delta)^2 = d\delta + \delta d \tag{2.94}$$

extends the Laplacian operator to forms. It maps k-forms to k-forms. A central result in harmonic analysis is the *Hodge decomposition* theorem, that states that given any k-form ω, can be split uniquely as

$$\omega = d\alpha + \delta\beta + \gamma, \tag{2.95}$$

where $\alpha \in \Omega^{k-1}$, $\beta \in \Omega^{k+1}$, and $\Delta\gamma = 0$

2.4.3 Example Hodge operator in $\mathbf{R}^2$
In $\mathbf{R}^2$,

$$\star dx = dy \qquad \star\, dy = -dx,$$

or, if one thinks of a matrix representation of $\star : \Omega(\mathbf{R}^2) \to \Omega(\mathbf{R}^2)$ in standard basis, we can write the above as

$$\star \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix}.$$

The reader might wish to peek at the symplectic matrix 5.50 in the discussion in chapter 5 on conformal mappings. Given functions $u = u(x,y)$ and $v = v(x,y)$, let $\omega = u\; dx - v\; dy$. Then,

$$\begin{aligned} d\omega &= -(u_y + v_x)\; dx \wedge dy, \\ d \star \omega &= (u_x - u_y)\; dx \wedge dy, \end{aligned} \qquad \text{hence} \qquad \begin{aligned} d\omega = 0 &\Rightarrow u_y = -v_x, \\ \star d\omega = 0 &\Rightarrow u_x = v_y. \end{aligned} \tag{2.96}$$

Thus, the equations $d\omega = 0$ and $d\star\omega = 0$ are equivalent to the Cauchy-Riemann equations for a holomorphic function $f(z) = u(x,y) + iv(x,y)$. On the other hand,

$$\begin{aligned} du &= u_x\; dx + u_y\; dy, \\ dv &= v_x\; dx + u_y\; dy, \end{aligned}$$

so the determinant of the Jacobian of the transformation $T : (x,y) \to (u,v)$), with the condition above on ω, is given by,

$$|J| = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = u_x^2 + u_y^2 = v_x^2 + v_y^2.$$

If $|J| \neq 0$, we can set $u_x = R\cos\phi, \quad u_y = R\sin\phi$, for some R and some angle ϕ. Then,

$$|J| = \begin{vmatrix} R & 0 \\ 0 & R \end{vmatrix} \begin{vmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{vmatrix}.$$

Thus, the transformation is given by the composition of a dilation and a rotation. A more thorough discussion of this topic is found in the section of conformal maps in chapter 5.

2.4.4 Example Hodge operator in $\mathbf{R}^3$

$$\begin{aligned}
\star dx^1 &= \epsilon^1{}_{jk} dx^j \wedge dx^k, \\
&= \frac{1}{2!}[\epsilon^1{}_{23} dx^2 \wedge dx^3 + \epsilon^1{}_{32} dx^3 \wedge dx^2], \\
&= \frac{1}{2!}[dx^2 \wedge dx^3 - dx^3 \wedge dx^2], \\
&= \frac{1}{2!}[dx^2 \wedge dx^3 + dx^2 \wedge dx^3], \\
&= dx^2 \wedge dx^3.
\end{aligned}$$

We leave it to the reader to complete the computation of the action of the $\star$ operator on the other basis forms. The results are

$$\begin{aligned}
\star dx^1 &= +dx^2 \wedge dx^3, \\
\star dx^2 &= -dx^1 \wedge dx^3, \\
\star dx^3 &= +dx^1 \wedge dx^2,
\end{aligned} \tag{2.97}$$

$$\begin{aligned}
\star(dx^2 \wedge dx^3) &= dx^1, \\
\star(-dx^3 \wedge dx^1) &= dx^2, \\
\star(dx^1 \wedge dx^2) &= dx^3,
\end{aligned} \tag{2.98}$$

and

$$\star\,(dx^1 \wedge dx^2 \wedge dx^3) = 1. \tag{2.99}$$

In particular, if $f : \mathbf{R}^3 \longrightarrow \mathbf{R}$ is any 0-form (a function), then,

$$\begin{aligned}
\star f &= f(dx^1 \wedge dx^2 \wedge dx^3), \\
&= f dV,
\end{aligned} \tag{2.100}$$

where dV is the volume form.

2.4.5 Example Let $\alpha = a_1 dx^1 a_2 dx^2 + a_3 dx^3$, and $\beta = b_1 dx^1 b_2 dx^2 + b_3 dx^3$. Then,

$$\begin{aligned}
\star(\alpha \wedge \beta) &= (a_2 b_3 - a_3 b_2) \star (dx^2 \wedge dx^3) + (a_1 b_3 - a_3 b_1) \star (dx^1 \wedge dx^3) + \\
&\quad (a_1 b_2 - a_2 b_1) \star (dx^1 \wedge dx^2), \\
&= (a_2 b_3 - a_3 b_2) dx^1 + (a_1 b_3 - a_3 b_1) dx^2 + (a_1 b_2 - a_2 b_1) dx^3, \\
&= (\mathbf{a} \times \mathbf{b})_i \, dx^i.
\end{aligned} \tag{2.101}$$

The previous examples provide some insight on the action of the $\wedge$ and $\star$ operators. If one thinks of the quantities dx^1, dx^2 and dx^3 as analogous to $\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$ and $\vec{\mathbf{k}}$, then it should be apparent that equations (2.97) are the differential geometry versions of the well-known relations

$$\begin{aligned}
\mathbf{i} &= \mathbf{j} \times \mathbf{k}, \\
\mathbf{j} &= -\mathbf{i} \times \mathbf{k}, \\
\mathbf{k} &= \mathbf{i} \times \mathbf{j}.
\end{aligned}$$

2.4.6 Example In Minkowski space the collection of all 2-forms has dimension $\binom{4}{2} = 6$. The Hodge $\star$ operator in this case splits $\Omega^2(M_{1,3})$ into two 3-dim subspaces $\Omega^2_{\pm}$, such that $\star : \Omega^2_{\pm} \longrightarrow \Omega^2_{\mp}$.
More specifically, Ω^2_{+} is spanned by the forms $\{dx^0 \wedge dx^1, dx^0 \wedge dx^2, dx^0 \wedge dx^3\}$, and Ω^2_{-} is spanned by the forms $\{dx^2 \wedge dx^3, -dx^1 \wedge dx^3, dx^1 \wedge dx^2\}$. The action of $\star$ on Ω^2_{+} is

$$\begin{aligned}
\star(dx^0 \wedge dx^1) = &\ \tfrac{1}{2}\epsilon^{01}{}_{kl}dx^k \wedge dx^l = & -dx^2 \wedge dx^3,\\
\star(dx^0 \wedge dx^2) = &\ \tfrac{1}{2}\epsilon^{02}{}_{kl}dx^k \wedge dx^l = & +dx^1 \wedge dx^3,\\
\star(dx^0 \wedge dx^3) = &\ \tfrac{1}{2}\epsilon^{03}{}_{kl}dx^k \wedge dx^l = & -dx^1 \wedge dx^2,
\end{aligned}$$

and on Ω^2_{-},

$$\begin{aligned}
\star(+dx^2 \wedge dx^3) = &\ \tfrac{1}{2}\epsilon^{23}{}_{kl}dx^k \wedge dx^l = & dx^0 \wedge dx^1,\\
\star(-dx^1 \wedge dx^3) = &\ \tfrac{1}{2}\epsilon^{13}{}_{kl}dx^k \wedge dx^l = & dx^0 \wedge dx^2,\\
\star(+dx^1 \wedge dx^2) = &\ \tfrac{1}{2}\epsilon^{12}{}_{kl}dx^k \wedge dx^l = & dx^0 \wedge dx^3.
\end{aligned}$$

In verifying the equations above, we recall that the Levi-Civita symbols that contain an index with value 0 in the up position have an extra minus sign as a result of raising the index with η^{00}. If $F \in \Omega^2(M)$, we will formally write $F = F_+ + F_-$, where $F_{\pm} \in \Omega^2_{\pm}$. We would like to note that the action of the dual operator on $\Omega^2(M)$ is such that $\star : \Omega^2(M) \longrightarrow \Omega^2(M)$, and $\star^2 = -1$. In a vector space a map like $\star$, with the property $\star^2 = -1$ is called a *linear involution* of the space. In the case in question, $\Omega^2_{\pm}$ are the eigenspaces corresponding to the +1 and -1 eigenvalues of this involution. It is also worthwhile to calculate the duals of 1-forms in $M_{1,3}$. The results are,

$$\begin{aligned}
\star dt &= -dx^1 \wedge dx^2 \wedge dx^3,\\
\star dx^1 &= +dx^2 \wedge dt \wedge dx^3,\\
\star dx^2 &= +dt \wedge dx^1 \wedge dx^3,\\
\star dx^3 &= +dx^1 \wedge dt \wedge dx^2.
\end{aligned} \tag{2.102}$$

2.4.2 Laplacian

Classical differential operators that enter in Green's and Stokes' theorems are better understood as special manifestations of the exterior differential and the Hodge $\star$ operators in $\mathbf{R}^3$. Here is precisely how this works:

1. Let $f : \mathbf{R}^3 \longrightarrow \mathbf{R}$ be a $\mathcal{C}^\infty$ function. Then

$$df = \frac{\partial f}{\partial x^j} dx^j = \nabla f \cdot d\mathbf{x}. \tag{2.103}$$

2. Let $\alpha = A_i dx^i$ be a 1-form in $\mathbf{R}^3$. Then

$$\begin{aligned}
(\star d)\alpha &= \frac{1}{2}\left(\frac{\partial A_i}{\partial x^j} - \frac{\partial A_i}{\partial x^j}\right) \star (dx^i \wedge dx^j)\\
&= (\nabla \times \mathbf{A}) \cdot d\mathbf{S}.
\end{aligned} \tag{2.104}$$

3. Let $\alpha = B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2$ be a 2-form in $\mathbf{R}^3$. Then

$$
\begin{aligned}
d\alpha &= (\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3})\, dx^1 \wedge dx^2 \wedge dx^3 \\
&= (\nabla \cdot \mathbf{B})\, dV.
\end{aligned} \tag{2.105}
$$

4. Let $\alpha = B_i dx^i$, then

$$(\star d \star)\, \alpha = \nabla \cdot \mathbf{B}. \tag{2.106}$$

5. Let f be a real valued function. Then the *Laplacian* is given by:

$$(\star\, d\, \star)\, df = \nabla \cdot \nabla f = \nabla^2 f. \tag{2.107}$$

The Laplacian definition here is consistent with 2.94 because in the case of a function f, that is, a 0-form, $\delta f = 0$ so $\Delta f = \delta df$. The results above can be summarized in terms of short exact sequence called the *de Rham complex* as shown in figure 2.4. The sequence is called exact because successive application of the differential operator gives zero. That is, $d \circ d = 0$. Since there are no 4-forms in $\mathbf{R}^3$, the sequence terminates as shown. If one starts with a function

$$
\begin{tikzcd}
\Omega^0(\mathbf{R}^3) \arrow[r, "d", "\text{Grad}"'] & \Omega^1(\mathbf{R}^3) \arrow[r, "d", "\text{Curl}"'] \arrow[r, bend left, "\star"] & \Omega^2(\mathbf{R}^3) \arrow[r, "d", "\text{Div}"'] & \Omega^3(\mathbf{R}^3)
\end{tikzcd}
$$

Fig. 2.4: de Rham Complex in $\mathbf{R}^3$

in $\Omega^0(\mathbf{R}^3)$, then $(d \circ d) f = 0$ just says that $\nabla \times \nabla f = 0$, as in the case of conservative vector fields. If instead, one starts with a one form α in $\Omega^1(\mathbf{R}^3)$, corresponding to a vector field $\mathbf{A}$, then $(d \circ d)\alpha = 0$ says that $\nabla \cdot (\nabla \times \mathbf{A}) = 0$, as in the case of incompressible vector fields. If one starts with a function, but instead of applying the differential twice consecutively, one “hops” in between with the Hodge operator, the result is the Laplacian of the function.

If we denote by R a simply connected closed region in Euclidean space whose boundary is δR , then in terms of forms, the fundamental theorem of calculus, Stokes’ theorem (See ref 2.81), and the divergence theorem in $\mathbf{R}^3$ can be expressed by a single generalized Stokes’ theorem.

$$\int_{\delta R} \omega = \int\!\!\int_R d\omega. \tag{2.108}$$

We find it irresistible to point out that if one defines a complex one-form,

$$\omega = f(z)\, dz, \tag{2.109}$$

where $f(z) = u(x, y) + iv(x, y)$, and where one assumes that u, v are differentiable with continuous derivatives, then the conditions introduced in equation

2.96 are equivalent to requiring that $d\omega = 0$. In other words, if the form is closed, then u and v satisfy the Cauchy-Riemann equations. Stokes' theorem then tells us that in a contractible region with boundary C, the line integral

$$\int_C \omega = \int_C f(z)\, dz = 0.$$

This is Cauchy's integral theorem. We should also point out the tantalizing resemblance of equations 2.96 to Maxwell's equations in the section that follows.

2.4.3 Maxwell Equations

The classical equations of Maxwell describing electromagnetic phenomena are

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 4\pi\rho & \qquad \nabla \times \mathbf{B} &= 4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} \\ \nabla \cdot \mathbf{B} &= 0 & \qquad \nabla \times \mathbf{E} &= -\frac{\partial B}{\partial t}, \end{aligned} \tag{2.110}$$

where we are using Gaussian units with $c = 1$. We would like to formulate these equations in the language of differential forms. Let $x^\mu = (t, x^1, x^2, x^3)$ be local coordinates in Minkowski's space $M_{1,3}$. Define the Maxwell 2-form F by the equation

$$F = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu, \quad (\mu, \nu = 0, 1, 2, 3), \tag{2.111}$$

where

$$F_{\mu\nu} = \begin{bmatrix} 0 & -E_x & -E_y & -E_y \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{bmatrix}. \tag{2.112}$$

Written in complete detail, Maxwell's 2-form is given by

$$\begin{aligned} F \quad = \quad & -E_x dt \wedge dx^1 - E_y dt \wedge dx^2 - E_z dt \wedge dx^3 + \\ & B_z dx^1 \wedge dx^2 - B_y dx^1 \wedge dx^3 + B_x dx^2 \wedge dx^3. \end{aligned} \tag{2.113}$$

We also define the source current 1-form

$$J = J_\mu dx^\mu = \rho dt + J_1 dx^1 + J_2 dx^2 + J_3 dx^3. \tag{2.114}$$

2.4.7 Proposition Maxwell's Equations (2.110) are equivalent to the equations

$$\begin{aligned} dF &= 0, \\ d \star F &= 4\pi \star J. \end{aligned} \tag{2.115}$$

Proof The proof is by direct computation using the definitions of the exterior derivative and the Hodge ⋆ operator.

$$
\begin{aligned}
dF \quad = \quad & -\frac{\partial E_x}{\partial x^2} \wedge dx^2 \wedge dt \wedge dx^1 - \frac{\partial E_x}{\partial x^3} \wedge dx^3 \wedge dt \wedge dx^1 + \\
& -\frac{\partial E_y}{\partial x^1} \wedge dx^1 \wedge dt \wedge dx^2 - \frac{\partial E_y}{\partial x^3} \wedge dx^3 \wedge dt \wedge dx^2 + \\
& -\frac{\partial E_z}{\partial x^1} \wedge dx^1 \wedge dt \wedge dx^3 - \frac{\partial E_z}{\partial x^2} \wedge dx^2 \wedge dt \wedge dx^3 + \\
& \frac{\partial B_z}{\partial t} \wedge dt \wedge dx^1 \wedge dx^2 - \frac{\partial B_z}{\partial x^3} \wedge dx^3 \wedge dx^1 \wedge dx^2 - \\
& \frac{\partial B_y}{\partial t} \wedge dt \wedge dx^1 \wedge dx^3 - \frac{\partial B_y}{\partial x^2} \wedge dx^2 \wedge dx^1 \wedge dx^3 + \\
& \frac{\partial B_x}{\partial t} \wedge dt \wedge dx^2 \wedge dx^3 + \frac{\partial B_x}{\partial x^1} \wedge dx^1 \wedge dx^2 \wedge dx^3.
\end{aligned}
$$

Collecting terms and using the antisymmetry of the wedge operator, we get

$$
\begin{aligned}
dF \quad = \quad & (\frac{\partial B_x}{\partial x^1} + \frac{\partial B_y}{\partial x^2} + \frac{\partial B_z}{\partial x^3})\, dx^1 \wedge dx^2 \wedge dx^3 + \\
& (\frac{\partial E_y}{\partial x^3} - \frac{\partial E_z}{\partial x^2} - \frac{\partial B_x}{\partial t})\, dx^2 \wedge dt \wedge dx^3 + \\
& (\frac{\partial E_z}{\partial x^1} - \frac{\partial E_x}{\partial dx^3} - \frac{\partial B_y}{\partial t})\, dt \wedge dx^1 \wedge x^3 + \\
& (\frac{\partial E_x}{\partial x^2} - \frac{\partial E_y}{\partial x^1} - \frac{\partial B_z}{\partial t})\, dx^1 \wedge dt \wedge dx^2.
\end{aligned}
$$

Therefore, $dF = 0$ iff

$$\frac{\partial B_x}{\partial x^1} + \frac{\partial B_y}{\partial x^2} + \frac{\partial B_y}{\partial x^3} = 0,$$

which is the same as

$$\nabla \cdot \mathbf{B} = 0,$$

and

$$
\begin{aligned}
\frac{\partial E_y}{\partial x^3} - \frac{\partial E_z}{\partial x^2} - \frac{\partial B_x}{\partial t} &= 0, \\
\frac{\partial E_z}{\partial x^1} - \frac{\partial E_x}{\partial x^3} - \frac{\partial B_y}{\partial t} &= 0, \\
\frac{\partial E_x}{\partial x^2} - \frac{\partial E_y}{\partial x^1} - \frac{\partial B_z}{\partial t} &= 0,
\end{aligned}
$$

which means that

$$-\nabla \times \mathbf{E} - \frac{\partial \mathbf{B}}{\partial t} = 0. \tag{2.116}$$

To verify the second set of Maxwell equations, we first compute the dual of the current density 1-form (2.114) using the results from example 2.4.1. We get

$$\star J = [-\rho dx^1 \wedge dx^2 \wedge dx^3 + J_1 dx^2 \wedge dt \wedge dx^3 + J_2 dt \wedge dx^1 \wedge dx^3 + J_3 dx^1 \wedge dt \wedge dx^2]. \tag{2.117}$$

We could now proceed to compute $d \star F$, but perhaps it is more elegant to notice that $F \in \Omega^2(M)$, and so, according to example (2.4.1), F splits into $F = F_+ + F_-$. In fact, we see from (2.112) that the components of F_+ are those of $-\mathbf{E}$ and the components of F_- constitute the magnetic field vector $\mathbf{B}$. Using the results of example (2.4.1), we can immediately write the components of $\star F$:

$$\begin{aligned} \star F &= \frac{1}{2!} B_x dt \wedge dx^1 + B_y dt \wedge dx^2 + B_z dt \wedge dx^3 + \\ & E_z dx^1 \wedge dx^2 - E_y dx^1 \wedge dx^3 + E_x dx^2 \wedge dx^3], \end{aligned} \tag{2.118}$$

or equivalently,

$$F^\star_{\mu\nu} = \begin{bmatrix} 0 & B_x & B_y & B_y \\ -B_x & 0 & E_z & -E_y \\ -B_y & -E_z & 0 & E_x \\ -B_z & E_y & -E_x & 0 \end{bmatrix}. \tag{2.119}$$

Effectively, the dual operator amounts to exchanging

$$\begin{aligned} \mathbf{E} &\longmapsto -\mathbf{B} \\ \mathbf{B} &\longmapsto +\mathbf{E}, \end{aligned}$$

in the left hand side of the first set of Maxwell equations. We infer from equations (2.116) and (2.117) that

$$\nabla \cdot \mathbf{E} = 4\pi\rho$$

and

$$\nabla \times \mathbf{B} - \frac{\partial E}{\partial t} = 4\pi \mathbf{J}.$$

Most standard electrodynamic textbooks carry out the computation entirely tensor components, To connect with this approach, we should mention that it $F^{\mu\nu}$ represents the electromagnetic tensor, then the dual tensor is

$$F^\star_{\mu\nu} = \frac{\sqrt{\det g}}{2} \epsilon_{\mu\nu\sigma\tau} F^{\sigma\tau}. \tag{2.120}$$

Since $dF = 0$, in a contractible region there exists a one form A such that $F = dA$. The form A is called the *4-vector potential.* The components of A are,

$$\begin{aligned} A &= A_\mu \, dx^\mu, \\ A_\mu &= (\phi, \mathbf{A}) \end{aligned} \tag{2.121}$$

where ϕ is the electric potential and $\mathbf{A}$ the magnetic vector potential. The components of the electromagnetic tensor F are given by

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}. \tag{2.122}$$

The classical electromagnetic Lagrangian is

$$L_{EM} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + J^\mu A_\mu, \tag{2.123}$$

with corresponding Euler-Lagrange equations

$$\frac{\partial}{\partial x^{\mu}}\left[\frac{\partial L}{\frac{\partial A_{\mu}}{\partial A_{\nu}}}\right]-\frac{\partial L}{\partial A_{\mu}}=0. \tag{2.124}$$

To carry out the computation we first use the Minkowski to write the Lagrangian with the indices down. The key is to keep in mind that $A_{\mu,\nu}$ are treated as independent variables, so the derivatives of $A_{\alpha,\beta}$ vanish unless $\mu=\alpha$ and $\nu=\beta$. We get,

$$\begin{aligned}
\frac{\partial L}{\partial(A_{\mu,\nu})} &= -\frac{1}{4}\frac{\partial L}{\partial(A_{\mu,\nu})}(F_{\alpha\beta}F^{\alpha\beta}),\\
&= -\frac{1}{4}\frac{\partial L}{\partial(A_{\mu,\nu})}(F_{\alpha\beta}F_{\lambda\sigma}\eta^{\alpha\lambda}\eta^{\beta\sigma}),\\
&= -\frac{1}{4}\eta^{\alpha\lambda}\eta^{\beta\sigma}[F_{\alpha\beta}(\delta^{\mu}_{\lambda}\delta^{\mu}_{\sigma}-\delta^{\mu}_{\sigma}\delta^{\mu}_{\lambda})+F_{\lambda\sigma}(\delta^{\mu}_{\alpha}\delta^{\mu}_{\beta}-\delta^{\mu}_{\beta}\delta^{\mu}_{\alpha}),\\
&= -\frac{1}{4}[\eta^{\alpha\mu}\eta^{\beta\nu}F_{\alpha\beta}+\eta^{\mu\lambda}\eta^{\nu\sigma}F_{\lambda\sigma}-\eta^{\alpha\nu}\eta^{\beta\mu}F_{\alpha\beta}-\eta^{\nu\lambda}\eta^{\mu\sigma}(F_{\lambda\sigma}],\\
&= -\frac{1}{4}[F^{\mu\nu}+F^{\mu\nu}-F^{\nu\mu}-F^{\nu\mu}],\\
&= -F^{\mu\nu}.
\end{aligned}$$

On the other hand,

$$\frac{\partial L}{\partial A_{\mu}}=J^{\nu}.$$

Therefore, the field equations are

$$\frac{\partial}{\partial x^{\mu}}F^{\mu\nu}=J^{\mu}. \tag{2.125}$$

The dual equations equivalent to the other pair of Maxwell equations is

$$\frac{\partial}{\partial x^{\mu}}\star F^{\mu\nu}=0.$$

In the gauge theory formulation of classical electrodynamics, the invariant expression for the Lagrangian is the square of the norm of the field F under the induced inner product

$$<F,F>=-\int(F\wedge\star F)\,d\Omega. \tag{2.126}$$

This the starting point to generalize to non-Abelian gauge theories.

Chapter 3

Connections

3.1 Frames

This chapter is dedicated to professor Arthur Fischer. In my second year as an undergraduate at Berkeley, I took the undergraduate course in differential geometry which to this day is still called Math 140. The driving force in my career was trying to understand the general theory of relativity, which was only available at the graduate level. However, the graduate course (Math 280 at the time) read that the only prerequisite was Math 140. So I got emboldened and enrolled in the graduate course taught that year by Dr. Fischer. The required book for the course was the classic by Adler, Bazin, Schiffer. I loved the book; it was definitely within my reach and I began to devour the pages with the great satisfaction that I was getting a grasp of the mathematics and the physics. On the other hand, I was completely lost in the course. It seemed as if it had nothing to do with the material I was learning on my own. Around the third week of classes, Dr. Fischer went through a computation with these mysterious operators, and upon finishing the computation he said if we were following, he had just derived the formula for the Christoffel symbols. Clearly, I was not following, they looked nothing like the Christoffel symbols I had learned from the book. So, with great embarrassment I went to his office and explained my predicament. He smiled, apologized when he did not need to, and invited me to 1-1 sessions for the rest of the two-semester course. That is how I got through the book he was really using, namely Abraham-Marsden. I am forever grateful.

As noted in Chapter 1, the theory of curves in $\mathbf{R}^3$ can be elegantly formulated by introducing orthonormal triplets of vectors which we called Frenet frames. The Frenet vectors are adapted to the curves in such a manner that the rate of change of the frame gives information about the curvature of the curve. In this chapter we will study the properties of arbitrary frames and their corresponding rates of change in the direction of the various vectors in the frame. These concepts will then be applied later to special frames adapted to surfaces.

3.1.1 Definition A coordinate *frame* in $\mathbf{R}^n$ is an n-tuple of vector fields $\{e_1, \ldots, e_n\}$ which are linearly independent at each point $\mathbf{p}$ in the space.

In local coordinates $\{x^1, \dots, x^n\}$, we can always express the frame vectors as linear combinations of the standard basis vectors

$$e_i = \sum_{j=1}^{n} A^j i \frac{\partial}{\partial x^j} = \partial_j A^j{}_i, \tag{3.1}$$

where $\partial_j = \frac{\partial}{\partial x^j}$. Placing the basis vectors ∂_j on the left is done to be consistent with the summation convention, keeping in mind that the differential operators do not act on the matrix elements. We assume the matrix $A = (A^j{}_i)$ to be nonsingular at each point. In linear algebra, this concept is called a change of basis, the difference being that in our case, the transformation matrix A depends on the position. A frame field is called *orthonormal* if at each point,

$$< e_i, e_j >= \delta_{ij}. \tag{3.2}$$

Throughout this chapter, we will assume that all frame fields are orthonormal. Whereas this restriction is not necessary, it is convenient because it results in considerable simplification in computions.

3.1.2 Proposition If $\{e_1, \dots, e_n\}$ is an orthonormal frame, then the transformation matrix is orthogonal (ie, $AA^T = I$)
Proof The proof is by direct computation. Let $e_i = \partial_j A^j{}_i$. Then

$$\begin{aligned}
\delta_{ij} &= < e_i, e_j >, \\
&= < \partial_k A^k{}_i, \partial_l A^l{}_j >, \\
&= A^k_i A^l{}_j < \partial_k, \partial_l >, \\
&= A^k_i A^l{}_j \delta_{kl}, \\
&= A^k_i A_{kj}, \\
&= A^k_i (A^T)_{jk}.
\end{aligned}$$

Hence

$$\begin{aligned}
(A^T)_{jk} A^k_i &= \delta_{ij}, \\
(A^T)^j{}_k A^k_i &= \delta^j{}_i, \\
A^T A &= I.
\end{aligned}$$

Given a frame $\{e_i\}$, we can also introduce the corresponding dual coframe forms θ^i by requiring that

$$\theta^i(e_j) = \delta^i_j. \tag{3.3}$$

Since the dual coframe is a set of 1-forms, they can also be expressed in local coordinates as linear combinations

$$\theta^i = B^i{}_k dx^k.$$

It follows from equation(3.3), that

$$\begin{aligned}
\theta^i(e_j) &= B^i_k dx^k(\partial_l A^l{}_j), \\
&= B^i{}_k A^l{}_j dx^k(\partial_l), \\
&= B^i{}_k A^l{}_j \delta^k_l, \\
\delta^i{}_j &= B^i{}_k A^k_j.
\end{aligned}$$

Therefore, we conclude that $BA = I$, so $B = A^{-1} = A^T$. In other words, when the frames are orthonormal, we have

$$\begin{aligned}
e_i &= \partial_k A^k_i, \\
\theta^i &= A^i{}_k dx^k.
\end{aligned} \tag{3.4}$$

3.1.3 Example Consider the transformation from Cartesian to cylindrical coordinates:

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = z. \tag{3.5}$$

Using the chain rule for partial derivatives, we have

$$\begin{aligned}
\frac{\partial}{\partial r} &= \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}, \\
\frac{\partial}{\partial \theta} &= -r\sin\theta\frac{\partial}{\partial x} + r\cos\theta\frac{\partial}{\partial y}, \\
\frac{\partial}{\partial z} &= \frac{\partial}{\partial z}.
\end{aligned}$$

The vectors $\frac{\partial}{\partial r}$, and $\frac{\partial}{\partial z}$ are clearly unit vectors.

To make the vector $\frac{\partial}{\partial \theta}$ a unit vector, it suffices to divide it by its length r. We can then compute the dot products of each pair of vectors and easily verify that the quantities

$$e_1 = \frac{\partial}{\partial r}, \quad e_2 = \frac{1}{r}\frac{\partial}{\partial \theta}, \quad e_3 = \frac{\partial}{\partial z}, \tag{3.6}$$

are a triplet of mutually orthogonal unit vectors and thus constitute an orthonormal frame. The surfaces with constant value for the coordinates r, θ and z respectively, represent a set of mutually orthogonal surfaces at each point. The frame vectors at a point are normal to these surfaces as shown in figure 3.1. Physicists often refer to these frame vectors as $\{\hat{\mathbf{r}}, \hat{\theta}, \hat{\mathbf{z}}\}$, or as $\{e_r, e_\theta, e_z.\}$.

3.1.4 Example For spherical coordinates (2.30)

$$\begin{aligned}
x &= r\sin\theta\cos\phi, \\
y &= r\sin\theta\sin\phi, \\
z &= r\cos\theta,
\end{aligned}$$

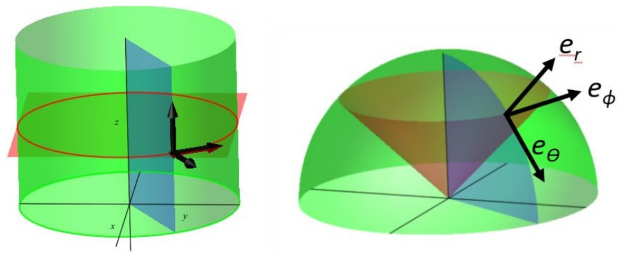


Fig. 3.1: Cylindrical and Spherical Frames.

the chain rule leads to

$$\begin{aligned}
\frac{\partial}{\partial r} &= \sin\theta\cos\phi\frac{\partial}{\partial x} + \sin\theta\sin\phi\frac{\partial}{\partial y} + \cos\theta\frac{\partial}{\partial z}, \\
\frac{\partial}{\partial \theta} &= r\cos\theta\cos\phi\frac{\partial}{\partial x} + r\cos\theta\sin\phi\frac{\partial}{\partial y} - r\sin\theta\frac{\partial}{\partial z}, \\
\frac{\partial}{\partial \phi} &= -r\sin\theta\sin\phi\frac{\partial}{\partial x} + r\sin\theta\cos\phi\frac{\partial}{\partial y}.
\end{aligned}$$

The vector $\frac{\partial}{\partial r}$ is of unit length but the other two need to be normalized. As before, all we need to do is divide the vectors by their magnitude. For $\frac{\partial}{\partial \theta}$, we divide by r and for $\frac{\partial}{\partial \phi}$, we divide by $r\sin\theta$. Taking the dot products of all pairs and using basic trigonometric identities, one can verify that we again obtain an orthonormal frame.

$$e_1 = e_r = \frac{\partial}{\partial r}, \quad e_2 = e_\theta = \frac{1}{r}\frac{\partial}{\partial \theta}, \quad e_3 = e_\varphi = \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}. \tag{3.7}$$

Furthermore, the frame vectors are normal to triply orthogonal surfaces, which in this case are spheres, cones and planes, as shown in figure 3.1. The fact that the chain rule in the two situations above leads to orthonormal frames is not coincidental. The results are related to the orthogonality of the level surfaces $x^i =$ constant. Since the level surfaces are orthogonal whenever they intersect, one expects the gradients of the surfaces to also be orthogonal. Transformations of this type are called triply orthogonal systems.

3.2 Curvilinear Coordinates

Orthogonal transformations, such as spherical and cylindrical coordinates, appear ubiquitously in mathematical physics, because the geometry of many problems in this discipline exhibit symmetry with respect to an axis or to the origin. In such situations, transformations to the appropriate coordinate system often result in considerable simplification of the field equations involved in the problem. It has been shown that the Laplace operator that appears in the potential, heat, wave, and Schrödinger field equations, is separable in exactly twelve orthogonal coordinate systems. A simple and efficient method to calculate the Laplacian in orthogonal coordinates can be implemented using differential forms.

3.2.1 Example In spherical coordinates the differential of arc length is given by (see equation 2.31) the metric:

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2.$$

Let

$$\begin{aligned} \theta^1 &= dr, \\ \theta^2 &= r d\theta, \\ \theta^3 &= r\sin\theta d\phi. \end{aligned} \tag{3.8}$$

Note that these three 1-forms constitute the dual coframe to the orthonormal frame derived in equation(3.7). Consider a scalar field $f = f(r, \theta, \phi)$. We now calculate the Laplacian of f in spherical coordinates using the methods of section 2.4.2. To do this, we first compute the differential df and express the result in terms of the coframe.

$$\begin{aligned} df &= \frac{\partial f}{\partial r} dr + \frac{\partial f}{\partial \theta} d\theta + \frac{\partial f}{\partial \phi} d\phi, \\ &= \frac{\partial f}{\partial r}\theta^1 + \frac{1}{r}\frac{\partial f}{\partial \theta}\theta^2 + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\theta^3. \end{aligned}$$

The components df in the coframe represent the gradient in spherical coordinates. Continuing with the scheme of section 2.4.2, we next apply the Hodge $\star$ operator. Then, we rewrite the resulting 2-form in terms of wedge products of coordinate differentials so that we can apply the definition of the exterior derivative.

$$\begin{aligned} \star df &= \frac{\partial f}{\partial r}\theta^2 \wedge \theta^3 - \frac{1}{r}\frac{\partial f}{\partial \theta}\theta^1 \wedge \theta^3 + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\theta^1 \wedge \theta^2, \\ &= r^2 \sin\theta \frac{\partial f}{\partial r} d\theta \wedge d\phi - r\sin\theta \frac{1}{r}\frac{\partial f}{\partial \theta} dr \wedge d\phi + r\frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi} dr \wedge d\theta, \\ &= r^2 \sin\theta \frac{\partial f}{\partial r} d\theta \wedge d\phi - \sin\theta\, \frac{\partial f}{\partial \theta} dr \wedge d\phi + \frac{1}{\sin\theta}\frac{\partial f}{\partial \phi} dr \wedge d\theta, \\ d \star df &= \frac{\partial}{\partial r}(r^2 \sin\theta \frac{\partial f}{\partial r}) dr \wedge d\theta \wedge d\phi - \frac{\partial}{\partial \theta}(\sin\theta \frac{\partial f}{\partial \theta}) d\theta \wedge dr \wedge d\phi + \\ &\quad \frac{1}{\sin\theta}\frac{\partial}{\partial \phi}(\frac{\partial f}{\partial \phi}) d\phi \wedge dr \wedge d\theta, \\ &= \left[\sin\theta \frac{\partial}{\partial r}(r^2 \frac{\partial f}{\partial r}) + \frac{\partial}{\partial \theta}(\sin\theta \frac{\partial f}{\partial \theta}) + \frac{1}{\sin\theta}\frac{\partial^2 f}{\partial \phi^2}\right] dr \wedge d\theta \wedge d\phi. \end{aligned}$$

Finally, rewriting the differentials back in terms of the coframe, we get

$$d \star df = \frac{1}{r^2 \sin\theta}\left[\sin\theta \frac{\partial}{\partial r}(r^2 \frac{\partial f}{\partial r}) + \frac{\partial}{\partial \theta}(\sin\theta \frac{\partial f}{\partial \theta}) + \frac{1}{\sin\theta}\frac{\partial^2 f}{\partial \phi^2}\right] \theta^1 \wedge \theta^2 \wedge \theta^3.$$

Therefore, the Laplacian of f is given by

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2\frac{\partial f}{\partial r}\right] + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial f}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}\right]. \tag{3.9}$$

The derivation of the expression for the spherical Laplacian by differential forms is elegant and leads naturally to the operator in Sturm-Liouville form.

The process above can be carried out for general orthogonal transformations. A change of coordinates $x^i = x^i(u^k)$ leads to an orthogonal transformation if in the new coordinate system u^k, the line metric

$$ds^2 = g_{11}(du^1)^2 + g_{22}(du^2)^2 + g_{33}(du^3)^2 \tag{3.10}$$

only has diagonal entries. In this case, we choose the coframe

$$\begin{aligned} \theta^1 &= \sqrt{g_{11}}du^1 = h_1 du^1, \\ \theta^2 &= \sqrt{g_{22}}du^2 = h_2 du^2, \\ \theta^3 &= \sqrt{g_{33}}du^3 = h_3 du^3. \end{aligned}$$

Classically, the quantities $\{h_1, h_2, h_3\}$ are called the weights. Please note that, in the interest of connecting to classical terminology, we have exchanged two indices for one and this will cause small discrepancies with the index summation convention. We will revert to using a summation symbol when these discrepancies occur. To satisfy the duality condition $\theta^i(e_j) = \delta^i_j$, we must choose the corresponding frame vectors e_i as follows:

$$\begin{aligned} e_1 &= \frac{1}{\sqrt{g_{11}}}\frac{\partial}{\partial u^1} = \frac{1}{h_1}\frac{\partial}{\partial u^1}, \\ e_2 &= \frac{1}{\sqrt{g_{22}}}\frac{\partial}{\partial u^2} = \frac{1}{h_2}\frac{\partial}{\partial u^2}, \\ e_3 &= \frac{1}{\sqrt{g_{33}}}\frac{\partial}{\partial u^3} = \frac{1}{h_3}\frac{\partial}{\partial u^3}. \end{aligned}$$

Gradient. Let $f = f(x^i)$ and $x^i = x^i(u^k)$. Then

$$\begin{aligned} df &= \frac{\partial f}{\partial x^k}dx^k, \\ &= \frac{\partial f}{\partial u^i}\frac{\partial u^i}{\partial x^k}dx^k, \\ &= \frac{\partial f}{\partial u^i}du^i, \\ &= \sum_i \frac{1}{h^i}\frac{\partial f}{\partial u^i}\theta^i. \\ &= e_i(f)\theta^i. \end{aligned}$$

As expected, the components of the gradient in the coframe θ^i are the just the frame vectors.

$$\nabla = \left(\frac{1}{h_1}\frac{\partial}{\partial u^1}, \frac{1}{h_2}\frac{\partial}{\partial u^2}, \frac{1}{h_3}\frac{\partial}{\partial u^3}\right). \tag{3.11}$$

Curl. Let $F = (F_1, F_2, F_3)$ be a classical vector field. Construct the corresponding 1-form $F = F_i\theta^i$ in the coframe. We calculate the curl using the dual of the exterior derivative.

$$
\begin{aligned}
F &= F_1\theta^1 + F_2\theta^2 + F_3\theta^3, \\
&= (h_1F_1)du^1 + (h_2F_2)du^2 + (h_3F_3)du^3, \\
&= (hF)_i du^i, \quad \text{where} \quad (hF)_i = h_iF_i. \\
dF &= \frac{1}{2}\left[\frac{\partial(hF)_i}{\partial u^j} - \frac{\partial(hF)_j}{\partial u^i}\right] du^i \wedge du^j, \\
&= \frac{1}{h_ih_j}\left[\frac{\partial(hF)_i}{\partial u^j} - \frac{\partial(hF)_j}{\partial u^i}\right] d\theta^i \wedge d\theta^j. \\
\star dF &= \epsilon^{ij}{}_k \left[\frac{1}{h_ih_j}[\frac{\partial(hF)_i}{\partial u^j} - \frac{\partial(hF)_j}{\partial u^i}]\right] \theta^k = (\nabla \times F)_k\theta^k.
\end{aligned}
$$

Thus, the components of the curl are

$$
\left(\frac{1}{h_2h_3}[\frac{\partial(h_3F_3)}{\partial u^2} - \frac{\partial(h_2F_2)}{\partial u^3}], \frac{1}{h_1h_3}[\frac{\partial(h_3F_3)}{\partial u^1} - \frac{\partial(h_1F_1)}{\partial u^3}], \frac{1}{h_1h_2}[\frac{\partial(h_1F_1)}{\partial u^2} - \frac{\partial(h_2F_2)}{\partial u^1}]\right).
$$

Divergence. As before, let $F = F_i\theta^i$ and recall that $\nabla \cdot F = \star d \star F$. The computation yields

$$
\begin{aligned}
F &= F_1\theta^1 + F_2\theta^2 + F_3\theta^3 \\
\star F &= F_1\theta^2 \wedge \theta^3 + F_2\theta^3 \wedge \theta^1 + F_3\theta^1 \wedge \theta^2 \\
&= (h_2h_3F_1)du^2 \wedge du^3 + (h_1h_3F_2)du^3 \wedge du^1 + (h_1h_2F_3)du^1 \wedge du^2 \\
d \star dF &= \left[\frac{\partial(h_2h_3F_1)}{\partial u^1} + \frac{\partial(h_1h_3F_2)}{\partial u^2} + \frac{\partial(h_1h_2F_3)}{\partial u^3}\right] du^1 \wedge du^2 \wedge du^3.
\end{aligned}
$$

Therefore,

$$
\nabla \cdot F = \star d \star F = \frac{1}{h_1h_2h_3}\left[\frac{\partial(h_2h_3F_1)}{\partial u^1} + \frac{\partial(h_1h_3F_2)}{\partial u^2} + \frac{\partial(h_1h_2F_3)}{\partial u^3}\right]. \tag{3.12}
$$

3.3 Covariant Derivative

In this section we introduce a generalization of directional derivatives. The directional derivative measures the rate of change of a function in the direction of a vector. We seek a quantity which measures the rate of change of a vector field in the direction of another.

3.3.1 Definition Given a pair (X, Y) of arbitrary vector field in $\mathbf{R}^n$, we associate a new vector field $\overline{\nabla}_X Y$, so that $\overline{\nabla}_X : \mathscr{X}(\mathbf{R}^n) \longrightarrow \mathscr{X}(\mathbf{R}^n)$. The quantity $\overline{\nabla}$ called a *Koszul connection* if it satisfies the following properties:

1. $\overline{\nabla}_{fX}(Y) = f\overline{\nabla}_X Y$,
2. $\overline{\nabla}_{(X_1+X_2)}Y = \overline{\nabla}_{X_1}Y + \overline{\nabla}_{X_2}Y$,

3. $\overline{\nabla}_X(Y_1+Y_2) = \overline{\nabla}_X Y_1 + \overline{\nabla}_X Y_2$,

4. $\overline{\nabla}_X fY = X(f)Y + f\overline{\nabla}_X Y$,

for all vector fields $X, X_1, X_2, Y, Y_1, Y_2 \in \mathscr{X}(\mathbf{R}^n)$ and all smooth functions f. Implicit in the properties, we set $\overline{\nabla}_X f = X(f)$. The definition states that the map $\overline{\nabla}_X$ is linear on X but behaves as a linear derivation on Y. For this reason, the quantity $\overline{\nabla}_X Y$ is called the *covariant derivative* of Y in the direction of X.

3.3.2 Proposition Let $Y = f^i \frac{\partial}{\partial x^i}$ be a vector field in $\mathbf{R}^n$, and let X another C^∞ vector field. Then the operator given by

$$\overline{\nabla}_X Y = X(f^i)\frac{\partial}{\partial x^i} \tag{3.13}$$

defines a Koszul connection.

Proof The proof just requires verification that the four properties above are satisfied, and it is left as an exercise.

The operator defined in this proposition is the standard connection compatible with the Euclidean metric. The action of this connection on a vector field Y yields a new vector field whose components are the directional derivatives of the components of Y.

3.3.3 Example Let

$$X = x\frac{\partial}{\partial x} + xz\frac{\partial}{\partial y},\; Y = x^2\frac{\partial}{\partial x} + xy^2\frac{\partial}{\partial y}.$$

Then,

$$\begin{aligned}
\overline{\nabla}_X Y &= X(x^2)\frac{\partial}{\partial x} + X(xy^2)\frac{\partial}{\partial y},\\
&= [x\frac{\partial}{\partial x}(x^2) + xz\frac{\partial}{\partial y}(x^2)]\frac{\partial}{\partial x} + [x\frac{\partial}{\partial x}(xy^2) + xz\frac{\partial}{\partial y}(xy^2)]\frac{\partial}{\partial y},\\
&= 2x^2\frac{\partial}{\partial x} + (xy^2 + 2x^2yz)\frac{\partial}{\partial y}.
\end{aligned}$$

3.3.4 Definition A Koszul connection $\overline{\nabla}_X$ is compatible with the metric $g(Y,Z)$ if

$$\overline{\nabla}_X < Y, Z > = < \overline{\nabla}_X Y, Z > + < Y, \overline{\nabla}_X Z > . \tag{3.14}$$

if $F : \mathbf{R}^n \to \mathbf{R}^n$ is an isometry so that $< F_*X, F_*Y > = < X, Y >$, then it is connection preserving in the sense

$$F_*(\overline{\nabla}_X Y) = \overline{\nabla}_{F_*X} F_*y. \tag{3.15}$$

In Euclidean space, the components of the standard frame vectors are constant, and thus their rates of change in any direction vanish. Let e_i be arbitrary frame field with dual forms θ^i. The covariant derivatives of the frame vectors in the directions of a vector X will in general yield new vectors. The new vectors must be linear combinations of the basis vectors as follows:

$$\begin{aligned}
\overline{\nabla}_X e_1 &= \omega^1{}_1(X)e_1 + \omega^2{}_1(X)e_2 + \omega^3{}_1(X)e_3, \\
\overline{\nabla}_X e_2 &= \omega^1{}_2(X)e_1 + \omega^2{}_2(X)e_2 + \omega^3{}_2(X)e_3, \\
\overline{\nabla}_X e_3 &= \omega^1{}_3(X)e_1 + \omega^2{}_3(X)e_2 + \omega^3{}_3(X)e_3. \qquad (3.16)
\end{aligned}$$

The coefficients can be more succinctly expressed using the compact index notation,

$$\overline{\nabla}_X e_i = e_j \omega^j{}_i(X). \qquad (3.17)$$

It follows immediately that

$$\omega^j{}_i(X) = \theta^j(\overline{\nabla}_X e_i). \qquad (3.18)$$

Equivalently, one can take the inner product of both sides of equation (3.17) with e_k to get

$$\begin{aligned}
< \overline{\nabla}_X e_i, e_k > &= < e_j \omega^j{}_i(X), e_k > \\
&= \omega^j{}_i(X) < e_j, e_k > \\
&= \omega^j{}_i(X) g_{jk}
\end{aligned}$$

Hence,

$$< \overline{\nabla}_X e_i, e_k >= \omega_{ki}(X) \qquad (3.19)$$

The left-hand side of the last equation is the inner product of two vectors, so the expression represents an array of functions. Consequently, the right-hand side also represents an array of functions. In addition, both expressions are linear on X, since by definition, $\overline{\nabla}_X$ is linear on X. We conclude that the right-hand side can be interpreted as a matrix in which each entry is a 1-forms acting on the vector X to yield a function. The matrix valued quantity $\omega^i{}_j$ is called the *connection form*. Sacrificing some inconsistency with the formalism of differential forms for the sake of connecting to classical notation, we sometimes write the above equation as

$$< d\mathbf{e}_i, \mathbf{e}_k >= \omega_{ki}, \qquad (3.20)$$

where $\{\mathbf{e}_i\}$ are vector calculus vectors forming an orthonormal basis.

3.3.5 Definition Let $\overline{\nabla}_X$ be a Koszul connection and let $\{e_i\}$ be a frame. The *Christoffel* symbols associated with the connection in the given frame are the functions $\Gamma^k{}_{ij}$ given by

$$\overline{\nabla}_{e_i} e_j = \Gamma^k{}_{ij} e_k \qquad (3.21)$$

The Christoffel symbols are the coefficients that give the representation of the rate of change of the frame vectors in the direction of the frame vectors themselves. Many physicists therefore refer to the Christoffel symbols as the connection, resulting in possible confusion. The precise relation between the Christoffel symbols and the connection 1-forms is captured by the equations,

$$\omega^k{}_i(e_j) = \Gamma^k{}_{ij}, \qquad (3.22)$$

or equivalently

$$\omega^k_{\ i} = \Gamma^k_{\ ij}\theta^j. \tag{3.23}$$

In a general frame in $\mathbf{R}^n$ there are n^2 entries in the connection 1-form and n^3 Christoffel symbols. The number of independent components is reduced if one assumes that the frame is orthonormal.
If $T = T^i e_i$ is a general vector field, then

$$\begin{aligned}\overline{\nabla}_{e_j} T &= \overline{\nabla}_{e_j}(T^i e_i)\\ &= T^i_{\ ,j}\, e_i + T^i \Gamma^k_{\ ji} e_k\\ &= (T^i_{\ ,j} + T^k \Gamma^i_{\ jk}) e_i,\end{aligned} \tag{3.24}$$

which is denoted classically as the covariant derivative

$$T^i_{\ \|j} = T^i_{\ ,j} + \Gamma^i_{\ jk} T^k. \tag{3.25}$$

Here, the comma in the subscript means regular derivative. The equation above is also commonly written as

$$\overline{\nabla}_{e_j} T^i = \overline{\nabla}_j T^i = T^i_{\ ,j} + \Gamma^i_{\ jk} T^k,$$

We should point out the accepted but inconsistent use of terminology. What is meant by the notation $\overline{\nabla}_j T^i$ above is not the covariant derivative of the vector but the tensor components of the covariant derivative of the vector; one more reminder that most physicists conflate a tensor with its components.

3.3.6 Proposition Let $\{e_i\}$ be an orthonormal frame and $\overline{\nabla}_X$ be a Koszul connection compatible with the metric . Then

$$\omega_{ji} = -\omega_{ij} \tag{3.26}$$

Proof Since it is given that $< e_i, e_j >= \delta_{ij}$, we have

$$\begin{aligned} 0 &= \overline{\nabla}_X < e_i, e_j >,\\ &= < \overline{\nabla}_X e_i, e_j > + < e_i, \overline{\nabla}_X e_j >,\\ &= < \omega^k_{\ i} e_k, e_j > + < e_i, \omega^k_{\ j} e_k >,\\ &= \omega^k_{\ i} < e_k, e_j > + \omega^k_{\ j} < e_i, e_k >,\\ &= \omega^k_{\ i} g_{kj} + \omega^k_{\ j} g_{ik},\\ &= \omega_{ji} + \omega_{ij}.\end{aligned}$$

thus proving that ω is indeed antisymmetric.
The covariant derivative can be extended to the full tensor field $\mathscr{T}^r_s(\mathbf{R}^n)$ by requiring that

a) $\overline{\nabla}_X : \mathscr{T}^r_s(\mathbf{R}^n) \to \mathscr{T}^r_s(\mathbf{R}^n)$,
b) $\overline{\nabla}_X(T_1 \otimes T_2) = \overline{\nabla}_X T_1 \otimes T_2 + T_1 \otimes \overline{\nabla}_X T_2$,
c) $\overline{\nabla}_X$ commutes with all contractions, $\overline{\nabla}_X(CT) = C(\overline{\nabla}_X)$.

Let us compute the covariant derivative of a one-form ω with respect to vector field X. The contraction of $\omega \otimes Y$ is the function $i_Y\omega = \omega(Y)$. Taking the covariant derivative, we have,

$$\begin{aligned}\overline{\nabla}_X(\omega(Y)) &= (\overline{\nabla}_X\omega)(Y) - \omega(\overline{\nabla}_X Y),\\ X(\omega(Y)) &= (\overline{\nabla}_X\omega)(Y) - \omega(\overline{\nabla}_X Y).\end{aligned}$$

Hence, the coordinate-free formula for the covariant derivative of one-form is,

$$(\overline{\nabla}_X\omega)(Y) = X(\omega(Y)) - \omega(\overline{\nabla}_X Y). \tag{3.27}$$

Let θ^i be the dual forms to e_i. We have

$$\overline{\nabla}_X(\theta^i \otimes e_j) = \overline{\nabla}_X\theta^i \otimes e_j + \theta^i \otimes \overline{\nabla}_X e_j.$$

The contraction of $i_{e_j}\theta^i = \theta^i(e_j) = \delta^i_j$, Hence, taking the contraction of the equation above, we see that the left-hand side becomes 0, and we conclude that,

$$(\overline{\nabla}_X\theta^i)(e_j) = -\theta^i(\overline{\nabla}_X e_j). \tag{3.28}$$

Let $\omega = T_i\theta^i$. Then,

$$\begin{aligned}(\overline{\nabla}_X\omega)(e_j) &= (\overline{\nabla}_X(T_i\theta^i))(e_j),\\ &= X(T_i)\theta^i(e_j) + T_i(\overline{\nabla}_X\theta^i)(e_j),\\ &= X(T_j) - T_i\theta^i(\overline{\nabla}_X e_j).\end{aligned} \tag{3.29}$$

If we now set $X = e_k$, we get,

$$\begin{aligned}(\overline{\nabla}_{e_k}\omega)(e_i) &= T_{j,k} - T_i\theta^i(\Gamma^l{}_{kj}e_l),\\ &= T_{j,k} - T_i\delta^i_l\Gamma^l{}_{kj},\\ &= T_{j,k} - \Gamma^i{}_{jk}T_i.\end{aligned}$$

Classically, we write

$$\overline{\nabla}_k T_j = T_{j\|k} = T_{j,k} - \Gamma^i{}_{jk}T_i. \tag{3.30}$$

In general, let T be a tensor of type $\binom{r}{s}$,

$$T = T^{i_1,\dots,i_r}_{j_1,\dots,j_s} e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes \theta^{j_1} \otimes \dots \theta^{j_s}. \tag{3.31}$$

Since we know how to take the covariant derivative of a function, a vector, and a one form, we can use Leibnitz rule for tensor products and property of the covariant derivative commuting with contractions, to get by induction, a formula for the covariant derivative of an $\binom{r}{s}$-tensor,

$$\begin{aligned}(\overline{\nabla}_X T)(\theta^{i_1},\dots,\theta^{i_r},e_{j_1},\dots e_{j_s}) &= X(T(\theta^{i_1},\dots,\theta^{i_r},e_{j_1},\dots,e_{j_s}))\\ &- T(\overline{\nabla}_X\theta^{i_1},\dots,\theta^{i_r},e_{j_1},\dots,e_{j_s}) - \cdots - T(\theta^{i_1},\dots,\overline{\nabla}_X\theta^{i_r},e_{j_1},\dots,e_{j_s})\dots\\ &- T(\theta^{i_1},\dots,\theta^{i_r},\overline{\nabla}_X e_{j_1},\dots,e_{j_s}) - \cdots - T(\theta^{i_1},\dots,\theta^{i_r},e_{j_1},\dots,\overline{\nabla}_X e_{j_s}).\end{aligned} \tag{3.32}$$

The covariant derivative picks up a term with a positive Christoffel symbol factor for each contravariant index and a term with a negative Christoffel symbol factor for each covariant index. Thus, for example, for a $\binom{1}{2}$ tensor, be components of the covariant derivative in classical notation are

$$\nabla_l T^i{}_{jk} = T^i{}_{jk\|l} = T^i{}_{jk,l} + \Gamma^i{}_{lh} T^h{}_{jk} - \Gamma^h{}_{jl} T^i{}_{hk} - \Gamma^h{}_{kl} T^i{}_{hj}. \tag{3.33}$$

In particular, if g is the metric tensor and X, Y, Z vector fields, we have

$$(\overline{\nabla}_X g)(Y, Z) = X(g(X, Y)) - g(\overline{\nabla}_X Y, Z) - g(X, \overline{\nabla}_X Z).$$

Thus, if we impose the condition $\overline{\nabla}_X g = 0$, the equation above reads

$$\overline{\nabla}_X < Y, Z > = < \overline{\nabla}_X Y, Z > + < Y, \overline{\nabla}_X Z > . \tag{3.34}$$

In other words, a connection is compatible with the metric just means that the metric is covariantly constant along any vector field.

In an orthonormal frame in $\mathbf{R}^n$ the number of independent coefficients of the connection 1-form is $(1/2)n(n-1)$ since by antisymmetry, the diagonal entries are zero, and one only needs to count the number of entries in the upper triangular part of the $n \times n$ matrix ω_{ij}. Similarly, the number of independent Christoffel symbols gets reduced to $(1/2)n^2(n-1)$. Raising one index with g^{ij}, we find that $\omega^i{}_j$ is also antisymmetric, so in $\mathbf{R}^3$ the connection equations become

$$\overline{\nabla}_X [e_1, e_2, e_3] = [e_1, e_2, e_3] \begin{bmatrix} 0 & \omega^1{}_2(X) & \omega^1{}_3(X) \\ -\omega^1{}_2(X) & 0 & \omega^2{}_3(X) \\ -\omega^1{}_3(X) & -\omega^2{}_3(X) & 0 \end{bmatrix} \tag{3.35}$$

Comparing the Frenet frame equation (1.39), we notice the obvious similarity to the general frame equations above. Clearly, the Frenet frame is a special case in which the basis vectors have been adapted to a curve, resulting in a simpler connection in which some of the coefficients vanish. A further simplification occurs in the Frenet frame, since in this case the equations represent the rate of change of the frame only along the direction of the curve rather than an arbitrary direction vector X. To elaborate on this transition from classical to modern notation, consider a unit speed curve $\beta(s)$. Then, as we discussed in section 1.15, we associate with the classical tangent vector $\mathbf{T} = \frac{d\mathbf{x}}{ds}$ the vector field $T = \beta'(s) = \frac{dx^i}{ds}\frac{\partial}{\partial x^i}$. Let $W = W(\beta(s)) = w^j(s)\frac{\partial}{\partial x^j}$ be an arbitrary vector field constrained to the curve. The rate of change of W along the curve is given by

$$\begin{aligned}
\overline{\nabla}_T W &= \overline{\nabla}_{(\frac{dx^i}{ds}\frac{\partial}{\partial x^i})}(w^j \frac{\partial}{\partial x^j}), \\
&= \frac{dx^i}{ds}\overline{\nabla}_{\frac{\partial}{\partial x^i}}(w^j \frac{\partial}{\partial x^j}) \\
&= \frac{dx^i}{ds}\frac{\partial w^j}{\partial x^i}\frac{\partial}{\partial x^j} \\
&= \frac{dw^j}{ds}\frac{\partial}{\partial x^j} \\
&= W'(s).
\end{aligned}$$

3.4 Cartan Equations

Perhaps, the most important contribution to the development of modern differential geometry, is the work of Cartan, culminating into the famous equations of structure discussed in this chapter.

First Structure Equation

3.4.1 Theorem Let $\{e_i\}$ be a frame with connection ${\omega^i}_j$ and dual coframe θ^i. Then

$$\Theta^i \equiv d\theta^i + {\omega^i}_j \wedge \theta^j = 0. \tag{3.36}$$

Proof Let

$$e_i = \partial_j {A^j}_i$$

be a frame, and let θ^i be the corresponding coframe. Since $\theta^i(e_j)$, we have

$$\theta^i = {(A^{-1})^i}_j dx^j.$$

Let X be an arbitrary vector field. Then

$$\begin{aligned}
\overline{\nabla}_X e_i &= \overline{\nabla}_X(\partial_j A^j_i). \\
e_j\omega^j_i(X) &= \partial_j X(A^j_i), \\
&= \partial_j d(A^j_i)(X), \\
&= e_k(A^{-1})^k_j d(A^j_i)(X). \\
\omega^k_i(X) &= (A^{-1})^k_j d(A^j_i)(X).
\end{aligned}$$

Hence,

$$\omega^k_i = (A^{-1})^k_j d(A^j_i),$$

or, in matrix notation,

$$\omega = A^{-1}dA. \tag{3.37}$$

On the other hand, taking the exterior derivative of θ^i, we find that

$$\begin{aligned}
d\theta^i &= d(A^{-1})^i{}_j \wedge dx^j, \\
&= d(A^{-1})^i{}_j \wedge A^j{}_k \theta^k, \\
d\theta &= d(A^{-1})A \wedge \theta.
\end{aligned}$$

However, since $A^{-1}A = I$, we have $d(A^{-1})A = -A^{-1}dA = -\omega$, hence

$$d\theta = -\omega \wedge \theta. \tag{3.38}$$

In other words

$$d\theta^i + \omega^i{}_j \wedge \theta^j = 0.$$

3.4.2 Example $SO(2, \mathbf{R})$

Consider the polar coordinates part of the transformation in equation 3.5. Then the frame equations 3.6 in matrix form are given by:

$$\begin{bmatrix} e_1, e_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3.39}$$

Thus, the attitude matrix

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.40}$$

is a rotation matrix in $\mathbf{R}^2$. The set of all such matrices forms a continuous group (*Lie group*) called $SO(2, \mathbf{R})$. In such cases, the matrix

$$\omega = A^{-1}dA \tag{3.41}$$

in equation 3.37 is called the *Maurer-Cartan* form of the group. An easy computation shows that for the rotation group $SO(2)$, the connection form is

$$\omega = \begin{bmatrix} 0 & -d\theta \\ d\theta & 0 \end{bmatrix}. \tag{3.42}$$

Second Structure Equation

Let θ^i be a coframe in $\mathbf{R}^n$ with connection $\omega^i{}_j$. Taking the exterior derivative of the first equation of structure and recalling the properties (2.66), we get

$$\begin{aligned}
d(d\theta^i) + d(\omega^i{}_j \wedge \theta^j) = 0, \\
d\omega^i{}_j \wedge \theta^j - \omega^i{}_j \wedge d\theta^j = 0.
\end{aligned}$$

Substituting recursively from the first equation of structure, we get

$$\begin{aligned} d\omega^i{}_j \wedge \theta^j - \omega^i{}_j \wedge (-\omega^j{}_k \wedge \theta^k) &= 0, \\ d\omega^i{}_j \wedge \theta^j + \omega^i{}_k \wedge \omega^k{}_j \wedge \theta^j &= 0, \\ (d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j) \wedge \theta^j &= 0, \\ d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j &= 0. \end{aligned}$$

3.4.3 Definition The curvature Ω of a connection ω is the matrix valued 2-form,

$$\Omega^i{}_j \equiv d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j. \tag{3.43}$$

3.4.4 Theorem Let θ be a coframe with connection ω in $\mathbf{R}^n$. Then the curvature form vanishes:

$$\Omega = d\omega + \omega \wedge \omega = 0. \tag{3.44}$$

Proof Given that there is a non-singular matrix A such that $\theta = A^{-1}dx$ and $\omega = A^{-1}dA$, we have

$$d\omega = d(A^{-1}) \wedge dA.$$

On the other hand,

$$\begin{aligned} \omega \wedge \omega &= (A^{-1}dA) \wedge (A^{-1}dA), \\ &= -d(A^{-1})A \wedge A^{-1}dA, \\ &= -d(A^{-1})(AA^{-1}) \wedge dA, \\ &= -d(A^{-1}) \wedge dA. \end{aligned}$$

Therefore, $d\omega = -\omega \wedge \omega$.
There is a slight abuse of the wedge notation here. The connection ω is matrix valued, so the symbol $\omega \wedge \omega$ is really a composite of matrix and wedge multiplication.

3.4.5 Example Sphere frame
The frame for spherical coordinates 3.7 in matrix form is

$$[e_r, e_\theta, e_\phi] = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right] \begin{bmatrix} \sin\theta\cos\phi & \cos\theta\cos\phi & -\sin\phi \\ \sin\theta\sin\phi & \cos\theta\sin\phi & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{bmatrix}.$$

Hence,

$$A^{-1} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix},$$

and

$$dA = \begin{bmatrix} \cos\theta\cos\phi\,d\theta - \sin\theta\sin\phi\,d\phi & -\sin\theta\cos\phi\,d\theta - \cos\theta\sin\phi\,d\phi & -\cos\phi\,d\phi \\ \cos\theta\sin\phi\,d\theta + \sin\theta\cos\phi\,d\phi & -\sin\theta\sin\phi\,d\theta + \cos\theta\cos\phi\,d\phi & -\sin\phi\,d\phi \\ -\sin\theta\,d\theta & -\cos\theta\,d\theta & 0 \end{bmatrix}.$$

Since the $\omega = A^{-1}dA$ is antisymmetric, it suffices to compute:

$$\begin{aligned}
\omega^1{}_2 &= [-\sin^2\theta\cos^2\phi - \sin^2\theta\sin^2\phi\ \ - \cos^2\theta]\,d\theta \\
&\qquad + [\sin\theta\cos\theta\cos\phi\sin\phi - \sin\theta\cos\theta\cos\phi\sin\phi]\,d\phi, \\
&= -d\theta, \\
\omega^1{}_3 &= [-\sin\theta\cos^2\phi - \sin\theta\sin^2\phi]\,d\phi = -\sin\theta\,d\phi, \\
\omega^2{}_3 &= [-\cos\theta\cos^2\phi - \cos\theta\sin^2\phi]\,d\phi = -\cos\theta\,d\phi.
\end{aligned}$$

We conclude that the matrix-valued connection one form is

$$\omega = \begin{bmatrix} 0 & -d\theta & -\sin\theta\,d\phi \\ d\theta & 0 & -\cos\theta\,d\phi \\ \sin\theta\,d\phi & \cos\theta\,d\phi & 0 \end{bmatrix}.$$

A slicker computation of the connection form can be obtained by a method of educated guessing working directly from the structure equations. We have that the dual one forms are:

$$\begin{aligned}
\theta^1 &= dr, \\
\theta^2 &= r\,d\theta, \\
\theta^3 &= r\sin\theta\,d\phi.
\end{aligned}$$

Then

$$\begin{aligned}
d\theta^2 &= -d\theta\wedge dr, \\
&= -\omega^2{}_1\wedge\theta^1 - \omega^2{}_3\wedge\theta^3.
\end{aligned}$$

So, on a first iteration we guess that $\omega^2{}_1 = d\theta$. The component $\omega^2{}_3$ is not necessarily 0 because it might contain terms with $d\phi$. Proceeding in this manner, we compute:

$$\begin{aligned}
d\theta^3 &= \sin\theta\,dr\wedge d\phi + r\cos\theta\,d\theta\wedge d\phi, \\
&= -\sin\theta\,d\phi\wedge dr - \cos\theta\,d\phi\wedge r\,d\theta, \\
&= -\omega^3{}_1\wedge dr\wedge\theta^1 - \omega^3{}_2\wedge\theta^2.
\end{aligned}$$

Now we guess that $\omega^3{}_1 = \sin\theta\,d\phi$, and $\omega^3{}_2 = \cos\theta\,d\phi$. Finally, we insert these into the full structure equations and check to see if any modifications need to be made. In this case, the forms we have found are completely compatible with the first equation of structure, so these must be the forms. The second equations of structure are much more straight-forward to verify. For example

$$
\begin{aligned}
d\omega^2{}_3 &= d(-\cos\theta\, d\phi),\\
&= \sin\theta\, d\theta \wedge d\phi,\\
&= -d\theta \wedge (-\sin\theta\, d\phi),\\
&= -\omega^2{}_1 \wedge \omega^1{}_3.
\end{aligned}
$$

Change of Basis

We briefly explore the behavior of the quantities Θ^i and $\Omega^i{}_j$ under a change of basis. Let e_i be frame in $M = \mathbf{R^n}$ with dual forms θ^i, and let $\overline{e}_i$ be another frame related to the first frame by an invertible transformation.

$$
\overline{e}_i = e_j B^j{}_i, \tag{3.45}
$$

which we will write in matrix notation as $\overline{e} = eB$. Referring back to the definition of connections (3.17), we introduce the covariant differential $\overline{\nabla}$ which maps vectors into vector-valued forms,

$$
\overline{\nabla} : \Omega^0(M, TM) \to \Omega^1(M, TM)
$$

given by the formula

$$
\begin{aligned}
\overline{\nabla} e_i &= e_j \otimes \omega^j{}_i\\
&= e_j \omega^j{}_i\\
\overline{\nabla} e &= e\,\omega
\end{aligned} \tag{3.46}
$$

where, once again, we have simplified the equation by using matrix notation. This definition is elegant because it does not explicitly show the dependence on X in the connection (3.17). The idea of switching from derivatives to differentials is familiar from basic calculus. Consistent with equation 3.20, the vector calculus notation for equation 3.46 would be

$$
d\mathbf{e_i} = \mathbf{e_j}\, \omega^j{}_i. \tag{3.47}
$$

However, we point out that in the present context, the situation is much more subtle. The operator $\overline{\nabla}$ here maps a vector field to a matrix-valued tensor of rank $\binom{1}{1}$. Another way to view the covariant differential is to think of $\overline{\nabla}$ as an operator such that if e is a frame, and X a vector field, then $\overline{\nabla} e(X) = \overline{\nabla}_X e$. If f is a function, then $\overline{\nabla} f(X) = \overline{\nabla}_X f = df(X)$, so that $\overline{\nabla} f = df$. In other words, $\overline{\nabla}$ behaves like a covariant derivative on vectors, but like a differential on functions. The action of the covariant differential also extends to the entire tensor algebra, but we do not need that formalism for now, and we delay discussion to section 6.4 on connections on vector bundles. Taking the exterior differential of (3.45) and using (3.46) recursively, we get

$$\begin{aligned}
\overline{\nabla}\overline{e} &= (\overline{\nabla}e)B + e(dB)\\
&= e\,\omega B + e(dB)\\
&= \overline{e}B^{-1}\omega B + \overline{e}B^{-1}dB\\
&= \overline{e}[B^{-1}\omega B + B^{-1}dB]\\
&= \overline{e}\,\overline{\omega}
\end{aligned}$$

provided that the connection $\overline{\omega}$ in the new frame $\overline{e}$ is related to the connection ω by the transformation law, (See 6.62)

$$\overline{\omega} = B^{-1}\omega B + B^{-1}dB. \tag{3.48}$$

It should be noted than if e is the standard frame $e_i = \partial_i$ in $\mathbf{R}^n$, then $\overline{\nabla}e = 0$, so that $\omega = 0$. In this case, the formula above reduces to $\overline{\omega} = B^{-1}dB$, showing that the transformation rule is consistent with equation (3.37). The transformation law for the curvature forms is,

$$\overline{\Omega} = B^{-1}\Omega B. \tag{3.49}$$

A quantity transforming as in 3.49 is said to be a *tensorial form of adjoint type.*

3.4.6 Example Suppose that B is a change of basis consisting of a rotation by an angle θ about e_3. The transformation is a an isometry that can be represented by the orthogonal rotation matrix

$$B = \begin{bmatrix} \cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1 \end{bmatrix}. \tag{3.50}$$

Carrying out the computation for the change of basis 3.48, we find:

$$\begin{aligned}
\overline{\omega}^1{}_2 &= \omega^1{}_2 - d\theta,\\
\overline{\omega}^1{}_3 &= \cos\theta\,\omega^1{}_3 + \sin\theta\,\omega^2{}_3,\\
\overline{\omega}^2{}_3 &= -\sin\theta\,\omega^1{}_3 + \cos\theta\,\omega^2{}_3.
\end{aligned} \tag{3.51}$$

The $B^{-1}dB$ part of the transformation only affects the $\omega^1{}_2$ term, and the effect is just adding $d\theta$ much like the case of the Maurer-Cartan form for $SO(2)$ above.

Chapter 4

Theory of Surfaces

4.1 Manifolds

4.1.1 Definition A *coordinate chart* or coordinate patch in $M \subset \mathbf{R}^3$ is a differentiable map $\mathbf{x}$ from an open subset V of $\mathbf{R}^2$ onto a set $U \subset \mathcal{M}$.

$$\begin{aligned} \mathbf{x} : V \subset \mathbf{R}^2 &\longrightarrow \mathbf{R}^3 \\ (u,v) &\overset{\mathbf{x}}{\longmapsto} (x(u,v), y(u,v), z(u,v)) \end{aligned} \tag{4.1}$$

Each set $U = \mathbf{x}(V)$ is called a *coordinate neighborhood* of M. We require that

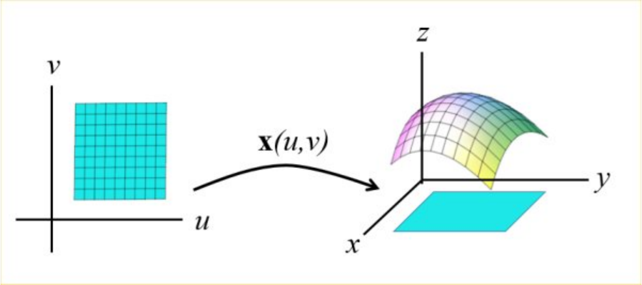


Fig. 4.1: Surface

the Jacobian of the map has maximal rank. In local coordinates, a coordinate chart is represented by three equations in two variables:

$$x^i = f^i(u^\alpha), \text{ where } i = 1,2,3, \quad \alpha = 1,2. \tag{4.2}$$

It will be convenient to use the tensor index formalism when appropriate, so that we can continue to take advantage of the Einstein summation convention. The assumption that the Jacobian $J = (\partial x^i/\partial u^\alpha)$ be of maximal rank allows one to invoke the implicit function theorem. Thus, in principle, one can locally solve for one of the coordinates, say x^3, in terms of the other two, to get an explicit function

$$x^3 = f(x^1, x^2). \tag{4.3}$$

The loci of points in $\mathbf{R}^3$ satisfying the equations $x^i = f^i(u^\alpha)$ can also be locally represented implicitly by an expression of the form

$$F(x^1, x^2, x^3) = 0. \tag{4.4}$$

4.1.2 Definition Let U_i and U_j be two coordinate neighborhoods of a point $p \in M$ with corresponding charts $\mathbf{x}(u^1, u^2) : V_i \longrightarrow U_i \subset \mathbf{R}^3$ and $\mathbf{y}(v^1, v^2) : V_j \longrightarrow U_j \subset \mathbf{R}^3$ with a non-empty intersection $U_i \cap U_j \neq \emptyset$. On the overlaps, the maps $\phi_{ij} = \mathbf{x}^{-1}\mathbf{y}$ are called transition functions or coordinate transformations. (See figure 4.2)

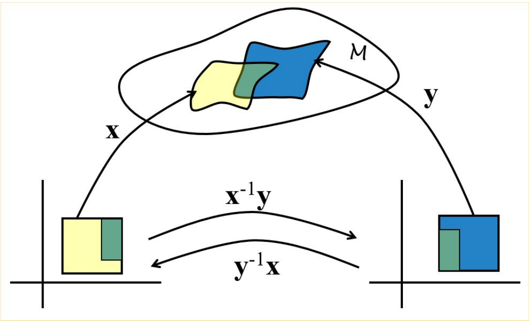


Fig. 4.2: Coordinate Charts

4.1.3 Definition A *differentiable manifold* of dimension 2, is a space M together with an indexed collection $\{U_\alpha\}_{\alpha \in I}$ of coordinate neighborhoods satisfying the following properties:

1. The neighborhoods $\{U_\alpha\}$ constitute an open cover M. That is, if $p \in M$, then p belongs to some chart.

2. For any pair of coordinate neighborhoods U_i and U_j with $U_i \cap U_j \neq \emptyset$, the transition maps ϕ_{ij} and their inverses are differentiable.

3. An indexed collection satisfying the conditions above is called an *atlas*. We require the atlas to be maximal in the sense that it contains all possible coordinate neighborhoods.

The overlapping coordinate patches represent different parametrizations for the same set of points in $\mathbf{R}^3$. Part (2) of the definition insures that on the overlap, the coordinate transformations are invertible. Part (3) is included for technical reasons, although in practice the condition is superfluous. A family of coordinate neighborhoods satisfying conditions (1) and (2) can always be extended to a maximal atlas. This can be shown from the fact that M inherits a subspace topology consisting of open sets which are defined by the intersection of open sets in $\mathbf{R}^3$ with M.

If the coordinate patches in the definition map from $\mathbf{R}^n$ to $\mathbf{R}^m$ $n < m$ we say that M is a n-dimensional *submanifold* embedded in $\mathbf{R}^m$. In fact, one could define an abstract manifold without the reference to the embedding space by starting with a topological space M that is locally Euclidean via homeomorphic coordinate patches and has a differentiable structure as in the definition above. However, it turns out that any differentiable manifold of dimension n can be embedded in $\mathbf{R}^{2n}$, as proved by Whitney in a theorem that is beyond the scope of these notes.

A 2-dimensional manifold embedded in $\mathbf{R}^3$ in which the transition functions are C^∞, is called a smooth surface. The first condition in the definition states that each coordinate neighborhood looks locally like a subset of $\mathbf{R}^2$. The second differentiability condition indicates that the patches are joined together smoothly as some sort of quilt. We summarize this notion by saying that a manifold is a space that is *locally Euclidean* and has a *differentiable structure*, so that the notion of differentiation makes sense. Of course, $\mathbf{R}^n$ is itself an n dimensional manifold.

The smoothness condition on the coordinate component functions $x^i(u^\alpha)$ implies that at any point $x^i(u_0^\alpha + h^\alpha)$ near a point $x^i(u_0^\alpha) = x^i(u_0, v_0)$, the functions admit a Taylor expansion

$$x^i(u_0^\alpha + h^\alpha) = x^i(u_0^\alpha) + h^\alpha \left(\frac{\partial x^i}{\partial u^\alpha}\right)_0 + \frac{1}{2!}h^\alpha h^\beta \left(\frac{\partial^2 x^i}{\partial u^\alpha \partial u^\beta}\right)_0 + \dots \tag{4.5}$$

Since the parameters u^α must enter independently, the Jacobian matrix

$$J \equiv \left[\frac{\partial x^i}{\partial u^\alpha}\right] = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{bmatrix}$$

must have maximal rank. At points where J has rank 0 or 1, there is a singularity in the coordinate patch.

4.1.4 Example Consider the local coordinate chart for the unit sphere obtained by setting $r = 1$ in the equations for spherical coordinates 2.30

$$\mathbf{x}(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta).$$

The vector equation is equivalent to three scalar functions in two variables:

$$\begin{aligned} x &= \sin\theta\cos\phi, \\ y &= \sin\theta\sin\phi, \\ z &= \cos\phi. \end{aligned} \tag{4.6}$$

Clearly, the surface represented by this chart is part of the sphere $x^2 + y^2 + z^2 = 1$. The chart cannot possibly represent the whole sphere because, although a sphere is locally Euclidean, (the earth is locally flat) there is certainly a topological difference between a sphere and a plane. Indeed, if one analyzes the coordinate chart carefully, one will note that at the North pole ($\theta = 0,\ z = 1$, the coordinates become singular. This happens because $\theta = 0$ implies that $x = y = 0$ regardless of the value of ϕ, so that the North pole has an infinite number of labels. In this coordinate patch, the Jacobian at the North Pole does not have maximal rank. To cover the entire sphere, one would need at least two coordinate patches. In fact, introducing an exactly analogous patch $\mathbf{y}(u.v)$ based on South pole would suffice, as long as in overlap around the equator functions $\mathbf{x}^{-1}\mathbf{y}$, and $\mathbf{y}^{-1}\mathbf{x}$ are smooth. One could conceive more elaborate coordinate patches such as those used in baseball and soccer balls.

The fact that it is required to have two parameters to describe a patch on a surface in $\mathbf{R}^3$ is a manifestation of the 2-dimensional nature of the surfaces. If one holds one of the parameters constant while varying the other, then the resulting 1-parameter equation describes a curve on the surface. Thus, for example, letting $\phi =$constant in equation (4.6), we get the equation of a meridian great circle.

4.1.5 Example *Surface of revolution*
Given a function $f(r)$, the coordinate chart

$$\mathbf{x}(r, \phi) = (r\cos\phi, r\sin\phi, f(r)) \qquad (4.7)$$

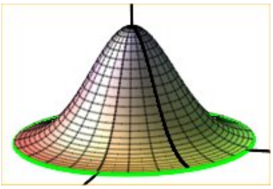


Fig. 4.3: Bell

represents a surface of revolution around the z-axis in which the cross section profile has the shape of the function. Horizontal cross-sections are circles of radius r. In figure 4.3, we have chosen the function $f(r) = e^{-r^2}$ to be a Gaussian, so the surface of revolution is bell-shaped. A lateral curve profile for $\phi = \pi/4$ is shown in black. We should point out that this parametrization of surfaces of revolution is fairly constraining because of the requirement of $z = f(r)$ to be a function. Thus, for instance, the parametrization will not work for surfaces of revolution generated by closed curves. In the next example, we illustrate how one easily get around this constraint.

4.1.6 Example *Torus*
Consider the surface of revolution generated by rotating a circle C of radius r around a parallel axis located a distance R from its center as shown in figure 4.4.
The resulting surface called a torus can be parametrized by the coordinate patch

$$\mathbf{x}(u, v) = ((R + r\cos u)\cos v, (R + r\cos u)\sin v, r\sin u). \qquad (4.8)$$

Here the angle u traces points around the z-axis, whereas the angle v traces points around the circle C. (At the risk of some confusion in notation, (the parameters in the figure are bold-faced; this is done solely for the purpose of visibility.) The projection of a point in the surface of the torus onto the xy-plane is located at a distance $(R + r\cos u)$ from the origin. Thus, the x

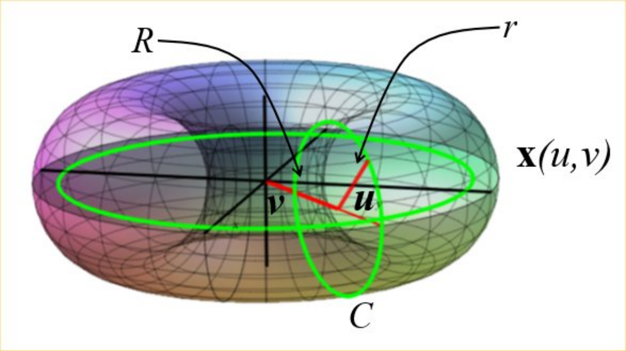


Fig. 4.4: Torus

and y coordinates of the point in the torus are just the polar coordinates of the projection of the point in the plane. The z-coordinate corresponds to the height of a right triangle with radius r and opposite angle u.

4.1.7 Example *Monge patch*
Surfaces in $\mathbf{R}^3$ are first introduced in vector calculus by a function of two variables $z = f(x, y)$. We will find it useful for consistency to use the obvious parametrization called an Monge patch

$$\mathbf{x}(u, v) = (u, v, f(u, v)). \tag{4.9}$$

4.1.8 Notation Given a parametrization of a surface in a local chart $\mathbf{x}(u, v) = \mathbf{x}(u^1, u^2) = \mathbf{x}(u^\alpha)$, we will denote the partial derivatives by any of the following notations:

$$\begin{aligned}
\mathbf{x}_u &= \mathbf{x}_1 = \frac{\partial \mathbf{x}}{\partial u}, & \mathbf{x}_{uu} &= \mathbf{x}_{11} = \frac{\partial^2 \mathbf{x}}{\partial u^2} \\
\mathbf{x}_v &= \mathbf{x}_2 = \frac{\partial \mathbf{x}}{\partial v}, & \mathbf{x}_{vv} &= \mathbf{x}_{22} = \frac{\partial^2 \mathbf{x}}{\partial v^2},
\end{aligned}$$

or more succinctly,

$$\mathbf{x}_\alpha = \frac{\partial \mathbf{x}}{\partial u^\alpha}, \quad \mathbf{x}_{\alpha\beta} = \frac{\partial^2 \mathbf{x}}{\partial u^\alpha \partial u^\beta} \tag{4.10}$$

4.2 The First Fundamental Form

Let $x^i(u^\alpha)$ be a local parametrization of a surface. Then, the Euclidean inner product in $\mathbf{R}^3$ induces an inner product in the space of tangent vectors at each point in the surface. This metric on the surface is obtained as follows:

$$\begin{aligned} dx^i &= \frac{\partial x^i}{\partial u^\alpha} du^\alpha, \\ ds^2 &= \delta_{ij} dx^i dx^j, \\ &= \delta_{ij} \frac{\partial x^i}{\partial u^\alpha} \frac{\partial x^j}{\partial u^\beta} du^\alpha du^\beta. \end{aligned}$$

Thus,

$$ds^2 = g_{\alpha\beta} du^\alpha du^\beta, \tag{4.11}$$

where

$$g_{\alpha\beta} = \delta_{ij} \frac{\partial x^i}{\partial u^\alpha} \frac{\partial x^j}{\partial u^\beta}. \tag{4.12}$$

We conclude that the surface, by virtue of being embedded in $\mathbf{R}^3$, inherits a natural metric (4.11) which we will call the *induced metric*. A pair $\{M, g\}$, where M is a manifold and $g = g_{\alpha\beta} du^\alpha \otimes du^\beta$ is a metric is called a *Riemannian manifold* if considered as an entity in itself, and a Riemannian submanifold of $\mathbf{R}^n$ if viewed as an object embedded in Euclidean space. An equivalent version of the metric (4.11) can be obtained by using a more traditional calculus notation:

$$\begin{aligned} d\mathbf{x} &= \mathbf{x}_u du + \mathbf{x}_v dv \\ ds^2 &= d\mathbf{x} \cdot d\mathbf{x} \\ &= (\mathbf{x}_u du + \mathbf{x}_v dv) \cdot (\mathbf{x}_u du + \mathbf{x}_v dv) \\ &= (\mathbf{x}_u \cdot \mathbf{x}_u) du^2 + 2(\mathbf{x}_u \cdot \mathbf{x}_v) du dv + (\mathbf{x}_v \cdot \mathbf{x}_v) dv^2. \end{aligned}$$

We can rewrite the last result as

$$ds^2 = E du^2 + 2F du dv + G dv^2, \tag{4.13}$$

where

$$\begin{aligned} E &= g_{11} = \mathbf{x}_u \cdot \mathbf{x}_u \\ F &= g_{12} = \mathbf{x}_u \cdot \mathbf{x}_v \\ &= g_{21} = \mathbf{x}_v \cdot \mathbf{x}_u \\ G &= g_{22} = \mathbf{x}_v \cdot \mathbf{x}_v. \end{aligned}$$

That is

$$g_{\alpha\beta} = \mathbf{x}_\alpha \cdot \mathbf{x}_\beta =< \mathbf{x}_\alpha, \mathbf{x}_\beta > .$$

4.2.1 Definition *First fundamental form*
The element of arc length,

$$ds^2 = g_{\alpha\beta} du^\alpha \otimes du^\beta, \tag{4.14}$$

is also called the first fundamental form.

We must caution the reader that this quantity is not a form in the sense of differential geometry since ds^2 involves the symmetric tensor product rather than the wedge product. The first fundamental form plays such a crucial role in the theory of surfaces that we will find it convenient to introduce a more modern version. Following the same development as in the theory of curves, consider a surface M defined locally by a function $q = (u^1, u^2) \longmapsto p = \alpha(u^1, u^2)$. We say that a quantity X_p is a tangent vector at a point $p \in M$ if X_p is a linear derivation on the space of C^∞ real-valued functions $\mathscr{F} = \{f | f : M \longrightarrow \mathbf{R}\}$ on the surface. The set of all tangent vectors at a point $p \in M$ is called the *tangent space* T_pM. As before, a vector field X on the surface is a smooth choice of a tangent vector at each point on the surface and the union of all tangent spaces is called the *tangent bundle* TM. Sections of the tangent bundle of M are consistently denoted by $\mathscr{X}(M)$. The coordinate chart map $\alpha : \mathbf{R}^2 \longrightarrow M \subset \mathbf{R}^3$ induces a *push-forward* map $\alpha_* : T\mathbf{R}^2 \longrightarrow TM$ which maps a vector V at each point in $T_q(\mathbf{R}^2)$ into a vector $V_{\alpha(q)} = \alpha_*(V_q)$ in $T_{\alpha(q)}M$, as illustrated in the diagram 4.5

$$\begin{tikzcd}
V \in T\mathbf{R}^2 \arrow[r, "\alpha_*"] \arrow[d] & T\mathcal{M} \arrow[d] & \\
q \in \mathbf{R}^2 \arrow[r, "\alpha"] & \mathcal{M} \in \mathbf{R}^3 \arrow[r, "f"] & \mathbf{R}
\end{tikzcd}$$

Fig. 4.5: Push-Forward

The action of the push-forward is defined by

$$\alpha_*(V)(f) \mid_{\alpha(q)} = V(f \circ \alpha) \mid_q . \tag{4.15}$$

Just as in the case of curves, when we revert back to classical notation to describe a surface as $x^i(u^\alpha)$, what we really mean is $(x^i \circ \alpha)(u^\alpha)$, where x^i are the coordinate functions in $\mathbf{R}^3$. Particular examples of tangent vectors on M are given by the push-forward of the standard basis of $T\mathbf{R}^2$. These tangent vectors which earlier we called $\mathbf{x}_\alpha$ are defined by

$$\alpha_*(\frac{\partial}{\partial u^\alpha})(f) \mid_{\alpha(u^\alpha)} = \frac{\partial}{\partial u^\alpha}(f \circ \alpha) \mid_{u^\alpha} .$$

In this formalism, the first fundamental form I is just the symmetric bilinear tensor defined by the induced metric,

$$I(X, Y) = g(X, Y) = < X, Y >, \tag{4.16}$$

where X and Y are any pair of vector fields in $\mathscr{X}(M)$.

Orthogonal Parametric Curves

Let V and W be vectors tangent to a surface M defined locally by a chart $\mathbf{x}(u^\alpha)$. Since the vectors $\mathbf{x}_\alpha$ span the tangent space of M at each point, the vectors V and W can be written as linear combinations,

$$\begin{aligned} V &= V^{\alpha}\mathbf{x}_{\alpha}, \\ W &= W^{\alpha}\mathbf{x}_{\alpha}. \end{aligned}$$

The functions V^{α} and W^{α} are the curvilinear components of the vectors. We can calculate the length and the inner product of the vectors using the induced Riemannian metric as follows:

$$\begin{aligned} \|V\|^2 &= < V, V > = < V^{\alpha}\mathbf{x}_{\alpha}, V^{\beta}\mathbf{x}_{\beta} > = V^{\alpha}V^{\beta} < \mathbf{x}_{\alpha}, \mathbf{x}_{\beta} >, \\ \|V\|^2 &= g_{\alpha\beta}V^{\alpha}V^{\beta}, \\ \|W\|^2 &= g_{\alpha\beta}W^{\alpha}W^{\beta}, \end{aligned}$$

and

$$\begin{aligned} < V, W > &= < V^{\alpha}\mathbf{x}_{\alpha}, W^{\beta}\mathbf{x}_{\beta} > = V^{\alpha}W^{\beta} < \mathbf{x}_{\alpha}, \mathbf{x}_{\beta} >, \\ &= g_{\alpha\beta}V^{\alpha}W^{\beta}. \end{aligned}$$

The angle θ subtended by the vectors V and W is the given by the equation

$$\begin{aligned} \cos\theta &= \frac{< V, W >}{\|V\| \cdot \|W\|}, \\ &= \frac{I(V,W)}{\sqrt{I(V,V)}\sqrt{I(W,W)}}, \\ &= \frac{g_{\alpha_1\beta_1}V^{\alpha_1}W^{\beta_1}}{\sqrt{g_{\alpha_2\beta_2}V^{\alpha_2}V^{\beta_2}}\sqrt{g_{\alpha_3\beta_3}W^{\alpha_3}W^{\beta_3}}}, \end{aligned} \tag{4.17}$$

where the numerical subscripts are needed for the α and β indices to comply with Einstein's summation convention.

Let $u^{\alpha} = \phi^{\alpha}(t)$ and $u^{\alpha} = \psi^{\alpha}(t)$ be two curves on the surface. Then the total differentials

$$du^{\alpha} = \frac{d\phi^{\alpha}}{dt}dt, \quad \text{and} \quad \delta u^{\alpha} = \frac{d\psi^{\alpha}}{dt}\delta t$$

represent infinitesimal tangent vectors (1.23) to the curves. Thus, the angle between two infinitesimal vectors tangent to two intersecting curves on the surface satisfies the equation:

$$\cos\theta = \frac{g_{\alpha_1\beta_1}du^{\alpha_1}\delta u^{\beta_1}}{\sqrt{g_{\alpha_2\beta_2}du^{\alpha_2}du^{\beta_2}}\sqrt{g_{\alpha_3\beta_3}\delta u^{\alpha_3}\delta u^{\beta_3}}}. \tag{4.18}$$

In particular, if the two curves happen to be the parametric curves, $u^1 =$ const. and $u^2 =$const., then along one curve we have $du^1 = 0$, with du^2 arbitrary, and along the second δu^1 is arbitrary and $\delta u^2 = 0$. In this case, the cosine of the angle subtended by the infinitesimal tangent vectors reduces to:

$$\cos\theta = \frac{g_{12}\delta u^1 du^2}{\sqrt{g_{11}(\delta u^1)^2}\sqrt{g_{22}(du^2)^2}} = \frac{g_{12}}{g_{11}g_{22}} = \frac{F}{\sqrt{EG}}. \tag{4.19}$$

A simpler way to obtain this result is to recall that parametric directions are given by $\mathbf{x}_u$ and $\mathbf{x}_v$, so

$$\cos\theta = \frac{<\mathbf{x}_u, \mathbf{x}_v>}{\|\mathbf{x}_u\| \cdot \|\mathbf{x}_v\|} = \frac{F}{\sqrt{EG}}. \tag{4.20}$$

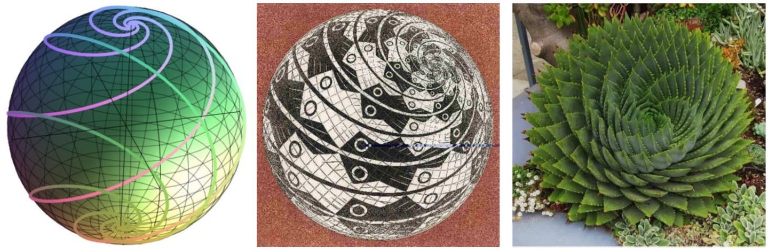
It follows immediately from the equation above that:

4.2.2 Proposition The parametric curves are orthogonal if $F = 0$.

Orthogonal parametric curves are an important class of curves, because locally the coordinate grid on the surface is similar to coordinate grids in basic calculus, such as in polar coordinates for which $ds^2 = dr^2 + r^2 d\theta^2$.

4.2.3 Examples a) Sphere

$$\begin{aligned}
\mathbf{x} &= (a\sin\theta\cos\phi, a\sin\theta\sin\phi, a\cos\theta),\\
\mathbf{x}_\theta &= (a\cos\theta\cos\phi, a\cos\theta\sin\phi, -a\sin\theta),\\
\mathbf{x}_\phi &= (-a\sin\theta\sin\phi, a\sin\theta\cos\phi, 0),\\
E &= \mathbf{x}_\theta\cdot\mathbf{x}_\theta = a^2,\\
F &= \mathbf{x}_\theta\cdot\mathbf{x}_\phi = 0,\\
G &= \mathbf{x}_\phi\cdot\mathbf{x}_\phi = a^2\sin^2\theta,\\
ds^2 &= a^2 d\theta^2 + a^2\sin^2\theta\, d\phi^2. \qquad (4.21)
\end{aligned}$$



a) Loxodromes b) Escher Drawing c) Aloe Polyphylla

There are many interesting curves on a sphere, but amongst these the *loxodromes* have a special role in history. A loxodrome is a curve that winds around a sphere making a constant angle with the meridians. In this sense, it is the spherical analog of a cylindrical helix and as such it is often called a spherical helix. The curves were significant in early navigation where they are referred as *rhumb* lines. As people in the late 1400's began to rediscover that earth was not flat, cartographers figured out methods to render maps on flat paper surfaces. One such technique is called the *Mercator* projection which is obtained by projecting the sphere onto a plane that wraps around the sphere as a cylinder tangential to the sphere along the equator.

As we will discuss in more detail later, a navigator travelling a constant bearing would be moving on a straight path on the Mercator projection map, but on the sphere it would be spiraling ever faster as one approached the poles. Thus, it became important to understand the nature of such paths. It appears as if the first quantitative treatise of loxodromes was carried in the mid 1500's by the portuguese applied mathematician Pedro Nuñes, who was chair of the department at the University of Coimbra.

As an application, we will derive the equations of loxodromes and compute the arc length. A general spherical curve can be parametrized in the form $\gamma(t) = \mathbf{x}(\theta(t), \phi(t))$. Let σ be the angle the curve makes with the meridians $\phi =$ constant. Then, recalling that $< \mathbf{x}_\theta, \mathbf{x}_\phi >= F = 0$, we have:

$$
\begin{aligned}
\gamma' &= \mathbf{x}_\theta \frac{d\theta}{dt} + \mathbf{x}_\phi \frac{d\phi}{dt}. \\
\cos\sigma &= \frac{< \mathbf{x}_\theta, \gamma' >}{\|\mathbf{x}_\theta\| \cdot \|\gamma'\|} = \frac{E\frac{d\theta}{dt}}{\sqrt{E}\frac{ds}{dt}} = a\frac{d\theta}{ds}. \\
a^2 d\theta^2 &= \cos^2\sigma\; ds^2, \\
a^2\sin^2\sigma\; d\theta^2 &= a^2\cos^2\sigma\sin^2\theta\; d\phi^2, \\
\sin\sigma\; d\theta &= \pm\cos\sigma\sin\theta\; d\phi, \\
\csc\theta\; d\theta &= \pm\cot\sigma\; d\phi.
\end{aligned}
$$

The convention used by cartographers, is to measure the angle θ from the equator. To better adhere to the history, but at the same time avoiding confusion, we replace θ with $\vartheta = \frac{\pi}{2} - \theta$, so that $\vartheta = 0$ corresponds to the equator. Integrating the last equation with this change, we get

$$
\begin{aligned}
\sec\vartheta\; d\vartheta &= \pm\cot\sigma\; d\phi \\
\ln\tan(\tfrac{\vartheta}{2} + \tfrac{\pi}{4}) &= \pm\cot\sigma(\phi - \phi_0).
\end{aligned}
$$

Thus, we conclude that the equations of loxodromes and their arc lengths are given by

$$\phi = \pm(\tan\sigma)\ln\tan(\tfrac{\vartheta}{2} + \tfrac{\pi}{4}) + \phi_0 \tag{4.22}$$

$$s = a(\theta - \theta_0)\sec\sigma, \tag{4.23}$$

where θ_0 and ϕ_0 are the coordinates of the initial position. Figure 4.2 shows four loxodromes equally distributed around the sphere.

Loxodromes were the bases for a number of beautiful drawings and woodcuts by M. C. Escher. figure 4.2 also shows one more beautiful manifestation of geometry in nature in a plant called Aloe Polyphylla. Not surprisingly, the plant has 5 loxodromoes which is a Fibonacci number. We will show later under the discussion of conformal (angle preserving) maps in section 5.2.2, that loxodromes map into straight lines making a constant angle with meridians in the Mercator projection (See Figure 5.9).

b) Surface of Revolution

$$\begin{aligned}
\mathbf{x} &= (r\cos\theta, r\sin\theta, f(r)),\\
\mathbf{x}_r &= (\cos\theta, \sin\theta, f'(r)),\\
\mathbf{x}_\theta &= (-r\sin\theta, r\cos\theta, 0),\\
E &= \mathbf{x}_r\cdot\mathbf{x}_r = 1+f'^2(r),\\
F &= \mathbf{x}_r\cdot\mathbf{x}_\theta = 0,\\
G &= \mathbf{x}_\theta\cdot\mathbf{x}_\theta = r^2,\\
ds^2 &= [1+f'^2(r)]dr^2 + r^2 d\theta^2.
\end{aligned}$$

As in figure 4.6, we have chosen a Gaussian profile to illustrate a surface of revolution. Since $F = 0$ the parametric lines are orthogonal. The picture shows that this is indeed the case. At any point of the surface, the analogs of meridians and parallels intersect at right angles.

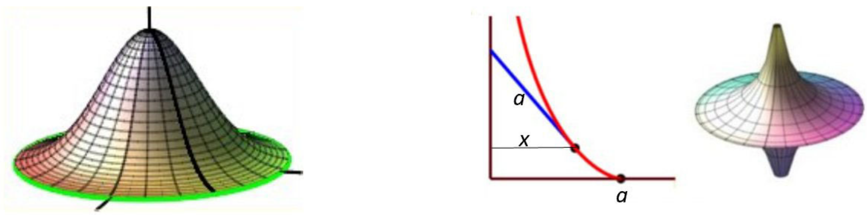


Fig. 4.6: Surface of Revolution and Pseudosphere

c) Pseudosphere

$$\begin{aligned}
\mathbf{x} &= (a\sin u\cos v, a\sin u\sin v, a(\cos u + \ln(\tan\frac{u}{2})),\\
E &= a^2\cot^2 u,\\
F &= = 0\\
G &= a^2\sin^2 u,\\
ds^2 &= a^2\cot^2 u\, du^2 + a^2\sin^2 u\, dv^2.
\end{aligned}$$

The *pseudosphere* is a surface of revolution in which the profile curve is a *tractrix.* The tractrix curve was originated by a problem posed by Leibnitz to the effect of finding the path traced by a point initially placed on the horizontal axis at a distance a from the origin, as it was pulled along the vertical axis by a taught string of constant length a, as shown in figure 4.6. The tractrix was later studied by Huygens in 1692. Colloquially this is the path of a reluctant dog at $(a, 0)$ dragged by a man walking up the z-axis. The tangent segment is the hypothenuse of a right triangle with base x and height $\sqrt{a^2 - x^2}$, so the slope is $dz/dx = -\sqrt{a^2 - x^2}/x$. Using the trigonometric substitution $x = a \sin u$, we get $z = a \int (\cos^2 u / \sin u)\, du$, which leads to the appropriate form for the profile of the surface of revolution. The pseudosphere was studied by Beltrami in 1868. He discovered that in spite of the surface extending asymptotically to infinity, the surface area is finite with $S = 4\pi a^2$ as in a sphere of the same radius, and the volume enclosed is half that sphere. We will have much more to say about this surface.

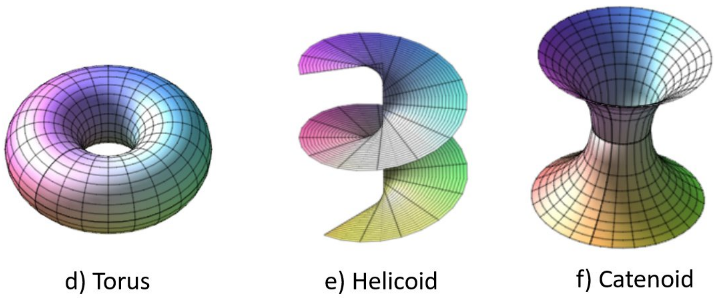


Fig. 4.7: Examples of Surfaces

d) Torus

$$
\begin{aligned}
\mathbf{x} &= ((b + a\cos u)\cos v, (b + a\cos u)\sin v, a\sin u) \quad \text{(See 4.8)},\\
E &= a^2,\\
F &= 0,\\
G &= (b + a\cos u)^2,\\
ds^2 &= a^2 du^2 + (b + a\cos u)^2 dv^2. \qquad (4.24)
\end{aligned}
$$

e) Helicoid

$$
\begin{aligned}
\mathbf{x} &= (u\cos v, u\sin v, av) \quad \text{Coordinate curves } u = c. \quad \text{are helices.}\\
E &= 1,\\
F &= 0,\\
G &= u^2 + a^2,\\
ds^2 &= du^2 + (u^2 + a^2) dv^2. \qquad (4.25)
\end{aligned}
$$

f) Catenoid

$$
\begin{aligned}
\mathbf{x} &= (u\cos v, u\sin v, c\cosh^{-1}\frac{u}{c}), \quad \text{This is a catenary of revolution.}\\
E &= \frac{u^2}{u^2 - c^2},\\
F &= 0,\\
G &= u^2,\\
ds^2 &= \frac{u^2}{u^2 - c^2} du^2 + u^2 dv^2, \qquad (4.26)
\end{aligned}
$$

g) Cone and Conical Helix

The equation $z^2 = \cot^2\alpha(x^2+y^2)$, represents a circular cone whose generator makes an angle α with the z-axis. In parametric form,

$$\begin{aligned}
\mathbf{x} &= (r\cos\phi, r\sin\phi, r\cot\alpha),\\
E &= \csc^2\alpha,\\
F &= 0,\\
G &= r^2,\\
ds^2 &= \csc^2\alpha\, dr^2 + r^2 d\phi^2. \quad (4.27)
\end{aligned}$$

A conical helix is a curve $\gamma(t) = \mathbf{x}(r(t), \phi(t))$, that makes a constant angle σ with the generators of the cone. Similar to the case of loxodromes, we have

$$\begin{aligned}
\gamma' &= \mathbf{x}_r \frac{dr}{dt} + \mathbf{x}_\phi \frac{d\phi}{dt}.\\
\cos\sigma &= \frac{<\mathbf{x}_r, \gamma'>}{\|\mathbf{x}_r\|\cdot\|\gamma'\|} = \frac{E\frac{dr}{dt}}{\sqrt{E}\frac{ds}{dt}} = \sqrt{E}\frac{dr}{ds}.\\
E\, dr^2 &= \cos^2\sigma\, ds^2,\\
\csc^2\alpha\, dr^2 &= \cos^2\sigma(\csc^2\alpha\, dr^2 + r^2 d\phi^2),\\
\csc^2\alpha\sin^2\sigma\, dr^2 &= r^2\cos^2\sigma\, d\phi^2,\\
\frac{1}{r}\, dr &= \cot\sigma\sin\alpha\, d\phi.
\end{aligned}$$

Therefore, the equations of a conical helix are given by

$$r = c\, e^{\cot\sigma\sin\alpha\phi}. \tag{4.28}$$

As shown in figure 4.8, a conical helix projects into the plane as a logarithmic spiral. Many sea shells and other natural objects in nature exhibit neatly such conical spirals. The picture shown here is that of lobatus gigas or caracol pala, previously known as strombus gigas. The particular one is included here with certain degree of nostalgia, for it has been a decorative item for decades in our family. The shell was probably found in Santa Cruz del Islote, Archipelago de San Bernardo, located in the Gulf of Morrosquillo in the Caribbean coast of Colombia. In this densely populated island paradise, which then enjoyed the pulchritude of enchanting coral reefs, the shells are now virtually extinct as the coral has succumbed to bleaching with rising temperatures of the waters. The shell shows a cut in the spire which the island natives use to sever the columellar muscle and thus release the edible snail.

Fig. 4.8: Conical Helix.

4.3 The Second Fundamental Form

Let $\mathbf{x} = \mathbf{x}(u^\alpha)$ be a coordinate patch on a surface M. Since $\mathbf{x}_u$ and $\mathbf{x}_v$ are tangential to the surface, we can construct a unit normal $\mathbf{n}$ to the surface by taking

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}. \tag{4.29}$$

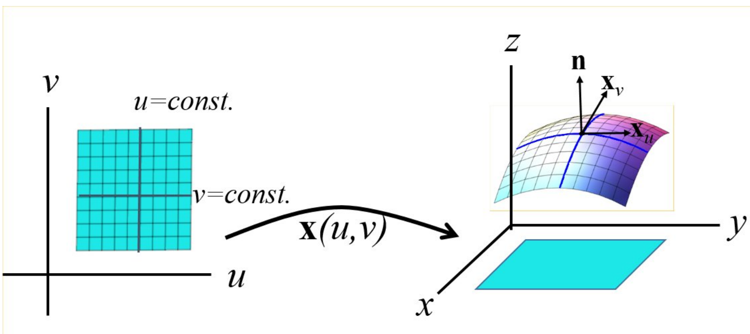


Fig. 4.9: Surface Normal

Now, consider a curve on the surface given by $u^\beta = u^\beta(s)$. Without loss of generality, we assume that the curve is parametrized by arc length s so that the curve has unit speed. Let $e = \{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ be the Frenet frame of the curve. Recall that the rate of change $\overline{\nabla}_T W$ of a vector field W along the curve correspond to the classical vector $\mathbf{w}' = \frac{d\mathbf{w}}{ds}$, so $\overline{\nabla} W$ is associated with the vector $d\mathbf{w}$. Thus the connection equation $\overline{\nabla} e = e\omega$ is given by

$$d[\mathbf{T}, \mathbf{N}, \mathbf{B}] = [\mathbf{T}, \mathbf{N}, \mathbf{B}] \begin{bmatrix} 0 & -\kappa\, ds & 0 \\ \kappa\, ds & 0 & -\tau\, ds \\ 0 & \tau\, ds & 0. \end{bmatrix} \tag{4.30}$$

Following ideas first introduced by Darboux and subsequently perfected by Cartan, we introduce a new orthonormal frame $f = \{\mathbf{T}, \mathbf{g}, \mathbf{n}, \}$ adapted to the surface, where at each point, $\mathbf{T}$ is the common tangent to the surface and to the curve on the surface, $\mathbf{n}$ is the unit normal to the surface and $\mathbf{g} = \mathbf{n} \times \mathbf{T}$. Since the two orthonormal frames must be related by a rotation that leaves the $\mathbf{T}$ vector fixed, we have $f = eB$, where B is a matrix of the form

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}. \tag{4.31}$$

We wish to find $\overline{\nabla} f = f\,\overline{\omega}$. A short computation using the change of basis equations $\overline{\omega} = B^{-1}\omega B + B^{-1}dB$ (see equations 3.48 and 3.51) gives:

$$d[\mathbf{T},\mathbf{g},\mathbf{n}] = [\mathbf{T},\mathbf{g},\mathbf{n}] \begin{bmatrix} 0 & -\kappa\cos\theta\,ds & -\kappa\sin\theta\,ds \\ \kappa\cos\theta\,ds & 0 & -\tau\,ds + d\theta \\ \kappa\sin\theta\,ds & \tau\,ds - d\theta & 0 \end{bmatrix}, \tag{4.32}$$

$$= [\mathbf{T},\mathbf{g},\mathbf{n}] \begin{bmatrix} 0 & -\kappa_g\,ds & -\kappa_n\,ds \\ \kappa_g\,ds & 0 & -\tau_g\,ds \\ \kappa_n\,ds & \tau_g\,ds & 0 \end{bmatrix}, \tag{4.33}$$

where:
$\kappa_n = \kappa\sin\theta$ is called the *normal curvature*,
$\kappa_g = \kappa\cos\theta$ is called the *geodesic curvature*; $\mathbf{K}_g = \kappa_g\mathbf{g}$ the geodesic curvature vector, and
$\tau_g = \tau - d\theta/ds$ is called the *geodesic torsion.*

We conclude that we can decompose $\mathbf{T}'$ and the curvature κ into their normal and surface tangent space components (see figure 4.10)

$$\mathbf{T}' = \kappa_n\mathbf{n} + \kappa_g\mathbf{g}, \tag{4.34}$$
$$\kappa^2 = \kappa_n^2 + \kappa_g^2. \tag{4.35}$$

The normal curvature κ_n measures the curvature of $\mathbf{x}(u^\alpha(s))$ resulting from the constraint of the curve to lie on a surface. The geodesic curvature κ_g measures the "sideward" component of the curvature in the tangent plane to the surface. Thus, if one draws a straight line on a flat piece of paper and then smoothly bends the paper into a surface, the line acquires some curvature. Since the line was originally straight, there is no sideward component of curvature so $\kappa_g = 0$ in this case. This means that the entire contribution to the curvature comes from the normal component, reflecting the fact that the only reason there is curvature here is due to the bend in the surface itself. In this sense, a curve on a surface for which the geodesic curvature vanishes at all points reflects locally the shortest path between two points. These curves are therefore called *geodesics* of the surface. The property of minimizing the path between two points is a local property. For example, on a sphere one would expect the geodesics to be great circles. However, travelling from Los Angeles to San Francisco along one such great circle, there is a short path and a very long one that goes around the earth.

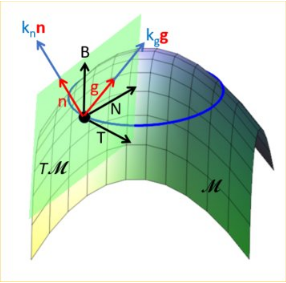


Fig. 4.10: Curvature

If one specifies a point $\mathbf{p} \in M$ and a direction vector $X_p \in T_pM$, one can geometrically envision the normal curvature by considering the equivalence class of all unit speed curves in M that contain the point $\mathbf{p}$ and whose tangent vectors line up with the direction of X. Of course, there are infinitely many such curves, but at an infinitesimal level, all these curves can be obtained by intersecting the surface with a "vertical" plane containing the vector X and the normal to M. All curves in this equivalence class have the same normal curvature and their geodesic curvatures vanish. In this sense, the normal curvature is more of a property pertaining to a direction on the surface at a point, whereas the geodesic curvature really depends on the curve itself. It might be impossible for a hiker walking on the undulating hills of the Ozarks to find a straight line trail, since the rolling hills of the terrain extend in all directions. It might be possible, however, for the hiker to walk on a path with zero geodesic curvature as long the same compass direction is maintained. We will come back to the Cartan structure equations associated with the Darboux frame, but for computational purposes, the classical approach is very practical.
Using the chain rule, we se that the unit tangent vector $\mathbf{T}$ to the curve is given by

$$\mathbf{T} = \frac{d\mathbf{x}}{ds} = \frac{d\mathbf{x}}{du^\alpha}\frac{du^\alpha}{ds} = \mathbf{x}_\alpha \frac{du^\alpha}{ds}. \tag{4.36}$$

To find an explicit formula for the normal curvature we first differentiate equation (4.36)

$$\begin{aligned}
\mathbf{T}' &= \frac{d\mathbf{T}}{ds}, \\
&= \frac{d}{ds}(\mathbf{x}_\alpha \frac{du^\alpha}{ds}), \\
&= \frac{d}{ds}(\mathbf{x}_\alpha)\frac{du^\alpha}{ds} + \mathbf{x}_\alpha \frac{d^2u^\alpha}{ds^2}, \\
&= (\frac{d\mathbf{x}_\alpha}{du^\beta}\frac{du^\beta}{ds})\frac{du^\alpha}{ds} + \mathbf{x}_\alpha \frac{d^2u^\alpha}{ds^2}, \\
&= \mathbf{x}_{\alpha\beta}\frac{du^\alpha}{ds}\frac{du^\beta}{ds} + \mathbf{x}_\alpha \frac{d^2u^\alpha}{ds^2}.
\end{aligned}$$

Taking the inner product of the last equation with the normal and noticing that $< \mathbf{x}_\alpha, \mathbf{n} >= 0$, we get

$$\begin{aligned}
\kappa_n &= < \mathbf{T}', \mathbf{n} >=< \mathbf{x}_{\alpha\beta}, \mathbf{n} > \frac{du^\alpha}{ds}\frac{du^\beta}{ds}, \\
&= \frac{b_{\alpha\beta}du^\alpha du^\beta}{g_{\alpha\beta}du^\alpha du^\beta},
\end{aligned} \tag{4.37}$$

where

$$b_{\alpha\beta} =< \mathbf{x}_{\alpha\beta}, \mathbf{n} > \tag{4.38}$$

4.3.1 Definition The expression

$$II = b_{\alpha\beta}\, du^\alpha \otimes du^\beta \tag{4.39}$$

is called the *second fundamental form* .

4.3.2 Proposition The second fundamental form is symmetric.
Proof In the classical formulation of the second fundamental form, the proof is trivial. We have $b_{\alpha\beta} = b_{\beta\alpha}$, since for a C^∞ patch $\mathbf{x}(u^\alpha)$, we have $\mathbf{x}_{\alpha\beta} = \mathbf{x}_{\beta\alpha}$, because the partial derivatives commute. We will denote the coefficients of the second fundamental form as follows:

$$\begin{aligned} e &= b_{11} =< \mathbf{x}_{uu}, \mathbf{n} >, \\ f &= b_{12} =< \mathbf{x}_{uv}, \mathbf{n} >, \\ &= b_{21} =< \mathbf{x}_{vu}, \mathbf{n} >, \\ g &= b_{22} =< \mathbf{x}_{vv}, \mathbf{n} >, \end{aligned}$$

so that equation (4.39) can be written as

$$II = edu^2 + 2f dudv + gdv^2. \tag{4.40}$$

It follows that the equation for the normal curvature (4.37) can be written explicitly as

$$\kappa_n = \frac{II}{I} = \frac{edu^2 + 2f dudv + gdv^2}{Edu^2 + 2F dudv + Gdv^2}. \tag{4.41}$$

We should pointed out that just as the first fundamental form can be represented as

$$I =< d\mathbf{x}, d\mathbf{x} >,$$

we can represent the second fundamental form as

$$II = -< d\mathbf{x}, d\mathbf{n} > .$$

To see this, it suffices to note that differentiation of the identity, $< \mathbf{x}_\alpha, \mathbf{n} >= 0$, implies that

$$< \mathbf{x}_{\alpha\beta}, \mathbf{n} >= -< \mathbf{x}_\alpha, \mathbf{n}_\beta > .$$

Therefore,

$$\begin{aligned} < d\mathbf{x}, d\mathbf{n} > &= < \mathbf{x}_\alpha du^\alpha, \mathbf{n}_\beta du^\beta >, \\ &= < \mathbf{x}_\alpha du^\alpha, \mathbf{n}_\beta du^\beta >, \\ &= < \mathbf{x}_\alpha, \mathbf{n}_\beta > du^\alpha du^\beta, \\ &= -< \mathbf{x}_{\alpha\beta}, \mathbf{n} > du^\alpha du^\beta, \\ &= -II. \end{aligned}$$

4.3.3 Definition Directions on a surface along which the second fundamental form

$$e\,du^2 + 2f\;du\,dv + g\;dv^2 = 0 \tag{4.42}$$

vanishes, are called *asymptotic directions*, and curves having these directions are called *asymptotic curves.* This happens for example when there are straight lines on the surface, as in the case of the intersection of the saddle $z = xy$ with the plane $z = 0$.

For now, we state without elaboration, that one can also define the third fundamental form by

$$III =< d\mathbf{n}, d\mathbf{n} >=< \mathbf{n}_\alpha, \mathbf{n}_\beta > du^\alpha du^\beta. \tag{4.43}$$

From a computational point a view, a more useful formula for the coefficients of the second fundamental formula can be derived by first applying the classical vector identity

$$(A \times B)\cdot(C \times D) = \begin{vmatrix} A\cdot C & A\cdot D \\ B\cdot C & B\cdot D \end{vmatrix}, \tag{4.44}$$

to compute

$$\begin{aligned} \|\mathbf{x}_u \times \mathbf{x}_v\|^2 &= (\mathbf{x}_u \times \mathbf{x}_v)\cdot(\mathbf{x}_u \times \mathbf{x}_v), \\ &= \det\begin{bmatrix} \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{bmatrix}, \\ &= EG - F^2. \end{aligned} \tag{4.45}$$

Consequently, the normal vector can be written as

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\sqrt{EG - F^2}}.$$

It follows that we can write the coefficients $b_{\alpha\beta}$ directly as triple products involving derivatives of $(\mathbf{x})$. The expressions for these coefficients are

$$\begin{aligned} e &= \frac{(\mathbf{x}_u\mathbf{x}_v\mathbf{x}_{uu})}{\sqrt{EG-F^2}}, \\ f &= \frac{(\mathbf{x}_u\mathbf{x}_v\mathbf{x}_{uv})}{\sqrt{EG-F^2}}, \\ g &= \frac{(\mathbf{x}_u\mathbf{x}_v\mathbf{x}_{vv})}{\sqrt{EG-F^2}}. \end{aligned} \tag{4.46}$$

4.3.4 Example Sphere
Going back to example 4.21, we have:

$$
\begin{aligned}
\mathbf{x}_{\theta\theta} &= (a\sin\theta\cos\phi, -a\sin\theta\sin\phi, -a\cos\theta),\\
\mathbf{x}_{\theta\phi} &= (-a\cos\theta\sin\phi, a\cos\theta\cos\phi, 0),\\
\mathbf{x}_{\phi\phi} &= (-a\sin\theta\cos\phi, -a\sin\theta\sin\phi, 0),\\
\mathbf{n} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta),\\
e &= \mathbf{x}_{\theta\theta}\cdot\mathbf{n} = -a,\\
f &= \mathbf{x}_{\theta\phi}\cdot\mathbf{n} = 0,\\
g &= \mathbf{x}_{\phi\phi}\cdot\mathbf{n} = -a\sin^2\theta,\\
II &= \frac{1}{a^2}I.
\end{aligned}
$$

The first fundamental form on a surface measures the square of the distance between two infinitesimally separated points. There is a similar interpretation of the second fundamental form as we show below. The second fundamental form measures the distance from a point on the surface to the tangent plane at a second infinitesimally separated point. To see this simple geometrical interpretation, consider a point $\mathbf{x}_0 = \mathbf{x}(u_0^\alpha) \in M$ and a nearby point $\mathbf{x}(u_0^\alpha + du^\alpha)$. Expanding on a Taylor series, we get

$$\mathbf{x}(u_0^\alpha + du^\alpha) = \mathbf{x}_0 + (\mathbf{x}_0)_\alpha du^\alpha + \frac{1}{2}(\mathbf{x}_0)_{\alpha\beta} du^\alpha du^\beta + \ldots.$$

We recall that the distance formula from a point $\mathbf{x}$ to a plane which contains $\mathbf{x}_0$ is just the scalar projection of $(\mathbf{x}-\mathbf{x}_0)$ onto the normal. Since the normal to the plane at $\mathbf{x}_0$ is the same as the unit normal to the surface and $<\mathbf{x}_\alpha, \mathbf{n}>= 0$, we find that the distance D is

$$
\begin{aligned}
D &= <\mathbf{x}-\mathbf{x}_0, \mathbf{n}>,\\
&= \frac{1}{2} <(\mathbf{x}_0)_{\alpha\beta}, \mathbf{n}> du^\alpha du^\beta,\\
&= \frac{1}{2} II_0.
\end{aligned}
$$

The first fundamental form (or, rather, its determinant) also appears in calculus in the context of calculating the area of a parametrized surface. If one considers an infinitesimal parallelogram subtended by the vectors $\mathbf{x}_u du$ and $\mathbf{x}_v dv$, then the *differential of surface area* is given by the length of the cross product of these two infinitesimal tangent vectors. That is,

$$
\begin{aligned}
dS &= \|\mathbf{x}_u \times \mathbf{x}_v\|\, dudv,\\
S &= \int\int \sqrt{EG - F^2}\, dudv.
\end{aligned}
$$

The second fundamental form contains information about the shape of the surface at a point. For example, the discussion above indicates that if $b = |b_{\alpha\beta}| = eg - f^2 > 0$ then all the neighboring points lie on the same side of the tangent plane, and hence, the surface is concave in one direction. If at a point on a surface $b > 0$, the point is called an elliptic point, if $b < 0$, the point is called hyperbolic or a saddle point, and if $b = 0$, the point is called parabolic.

4.4 Curvature

The concept of curvature and its relation to the fundamental forms, constitute the central object of study in differential geometry. One would like to be able to answer questions such as "what quantities remain invariant as one surface is smoothly changed into another?" There is certainly something intrinsically different between a cone, which we can construct from a flat piece of paper, and a sphere, which we cannot. What is it that makes these two surfaces so different? How does one calculate the shortest path between two objects when the path is constrained to lie on a surface?

These and questions of similar type can be quantitatively answered through the study of curvature. We cannot overstate the importance of this subject; perhaps it suffices to say that, without a clear understanding of curvature, there would be no general theory of relativity, no concept of black holes, and even more disastrous, no Star Trek.

The notion of curvature of a hypersurface in $\mathbf{R}^n$ (a surface of dimension $n-1$) begins by studying the covariant derivative of the normal to the surface. If the normal to a surface is constant, then the surface is a flat hyperplane. Variations in the normal are indicative of the presence of curvature. For simplicity, we constrain our discussion to surfaces in $\mathbf{R}^3$, but the formalism we use is applicable to any dimension. We will also introduce in the modern version of the second fundamental form.

4.4.1 Classical Formulation of Curvature

The normal curvature κ_n at any point on a surface measures the deviation from flatness as one moves along a direction tangential to the surface at that point. The direction can be taken as the unit tangent vector to a curve on the surface. We seek the directions in which the normal curvature attains the extrema. For this purpose, let the curve on the surface be given by $v = v(u)$ and let $\lambda = \frac{dv}{du.}$ Then we can write the normal curvature 4.41 in the form

$$\kappa_n = \frac{II^*}{I^*} = \frac{e + 2f\lambda + g\lambda^2}{E + 2F\lambda + G\lambda^2}, \tag{4.47}$$

where II^* and I^* are the numerator and denominator respectively. To find the extrema, we take the derivative with respect to λ and set it equal to zero. The resulting fraction is zero only when the numerator is zero, so from the quotient rule we get

$$I^*(2f + 2g\lambda) - II^*(2F + 2G\lambda) = 0.$$

It follows that,

$$\kappa_n = \frac{II^*}{I^*} = \frac{f + g\lambda}{F + G\lambda}. \tag{4.48}$$

On the other hand, combining with equation 4.47 we have,

$$\kappa_n = \frac{(e + f\lambda) + \lambda(f + g\lambda)}{(E + F\lambda) + \lambda(F + G\lambda)} = \frac{f + g\lambda}{F + G\lambda}.$$

This can only happen if

$$\kappa_n = \frac{f + g\lambda}{F + G\lambda} = \frac{e + f\lambda}{E + F\lambda}. \tag{4.49}$$

Equation 4.49 contains a wealth of information. On one hand, we can eliminate κ_n which leads to the quadratic equation for λ

$$(Fg - gF)\lambda^2 + (Eg - Ge)\lambda + (Ef - Fe) = 0.$$

Recalling that $\lambda = dv/du$, and noticing that the coefficients resemble minors of a 3×3 matrix, we can elegantly rewrite the equation as

$$\begin{Vmatrix} du^2 & -du\,dv & dv^2 \\ E & F & G \\ e & f & g \end{Vmatrix} = 0. \tag{4.50}$$

Equation 4.50 determines two directions $\frac{du}{dv}$ along which the normal curvature attains the extrema, except for special cases when either $b_{\alpha\beta} = 0$, or $b_{\alpha\beta}$ and $g_{\alpha\beta}$ are proportional, which would cause the determinant to be identically zero. These two directions are called *principal directions* of curvature, each associated with an extremum of the normal curvature. We will have much more to say about these shortly.

On the other hand, we can write equations 4.49 in the form

$$\begin{cases} (e - E\kappa_n) + \lambda(f - F\kappa_n) = 0, \\ (f - F\kappa_n) + \lambda(g - G\kappa_n) = 0. \end{cases}$$

Solving each equation for λ we can eliminate λ instead, and we are lead to a quadratic equation for κ_n which we can write as

$$\begin{Vmatrix} e - E\kappa_n & f - F\kappa_n \\ f - F\kappa_n & g - G\kappa_n \end{Vmatrix} = 0. \tag{4.51}$$

It is interesting to note that equation 4.51 can be written as

$$\left\| \begin{bmatrix} e & f \\ f & g \end{bmatrix} - \kappa_n \begin{bmatrix} E & F \\ F & G \end{bmatrix} \right\| = 0.$$

In other words, the extrema for the values of the normal are the solutions of the equation

$$\| b_{\alpha\beta} - \kappa_n g_{\alpha\beta} \| = 0. \tag{4.52}$$

Had it been the case that $g_{\alpha\beta} = \delta_{\alpha\beta}$, the reader would recognize this as a eigenvalue equation for a symmetric matrix giving rise to two invariants, that is, the trace and the determinant of the matrix. We will treat this formally in the next section. The explicit quadratic expression for the extrema of κ_n is

$$(EG - F^2)\kappa_n^2 - (Eg - 2Ff + Ge)\kappa_n + (eg - f^2) = 0.$$

We conclude there are two solutions κ_1 and κ_2 such that

$$K = \kappa_1 \kappa_2 = \frac{eg - f^2}{EG - F^2}, \tag{4.53}$$

and

$$M = \tfrac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2}\frac{Eg - 2Ff + Ge}{EF - G^2}.. \tag{4.54}$$

The quantity K is called the *Gaussian curvature* and M is called the *mean curvature*. To understand better the deep significance of the last two equations, we introduce the modern formulation which will allow is to draw conclusions from the inextricable connection of these results with the linear algebra spectral theorem for symmetric operators.

4.4.2 Covariant Derivative Formulation of Curvature

4.4.1 Definition Let X be a vector field on a surface M in $\mathbf{R}^3$ and let N be the normal vector. The map L, given by

$$LX = -\overline{\nabla}_X N, \tag{4.55}$$

is called the *Weingarten map*. Some authors call this the *shape operator*. The same definition applies if M is an n-dimensional hypersurface in $\mathbf{R}^{n+1}$.

Here, we have adopted the convention to overline the operator $\overline{\nabla}$ when it refers to the ambient space. The Weingarten map is natural to consider, since it represents the rate of change of the normal in an arbitrary direction tangential to the surface, which is what we wish to quantify.

4.4.2 Definition The *Lie bracket* $[X, Y]$ of two vector fields X and Y on a surface M is defined as the commutator,

$$[X, Y] = XY - YX, \tag{4.56}$$

meaning that if f is a function on M, then $[X, Y](f) = X(Y(f)) - Y(X(f))$.

4.4.3 Proposition The Lie bracket of two vectors $X, Y \in \mathscr{X}(M)$ is another vector in $\mathscr{X}(M)$.

Proof If suffices to prove that the bracket is a linear derivation on the space of C^∞ functions. Consider vectors $X, Y \in \mathscr{X}(M)$ and smooth functions f, g in M. Then,

$$\begin{aligned}
[X, Y](f + g) &= X(Y(f + g)) - Y(X(f + g)), \\
&= X(Y(f) + Y(g)) - Y(X(f) + X(g)), \\
&= X(Y(f)) - Y(X(f)) + X(Y(g)) - Y(X(g)), \\
&= [X, Y](f) + [X, Y](g),
\end{aligned}$$

and

$$
\begin{aligned}
[X,Y](fg) &= X(Y(fg)) - Y(X(fg)), \\
&= X[fY(g) + gY(f)] - Y[fX(g) + gX(f)], \\
&= X(f)Y(g) + fX(Y(g)) + X(g)Y(f) + gX(Y(f)), \\
&\quad -Y(f)X(g) - f(Y(X(g)) - Y(g)X(f) - gY(X(f)), \\
&= f[X(Y(g)) - (Y(X(g))] + g[X(Y(f)) - Y(X(f))], \\
&= f[X,Y](g) + g[X,Y](f).
\end{aligned}
$$

4.4.4 Proposition The Weingarten map is a linear transformation on $\mathscr{X}(M)$.
Proof Linearity follows from the linearity of $\overline{\nabla}$, so it suffices to show that $L : X \longrightarrow LX$ maps $X \in \mathscr{X}(M)$ to a vector $LX \in \mathscr{X}(M)$. Since N is the unit normal to the surface, $< N, N >= 1$, so any derivative of $< N, N >$ is 0. Assuming that the connection is compatible with the metric,

$$
\begin{aligned}
\overline{\nabla}_X < N, N > &= < \overline{\nabla}_X N, > + < N, \overline{\nabla}_X N >, \\
&= 2 < \overline{\nabla}_X N, N >, \\
&= 2 < -LX, N >= 0.
\end{aligned}
$$

Therefore, LX is orthogonal to N; hence, it lies in $\mathscr{X}(M)$.

In the preceding section, we gave two equivalent definitions $< d\mathbf{x}, d\mathbf{x} >$, and $< X, Y >$ of the first fundamental form. We will now do the same for the second fundamental form.

4.4.5 Definition The *second fundamental form* is the bilinear map

$$
II(X, Y) =< LX, Y > . \tag{4.57}
$$

4.4.6 Remark The two definitions of the second fundamental form are consistent. This is easy to see if one chooses X to have components $\mathbf{x}_\alpha$ and Y to have components $\mathbf{x}_\beta$. With these choices, LX has components $-\mathbf{n}_a$ and $II(X, Y)$ becomes $b_{\alpha\beta} = - < \mathbf{x}_\alpha, \mathbf{n}_\beta >$.

We also note that there is a third fundamental form defined by

$$
III(X, Y) =< LX, LY >=< L^2 X, Y > . \tag{4.58}
$$

In classical notation, the third fundamental form would be denoted by $< d\mathbf{n}, d\mathbf{n} >$. As one would expect, the third fundamental form contains third order Taylor series information about the surface.

4.4.7 Definition The *torsion* of a connection $\overline{\nabla}$ is the operator T such that $\forall X, Y$,

$$
T(X, Y) = \overline{\nabla}_X Y - \overline{\nabla}_Y X - [X, Y]. \tag{4.59}
$$

A connection is called *torsion-free* if $T(X,Y)=0$. In this case,

$$\overline{\nabla}_X Y - \overline{\nabla}_Y X = [X,Y].$$

We will elaborate later on the importance of torsion-free connections. For the time being, it suffices to assume that for the rest of this section, all connections are torsion-free. Using this assumption, it is possible to prove the following important theorem.

4.4.8 Theorem The Weingarten map is a self-adjoint endomorphism on $\mathscr{X}(M)$.

Proof We have already shown that $L : \mathscr{X}M \longrightarrow \mathscr{X}M$ is a linear map. Recall that an operator L on a linear space is self-adjoint if $< LX, Y > = < X, LY >$, so the theorem is equivalent to proving that the second fundamental form is symmetric ($II[X,Y] = II[Y,X]$). Computing the difference of these two quantities, we get

$$\begin{aligned} II(X,Y) - II(Y,X) &= < LX, Y > - < LY, X >, \\ &= < -\overline{\nabla}_X N, Y > - < -\overline{\nabla}_Y N, X > . \end{aligned}$$

Since $< X, N > = < Y, N > = 0$ and the connection is compatible with the metric, we know that

$$\begin{aligned} < -\overline{\nabla}_X N, Y > &= < N, \overline{\nabla}_X Y >, \\ < -\overline{\nabla}_Y N, X > &= < N, \overline{\nabla}_Y X >, \end{aligned}$$

hence,

$$\begin{aligned} II(X,Y) - II(Y,X) &= < N, \overline{\nabla}_Y X > - < N, \overline{\nabla}_X Y >, \\ &= < N, \overline{\nabla}_Y X - \overline{\nabla}_X Y >, \\ &= < N, [X,Y] >, \\ &= 0 \qquad (\text{iff } [X,Y] \in T(M)). \end{aligned}$$

The central theorem of linear algebra is the spectral theorem. In the case of real, self-adjoint operators, the spectral theorem states that given the eigenvalue equation for a symmetric operator

$$LX = \kappa X, \tag{4.60}$$

on a vector space with a real inner product, the eigenvalues are always real and eigenvectors corresponding to different eigenvalues are orthogonal. Here, the vector spaces in question are the tangent spaces at each point of a surface in $\mathbf{R}^3$, so the dimension is 2. Hence, we expect two eigenvalues and two eigenvectors:

$$\begin{aligned} LX_1 &= \kappa_1 X_1 \qquad (4.61) \\ LX_2 &= \kappa_2 X_2. \qquad (4.62) \end{aligned}$$

4.4.9 Definition The eigenvalues κ_1 and κ_2 of the Weingarten map L are called the *principal curvatures* and the eigenvectors X_1 and X_2 are called the *principal directions.*

Several possible situations may occur, depending on the classification of the eigenvalues at each point p on a given surface:

1. If $\kappa_1 \neq \kappa_2$ and both eigenvalues are positive, then p is called an *elliptic point.*

2. If $\kappa_1\kappa_2 < 0$, then p is called a *hyperbolic point.*

3. If $\kappa_1 = \kappa_2 \neq 0$, then p is called an *umbilic point.*

4. If $\kappa_1\,\kappa_2 = 0$, then p is called a *parabolic point.*

It is also known from linear algebra, that in a vector space of dimension two, the determinant and the trace of a self-adjoint operator are the only invariants under an adjoint (similarity) transformation. Clearly, these invariants are important in the case of the operator L, and they deserve special names. In the case of a hypersurface of n-dimensions, there would n eigenvalues, counting multiplicities, so the classification of the points would be more elaborate

4.4.10 Definition The determinant $K = \det(L)$ is called the *Gaussian curvature* of M and $H = \frac{1}{2}\mathrm{Tr}(L)$ is called the *mean curvature* .

Since any self-adjoint operator is diagonalizable and in a diagonal basis the matrix representing L is $\mathrm{diag}(\kappa_1, \kappa_2)$, if follows immediately that

$$\begin{aligned} K &= \kappa_1\kappa_2, \\ H &= \frac{1}{2}(\kappa_1 + \kappa_2). \end{aligned} \tag{4.63}$$

An alternative definition of curvature is obtained by considering the unit normal as a map $N : M \to S^2$, which maps each point p on the surface M, to the point on the sphere corresponding to the position vector N_p. The map is called the *Gauss map.*

4.4.11 Examples

1. The Gauss map of a plane is constant. The image is a single point on S^2.

2. The image of the Gauss map of a circular cylinder is a great circle on S^2.

3. The Gauss map of the top half of a circular cone sends all points on the cone into a circle. We may envision this circle as the intersection of the cone and a unit sphere centered at the vertex.

4. The Gauss map of a circular hyperboloid of one sheet misses two antipodal spherical caps with boundaries corresponding to the circles of the asymptotic cone.

5. The Gauss map of a catenoid misses two antipodal points.

The Weingarten map is minus the derivative $N_* = dN$ of the Gauss map. That is, $LX = -N_*(X)$.

4.4.12 Proposition Let X and Y be any linearly independent vectors in $\mathscr{X}(M)$. Then

$$
\begin{aligned}
LX \times LY &= K(X \times Y), \\
(LX \times Y) + (X \times LY) &= 2H(X \times Y).
\end{aligned} \tag{4.64}
$$

Proof Since $LX, LY \in \mathscr{X}(M)$, they can be expressed as linear combinations of the basis vectors X and Y.

$$
\begin{aligned}
LX &= a_1 X + b_1 Y, \\
LY &= a_2 X + b_2 Y.
\end{aligned}
$$

computing the cross product, we get

$$
\begin{aligned}
LX \times LY &= \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} X \times Y, \\
&= \det(L)(X \times Y).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
(LX \times Y) + (X \times LY) &= (a_1 + b_2)(X \times Y), \\
&= \mathrm{Tr}(L)(X \times Y), \\
&= (2H)(X \times Y).
\end{aligned}
$$

4.4.13 Proposition

$$
\begin{aligned}
K &= \frac{eg - f^2}{EG - F^2}, \\
H &= \frac{1}{2}\frac{Eg - 2Ff + eG}{EG - F^2}.
\end{aligned} \tag{4.65}
$$

Proof Starting with equations (4.64), take the inner product of both sides with $X \times Y$ and use the vector identity (4.44). We immediately get

$$
K = \frac{\begin{vmatrix} < LX, X > & < LX, Y > \\ < LY, X > & < LX, X > \end{vmatrix}}{\begin{vmatrix} < X, X > & < X, Y > \\ < Y, X > & < Y, Y > \end{vmatrix}}, \tag{4.66}
$$

$$
2H = \frac{\begin{vmatrix} < LX, X > & < LX, Y > \\ < Y, X > & < Y, Y > \end{vmatrix} + \begin{vmatrix} < X, X > & < X, Y > \\ < LY, X > & < LY, Y > \end{vmatrix}}{\begin{vmatrix} < X, X > & < X, Y > \\ < Y, X > & < Y, Y > \end{vmatrix}}. \tag{4.67}
$$

The result follows by taking $X = \mathbf{x}_u$ and $Y = \mathbf{x}_v$. Not surprisingly, this is in complete agreement with the classical formulas for the Gausssian curvature (equation 4.53) and for the mean curvature (equation 4.54.

If we denote by g and b the matrices of the fundamental forms whose components are $g_{\alpha\beta}$ and $b_{\alpha\beta}$ respectively, we can write the equations for the curvatures as:

$$K = \det\left(\frac{b}{g}\right) = \det(g^{-1}b), \tag{4.68}$$

$$2H = \mathrm{Tr}\left(\frac{b}{g}\right) = \mathrm{Tr}(g^{-1}b) \tag{4.69}$$

4.4.14 Example Sphere
From equations 4.21 and 4.3 we see that $K = 1/a^2$ and $H = 1/a$. This is totally intuitive since one would expect $\kappa_1 = \kappa_2 = 1/a$ because the normal curvature in any direction should equal the curvature of great circle. This means that a sphere is a surface of constant curvature and every point of a sphere is an umbilic point. This is another way to think of the symmetry of the sphere in the sense that an observer at any point sees the same normal curvature in all directions.

The next theorem due to Euler gives a characterization of the normal curvature in the direction of an arbitrary unit vector X tangent to the surface M at a given point.

4.4.15 Theorem (Euler) Let X_1 and X_2 be unit eigenvectors of L and let $X = (\cos\theta)X_1 + (\sin\theta)X_2$. Then

$$II(X, X) = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta. \tag{4.70}$$

Proof Simply compute $II(X, X) =< LX, X >$, using the fact the $LX_1 = \kappa_1 X_1$, $LX_2 = \kappa_2 X_2$, and noting that the eigenvectors are orthogonal. We get

$$\begin{aligned} < LX, X > &= < (\cos\theta)\kappa_1 X_1 + (\sin\theta)\kappa_2 X_2, (\cos\theta)X_1 + (\sin\theta)X_2 > \\ &= \kappa_1 \cos^2\theta < X_1, X_1 > + \kappa_2 \sin^2\theta < X_2, X_2 > \\ &= \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta. \end{aligned}$$

4.4.16 Theorem The first, second and third fundamental forms satisfy the equation

$$III - 2H\, II + K I = 0 \tag{4.71}$$

Proof The proof follows immediately from the fact that for a symmetric 2 by 2 matrix A, the characteristic polynomial is $\kappa^2 - \mathrm{tr}(A)\kappa + \det(A) = 0$, and from the Cayley-Hamilton theorem stating that the matrix is annihilated by its characteristic polynomial.

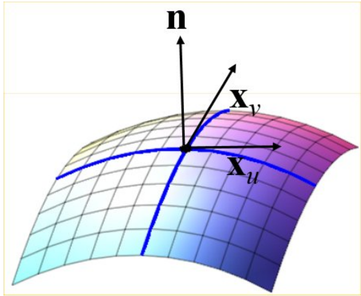


Fig. 4.11: Surface Frame.

4.5 Fundamental Equations

4.5.1 Gauss-Weingarten Equations

As we have just seen for example, the Gaussian curvature of sphere of radius a is $1/a^2$. To compute this curvature we had to first compute the coefficients of the second fundamental form, and therefore, we first needed to compute the normal to the surface in $\mathbf{R}^3$. The computation therefore depended on the particular coordinate chart parametrizing the surface.

However, it would reasonable to conclude that the curvature of the sphere is an intrinsic quantity, independent of the embedding in $\mathbf{R}^3$. After all, a "two-dimensional" creature such as ant moving on the surface of the sphere would be constrained by the curvature of the sphere independent of the higher dimension on which the surface lives. This mode of thinking lead the brilliant mathematicians Gauss and Riemann to question if the coefficients of the second fundamental form were functionally computable from the coefficients of the first fundamental form. To explore this idea, consider again the basis vectors at each point of a surface consisting of two tangent vectors and the normal, as shown in figure 4.11. Given a coordinate chart $\mathbf{x}(u^\alpha)$, the vectors $\mathbf{x}_\alpha$ live on the tangent space, but this is not necessarily true for the second derivative vectors $\mathbf{x}_{\alpha\beta}$. Here, $\mathbf{x}(u^\alpha)$ could refer to a coordinate patch in any number of dimensions, so all the tensor index formulas that follow, apply to surfaces of codimension 1 in $\mathbf{R}^n$. The set of vectors $\{\mathbf{x}_\alpha, \mathbf{n}\}$ constitutes a basis for $\mathbf{R}^n$ at each point on the surface, we can express the vectors $\mathbf{x}_{\alpha\beta}$ as linear combinations of the basis vectors. Therefore, there exist coefficients $\Gamma^\gamma_{\alpha\beta}$ and $c_{\alpha\beta}$ such that,

$$\mathbf{x}_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta}\mathbf{x}_\gamma + c_{\alpha\beta}\mathbf{n}. \tag{4.72}$$

Taking the inner product of equation 4.72 with $\mathbf{n}$, noticing that the latter is a unit vector orthogonal to $\mathbf{x}_\gamma$, we find that $c_{\alpha\beta} =< \mathbf{x}_{\alpha\beta}, \mathbf{n} >$, and hence these are just the coefficients of the second fundamental form. In other words, equation 4.72 can be written as

$$\mathbf{x}_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta}\mathbf{x}_\gamma + b_{\alpha\beta}\mathbf{n}. \tag{4.73}$$

Equation 4.73 together with equation 4.76 below, are called the formulæ of *Gauss*. The covariant derivative formulation of the equation of Gauss follows in a similar fashion. Let X and Y be vector fields tangent to the surface. We decompose the covariant derivative of Y in the direction of X into its tangential and normal components

$$\overline{\nabla}_X Y = \nabla_X Y + h(X, Y)N.$$

But then,

$$\begin{aligned} h(X, Y) &= < \overline{\nabla}_X Y, N >, \\ &= - < Y, \overline{\nabla}_X N >, \\ &= - < Y, LX, >, \\ &= - < LX, Y >, \\ &= II(X, Y). \end{aligned}$$

Thus, the coordinate independent formulation of the equation of Gauss reads

$$\overline{\nabla}_X Y = \nabla_X Y + II(X, Y)N. \tag{4.74}$$

The quantity $\nabla_X Y$ represents a covariant derivative on the surface, so in that sense, it is intrinsic to the surface. If $\alpha(s)$ is a curve on the surface with tangent $T = \alpha'(s)$, we say that a vector field Y is *parallel-transported* along the curve if $\nabla_T Y = 0$. This notion of parallelism refers to parallelism on the surface, not the ambient space. To illustrate by example, Figure 4.12 shows a vector field Y tangent to a sphere along the circle with azimuthal angle $\theta = \pi/3$. The circle has unit tangent $T = \alpha'(s)$, and at each point on the circle, the vector Y points North. To the inhabitants of the sphere, the vector Y appears parallel-transported on the surface along the curve, that is $\nabla_T Y = 0$. However, Y is clearly not parallel-transported in the ambient $\mathbf{R}^3$ space with respect to the connection $\overline{\nabla}$.

The torsion T of the connection ∇ is defined exactly as before (See equation 4.59).

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

Also, as in definition 3.14, the connection is compatible with the metric on the surface if

$$\nabla_X < Y, Z > = < \nabla_X Y, Z > + < Y, \nabla_X Z > .$$

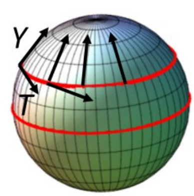


Fig. 4.12:

A torsion-free connection that is compatible with the metric is called a *Levi-Civita* connection.

4.5.1 Proposition A Levi-Civita connection preserves length and angles under parallel transport.

Proof Let $T = \alpha'(t)$ be tangent to curve $\alpha(T)$, and X and Y be parallel-transported along α. By definition, $\nabla_T X = \nabla_T Y = 0$. Then

$$\begin{aligned}
\nabla_T < X, X > &=< \nabla_T X, X > + < X, \nabla_T X >, \\
&= 2 < \nabla_T X, X >= 0, \\
&\Rightarrow \|X\| = \text{constant}. \\
\nabla_T < X, Y > &=< \nabla_T X, Y > + < X, \nabla_T Y >= 0, \\
&\Rightarrow < X, Y >= \text{constant}. \quad \text{So,} \\
\cos\theta &= \frac{< X, Y >}{\|X\| \cdot \|Y\|} = \text{constant}.
\end{aligned}$$

If one takes $\{e_\alpha\}$ to be a basis of the tangent space, the components of the connection in that basis are given by the familiar equation

$$\nabla_{e_\alpha e_\beta} = \Gamma^\gamma{}_{\alpha\beta} e_\gamma.$$

The Γ's here are of course the same Christoffel symbols in the equation of Gauss 4.73. We have the following important result:

4.5.2 Theorem In a manifold $\{M, g\}$ with metric g, there exists a unique Levi-Civita connection.

The proof is implicit in the computations that follow leading to equation 4.76, which express the components uniquely in terms of the metric. The entire equation (4.73) must be symmetric on the indices $\alpha\beta$, since $\mathbf{x}_{\alpha\beta} = \mathbf{x}_{\beta\alpha}$, so $\Gamma^\gamma_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta}$ is also symmetric on the lower indices. These quantities are called the *Christoffel symbols of the first kind.* Now we take the inner product with $\mathbf{x}_\sigma$ to deduce that

$$\begin{aligned}
< \mathbf{x}_{\alpha\beta}, \mathbf{x}_\sigma > &= \Gamma^\gamma_{\alpha\beta} < \mathbf{x}_\gamma, \mathbf{x}_\sigma >, \\
&= \Gamma^\gamma_{\alpha\beta} g_{\gamma\sigma}, \\
&= \Gamma_{\alpha\beta\sigma};
\end{aligned}$$

where we have lowered the third index with the metric on the right hand side of the last equation. The quantities $\Gamma_{\alpha\beta\sigma}$ are called *Christoffel symbols of the second kind.* Here we must note that not all indices are created equal. The Christoffel symbols of the second kind are only symmetric on the first two indices. The notation $\Gamma_{\alpha\beta\sigma} = [\alpha\beta, \sigma]$ is also used in the literature.

The Christoffel symbols can be expressed in terms of the metric by first noticing that the derivative of the first fundamental form is given by (see equation 3.34)

$$\begin{aligned}
g_{\alpha\gamma,\beta} &= \frac{\partial}{\partial u^\beta} < \mathbf{x}_\alpha, \mathbf{x}_\gamma >, \\
&=< \mathbf{x}_{\alpha\beta}, \mathbf{x}_\gamma > + < \mathbf{x}_\alpha, \mathbf{x}_{\gamma\beta}, >, \\
&= \Gamma_{\alpha\beta\gamma} + \Gamma_{\gamma\beta\alpha}.
\end{aligned}$$

Taking other cyclic permutations of this equation, we get

$$\begin{aligned}
g_{\alpha\gamma,\beta} &= \Gamma_{\alpha\beta\gamma} + \Gamma_{\gamma\beta\alpha}, \\
g_{\beta\gamma,\alpha} &= \Gamma_{\alpha\beta\gamma} + \Gamma_{\gamma\alpha\beta}, \\
g_{\alpha\beta,\gamma} &= \Gamma_{\alpha\gamma\beta} + \Gamma_{\gamma\beta\alpha}.
\end{aligned}$$

Adding the first two and subtracting the third of the equations above, and recalling that the Γ's are symmetric on the first two indices, we obtain the formula

$$\Gamma_{\alpha\beta\gamma} = \frac{1}{2}(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}). \tag{4.75}$$

Raising the third index with the inverse of the metric, we also have the following formula for the Christoffel symbols of the first kind (hereafter, Christoffel symbols refer to the symbols of the first kind, unless otherwise specified.)

$$\Gamma^{\sigma}_{\alpha\beta} = \frac{1}{2}g^{\sigma\gamma}(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}). \tag{4.76}$$

The Christoffel symbols are clearly symmetric in the lower indices

$$\Gamma^{\sigma}_{\alpha\beta} = \Gamma^{\sigma}_{\beta\alpha}. \tag{4.77}$$

Unless otherwise specified, a connection on $\{M, g\}$ refers to the unique Levi-Civita connection.

We derive a well-known formula for the Christoffel symbols for the case ${\Gamma^{\alpha}}_{\alpha\beta}$. From 4.76 we have:

$$\Gamma^{\alpha}_{\alpha\beta} = \frac{1}{2}g^{\alpha\gamma}(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}).$$

On the other hand,

$$g^{\alpha\gamma}g_{\beta\gamma,\alpha} = g^{\alpha\gamma}g_{\alpha\beta,\gamma}$$

as can be seen by switching the repeated indices of summation α and σ, and using the symmetry of the metric. The equation reduces to

$$\Gamma^{\alpha}_{\alpha\beta} = \frac{1}{2}g^{\alpha\gamma}g_{\alpha\gamma,\beta}$$

Let A be the cofactor transposed matrix of g. From the linear algebra formula for the expansion of a determinant in terms of cofactors we can get an expression an expression for the inverse of the metric as follows:

$$\begin{aligned}
\det(g) &= g_{\alpha\gamma}A^{\alpha\gamma}, \\
\frac{\partial\det(g)}{\partial g_{\alpha\gamma}} &= A^{\alpha\gamma}, \\
g^{\alpha\gamma} &= \frac{A^{\alpha\gamma}}{\det(g)}, \\
&= \frac{1}{\det(g)}\frac{\partial\det(g)}{\partial g_{\alpha\gamma}}
\end{aligned}$$

so that

$$\begin{aligned}
\Gamma^{\alpha}_{\alpha\beta} &= \frac{1}{2\det(g)} \frac{\partial \det(g)}{\partial g_{\alpha\gamma}} \frac{\partial}{\partial u^{\beta}} g_{\alpha\gamma}, \\
&= \frac{1}{2\det(g)} \frac{\partial}{\partial u^{\beta}} (\det(g)). \qquad (4.78)
\end{aligned}$$

Using this result we can also get a tensorial version of the divergence of the vector field $X = v^{\alpha} e_{\alpha}$ on the manifold. Using the classical covariant derivative formula 3.25 for the components v^{α}, we define:

$$\mathrm{Div}\, X = \nabla \cdot X = {v^{\alpha}}_{\|\alpha} \qquad (4.79)$$

We get

$$\begin{aligned}
\mathrm{Div}\, X &= {v^{\alpha}}_{,\alpha} + {\Gamma^{\alpha}}_{\alpha\gamma}\, v^{\gamma}, \\
&= \frac{\partial}{\partial u^{\alpha}} v^{\alpha} + \frac{1}{2\det(g)} \frac{\partial}{\partial u^{\gamma}} (\det(g)) v^{\gamma}, \\
&= \frac{1}{\sqrt{\det(g)}} \frac{\partial}{\partial u^{\alpha}} (\sqrt{\det(g)} v^{\alpha}). \qquad (4.80)
\end{aligned}$$

If f is a function on the manifold, $df = f_{,\beta}\, du^{\beta}$ so the contravariant components of the gradient are

$$(\nabla f)^{\alpha} = g^{\alpha\beta} f_{,\beta}\,. \qquad (4.81)$$

Combining with equation above, we get a second order operator

$$\begin{aligned}
\Delta f &= \mathrm{Div}(\mathrm{Grad} f), \\
&= \frac{1}{\sqrt{\det(g)}} \frac{\partial}{\partial u^{\alpha}} (\sqrt{\det(g)} g^{\alpha\beta} f_{,\beta}) \qquad (4.82)
\end{aligned}$$

The quantity Δ is called the *Laplace-Beltrami* operator on a function and it generalizes the Laplacian of functions in $\mathbf{R}^n$ to functions on manifolds.

4.5.3 Example Laplacian in Spherical Coordinates
The metric in spherical coordinates is $ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2$, so

$$g_{\alpha\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{bmatrix}, \qquad g^{\alpha\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{i}{r^2 \sin^2\theta} \end{bmatrix}, \qquad \sqrt{\det(g)} = r^2 \sin\theta.$$

The Laplace-Beltrami formula gives,

$$\begin{aligned}
\Delta f &= \frac{1}{\sqrt{\det g}} \left[\frac{\partial}{\partial u^1} (\sqrt{\det g}\, g^{11} \frac{\partial f}{\partial u^1}) + \frac{\partial}{\partial u^2} (\sqrt{\det g}\, g^{22} \frac{\partial f}{\partial u^2}) + \frac{\partial}{\partial u^3} (\sqrt{\det g}\, g^{33} \frac{\partial f}{\partial u^3}) \right], \\
&= \frac{1}{r^2 \sin\theta} \left[\frac{\partial}{\partial r} (r^2 \sin\theta \frac{\partial f}{\partial r}) + \frac{\partial}{\partial \theta} (r^2 \sin\theta \frac{1}{r^2} \frac{\partial f}{\partial \theta}) + \frac{\partial}{\partial \phi} (r^2 \sin\theta \frac{1}{r^2 \sin^2\theta} \frac{\partial f}{\partial \phi}) \right], \\
&= \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 f}{\partial \phi^2}.
\end{aligned}$$

The result is the same as the formula for the Laplacian 3.9 found by differential form methods.

4.5.4 Example

As an example we unpack the formula for Γ^1_{11}. First, note that $\det(g) = \|g_{\alpha\beta}\| = EG - F^2$. From equation 4.76 we have

$$\begin{aligned}
\Gamma^1_{11} &= \frac{1}{2} g^{1\gamma}(g_{1\gamma,1} + g_{1\gamma,1} - g_{11,\gamma}), \\
&= \frac{1}{2} g^{1\gamma}(2g_{1\gamma,1} - g_{11,\gamma}), \\
&= \frac{1}{2}[g^{11}(2g_{11,1} - g_{11,1}) + g^{12}(2g_{12,1} - g_{11,2})], \\
&= \frac{1}{2\det(g)}[GE_u - F(2F_u - FE_v)], \\
&= \frac{GE_u - 2FF_u + FE_v}{2(EG - F^2)}.
\end{aligned}$$

Due to symmetry, there are five other similar equations for the other Γ's. Proceeding as above, we can derive the entire set.

$$\begin{aligned}
\Gamma^1_{11} &= \frac{GE_u - 2FF_u + FE_v}{2(EG - F^2)} & \Gamma^2_{11} &= \frac{2EF_u - EE_v - FE_u}{2(EG - F^2)} \\
\Gamma^1_{12} &= \frac{GE_v - FG_u}{2(EG - F^2)} & \Gamma^2_{12} &= \frac{EG_u - FE_v}{2(EG - F^2)} \\
\Gamma^1_{22} &= \frac{2GF_v - GG_u - FG_v}{2(EG - F^2)} & \Gamma^2_{22} &= \frac{EG_v - 2FF_v + FG_u}{2(EG - F^2)}.
\end{aligned} \tag{4.83}$$

They are a bit messy, but they all simplify considerably for orthogonal systems, in which case $F = 0$. Another reason why we like those coordinate systems.

4.5.5 Example Harmonic functions.

A function h on a surface in $\mathbf{R}^3$ is called *harmonic* if it satisfies:

$$\Delta\, h = 0. \tag{4.84}$$

Noticing that the matrix components of the inverse of the metric are given by

$$g^{\alpha\beta} = \frac{1}{\det(g)} \begin{bmatrix} G & -F \\ -F & E \end{bmatrix} \tag{4.85}$$

we get immediately from 4.82, the classical Laplace-Beltrami equation for surfaces,

$$\Delta h = \frac{1}{\sqrt{EG - F^2}} \left\{ \frac{\partial}{\partial u} \left[\frac{Gh_{,u} - Fh_{,v}}{\sqrt{EG - F^2}} \right] + \frac{\partial}{\partial v} \left[\frac{Eh_{,v} - Fh_{,u}}{\sqrt{EG - F^2}} \right] \right\} = 0. \tag{4.86}$$

If the coordinate patch is orthogonal so that $F = 0$, the equation reduces to:

$$\frac{\partial}{\partial u}\left[\frac{\sqrt{G}}{\sqrt{E}}\frac{\partial h}{\partial u}\right] + \frac{\partial}{\partial v}\left[\frac{\sqrt{E}}{\sqrt{G}}\frac{\partial h}{\partial v}\right] = 0 \tag{4.87}$$

If in addition $E = G = \lambda^2$ so that the metric has the form,

$$ds^2 = \lambda^2\,(du^2 + dv^2), \tag{4.88}$$

then,

$$\Delta h = \frac{1}{\lambda^2}\left[\frac{\partial^2 h}{\partial u^2} + \frac{\partial^2 h}{\partial v^2}\right]. \tag{4.89}$$

Hence, $\Delta^2 h = 0$ is equivalent to $\nabla^2 h = 0$, where ∇^2 is the Euclidean Laplacian. (Please compare to the discussion on the isothermal coordinates example 4.5.14.) Two metrics that differ by a multiplicative factor are called conformally related. The result here means that the Laplacian is conformally invariant under this conformal transformation. This property is essential in applying the elegant methods of complex variables and conformal mappings to solve physical problems involving the Laplacian in the plane.

For a surface $z = f(x, y)$, which we can write as a Monge patch $\mathbf{x} =< x, y, f(x, y) >$, we have $E = 1 + f_x^2$, $F = 2f_x f_y$ and $G = 1 + f_y^2$=0. A short computation shows that in this case, the Laplace-Beltrami equation can be written as, (compare to equation 5.43)

$$\Delta h = \frac{1}{\sqrt{1 + f_x^2 + f_y^2}}\left\{\frac{\partial}{\partial x}\left[\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}\right] + \frac{\partial}{\partial y}\left[\frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}\right]\right\} = 0,$$

or in terms of the Euclidean $\mathbf{R}^2$ del operator $\nabla =< \frac{\partial}{\partial x}, \frac{\partial}{\partial y} >$,

$$\frac{\partial}{\partial x}\left[\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}\right] + \frac{\partial}{\partial y}\left[\frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}\right] = 0,$$

$$\nabla \cdot \left[\frac{\nabla f}{\sqrt{1 + \|\nabla f\|^2}}\right] = 0. \tag{4.90}$$

4.5.2 Curvature Tensor, Gauss's Theorema Egregium

A fascinating set of relations can be obtained simply by equating $\mathbf{x}_{\beta\gamma\delta} = \mathbf{x}_{\beta\delta\gamma}$. First notice that we can also write $\mathbf{n}_\alpha$ in terms of the frame vectors. This is by far easier since $< \mathbf{n}, \mathbf{n} >= 1$ implies that $< \mathbf{n}_\alpha, \mathbf{n} >= 0$, so $\mathbf{n}_\alpha$ lies on the tangent plane and it is therefore a linear combination the tangent vectors. As before, we easily verify that the coefficients are the second fundamental form with a raised index

$$\mathbf{n}_\alpha = -b^\gamma_\alpha \mathbf{x}_\gamma. \tag{4.91}$$

These are called the formulæ of *Weingarten.*

Differentiating the equation of Gauss and recursively using the formulas of Gauss 4.73 and Weingarten 4.91 to write all the components in terms of the frame, we get

$$
\begin{aligned}
\mathbf{x}_{\beta\delta} &= \Gamma^{\alpha}_{\beta\delta}\mathbf{x}_{\alpha} + b_{\beta\delta}\mathbf{n}, \\
\mathbf{x}_{\beta\delta\gamma} &= \Gamma^{\alpha}_{\beta\delta,\gamma}\mathbf{x}_{\alpha} + \Gamma^{\alpha}_{\beta\delta}\mathbf{x}_{\alpha\gamma} + b_{\beta\delta,\gamma}\mathbf{n} + b_{\beta\delta}\mathbf{n}_{\gamma} \\
&= \Gamma^{\alpha}_{\beta\delta,\gamma}\mathbf{x}_{\alpha} + \Gamma^{\alpha}_{\beta\delta}[\Gamma^{\mu}_{\alpha\gamma}\mathbf{x}_{\mu} + b_{\alpha\gamma}\mathbf{n}] + b_{\beta\delta,\gamma}\mathbf{n} - b_{\beta\delta}b^{\alpha}_{\gamma}\mathbf{x}_{\alpha} \\
\mathbf{x}_{\beta\delta\gamma} &= [\Gamma^{\alpha}_{\beta\delta,\gamma} + \Gamma^{\mu}_{\beta\delta}\Gamma^{\alpha}_{\mu\gamma} - b_{\beta\delta}b^{\alpha}_{\gamma}]\mathbf{x}_{\alpha} + [\Gamma^{\alpha}_{\beta\delta}b_{\alpha\gamma} + b_{\beta\delta,\gamma}]\mathbf{n}, \qquad (4.92) \\
\mathbf{x}_{\beta\gamma\delta} &= [\Gamma^{\alpha}_{\beta\gamma,\delta} + \Gamma^{\mu}_{\beta\gamma}\Gamma^{\alpha}_{\mu\delta} - b_{\beta\gamma}b^{\alpha}_{\delta}]\mathbf{x}_{\alpha} + [\Gamma^{\alpha}_{\beta\gamma}b_{\alpha\delta} + b_{\beta\gamma,\delta}]\mathbf{n}. \qquad (4.93)
\end{aligned}
$$

The last equation above was obtained from the preceding one just by permuting δ and γ. Subtracting that last two equations and setting the tangential component to zero we get

$$R^{\alpha}{}_{\beta\gamma\delta} = b_{\beta\delta}b^{\alpha}_{\gamma} - b_{\beta\gamma}b^{\alpha}_{\delta}, \tag{4.94}$$

where the components of the *Riemann tensor* R are defined by

$$R^{\alpha}_{\beta\gamma\delta} = \Gamma^{\alpha}_{\beta\delta,\gamma} - \Gamma^{\alpha}_{\beta\gamma,\delta} + \Gamma^{\mu}_{\beta\delta}\Gamma^{\alpha}_{\gamma\mu} - \Gamma^{\mu}_{\beta\gamma}\Gamma^{\alpha}_{\delta\mu}. \tag{4.95}$$

Technically we are not justified at this point in calling R a tensor since we have not established yet the appropriate multi-linear features that a tensor must exhibit. We address this point in a later chapter. Lowering the index above we get

$$R_{\alpha\beta\gamma\delta} = b_{\beta\delta}b_{\alpha\gamma} - b_{\beta\gamma}b_{\alpha\delta}. \tag{4.96}$$

4.5.6 *Theorema egregium* Let M be a smooth surface in $\mathbf{R}^3$. Then,

$$K = \frac{R_{1212}}{\det(g)}. \tag{4.97}$$

Proof Let $\alpha = \gamma = 1$ and $\beta = \delta = 2$ above. The equation then reads

$$
\begin{aligned}
R_{1212} &= b_{22}b_{11} - b_{21}b_{12}, \\
&= (eg - f^2), \\
&= K(EF - G^2), \\
&= K\det(g)
\end{aligned}
$$

The remarkable result is that the Riemann tensor and hence the Gaussian curvature does not depend on the second fundamental form but only on the coefficients of the metric. Thus, the Gaussian curvature is an intrinsic quantity independent of the embedding, so that two surfaces that have the same first fundamental form have the same curvature. In this sense, the Gaussian curvature is a bending invariant!

Setting the normal components equal to zero gives

$$\Gamma^{\alpha}_{\beta\delta} b_{\alpha\gamma} - \Gamma^{\alpha}_{\beta\gamma} b_{\alpha\delta} + b_{\beta\delta,\gamma} - b_{\beta\gamma,\delta} = 0 \tag{4.98}$$

These are called the *Codazzi* (or *Codazzi-Mainardi*) equations.

Computing the Riemann tensor is labor intensive since one must first obtain all the non-zero Christoffel symbols as shown in the example above. Considerable gain in efficiency results from a form computation. For this purpose, let $\{e_1, e_2, e_3\}$ be a Darboux frame adapted to the surface M, with $e_3 = \mathbf{n}$. Let $\{\theta^1, \theta^2, \theta^3\}$ be the corresponding orthonormal dual basis. Since at every point, a tangent vector $X \in TM$ is a linear combination of $\{e_1, e_2\}$, we see that $\theta^3(X) = 0$ for all such vectors. That is, $\theta^3 = 0$ on the surface. As a consequence, the entire set of the structure equations is

$$
\begin{aligned}
d\theta^1 &= -\omega^1{}_2 \wedge \theta^2, && && (4.99)\\
d\theta^2 &= -\omega^2{}_1 \wedge \theta^1, && && (4.100)\\
d\theta^3 &= -\omega^3{}_1 \wedge \theta^1 - \omega^3{}_2 \wedge \theta^2 = 0, && && (4.101)\\
d\omega^1{}_2 &= -\omega^1{}_3 \wedge \omega^3{}_2, && \text{Gauss Equation} && (4.102)\\
d\omega^1{}_3 &= -\omega^1{}_2 \wedge \omega^2{}_3, && \text{Codazzi Equations} && (4.103)\\
d\omega^2{}_3 &= -\omega^2{}_1 \wedge \omega^1{}_3. && && (4.104)
\end{aligned}
$$

The key result is the following theorem

4.5.7 *Curvature form equations*

$$
\begin{aligned}
d\omega^1{}_2 &= K\,\theta^1 \wedge \theta^2, && (4.105)\\
\omega^1{}_3 \wedge \theta^2 + \omega^2{}_3 \wedge \theta^1 &= -2H\ \theta^1 \wedge \theta^2. && (4.106)
\end{aligned}
$$

Proof By applying the Weingarten map to the basis vector $\{e_1, e_2\}$ of TM, we find a matrix representation of the linear transformation:

$$
\begin{aligned}
Le_1 &= -\nabla_{e_1} e_3 = -\omega^1{}_3(e_1)e_1 - \omega^2{}_3(e_1)e_2,\\
Le_2 &= -\nabla_{e_2} e_3 = -\omega^1{}_3(e_2)e_1 - \omega^2{}_3(e_2)3_2.
\end{aligned}
$$

Recalling that ω is antisymmetric, we find:

$$
\begin{aligned}
K = \det(L) &= -[\omega^1{}_3(e_1)\omega^3{}_2(e_2) - \omega^1{}_3(e_2)\omega^3{}_2(e_1)],\\
&= -(\omega^1{}_3 \wedge \omega^3{}_2)(e_1, e_2),\\
&= d\omega^1{}_2(e_1, e_2).
\end{aligned}
$$

Hence

$$d\omega^1{}_2 = K\,\theta^1 \wedge \theta^2.$$

Similarly, recalling that $\theta^1(e_j) = \delta^i_j$, we have

$$
\begin{aligned}
(\omega^1_{\ 3} \wedge \theta^2 + \omega^3_{\ 2} \wedge \theta^1)(e_1, e_2) &= \omega^1_{\ 3}(e_1) - \omega^3_{\ 2}(e_2), \\
&= \omega^1_{\ 3}(e_1) + \omega^2_{\ 3}(e_2), \\
&= \mathrm{Tr}(L) = -2H.
\end{aligned}
$$

4.5.8 Definition A point of a surface at which $K = 0$ is called a *planar point.* A surface with $K = 0$ at all points is called a *flat* or *Gaussian flat* surface. A surface on which $H = 0$ at all points is called a *minimal* surface.

4.5.9 Example Sphere Since the first fundamental form is $I = a^2\,d\theta^2 + a^2 \sin^2\theta\,d\phi^2$, we have

$$
\begin{aligned}
\theta^1 &= a\,d\theta, \\
\theta^2 &= a \sin\theta\,d\phi, \\
d\theta^2 &= a\cos\theta\,d\theta \wedge d\phi, \\
&= -\cos\theta\,d\phi \wedge \theta^1 = -\omega^2_{\ 1} \wedge \theta^1, \\
\omega^2_{\ 1} &= \cos\theta\,d\phi = -\omega^1_{\ 2}, \\
d\omega^1_{\ 2} &= \sin\theta\,d\theta \wedge d\phi = \frac{1}{a^2}\,(a\,d\theta) \wedge (a\sin\theta\,d\phi), \\
&= \frac{1}{a^2}\theta^1 \wedge \theta^2, \\
K &= \frac{1}{a^2}.
\end{aligned}
$$

4.5.10 Example Torus

Using the the parametrization (See 4.24),

$$\mathbf{x} = ((b + a\cos\theta)\cos\phi, (b + a\cos\theta)\sin\phi, a\sin\theta),$$

the first fundamental form is

$$ds^2 = a^2 d\theta^2 + (b + a\cos\theta)^2 d\phi^2.$$

Thus, we have:

$$
\begin{aligned}
\theta^1 &= a\, d\theta, \\
\theta^2 &= (b + a\cos\theta)\, d\phi, \\
d\theta^2 &= -a\sin\theta\, d\theta \wedge d\phi, \\
&= \sin\theta\, d\phi \wedge \theta^1 = -\omega^2{}_1 \wedge \theta^1, \\
\omega^2{}_1 &= -\sin\theta\, d\phi = -\omega^1{}_2, \\
d\omega^1{}_2 &= \cos\theta\, d\theta \wedge d\phi = \frac{\cos\theta}{a(b + a\cos\theta)}\, (a\, d\theta) \wedge [(a + b\cos\theta)\, d\phi], \\
&= \frac{\cos\theta}{a(b + a\cos\theta)} \theta^1 \wedge \theta^2, \\
K &= \frac{\cos\theta}{a(b + a\cos\theta)}.
\end{aligned}
$$

This result makes intuitive sense.

When $\theta = 0$, the points lie on the outer equator, so $K = \dfrac{1}{a(b+a)} > 0$ is the product of the curvatures of the generating circle and the outer equator circle. The points are elliptic.

When $\theta = \pi/2$, the points lie on the top of the torus, so $K = 0$. The points are parabolic.

When $\theta = \pi$, the points lie on the inner equator, so $K = \dfrac{-1}{a(b-a)} < 0$ is the product of the curvatures of the generating circle and the inner equator circle. The points are hyperbolic.

4.5.11 **Example** Orthogonal parametric curves

The examples above have the common feature that the parametric curves are orthogonal and hence $F = 0$. Using the same method, we can find a general formula for such cases. Since the first fundamental form is given by

$$I = Edu^2 + Gdv^2.$$

We have:

$$
\begin{aligned}
\theta^1 &= \sqrt{E}\, du, \\
\theta^2 &= \sqrt{G}\, dv, \\
d\theta^1 &= (\sqrt{E})_v\, dv \wedge du = -(\sqrt{E})_v\, du \wedge dv, \\
&= -\frac{(\sqrt{E})_v}{\sqrt{G}}\, du \wedge \theta^2 = -\omega^1{}_2 \wedge \theta^2, \\
d\theta^2 &= (\sqrt{G})_u\, du \wedge dv = -(\sqrt{G})_u\, dv \wedge du \\
&= -\frac{(\sqrt{G})_u}{\sqrt{E}}\, dv \wedge \theta^2 = -\omega^2{}_1 \wedge \theta^1, \\
\omega^1{}_2 &= \frac{(\sqrt{E})_v}{\sqrt{G}}\, du - \frac{(\sqrt{G})_u}{\sqrt{E}}\, dv \\
d\omega^1{}_2 &= -\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right) du \wedge dv + \frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right) dv \wedge du, \\
&= -\frac{1}{\sqrt{EG}}\left[\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right)\right] \theta^1 \wedge \theta^2.
\end{aligned}
$$

Therefore, the Gaussian Curvature of a surface mapped by a coordinate patch in which the parametric lines are orthogonal is given by:

$$
K = -\frac{1}{\sqrt{EG}}\left[\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right)\right]. \tag{4.107}
$$

Again, to connect with more classical notation, if a surface described by a coordinate patch $\mathbf{x}(u,v)$ has first fundamental for given by $I = E\, du^2 + G\, dv^2$, then

$$
\begin{aligned}
d\mathbf{x} &= \mathbf{x}_u\, du + \mathbf{x}_v\, dv, \\
&= \frac{\mathbf{x}_u}{\sqrt{E}}\sqrt{E}\, du + \frac{\mathbf{x}_v}{\sqrt{G}}\sqrt{G}\, dv, \\
&= \frac{\mathbf{x}_u}{\sqrt{E}}\theta^1 + \frac{\mathbf{x}_v}{\sqrt{G}}\theta^2, \\
d\mathbf{x} &= \mathbf{e}_1\, \theta^1 + \mathbf{e}_2\, \theta^2,
\end{aligned} \tag{4.108}
$$

where

$$
\mathbf{e}_1 = \frac{\mathbf{x}_u}{\sqrt{E}}, \quad \mathbf{e}_2 = \frac{\mathbf{x}_v}{\sqrt{G}}.
$$

Thus, when the parametric curves are orthogonal, the triplet $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e_3} = \mathbf{n}\}$ constitutes a moving orthonormal frame adapted to the surface. The awkwardness of combining calculus vectors and differential forms in the same equation is mitigated by the ease of jumping back and forth between the classical and the modern formalism. Thus, for example, covariant differential of the normal in 4.104 can be rewritten without the arbitrary vector in the operator LX as shown:

$$\overline{\nabla}_X e_3 = \omega^1{}_3(X)\, e_1 + \omega^2{}_3(X)\, e_2, \tag{4.109}$$
$$d\mathbf{e}_3 = \mathbf{e}_1\, \omega^1{}_3 + \mathbf{e}_2\, \omega^2{}_3 = 0, \tag{4.110}$$

The equation just expresses the fact that the components of the Weingarten map, that is, the second fundamental form in this basis, can be written as some symmetric matrix given by:

$$\begin{aligned} \omega^1{}_3 &= l\, \theta^1 + m\, \theta^2, \\ \omega^2{}_3 &= m\, \theta^1 + n\, \theta^2. \end{aligned} \tag{4.111}$$

If $E = 1$, we say that the metric

$$ds^2 = du^2 + G(u,v) dv^2, \tag{4.112}$$

is in *geodesic coordinates.* In this case, the equation for curvature reduces even further to:

$$K = -\frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{G}}{\partial u^2}. \tag{4.113}$$

The case is not as special as it appears at first. The change of parameters

$$\hat{u}' = \int_0^u \sqrt{E}\, du$$

results on $d\hat{u}^2 = E\, du^2$, and thus it transforms an orthogonal system to one with $E = 1$. The parameters are reminiscent of polar coordinates $ds^2 = dr^2 + r^2\, d\phi^2$. Equation 4.113 is called *Jacobi's differential equation* for geodesic coordinates.

A slick proof of the theorema egregium can be obtained by differential forms. Let $F : M \to \widetilde{\mathcal{M}}$ be an isometry between two surfaces with metrics g and $\tilde{g}$ respectively. Let $\{e_\alpha\}$ be an orthonormal basis for dual basis $\{\theta^\alpha\}$. Define $\tilde{e}_\alpha = F_* e_\alpha$. Recalling that isometries preserve inner products, we have

$$< \tilde{e}_\alpha, \tilde{e}_\beta >=< F_* e_\alpha, F_* e_\beta >=< e_\alpha, e_\beta >= \delta_{\alpha\beta}.$$

Thus, $\{\tilde{e}_\alpha\}$ is also an orthonormal basis of the tangent space of $\widetilde{\mathcal{M}}$. Let $\tilde{\theta}^\alpha$ be the dual forms and denote with tilde's the connection forms and Gaussian curvature of $\widetilde{\mathcal{M}}$.

4.5.12 Theorem (Theorema egregium)
a) $F^* \tilde{\theta}_\alpha = \theta_\alpha$,
b) $F^* \tilde{\omega}^\alpha{}_\beta = \omega^\alpha{}_\beta$,
c) $F^* \tilde{K} = K$.

Proof

a) It suffices to show that the forms agree on basis vectors. We have

$$\begin{aligned} F^* \tilde{\theta}_\alpha(e_\beta) &= \tilde{\theta}_\alpha(F_* e_\beta), \\ &= \tilde{\theta}_\alpha(\tilde{e}_\beta), \\ &= \delta^\alpha{}_\beta, \\ &= \theta(e_\beta). \end{aligned}$$

b) We compute the pull-back of the first structure equation in $\widetilde{M}$:

$$\begin{aligned} d\tilde{\theta}^\alpha + \tilde{\omega}^\alpha{}_\beta \wedge \tilde{\theta}^\beta &= 0, \\ F^* d\tilde{\theta}^\alpha + F^*\tilde{\omega}^\alpha{}_\beta \wedge F^*\tilde{\theta}^\beta &= 0, \\ d\theta^\alpha + F^*\tilde{\omega}^\alpha{}_\beta \wedge \theta^\beta &= 0, \end{aligned}$$

The connection forms are defined uniquely by the first structure equation, so

$$F^*\tilde{\omega}^\alpha{}_\beta = \omega^\alpha{}_\beta$$

.

c) In a similar manner, we compute the pull-back of the curvature equation:

$$\begin{aligned} d\tilde{\omega}^1{}_2 &= \tilde{K}\,\tilde{\theta}^1 \wedge \tilde{\theta}^2, \\ F^* d\tilde{\omega}^1{}_2 &= (F^*\tilde{K})\, F^*\tilde{\theta}^1 \wedge F^*\tilde{\theta}^2, \\ dF^*\tilde{\omega}^1{}_2 &= (F^*\tilde{K})\, F^*\tilde{\theta}^1 \wedge F^*\tilde{\theta}^2, \\ d\omega^1{}_2 &= (F^*K)\,\theta^1 \wedge \theta^2, \end{aligned}$$

So again by uniqueness, $F^*K = K$.

4.5.13 Example Catenoid - Helicoid

Perhaps the most celebrated manifestation of the theorema egregium, is that of mapping between a helicoid M and a catenoid $\widetilde{M}$. Let $a = c$, and label the coordinate patch for the former as $\mathbf{x}(u^\alpha)$ and $\mathbf{y}(\tilde{u}^\alpha)$ for the latter. The first fundamental forms are given as in 4.25 and 4.26.

$$\begin{aligned} ds^2 &= du^2 + (u^2 + a^2)\, dv^2, & & E = 1, \quad G = u^2 + a^2, \\ d\tilde{s}^2 &= \frac{\tilde{u}^2}{\tilde{u}^2 - a^2}\, d\tilde{u}^2 + \tilde{u}^2\, d\tilde{v}^2 & \text{with} \quad & \tilde{E} = \frac{\tilde{u}^2}{\tilde{u}^2 - a^2}, \quad \tilde{G} = \tilde{u}^2. \end{aligned}$$

Let $F : M \to \widetilde{\mathbf{M}}$ be the mapping $\mathbf{y} = F\mathbf{x}$, defined by $\tilde{u}^2 = u^2 + a^2$ and $\tilde{v} = v$. Since $\tilde{u}\; d\tilde{u} = u\; du$, we have $\tilde{u}^2\; d\tilde{u}^2 = u^2\; du^2$ which shows that the mapping preserves the metric and hence it is an isometry. The Gaussian curvatures K and $\tilde{K}$ follow from an easy computation using formula 4.107.

$$K = \frac{-1}{\sqrt{u^2 + a^2}} \frac{\partial}{\partial u}\left(\frac{\partial}{\partial u}\sqrt{u^2 + a^2}\right) = \frac{a^2}{(u^2 + a^2)^2}, \tag{4.114}$$

$$\tilde{K} = -\frac{\sqrt{\tilde{u}^2 - a^2}}{\tilde{u}^2} \frac{\partial}{\partial \tilde{u}}\left(\frac{\sqrt{\tilde{u}^2 - a^2}}{\tilde{u}}\right) = -\frac{a^2}{\tilde{u}^4}. \tag{4.115}$$

It is immediately evident by substitution that as expected $F^*\tilde{K} = K$. Figure 4.13 shows several stages of a one-parameter family M_t of isometries deforming a catenoid into a helicoid. The one-parameter family of coordinate patches chosen is

$$\mathbf{z}_t = (\cos t)\,\mathbf{x} + (\sin t)\,\mathbf{y} \tag{4.116}$$

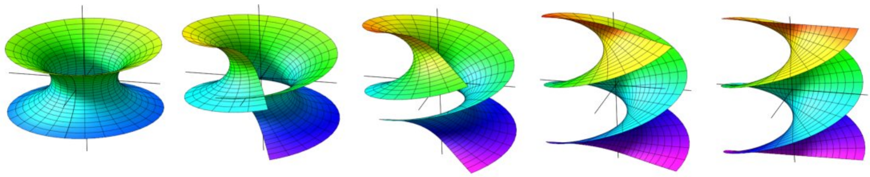


Fig. 4.13: Catenoid - Helicoid isometry

Writing the equation of the coordinate patch $\mathbf{z}_t$ in complete detail, one can compute the coefficients of the fundamental forms and thus establish the family of surfaces has mean curvature H independent of the parameter t, and in fact $H = 0$ for each member of the family. We will discuss at a later chapter the geometry of surfaces of zero mean curvature.

4.5.14 Example Isothermal coordinates.
Consider the case in which the metric has the form

$$ds^2 = \lambda^2\ (du^2 + dv^2), \tag{4.117}$$

so that $E = G = \lambda^2,\ F = 0$. A metric in this form is said to be in *isothermal coordinates.* Inserting into equation 4.107, we get

$$\begin{aligned} K &= -\frac{1}{\lambda^2}\left[\frac{\partial}{\partial u}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial v}\right)\right], \\ &= -\frac{1}{\lambda^2}\left[\frac{\partial}{\partial u}\frac{\partial}{\partial u}(\ln\lambda) + \frac{\partial}{\partial v}\frac{\partial}{\partial v}(\ln\lambda)\right]. \end{aligned}$$

Hence,

$$K = -\frac{1}{\lambda^2}\nabla^2(\ln\lambda). \tag{4.118}$$

The tantalizing appearance of the Laplacian in this coordinate system gives an inkling that there is some complex analysis lurking in the neighborhood. Readers acquainted with complex variables will recall that the real and imaginary parts of holomorphic functions satisfy Laplace's equations and that any holomorphic function in the complex plane describes a conformal map. In anticipation of further discussion on this matter, we prove the following:

4.5.15 Theorem Define the mean curvature vector $\mathbf{H} = H\mathbf{n}$. If $\mathbf{x}(u, v)$ is an isothermal parametrization of a surface, then

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2\lambda^2\mathbf{H}. \tag{4.119}$$

Proof Since the coordinate patch is isothermal, $E = G = \lambda^2$ and $F = 0$. Specifically, we have $<\mathbf{x}_u, \mathbf{x}_u> = <\mathbf{x}_v, \mathbf{x}_v>$, and $<\mathbf{x}_u, \mathbf{x}_v> = 0$. Differentiation then gives:

$$\begin{aligned}
< \mathbf{x}_{uu}, \mathbf{x}_u > &= < \mathbf{x}_v, \mathbf{x}_{vu} >, \\
&= - < \mathbf{x}_{vv}, \mathbf{x}_u >, \\
< \mathbf{x}_{uu} + \mathbf{x}_{vv}, \mathbf{x}_u > &= 0.
\end{aligned}$$

In the same manner,

$$\begin{aligned}
< \mathbf{x}_{vv}, \mathbf{x}_v > &= < \mathbf{x}_u, \mathbf{x}_{uv} >, \\
&= - < \mathbf{x}_{uu}, \mathbf{x}_v >, \\
< \mathbf{x}_{uu} + \mathbf{x}_{vv}, \mathbf{x}_v > &= 0.
\end{aligned}$$

If follows that $\mathbf{x}_{uu} + \mathbf{x}_{vv}$ is orthogonal to the surface and points in the direction of the normal $\mathbf{n}$. On the other hand,

$$\begin{aligned}
\frac{Eg + Ge}{2EG} &= H, \\
\frac{g + e}{2\lambda^2} &= H, \\
e + g &= 2\lambda^2 H, \\
< \mathbf{x}_{uu} + \mathbf{x}_{vv}, \mathbf{n} > &= 2\lambda^2 H, \\
\mathbf{x}_{uu} + \mathbf{x}_{vv} &= 2\lambda^2 \mathbf{H}.
\end{aligned}$$

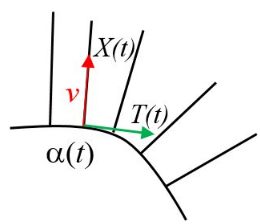
Chapter 5

Geometry of Surfaces

5.1 Surfaces of Constant Curvature

5.1.1 Ruled and Developable Surfaces

We present a brief discussion of surfaces of constant curvature $K = 0$. since $K = k_1 k_2$, a surface with zero Gaussian curvature at each point must have a principal direction with zero normal curvature, that is, either $k_1 = 0$ or $k_2 = 0$. It is therefore a necessary condition for a surface to have $K = 0$, that at each point there be a principal direction which is a straight line. A surface having this property of containing a straight line or segment of a straight line at each point is called a ruled surface. We may think of a ruled surface as a surface generated by the motion of a straight line. Given a point p on a ruled surface, let $\alpha(t)$ be a curve with $\alpha(0) = p$, and let $X(t)$ be a unit vector field on the curve and pointing along the lines at their intersection points with the curve. On can then parametrize the surface near p by a coordinate patch



$$\mathbf{y}(t, v) = \alpha(t) + vX(t),$$

as shown in the figure.

Having a straight line passing through each point in the surface is a necessary but not sufficient condition to ensure that $K = 0$, as illustrated by the following examples:

1) Saddle. Consider the saddle $z = xy$ which is trivially parametrized by the coordinate patch $\mathbf{y}(u, v) = (u, v, uv)$. The patch can be written as $\mathbf{y}(u, v) = (u, 0, 0) + v(0, 1, u)$ or as $\mathbf{y}(u, v) = (0, v, 0) + u(1, 0, v)$, so that the surface is doubly-ruled as shown in figure 5.1(a). The rulings are the coordinate curves $u=$ constant, and $v=$ constant. This neat fact is reflected in some architectural designs of simple structures with roofs made of straight slabs arranged in the

shape of a hyperbolic paraboloid. A short computation gives

$$K = -(1+u^2+v^2)^{-2} = -(1+x^2+y^2)^{-2}$$

2) Hyperboloid. A common calculus example of a doubly ruled surface is given by circular hyperboloids of one sheet. Consider the circle $\alpha(u) = (\cos u, \sin u, 0)$, and the vector field $X(u) = \dot{\alpha} + \mathbf{k}$ which points at a constant skew angle $\pi/4$. Then the (x, y, z) coordinates in the parametrization

$$\begin{aligned}
\mathbf{y}(u,v) &= \alpha(u) + vX(u), \\
&= (\cos u, \sin u, 0) + v(-\sin u, \cos u, 1), \\
&= (\cos u - v\sin u, \sin u + v\cos u, v),
\end{aligned}$$

satisfy the equation $x^2+y^2-z^2 = 1$; that is, the surface is a circular hyperboloid of one sheet. The coordinate curves u =constant are straight line generators. If instead, we choose $X(u) = -\dot{\alpha} + \mathbf{k}$, we get the same surface, but with an orthogonal set of line-generators as shown in figure 5.1(b). This is an example of a surface in which the asymptotic curves are orthogonal at each point. Tangent planes to the surface at any point in the circle $x^2 + y^2 = 1$ at the throat, intersect the hyperboloid in a pair of line generators. The Gaussian curvature is also negative and is given by

$$K = -(1+2v^2)^{-2} = -(1+2z^2)^{-2}.$$

The double-ruled nature of the circular hyperboloid has been exploited by civil engineers in the design of heavy-duty gears with long teeth engaging along the lines. The double-ruling is also advantageous for the construction of the metal frame a type of tower to used in nuclear reactors.

3) Möbius Band. The formal definition of an *orientable Surface M* is that there exists a two-form that is non-zero at every point of M. The idea is that the 2-form represents the oriented differential of surface area $dS = \sqrt{\det g}\, du \wedge dv$. For the present purpose, an intuitive characterization is that there exists a unit normal vector field on M. The Möbius Band is the most famous example of a non-orientable surface. It can be parametrized by the coordinate patch

$$\begin{aligned}
\mathbf{y}(u,v) &= \alpha(u) + vX(u), \\
\alpha(u) &= (\cos 2u, \sin 2u, 0), \\
X(u) &= (\cos u \cos 2u, \cos u \sin 2u, \sin u), \\
\mathbf{x}(u,v) &= (\cos 2u + v\cos u \cos 2u, \sin 2u + v\cos u \sin 2u, v \sin u). \qquad (5.1)
\end{aligned}$$

The curve $\alpha(u)$ is a circle, and the vector $X(u)$ on the circle points in a direction that winds around by an angle π in one revolution. In the rendition of the surface shown in figure 5.1, the parameter v is restricted to $[-0.2, 0.2]$. As is evident from the graph, the generating line segment flips after one turn, resulting on a one-sided surface. Indeed, at $u = 0$ the generating line segment is $vX(0) = v(1,0,0)$ but after one revolution at $u = \pi$, the generating line segment given by $vX(\pi) = v(-1,0,0)$, points in the opposite direction. We can interpret the topology of the surface as a rectangle with a pair of opposite sides identified (see figure 9.2). Clearly, it is impossible to choose a well-defined normal vector field. The Gaussian curvature is somewhat messy, but the computation shows that K is negative everywhere.

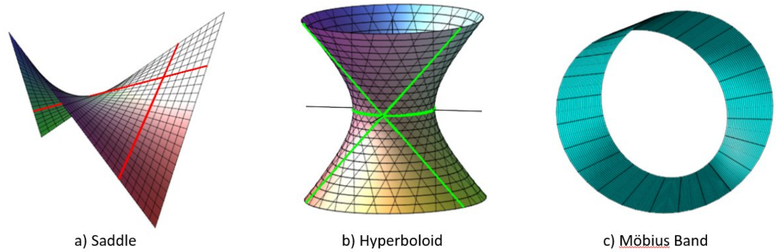


Fig. 5.1: Examples of Ruled Surfaces

Let M be a ruled surface with unit normal N and let X be a unit vector field tangent along the straight lines that generate the surface. The straight lines are geodesic in $\mathbf{R}^3$ so $\overline{\nabla}_X X = 0$. By Gauss's equation 4.74, $\nabla_X X = 0$, so the line is also a geodesic on the surface and $< LX, X >= 0$, that is, the generator lines are asymptotic. Let Y be a unit tangent vector field orthogonal to X so that the pair constitutes an orthogonal basis of the tangent space at each point. Then

$$\begin{aligned} K &=< LX, X >< LY, Y > - < LX, Y >^2, \\ &= - < LX, Y >^2 = -f^2, \end{aligned}$$

so we conclude that $K \leq 0$. If the vectors X and Y are not an orthogonal basis, the result must be modified as in equation 4.66, which gives,

$$K = -\frac{f^2}{EG - F^2}. \tag{5.2}$$

The general formula for the Gaussian curvature of a ruled surface is obtained by a straightforward computation. We have:

$$\begin{aligned} \mathbf{y}_u &= \alpha' + vX', \\ \mathbf{y}_v &= X, \\ \mathbf{y}_u \times \mathbf{y}_v &= (\alpha' + vX') \times X, \\ EG - F^2 &= \|\mathbf{y}_u \times \mathbf{y}_v\| = \|X \times (\alpha' + vX')\|. \\ \mathbf{y}_{uv} &= X', \quad \mathbf{y}_{vv} = 0. \end{aligned}$$

Hence, $g =< \mathbf{y}_{vv}, N >= 0$, and $f =< \mathbf{y}_{uv}, N >= (\alpha' X X')/\sqrt{EG - F^2}$, where we are using the notation for the triple product. The resulting curvature is:

$$K - \frac{(\alpha' X X')}{\|X \times (\alpha' + vX')\|^4} \tag{5.3}$$

We may choose the orthogonal trajectories $\alpha(u)$ to be integral curves of Y parametrized by arc length, so that $Y = \alpha'(u) = T$. Since by choice, T and X are orthogonal unit vectors tangent to the surface, $N = T \times X$ is normal to the surface and we have an orthonormal frame at each point. The covariant derivatives of the frame with respect to T along the one-parameter curves α is just derivative with respect to the parameter, so we get a Frenet-like frame

$$\begin{array}{rcrrr} T' & = & & c_1 X & + \; c_2 N \\ X' & = & -c_1 T & & + \; c_3 N \; . \\ N' & = & -c_2 T & -c_3 X & \end{array} \tag{5.4}$$

A one-line computation gives:

$$c_3 = - < X, N' > = < N, X' > = (TXX').$$

The function

$$p(u) = \frac{(TXX')}{\|X'\|^2} = \frac{c_3}{{c_1}^2 + {c_3}^2},$$

is called the *distribution parameter*. Substituting 5.4 into 5.3, we rewrite the Gaussian curvature as

$$\begin{aligned} K &= -\frac{(TXX')}{\|X \times (T + vX')\|^4}, \\ &= -\frac{{c_3}^2}{[1 - 2c_1 v + {c_1}^2 v^2 + v^2 {c_3}^2]^2}. \end{aligned}$$

The special curve along which $c_1 = 0$ that is, $(T', X) = 0$ is called the *stricture curve*. Using the parametrization with the base curve being the stricture curve, we have $p(u) = 1/c_3$ and

$$K = -\frac{{c_3}^2}{(1 + v^2 c_3(2)} = \frac{p^2(u)}{(p^2(u) + v^2))}. \tag{5.5}$$

A beautiful example is the hyperboloid of revolution in figure 5.1(b). The circle $\alpha(t)$ at the throat used to generate the surface is the stricture curve. It turns out that X does not need to be orthogonal to T as it is the case here, as long as $\|x\| = 1$ and $(T', X) = 0$.

A ruled surface is called a *developable surface* if in addition, $LX = \overline{\nabla}_X N = 0$, that is, the normal vector is parallel along the generating lines. Then equality holds in $K \leq 0$ and we have the following theorem

5.1.1 Theorem A necessary and sufficient condition for surface to be developable is to have Gaussian curvature $K = 0$.

It is surprising that the general case of a closed and connected surface with $K = 0$ to be developable was not proved until 1961 in a short paper by Massey. A particularly interesting type of developable surfaces are those in which the vector X is taken to be the tangent vector T of the curve α itself. A surface with this property is called a *tangential developable*.

5.1.2 Example Developable helicoid. A helicoid

$$\mathbf{x}(u,v) = (u\cos v, u\sin v, v)$$

can be written in the form $\mathbf{x} = \alpha(v) + uX(v)$, where $\alpha(v) = (0,0,v)$ and $X(v) = (\cos v, \sin v, 0)$, so it is a ruled surface. The surface has negative curvature as computed in 4.114 and the stricture curve is the z-axis. A neat related surface is obtained by the tangential developable of a helix. We choose $\alpha(u) = (\cos u, \sin u, u)$ and $X = T = (-\sin u, \cos u, 1)$ so that

$$\mathbf{x}(u,v) = (\cos u - v\sin u, \sin u + v\cos u, u + v). \tag{5.6}$$

Since this is a flat surface having $K = 0$ it is isometric to a plane. Indeed, if one takes a thin cardboard annulus with a slit in the xy-plane with the appropriate radius, one can bend the annulus around a cylinder by lifting one edge of the slit, thus creating a ribbon that wraps around the cylinder as shown in figure 5.2. A magnificent architectural example is exhibited by base of the spiral staircase near the pyramid of the Louvre museum. For the maple-generated image, a small numerical computation was carried to figure out the vertical shift and radius of the helicoid so that the staircase and the supporting developable match at the helix of intersection.

Fig. 5.2: Developable Helicoid

5.1.2 Surfaces of Constant Positive Curvature

In this section we prove a few global theorems. We assume the reader is acquainted with the notion of a compact space. In particular, in $\mathbf{R}^n$ a compact set is closed and bounded so it is contained in a ball of sufficiently large radius, centered at the origin. We are concerned with compact manifolds, which by definition are locally Euclidean and have a differentiable structure. Thus, a compact surface in $\mathbf{R}^3$ cannot have any edges or creases, and the tangent space is well defined at all points.

5.1.3 Theorem In any compact surface in $\mathbf{R}^3$ there exists at least one point p at which $K(p) > 0$.

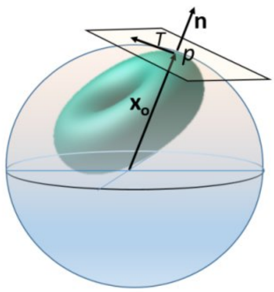


Fig. 5.3: Compact Surface

Proof Let M be compact. Consider the function $f : M \to \mathbf{R}$ defined by $f = \|\mathbf{x}\|$, where $\mathbf{x}$ are local coordinates of a point on the surface. This is a continuous function on compact space, so it attains a maximum at a least one point p. The geometric interpretation of p is that it is farthest away from the origin. The intuition about this theorem is simply that near the point p, the surface is entirely on one side of the tangent plane as shown in figure 5.3, so the principal curvatures have the same sign. We make this formal. Let R be the distance from p to the origin and construct a sphere of radius R centered at the origin. The sphere will be tangential to the surface at p. Given a unit tangent T, let $\alpha(t)$ be unit speed integral curve near p, that is, $\alpha(0) = p$ and $T = \alpha'(0)$. The composite function $f(\alpha(t))$ also has a maximum at p, so by the second derivative test, we have $[f(\alpha)]'(0) = 0$, and $[f(\alpha)]''(0) < 0$. By the definition of f, $f(\alpha) = \|\alpha^2\| = (\alpha, \alpha)$, so $[f(\alpha)]'(0) = 2(\alpha, \alpha')(0) = 0$. We conclude that the position vector $\mathbf{x}_o = \alpha(0)$ of the point p is orthogonal to T. Since this is true for any such T, the vector $\mathbf{x}_o$ is also normal to the surface, so that the unit normal is $\mathbf{n} = \mathbf{x}_o/R$. Computing the second derivative we get

$$\begin{aligned}
\tfrac{1}{2}[f(\alpha)]''(0) &=< \alpha, \alpha' >' (0), \\
&=< T, T > + < \mathbf{x}_o, \alpha''(0) \ >, \\
&= 1 + R < \mathbf{n}, \alpha''(0) \ >< 0,
\end{aligned}$$

But clearly $< \mathbf{n}, \alpha''(0) \ >$ is the normal curvature along T so

$$k_n(p) < -\frac{1}{R},$$

Again, since T was arbitrary, the normal curvature is less than $-1/R$ in any direction, a geometric indication that the surface is bending inward more than the sphere as intuitively shown by the picture. Therefore

$$K(p) = \kappa_1 \kappa_2 > \frac{1}{R^2} > 0.$$

5.1.4 Theorem In a surface in which the coordinate directions are chosen to be the principal directions of curvature, the Codazzi equations are

$$
\begin{aligned}
\frac{\partial \kappa_1}{\partial u} &= \frac{1}{2}\frac{E_v}{E}(\kappa_2 - \kappa_1), \\
\frac{\partial \kappa_2}{\partial u} &= \frac{1}{2}\frac{G_u}{G}(\kappa_1 - \kappa_2)
\end{aligned} \tag{5.7}
$$

Proof Let X and Y be eigenvectors of L, so that

$$
\begin{aligned}
LX = \kappa_1 X, \quad \kappa_1 &= \frac{< LX, X >}{< X, X >} = \frac{e}{E}, \\
LY = \kappa_2 Y, \quad \kappa_1 &= \frac{< LY, Y >}{< Y, Y >} = \frac{g}{G}.
\end{aligned}
$$

If $\kappa_1 \neq \kappa_2$, the eigenvectors are orthogonal, so taken them as the parametric directions means that $F = f = 0$. The Codazzi equations 4.98 are obtained by setting to zero the normal component of $\mathbf{x}_{\alpha\beta\gamma} - \mathbf{x}_{\beta\alpha\gamma} = 0$. In terms of the covariant derivative formulation of the Gauss 4.74 with $X = e_1 = \mathbf{x}_u$, $Y = e_2 = \mathbf{x}_v$, $Z = e_\gamma$, the normal component of $(\overline{\nabla}_X \overline{\nabla}_Y - \overline{\nabla}_Y \overline{\nabla}_X)Z = 0$ result in the equations of Codazzi in the form: (See 6.22)

$$
\begin{aligned}
< \overline{\nabla}_X LY - \overline{\nabla}_X LX, Z > &= 0, \\
\overline{\nabla}_X LY - \overline{\nabla}_X LX &= 0
\end{aligned}
$$

We proceed to expand this equation:

$$
\begin{aligned}
\overline{\nabla}_{e_1} Le_2 - \overline{\nabla}_{e_2} Le_1 &= 0, \\
\overline{\nabla}_{e_1}(\kappa_2 e_2) - \overline{\nabla}_{e_2}(\kappa_1 e_1) &= 0, \\
\frac{\partial \kappa_2}{\partial u} e_2 + \kappa_2 {\Gamma^\alpha}_{12} e_\alpha - \frac{\partial \kappa_1}{\partial v} e_1 - \kappa_1 {\Gamma^\alpha}_{21} e_\alpha &= 0.
\end{aligned}
$$

Setting the e_1 and e_2 components to zero, we get:

$$
\begin{aligned}
\frac{\partial \kappa_2}{\partial_u} &= \kappa_1 {\Gamma^2}_{21} - \kappa_2 {\Gamma^2}_{12} = (\kappa_1 - \kappa_2){\Gamma^2}_{12}, \\
\frac{\partial \kappa_1}{\partial_v} &= \kappa_2 {\Gamma^1}_{12} - \kappa_1 {\Gamma^1}_{21} = (\kappa_2 - \kappa_1){\Gamma^1}_{21}.
\end{aligned}
$$

The result follows immediately from the expressions for the Christoffel symbols 4.83 after setting $F = 0$.

5.1.5 Proposition (Hilbert) Let p be a non-umbilic point and $\kappa_1(p) > \kappa_2(p)$. If κ_1 has a local maximum at p and κ_2 has a local minimum at p, the $K(p) < 0$.
Proof Take the asymptotic curves as parametric curves as in the preceding proposition. Suppose $\kappa_1(p) > \kappa_2(p)$ and that the principal curvatures are local extrema. Then $(\kappa_1)_u = (\kappa_2)_v = 0$ so by equation 5.7, we have $E_v = G_u = 0$.

Applying the second derivative test by differentiating 5.7 at p we get

$$
\begin{aligned}
(\kappa_1)_{vv} &= \frac{1}{2}\frac{E_{vv}}{E}(\kappa_2 - \kappa_1) \leq 0, \\
(\kappa_2)_{uu} &= \frac{1}{2}\frac{G_{uu}}{G}(\kappa_1 - \kappa_2) \geq 0,
\end{aligned}
$$

which implies that $E_{vv} \geq 0$, end $G_{uu} \geq 0$. On the other hand, noting as above that $E_v = G_u = 0$, the Gaussian curvature formula 4.107 gives

$$
K = -\frac{1}{2EG}(E_{vv} + G_{uu}) \leq 0.
$$

5.1.6 Theorem (Liebmann) A compact manifold M in $\mathbf{R}^3$ of constant Gaussian curvature K is a sphere of radius R with $K = 1/R^2$.

Proof Since M is compact, there is at least one point at which $K > 0$, and since K is constant, $K > 0$ everywhere. We prove by contradiction that all points are umbilic. Suppose there exists an non-umbilic point. Without loss of generality, we assume that the larger principal curvature is κ_1. The principal curvatures are continuous functions in a compact space, so there is a point p at which κ_1 is maximum. Since $K = \kappa_1\kappa_2$ =constant then at p, κ_2 is a minimum. By Hilbert's theorem above, $K(p) < 0$ which is a contradiction. So M is a sphere so some radius R and $K = 1/R^2$.

5.1.3 Surfaces of Constant Negative Curvature

The geometry of surfaces of constant negative curvature is very rich and it has a number of neat applications to physics. If $K < 0$, then it must be the case that the principal curvatures κ_1 and κ_2 have different signs. All points on the surface are hyperbolic, and by Hilbert's theorem there are no compact surfaces of constant negative curvature. In addition, since $\kappa_1 \neq \kappa_2$, there always exist orthogonal asymptotic curves with asymptotic directions along the eigenvectors of the second fundamental form. The prototype of a surface of constant negative curvature is the pseudosphere introduced in equation 4.24, which we repeat here for convenience.

$$
\mathbf{x}(u, v) = (a \sin u \cos v, a \sin u \sin v, a(\cos u + \ln(\tan \frac{u}{2})).
$$

To compute the Gaussian curvature we first verify that the first fundamental form is as stated in 4.24. We have:

$$\begin{aligned}
\mathbf{x}_u &= (a\cos u\cos v, a\cos u\sin v, a\frac{\cos^2 u}{\sin u}),\\
\mathbf{x}_v &= (-a\sin u\sin v, a\sin u\cos v, 0),\\
E &= a^2\cos^2 u + a^2\frac{\cos^4 u}{\sin^2 u},\\
&= a^2\cos^2 u(1+\frac{\cos^2 u}{\sin^2 u}),\\
&= a^2\cot^2 u,\\
F &= 0,\\
G &= a^2\sin^2 u.
\end{aligned}$$

So the parametric curves are orthogonal, and

$$I = a^2\cot^2 u\, du^2 + a^2\sin^2 u\, dv^2.$$

Inserting into formula 4.107, we get

$$\begin{aligned}
K &= -\frac{1}{a^2\cos u}\left[\frac{\partial}{\partial u}\left(\frac{\sin u}{a\cos u}\frac{\partial}{\partial u}(a\sin u)\right)\right],\\
&= -\frac{1}{a^2\cos u}\frac{\partial}{\partial u}(\sin u),\\
&= -\frac{1}{a^2}.
\end{aligned} \tag{5.8}$$

Another common parametrization of the pseudosphere is obtained by the substitution

$$\mu = a\ln\tan(\tfrac{u}{2}). \tag{5.9}$$

Without real loss of generality, we set $a = 1$, so $e^\mu = \tan(u/2)$. The substitution is somewhat related to the classical Gudermannian. We have:

$$\begin{aligned}
\operatorname{sech}\mu &= \frac{2}{e^{u/2}+e^{-u/2}},\\
&= \frac{2}{\tan(u/2)+\cot(u/2)},\\
&= 2\sin(u/2)\cos(u/2),\\
&= \sin u,
\end{aligned}
\qquad \text{and} \qquad
\begin{aligned}
\tanh\mu &= \frac{e^{u/2}-e^{-u/2}}{e^{u/2}+e^{-u/2}},\\
&= \frac{\tan(u/2)-\cot(u/2)}{\tan(u/2)+\cot(u/2)},\\
&= \sin^2(u/2)-\cos^2(u/2),\\
&= -\cos u
\end{aligned}$$

In simplifying the equations above we multiplied top and bottom of the fractions by $\sin(u/2)\cos(u/2)$. In terms of the parameter μ, the coordinate patch for the pseudosphere becomes

$$\mathbf{x}(\mu, v) = (a\operatorname{sech}\mu\cos v, a\operatorname{sech}\mu\sin v, a(\mu-\tanh\mu)), \tag{5.10}$$

and the fundamental forms are:

$$\begin{aligned} I &= \tanh^2\mu\, d\mu^2 + \operatorname{sech}^2\mu\, dv^2, \\ II &= -\operatorname{sech}\mu\, d\mu^2 + \operatorname{sech}\mu \tanh\mu\, dv^2. \end{aligned}$$

Using this latter parametrization we compute the surface area and volume of the top half of the pseudosphere.

$$\begin{aligned} S &= \int\int \sqrt{EG - F^2}, \\ &= \int_0^{\pi}\int_0^{\infty} (a \operatorname{sech}\mu)(a \tanh\mu)\, d\mu\, dv, \\ &= 4\pi a^2, && (5.11) \\ V &= \pi a^3 \int_{-\infty}^{\infty} \operatorname{sech}^2\mu \tanh^2\mu\, d\mu, && (5.12) \\ &= \tfrac{2}{3}\pi a^3. && (5.13) \end{aligned}$$

It is interesting to note that the surface area is exactly the same as that of a sphere of radius a, whereas the volume of revolution is half the volume of the sphere. Without loss of understanding of the geometry, for the rest of this section we set $a = 1$, so that $K = -1$ We have the following theorem:

5.1.7 Theorem Let M be a surface with constant negative curvature $K = -1$. If the parametric curves are chosen to be the asymptotic directions, there exists some quantity ω so that the first fundamental form can be written as:

$$I = \cos^2\omega\, du^2 + \sin^2\omega\, dv^2, \tag{5.14}$$

Proof The proof amounts to analyzing the integrability conditions represented by the Codazzi equations. In the brilliant book by Eisenhart [19], the author writes down the Codazzi equations and notes that the choice of E and G above are solutions of the equations. We take a more humble approach and show the steps on how to find a solution using the Cartan formalism. We have seen that using the asymptotic curves as parametric curves means that $F = f = 0$, $\kappa_1 = E/e$ and $\kappa_1 = G/e$. Let $E = \alpha^2$ and $G = \beta^2$ so that the first fundamental form is:

$$I = \alpha^2\, du^2 + \beta^2\, dv^2.$$

We choose $\theta^1 = \alpha\, du$, and $\theta^2 = \beta\, dv$, as basis for the cotangent space dual to $\{e_1, e_1\}$. Let e_3 be the unit normal to the surface. We have:

$$\begin{aligned} Le_1 &= \kappa_1 e_1 = \overline{\nabla}_{e_1} e_3 = \omega^i{}_3(e_1)e_i, \\ Le_2 &= \kappa_2 e_2 = \overline{\nabla}_{e_2} e_3 = \omega^i{}_3(e_2)e_i, \end{aligned}$$

so,

$$\begin{aligned} \omega^1{}_3 &= \kappa_1\theta^1 = \kappa_1\alpha\, du, \\ \omega^2{}_3 &= \kappa_2\theta^2 = \kappa_2\beta\, dv. \end{aligned}$$

On the other hand,

$$
\begin{aligned}
d\theta^1 &= \alpha_v\, dv \wedge du = -(\frac{\alpha_v}{\beta}\, du) \wedge \theta^2 = -\omega^1{}_2 \wedge \theta^2, \\
d\theta^2 &= \beta_u\, du \wedge dv = -(\frac{\beta_u}{\alpha}\, dv) \wedge \theta^1 = -\omega^2{}_1 \wedge \theta^1, \\
\omega^1{}_2 &= \frac{\alpha_v}{\beta}\, du - \frac{\beta_u}{\alpha}\, dv.
\end{aligned} \tag{5.15}
$$

Recall the Codazzi equations 4.104

$$
\begin{aligned}
d\omega^1_{\ 3} &= -\omega^1_{\ 2} \wedge \omega^2_{\ 3}, \\
d\omega^2_{\ 3} &= -\omega^2_{\ 1} \wedge \omega^1_{\ 3}.
\end{aligned}
$$

Inserting the connection forms into the first Codazzi equation gives

$$
(\kappa_1\alpha)_v\, dv \wedge du + [\frac{\alpha_v}{\beta}\, du - \frac{\beta_u}{\alpha}\, dv] \wedge \kappa_2\beta\, dv = 0.
$$

Since $\kappa_1\kappa_2 = -1$, we can eliminate κ_2 and solve for α.

$$
\begin{aligned}
(\kappa_1\alpha)_v - \alpha_v\kappa_2 &= 0, \\
(\kappa_1 - \kappa_2)\alpha_v + (\kappa_1)_v\alpha &= 0, \\
\frac{\alpha_v}{\alpha} &= -\frac{(\kappa_1)_v}{\kappa_1 - \kappa_2}, \\
&= -\frac{(\kappa_1)_v}{\kappa_1 + (1/\kappa_1)}, \\
&= -\frac{\kappa_1(\kappa_1)_v}{\kappa_1{}^2 + 1}, \\
\frac{\partial}{\partial v}(\ln\alpha) &= -\frac{\partial}{\partial v}\ln[(\kappa_1{}^2 + 1)^{1/2}].
\end{aligned}
$$

We may set

$$
\begin{aligned}
\kappa_1 &= \tan\omega, \\
\kappa_2 &= -\cot\omega,
\end{aligned} \tag{5.16}
$$

for some ω. Then $(\kappa_1{}^2 + 1)^{1/2} = \sec\omega$, so

$$
\frac{\partial}{\partial v}(\ln\alpha) = -\frac{\partial}{\partial v}\ln(\sec\omega) = \frac{\partial}{\partial v}\ln(\cos\omega).
$$

We choose the simplest solution $\alpha = \cos\omega$. By a completely analogous computation using the second Codazzi equation, we get $\beta = \sin\omega$ and that proves the theorem.

5.1.8 Theorem If a surface with $K = -1$ has first fundamental form written as $I = \cos^2\omega\ du^2 + \sin^2\omega\ du^2$, then ω satisfies the so-called sine-Gordon equation (SGE):

$$\omega_{uu} - \omega_{vv} = \sin\omega\cos\omega. \tag{5.17}$$

Proof Here $E = \cos^2\omega$ and $G = \sin^2\omega$. The theorem follows immediately from inserting these into the Gauss curvature equation in orthogonal coordinates 4.107 and setting $K = -1$

$$K = -\frac{\omega_{uu} - \omega_{vv}}{\sin\omega\cos\omega} = -1$$

The following transformation is often made:

$$u = \hat{u} + \hat{v} \qquad u = \hat{u} - \hat{v}.$$

A quick computation yields a transformed fundamental form

$$\hat{I} = d\hat{u}^2 + 2\cos\hat{\omega}\ d\hat{u}d\hat{v} + d\hat{v}^2, \tag{5.18}$$

where $\omega = \hat{\omega}/2$. A coordinate system in which the first fundamental form is of this type is called a *Tchebychev patch* (eventually one has to make a choice on how to transliterate from the Cyrillic alphabet). The corresponding curvature equation is

$$\hat{\omega}_{uv} = \sin\hat{\omega} \tag{5.19}$$

The sine-Gordon equation is one of class of very special type of nonlinear partial differential equations which admit *soliton* solutions. This is an incredibly rich area of research that would take us into whole new branch of mathematics. We constrain our discussion to certain transformations that allow one to obtain new solutions from known solutions, and associate these with pseudospherical surfaces, that is, surfaces in $\mathbf{R}^3$ with constant negative curvature. We note that if in the Sine-Gordon equation 5.17 one sets $v = t$ where t is a time parameter, what we have is a non-linear wave equation with speed $v = 1$. The reader will then recognize the transformation $u = \hat{u} + \hat{t}$ $u = \hat{u} - \hat{t}$ as the equations of characteristics. It is thus not surprising that the equation has solutions of the form $f(u - t)$.

5.1.4 Bäcklund Transforms

5.1.9 Definition Let M by a surface with $K = -1$ and let $F : M \to \hat{M}$ be a map to another surface $\hat{M}$. Let $\hat{p} = F(p)$ and $N(p)$ and $\hat{N}(\hat{p})$ be the unit normals at p and $\hat{p}$ respectively. $\hat{M}$ is called a *Bäcklund transform* (BT) of M with constant angle of inclination σ, if for all p:

a) The angle between N and $\hat{N}$ is σ,
b) The distance λ between p and $\hat{p}$ is $\sin\sigma$,
c) the segment $\overline{p\hat{p}}$ is tangent to $\mathcal{M}$ at p.

Bäcklund proved in 1883, that F maps pseudospherical surfaces to pseudospherical surfaces and asymptotic lines to asymptotic lines. He also showed that given

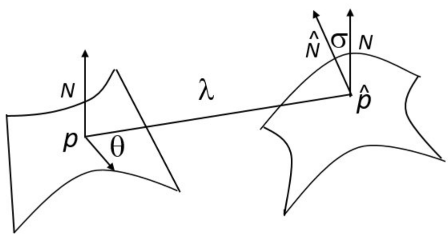


Fig. 5.4: Bäcklund Transform

any unit tangent vector at p that is not an asymptotic direction, a BT exists with $\overline{p\hat{p}}$ in the direction of that tangent. The idea behind the proof of the BT theorem is basically to quantify the transformation from an orthonormal frame at p to the orthonormal frame at $\hat{p}$, find the conditions required for $\hat{K} = -1$, and write down the integrability conditions for the Cartan structure equations. The transformation consists of a rotation by an angle σ, a translation from p to $\hat{p}$ and a rotation by an angle θ to align the frame with segment $\overline{p\hat{p}}$. This could be done all at once, but we prefer to carry out the process in two stages. In the first stage, we apply the translation of the frame assuming that the segment joining p and $\hat{p}$ is parallel to the basis vector e_1 at p, followed by a rotation by an angle σ around the e_1 direction. We use this to seek conditions to guarantee that $\hat{K} = -1$. In stage two, we apply a rotation by an angle θ in the tangent plane.

5.1.10 Theorem Let M have Gaussian curvature $K = -1$, and let $\mathbf{x}(u, v)$ a coordinate patch for M so that $I = E\,du^2 + G\,dv^2$. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 = \mathbf{n}\}$ be an orthonormal frame aligned with the asymptotic directions. Denote the frame and Cartan forms at $\hat{p}$ with hats. Consider the transformation $\hat{\mathbf{x}} = \mathbf{x} + \lambda\mathbf{e}_1$ along with a rotation by an angle σ around the $\mathbf{e}_1$ axis. Then $\hat{K} = -1$ if and only if $\lambda = \sin\sigma$.

Proof A rotation by an angle σ around $\mathbf{e}_1$ leaves the tangent vector e_1 and its dual form θ^1 fixed. The rotation has a matrix representation as shown below.

$$\begin{bmatrix} \hat{\mathbf{e}_1} \\ \hat{\mathbf{e}_2} \\ \hat{\mathbf{e}_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\sigma & -\sin\sigma \\ 0 & \sin\sigma & \cos\sigma \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{bmatrix}. \tag{5.20}$$

We compute the Cartan frame equations. We have: The coframe forms are $\theta^1 = \sqrt{E}\,du$, $\theta^2 = \sqrt{G}\,dv$ and $\theta^3 = 0$ on M. Thus

$$\begin{aligned} d\mathbf{x} &= \mathbf{x}_u\,du + \mathbf{x}_v\,dv, \\ &= \frac{\mathbf{x}_u}{\sqrt{E}}\sqrt{e}\,du + \frac{\mathbf{x}_v}{\sqrt{G}}\sqrt{e}\,dv, \\ &= \frac{\mathbf{x}_u}{\sqrt{E}}\,\theta^1 + \frac{\mathbf{x}_v}{\sqrt{G}}\,\theta^2, \\ &= \mathbf{e}_1\theta^1 + \mathbf{e}_2\theta^2. \end{aligned}$$

We now compute $d\hat{\mathbf{x}}$ taking into account the rotation and then the translation:

$$\begin{aligned}
d\hat{\mathbf{x}} &= \hat{\mathbf{e}}_1\hat{\theta}^1 + \hat{\mathbf{e}}_2\hat{\theta}^2, \\
&= \mathbf{e}_1\theta^1 + [\cos\sigma\; \mathbf{e}_2 - \sin\sigma \mathbf{e}_3]\; \hat{\theta}^2, \\
d\hat{\mathbf{x}} &= d\mathbf{x} + \lambda d\mathbf{e}_1, \\
&= \mathbf{e}_1\theta^1 + \mathbf{e}_2\theta^2 + \lambda(\mathbf{e}_2\,\omega^2{}_1 + \mathbf{e}_3\,\omega^3{}_1)
\end{aligned}$$

Equating the coefficients of $\mathbf{e}_2$ and $\mathbf{e}_3$ in two equations above, we get

$$\begin{aligned}
\cos\sigma\; \hat{\theta}^2 &= \theta^2 + \lambda\omega^2{}_1, \\
-\sin\sigma\; \hat{\theta}^2 &= \lambda\,\omega^3{}_1.
\end{aligned} \tag{5.21}$$

Recall from equation 4.111, that $\omega^1{}_3 = l\theta^1 + m\theta^2$ and $\omega^2{}_3 = m\theta^1 + n\theta^2$ yield symmetric matrix components of the second fundamental form in the given basis. Using this fact and wedging with $\hat{\theta}^1$ the second equation above, we get

$$\begin{aligned}
-\sin\sigma\; \hat{\theta}^1 \wedge \hat{\theta}^2 &= \lambda\; \hat{\theta}^1 \wedge \omega^3{}_1, \\
&= \lambda\; \theta^1 \wedge (l\; \theta^1 + m\; \theta^2), \\
\hat{\theta}^1 \wedge \hat{\theta}^2 &= -\frac{\lambda m}{\sin\sigma}\theta^1 \wedge \theta^2.
\end{aligned} \tag{5.22}$$

Multiplying the first equation in 5.21 by $\sin\sigma$, the second by $\cos\sigma$, and adding, we get

$$\sin\sigma[\theta^2 + \lambda\;\omega^2{}_1] = -\lambda\;\cos\sigma\;\omega^3{}_1, \tag{5.23}$$

$$\theta^2 = -\frac{\lambda}{\sin\sigma}\;[\sin\sigma\;\omega^2{}_1 + \cos\sigma\;\omega^3{}_1]. \tag{5.24}$$

Next, we compute $\hat{\omega}_{32}$:

$$\begin{aligned}
\hat{\omega}_{32}(X) &=< \overline{\nabla}_X\hat{e}_2, \hat{e}_3 >, \\
&=< \cos\sigma\overline{\nabla}_X e_3 - \sin\sigma\overline{\nabla}_X e_3, \sin\sigma\; e_2 + \cos\sigma\; e_3 >, \\
&= \cos^2\sigma < \overline{\nabla}_X e_2, e_3 > - \sin^2\sigma < \overline{\nabla}_X e_3, e_2 >, \\
&= (\cos^2\sigma + \sin^2\sigma) < \overline{\nabla}_X e_2, e_3 >, \\
&= \omega_{32}(X), \\
\hat{\omega}^3_{\;2} &= \omega^3{}_2 = -m\;\theta^1 - n\;\theta^2
\end{aligned}$$

By the same process, we calculate $\hat{\omega}_{31}$:

$$\begin{aligned}
\hat{\omega}_{31}(X) &=< \overline{\nabla}_X\hat{e}_1, \hat{e}_3 >, \\
&=< \overline{\nabla}_X e^3 + \overline{\nabla}_X e^3, \sin\sigma\; e^2 + \cos\sigma\; e^3 >, \\
&=< \omega^2{}_1(X)e_2 + \omega^3{}_1(X)e_3, \sin\sigma\; e_2 + \cos\sigma\; e_3 >, \\
&= \sin\sigma\omega_{21}(X) + \sin\sigma\omega_{31}(X), \\
\hat{\omega}^3_{\;1} &= \frac{\sin\sigma}{\lambda}\theta^2.
\end{aligned}$$

Finally, putting these results together, we get

$$\begin{aligned} d\hat{\omega}^1{}_2 &= -\hat{\omega}^1_3 \wedge \hat{\omega}^3_2, \\ &= -\left[\frac{m\sin\sigma}{\lambda}\right]\ \theta^1 \wedge \theta^2, \\ &= \left[\frac{m\sin\sigma}{\lambda}\right]\left[\frac{\sin\sigma}{m\,\lambda}\right]\hat{\theta}^1 \wedge \hat{\theta}^2, \\ &= \left[\frac{\sin\sigma}{\lambda}\right]^2\ \hat{\theta}^1 \wedge \hat{\theta}^2. \end{aligned}$$

Hence

$$K = -\left[\frac{\sin\sigma}{\lambda}\right]^2 = -1$$

if and only if $\lambda = \pm\sin\sigma$. We choose $\lambda = \sin\sigma$

The conclusion of the theorem explains the condition in the definition of a BT that requires this equation to hold. With this condition, equation 5.23 takes the form:

$$\omega^1{}_2 = \cot\sigma\ \omega^3{}_1 + \csc\sigma\ \theta^2. \tag{5.25}$$

We move to stage two of the BT process.

5.1.11 Theorem Let M be a pseudospherical surface with first fundamental form as in 5.14, and let $F : M \to \hat{M}$ be a BT with angle of inclination σ. If the segment $\overline{p\hat{p}}$ makes a constant angle α with the basis vector e_1 at each point $p \in M$, then

$$\begin{aligned} \sin\sigma(\theta_u + \omega_v) &= \quad\sin\theta\cos\omega - \cos\sigma\cos\theta\sin\omega, \\ \sin\sigma(\theta_v + \omega_u) &= -\cos\theta\sin\omega + \cos\sigma\sin\theta\cos\omega. \end{aligned} \tag{5.26}$$

Proof Suppose $\overline{p\hat{p}}$ makes an angle θ with the tangent vector e_1. In this case we first perform a rotation of axis around the normal vector e_3 to align the frame with e_1. The rotation can be represented by a matrix $\overline{e}_i = e_j\, B^j{}_i$

$$B = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.27}$$

The effect on the Cartan frame is much easier establish since all we have to do is apply the change of basis formula 3.48 as shown in 3.51.

$$\begin{aligned} \overline{\omega}^1{}_2 &= \omega^1{}_2 - d\theta, \\ \overline{\omega}^1{}_3 &= \cos\theta\ \omega^1{}_3 + \sin\theta\ \omega^2{}_3, \\ \overline{\omega}^2{}_3 &= -\sin\theta\ \omega^1{}_3 + \cos\theta\ \omega^2{}_3. \end{aligned} \tag{5.28}$$

The dual forms transform with $A = B^{-1} = B^T$, that is $\overline{\theta}^i = A^i{}_j\theta^j$. In particular,

$$\overline{\theta}^2 = -\sin\theta\ \theta^1 + \cos\theta\ \theta^2. \tag{5.29}$$

If we start with a pseudospherical surface with $I = \cos^2\omega\ du^2 + \sin^2\omega\ dv^2$, the Cartan forms are:

$$\begin{aligned}
\theta^1 &= \cos\omega\ du, \quad \theta^2 = \sin\omega\ dv,\\
\omega^1{}_3 &= \sin\omega\ du, \quad \omega^2{}_3 = -\cos\omega\ dv\\
\omega^1{}_2 &= -\omega_v\ du - \omega_u\ dv.
\end{aligned}$$

The BT transformation is the composition of the stage one and stage two. This means that we must subject the Cartan forms to the change of basis by a rotation by θ, as described by equations 5.28 and 5.29, followed by substitution into equation 5.25. We get:

$$(\omega^1{}_2 - d\theta) = \cot\sigma\ (\cos\theta\ \omega^1{}_3 + \sin\theta\ \omega^2{}_3) + \csc\sigma(-\sin\theta\ \theta^1 + \cos\theta\ \theta^2).$$

We extract two formulas obtained by equating the coefficients of du and dv respectively.

$$\begin{aligned}
-\omega_v - \theta_u &= \quad \cot\sigma\cos\theta\sin\omega - \csc\sigma\sin\theta\cos\omega,\\
-\omega_u - \theta_v &= -\cot\sigma\sin\theta\cos\omega + \csc\sigma\cos\theta\sin\omega.
\end{aligned}$$

The theorem follows by multiplying these equations by $\sin\sigma$, and rearranging terms. The system of equations 5.26 is the *classical Bäcklund transform.* In the special case in which $\sigma = \pi/2$, the angle between the normals e_3 and $\hat{e}_3$ is a right angle, so $\hat{e}_3$ is parallel to a tangent vector of M. This is called a *Bianchi transform.* Equations 5.26 then reduce to the much simpler system:

$$\begin{aligned}
\theta_u + \omega_v &= \quad \sin\theta\cos\omega,\\
\theta_v + \omega_u &= -\cos\theta\sin\omega.
\end{aligned} \tag{5.30}$$

We can rewrite the BT-equations in the so-called asymptotic coordinates. Let

$$\begin{aligned}
u &= x + t, \\
v &= x - t,
\end{aligned} \qquad \text{so that} \qquad \begin{aligned}
x &= \frac{1}{2}(u + v),\\
t &= \frac{1}{2}(u - v).
\end{aligned}$$

By the chain rule, we have:

$$\begin{aligned}
\theta_u &= \frac{1}{2}(\theta_x + \theta_t),\\
\theta_v &= \frac{1}{2}(\theta_x - \theta_t),
\end{aligned} \qquad \text{and} \qquad \begin{aligned}
\omega_u &= \frac{1}{2}(\omega_x + \omega_t),\\
\omega_v &= \frac{1}{2}(\theta_x - \theta_t).
\end{aligned}$$

Adding and subtraction equations 5.26, the system reduces to

$$\begin{aligned}
\theta_x + \omega_x &= \frac{1 + \cos\sigma}{\sin\sigma}\sin(\theta - \omega),\\
\theta_t - \omega_t &= \frac{1 - \cos\sigma}{\sin\sigma}\sin(\theta + \omega),
\end{aligned}$$

which we can rewrite as

$$\begin{aligned} \theta_t &= \quad \omega_t + s\sin(\theta+\omega), \\ \theta_x &= -\omega_x + \frac{1}{s}\sin(\theta-\omega),, \end{aligned} \tag{5.31}$$

where $s = \tan(\sigma/2.)$ We denote the system of BT equations 5.31 by the notation $F = F(\omega, \theta, s)$.

Given a pseudospherical Surface S associated with a solution ω of the sine-Gordon equation, the transform $F = F(\omega, \theta, \sigma)$ produces a new solution θ associated with a new pseudospherical surface S'. Of course, the process can be iterated to produce new surfaces and new solutions. The neat thing is that further iterations can be carried out algebraically without the need to solve more differential equations. This remarkable result is encapsulated in the following theorem (see [19]).

5.1.12 Theorem (Bianchi permutability) Let $\{S, \omega\}$ be the pair consisting of a pseudospherical surface corresponding to a SGE solution ω. Suppose that $\sin^2\theta_1 \neq \sin^2\theta_2$, and that $\{S_1, \theta_1\}$ and $\{S_2, \theta_2\}$ are pseudospherical surfaces generated respectively from surface S by BT's

$$\begin{aligned} S &\xrightarrow{F(\omega,\theta_1,s_1)} S_1, \\ S &\xrightarrow{F(\omega,\theta_2,s_2)} S_2. \end{aligned}$$

Then, the pair $\{S', \Omega\}$ consisting of a pseudospherical surface S' with SGE solution Ω can be found algebraically by requiring the compatibility of BT's

$$\begin{aligned} S_1 &\xrightarrow{F(\theta_1,\Omega,s_2)} S', \\ S_2 &\xrightarrow{F(\theta_2,\Omega,s_1)} S'. \end{aligned}$$

Proof It suffices to use only one of each pair of BT's. By assumption,

$$\begin{aligned} (\theta_1)_t &= \omega_t + s_1\sin(\theta_1+\omega) \\ (\theta_2)_t &= \omega_t + s_2\sin(\theta_2+\omega) \end{aligned} \quad \text{and} \quad \begin{aligned} \Omega_t &= (\theta_1)_t + s_2\sin(\Omega+\theta_1) \\ \Omega_t &= (\theta_2)_t + s_1\sin(\Omega+\theta_2) \end{aligned}$$

Adding the difference of the two equations on the left with the difference of the two equations on the right, we see that all derivatives cancel, and we get:

$$s_1[\sin(\theta_1+\omega) - \sin(\Omega+\theta_2)] + s_2[\sin(\Omega+\theta_1) - \sin(\theta_2+\omega)] = 0.$$

If we have quantities A and B such that $As_1 + Bs_2 = 0$, then

$$\begin{aligned} \frac{s_1}{s_2} &= -\frac{B}{A}, \\ 1 + \frac{s_1}{s_2} &= \frac{s_2+s_1}{s_2} = \frac{A-B}{A}, \\ 1 - \frac{s_1}{s_2} &= \frac{s_2-s_1}{s_2} = \frac{A+B}{A}, \\ \frac{s_2+s_1}{s_2-s_1} &= \frac{A-B}{A+B}. \end{aligned}$$

Applying this to the equation above, we have

$$\begin{aligned}\frac{s_2+s_1}{s_2-s_1} &= \frac{[\sin(\theta_1+\omega)-\sin(\Omega+\theta_2)]-[\sin(\Omega+\theta_1)-\sin(\theta_2+\omega)]}{[\sin(\theta_1+\omega)-\sin(\Omega+\theta_2)]+[\sin(\Omega+\theta_1)-\sin(\theta_2+\omega)]},\\ &= \frac{[\sin(\theta_1+\omega)-\sin(\Omega+\theta_1)]+[\sin(\theta_2+\omega)-\sin(\Omega+\theta_2)]}{[\sin(\theta_1+\omega)+\sin(\Omega+\theta_1)]-[\sin(\theta_2+\omega)+\sin(\Omega+\theta_2)]},\end{aligned}$$

Using the sum-product formulas for sine functions we rewrite the equation as,

$$\begin{aligned}\frac{s_2+s_1}{s_2-s_1} &= \frac{2\sin[\frac{1}{2}(\omega-\Omega)]\cos[\frac{1}{2}(\Omega+\omega+2\theta_1)]+2\sin[\frac{1}{2}(\omega-\Omega)]\cos[\frac{1}{2}(\Omega+\omega+2\theta_2)]}{2\cos[\frac{1}{2}(\omega-\Omega)]\sin[\frac{1}{2}(\Omega+\omega+2\theta_1)]-2\cos[\frac{1}{2}(\omega-\Omega)]\sin[\frac{1}{2}(\Omega+\omega+2\theta_2)]},\\ &= \frac{2\sin[\frac{1}{2}(\Omega-\omega)]\{\cos[\frac{1}{2}(\Omega+\omega+2\theta_1)]+\cos[\frac{1}{2}(\Omega+\omega+2\theta_2)]\}}{2\cos[\frac{1}{2}(\omega-\Omega)]\{\sin[\frac{1}{2}(\Omega+\omega+2\theta_1)]-\sin[\frac{1}{2}(\Omega+\omega+2\theta_2)]\}},\end{aligned}$$

Now, using the sum-product formulas again, we get,

$$\begin{aligned}\frac{s_2+s_1}{s_2-s_1} &= \frac{4\sin[\frac{1}{2}(\Omega-\omega)]\{\cos[\frac{1}{4}(2\theta_1-2\theta_2)]\cos[\frac{1}{4}(2\Omega+2\omega+2\theta_1+2\theta_2)]\}}{4\cos[\frac{1}{2}(\omega-\Omega)]\{\sin(\frac{1}{4}(2\theta_1-2\theta_2))\cos[\frac{1}{4}(2\Omega+2\omega+2\theta_1+2\theta_2)]\}},\\ &= \frac{4\sin[\frac{1}{2}(\Omega-\omega)]\cos[\frac{1}{2}(\theta_1-\theta_2)]}{4\cos[\frac{1}{2}(\omega-\Omega)]\sin[\frac{1}{2}(\theta_2-\theta_1)]},\end{aligned}$$

We conclude that,

$$\tan\left(\frac{\Omega-\omega}{2}\right)=\frac{s_2+s_1}{s_2-s_1}\tan\left(\frac{\theta_2-\theta_1}{2}\right) \tag{5.32}$$

It is easy to write coordinate patch equations for a BT. The vector $\hat{\mathbf{x}}-\mathbf{x}$ must be a vector of length $\sin\sigma$ which is tangent to M at each point p and makes and angle θ with e_1. Therefore, we must have

$$\hat{\mathbf{x}}-\mathbf{x}=\sin\sigma[\cos\theta\; e_1+\sin\theta\; e_2].$$

But, $e_1=\mathbf{x}_u/\sqrt{E}$ and $e_2=\mathbf{x}_v/\sqrt{G}$, so we have:

$$\hat{\mathbf{x}}=\mathbf{x}+\sin\sigma\left[\frac{\cos\theta}{\cos\omega}\mathbf{x}_u+\frac{\sin\theta}{\sin\omega}\mathbf{x}_v\right]. \tag{5.33}$$

5.1.13 Example *Pseudosphere and one-soliton solution*

If we are willing to sacrifice a bit of rigor for the sake of intuition, we can motivate the derivation of the standard parametric representation of the pseudosphere directly from the Bianchi transform. Recall that in the steps leading to the SGE, we chose the principal curvatures (see equation 5.16) such that $\kappa_1=\tan\omega$ and $\kappa_2=-\cot\omega$. Then, as ω approaches zero, κ_1 also approaches zero, while κ_2 becomes arbitrarily large, so as to maintain $K=-1$. The result is a degenerate surface that collapses onto a straight line. We may think of it as an infinitely long trumpet of infinitesimal diameter. As such, we pick a degenerate patch of the form

$$\mathbf{x}(u,v)=(0,0,u)$$

We set $e_1 = (0,0,1)$ and, anticipating a surface of revolution, we pick $e_2 = (\cos\phi, \sin\phi, 0)$. Recalling the BT coordinate patch in equation 5.33

$$\hat{\mathbf{x}} = \mathbf{x} + \sin\sigma[\cos\theta\ e_1 + \sin\theta\ e_2], \tag{5.34}$$

we consider the case with $\sin\sigma = 1$. With ω arbitrarily close to zero, the Bianchi transform equations 5.30 become,

$$\theta_u = \sin\theta,\ \theta_v = 0.$$

We set $v = \phi$ and without loss of generality, we pick the constant of integration in the first equation above to be 0. The elementary integral gives immediately the stationary one-soliton solution

$$u = \ln(\tan\tfrac{\theta}{2}), \quad \text{or,} \quad \theta = 2\tan^{-1}(e^u).$$

The coordinate patch for the corresponding surface then gives,

$$\begin{aligned}\hat{\mathbf{x}} &= (0,0,\ln(\tan\tfrac{\theta}{2})) + \cos\theta(0,0,1) + \sin\theta(\cos\phi, \sin\phi, 0),\\ &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta + \ln(\tan\tfrac{\theta}{2}))\end{aligned}$$

which agrees with the parametrization of the pseudosphere given in equation 4.24. In other words, the angle θ is precisely as shown in figure 4.6, consistent with the geometry of the Bianchi transform stating that the segment joining the corresponding points is tangential to the surface generated.

5.1.14 Example *Dini's surface*
Another well-known surface that can be obtained by BT results in "twisting" the pseudosphere in a helicoidal manner. Dini's surface is obtained by removing the special condition $\sin\sigma = 1$. Again, if one begins with the trivial solution $\omega = 0$ of the Sine-Gordon equation 5.19, the BT equations 5.31 reduce to,

$$\begin{aligned}\theta_t &= s\sin\theta,\\ \theta_x &= \frac{1}{s}\sin\theta.\end{aligned}$$

Integrating, we get

$$\ln(\tan\tfrac{\theta}{2}) = sx + \frac{1}{s}t, \tag{5.35}$$

where we have set the constant of integration to 0. Solving for θ, we are lead to the moving one-soliton solution

$$\theta(x,t) = 2\tan^{-1}(e^{sx+\frac{1}{s}t}),$$

We carry out the parametrization using the BT equations in 5.26. Again, assuming ω is arbitrarily close to zero, the equations reduce to

$$\begin{aligned}(\sin\sigma)\theta_u &= \sin\theta, \\ (\sin\sigma)\theta_v &= \cos\sigma\sin\theta.\end{aligned} \quad \text{or} \quad \begin{aligned}(\csc\theta)\theta_u &= 1/\sin\sigma\\ (\csc\theta)\theta_v &= \cos\sigma/\sin\sigma\end{aligned}$$

Rewriting the equations in differential form

$$d\,[\ln\tan(\tfrac{\theta}{2})] = \frac{1}{\sin\sigma}du + \frac{\cos\sigma}{\sin\sigma}dv,$$

we can integrate immediately,

$$\ln\tan(\tfrac{\theta}{2}) = \frac{u + v\cos\sigma}{\sin\sigma} = \chi. \tag{5.36}$$

Here we have used χ to denote the expression on the right hand side and we set the integration constant to 0. In terms of these coordinates, the moving soliton solution is

$$\theta(u,v)) = 2\tan^{-1}(e^{\chi}).$$

Using the same degenerate patch 5.34, the parametrization of the surface becomes

$$\begin{aligned}\hat{\mathbf{x}} &= (0,0,u) + \sin\sigma[\cos\theta(0,0,1) + \sin\theta(\cos\phi, \sin\phi, 0)],\\ &= (\sin\sigma\sin\theta\cos\phi, \sin\sigma\sin\theta\sin\phi, u - \sin\sigma\cos\theta)).\end{aligned}$$

Finally, using the results in equations 5.10, we rewrite the parametrization of Dini's surface as

$$\hat{\mathbf{x}} = (\sin\sigma\cos\phi\,\mathrm{sech}\,\chi, \sin\sigma\sin\theta\sin\phi\,\mathrm{sech}\,\chi, u - \sin\sigma\tanh\chi)). \tag{5.37}$$

Notice that as expected, when $\sigma = \pi/2$, that is, when $\sin\sigma = 1$, we get $\chi = u$, and the equation reduces to a pseudosphere (see equation 5.10). Another common parametrization of Dini's surface in which the geometry is more intuitive is given by:

$$\mathbf{x}(u,v) = (\ a\cos u\cos v, a\cos u\sin v, a(\cos u + \ln\tan(\frac{u}{2})) + bv\ ). \tag{5.38}$$

This surface has curvature $K = -1$ when $a^2 + b^2 = 1$, and it has an unfolding infundibular shape, as shown in figure 5.5, with parameters $u \in [0,2],\ v \in [0,4\pi]$ and $a = 1,\ b = 0.5, 0.2$. The surface is essentially generated by revolving the tractrix profile curve of the pseudosphere about the central axis, while at the same time translating the curve at a constant rate parallel to the axis. The meridians traced by the parametric curves u =constant are helices. When $b = 0$, the equation gives a pseudosphere.

5.1.15 Example *Kuen surface*

Applying the permutability theorem 5.32 to solution 5.1.4 we obtain immediately the two-soliton solution

$$\Omega = 2\tan^{-1}\left[\frac{s_2 + s_1}{s_2 - s_1}\frac{e^{s_1 t + \frac{1}{s_1}t} - e^{s_2 t + \frac{1}{s_2}t}}{1 + e^{(s_1+s_2)t + (\frac{1}{s_1} + \frac{1}{s_2})t}}\right]. \tag{5.39}$$

In this example, we perform a Bianchi transformation of a one-soliton pseudosphere to obtain a Kuen surface which is associated with a two-soliton solution. We begin with the parametrization of a pseudosphere given by 5.10 with $a = 1$,

$$\mathbf{x}(u,v) = (\mathrm{sech}\,\mu\cos v, \mathrm{sech}\,\mu\sin v, (\mu - \tanh\mu)).$$

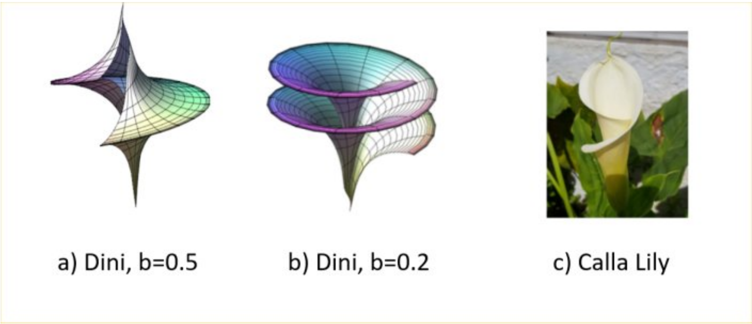


Fig. 5.5: Dini's Surface

Let $\omega = 2\tan^{-1}(e^{\mu})$, so that

$$\mu = \ln\tan(\omega/2),$$

Then,

$$\begin{aligned} \omega &= \quad 2\tan^{-1}(e^{\mu}), \\ \sin\omega &= \quad \operatorname{sech}\mu, \\ \cos\omega &= -\tanh\mu. \end{aligned}$$

We will find θ by solving the Bianchi equations 5.30. We compute:

$$\begin{aligned} \omega_{\mu} &= \frac{2e^{\mu}}{1+e^{2\mu}} = \frac{2}{e^{\mu}+e^{\mu}} = \operatorname{sech}\mu, \\ \omega_{v} &= 0. \end{aligned}$$

Substituting into the Bianchi equations, we get:

$$\begin{cases} \theta_{\mu} &= -\sin\theta\tanh\mu, \\ \theta_{v} &= -\cos\theta\operatorname{sech}\mu - \operatorname{sech}\mu = -\operatorname{sech}\mu(1+\cos\theta) = -2\cos^2(\frac{\theta}{2})\operatorname{sech}\mu. \end{cases}$$

Separate variables

$$\begin{cases} \csc\theta\ \theta_{\mu} &= -\tanh\mu, \\ \frac{1}{2}\sec^2(\frac{\theta}{2})\ \theta_{v} &= -\operatorname{sech}\mu, \end{cases}$$

and integrate. The result is:

$$\begin{cases} \tan(\frac{\theta}{2}) = -h_1(v)\operatorname{sech}\mu, \\ \tan(\frac{\theta}{2}) = -v\operatorname{sech}\mu + h_2(\mu), \end{cases}$$

where h_1 and h_2 are the arbitrary functions of integration. Consistency of the equations requires $h_1 = 1$ and $h_2 = 0$. The solution is therefore

$$\begin{aligned} \tan(\tfrac{\theta}{2}) &= -v\operatorname{sech}\mu = \frac{-v}{\cosh\mu}, \\ \theta &= 2\tan^{-1}(-v\operatorname{sech}\mu). \end{aligned}$$

Only the cosine and the sine of the angle θ enter into the Bianchi coordinate patch. Thinking of $\tan(\frac{\theta}{2})$ as the ratio of the opposite over the adjacent side of the right triangle with hypothenuse $\sqrt{\cosh^2\mu + v^2}$, we can compute the sine and cosine from the double angle formulas:

$$\begin{aligned}\cos\theta &= \cos^2(\tfrac{\theta}{2}) - \sin^2(\tfrac{\theta}{2}) = \frac{\cosh^2\mu - v^2}{\cosh^2\mu + v^2},\\ \sin\theta &= 2\sin(\tfrac{\theta}{2})\cos(\tfrac{\theta}{2}) = \frac{-2v\cosh\mu}{\cosh^2\mu + v^2}.\end{aligned}$$

It remains to go through the algebraic gymnastics of computing the coordinate patch:

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{x} + \frac{\cos\theta}{\cos\omega}\mathbf{x}_u + \frac{\sin\theta}{\sin\omega}\mathbf{x}_v,\\ &= (\,\operatorname{sech}\mu\cos v, \operatorname{sech}\mu\sin v, (\mu - \tanh\mu)\,)\\ &\quad - \frac{\cos\theta}{\tanh\mu}(\,-\operatorname{sech}\mu\tanh\mu\cos v, -\operatorname{sech}\tanh\mu\sin v, \tanh^2\mu\,)\\ &\quad + \frac{\sin\theta}{\operatorname{sech}\mu}(\,-\operatorname{sech}\mu\sin v, \operatorname{sech}\mu\cos v, 0\,),\\ &= (\,\operatorname{sech}\mu\cos v, \operatorname{sech}\mu\sin v, (\mu - \tanh\mu)\,)\\ &\quad + \cos\theta(\,\operatorname{sech}\mu\cos v, \operatorname{sech}\sin v, -\tanh\mu\,)\\ &\quad + \sin\theta(\,-\sin v, \cos v, 0\,),\end{aligned}$$

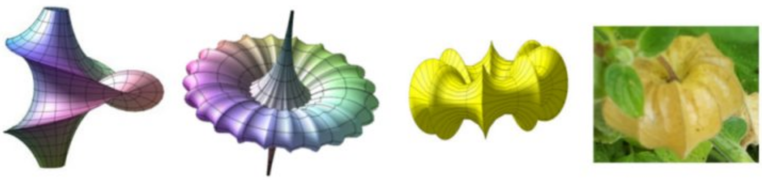


Fig. 5.6: Surfaces with K= -1

The x-component of $\mathbf{x}$ is:

$$\begin{aligned}\mathbf{x}_{(x)} &= (1+\cos\theta)\operatorname{sech}\mu\cos v - \sin\theta\sin v\\ &= \left[1 + \frac{\cosh^2\mu - v^2}{\cosh^2\mu + v^2}\right]\cos v + \left[\frac{2v\cosh\mu}{\cosh^2\mu + v^2}\right]\sin v,\\ &= \left[\frac{2\cosh^2\mu}{\cosh^2\mu + v^2}\right]\cos v + \left[\frac{2v\cosh\mu}{\cosh^2\mu + v^2}\right]\sin v,\\ &= \frac{2\cosh\mu(\cos v + v\sin v)}{\cosh^2\mu + v^2}.\end{aligned}$$

The computation of the other two components is left as an exercise. The result is

$$\mathbf{x}(u,v) = \left(\frac{2\cosh\mu(\cos v - v\sin v)}{\cosh^2\mu + v^2}, \frac{2\cosh\mu(\sin v - v\cos v)}{\cosh^2\mu + v^2}, \mu - \frac{2\sinh 2\mu}{\cosh^2\mu + v^2} \right) \tag{5.40}$$

The Kuen surface in figure 5.6a is plotted with parameters $u \in [-1.4, 1.4]$ and $v \in [-4, 4]$.

As noted by Terng and Uhnlenbeck [38], if in the 2-soliton equation 5.39 one sets $s_1 = e^{i\theta}$ and $s_2 = -e^{-i\theta}$, we get a real-valued solution

$$\phi = 2\tan^{-1}\left[\frac{\sin\theta\sin(\eta\cos\theta)}{\cos\theta\cosh(\xi\sin\theta)}\right], \tag{5.41}$$

where $\xi = x - t$ and $\eta = x + t$. This is a periodic solution called a breather. A rendition of the surface associated with this solution is shown in figure 5.6, using the parametrization derived by Rogers and Schief [31].

5.2 Minimal Surfaces

5.2.1 Minimal Area Property

In an earlier chapter we defined a minimal surface to be a surface of mean curvature $H = 0$. From the formula for the mean curvature 4.65, a surface in $\mathbf{R}^3$ is minimal if

$$2H = \mathrm{Tr}(g^{-1}b) = 0, \quad \text{which implies} \quad Eg - 2Ff + Ge = 0.$$

For historical reasons we first consider this condition for a surface with equation $z = f(x, y)$. We rewrite in parametric form using a Monge coordinate patch

$$\mathbf{x}(x,y) = (x, y, f(x,y)\ )$$

A quick computation yields the coefficients of the first and second fundamental forms.

$$\begin{aligned} E &= 1 + {f_x}^2, & & & e &= f_{xx}/D, \\ G &= 1 + {f_y}^2, & &\text{and} & g &= f_{yy}/D, \\ F &= f_x f_y, & & & f &= f_{xy}/D, \end{aligned}$$

where $D = \sqrt{EG - F^2} = \sqrt{1 + {f_x}^2 + {f_y}^2}$. The surface has Gaussian curvature $K = 0$ if $f(x, y)$ satisfies the Monge-Ampere equation

$$f_{xx}f_{yy} - f_{xy}^2 = 0.$$

We have already determined that the solutions are developable surfaces. The condition $H = 0$ for having a minimal surface is that $f(x, y)$ satisfies the quasi-linear differential equation:

$$(1 + {f_x}^2)f_{xx} - 2f_x f_y f_{xy} + (1 + {f_y}^2)f_{yy} = 0. \tag{5.42}$$

Using the notation $p = f_x,\ q = f_y$ the condition that the surface area over a region be minimal, follows from the variational equation:

$$\begin{aligned}\delta\int\!\!\int\sqrt{EG-F^2}\,dy\,dx &= \delta\int\!\!\int\sqrt{1+p^2+q^2}\,dy\,dx,\\ &= \delta\int\!\!\int F(z,x,y,p,q)\,dy\,dx = 0.\end{aligned}$$

The Euler-Lagrange equation for this functional is

$$\nabla\cdot\left[\frac{\nabla f}{1+|\nabla f|^2}\right] = \frac{d}{dx}\frac{p}{\sqrt{1+p^2+q^2}} + \frac{d}{dy}\frac{q}{\sqrt{1+p^2+q^2}} = 0. \tag{5.43}$$

It was proved by Lagrange in 1762, that these are equivalent to the condition $H = 0$ exemplified in equation 5.42, but he was unable to find non-trivial solutions. In 1776 Meusnier showed that the catenoid and the helicoid satisfied Euler-Lagrange equations 5.43 and thus had zero mean curvature.

5.2.1 Example *Scherk's surface*

The catenoid and the helicoid remained the only known minimal surfaces until 1830, when Scherk obtained a new solution under the assumption that $f(x, y)$ has the special form.

$$f(x,y) = U(x) + V(y)$$

In this case, the minimal surface equation 5.42 can be separated

$$\begin{aligned}U''(1-V'^2) + V''(1+U'^2) &= 0,\\ \frac{U''}{1+U'^2} = -\frac{V''}{1+V'^2} &= C.\end{aligned}$$

The ordinary differential equations only involve the first and second derivatives of the variables, so they can be easily integrated. First, we let $R = U'$, and solve for U, setting the integration constants to zero.

$$\begin{aligned}\frac{R'}{1+R^2} &= C,\\ \int\frac{R}{1+R^2}dR &= \int C\,dx,\\ \tan^{-1}R &= Cx,\\ U' &= \tan(Cx)\\ U &= \ln[\sec Cx].\end{aligned}$$

The integral for V is done the same way, and the result is $V = -\tan[\ln\sec y]$, so

$$f(x,y) = \ln[\sec Cx] - \ln[\sec Cy].$$

Setting $C = 1$ and rewriting in terms of cosines, we get

$$f(x,y) = \ln\frac{\cos y}{\cos x}. \tag{5.44}$$

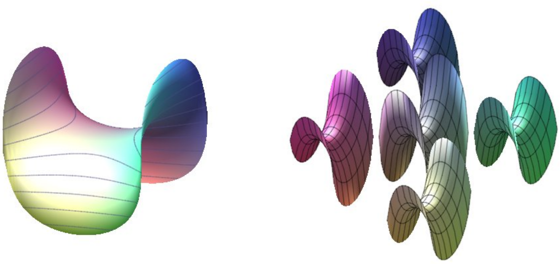


Fig. 5.7: Scherk's Surface.

This is the classical, doubly periodic *Scherk surface* which we render in figure 5.7.

Next, we prove the area-minimizing property for more general surfaces with zero mean curvature. The integrand in the formula for surface area is the square root of the determinant of the first fundamental form, so to perform a variation, we will need to take derivatives of determinants. The main idea in obtaining a formula for the derivative of a determinant rests on a neat result from linear algebra which at the risk of digressing a bit, it is worth proving now. This theorem due to Jacobi, will be most important later when we discuss the exponential map in the context of Lie groups and Lie algebras (See section 7.2.1).

5.2.2 Theorem Let A be a square matrix. Then

$$\det e^A = e^{\operatorname{Tr} A}. \tag{5.45}$$

Proof First consider the case where A is a diagonal $n \times n$ matrix $A = \operatorname{diag}[\kappa_1, \kappa_1, \ldots, \kappa_n]$ by an . Defining e^A by the exponential power series (See equation 7.52), it is immediately verified that:

$$\begin{aligned} e^A &= \operatorname{diag}[e^{\kappa_1}, e^{\kappa_2}, \ldots, e^{\kappa_n}], \\ \det e^A &= e^{\kappa_1} e^{\kappa_2} \ldots e^{\kappa_n}, \\ &= e^{\operatorname{Tr} A}. \end{aligned}$$

Next, consider the case where A is diagonalizable. Then, there exists a similarity transformation Q such that $A = Q^{-1}DQ$, where D is the diagonal matrix with the eigenvalues along the diagonal, $D = \operatorname{diag}[\kappa_1, \kappa_1, \ldots, \kappa_n]$. We recall that determinant and trace are invariant under similarity transformations. We have

$$\begin{aligned} A^2 &= (Q^{-1}DQ)(Q^{-1}DQ), \\ &= Q^{-1}D(QQ^{-1})DQ, \\ &= Q^{-1}DIDQ, \\ &= Q^{-1}D^2Q. \end{aligned}$$

By induction, we can easily prove that $A^n = Q^{-1}D^nQ$ and hence,

$$\begin{aligned}
e^A &= \sum_{n=0}^{\infty} \frac{1}{n!} A^n, \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} Q^{-1}D^nQ, \\
&= Q^{-1}(\sum_{n=0}^{\infty} \frac{1}{n!} D^n)Q, \\
&= Q^{-1}e^DQ, \\
\det A &= \det Q^{-1}(\det e^D) \det Q, \\
&= \det e^D = e^{\mathrm{Tr}\, D} = e^{\mathrm{Tr}\, A}.
\end{aligned}$$

Finally, A might not diagonalizable, but it can be reduced to canonical Jordan form J by a similarity transformation $A = Q^{-1}JQ$. The Jordan matrix J has the eigenvalues along the diagonal. with a block structure of the form $J_\lambda = \lambda I + N$, where N is nilpotent, it remains true that $\det e^J = e^{\mathrm{Tr}\, J}$. Thus the argument above holds in this case as well.

Back to our topic. If in equation 5.45 we replace A by $\ln A$, we get,

$$\det A = e^{\mathrm{Tr}(\ln A)}. \tag{5.46}$$

Suppose that instead of single matrix A, we have a C^∞, one-parameter family of matrices A_t. Then, we can use the equation above to differentiate with respect to t.

$$\begin{aligned}
\frac{d}{dt} \det A_t &= e^{\mathrm{Tr}(\ln A_t)} \frac{d}{dt}(\mathrm{Tr}(\ln A_t)), \quad \text{hence,} \\
&= (\det A_t)\, \mathrm{Tr}\, \frac{d}{dt} \ln A_t, \\
&= (\det A_t)\mathrm{Tr}({A_t}^{-1} \frac{dA_t}{dt})
\end{aligned} \tag{5.47}$$

We unpack this formula for a special kind of variation defined as follows:

5.2.3 Definition Let $\mathbf{x}(u, v) : U \to \mathbf{R}^3$ be a coordinate patch for a surface $\{M, g\}$ defined over a set $U \subset \mathbf{R}^2$, and let $\phi : U \to \mathbf{R}$ be a C^∞ function. A *normal deformation* of the surface is a one-parameter family of surfaces M_t with coordinate patches of the form $\mathbf{x}_t(u, v) = \mathbf{x}(u, v) + t\phi(u, v)\mathbf{n}$, where t is a small parameter, $t \in [-\epsilon, \epsilon]$, and $\mathbf{n}$ is the unit normal.

Let g_t be the matrix of the first fundamental form induced on the surface M_t and let $\det(g_t)$ denote the determinant. The elements of surface area are $dS_t = \sqrt{\det(g_t)}\, du \wedge dv$, and the areas are given by:

$$A_t = \int\int_U dS_t = \int\int_u \sqrt{\det(g_t)}\, du \wedge dv. \tag{5.48}$$

At $t = 0$, $g_0 = g$ and $dS_o = dS$ represent the metric and the differential of surface area element of M. We have the following theorem:

5.2.4 Theorem The variation of surface area satisfies

$$A_t'(0) = -2\int\int_U \phi H dS, \tag{5.49}$$

so that $A_t'(0) = 0$ if, and only if, $H = 0$.

Proof The proof is by computation, using the formula 5.47 for derivatives of determinants.

$$\begin{aligned}
\frac{d}{dt}A_t &= \frac{d}{dt}\int\int_U dS_t = \int\int_U \frac{d}{dt}dS_t, \\
&= \int\int_U \frac{1}{2\sqrt{\det(g_t)}}\frac{d}{dt}\det(g_t)\, du\wedge dv, \\
&= \int\int_U \frac{1}{2\sqrt{\det(g_t)}}\det(g_t)\mathrm{Tr}({g_t}^{-1}\frac{dg_t}{dt})\, du\wedge dv, \\
&= \frac{1}{2}\int\int_U \mathrm{Tr}({g_t}^{-1}\frac{dg_t}{dt})\, dS_t.
\end{aligned}$$

It remains to compute the derivative of the one-parameter family of first fundamental forms along the normal deformation.

$$\begin{aligned}
d\mathbf{x}_t &= d\mathbf{x} + t(\phi\, d\mathbf{n} + d\phi\, \mathbf{n}), \\
< d\mathbf{x}_t, d\mathbf{x}_t > &= < d\mathbf{x}, d\mathbf{x} > + 2t\phi\ < d\mathbf{x}, d\mathbf{n} > + \mathscr{O}(t^2), \\
I_t &= I - 2\phi\, t\, II + \mathscr{O}(t^2), \\
(g_t)_{\alpha\beta} &= g_{\alpha\beta} - 2\phi\, t\, b_{\alpha\beta} + \mathscr{O}(t^2).
\end{aligned}$$

Therefore,

$$\frac{d}{dt}g_t|_{t=0} = -2\phi\, b$$

We deduce that

$$A_t'(0) = -\int\int_U \phi\mathrm{Tr}(g^{-1}b)\, dS = -2\int\int_U \phi H dS,$$

for any function ϕ, so this concludes the proof. A more complete proof would include analysis of the second variation, but this will not be treated in these notes.

5.2.5 Example Surface of revolution

Let M be a surface of revolution that is also a minimal surface. The standard coordinate patch is given by 4.7

$$\mathbf{x}(r,\phi) = (r\cos\phi, r\sin\phi, f(r))$$

with fundamental form coefficients

$$
\begin{aligned}
E &= 1 + f'^2, & & & e &= f''/\sqrt{1 + f'^2}, \\
F &= 0, & &\text{and} & f &= 0, \\
G &= r^{,} & & & g &= rf'/\sqrt{1 + f'^2}.
\end{aligned}
$$

For this to be a minimal surface we must have:

$$
\begin{aligned}
H = Eg + Ge &= 0, \\
rf'(1 + f'^2) + r^2 f'' &= 0.
\end{aligned}
$$

The equation is easily integrated. Let $p = f'$, separate variables and integrate by partial fractions:

$$
\begin{aligned}
rp' &= p(1 + p^2), \\
\frac{1}{p(1+p^2)} dp &= -\frac{1}{r} dr, \\
\frac{p}{\sqrt{1+p^2}} &= \frac{A}{r},
\end{aligned}
$$

where A is a constant of integration. Squaring both sides and solving for $p = f'$ we get:

$$
\begin{aligned}
p^2 r^2 &= A^2(1 + p^2), \\
p = f' &= \frac{A}{\sqrt{r^2 - 1}}, \\
f &= A \cosh^{-1} r.
\end{aligned}
$$

The conclusion is that a catenoid is the only minimal surface of revolution. The mean curvature is not a bending invariant because it depends specifically on the second fundamental form. However, it is notable by a short computation, that the deformation described in equation 4.116 that bends a catenoid into a helicoid, results on a one-parameter family of minimal surfaces $\mathbf{z}_t$, independent of t. A surface of type $\mathbf{x} = (r\cos\phi, r\sin\phi, h(\phi))$ is a called a *conoid.* The helicoid is the only minimal surface that is also a conoid.

5.2.2 Conformal Mappings

In this section we explore the connection between minimal surfaces and conformal maps. For this purpose, it will be useful to insert a pedestrian review of some basic concepts in complex variables. We denote by $\mathbf{C}$ the usual vector space of complex numbers of the form $z = x + iy$. Complex numbers can also be represented by antisymmetric matrices

$$
\begin{bmatrix} x & y \\ -y & x \end{bmatrix},
$$

in which the binary algebraic operations are mapped to matrix operations. Thus, for example, multiplying two complex numbers $z_1 z_2$, is equivalent to multiplying the corresponding matrices. By Euler's formula, any complex number of unit length can be written in the form $z = e^{i\theta} = \cos\theta + i\sin\theta$. The matrix version of a unit vector is a rotation matrix

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

The set all such matrices forms a group called $SO(2)$. There are two special elements in this set that comprise a basis for the vector space, the identity matrix I and the symplectic matrix

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{5.50}$$

The former corresponds to a rotation matrix with $\theta = 0$ and the latter to a rotation by $\theta = \pi/2$. Clearly $J^2 = -I$, showing that J plays the role of the imaginary number i in the matrix representation.

A fundamental result from complex variables is that if a function $w = f(z) = u + iv$ is differentiable in the complex sense (i.e. holomorphic), then the following properties hold:

1. The real and imaginary parts satisfy the Cauchy-Riemann equations $u_x = v_y$; $u_y = -v_x$.

2. The functions u and v are (conjugate) harmonic: $\nabla^2 u = \nabla^2 v = 0$.

3. the families of curves $u(x, y)$ =constant and $v(x, y)$ =constant, are mutually orthogonal. That is, $(\text{Grad } u, \text{Grad } v) = 0$. In the context of heat flow these curves are called isothermal lines.

4. The map is conformal, that is, it preserves angles. The conformal factor is given by the Jacobian $|f'(z)|^2 = |\nabla u|^2 = |\nabla v|^2$. In vector component notation, if $z = (x, y)^T$ and $h = (k_1, k_2)^T$, then by differentiability,

$$f(z + h) = f(z) + Df(h) + \epsilon,$$

   where Df is the Jacobian map f' and $\epsilon \to 0$ as $h \to 0$. From the Cauchy Riemann equations at any given point, we have

$$\begin{aligned} Df(h) &= \begin{bmatrix} x_u & x_v \\ y_u & y_v \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \\ &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \\ &= \sqrt{a^2 + b^2} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \end{aligned}$$

   for some numbers a, b and some angle θ. Thus, the transformation consists locally of a dilation and a rotation.

5. Let $z = x + iy$ and $\overline{z} = x - iy$. The chain rule gives:
$$\partial_z \equiv \frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \partial_{\overline{z}} \equiv \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

6. $w = f(z)$ is holomorphic if and only if $\partial_{\overline{z}} f = \frac{\partial f}{\partial \overline{z}} = 0$. This is equivalent to the Cauchy-Riemann equations.

7. $I = dx^2 + dy^2 = dz d\overline{z}$, and $\Delta f \equiv f_{xx} + f_{yy} = 4\frac{\partial^2 f}{\partial z \partial \overline{z}.}$

Since $\mathbf{R}^2$ as a vector space is isomorphic to $\mathbf{C}$, we can extend the complex structure to a surface M in $\mathbf{R}^3$ by requiring that the locally Euclidean property be replaced by complex coordinate patches that are holomorphic. This can actually be done intrinsically for an oriented surface by introducing a $(1,1)$ tensor $J : TM \to TM$ so that for an orthogonal basis $\{e_1, e_2\}$ of the tangent space, $J(e_1) = e_2$ and $J(e_2) = -e_1$. This results on a matrix representation of J at each point identical to the symplectic matrix 5.50 and it represents a rotation by $\pi/2$. The tensor J can always be introduced in any coordinate patch by starting with $\mathbf{x}_u$ or $\mathbf{x}_v$ and using the Gram-Schmidt process to find an orthogonal vector. An easy computation gives:
$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_u, \quad \mathbf{v}_2 = \mathbf{x}_v - \frac{F}{E}\mathbf{x}_u = \frac{E\mathbf{x}_v - F\mathbf{x}_u}{E}, \\ \mathbf{w}_1 &= \mathbf{x}_v, \quad \mathbf{w}_2 = \mathbf{x}_u - \frac{F}{G}\mathbf{x}_v = \frac{G\mathbf{x}_u - F\mathbf{x}_v}{G}. \end{aligned} \tag{5.51}$$
Orientation is preserved if the differential of surface satisfies $dS(X, J(X)) > 0$ for all $X \neq 0$. Since J represents a rotation by $\pi/2$, there must be constants c_1 and c_2 such that $J\mathbf{x}_u = c_1(E\mathbf{x}_v - F\mathbf{x}_u)$ and $J\mathbf{x}_v = c_2(F\mathbf{x}_v - G\mathbf{x}_u)$. Setting $\|\mathbf{x}_u\|^2 = \|J\mathbf{x}_u\|^2$ and $\|\mathbf{x}_v\|^2 = \|J\mathbf{x}_v\|^2$ we find that $c_1 = c_2 = 1/\sqrt{\det g}$, hence, the components of J in the parametric coordinate frame are
$$\begin{aligned} J\mathbf{x}_u &= \frac{E\mathbf{x}_v - F\mathbf{x}_u}{\sqrt{EG - F^2}}, \\ J\mathbf{x}_v &= \frac{F\mathbf{x}_v - G\mathbf{x}_u}{\sqrt{EG - F^2}}. \end{aligned} \tag{5.52}$$

5.2.6 Definition Let $\{M, g\}$ and $\{M', g'\}$ be Riemannian manifolds. A map $F : M \to M'$ is called *conformal* if there exists a function λ, such that
$$< F_*X, F_*Y >' = \lambda^2 < X, Y >, \tag{5.53}$$
for all tangent vectors. If $\lambda = 1$, the map is an isometry.
Conformal maps on manifolds as defined above have the following properties:

1. $\|F_*X\|'^2 = |\lambda|^2 \|X\|^2$. The reason is that $\|X\|^2 = < X, X >$.

2. F preserves angles, since at each point, the λ's cancel in the formula for the cosine of the angle as in 1.51 and 4.17.

3. $F_*(J(X)) = J(F_*(X))$. This is easily shown by applying the maps to a basis and using the fact that F preserves angles and J represents a rotation.

5.2.3 Isothermal Coordinates

An isothermal system as in 4.117 in which the metric takes the form $ds^2 = \lambda^2\,(du^2 + dv^2)$ can be considered as a map from $U \subset \mathbf{R}^2$ as a plane surface, to M. This is an example of the most basic conformal map. As previously stated in theorem 4.5.15, the coordinate patch in this case satisfies the equation,

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2\lambda^2 \mathbf{H}.$$

The conclusion of this theorem can be restated by saying that that surface in isothermal coordinates is a minimal surface if and only if its coordinate patch functions are harmonic in the usual Euclidean sense.

5.2.7 Definition Given an isothermal, conformal patch $\mathbf{x}(u, v)$, we call a map $\mathbf{y}(u, v)$ a *conjugate patch* if it satisfies the Cauchy-Riemann equations

$$\mathbf{x}_u = \mathbf{y}_v, \quad \mathbf{x}_v = -\mathbf{y}_u. \tag{5.54}$$

In complex variables, the real and imaginary parts of a holomorphic function are conjugate harmonic functions. Given one of these, one can determine the other by integrating the Cauchy-Riemann equations. In a similar way, given an isothermal conformal patch, we may determine a conjugate patch and this conjugate patch is also isothermal and conformal. The conjugate patches can be rendered as the real and imaginary parts of a complex holomorphic patch.

5.2.8 Definition Given conjugate harmonic patches $\mathbf{x}$ and $\mathbf{y}$, we define the *associated family* to be the one-parameter family

$$\mathbf{z}_t = \Re[e^{it}(\mathbf{x} + i\mathbf{y})] = (\cos t)\mathbf{x} + (\sin t)\mathbf{y}. \tag{5.55}$$

5.2.9 *Existence of isothermal coordinates*
The existence of isothermal coordinates is more subtle and requires some harmonic analysis. Given $\{M, g\}$ with a coordinate patch $\mathbf{x}(u, v)$, we seek a map $F(p) = (h(p), k(p))$, in which the metric is isothermal. The condition that the parametric curves be orthogonal means we must have $(\text{Grad } h, \text{Grad } k) = 0$. Using the definition of the gradient 2.26, and recalling the components of the inverse of the metric 4.85, we get

$$\begin{aligned}
(\text{Grad } h, \text{Grad } k) &= g_{\alpha\beta}(\nabla h)^\alpha (\nabla k)^\beta, \\
&= (\nabla h)^\alpha (\nabla k)_\alpha, \\
&= \frac{1}{\det(g)}[k_u(Gh_u - Fh_v) + k_v(Eh_v - Fh_u)], \\
&= \frac{1}{\det(g)}[k_u(Gh_u - Fh_v) - k_v(Fh_u - Eh_v)]
\end{aligned}$$

The equation $(\text{Grad } h, \text{Grad } k) = 0$ holds if there exists a function λ so that

$$k_u = \lambda(Fh_u - Eh_v), \quad k_v = \lambda(Gh_u - Fh_v).$$

To get $|\nabla k|^2 = \nabla h|^2$ we set $\lambda = 1/\sqrt{\det(g)}$. The integrability condition $k_{uv} = k_{vu}$ yields the classical Laplace-Beltrami equation 4.86

$$\frac{\partial}{\partial u}\left[\frac{Gh_u - Fh_v}{\sqrt{EG-F^2}}\right] + \frac{\partial}{\partial v}\left[\frac{Eh_v - Fh_u}{\sqrt{EG-F^2}}\right] = 0.$$

Hence, the existence of non-trivial solutions is tied to the harmonic analysis on the existence of solutions of the elliptic operator.

A more classical approach for the analytic case was obtained by Gauss by the simple but ingenious method of factoring the first fundamental form $ds^2 = E\,du^2 + 2F\,dudv + G\,dv^2$,

$$ds^2 = \left[\sqrt{E}du + \frac{F + i\sqrt{EG-F^2}}{\sqrt{E}}dv\right]\left[\sqrt{E}du + \frac{F - i\sqrt{EG-F^2}}{\sqrt{E}}dv\right].$$

The idea is to find coordinates h and k and a conformal factor λ, such that

$$ds^2 = \lambda^2\,dh\,dk.$$

This would suffice, since having found h and k, one could set $h = \phi + i\psi$ and $k = \phi - i\psi$. Following Eisenhart [19], let t_1 and t_2 be integrating factors of the system

$$\begin{aligned} t_1\left(\sqrt{E}du + \frac{F + i\sqrt{EG-F^2}}{\sqrt{E}}dv\right) &= dh = h_u\,du + h_v\,dv, \\ t_2\left(\sqrt{E}du + \frac{F - i\sqrt{EG-F^2}}{\sqrt{E}}dv\right) &= dk = k_u\,du + k_v\,dv, \end{aligned}$$

where $\lambda^2 = 1/(t_1t_2)$. The first equation above is equivalent to,

$$t_1\sqrt{E} = h_u, \quad t_1\left(\frac{F+iH}{\sqrt{E}}\right) = h_v,$$

where $H = \sqrt{\det(g)} = \sqrt{EG-F^2}$. Solving for t_1 in the first equation and substituting on the second, we get

$$h_u(F + iH) = Eh_v, \tag{5.56}$$

$$ih_u = \frac{Eh_v - Fh_u}{H}. \tag{5.57}$$

On the other hand, multiplying the first equation above by $F - iH$ we get

$$\begin{aligned} h_u(F^2 + H^2) &= E[F - iH]h_v, \\ h_u(F^2 + EG - F^2) &= E(F - iH)h_v, \\ EGh_u &= E(F - iH)h_v. \end{aligned}$$

Hence

$$ih_v = \frac{Fh_v - Gh_u}{H}. \tag{5.58}$$

The integrability condition $h_{uv} = h_{vu}$ for equations 5.57 and 5.58 gives again the Laplace Beltrami equation

$$\frac{\partial}{\partial u}\left[\frac{Gh_u - Fh_v}{\sqrt{EG - F^2}}\right] + \frac{\partial}{\partial v}\left[\frac{Eh_v - Fh_u}{\sqrt{EG - F^2}}\right] = 0.$$

By a completely analogous computation, one can verify that k satisfies the same integrability condition.

The existence of isothermal coordinates is closely related to the existence of complex structures. Even for the case of 2-dimensional manifolds, if the topological conditions on the metric are weakened, the problem is not simple (See: [6]).

5.2.10 Exercise Let $ds^2 = E\,du^2 + 2F\,dudv + F\,dv^2 = \nu|dz + \mu d\overline{z}|^2$, where $\nu > 0$, $|\mu| < 1$, and $z = u + iv$.
1) Substitute du and dv in terms of dz and $d\overline{z}$. Equate coefficients to show that,

$$\mu = \frac{1}{4\nu}(E - G + 2Fi), \quad \nu(1 + \mu\overline{\mu}) = \frac{1}{2}(E + G).$$

2) Solve the quadratic equation for ν and thus show that,

$$\nu = \frac{1}{4}(E + G + 2\sqrt{EG - F^2}).$$

3) We want $ds^2 = \lambda^2 dh d\overline{h} = \lambda^2|dh|^2$, where $h = h(z, \overline{z})$. Show that h exists if, and only if, it satisfies the Beltrami equation

$$\frac{\partial h}{\partial \overline{z}} = \mu\frac{\partial h}{\partial z}.$$

5.2.11 Example Consider the unit sphere with $ds^2 = d\theta^2 + \sin^2\theta\ d\phi^2$. We rewrite the metric as

$$ds^2 = \sin^2\theta\left(\frac{1}{\sin^2\theta}d\theta^2 + d\phi^2\right),$$

and set

$$du = \frac{1}{\sin\theta}d\theta, \quad dv = d\phi.$$

Integrating, we see that the transformation

$$u = \ln|\tan(\tfrac{\theta}{2})|, \quad v = \phi.$$

gives the metric manifestly in isothermal coordinates

$$ds^2 = \sin^2\theta(du^2 + dv^2). \tag{5.59}$$

The example is technically not complete since the conformal factor is not in terms of the new variables. This is not hard to do, but the geometry will become more clear in the context of the sterographic projection.

5.2.4 Stereographic Projection

Consider the unit sphere $S^2 : x^2 + y^2 + z^2 = 1$. for each point $P(x, y, z)$ on the sphere, we draw a line segment from the north pole to P and extend the segment until it intersects the xy-plane, viewed as a copy of the complex plane, at point $\zeta = X + iY$. By simple ratio and proportions of corresponding sides of similar triangles, as partially shown in 5.8, we have,

$$\begin{aligned} X &= \frac{x}{1-z}, \\ Y &= \frac{y}{1-z}, \quad \text{so that} \\ \zeta &= \frac{x+iy}{1-z} \end{aligned} \tag{5.60}$$

The map $\pi : S^2 \to \mathbf{C}$ projects each point P except for the north pole to the unique complex number ζ given by the equation above. The closer the point P is to the north pole, the larger the norm of ζ. This gives rise to the geometric interpretation that under the stereographic projection, the north pole correspond to the point at infinity in the complex plane. To find the inverse π^{-1} of the projection, first notice that:

$$\begin{aligned} \zeta\bar{\zeta} &= \frac{x^2+y^2}{(1-z)^2}, \\ \zeta\bar{\zeta}+1 &= \frac{x^2+y^2}{(1-z)^2} + 1, \\ &= \frac{x^2+y^2+(1-z)^2}{(1-z)^2}, \\ &= \frac{2-2z}{(1-z)^2}, \\ &= \frac{2}{1-z}. \end{aligned}$$

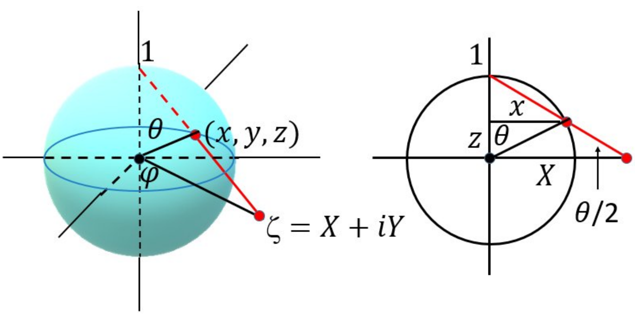


Fig. 5.8: Stereographic Projection

Combining that last equation with 5.60, we find the x, y and z coordinates on the sphere,

$$\begin{aligned}\pi^{-1}(\zeta) &= (x, y, z), \\ &= \left(\frac{\zeta+\bar{\zeta}}{\zeta\bar{\zeta}+1}, \frac{\zeta-\bar{\zeta}}{i(\zeta\bar{\zeta}+1)}, \frac{\zeta\bar{\zeta}-1}{\zeta\bar{\zeta}+1}\right)\end{aligned} \tag{5.61}$$

The existence of the inverse map shows that the sphere is locally diffeomorphic to the complex plane. If η represents the complex coordinate arising from the stereographic projection from the south pole, then the transition functions on the overlap of the coordinate patches is given by $\zeta = 1/\eta$ and $\eta = 1/\zeta$. Thus, S^2 may be considered as a complex manifold called the *Riemann sphere*. In polar coordinates $X = R\cos\phi$, $Y = R\sin\phi$, so, by ratio and proportion applied to similar triangles as in figure 5.8, we see that $\|\zeta\| = R = \cot(\frac{\theta}{2})$, and hence

$$\begin{aligned}\zeta &= \cot(\tfrac{\theta}{2})e^{i\phi}, \\ \eta &= \tan(\tfrac{\theta}{2})e^{-i\phi}.\end{aligned} \tag{5.62}$$

The short computation that follows gives the complex metric of the Riemann sphere in terms of the standard metric in spherical coordinates.

$$\begin{aligned}d\zeta &= -\tfrac{1}{2}\csc^2(\tfrac{\theta}{2})e^{i\phi}\,d\theta + i\cot\tfrac{\theta}{2}e^{i\phi}\,d\phi, \\ d\bar{\zeta} &= -\frac{1}{2}\csc^2(\tfrac{\theta}{2})e^{-i\phi}\,d\theta - i\cot\tfrac{\theta}{2}e^{-i\phi}\,d\phi, \\ d\zeta d\bar{\zeta} &= \tfrac{1}{4}\csc^4(\tfrac{\theta}{2})\,d\theta^2 + \cot^2(\tfrac{\theta}{2})\,d\phi^2, \\ &= \frac{1}{4\sin^4(\frac{\theta}{2})}(d\theta^2 + \sin^2\theta d\phi^2).\end{aligned}$$

Using equation 5.62 again, we find:

$$1 + \zeta\bar{\zeta} = 1 + \cot^2\tfrac{\theta}{2} = \csc^2(\tfrac{\theta}{2}),$$

therefore,

$$ds^2 = \frac{4\,d\zeta d\bar{\zeta}}{(1+\zeta\bar{\zeta})^2} = d\theta^2 + \sin^2\theta\,d\phi^2. \tag{5.63}$$

This shows that the map is conformal. Looking back at equation 5.59 for the metric of the sphere in isothermal coordinates given by the transformation

$$u = \ln|\tan(\tfrac{\theta}{2})|, \quad v = \phi,$$

set $\alpha = u - iv$, and

$$\eta = e^{\alpha} = \tan(\tfrac{\theta}{2})e^{-i\phi}.$$

Thus, we see that the transformation is effectively the stereographic projection from the south pole. The complex version of the first fundamental form is called the *Fubini-Study* metric of the Riemann sphere. If we let K be the function

$$K = \ln(1 + \zeta\bar{\zeta}), \tag{5.64}$$

then we can write the components g of the complex metric as

$$g = 4\partial_\zeta \partial_{\bar{\zeta}} K.$$

This formula generalizes to $\mathbf{CP}^n$, for which

$$\begin{aligned} K &= \ln(1 + \delta_{ij} z^i \bar{z}^j), \\ ds^2 &= g_{i\bar{j}} dz^i d\bar{z}^j, \\ g_{i\bar{j}} &= \partial_{z^i} \partial_{\bar{z}^j} K. \end{aligned}$$

The function K is called the *Kähler potential.* Let the ∂ and $\bar{\partial}$ be the *Dolbeault* operators that give the natural extension of exterior derivative in several complex variables; that is, given a differential form if $\alpha = f_{IJ} dz^I \wedge d\bar{z}^J$ for multi-indices I and J, we define

$$\begin{aligned} \partial\alpha &= \frac{\partial f_{IJ}}{\partial z^k} dz^k \wedge dz^I \wedge d\bar{z}^J, \\ \bar{\partial}\alpha &= \frac{\partial f_{IJ}}{\partial \bar{z}^k} d\bar{z}^k \wedge dz^I \wedge d\bar{z}^J. \end{aligned}$$

Then we obtain a natural, a non-degenerate 2-form

$$\begin{aligned} \omega_K &= i\partial\bar{\partial} K, \\ &= i \frac{d\zeta \wedge d\bar{\zeta}}{(1 + \zeta\bar{\zeta})^2}. \end{aligned} \tag{5.65}$$

is called the *Kähler Form* of the sphere S^2. Some authors include a conventional factor of 1/2, but the factor of i is required for compatibility of the Hermitian and the Riemannian structure. Thus, in terms of differential geometry structures in dimension 2, this is as good as it gets. The two-sphere has a natural Riemannian structure inherited by the induced metric from $\mathbf{R}^3$, a complex structure induced by the stereographic projection, and a natural symplectic 2-form given by the Kähler form above. This is summarized by saying that we have the structure of a *Kähler manifold.* As we will see later, the Kähler form 5.65 is at the center of the construction of Dirac monopoles and instanton bundles.

The formulas for the stereographic projection 5.60 naturally extrapolate to spheres in all dimensions. If the unit sphere $S^n \in \mathbf{R}^{n+1}$ is given by equation

$$(x^1)^2 + (x^2)^2 + \cdots + (x^{n+1})^2 = 1$$

then the coordinates X^i of the projection onto $\mathbf{R}^n$ from the north pole $(0, 0, \ldots, 1)$ are given by

$$X^i = \frac{x^i}{1 - x^{n+1}}, \quad i = 1, \ldots, n$$

In the case of the unit circle S^1, the quantity ζ is a real number. Since we are using θ as the azimuthal angle, we have the unconventional parametrization of the unit circle $x = \sin\theta,\ y = \cos\theta$. Then the formulas for π^{-1} and the metric read:

$$x = \frac{2\zeta}{\zeta^2+1}, \quad y = \frac{\zeta^2-1}{\zeta^2+1}, \quad d\theta = \frac{2\zeta}{\zeta^2+1}d\zeta. \tag{5.66}$$

Hence, the substitution $\zeta = \cot(\frac{\theta}{2})$ transforms integrals of rational functions of sines and cosines into rational functions of ζ. This is the source of the now infamous *Weierstrass* substitution, which is more commonly written in terms of the polar angle, $t = \tan(\frac{\theta}{2})$. It is by this method that one obtains the integral 5.67, and the neat formula

$$\int \csc\theta\, d\theta = \ln|\tan(\tfrac{\theta}{2})|,$$

that often appear in the theory of curves and surfaces. The elegant Weierstrass substitution is, for the wrong reasons, no longer taught as a technique of integration in the standard second semester calculus course. This author does not disagree with those calculus reformers that pushed the topic out of the curriculum at the time, but strongly disagrees with the reasons given to the effect, that they never again encountered this in their work. There should have been a geometer in the committee. The stereographic projection is the starting point for the entire theory of spinors.

As an added bonus, if in equation 5.66 we pick $\zeta = \frac{m}{n}$ to be any rational number on the real line, the inverse stereographic map yields a rational point in the unit circle with entries giving Euclid's formula for Pythagorean triplets

$$\left(\frac{2mn}{n^2+m^2}, \frac{n^2-m^2}{n^2+m^2}.\right)$$

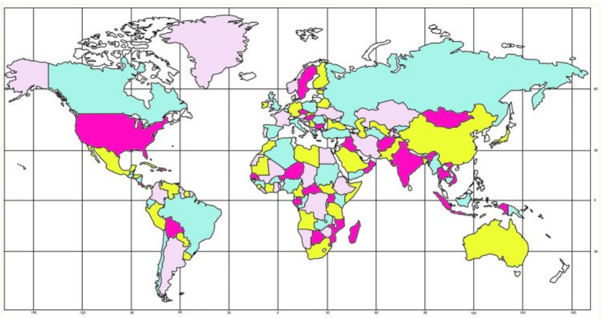


Fig. 5.9: Mercator Projection

5.2.12 Example *Mercator Projection*
Given a surface of revolution 4.7 with metric

$$\begin{aligned} ds^2 &= (1+f'^2)\,dr^2 + r^2\,d\phi^2, \\ &= r^2\left[\frac{1+f'^2}{r^2}\,dr^2 + d\phi^2\right], \end{aligned}$$

we set

$$d\hat{y} = \frac{\sqrt{1+f'^2}}{r}dr, \quad d\hat{x} = d\phi$$
$$\hat{y} = \int \frac{\sqrt{1+f'^2}}{r}dr, \quad \hat{x} = \phi.$$

This is a conformal map into a plane with Cartesian coordinates $(\hat{x}, \hat{y})$ in which the meridians map to $\hat{x}$ =constant, and the horizontal circles on the surface of revolution map to parallels $\hat{y}$ =constant. In particular, for a sphere in which $f(r) = \sqrt{a^2 - r^2}$, the substitution $r = a\sin\theta$ yields

$$\hat{y} = \int \sec\theta\, d\theta = \ln|\tan(\tfrac{\theta}{2} + \tfrac{\pi}{4})|. \tag{5.67}$$

Comparing with equation 4.23, we see that in this projection, loxodromes on the sphere map to straight lines on the plane. The projection is more faithful near the equator as shown in figure 5.9. The map is depicted with a four-color scheme, which it is always possible because of a famous theorem.

5.2.5 Minimal Surfaces by Conformal Maps

Consider a surface $\{M, g\}$ with coordinate patch $\mathbf{x}(u, v)$. Let $\zeta = u + iv$, and $\overline{\zeta} = u - iv$, so that $d\zeta = du + idv$, $d\overline{\zeta} = du - idv$. The composite map gives the patch as

$$\mathbf{x}(\zeta, \overline{\zeta}) = (\ x^1(\zeta, \overline{\zeta}), x^2(\zeta, \overline{\zeta}), x^3(\zeta, \overline{\zeta})\ ),$$

The complex derivative with respect to ζ of the patch is given by:

$$\begin{aligned}\mathbf{x}_\zeta = \frac{\partial \mathbf{x}}{\partial \zeta} &= \tfrac{1}{2}(\mathbf{x}_u - i\mathbf{x}_v),\\ &= (\tfrac{\partial x^1}{\partial \zeta}, \tfrac{\partial x^2}{\partial \zeta}, \tfrac{\partial x^3}{\partial \zeta}),\\ &= (x^1_\zeta, x^2_\zeta, x^3_\zeta).\end{aligned}$$

Let $\phi = 2\mathbf{x}_\zeta$, and $\phi^2 = 4(\mathbf{x}_\zeta, \mathbf{x}_\zeta)$. Since we have chosen ϕ to depend only on ζ, we have $\partial_{\overline{\zeta}}\phi = 0$, so ϕ is holomorphic. We have the following theorem:

5.2.13 Theorem A holomorphic patch $\mathbf{x}$ is isothermal if and only if

$$\phi^2 = (\phi^1)^2 + (\phi^2)^2 + (\phi^3)^2 = 0. \tag{5.68}$$

Proof The proof is by direct computation.

$$\begin{aligned}\phi^2 &= 4(\partial \mathbf{x}/\partial \zeta, \partial \mathbf{x}/\partial \zeta),\\ &= (\mathbf{x}_u - i\mathbf{x}_v, \mathbf{x}_u - i\mathbf{x}_v),\\ &= (\mathbf{x}_u, \mathbf{x}_u) - (\mathbf{x}_v, \mathbf{x}_v) - 2i(\mathbf{x}_u, \mathbf{x}_v),\\ &= E - G - 2iF.\end{aligned}$$

The theorem follows because in a isothermal patch, $E = G$ and $F = 0$, so $\phi^2 = 0$.

5.2.14 Remark The function ϕ is a complex function, so we must be careful not to confuse ϕ^2 as defined above with $|\phi|^2 = \phi\overline{\phi}$. The latter is given by

$$\begin{aligned} |\phi|^2 &= 4(\mathbf{x}_\zeta, \overline{\mathbf{x}}_\zeta), \\ &= (\mathbf{x}_u - i\mathbf{x}_v, \mathbf{x}_u + i\mathbf{x}_v), \\ &= (\mathbf{x}_u, \mathbf{x}_u) + (\mathbf{x}_v, \mathbf{x}_v), \\ &= E + G, \end{aligned} \tag{5.69}$$

so if the patch is isothermal, $E = G = \frac{1}{2}|\phi|^2$

We now have a process to construct minimal surfaces. We need to find a holomorphic patch ϕ with $\phi^2 = 0$. To construct such patches we introduce a special class of complex curves.

5.2.15 Definition A complex curve $\phi : U \subset \mathbf{C} \to \mathbf{R}^2$ is called a *minimal curve* or an *isotropic curve* if the differential of arc length

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 = 0.$$

To find such curves we once again we take advantage of the wonderful ingenuity of the leading mathematicians in the 1800's. Factoring over the complex, we have

$$\begin{aligned} (dx^1 + i\,dx^2)(dx^1 - i\,dx^2) + (dx^3)^2 &= 0, \\ (dx^1 + i\,dx^2)(dx^1 - i\,dx^2) &= -(dx^3)^2, \\ \frac{dx^1 + i\,dx^2}{dx3} &= -\frac{dx^3}{dx^1 - i\,dx^2}. \end{aligned}$$

Setting the left hand side equal to some function $-\tau$ we get a pair of differential equations

$$\begin{aligned} \frac{dx^1}{dx^3} + i\frac{dx^2}{dx^3} &= -\tau, \\ \frac{dx^1}{dx^3} + i\frac{dx^2}{dx^3} &= \frac{1}{\tau}. \end{aligned}$$

If we add the two equations we get:

$$\begin{aligned} \frac{dx^1}{dx^3} &= \frac{1}{\tau} - \tau, \\ &= \frac{1-\tau^2}{\tau}, \\ \frac{dx^1}{1-\tau^2} &= \frac{dx^3}{2\tau}. \end{aligned}$$

Subtracting the equations instead, we get:

$$
\begin{aligned}
2i\frac{dx^2}{dx^3} &= -\tau - \frac{1}{\tau}, \\
&= -\frac{1+\tau^2}{\tau}, \\
\frac{dx^1}{i(1-\tau^2)} &= \frac{dx^3}{2\tau}.
\end{aligned}
$$

Thus, we are lead to the following equation:

$$
\frac{dx^1}{1-\tau^2} = \frac{dx^2}{i(1+\tau^2)} = \frac{dx^3}{2\tau} = F(\tau)\, d\tau \tag{5.70}
$$

for some arbitrary analytic function F. At the end, we want real geometric objects, so we integrate and extract the real part.

$$
\begin{aligned}
x^1 &= \Re \int^z (1-\tau^2)F(\tau)\, d\tau, \\
x^2 &= \Re \int^z i(1+\tau^2)F(\tau)\, d\tau, \\
x^3 &= \Re \int^z 2\tau F(\tau)\, d\tau.
\end{aligned} \tag{5.71}
$$

This choice of coordinates satisfies the condition $\phi^2 = 0$ for an isothermal patch. Although a bit redundant, this can be verified immediately.

$$
\begin{aligned}
\phi &= F(1-\tau^2, i(1+\tau^2), 2\tau) \\
\phi^2 &= F^2[(1-\tau^2)^2 + i^2(1+\tau^2)^2 + 4\tau^2] = 0.
\end{aligned}
$$

Equations 5.71 are the classical *Weierstrass coordinates* from which one ascertains that a holomorphic function $F(\tau)$ gives rise to a minimal surface.

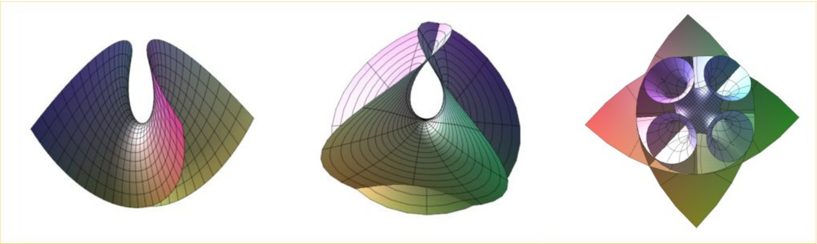


Fig. 5.10: Enneper Surface

5.2.16 Example *Enneper surface*
This is the surface corresponding to a Weierstrass patch with the simplest, nontrivial holomorphic function, namely a constant $F(\tau) = a$ which choose to be

$a = 3$. The surface was discovered by Enneper in 1864. Integration of 5.71 with $F(\tau) = 3$ gives,

$$\begin{aligned} x^1 &= \Re(3\tau - \tau^3) = 3u + 3uv^2 - u^3, \\ x^2 &= \Re\left[i(3\tau + \tau^3)\right] = -3v - 3u^2v + v^3, \\ x^3 &= \Re(3\tau^2) = 3(u^2 - v^2), \end{aligned}$$

resulting in the coordinate patch

$$\mathbf{x}(u, v) = (3u + 3uv^2 - u^3, -3v - 3u^2v + v^3, 3(u^2 - v^2)). \tag{5.72}$$

If one sets $\tau = re^{\phi}$ a polar coordinate parametrization is obtained

$$\mathbf{x}(r, \phi) = (3r\cos\phi - r^3\cos 3\phi, -3r\sin\phi - 3r^3\sin 3\phi, 3r^2\cos 2\phi). \tag{5.73}$$

Figure 5.10 shows some Enneper surfaces, the first with u and v ranging from 0 to 1.2 and the second for $r \in [0, 2]$, $\phi \in [\pi, \pi]$. The surface is self-intersecting if the range is big enough. In the polar parametrization $K = -(4/9)/(1 + r^2)^{-4}$ and of course $H = 0$. Using an advanced algebraic geometry technique called Gröbner basis, one can eliminate the parameters and show that the surface is algebraic, and that it can be written implicitly in terms of a ninth degree polynomial. As will be discussed later in equation 5.79, there is an alternate way to write the Weierstrass parametrization. In this alternate formulation, the classical Enneper surface is generated by $f = 1$ and $g = \sigma$. A generalization of the Enneper surface is obtained by taking $g(\sigma) = \sigma^n$, where $n+1$ is the number of "flaps" bending up. The surface on the right of figure 5.10 corresponds to the case $n = 3$.

Setting $r =$constant in the polar form of the Enneper surface, we get curves that are not spherical curves, but they are close. On the left of Figure 5.11 we display the intersection of an Enneper surface with a sphere.

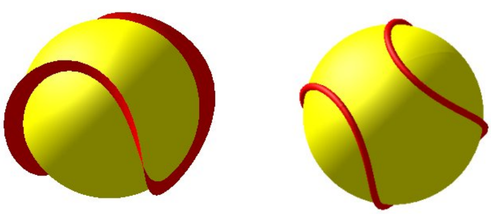


Fig. 5.11: Baseball Seam

This raises the fun problem of finding a curve that is actually spherical and resembles the seam of a tennis ball or a baseball. We explore this possibility with a curve of the form

$$\mathbf{x}(r, \phi) = (a\cos\phi - b\cos 3\phi, a\sin\phi + b\sin 3\phi, c\cos 2\phi),$$

and requiring it to have a constant norm. A short computation using the sum and half-angle formulas for cosine, gives

$$\begin{aligned}\|\mathbf{x}\|^2 &= a^2 + b^2 - 2ab\cos 4\phi + c^2\cos^2 2\phi,\\ &= a^2 + b^2 - 2ab\cos 4\phi + \tfrac{c^2}{2}(1+\cos 4\phi),\\ &= a^2 + b^2 + \tfrac{c^2}{2} + (\tfrac{c^2}{2} - 2ab)\cos 4\phi.\end{aligned}$$

To make this length constant, independently of ϕ, we set $c^2 = 4ab$, which remarkably, leads to the norm being a perfect square $\|\mathbf{x}\|^2 = (a+b)^2$. A nice choice leading to all integer coefficients is $a = 9,\ b = 4$ which gives $c = 12$. The graph shown on the right in figure 5.11 shows that this results on a reasonable shape for the seam of a baseball. Personally, I find this more gratifying than the baseball I hand-made as a kid because I could not afford one. I lost that ball in the first pitch when an older boy hit a home run into a pig sty.

5.2.17 Example Catenoid
As we would expect, the ubiquitous catenoid can be obtained from the Weierstrass parametrization. Anticipating a coordinate patch with cosh functions, we choose $F(\tau) = -a/\tau^2$ and then let $\tau = e^w = e^{u+iv}$. Integration, followed by an application of the summation formulae for hyperbolic functions, we get:

$$\begin{aligned}x^1 &= -a\,\Re\int[(1/\tau^2) - 1]\,d\tau,\\ &= -a\,\Re\,[(-1/\tau) - \tau],\\ &= a\,\Re\,[e^w + e^{-w}] = a\,\Re\,[\cosh w],\\ &= a\,\Re\,[\cosh(u+iv)],\\ &= a\cosh u\cos v.\end{aligned}$$

In a similar manner,

$$\begin{aligned}x^2 &= -a\,\Re\int i[(1/\tau^2) + 1]\,d\tau],\\ &= -a\,\Re\,\{i[(-1/\tau) + \tau)]\},\\ &= -a\,\Re\,\{i[e^w - e^{-w}]\} = -a\,\Re[i\sinh w],\\ &= -a\,\Re\,[i\sinh(u+iv)],\\ &= a\cosh u\sin v.\end{aligned}$$

The last integration is a one-liner

$$x^3 = -2a\,\Re[\ln\tau] = -2a\,\Re(w) = -2au.$$

The result is the catenoid

$$\mathbf{x}(u,v) = (a\cosh u\cos v, a\cosh u\sin v, -2au).$$

5.2.18 Example *Henneberg Surface*

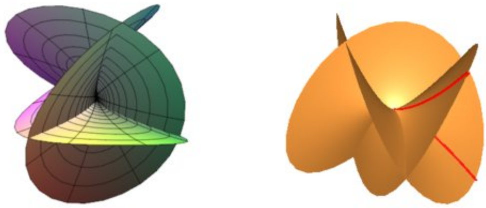


Fig. 5.12: Henneberg Surface

This surface was discovered in 1875 by the German mathematician Lebrecht Henneberg. The surface can be obtained from Weierstrass coordinates 5.71 by choosing $F(\tau) = 1-1/\tau^4$ and then letting $\tau = e^w = e^{u+iv}$. In a manner similar to the computation above, we integrate the equations and follow by an applying the summation formulæfor hyperbolic functions.

$$\begin{aligned}
x^1 &= \Re \int [(1-\tau^2)(1-1/\tau^4)]\, d\tau, \\
&= \Re \int [1+1/\tau^2-\tau^2-1/\tau^4)]\, d\tau, \\
&= \Re\, [(\tau-1/\tau)-(1/3)(\tau^3-1/\tau^3),] \\
&= 2\,\Re\, [\sinh w-(1/3)\sinh 3w], \\
&= 2\,\Re\, [\sinh(u+iv)-(1/3)\sinh 3(u+iv)], \\
&= 2\sinh u\cos v-(2/3)(\sinh 3u\cos 3v).
\end{aligned}$$

The integral for x^2 gives ,

$$\begin{aligned}
x^2 &= \Re \int i[(1+\tau^2)(1-1/\tau^4)]\, d\tau, \\
&= \Re \int i[1-1/\tau^2+\tau^2-1/\tau^4)]\, d\tau, \\
&= \Re\, \{i[(\tau+1/\tau)+(1/3)(\tau^3+1/\tau^3)]\}, \\
&= 2\,\Re\, \{i[\cosh w+(1/3)\cosh 3w]\}, \\
&= 2\,\Re\, \{i[\cosh(u+iv)+(1/3)\cosh 3(u+iv)]\}, \\
&= -2\sinh u\sin v-(2/3)(\sinh 3u\sin 3v).
\end{aligned}$$

The last integration is a bit simpler

$$
\begin{aligned}
x^3 &= \Re \int [2\tau(1 - 1/\tau^4)]\, d\tau, \\
&= \Re \int [2\tau - 2/\tau^3]\, d\tau, \\
&= \Re\, [(\tau^2 + 1/\tau^2)], \\
&= 2\, \Re\, [\cosh 2w], \\
&= 2\, \Re\, [\cosh 2(u + iv)], \\
&= 2 \cosh 2u \cos 2v.
\end{aligned}
$$

Neglecting the factor of 2 and using a common abbreviation for the hyperbolic functions, the result is:

$$
\mathbf{x}(u, v) = (\text{sh}\ u \cos v - \tfrac{1}{3}(\text{sh}\ 3u \cos 3v), -\text{sh}\ u \sin v - \tfrac{1}{3}(\text{sh}\ 3u \sin 3v), \text{ch}\ 2u \cos 2v).
$$

The curve given by $v = 0$ is a geodesic in the shape of a semicubical parabola which is the intersection of the surface with the plane $y = 0$. This is a reflection of the fact the a Hennerberg surface is a Björling surface, that is, a minimal surface that contains a given curve with prescribed normal. It turns out that the Hennerberg surface is a model representing an immersion in $\mathbf{R}^3$ of the projective plane.

The curvature for a minimal surface in Weierstrass coordinates can be calculated from equation 4.118. Using the fact that $E = G = \frac{1}{2}|\phi\,|^2$, we have,

$$
\begin{aligned}
\phi &= F(1 - \tau^2, i(1 + \tau^2), 2\tau), \\
|\phi|^2 &= |F|^2[(1 - \tau^2)(1 - \overline{\tau}^2) + (1 + \tau^2)(1 + \overline{\tau}^2) + 4\tau\overline{\tau}], \\
&= |F|^2(2 + 2\tau^2\overline{\tau}^2 + 4\tau\overline{\tau}), \\
&= 2|F|^2(1 + |\tau|^2)^2.
\end{aligned}
$$

The conformal factor is $E = \lambda^2 = \frac{1}{2}|\phi|^2$, so the Gaussian curvature is given by

$$
\begin{aligned}
K &= -\frac{1}{\lambda^2}\nabla^2(\ln \lambda), \\
&= -4\frac{1}{\lambda^2}\frac{\partial^2}{\partial\overline{\tau}\partial\tau}[\ln(|F|(1 + |\tau|^2))], \\
&= -4\frac{1}{\lambda^2}\frac{\partial^2}{\partial\overline{\tau}\partial\tau}[\tfrac{1}{2}(\ln F + \ln \overline{F}) + \ln(1 + |\tau|^2)], \\
&= -4\frac{1}{\lambda^2}\frac{\partial^2}{\partial\overline{\tau}\partial\tau}[\ln(1 + |\tau|^2)], \\
&= -4\frac{1}{\lambda^2}\frac{\partial}{\partial\overline{\tau}}[\frac{\overline{\tau}}{(1 + |\tau|^2)}], \\
&= -4\frac{1}{\lambda^2}\frac{(1 + \tau\overline{\tau}) - \tau\overline{\tau}}{(1 + |\tau|^2)^2} = -4\frac{1}{\lambda^2}\frac{(1}{(1 + |\tau|^2)^2}.
\end{aligned}
$$

So, the result is,

$$K == \frac{-4}{|F|^2(1+|\tau|^2)^4} = \frac{-4}{|F|^2(1+u^2+v^2)^4}. \tag{5.74}$$

There is an equivalent formulation of Weierstrass parametrization that also appears frequently in the literature. From equation 5.68, the holomorphic patch $\mathbf{x}$ is isothermal if

$$(\phi^1)^2 + (\phi^2)^2 + (\phi^3)^2 = 0. \tag{5.75}$$

Proceeding along the same lines as in the computation of isotropic coordinates, we have

$$\begin{aligned}(\phi^1 + i\phi^2)(\phi^1 - i\phi^2) &= -(\phi^3)^2,\\ (\phi^1 + i\phi^2) &= -\frac{(\phi^3)^2}{(\phi^1 - i\phi^2)}.\end{aligned} \tag{5.76}$$

Set

$$f = (\phi^1 - i\phi^2), \quad g = \frac{\phi^3}{(\phi^1 - i\phi^2)}. \tag{5.77}$$

Here, f is holomorphic, g is meromorphic, but $fg^2 = -(\phi^1 + i\phi^2)$ is holomorphic. Since we also have $fg = \phi^3$, we can easily solve for the components of ϕ in terms of f and g. The result is,

$$\begin{aligned}\phi^1 &= \frac{1}{2}f(1-g^2),\\ \phi^2 &= \frac{i}{2}f(1+g^2),\\ \phi^3 &= fg.\end{aligned} \tag{5.78}$$

Since $\phi = \mathbf{x}_u - i\mathbf{x}_v$, we are led to an alternative Weierstrass parametrization

$$\begin{aligned}x &= \Re\int^z \frac{1}{2}f(\sigma)(1-g(\sigma)^2)\,d\sigma,\\ y &= \Re\int^z \frac{i}{2}f(\sigma)(1+g(\sigma)^2)\,d\sigma,\\ z &= \Re\int^z f(\sigma)g(\sigma)\,d\sigma.\end{aligned} \tag{5.79}$$

For easy reference, we denote this parametrization by the notation $\mathbf{x} = \mathcal{W}(f,g)$. To see the relationship to the equations 5.71, consider the case when g is holomorphic and g^{-1} is holomorphic. Then, can use g as the complex variable, and we set $g = \tau$, so that $dg = d\tau$. Let $F = \frac{1}{2}f/\frac{dg}{d\sigma} = \frac{1}{2}f\frac{d\sigma}{dg}$. Then $F(\tau)d\tau = \frac{1}{2}f(\sigma)d\sigma$ and we have recovered equation 5.71. Choosing minus the imaginary parts in equation 5.79 yields *conjugate minimal surfaces*, given by the conjugate harmonic patch $\mathbf{y}$ as defined by equation 5.54.

A remarkable result can be obtained as follows. As a complex variable, g can be mapped into the unit sphere S^2 by the inverse stereographic projection

(See equation 5.61)

$$\pi^{-1}(g) = \frac{(g+\overline{g}, -i(g-\overline{g}), g\overline{g}-1)}{(g\overline{g}+1)}. \tag{5.80}$$

On the other hand, the real and imaginary parts of $\phi = \mathbf{x}_u - i\mathbf{x}_v$ comprise two orthogonal tangent vectors to the surface. We compute the dot product

$$\begin{aligned}
(\phi, \pi^{-1}g) &= \frac{f}{2(g\overline{g}+1)}[(1-g^2)(g+\overline{g}) + (1+g^2)(g-\overline{g}) + g(|g|^2-1)], \\
&= \frac{f}{2(g\overline{g}+1)}[(g+\overline{g}) + (g-\overline{g}) - g^2(g+\overline{g}-g+\overline{g}) + 2g|g|^2 - 2g], \\
&= \frac{f}{2(g\overline{g}+1)}[2g - 2g^2\overline{g} + 2g|g|^2 - 2g], \\
&= 0.
\end{aligned}$$

We conclude that $\pi^{-1}g$ is a unit normal to the surface, and thus we have the following theorem,

5.2.19 Theorem If $\mathbf{x}(\zeta, \overline{\zeta})$ is an isothermal holomorphic patch with $\phi = 2\mathbf{x}_\zeta = \mathbf{x}_u - i\mathbf{x}_v$, then the function $g = \phi^3/(\phi^1 - i\phi^2)$ in the Weierstrass parametrization $\mathcal{W}(f, g)$ of a minimal surface, is the stereographic projection π of the Gauss map. That is,

$$\pi \circ N = g, \tag{5.81}$$

where as usual, $N = (\mathbf{x}_u \times \mathbf{x}_v)/\|\mathbf{x}_u \times \mathbf{x}_v\|$.
Rewriting the expression for the Gaussian curvature 5.74 in terms of f and g, we get,

$$K = -\left[\frac{4|g'|}{|f|(1+|g|^2)}\right]^2. \tag{5.82}$$

We see immediately that the one-parameter family of associated patches $\mathbf{z}_t$ given by

$$\mathbf{z}_t = \Re(e^{it}\phi), \tag{5.83}$$

results on a family of isometric minimal surfaces. This provides another way to view the deformation of a catenoid into a helicoid described earlier in equation 4.116.

5.2.20 Example *Bour surface*
The surface $\mathcal{W}(f, g)$, with $f = 1$ and $g = \sqrt{\sigma}$ was first discussed in 1861 by Bour, who subsequently was awarded the mathematics prize of the French Academy of Sciences. The integrals are completely elementary,

$$\begin{aligned}
x^1 &= \tfrac{1}{2}\Re \int^z (1-\sigma)\, d\sigma = \tfrac{1}{2}\Re\,[z - \tfrac{1}{2}z^2], \\
x^2 &= \tfrac{i}{2}\Re \int^z (1+\sigma)\, d\sigma = \tfrac{1}{2}\Re\,[z + \tfrac{1}{2}z^2], \\
x^3 &= \tfrac{2}{3}\Re\,[z^{3/2}].
\end{aligned}$$

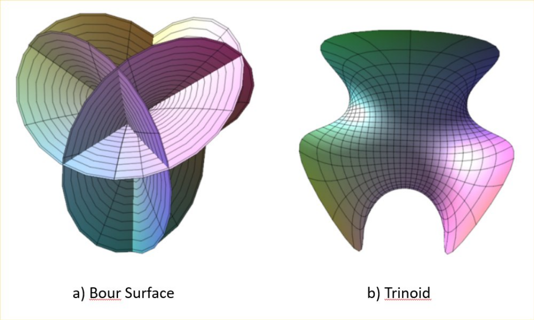


Fig. 5.13: Bour and Trinoid Surfaces

Converting to polar form, we get the parametrization.

$$\mathbf{x}(r,\phi) = (\tfrac{1}{2}r\cos\phi - \tfrac{1}{4}r^2\cos 2\phi, -\tfrac{1}{2}r\sin\phi + \tfrac{1}{4}r^2\sin 2\phi, \tfrac{2}{3}r^{3/2}\cos(\tfrac{2}{3}\phi).) \quad (5.84)$$

The surface in rendered in figure 5.13a, with $r \in [0,4]$.

5.2.21 Example *Trinoid surface*

This surface is part of family of surfaces indexed by $f = \frac{1}{(\sigma^k-1)^2}$, $g = \sigma^{k-1}$, where $k = 2,3\ldots.$ These remarkable surfaces have neat topologies discovered by Jorge and Meeks in 1983. The binoid case $k = 2$ is just a catenoid. The trinoid corresponds to $k = 3$. The integrals for the trinoid yield the following parametrization

$$\begin{aligned}
x^1 &= \Re\left[-\frac{2}{9}\ln(z-1) + \frac{z}{6(z^2+z+1)} + \frac{1}{9}\ln(z^2+z+1)\right],\\
x^2 &= -\frac{1}{9}\mathscr{I}m\left[\frac{1}{z-1} + \frac{z+2}{2(z^2+z+1)} - 2\sqrt{3}\tan^{-1}(\tfrac{1}{\sqrt{3}}(2z+1))\right].\\
x^3 &= -\Re\left[\frac{1}{3(z^3-1)}\right],
\end{aligned}$$

where $z = u + iv$. The algebra involved in finding the real coordinate patch is messy, so the task is best left to a computer algebra system. Converting to polar coordinates by letting $z = re^{i\phi}$ improves rendering the surface plot. As shown in the portion of the surface figure 5.13b, the surface has three openings, meaning that the Gauss map misses three points on S^2. In the same manner, the k-noid is topologically equivalent to a sphere with k points removed.

For convenient reference, we include the following table showing the choices of f and g yielding the listed minimal surfaces.

f	g	**Surface**	**Author, Date**
$-e^{it}/2,\ t=0$	$e^{-\sigma}$	Catenoid	Euler, 1740
$-e^{it/2},\ t=\pi/2$	$e^{-\sigma}$	Helicoid	Meusnier, 1770
$2/(1-\sigma^2)$	σ	Scherk	Scherk, 1834
1	$\sqrt{\sigma}$	Bour	Bour, 1861
1	σ	Enneper	Enneper, 1863
$2(1-\sigma^{-4})$	σ	Henneberg	Henneberg, 1875
$(\sigma^3-1)^{-2}$	σ^2	Trinoid	Jorge and Meeks, 1983
$\wp$	$A/\wp'$	Costa	Costa, 1996

If I had to pick a minimal surface to represent the "orchid" which is the national flower of my native country, I would pick the Costa surface. This surface was not discovered until 1982, triggering a renewed interest on minimal surfaces with non-trivial topologies. A parametrization was not produced until 1996. The integrals involve Weierstrass elliptic functions, which probably would have been most pleasing to Weierstrass, but requires more advanced knowledge than is typically covered on introductory courses on complex variables. The Weierstrass elliptic function $\wp(z)$ is implemented in Maple as,

$$\text{WeierstrassP}(z, g2, g3) = \frac{1}{z^2} + \sum_{\omega}\left(\frac{1}{z-\omega^2} - \frac{1}{\omega}\right). \tag{5.85}$$

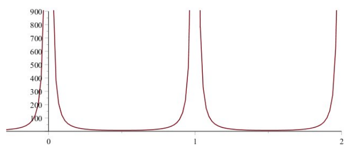


Fig. 5.14: Weierstrass Elliptic Function

The functions $\wp$ are meromorphic with a pole of order 2 at the origin. They are doubly periodic over an arbitrary lattice $\{\omega = 2m\omega_1 + 2n\omega_2 |\ m, n \in \mathbf{Z}\}$ with periods ω_1 and ω_2. The quantities $g2$ and $g2$ are called the invariants and are related to ω_1 and ω_2. In the Maple implementation, the periods are set to $\omega_1 = \omega_2 = \frac{1}{2}$, which does not result in any real loss of generality. A salient property of the invariants is that they link $\wp$ to the cubic differential equation

$$(\wp')^2 = 4(\wp)^3 - g_2\wp - g_3. \tag{5.86}$$

Costa's surface is generated by setting $f = \wp$ and $g = A/\wp'$.

Following Alfred Grey, as shown in Eric Weisstein's Wolfram Mathworld [40], we pick $g_2 = 189.07272$, $g_3 = 0$ and $A = \sqrt{2\pi g_2}$. As we have done before, we first try simply inserting the choices of f and g into $\mathcal{W}(f, g)$, integrating with maple with the substitution $z = u + iv$, and then extracting the real part. The result is a bit disconcerting as one obtains expressions with Weierstrass ζ

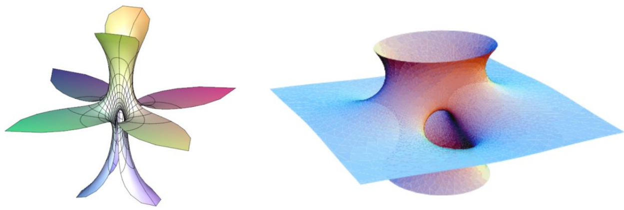


Fig. 5.15: Costa's Minimal Surface

function with very large coefficients of the order of 10^6. The problem becomes immediately clear by observing a plot of the real part $\wp$ as shown in figure 5.14. The $\mathcal{W}(f,g)$ coordinates are integrated by default from 0 to z, so the problem is caused by the pole of order 2 at the origin. Still, one may persist by proceeding to plot the surface by restricting the values of u and v to stay away from the singularities. We chose the ranges to go from 0.02 to 0.98. Surprisingly, Maple takes some time to compute, but it renders the beautiful flowery-shaped curve that appears at the left on figure 5.15. While aesthetically pleasing, the figure is topologically inaccurate. A much more rigorous analysis such as done by A. Grey in deriving the coordinates quoted at the Mathworld site, lead to the widely diffused picture of Costa's surface that appears on the right. Costa's surface is topologically equivalent to a torus (genus=1) with three points removed.

Chapter 6

Riemannian Geometry

6.1 Riemannian Manifolds

In the definition of manifolds introduced in section 4.1, it was implicitly assumed manifolds were embedded (or immersed) in $\mathbf{R}^n$. As such, they inherited a natural metric induced by the standard Euclidean metric of $\mathbf{R}^n$, as shown in section 4.2. For general manifolds it is more natural to start with a topological space M, and define the coordinate patches as pairs $\{U_i, \phi_i\}$, where $\{U_i\}$ is an open cover of M with local homeomorphisms

$$\phi_i : U_i \subset M \to \mathbf{R}^n.$$

If $p \in U_i \cap U_j$ is a point in the non-empty intersection of two charts, we require that the overlap map $\phi_{ij} = \phi_i \phi_j^{-1} : \mathbf{R}^n \to \mathbf{R^n}$ be a diffeomorphism. The local coordinates on patch $\{U, \phi\}$ are given by $(x^1, \ldots, x^n)$, where

$$x^i = u^i \circ \phi,$$

and $u^i : \mathbf{R}^n \to \mathbf{R}$ are the projection maps on each slot. The concept is the same as in figure 4.2, but, as stated, we are not assuming a priori that M is embedded (or immersed) in Euclidean space. If in addition the space is equipped with a metric, the space is called a Riemannian manifold. If the signature of the metric is of type $g = \mathrm{diag}(1, 1, \ldots, -1, -1)$, with p '+' entries and q '-' entries, we say that M is a *pseudo-Riemannian* manifold of type (p, q). As we have done with Minkowski's space, we switch to Greek indices x^μ for local coordinates of curved space-times. We write the Riemannian metric as

$$ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu. \tag{6.1}$$

We will continue to be consistent with earlier notation and denote the tangent space at a point $p \in M$ as T_pM, the tangent bundle as TM, and the space of vector fields as $\mathscr{X}(M)$. Similarly, we denote the space of differential k-forms by $\Omega^k(M)$, and the set of type $\binom{r}{s}$ tensor fields by $\mathscr{T}^r_s(M)$.

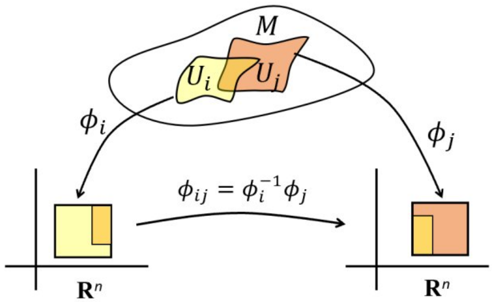


Fig. 6.1: Coordinate Charts

Product Manifolds

Suppose that M_1 and M_2 are differentiable manifolds of dimensions m_1 and m_2 respectively. Then, $M_1 \times M_2$ can be given a natural manifold structure of dimension $n = m_1 + m_2$ induced by the product of coordinate charts. That is, if (ϕ_{i_1}, U_{i_1}) is a chart in M_1 in a neighborhood of $p_1 \in M_1$, and (ϕ_{i_2}, U_{i_2}) is a chart in a neighborhood of $p_2 \in M_2$ in M_2, then the map

$$\phi_{i_1} \times \phi_{i_2} : U_{i_1} \times U_{i_2} \to \mathbf{R}^n$$

defined by

$$(\phi_{i_1} \times \phi_{i_2})(p_1, p_2) = (\phi_{i_1}(p_1), \phi_{i_2}(p_2)),$$

is a coordinate chart in the product manifold. An atlas constructed from such charts, gives the differentiable structure. Clearly, $M_1 \times M_2$ is locally diffeomorphic to $\mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$. To discuss the tangent space of a product manifold, we recall from linear algebra, that given two vector spaces V and W, the *direct sum* $V \oplus W$ is the vector space consisting of the set of ordered pairs

$$V \oplus W = \{(v, w) : v \in V, \quad w \in W\},$$

together with the vector operations

$$\begin{aligned}(v_1, w_1) + (v_2, w_2) &= (v_1 + v_2, w_1 + w_2), \quad \text{for all, } \; v_1, v_2 \in V; \;\; w_1, w_2 \in W,\\ k(v, w) &= (kv, kw), \quad \text{for all } \; k \in \mathbf{R}\end{aligned}$$

People often say that one cannot add apples and peaches, but this is not a problem for mathematicians. For example, 3 apples and 2 peaches plus 4 apples and 6 peaches is 7 apples and 8 peaches. This is the basic idea behind the direct sum. We now have the following theorem:

6.1.1 Theorem Let $(p_1, p_2) \in M_1 \times M_2)$, then there is a vector space isomorphism

$$T_{(p_1, p_2)}(M_1 \times M_2) \cong T_{p_1} M_1 \oplus T_{p_2} M_2.$$

Proof The proof is adapted from [18]. Given $X_1 \in T_{p_1}M_1$ and $X_2 \in T_{p_2}M_2$, let

$$\begin{aligned}
&x_1(t) \text{ be a curve with } x_1(0) = p_1 \text{ and } x_1'(0) = X_1,\\
&x_2(t) \text{ be a curve with } x_2(0) = p_2 \text{ and } x_2'(0) = X_2.
\end{aligned}$$

Then, we can associate

$$(X_1, X_2) \in T_{p_1}M_1 \oplus T_{p_2}M_2$$

with the vector $X \in T_{(p_1,p_2)}(M_1 \times M_2)$, which is tangent to the curve $x(t) = (x_1(t), x_2(t))$, at the point (p_1, p_2). In the simplest possible case where the product manifold is $\mathbf{R}^2 = \mathbf{R}^1 \times \mathbf{R}^1$, the vector X would be the velocity vector $X = x'(t)$ of the curve $x(t)$. It is convenient to introduce the inclusion maps

$$\begin{array}{lll}
i_{p_2}: M_1 & \searrow & \\
& & (M_1 \times M_2)\\
i_{p_1}: M_2, & \nearrow &
\end{array}$$

defined by,

$$\begin{aligned}
i_{p_2}(p) &= (p, p_2), \quad \text{for } p \in M_1,\\
i_{p_1}(q) &= (p_1, q), \quad \text{for } q \in M_2
\end{aligned}$$

The image of the vectors X_1 and X_2 under the push-forward of the inclusion maps

$$\begin{array}{lll}
i_{p_2*}: T_{p_1}M_1 & \searrow & \\
& & T_{(p_1,p_2)}(M_1 \times M_2)\\
i_{p_1*}: T_{p_2}M_2, & \nearrow &
\end{array}$$

yield vectors $\overline{X}_1$ and $\overline{X}_2$, given by,

$$\begin{aligned}
i_{p_2*}(X_1) &= \overline{X}_1 = (x_1'(t), p_2),\\
i_{p_1*}(X_2) &= \overline{X}_1 = (p_1, x_2'(t)).
\end{aligned}$$

Then, it is easy to show that,

$$X = i_{p_2*}(X_1) + i_{p_1*}(X_2).$$

Indeed, if f is a smooth function $f : M_1 \times M_2 \to \mathbf{R}$, we have,

$$\begin{aligned}
X(f) &= \frac{d}{dt}(f(x_1(t), x_2(t))|_{t=0},\\
&= \frac{d}{dt}(f(x_1(t), x_2(0))|_{t=0} + \frac{d}{dt}(f(x_1(0), x_2(t))|_{t=0},\\
&= \overline{X}_1(f) + \overline{X}_2(f).
\end{aligned}$$

More generally, if

$$\varphi : M_1 \times M_2 \to N$$

is a smooth manifold mapping, then we have a type of product rule formula for the Jacobian map,

$$\begin{aligned}\varphi_* X &= \varphi_*(i_{p_2 *}(X_1)) + \varphi_*(i_{p_1 *}(X_2)),\\ &= (\varphi \circ i_{p_2})_* X_1 + (\varphi \circ i_{p_1})_* X_2 \end{aligned} \tag{6.2}$$

This formula will be useful in the treatment of principal fiber bundles, in which case we have a bundle space E, and a Lie group G acting on the right by a product manifold map $\mu : E \times G \to E$.

6.2 Submanifolds

A *Riemannian submanifold* is a subset of a Riemannian manifold that is also Riemannian. The most natural example is a hypersurface in $\mathbf{R}^n$. If $(x^1, x^2 \dots x^n)$ are local coordinates in $\mathbf{R}^n$ with the standard metric, and the surface M is defined locally by functions $x^i = x^i(u^\alpha)$, then M together with the induced first fundamental form 4.12, has a canonical Riemannian structure. We will continue to use the notation $\overline{\nabla}$ for a connection in the ambient space and ∇ for the connection on the surface induced by the tangential component of the covariant derivative

$$\overline{\nabla}_X Y = \nabla_X Y + H(X, Y), \tag{6.3}$$

where $H(X, Y)$ is the component in the normal space. In the case of a hypersurface, we have the classical Gauss equation 4.74

$$\overline{\nabla}_X Y = \nabla_X Y + II(X, Y) N \tag{6.4}$$

$$= \nabla_X Y + < LX, Y > N, \tag{6.5}$$

where $LX = -\overline{\nabla}_X N$ is the Weingarten map. If M is a submanifold of codimension $n-k$, then there are k normal vectors N_k and k classical second fundamental forms $II_k(X, Y)$, so that $H(X, Y) = \sum_k II_k(X, Y) N_k$.

As shown by the theorema egregium, the curvature of a surface in $\mathbf{R}^3$ depends only on the first fundamental form, so the definition of Gaussian curvature as the determinant of the second fundamental form does not even make sense intrinsically. One could redefine K by Cartan's second structure equation as it was used to compute curvatures in Chapter 4, but what we need is a more general definition of curvature that is applicable to any Riemannian manifold. The concept leading to the equations of the theorema egregium involved calculation of the difference of second derivatives of tangent vectors. At the risk of being somewhat misleading, figure 4.95 illustrates the concept. In this figure, the vector field X consists of unit vectors tangent to parallels on the sphere, and the vector field Y are unit tangents to meridians. If an arbitrary tangent vector Z is parallel-transported from one point on an spherical triangle to the diagonally opposed point, the result depends on the path taken. Parallel transport of Z

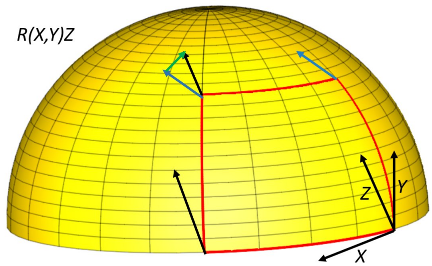


Fig. 6.2: R(X,Y)Z

along X followed by Y, would yield a different outcome that parallel transport along Y followed by parallel transport along X. The failure of the covariant derivatives to commute is a reflection of the existence of curvature. Clearly, the analogous parallel transport by two different paths on a rectangle in $\mathbf{R}^n$ yield the same result. This fact is the reason why in elementary calculus, vectors are defined as quantities that depend only on direction and length. As indicated, the picture is misleading, because, covariant derivatives, as is the case with any other type of derivative, involve comparing the change of a vector under infinitesimal parallel transport. The failure of a vector to return to itself when parallel-transported along a closed path is measured by an entity related to the curvature called the *holonomy* of the connection. Still, the figure should help motivate the definition that follows.

6.2.1 Definition On a Riemannian manifold with connection ∇, the curvature R and the torsion T are defined by:

$$R(X,Y) = \nabla_X\nabla_Y - \nabla_Y\nabla_X - \nabla_{[X,Y]}, \tag{6.6}$$
$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]. \tag{6.7}$$

6.2.2 Theorem The Curvature R is a tensor. At each point $p \in M$, $R(X,Y)$ assigns to each pair of tangent vectors, a linear transformation from T_pM into itself.

Proof Let $X, Y, Z \in \mathscr{X}(M)$ be vector fields on M. We need to establish that R is muiltilinear. Since clearly $R(X,Y) = -R(Y,X)$, we only need to establish linearity on two slots. Let f be a C^∞ function. Then,

$$\begin{aligned}
R(fX,Y) &= \nabla_{fX}\nabla_Y Z - \nabla_Y\nabla_{fX}Z - \nabla_{[fX,Y]}Z,\\
&= f\nabla_X\nabla_Y Z - \nabla_Y(f\nabla_X)Z - \nabla_{[fXY-Y(fX)]}Z,\\
&= f\nabla_X\nabla_Y Z - Y(f)\nabla_X)Z - f\nabla_Y\nabla_X Z - \nabla_{fXY}Z + \nabla_{(Y(f)X+fYX)}Z,\\
&= f\nabla_X\nabla_Y Z - Y(f)\nabla_X)Z - f\nabla_Y\nabla_X Z - f\nabla_{XY}Z + \nabla_{Y(f)X}Z + \nabla_{fYX}Z,\\
&= f\nabla_X\nabla_Y Z - Y(f)\nabla_X)Z - f\nabla_Y\nabla_X Z - f\nabla_{XY}Z + Y(f)\nabla_X Z + f\nabla_{YX}Z,\\
&= f\nabla_X\nabla_Y Z - f\nabla_Y\nabla_X Z - f(\nabla_{XY}Z - \nabla_{YX})Z,\\
&= fR(X,Y)Z.
\end{aligned}$$

Similarly, recalling that $[X,Y] \in \mathscr{X}$, we get:

$$
\begin{aligned}
R(X,Y)(fZ) &= \nabla_X \nabla_Y (fZ) - \nabla_Y \nabla_X (fZ) - \nabla_{[X,Y]}(fZ), \\
&= \nabla_X (Y(f)Z) + f\nabla_Y Z) - \nabla_Y (X(f)Z + f\nabla_X Z) - [X,Y](f)Z) - f\nabla_{[X,Y]}Z, \\
&= XY(f)Z) + Y(f)\nabla_X Z + X(f)\nabla_Y Z + f\nabla_X \nabla_Y Z - \\
&\quad\; YX(f)Z) - X(f)\nabla_Y Z - Y(f)\nabla_X Z - f\nabla_Y \nabla_X Z - \\
&\quad\; [X,Y](f)Z) - f\nabla_{[X,Y]}(Z), \\
&= fR(X,Y)Z.
\end{aligned}
$$

We leave it as an almost trivial exercise to check linearity over addition in all slots.

6.2.3 Theorem The torsion T is also a tensor.
Proof Since $T(X,Y) = -T(Y,X)$, it suffices to prove linearity on one slot. Thus,

$$
\begin{aligned}
T(fX,Y) &= \nabla fXY - \nabla_Y(fX) - [fX,Y], \\
&= f\nabla_X Y - Y(f)X - f\nabla_Y X - fXY + Y(fX), \\
&= f\nabla_X Y - Y(f)X - f\nabla_Y X - fXY + Y(f)X + fYX, \\
&= f\nabla_X Y - f\nabla YX - f[X,Y], \\
&= fT(X,Y).
\end{aligned}
$$

Again, linearity over sums is clear.

6.2.4 Theorem In a Riemannian manifold there exist a unique torsion free connection called the *Levi-Civita connection*, that is compatible with the metric. That is:

$$
[X,Y] = \nabla_X Y - \nabla_Y X, \tag{6.8}
$$

$$
\nabla_X < Y,Z >= < \nabla_X Y, Z > + < Y, \nabla_X Z > . \tag{6.9}
$$

Proof The proof parallels the computation leading to equation 4.76.Let ∇ be a connection compatible with the metric. By taking the three cyclic derivatives of the inner product, and subtracting the third from the sum of the first two

$$
\begin{aligned}
&(a) \qquad \nabla_X < Y,Z >= < \nabla_X Y, Z > + < Y, \nabla_X Z >, \\
&(b) \qquad \nabla_Y < X,Z >= < \nabla_Y X, Z > + < X, \nabla_Y Z >, \\
&(c) \qquad \nabla_Z < X,Y >= < \nabla_Z X, Y > + < X, \nabla_Z Y >, \\
&\qquad (a) + (b) - (c) = < \nabla_X Y, Z > + < \nabla_Y X, Z > + < [X,Z], Y > + < [Y,Z], X > \\
&\qquad\qquad\qquad\quad = 2 < \nabla_X Y, Z > + < [Y,X], Z > + < [X,Z], Y > + < [Y,Z], X >
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
< \nabla_X Y, Z >= \tfrac{1}{2}\{ &\nabla_X < Y,Z > + \nabla_Y < X,Z > - \nabla_Z < X,Y > \\
&+ < [X,Y], Z > + < [Z,X], Y > + < [Z,Y], X > \}.
\end{aligned} \tag{6.10}
$$

The bracket of any two vector fields is a vector field, so the connection is unique since it is completely determined by the metric. In disguise, this is the formula in local coordinates for the Christoffel symbols 4.76. This follows immediately by choosing $X = \partial/\partial x^\alpha$, $Y = \partial/\partial x^\beta$ and $Z = \partial/\partial x^\gamma$. Conversely, if one defines $\nabla_X Y$ by equation 6.10, a long but straightforward computation with lots of cancellations, shows that this defines a connection compatible with the metric.

As before, if $\{e_\alpha\}$ is a frame with dual frame $\{\theta^a\}$, we define the connection forms ω, Christoffel symbols Γ and torsion components in the frame by

$$\nabla_X e_\beta = \omega^\gamma{}_\beta(X)\ e_\gamma, \tag{6.11}$$

$$\nabla_{e_\alpha} e_\beta = \Gamma^\gamma_{\alpha\beta}\ e_\gamma, \tag{6.12}$$

$$T(e_\alpha, e_\beta) = T^\gamma{}_{\alpha\beta}\ e_\gamma. \tag{6.13}$$

As was pointed out in the previous chapter, if the frame is an orthonormal frame such as the coordinate frame $\{\partial/\partial x^\mu\}$ for which the bracket is zero, then $T = 0$ implies that the Christoffel symbols are symmetric in the lower indices.

$$T^\gamma{}_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta} - \Gamma^\gamma_{\beta\alpha} = 0.$$

For such a coordinate frame, we can compute the components of the Riemann tensors as follows:

$$\begin{aligned}
R(e_\gamma, e_\beta)\ e_\delta &= \nabla_{e_\gamma}\nabla_{e_\beta} e_\delta - \nabla_{e_\beta}\nabla_{e_\gamma} e_\delta, \\
&= \nabla_{e_\gamma}(\Gamma^\alpha_{\beta\delta} e_\alpha) - \nabla_{e_\beta}(\Gamma^\alpha_{\gamma\delta} e_\alpha), \\
&= \Gamma^\alpha_{\beta\delta,\gamma} e_\alpha + \Gamma^\alpha_{\beta\delta}\Gamma^\mu_{\gamma\alpha} e_\mu - \Gamma^\alpha_{\gamma\delta,\beta} e_\alpha - \Gamma^\alpha_{\gamma\delta}\Gamma^\mu_{\beta\alpha} e_\mu, \\
&= [\Gamma^\alpha_{\beta\delta,\gamma} - \Gamma^\alpha_{\beta\gamma,\delta} + \Gamma^\mu_{\beta\delta}\Gamma^\alpha_{\gamma\mu} - \Gamma^\mu_{\beta\gamma}\Gamma^\alpha_{\delta\mu}] e_\alpha, \\
&= R^\alpha{}_{\beta\gamma\delta}\ e_\alpha,
\end{aligned}$$

where the components of the Riemann Tensor are defined by:

$$R^\alpha{}_{\beta\gamma\delta} = \Gamma^\alpha_{\beta\delta,\gamma} - \Gamma^\alpha_{\beta\gamma,\delta} + \Gamma^\mu_{\beta\delta}\Gamma^\alpha_{\gamma\mu} - \Gamma^\mu_{\beta\gamma}\Gamma^\alpha_{\delta\mu}. \tag{6.14}$$

Let $X = X^\mu e_\mu$ be and $\alpha = X_\mu \theta^\mu$ be a covariant and a contravariant vector field respectively. Using the notation $\nabla_\alpha = \nabla_{e_\alpha}$ it is almost trivial to compute the covariant derivatives. The results are,

$$\begin{aligned}
\nabla_\beta X &= (X^\mu_{,\beta} + X^\nu \Gamma^\mu{}_{\beta\nu}) e_\mu, \\
\nabla_\beta \alpha &= (X_{\mu,\beta} - X^\nu \Gamma^\nu{}_{\beta\mu}) \theta^\mu,
\end{aligned} \tag{6.15}$$

We show the details of the first computation, and leave the second one as an easy exercise

$$\nabla_\beta X = \nabla_\beta (X^\mu e_\mu), \tag{6.16}$$

$$= X^\mu_{,\beta} e_\mu + X^\mu \Gamma^\delta_{\beta\mu} e_\delta, \tag{6.17}$$

$$= (X^\mu_{,\beta} + X^\nu \Gamma^\mu{}_{\beta\nu}) e_\mu. \tag{6.18}$$

In classical notation, the covariant derivatives $X^{\mu}{}_{\|\beta}$ and $X_{\mu\|\beta}$ are given in terms of the tensor components,

$$\begin{aligned} X^{\mu}{}_{\|\beta} &= X^{\mu}_{,\beta} + X^{\nu}\Gamma^{\mu}{}_{\beta\nu}, \\ X_{\mu\|\beta} &= X_{\mu,\beta} - X_{\nu}\Gamma^{\nu}{}_{\beta\mu}. \end{aligned} \tag{6.19}$$

It is also straightforward to establish the *Ricci identities*

$$\begin{aligned} X^{\mu}{}_{\|\alpha\beta} - X^{\mu}{}_{\|\beta\alpha} &= \quad X^{\nu}R^{\mu}{}_{\nu\alpha\beta}, \\ X_{\mu\|\alpha\beta} - X_{\mu\|\beta\alpha} &= -X_{\nu}R^{\nu}{}_{\mu\alpha\beta}. \end{aligned} \tag{6.20}$$

Again, we show the computation for the first identity and leave the second as a exercise. We take the second derivative. and then reverse the order,

$$\begin{aligned} \nabla_\alpha\nabla_\beta X &= \nabla_\alpha(X^{\mu}_{,\beta}e_\mu + X^{\nu}\Gamma^{\mu}{}_{\beta\nu}e_\mu), \\ &= X^{\mu}_{,\beta\alpha}e_\mu + X^{\mu}_{,\beta}\Gamma^{\delta}_{\alpha\mu}e_\delta + X^{\nu,\alpha}\Gamma^{\mu}_{\beta\mu}e_\nu + X^{\nu}\Gamma^{\mu}_{\beta\nu,\alpha}e_\mu + X^{\nu}\Gamma^{\mu}_{\beta\nu}\Gamma^{\delta}_{\alpha\nu}e_\delta, \\ \nabla_\alpha\nabla_\beta X &= (X^{\mu}_{,\beta\alpha} + X^{\nu}_{,\beta}\Gamma^{\mu}_{\alpha\nu} + X^{\nu}_{,\alpha}\Gamma^{\nu}_{\beta\nu} + X^{\nu}\Gamma^{\mu}_{\beta\nu,\alpha} + X^{\nu}\Gamma^{\delta}_{\beta\nu}\Gamma^{\mu}_{\alpha\delta})e_\mu, \\ \nabla_\beta\nabla_\alpha X &= (X^{\mu}_{,\alpha\beta} + X^{\nu}_{,\alpha}\Gamma^{\mu}_{\beta\nu} + X^{\nu}_{,\beta}\Gamma^{\nu}_{\alpha\nu} + X^{\nu}\Gamma^{\mu}_{\alpha\nu,\beta} + X^{\nu}\Gamma^{\delta}_{\alpha\nu}\Gamma^{\mu}_{\beta\delta})e_\mu. \end{aligned}$$

Subtracting the last two equations, only the last two terms of each survive, and we get the desired result,

$$\begin{aligned} 2\nabla_{[\alpha}\nabla_{\beta]}(X) &= X^{\nu}(\Gamma^{\mu}_{\beta\nu,\alpha} - \Gamma^{\mu}_{\alpha\nu,\beta} + \Gamma^{\delta}_{\beta\nu}\Gamma^{\mu}_{\alpha\delta} - \Gamma^{\delta}_{\alpha\nu}\Gamma^{\mu}_{\beta\delta})e_\mu, \\ 2\nabla_{[\alpha}\nabla_{\beta]}(X^{\mu}e_\mu) &= \quad (X^{\nu}R^{\mu}{}_{\nu\alpha\beta})e_\mu. \end{aligned}$$

In the lietrature, many authors use the notation $\nabla_\beta X^\mu$ to denote the covariant derivative $X^{\mu}{}_{\|\beta}$, but it is really an (excusable) abuse of notation that arises from thinking of tensors as the components of the tensors. The Ricci identities are the basis for the notion of holonomy, namely, the simple interpretation that the failure of parallel transport to commute along the edges of an rectangle, indicates the presence of curvature. With more effort with repeated use of Leibnitz rule, one can establish more elaborate Ricci identities for higher order tensors. If one assumes zero torsion, the Ricci identities of higher order tensors just involve more terms with the curvature. It the torsion is not zero, there are additional terms involving the torsion tensor; in this case it is perhaps a bit more elegant to use the covariant differential introduced in the next section, so we will postpone the computation until then.

The generalization of the theorema egregium to manifolds comes from the same principle of splitting the curvature tensor of the ambient space into the tangential on normal components. In the case of a hypersurface with normal

N and tangent vectors X, Y, Z, we have:

$$
\begin{aligned}
\overline{R}(X,Y)Z &= \overline{\nabla}_X \overline{\nabla}_Y Z - \overline{\nabla}_Y \overline{\nabla}_X Z - \overline{\nabla}_{[X,Y]} Z, \\
&= \overline{\nabla}_X (\nabla_Y Z + < LY, Z > N) - \overline{\nabla}_Y (\nabla_X Z + < LX, Z > N) - \overline{\nabla}_{[X,Y]} Z, \\
&\quad \nabla_X \nabla_{\overline{Y}} Z + < LX, \nabla_Y Z > N + X < LY, Z > N + < LY, Z > LX - \\
&\quad \nabla_Y \nabla_X Z - < LY, \nabla_Y Z > N - Y < LX, Z > N - < LX, Z > LY - \\
&\quad \nabla_{[X,Y]} Z - < L([X,Y]), Z > N, \\
&\quad \nabla_X \nabla_{\overline{Y}} Z + < LX, \nabla_Y Z > N + X < LY, Z > N + < LY, Z > LX - \\
&\quad \nabla_Y \nabla_X Z - < LY, \nabla_Y Z > N - Y < LX, Z > N - < LX, Z > LY - \\
&\quad \nabla_{[X,Y]} Z - < L([X,Y]), Z > N, \\
&= \nabla_X \nabla_Y Z + < LX, \nabla_Y Z > N + < \nabla_X LY, Z > N + < LY, \nabla_X Z > N + < LY, Z > LX - \\
&\quad \nabla_Y \nabla_X Z - < LY, \nabla_Y Z > N - < \nabla_Y LX, Z > N - < LX, \nabla_Y Z > N - < LX, Z > LY - \\
&\quad \nabla_{[X,Y]} Z - < L([X,Y]), Z > N, \\
&= R(X,Y)Z + < LY, Z > LX - < LX, Z > LY + \\
&\quad \{< \nabla_X LY, Z > - < \nabla_Y LX, Z > - < L([X,Y]), Z >\} N.
\end{aligned}
$$

If the ambient space is $\mathbf{R}^n$, the curvature tensor $\overline{R}$ is zero, so we can set the horizontal and normal components in the right to zero. Noting that the normal component is zero for all Z, we get:

$$R(X,Y)Z + < LY, Z > LX - < LX, Z > LY = 0, \tag{6.21}$$

$$\nabla_X LY - \nabla_Y LX - L([X,Y]) = 0. \tag{6.22}$$

In particular, if $n = 3$, and at each point in the surface, the vectors X and Y constitute an a basis of the tangent space, we get the coordinate-free theorema egregium

$$K = < R(X,Y)X, Y > = < LX, X >< LY, Y > - < LY, X >< LX, Y > = \det(L). \tag{6.23}$$

The expression 6.22 is the coordinate-independent version of the equation of Codazzi.

We expect the covariant definition of the torsion and curvature tensors to be consistent with the formalism of Cartan.

6.2.5 Theorem Equations of Structure.

$$\Theta^\alpha = d\theta^\alpha + \omega^\alpha{}_\beta \wedge \theta^\beta, \tag{6.24}$$

$$\Omega^\alpha{}_\beta = d\omega^\alpha{}_\beta + \omega^\alpha{}_\gamma \wedge \omega^\gamma{}_\beta. \tag{6.25}$$

To verify this is the case, we define:

$$T(X,Y) = \Theta^\alpha(X,Y) e_\alpha, \tag{6.26}$$

$$R(X,Y) e_\beta = \Omega^\alpha{}_\beta(X,Y) e_\alpha. \tag{6.27}$$

Recalling that any tangent vector X can be expressed in terms of the basis as

$X = \theta^\alpha(X)\, e_\alpha$, we can carry out a straight-forward computation:

$$
\begin{aligned}
\Theta^\alpha(X,Y)\, e_\alpha &= T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y], \\
&= \nabla_X(\theta^\alpha(Y))\, e_\alpha - \nabla_Y(\theta^\alpha(X))\, e_\alpha - \theta^\alpha([X,Y])\, e_\alpha, \\
&= X(\theta^\alpha(Y))\, e_\alpha + \theta^\alpha(Y)\, \omega^\beta{}_\alpha(X)\, e_\beta - Y(\theta^\alpha(X))\, e_\alpha \\
&\quad - \theta^\alpha(X)\, \omega^\beta{}_\alpha(Y)\, e_\beta - \theta^\alpha([X,Y])\, e_\alpha, \\
&= \{X(\theta^\alpha(Y)) - Y(\theta^\alpha(X)) - \theta^\alpha([X,Y]) + \omega^\alpha{}_\beta(X)(\theta^\beta(Y) - \omega^\alpha{}_\beta(Y)(\theta^\beta(X)\}e_\alpha, \\
&= \{(d\theta^\alpha + \omega^\alpha{}_\beta \wedge \theta^\beta)(X,Y)\}e_\alpha,
\end{aligned}
$$

where we have introduced a coordinate-free definition of the differential of the one form θ by

$$d\theta(X,Y) = X(\theta(Y)) - y(\theta(X)) - \theta([X,Y]). \tag{6.28}$$

It is easy to verify that this definition of the differential of a one form satisfies all the required properties of the exterior derivative, and that it is consistent with the coordinate version of the differential introduced in Chapter 2. We conclude that

$$\Theta^\alpha = d\theta^\alpha + \omega^\alpha{}_\beta \wedge \theta^\beta, \tag{6.29}$$

which is indeed the first Cartan equation of structure. Proceeding along the same lines, we compute:

$$
\begin{aligned}
\Omega^\alpha{}_\beta(X,Y)\, e_\alpha &= \nabla_X \nabla_Y\, e_\beta - \nabla_Y \nabla_X\, e_\beta - \nabla_{[X,Y]}\, e_\beta, \\
&= \nabla_X(\omega^\alpha{}_\beta(Y)\, e_\alpha) - \nabla_Y(\omega^\alpha{}_\beta(X)\, e_\alpha) - \omega^\alpha{}_\beta([X,Y])\, e_\alpha, \\
&= X(\omega^\alpha{}_\beta(Y))\, e_\alpha + \omega^\alpha{}_\beta(Y)\, \omega^\gamma{}_\alpha(X)\, e_\gamma - Y(\omega^\alpha{}_\beta(X))\, e_\alpha \\
&\quad - \omega^\alpha{}_\beta(X)\, \omega^\gamma{}_\alpha(Y)\, e_\gamma - \omega^\alpha{}_\beta([X,Y])\, e_\alpha \\
&= \{(d\omega^\alpha{}_\beta + \omega^\alpha{}_\gamma \wedge \omega^\gamma{}_\beta)(X,Y)\}e_\alpha,
\end{aligned}
$$

thus arriving at the second equation of structure

$$\Omega^\alpha{}_\beta = d\omega^\alpha{}_\beta + \omega^\alpha{}_\gamma \wedge \omega^\gamma{}_\beta. \tag{6.30}$$

The quantities connection and curvature forms are matrix-valued. Using matrix multiplication notation, we can abbreviate the equations of structure as

$$
\begin{aligned}
\Theta &= d\theta + \omega \wedge \theta, \\
\Omega &= d\omega + \omega \wedge \omega.
\end{aligned} \tag{6.31}
$$

Taking the exterior derivative of the structure equations gives some interesting results. Here is the first computation,

$$
\begin{aligned}
d\Theta &= d\omega \wedge \theta - \omega \wedge d\theta, \\
&= d\omega \wedge \theta - \omega \wedge (\Theta - \omega \wedge \theta), \\
&= d\omega \wedge \theta - \omega \wedge \Theta + \omega \wedge \omega\theta, \\
&= (d\omega + \omega \wedge \omega) \wedge \theta - \omega \wedge \Theta, \\
&= \Omega \wedge \theta - \omega \wedge \Theta,
\end{aligned}
$$

so,

$$d\Theta + \omega \wedge \Theta = \Omega \wedge \theta. \tag{6.32}$$

Similarly, taking d of the second structure equation we get,

$$\begin{aligned} d\Omega &= d\omega \wedge \omega + \omega \wedge d\omega, \\ &= (\Omega - \omega \wedge \omega) \wedge \omega + \omega \wedge (\Omega - \omega \wedge \omega). \end{aligned}$$

Hence,

$$d\Omega = \Omega \wedge \omega - \omega \wedge \Omega. \tag{6.33}$$

Equations 6.32 and 6.33 are called the first and second Bianchi identities. The relationship between the torsion and Riemann tensor components with the corresponding differential forms are given by

$$\begin{aligned} \Theta^\alpha &= \tfrac{1}{2} T^\alpha{}_{\gamma\delta}\, \theta^\gamma \wedge \theta^\delta, \\ \Omega^\alpha{}_\beta &= \tfrac{1}{2} R^\alpha{}_{\beta\gamma\delta}\, \theta^\gamma \wedge \theta^\delta. \end{aligned} \tag{6.34}$$

In the case of a non-coordinate frame in which the Lie bracket of frame vectors does not vanish, we first write them as linear combinations of the frame

$$[e_\beta, e_\gamma] = C^\alpha_{\beta\gamma}\, e_\alpha. \tag{6.35}$$

The components of the torsion and Riemann tensors are then given by

$$\begin{aligned} T^\alpha{}_{\beta\gamma} &= \Gamma^\alpha_{\beta\gamma} - \Gamma^\alpha_{\gamma\beta} - C^\alpha_{\beta\gamma}, \\ R^\alpha{}_{\beta\gamma\delta} &= \Gamma^\alpha_{\beta\delta,\gamma} - \Gamma^\alpha_{\beta\gamma,\delta} + \Gamma^\mu_{\beta\delta}\Gamma^\alpha_{\gamma\mu} - \Gamma^\mu_{\beta\gamma}\Gamma^\alpha_{\delta\mu} - \Gamma^\alpha_{\beta\mu}C^\mu_{\gamma\delta} - \Gamma^\alpha{}_{\sigma\beta}C^\sigma{}_{\gamma\delta}. \end{aligned} \tag{6.36}$$

The Riemann tensor for a torsion-free connection has the following symmetries;

$$\begin{aligned} R(X,Y) &= -R(Y,X), \\ < R(X,Y)Z, W > &= - < R(X,Y)W, Z >, \\ R(X,Y)Z + R(Z,X)Y + R(Y,Z)X &= 0. \end{aligned} \tag{6.37}$$

In terms of components, the Riemann Tensor symmetries can be expressed as

$$\begin{aligned} R_{\alpha\beta\gamma\delta} &= -R_{\alpha\beta\delta\gamma} = -R_{\beta\alpha\gamma\delta}, \\ R_{\alpha\beta\gamma\delta} &= R_{\gamma\delta\alpha\beta}, \\ R_{\alpha\beta\gamma\delta} + R_{\alpha\gamma\delta\beta} + R_{\alpha\delta\beta\gamma} &= 0. \end{aligned} \tag{6.38}$$

The last cyclic equation is the tensor version of the first Bianchi Identity with 0 torsion. It follows immediately from setting $\Omega \wedge \theta = 0$ and taking a cyclic permutation of the antisymmetric indices $\{\beta, \gamma, \delta\}$ of the Riemann tensor. The symmetries reduce the number of independent components in an n-dimensional manifold from n^4 to $n^2(n^2-1)/12$. Thus, for a 4-dimensional space, there are at most 20 independent components. The derivation of the tensor version of the second Bianchi identity from the elegant differential forms version, takes a bit more effort. In components the formula

$$d\Omega = \Omega \wedge \omega - \omega \wedge \Omega$$

reads,

$$R^{\alpha}{}_{\beta\kappa\lambda;\mu}\,\theta^{\mu} \wedge \theta^{\kappa} \wedge \theta^{\lambda} = (\Gamma^{\rho}{}_{\mu\beta} R^{\alpha}{}_{\rho\kappa\lambda} - \Gamma^{\alpha}{}_{\mu\rho} R^{\rho}{}_{\beta\kappa\lambda})\theta^{\mu} \wedge \theta^{\kappa} \wedge \theta^{\lambda},$$

where we used the notation,

$$\nabla_{\mu} R^{\alpha}{}_{\beta\kappa\lambda} = R^{\alpha}{}_{\beta\kappa\lambda;\mu}.$$

Taking a cyclic permutation on the antisymmetric indices κ, λ, μ, and using some index gymnastics to show that the right hand becomes zero, the tensor version of the second Bianchi identity for zero torsion becomes

$$R^{\alpha}{}_{\beta[\kappa\lambda;\mu]} = 0 \tag{6.39}$$

6.3 Sectional Curvature

Let $\{M, g\}$ be a Riemannian manifold with Levi-Civita connection ∇ and curvature tensor $R(X, Y)$. In local coordinates at a point $p \in M$ we can express the components

$$R = R_{\mu\nu\rho\sigma}\, dx^{\mu} dx^{\nu} dx^{\rho} dx^{\sigma}$$

of a covariant tensor of rank 4. With this in mind, we define a multilinear function

$$R : T_p(M) \otimes T_p(M) \otimes T_p(M) \otimes T_p(M) \to \mathbf{R},$$

by

$$R(W, Y, X, Z) =< W, R(X, Y)Z > \tag{6.40}$$

In this notation, the symmetries of the tensor take the form,

$$\begin{aligned} R(W, X, Y, Z) &= -R(W, Y, X, Z), \\ R(W, X, Y, Z) &= -R(Z, Y, X, W) \\ R(W, X, Y, Z) + R(W, Z, X, Y) + R(X, Y, Z, X) &= 0. \end{aligned} \tag{6.41}$$

From the metric, we can also define a multilinear function

$$G(W, Y, X, Z) =< Z, Y >< X, W > - < Z, X >< Y, W > .$$

Now, consider any 2-dimensional plane $V_p \subset T_p(M)$ and let $X, Y \in V$ be linearly independent. Then,

$$G(X, Y, X, Y) =< X, X >< Y, Y > - < X, Y >^2$$

is a bilinear form that represents the area of the parallelogram spanned by X and Y. If we perform a linear, non-singular change of of coordinates,

$$X' = aX + bY, \quad y' = cX + dY, \quad ad - bc \neq 0,$$

then both, $G(X, Y.X, Y)$ and $R(X, Y, X, Y)$ transform by the square of the determinant $D = ad - bc$, so the ratio is independent of the choice of vectors. We define the *sectional curvature* of the subspace V_p by

$$\begin{aligned} K(V_p) &= \frac{R(X, Y, X, Y)}{G(X, Y, X, Y)}, \\ &= \frac{R(X, Y, X, Y)}{< X, X >< Y, Y > - < X, Y >^2} \end{aligned} \tag{6.42}$$

The set of values of the sectional curvatures for all planes at $T_p(M)$ completely determines the Riemannian curvature at p. For a surface in $\mathbf{R}^3$ the sectional curvature is the Gaussian curvature, and the formula is equivalent to the theorema egregium. If $K(V_p)$ is constant for all planes $V_p \in T_p(M)$ and for all points $p \in M$, we say that M is a space of *constant curvature.* For a space of constant curvature k, we have

$$R(X, Y)Z = k(< Z, Y > X- < Z, X > Y) \tag{6.43}$$

In local coordinates, the equation gives

$$R_{\mu\nu\rho\sigma} = k(g_{\nu\sigma}g_{\mu\rho} - g_{\nu\gamma}g_{\mu\sigma}). \tag{6.44}$$

6.3.1 Example

The model space of manifolds of constant curvature is a quadric hypersurface M of $\mathbf{R}^{n+1}$ with metric

$$ds^2 = \epsilon k^2 dt^2 + (dy^1)^2 + \cdots + dy^n)^2,$$

given by the equation

$$M : \epsilon k^2 t^2 + (y^1)^2 + \ldots (y^n)^2 = \epsilon k^2, \quad t \neq 0,$$

where k is a constant and $\epsilon = \pm 1$. For the purposes of this example it will actually be simpler to completely abandon the summation convention. Thus, we write the quadric as

$$\epsilon k^2 t^2 + \Sigma_i (y^i)^2 = \epsilon k^2.$$

If $k = 0$, the space is flat. If $\epsilon = 1$, let $(y^0)^2 = k^2 t^2$ and the quadric is isometric to a sphere of constant curvature $1/k^2$. If $\epsilon = -1$, $\Sigma_i (x^i)^2 = -k^2(1 - t^2) > 0$, then $t^2 < 1$ and the surface is a hyperboloid of two sheets. Consider the mapping from $(\mathbf{R})^{\mathbf{n+1}}$ to $\mathbf{R}^n$ given by

$$x^i = y^i/t.$$

We would like to compute the induced metric on the the surface. We have

$$-k^2t^2 + \Sigma_i(y^i)^2 = -k^2t^2 + t^2\Sigma_i(x^i)^2 = -k^2$$

so

$$t^2 = \frac{-k^2}{-k^2 + \Sigma_i(x^i)^2}.$$

Taking the differential, we get

$$t\;dt = \frac{-k^2\Sigma_i(x^i dx^i)}{-k^2 + \Sigma_i(x^i)^2}.$$

Squaring and dividing by t^2 we also have

$$dt^2 = \frac{-k^2(\Sigma_i x^i\,dx^i)^2}{(-k^2 + \Sigma_i(x^i)^2)^3}.$$

From the product rule, we have $dy^i = x^i\,dt + t\,dx^i$, so the metric is

$$\begin{aligned}
ds^2 &= -k^2\,dt^2 + [\Sigma(x^i)^2]dt^2 + 2t\,dt\,\Sigma_i(x^i\,dx^i) + t^2\Sigma(dx^i)^2,\\
&= [-k^2 + \Sigma_i(x^i)^2]\,dt^2 + 2t\,dt\,\Sigma_i(x^i\,dx^i) + t^2\Sigma(dx^1)^2,\\
&= \frac{-k^2[\Sigma_i(x^i dx^i)]^2}{[-k^2 + \Sigma_i(x^i)^2]^2} + \frac{2k^2[\Sigma_i(x^i dx^i)]^2}{[-k^2 + \Sigma_i(x^i)^2]^2} + \frac{-k^2\Sigma_i(dx^i)^2}{-k^2 + \Sigma_i(dx^i)^2},\\
&= k^2\,\frac{[k^2 - \Sigma_i(x^i)^2]\Sigma_i(dx^i)^2 - (\Sigma_i(x^i\,dx^i))^2}{[k^2 - \Sigma_i(x^i)^2]^2}
\end{aligned}$$

It is not obvious, but in fact, the space is also of constant curvature $(-1/k^2)$. For an elegant proof, see [18]. When $n = 4$ and $\epsilon = -1$, the group leaving the metric

$$ds^2 = -k^2dt^2 + (dy^1)^2 + (dy^2)^2 + (dy^3)^2 + (dy^4)^2$$

invariant, is the Lorentz group $O(1, 4)$. With a minor modification of the above, consider the quadric

$$M : -k^2t^2 + (y^1)^2 + \ldots (y^4)^2 = k^2.$$

In this case, the quadric is a hyperboloid of one sheet, and the submanifold with the induced metric is called the *de Sitter* space. The isotropy subgroup that leaves $(1, 0, 0, 0, 0)$ fixed is $O(1, 3)$ and the manifold is diffeomorphic to $O(1, 4)/O(1, 3)$. Many alternative forms of the de Sitter metric exist in the literature. One that is particularly appealing is obtained as follows. Write the metric in ambient space as

$$ds^2 = -(dy^0)^2 + (dy^1)^2 + (dy^2)^2 + (dy^3)^2 + (dy^4)^2$$

with the quadric given by

$$M : -(y^0)^2 + (y^1)^2 + \ldots (y^4)^2 = k^2.$$

Let

$$\sum_{i=1}^{4}(x^i)^2 = 1$$

so M represents a unit sphere S^3. Introduce the coordinates for M

$$\begin{aligned} y^0 &= k\sinh(\tau/k), \\ y^i &= k\cosh(\tau/k). \end{aligned}$$

Then, we have

$$\begin{aligned} dy^0 &= \cosh(\tau/k)\,d\tau, \\ dy^i &= \sinh(\tau/k)x^i\,d\tau + k\cosh(\tau/k)\,dx^i. \end{aligned}$$

The induced metric on M becomes,

$$\begin{aligned} ds^2 &= -[\cosh^2(\tau/k) - \sinh^2(\tau/k)\Sigma_i(x^i)^2]\,d\tau + \cosh^2(\tau/k)\Sigma_i(dx^i)^2, \\ &= -d\tau^2 + \cosh^2(\tau/k)d\Omega^2, \end{aligned}$$

where $d\Omega$ is the volume form for S^3. The most natural coordinates for the volume form are the Euler angles and Cayley-Klein parameters. The interpretation of this space-time is that we have a spatial 3-sphere which propagates in time by shrinking to a minimum radius at the throat of the hyperboloid, followed by an expansion. Being a space of constant curvature, the Ricci tensor is proportional to the metric, so this is an Einstein manifold.

6.4 Big D

In this section we discuss the notion of a connection on a vector bundle E. Let M be a smooth manifold and as usual we denote by $T^r_s(p)$ the vector space of type $\binom{r}{s}$ tensors at a point $p \in M$. The formalism applies to any vector bundle, but in this section we are primarily concerned with the case where E is the tensor bundle $E = T^r_s(M)$. Sections $\Gamma(E) = \mathscr{T}^r_s(M)$ of this bundle are called tensor fields on M. For general vector bundles, we use the notation $s \in \Gamma(E)$ for the sections of the bundle. The section that maps every point of M to the zero vector, is called the *zero section*. Let $\{e_\alpha\}$ be an orthonormal frame with dual forms $\{\theta^\alpha\}$. We define the space $\Omega^p(M, E)$ tensor-valued p-form as sections of the bundle,

$$\Omega^p(M, E) = \Gamma(E \otimes \Lambda^p(M)). \tag{6.45}$$

As in equation 2.63, a tensor-valued p form is a tensor of type $\binom{r}{s+p}$ with components,

$$T = T^{\alpha_1,\ldots\alpha_r}_{\beta_1,\ldots\beta_s,\gamma_1,\ldots,\gamma_p} e_{\alpha_1} \otimes \ldots e_{\alpha_r} \otimes \theta^{\beta_1} \otimes \cdots \otimes \theta^{\beta_s} \wedge \theta^{\gamma_1} \wedge \ldots \wedge \theta^{\gamma_p}. \tag{6.46}$$

A tensor-valued 0-form is just a regular tensor field $T \in \mathscr{T}^r_s(M)$ The main examples of tensor-valued forms are the torsion and the curvature forms

$$\begin{aligned}\Theta &= \Theta^\alpha \otimes e_\alpha, \\ \Omega &= \Omega^\alpha{}_\beta \otimes e_\alpha \otimes \theta^\beta.\end{aligned} \tag{6.47}$$

The tensorial components of the torsion tensor, would then be written as

$$\begin{aligned}T &= T^\alpha{}_{\beta\gamma} e_\alpha \otimes \theta^\beta \otimes \theta^\gamma, \\ &= \frac{1}{2} T^\alpha{}_{\beta\gamma} e_\alpha \otimes \theta^\beta \wedge \theta^\gamma, \\ &= e_\alpha \otimes (\tfrac{1}{2} T^\alpha{}_{\beta\gamma} \theta^\beta \wedge \theta^\gamma).\end{aligned}$$

since the tensor is antisymmetric in the lower indices. Similarly, the tensorial components of the curvature are

$$\begin{aligned}\Omega &= \frac{1}{2} R^\alpha{}_{\beta\gamma\delta}\; e_\alpha \otimes \theta^\beta \otimes \theta^\gamma \wedge \theta^\delta, \\ &= \; e_\alpha \otimes \theta^\beta \otimes (\tfrac{1}{2} R^\alpha{}_{\beta\gamma\delta} \theta^\gamma \wedge \theta^\delta).\end{aligned}$$

The connection forms

$$\omega = \omega^\alpha{}_\beta \otimes e_\alpha \otimes \theta^\beta \tag{6.48}$$

are matrix-valued, but they are not tensorial forms. If T is a type $\binom{r}{s}$ tensor field, and α a p-form, we can write a tensor-valued p-form as $T \otimes \alpha \in \Omega^p(M, E)$ is. We seek an operator that behaves like a covariant derivative ∇ for tensors and exterior derivative d for forms.

6.4.1 Linear Connections

Given a vector field X and a smooth function f, we define a *linear connection* as a map

$$\nabla_X : \Gamma(T^r_s) \to \Gamma(T^r_s)$$

with the following properties

1) $\nabla_X(f) = X(f)$,
1) $\nabla_{fX} T = f D_X T$,
2) $\nabla_{X+Y} T = \nabla_X T + \nabla_Y T$, for all $X, Y \in \mathscr{X}(M)$,
3) $\nabla_X(T_1 + T_2) = \nabla_X T_1 + \nabla_X T_2$, for $T_1, T_2 \in \Gamma(T^r_s)$,
4) $\nabla_X(fT) = X(f)T + f\nabla_X T$.

If instead of the tensor bundle we have a general vector bundle E, we replace the tensor fields in the definition above by sections $s \in \Gamma(E)$ of the vector bundle. The definition induces a derivation on the entire tensor algebra satisfying the additional conditions,

5) $\nabla_X(T_1 \otimes T_2) = \nabla_X T_1 \otimes T_2 + T_1 \otimes \nabla_X T_2$,
6) $\nabla_X \circ C = C \circ \nabla_X$, for any contraction C.

The properties are the same as a Koszul connection, or covariant derivative for tensor-valued 0 forms T. Given an orthonormal frame, consider the identity tensor,

$$I = \delta^\alpha{}_\beta\; e_\alpha \otimes \theta^\beta, \tag{6.49}$$

and take the covariant derivative ∇_X. We get

$$\begin{aligned}
\nabla_X e_\alpha \otimes \theta^\alpha + e_\alpha \otimes \nabla_X \theta^\alpha &= 0, \\
e_\alpha \otimes \nabla_X \theta^\alpha &= -\nabla_X e_\alpha \otimes \theta^\alpha, \\
&= -e_\beta\, \omega^\beta{}_\alpha(X) \otimes \theta^\alpha, \\
e_\beta \otimes \nabla_X \theta^\beta &= -e_\beta\, \omega^\beta{}_\alpha(X) \otimes \theta^\alpha,
\end{aligned}$$

which implies that,

$$\nabla_X \theta^\beta = -\omega^\beta{}_\alpha(X)\theta^\alpha. \tag{6.50}$$

Thus as before, since we have formulas for the covariant derivative of basis vectors and forms, we are led by induction to a general formula for the covariant derivative of an $\binom{r}{s}$-tensor given mutatis mutandis by the formula 3.32. In other words, the covariant derivative of a tensor acquires a term with a multiplicative connection factor for each contravariant index and a negative term with a multiplicative connection factor for each covariant index.

6.4.1 Definition A connection ∇ on the vector bundle E is a map

$$\nabla : \Gamma(M.E) \to \Gamma(M, E \otimes T^*(M))$$

which satisfies the following conditions

a) $\nabla(T_1 + T_2) = \nabla T_1 + \nabla T_2, \quad T_1, T_2 \in \Gamma(E))$,

b) $\nabla(fT) = df \otimes T + f\nabla T$,

c) $\nabla_X T = i_X \nabla T$.

As a reminder of the definition of the inner product i_X, condition (c) is equivalent to the equation,

$$\nabla T(\theta^1, \ldots, \theta^r, X, X_1, \ldots, X_s) = (\nabla_X T)(\theta^1, \ldots, \theta^r, X_1, \ldots, X_s).$$

In particular, if X is vector field, then, as expected

$$\nabla X(Y) = \nabla_X Y,$$

The operator ∇ is called the *covariant differential.* Again, for a general vector bundles, we denote the sections by $s \in \Gamma(E)$ and the covariant differential by ∇s.

6.4.2 Affine Connections

A connection on the tangent bundle $T(M)$ is called an *affine connection.* In a local frame field e, we may assume that the connection is represented by a matrix of one-forms ω

$$\begin{aligned}
\nabla e_\beta &= e_\alpha \otimes \omega^\alpha{}_\beta, \\
\nabla e &= e \otimes \omega.
\end{aligned} \tag{6.51}$$

The tensor multiplication symbol is often omitted when it is clear in context. Thus, for example, the connection equation is sometimes written as $\nabla e = e\omega$.

In a local coordinate system $\{x_1, \ldots, x^n\}$, with basis vectors $\partial_\mu = \frac{\partial}{\partial x^\mu}$ and dual forms dx^μ, we have,

$$\omega^\alpha{}_{\beta\mu} = \Gamma^\alpha_{\beta\mu}\, dx^\mu.$$

From equation 6.50, it follows that

$$\nabla\theta^\alpha = -\omega^\alpha{}_\beta \otimes \theta^\beta. \tag{6.52}$$

We need to be a bit careful with the dual forms θ^α. We can view them as a vector-valued 1-form

$$\theta = e_\theta \otimes \theta^\alpha,$$

which has the same components as the identity $\binom{1}{1}$ tensor. This is a kind of a odd creature. About the closest analog to this entity in classical terms is the differential of arc-length

$$d\mathbf{x} = \mathbf{i}\, dx + \mathbf{j}\, dy + \mathbf{k}\, dz,$$

which is sort of a mixture of a vector and a form. The vector of differential forms would then be written as a column vector.
In a frame $\{e_\alpha\}$, the covariant differential of tensor-valued 0-form T is given by

$$\nabla T = \nabla_{e_\alpha} T \otimes \theta^\alpha \equiv \nabla_\alpha T \otimes \theta^\alpha.$$

In particular, if $X = v^\alpha e_\alpha$, we get,

$$\begin{aligned}
\nabla X = \nabla_\beta X \otimes \theta^\beta &= \nabla_\beta(v^\alpha e_\alpha) \otimes \theta^\beta, \\
&= (\nabla_\beta(v^\alpha)\, e_\alpha + v^\alpha \Gamma^\gamma_{\beta\alpha}\, e_\gamma) \otimes \theta^\beta, \\
&= (v^\alpha_{,\beta} + v^\gamma \Gamma^\alpha_{\beta\gamma}) e_\alpha \otimes \theta^\beta \\
&= v^\alpha_{\|\beta}\, e_\alpha \otimes \theta^\beta,
\end{aligned}$$

where we have used the classical symbols

$$v^\alpha{}_{\|\beta} = v^\alpha{}_{,\beta} + \Gamma^\alpha_{\beta\gamma}\, v^\gamma, \tag{6.53}$$

for the covariant derivative components $v^\alpha_{\|\beta}$ and the comma to abbreviate the directional derivative $\nabla_\beta(v^\alpha)$. Of course, the formula is in agreement with equation 3.25. ∇X is a $\binom{1}{1}$-tensor.
Similarly, for a covariant vector field $\alpha = v_\alpha \theta^\alpha$, we have

$$\begin{aligned}
\nabla\alpha &= \nabla(v_\alpha \otimes \theta^\alpha) \\
&= \nabla v_\alpha \otimes \theta^\alpha - v_\beta\, \omega^\beta{}_\alpha \otimes \theta^\alpha, \\
&= (\nabla_\gamma v_\alpha\, \theta^\gamma - v_\beta \Gamma^\beta_{\alpha\gamma} \theta^\gamma) \otimes \theta^\alpha, \\
&= (v_{\alpha,\gamma} - \Gamma^\alpha_{\beta\gamma}\, v_\alpha)\, \theta^\gamma \otimes \theta^\alpha,
\end{aligned}$$

hence,

$$v_{\alpha\|\beta} = v_{\alpha,\gamma} - \Gamma^\alpha_{\beta\gamma}\, v_\alpha. \tag{6.54}$$

As promised earlier, we now prove the Ricci identities for contravariant and covariant vectors when the torsion is not zero. Ricci Identities with torsion. The results are,

$$\begin{aligned} X^{\mu}{}_{\|\alpha\beta} - X^{\mu}{}_{\|\beta\alpha} &= \quad X^{\nu} R^{\mu}{}_{\nu\alpha\beta} - X^{\mu}{}_{,\nu} T^{\nu}{}_{\alpha\beta}, \\ X_{\mu\|\alpha\beta} - X_{\mu\|\beta\alpha} &= -X_{\nu} R^{\nu}{}_{\mu\alpha\beta} - X_{\mu,\nu} T^{\nu}{}_{\alpha\beta}, \end{aligned} \tag{6.55}$$

We prove the first one. Let $X = X^{\mu} e_{\mu}$. We have

$$\begin{aligned} \nabla X &= \nabla_{\beta} X \otimes \theta^{\beta}, \\ \nabla^2 X &= \nabla(\nabla_{\beta} X \otimes \theta^{\beta}), \\ &= \nabla(\nabla_{\beta} X) \otimes \theta^{\beta} + \nabla_{\beta} X \otimes \nabla \theta^{\beta}, \\ &= \nabla_{\alpha} \nabla_{\beta} X \otimes \theta^{\beta} \otimes \theta^{\alpha} - \nabla_{\beta} \otimes \omega^{\beta}{}_{\alpha} \otimes \theta^{\alpha}, \\ &= \nabla_{\alpha} \nabla_{\beta} X \otimes \theta^{\beta} \otimes \theta^{\alpha} - \nabla_{\mu} X \otimes \Gamma^{\mu}_{\alpha\beta} \theta^{\beta} \otimes \theta^{\alpha}, \\ \nabla^2 X &= (\nabla_{\alpha} \nabla_{\beta} X - \nabla_{\mu} X \, \Gamma^{\mu}_{\alpha\beta})\, \theta^{\beta} \otimes \theta^{\alpha}. \end{aligned}$$

On the other hand, we also have $\nabla X = \nabla_{\alpha} X \otimes \theta^{\alpha}$, so we can compute ∇^2 by differentiating in the reverse order to get the equivalent expression,

$$\nabla^2 X = (\nabla_{\beta} \nabla_{\alpha} X - \nabla_{\mu} X \, \Gamma^{\mu}_{\beta\alpha})\, \theta^{\alpha} \otimes \theta^{\beta}.$$

Subtracting the last two equations we get an alternating tensor, or a two-form that we can set equal to zero. For lack of a better notation we call this form $[\nabla, \nabla]$. The notations $Alt(\nabla^2)$ and $\nabla \wedge \nabla$ also appear in the literature. We get

$$\begin{aligned} [\nabla, \nabla] &= [\nabla_{\alpha} \nabla_{\beta} - \nabla_{\beta} \nabla_{\alpha}) X - \nabla_{\mu} X (\Gamma^{\mu}_{\alpha\beta} - \Gamma^{\mu}_{\beta\alpha})] \theta^{\beta} \wedge \theta^{\alpha}, \\ &= [\nabla_{\alpha} \nabla_{\beta} - \nabla_{\beta} \nabla_{\alpha} - \nabla_{[e_{\alpha}, \beta]}) X + \nabla_{[e_{\alpha}, \beta]} X - \nabla_{\mu} X (\Gamma^{\mu}_{\alpha\beta} - \Gamma^{\mu}_{\beta\alpha})] \theta^{\beta} \wedge \theta^{\alpha}, \\ &= [R(e_{\alpha}, e_{\beta}) X + C^{\mu}_{\alpha\beta} \nabla_{\mu} X - \nabla_{\mu} X (\Gamma^{\mu}_{\alpha\beta} - \Gamma^{\mu}_{\beta\alpha}] \theta^{\beta} \wedge \theta^{\alpha}, \\ &= [R(e_{\alpha}, e_{\beta}) X + C^{\mu}_{\alpha\beta} - \nabla_{\mu} X (\Gamma^{\mu}_{\alpha\beta} - \Gamma^{\mu}_{\beta\alpha} - C^{\mu}_{\alpha\beta}] \theta^{\beta} \wedge \theta^{\alpha}, \\ &= \tfrac{1}{2} (X^{\nu} R^{\mu}{}_{\nu\alpha\beta} - \nabla_{\mu} X \, T^{\mu}{}_{\alpha\beta}) \theta^{\beta} \wedge \theta^{\alpha}. \end{aligned}$$

6.4.3 Exterior Covariant Derivative

Since we know how to take the covariant differential of the basis vectors, covectors, and tensor products thereof, an affine connection on the tangent bundle induces a covariant differential on the tensor bundle. It is easy to get a formula by induction for the covariant differential of a tensor-valued 0-form. A given connection can be extended in a unique way to to tensor-valued p-forms. Just as with the wedge product of a 0-form f with a p-form α for which identify $f\alpha$ with $f \otimes \alpha = f \wedge \alpha$, we write a tensor-valued p form as $T \otimes \alpha = T \wedge \alpha$, where T is a type $\binom{r}{s}$ tensor. We define the *exterior covariant derivative*

$$D : \Omega^{p}(M, E) \to \Omega^{p+1}(M, E)$$

by requiring that,

$$\begin{aligned} D(T\otimes\alpha) &= D(T\wedge\alpha), \\ &= \nabla T\wedge\alpha + (-1)^p T\wedge d\alpha. \end{aligned} \tag{6.56}$$

it is instructive to show the details of the computation of the exterior covariant derivative of the vector valued one forms θ and Θ, and the $\binom{1}{1}$ tensor-valued 2-form Ω. The results are

$$\begin{aligned} D\theta^\alpha &= d\theta^\alpha + \omega^\alpha{}_\beta\wedge\theta^\beta, \\ D\Theta^\alpha &= d\Theta^\alpha + \omega^\alpha{}_\beta\wedge\Theta^\beta, \\ D\Omega^\alpha{}_\beta &= d\Omega^\alpha{}_\beta + \omega^\alpha{}_\gamma\wedge\Omega^\alpha{}_\gamma - \Omega^\alpha{}_\gamma\wedge\omega^\gamma{}_\beta \end{aligned} \tag{6.57}$$

The first two follow immediately, we compute the third. We start by writing

$$\begin{aligned} \Omega &= \Omega^\alpha{}_\beta\; e_\alpha\otimes\theta^\beta, \\ &= (e_\alpha\otimes\theta^\beta)\wedge\Omega^\alpha{}_\beta. \end{aligned}$$

Then,

$$\begin{aligned} D\Omega &= D(e_\alpha\otimes\theta^\beta)\wedge\Omega^\alpha{}_\beta + (-1)^2(e_\alpha\otimes\theta^\beta)\wedge d\Omega^\alpha{}_\beta, \\ &= (De_\alpha\otimes\theta^\beta + e_\alpha\otimes D\theta^\beta)\wedge\Omega^\alpha{}_\beta + (e_\alpha\otimes\theta^\beta)\wedge d\Omega^\alpha{}_\beta, \\ &= (e_\gamma\otimes\omega^\gamma{}_\theta\otimes\theta^\beta + e_\alpha\otimes\omega^\beta{}_\gamma\otimes\theta^\gamma)\wedge)\Omega^\alpha{}_\beta + (e_\alpha\otimes\theta^\beta)\wedge d\Omega^\alpha{}_\beta, \\ &= (e_\alpha\otimes\theta^\beta)\wedge(d\Omega^\alpha{}_\beta + \omega^\alpha{}_\gamma\wedge\Omega^\gamma{}_\beta - \omega^\alpha{}_\gamma\wedge\Omega^\gamma{}_\beta). \end{aligned}$$

In the last step we had to relabel a couple of indices so that we could factor out $(e_\alpha\otimes\theta^\beta)$. The pattern should be clear. We get an exterior derivative for the forms, an $\omega\wedge\Omega$ term for the contravariant index and an $\Omega\wedge\omega$ term with the appropriate sign, for the covariant index. Here the computation gives

$$\begin{aligned} D\Omega^\alpha{}_\beta &= d\Omega^\alpha{}_\beta + \omega^\alpha{}_\gamma\wedge\Omega^\gamma{}_\beta - \omega^\alpha{}_\gamma\wedge\Omega^\gamma{}_\beta, \quad \text{or} \\ D\Omega &= d\omega + \omega\wedge\Omega - \Omega\wedge\omega. \end{aligned} \tag{6.58}$$

This means that we can write the equations of structure as

$$\begin{aligned} \Theta &= D\theta, \\ \Omega &= d\omega + \omega\wedge\omega, \end{aligned} \tag{6.59}$$

and the Bianchi's identities as

$$\begin{aligned} D\Theta &= \Omega\wedge\theta, \\ D\Omega &= 0 \end{aligned} \tag{6.60}$$

With apologies for the redundancy, we reproduce the change of basis formula 3.49. Let $e' = eB$ be an orthogonal change of basis. Then

$$\begin{aligned} De' &= e\otimes dB + DeB, \\ &= e\otimes dB + (e\otimes\omega)B, \\ &= e'\otimes(B^{-1}dB + B^{-1}\omega B), \\ &= e'\otimes\omega', \end{aligned}$$

where,

$$\omega' = B^{-1}dB + B^{-1}\omega B. \tag{6.61}$$

Multiply the last equation by B and take the exterior derivative d. We get.

$$\begin{aligned} B\omega' &= dB + \omega B, \\ Bd\omega' + dB \wedge \omega' &= d\omega B - \omega \wedge dB, \\ Bd\omega' + (B\omega' - \omega B) \wedge \omega' &= d\omega B - \omega \wedge (\omega' B - \omega B), \\ B(d\omega' + \omega' \wedge \omega') &= (d\omega + \omega \wedge \omega)B, \end{aligned}$$

Setting $\Omega = d\omega + \omega \wedge \omega$, and $\Omega' = d\omega' + \omega' \wedge \omega'$, the last equation reads,

$$\Omega' = B^{-1}\Omega B. \tag{6.62}$$

As pointed out after equation 3.49, the curvature is a tensorial form of adjoint type. The transformation law above for the connection has an extra term, so it is not tensorial. It is easy to obtain the classical transformation law for the Christoffel symbols from equation 6.61. Let $\{x^\alpha\}$ be coordinates in a patch (ϕ_α, U_α), and $\{y^\beta\}$ be coordinates on a overlapping patch (ϕ_β, U_β). The transition functions $\phi_{\alpha\beta}$ are given by the Jacobian of the change of coordinates,

$$\begin{aligned} \frac{\partial}{\partial y^\beta} &= \frac{\partial x^\alpha}{\partial y^\beta}\frac{\partial}{\partial x^\alpha}, \\ \phi_{\alpha\beta} &= \frac{\partial x^\alpha}{\partial y^\beta}. \end{aligned}$$

Inserting the connection components ${\omega'^\alpha}_\beta = \Gamma'^\alpha_{\beta\gamma}dy^\gamma$, into the change of basis formula 6.61, with $B = \phi_{\alpha\beta}$, we get[1],

$$\begin{aligned} {\omega'^\alpha}_\beta &= {(B^{-1})^\alpha}_\kappa d{B^\kappa}_\beta + {(B^{-1})^\alpha}_\kappa \, {\omega^\kappa}_\lambda {B^\lambda}_\beta, \\ &= \frac{\partial y^\alpha}{\partial x^\kappa} d\left(\frac{\partial x^\kappa}{\partial y^\beta}\right) + \frac{\partial y^\alpha}{\partial x^\kappa}\, {\omega^\kappa}_\lambda \, \frac{\partial x^\lambda}{\partial y^\beta}, \\ \Gamma'^\alpha_{\beta\gamma} dy^\gamma &= \frac{\partial y^\alpha}{\partial x^\kappa}\frac{\partial^2 x^\kappa}{\partial y^\sigma \partial y^\beta} dy^\sigma + \frac{\partial y^\alpha}{\partial x^\kappa}\, \Gamma^\kappa_{\lambda\sigma} dx^\sigma \, \frac{\partial x^\lambda}{\partial y^\beta}, \\ \Gamma'^\alpha_{\beta\gamma} &= \frac{\partial y^\alpha}{\partial x^\kappa}\frac{\partial^2 x^\kappa}{\partial y^\gamma \partial y^\beta} + \frac{\partial y^\alpha}{\partial x^\kappa}\, \Gamma^\kappa_{\lambda\sigma} \frac{\partial x^\sigma}{\partial y^\gamma}\, \frac{\partial x^\lambda}{\partial y^\beta}. \end{aligned}$$

Thus, we retrieve the classical transformation law for Christoffel symbols that one finds in texts on general relativity.

$$\Gamma'^\alpha_{\beta\gamma} = \Gamma^\kappa_{\lambda\sigma} \frac{\partial y^\alpha}{\partial x^\kappa}\, \frac{\partial x^\sigma}{\partial y^\gamma}\, \frac{\partial x^\lambda}{\partial y^\beta} + \frac{\partial y^\alpha}{\partial x^\kappa}\frac{\partial^2 x^\kappa}{\partial y^\gamma \partial y^\beta}. \tag{6.63}$$

[1] We use this notation reluctantly, to be consistent with most literature. The notation results in violation of the index notation. We really should be writing ${\phi^\alpha}_\beta$, since in this case, the transition functions are matrix-valued.

6.4.4 Parallelism

When first introduced to vectors in elementary calculus and physics courses, vectors are often described as entities characterized by a direction and a length. This primitive notion, that two such entities in $\mathbf{R}^n$ with the same direction and length represent the same vector, regardless of location, is not erroneous in the sense that parallel translation of a vector in $\mathbf{R}^n$ does not change the attributes of a vector as described. In elementary linear algebra, vectors are described as n-tuples in $\mathbf{R}^n$ equipped with the operations of addition and multiplication by scalar, and subject to eight vector space properties. Again, those vectors can be represented by arrows which can be located anywhere in $\mathbf{R}^n$ as long as they have the same components. This is another indication that parallel transport of a vector in $\mathbf{R}^n$ is trivial, a manifestation of the fact the $\mathbf{R}^n$ is a flat space. However, in a space that is not flat, such as s sphere, parallel transport of vectors is intimately connected with the curvature of the space. To elucidate this connection, we first describe parallel transport for a surface in $\mathbf{R}^3$.

6.4.2 Definition Let $u^\alpha(t)$ be a curve on a surface $\mathbf{x} = \mathbf{x}(u^\alpha)$, and let $V = \alpha'(t) = \alpha_*(\frac{d}{dt})$ be the velocity vector as defined in 1.25. A vector field Y is called *parallel* along α if

$$\nabla_V Y = 0,$$

as illustrated in figure 6.2. The notation

$$\frac{DY}{dt} = \nabla_V Y$$

is also common in the literature. The vector field $\nabla_V V$ is called the *geodesic vector field*, and its magnitude is called the geodesic curvature κ_g of α. As usual, we define the speed v of the curve by $\|V\|$ and the unit tangent $T = V/\|V\|$, so that $V = vT$. We assume $v > 0$ so that T is defined on the domain of the curve. The arc length s along the curve is the related to the speed by the equation $v = ds/dt$.

6.4.3 Definition A curve $\alpha(t)$ with velocity vector $V = \alpha'(t)$ is called a *geodesic* or *self-parallel* if $\nabla_V V = 0$.

6.4.4 Theorem A curve $\alpha(t)$ is geodesic iff
a) $v = \|V\|$ is constant along the curve and,
b) either $\nabla_T T = 0$, or $\kappa_g = 0$.

Proof Expanding the definition of the geodesic vector field:

$$\begin{aligned}\nabla_V V &= \nabla_{vT}(vT), \\ &= v\nabla_T(vT), \\ &= v\frac{dv}{dt}T + v^2\nabla_T T, \\ &= \tfrac{1}{2}\frac{d}{dt}(v^2)T + v^2\nabla_T T\end{aligned}$$

We have $< T, T >= 1$, so $< \nabla_T T, T >= 0$ which shows that $\nabla_T T$ is orthogonal to T. We also have $v > 0$. Since both the tangential and the normal components need to vanish, the theorem follows.

If M is a hypersurface in $\mathbf{R}^n$ with unit normal $\mathbf{n}$, we gain more insight on the geometry of geodesics as a direct consequence of the discussion above. Without real loss of generality consider the geometry in the case of $n = 3$. Since α is geodesic, we have $\|\alpha'\|^2 =< \alpha', \alpha' >=$constant. Differentiation gives $< \alpha', \alpha'' >= 0$, so that the acceleration α'' is orthogonal to α'. Comparing with equation 4.34 we see that $T' = \kappa_n \mathbf{n}$, which reinforces the fact that the entire curvature of the curve is due to the normal curvature of the surface as a submanifold of the ambient space. In this sense, inhabitants constrained to live on the surface would be unaware of this curvature, and to them, geodesics would appear locally as the straightest path to travel. Thus, for a sphere in $\mathbf{R}^3$ of radius a, the acceleration α'' of a geodesic only has a normal component, and the normal curvature is $1/a$. That is, the geodesic must lie along a great circle.

6.4.5 Theorem Let $\alpha(t)$ by curve with velocity V. For each vector Y in the tangent space restricted to the curve, there is a unique vector field $Y(t)$ locally obtained by parallel transport.

Proof We choose local coordinates with frame field $\{e_\alpha = \frac{\partial}{\partial u^\alpha}\}$. We write the components of the vector fields in terms of the frame

$$\begin{aligned} Y &= y^\beta \frac{\partial}{\partial u^\beta}, \\ V &= \frac{du^\alpha}{dt}\frac{\partial}{\partial u^\alpha}. \quad \text{then,} \\ \nabla_T V &= \nabla_{\dot{u^\alpha} e_\alpha}(y^\beta e_\beta), \\ &= \dot{u^\alpha} \nabla_{e^\alpha}(y^\beta e_\beta), \\ &= \frac{du^\alpha}{dt}\frac{\partial y^\beta}{\partial u^\alpha} + \dot{u^\alpha} y^\beta \Gamma^\gamma{}_{\alpha\beta} e_\gamma, \\ &= [\frac{dy^\gamma}{dt} + y^\beta \frac{du^\alpha}{dt}\Gamma^\gamma{}_{\alpha\beta}] e_\gamma. \end{aligned}$$

So, Y is parallel along the curve iff,

$$\frac{dy^\gamma}{dt} + y^\beta \frac{du^\alpha}{dt}\Gamma^\gamma{}_{\alpha\beta} = 0. \tag{6.64}$$

The existence and uniqueness of the coefficients y^β that define Y are guaranteed by the theorem on existence and uniqueness of differential equations with appropriate initial conditions.

We derive the equations of geodesics by an almost identical computation.

$$
\begin{aligned}
\nabla_V V &= \nabla_{\dot{u}^\alpha e_\alpha}[\dot{u}^\beta e_\beta], \\
&= \dot{u}^\alpha \nabla_{e_\alpha}[\dot{u}^\beta e_\beta], \\
&= \dot{u}^\alpha [\frac{\partial \dot{u}^\beta}{\partial u^\alpha} e_\beta + \dot{u}^\beta \nabla_{e_\alpha} e_\beta], \\
&= \dot{u}^\alpha \frac{\partial \dot{u}^\beta}{\partial u^\alpha} e_\beta + \dot{u}^\alpha \dot{u}^\beta \nabla_{e_\alpha} e_\beta, \\
&= \frac{du^\alpha}{dt} \frac{\partial \dot{u}^\beta}{\partial u^\alpha} e_\beta + \dot{u}^\alpha \dot{u}^\beta \Gamma^\sigma_{\alpha\beta} e_\sigma, \\
&= \ddot{u}^\beta e_\beta + \dot{u}^\alpha \dot{u}^\beta \Gamma^\sigma_{\alpha\beta} e_\sigma, \\
&= [\ddot{u}^\sigma + \dot{u}^\alpha \dot{u}^\beta \Gamma^\sigma_{\alpha\beta}] e_\sigma.
\end{aligned}
$$

Thus, the equation for geodesics becomes

$$\ddot{u}^\sigma + \Gamma^\sigma_{\alpha\beta} \dot{u}^\alpha \dot{u}^\beta = 0. \tag{6.65}$$

The existence and uniqueness theorem for solutions of differential equations leads to the following theorem

6.4.6 Theorem Let p be a point in M and V a vector T_pM. Then, for any real number t_0, there exists a number δ and a curve $\alpha(t)$ defined on $[t_0-\delta, t_0+\delta]$, such that $\alpha(t_0) = p$, $\alpha'(t_0) = V$, and α is a geodesic.

For a general vector bundles E over a manifold M, a section $s \in \Gamma(E)$ of a vector bundle is called a *parallel section* if

$$\nabla s = 0. \tag{6.66}$$

We discuss the length minimizing properties geodesics in section 6.6 and provide a number of examples for surfaces in $\mathbf{R}^3$ and for Lorentzian manifolds. Since geodesic curves have zero acceleration, in Euclidean space they are straight lines. In Einstein's theory of relativity, gravitation is a fictitious force caused by the curvature of space time, so geodesics represent the trajectory of free particles.

6.5 Lorentzian Manifolds

The formalism above refers to Riemannian manifolds, for which the metric is positive definite, but it applies just as well to pseudo-Riemannian manifolds. A 4-dimensional manifold $\{M, g\}$ is called a *Lorentzian manifold* if the metric has signature $(+ - --)$. Locally, a Lorentzian manifold is diffeomorphic to Minkowski's space which is the model space introduced in section 2.2.3. Some authors use signature $(- + ++)$.

For the purposes of general relativity, we introduce the symmetric tensor *Ricci tensor* $R_{\beta\delta}$ by the contraction

$$R_{\beta\delta} = R^\alpha{}_{\beta\alpha\delta}, \tag{6.67}$$

and the *scalar curvature* R by

$$R = R^{\alpha}{}_{\beta}. \tag{6.68}$$

The traceless part of the Ricci tensor

$$G_{\alpha\beta} = R_{\alpha\beta} - \tfrac{1}{2} R g_{\alpha\beta}, \tag{6.69}$$

is called the *Einstein tensor.* The Einstein field equations (without a cosmological constant) are

$$G_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta}, \tag{6.70}$$

where T is the *stress energy tensor* and G is the gravitational constant. As I first learned from one of my professors Arthur Fischer, the equation states that curvature indicates the presence of matter, and matter tells the space how to curve. Einstein equations with cosmological constant Λ are,

$$R_{\alpha\beta} - \tfrac{1}{2} R g_{\alpha\beta} + \Lambda g_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta} \tag{6.71}$$

Fig. 6.3: Gravity

A space time which satisfies

$$R_{\alpha\beta} = 0 \tag{6.72}$$

is called *Ricci-flat.* A space which the Ricci tensor is proportional to the metric,

$$R_{\alpha\beta} = k g_{\alpha\beta} \tag{6.73}$$

is called an *Einstein manifold*

6.5.1 Example: Vaidya Metric

This example of a curvature computation in four-dimensional space-time is due to W. Israel. It appears in his 1978 notes on Differential Forms in General Relativity, but the author indicates the work arose 10 years earlier from a seminar at the Dublin Institute for Advanced Studies. The most general, spherically symmetric, static solution of the Einstein vacuum equations is the Schwarzschild metric [2]

$$ds^2 = \left[1 - \tfrac{2GM}{r}\right] dt^2 - \frac{1}{\left[1 - \tfrac{2GM}{r}\right]} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2 \tag{6.74}$$

It is convenient to set $m = GM$ and introduce the retarded coordinate transformation

$$t = u + r + 2m \ln(\tfrac{r}{2m} - 1),$$

so that,

$$dt = du + \frac{1}{\left[1 - \tfrac{2m}{r}\right]} dr.$$

Substitution for dt above gives the metric in outgoing *Eddington-Finkelstein* coordinates,

$$ds^2 = 2drdu + [1 - \tfrac{2m}{r}]\, du^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2. \tag{6.75}$$

In these coordinates it is evident that the event horizon $r = 2m$ is not a real singularity. The Vaidya metric is the generalization

$$ds^2 = 2drdu + [1 - \tfrac{2m(u)}{r}]\, du^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2, \tag{6.76}$$

where $m(u)$ is now an arbitrary function. The geometry described by the Vaidya solution to Einstein equations, represents the gravitational field in the exterior of a radiating, spherically symmetric star. In all our previous curvature computations by differential forms, the metric has been diagonal; this is an instructive example of one with a non-diagonal metric. The first step in the curvature computation involves picking out a basis of one-forms. The idea is to pick out the forms so that in the new basis, the metric has constant coefficients. One possible choice of 1-forms is

$$\begin{aligned}
\theta^0 &= du, \\
\theta^1 &= dr + \tfrac{1}{2}[1 - \tfrac{2m(u)}{r}]\, du, \\
\theta^2 &= r\, d\theta, \\
\theta^3 &= r \sin\theta \, d\phi.
\end{aligned} \tag{6.77}$$

In terms of these forms, the line element becomes

$$ds^2 = g_{\alpha\beta}\theta^\alpha\theta^\beta = 2\theta^0\theta^1 - (\theta^2)^2 - (\theta^3)^2,$$

where

$$g_{01} = g_{10} = -g_{22} = -g_{33} = 1,$$

while all the other $g_{\alpha\beta} = 0$. In the coframe, the metric has components:

$$g_{\alpha\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{6.78}$$

[2]The Schwarzschild radius is $r = \frac{2GM}{c^2}$, but here we follow the common convention of setting $c = 1$.

Since the coefficients of the metric are constant, the components $\omega_{\alpha\beta}$ of the connection will be antisymmetric. This means that

$$\omega_{00} = \omega_{11} = \omega_{22} = \omega_{33} = 0.$$

We thus conclude that

$$\begin{aligned}
{\omega^1}_0 &= g^{10}\omega_{00} = 0,\\
{\omega^0}_1 &= g^{01}\omega_{11} = 0,\\
{\omega^2}_2 &= g^{22}\omega_{22} = 0,\\
{\omega^3}_3 &= g^{33}\omega_{33} = 0.
\end{aligned}$$

To compute the connection, we take the exterior derivative of the basis 1-forms. The result of this computation is

$$\begin{aligned}
d\theta^0 &= 0,\\
d\theta^1 &= -d[\tfrac{m}{r}du] = \frac{m}{r^2}\,dr\wedge du = \frac{m}{r^2}\,\theta^1\wedge\theta^0,\\
d\theta^2 &= dr\wedge d\theta = \frac{1}{r}\theta^1\wedge\theta^2 - \frac{1}{2r}[1-\tfrac{2m}{r}]\,\theta^0\wedge\theta^2,\\
d\theta^3 &= \sin\theta\,dr\wedge d\phi + r\cos\theta\,d\theta\wedge d\phi,\\
&= \frac{1}{r}\theta^1\wedge\theta^3 - \frac{1}{2}[1-\tfrac{2m}{r}]\,\theta^0\wedge\theta^3 + \frac{1}{r}\cot\theta\,\theta^2\wedge\theta^3.
\end{aligned}\tag{6.79}$$

For convenience, we write below the first equation of structure [6.24] in complete detail.

$$\begin{aligned}
d\theta^0 &= {\omega^0}_0\wedge\theta^0 + {\omega^0}_1\wedge\theta^1 + {\omega^0}_2\wedge\theta^2 + {\omega^0}_3\wedge\theta^3,\\
d\theta^1 &= {\omega^1}_0\wedge\theta^0 + {\omega^1}_1\wedge\theta^1 + {\omega^1}_2\wedge\theta^2 + {\omega^1}_3\wedge\theta^3,\\
d\theta^2 &= {\omega^2}_0\wedge\theta^0 + {\omega^2}_1\wedge\theta^1 + {\omega^2}_2\wedge\theta^2 + {\omega^2}_3\wedge\theta^3,\\
d\theta^3 &= {\omega^3}_0\wedge\theta^0 + {\omega^3}_1\wedge\theta^1 + {\omega^3}_2\wedge\theta^2 + {\omega^3}_3\wedge\theta^3.
\end{aligned}\tag{6.80}$$

Since the ω's are one-forms, they must be linear combinations of the θ's. Comparing Cartan's first structural equation with the exterior derivatives of the coframe, we can start with the initial guess for the connection coefficients below:

$$\begin{array}{llll}
{\omega^1}_0 = 0, & {\omega^1}_1 = \dfrac{m}{r^2}\,\theta^0, & {\omega^1}_2 = A\,\theta^2, & {\omega^1}_3 = B\,\theta^3,\\
{\omega^2}_0 = -\dfrac{1}{2}[1-\dfrac{2m}{r}]\,\theta^2, & {\omega^2}_1 = \dfrac{1}{r}\,\theta^2, & {\omega^2}_2 = 0, & {\omega^2}_3 = C\,\theta^3,\\
{\omega^3}_0 = -\dfrac{1}{2}[1-\dfrac{2m}{r}]\,\theta^3, & {\omega^3}_1 = \dfrac{1}{r}\,\theta^3, & {\omega^3}_2 = \dfrac{1}{r}\cot\theta\,\theta^3, & {\omega^3}_3 = 0.
\end{array}$$

Here, the quantities A, B, and C are unknowns to be determined. Observe that these are not the most general choices for the ω's. For example, we could have added a term proportional to θ^1 in the expression for ${\omega^1}_1$, without affecting the validity of the first structure equation for $d\theta^1$. The strategy is to interactively tweak the expressions until we set of forms completely consistent with Cartan's structure equations.

We now take advantage of the skewsymmetry of $\omega_{\alpha\beta}$, to determine the other components. The find A, B and C, we note that

$$
\begin{aligned}
{\omega^1}_2 &= g^{10}\omega_{02} = -\omega_{20} = {\omega^2}_0, \\
{\omega^1}_3 &= g^{10}\omega_{03} = -\omega_{30} = {\omega^3}_0, \\
{\omega^2}_3 &= g^{22}\omega_{23} = \omega_{32} = -{\omega^3}_2.
\end{aligned}
$$

Comparing the structure equations 6.80 with the expressions for the connection coefficients above, we find that

$$
A = -\frac{1}{2}[1 - \frac{2m}{r}], \quad B = -\frac{1}{2}[1 - \frac{2m}{r}], \quad C = -\frac{1}{r}\cot\theta. \tag{6.81}
$$

Similarly, we have

$$
\begin{aligned}
{\omega^0}_0 &= -{\omega^1}_1, \\
{\omega^0}_2 &= {\omega^2}_1, \\
{\omega^0}_3 &= {\omega^3}_1,
\end{aligned}
$$

hence,

$$
\begin{aligned}
{\omega^0}_0 &= -\frac{m}{r^2}\theta^0, \\
{\omega^0}_2 &= -\frac{1}{r}\theta^2, \\
{\omega^0}_3 &= \frac{1}{r}\theta^3.
\end{aligned}
$$

It is easy to verify that our choices for the ω's are consistent with first structure equations, so by uniqueness, these must be the right values.

There is no guesswork in obtaining the curvature forms. All we do is take the exterior derivative of the connection forms and pick out the components of the curvature from the second Cartan equations [6.25]. Thus, for example, to obtain ${\Omega^1}_1$, we proceed as follows.

$$
\begin{aligned}
{\Omega^1}_1 &= d{\omega^1}_1 + {\omega^1}_1 \wedge {\omega^1}_1 + {\omega^1}_2 \wedge {\omega^2}_1 + {\omega^1}_3 \wedge {\omega^3}_1, \\
&= d[\frac{m}{r^2}\theta^0] + 0 - \frac{1}{2r^2}[1 - \frac{2m}{r}]\ {\omega^1}_3 \wedge {\omega^3}_1 + (\theta^2 \wedge \theta^2 + \theta^3 \wedge \theta^3), \\
&= -\frac{2m}{r^3} dr \wedge \theta^0, \\
&= -\frac{2m}{r^3}\theta^1 \wedge \theta^0.
\end{aligned}
$$

The computation of the other components is straightforward and we just present the results.

$$
{\Omega^1}_2 = -\frac{1}{r^2}\frac{dm}{du}\theta^2 \wedge \theta^0 - \frac{m}{r^3}\theta^1 \wedge \theta^2,
$$

$$\Omega^1{}_3 = -\frac{1}{r^2}\frac{dm}{du}\theta^3 \wedge \theta^0 - \frac{m}{r^3}\theta^1 \wedge \theta^3,$$
$$\Omega^2{}_1 = \frac{m}{r^3}\theta^2 \wedge \theta^0,$$
$$\Omega^3{}_1 = \frac{m}{r^3}\theta^3 \wedge \theta^0,$$
$$\Omega^2{}_3 = \frac{2m}{r^3}\theta^2 \wedge \theta^3.$$

By antisymmetry, these are the only independent components. We can also read the components of the full Riemann curvature tensor from the definition

$$\Omega^\alpha{}_\beta = \frac{1}{2}R^\alpha{}_{\beta\gamma\delta}\theta^\gamma \wedge \theta^\delta. \tag{6.82}$$

Thus, for example, we have

$$\Omega^1{}_1 = \frac{1}{2}R^1{}_{1\gamma\delta}\theta^\gamma \wedge \theta^\delta,$$

hence

$$R^1{}_{101} = -R^1{}_{110} = \frac{2m}{r^3};\ \text{other}\ \ R^1{}_{1\gamma\delta} = 0.$$

Using the antisymmetry of the curvature forms, we see, that for the Vaidya metric $\Omega^1{}_0 = \Omega_{00} = 0$, $\Omega^2{}_0 = -\Omega^1{}_2$, etc., so that

$$\begin{aligned} R_{00} &= R^2{}_{020} + R^3{}_{030} \\ &= R^1{}_{220} + R^1{}_{330} \end{aligned}$$

Substituting the relevant components of the curvature tensor, we find that

$$R_{00} = 2\frac{1}{r^2}\frac{dm}{du} \tag{6.83}$$

while all the other components of the Ricci tensor vanish. As stated earlier, if m is constant, we get the Ricci flat Schwarzschild metric.

6.6 Geodesics

Geodesics were introduced in the section on parallelism. The equation of geodesics on a manifold given by equation 6.65 involves the Christoffel symbols. Whereas it is possible to compute all the Christoffel symbols starting with the metric as in equation 4.76, this is most inefficient, as it is often the case that many of the Christoffel symbols vanish. Instead, we show next how to obtain the geodesic equations by using variational principles

$$\delta \int L(u^\alpha, \dot{u}^\alpha, s)\, ds = 0, \tag{6.84}$$

to minimize the arc length. Then we can pick out the non-vanishing Christoffel symbols from the geodesic equation. Following the standard methods of Lagrangian mechanics, we let u^α and $\dot{u}^\alpha$ be treated as independent (canonical) coordinates and choose the Lagrangian in this case to be

$$L = g_{\alpha\beta}\dot{u}^\alpha\dot{u}^\beta. \tag{6.85}$$

The choice will actually result in minimizing the square of the arc length, but clearly this is an equivalent problem. It should be observed that the Lagrangian is basically a multiple of the kinetic energy $\frac{1}{2}mv^2$. The motion dynamics are given by the Euler-Lagrange equations.

$$\frac{d}{ds}\left(\frac{\partial L}{\partial \dot{u}^\gamma}\right) - \frac{\partial L}{\partial u^\gamma} = 0. \tag{6.86}$$

Applying this equations keeping in mind that $g_{\alpha\beta}$ is the only quantity that depends on u^α, we get:

$$\begin{aligned}
0 &= \tfrac{d}{ds}[g_{\alpha\beta}\delta^\alpha_\gamma\dot{u}^\beta + g_{\alpha\beta}\dot{u}^\alpha\delta^\beta_\gamma] - g_{\alpha\beta,\gamma}\dot{u}^\alpha\dot{u}^\beta \\
&= \tfrac{d}{ds}[g_{\gamma\beta}\dot{u}^\beta + g_{\alpha\gamma}\dot{u}^\alpha] - g_{\alpha\beta,\gamma}\dot{u}^\alpha\dot{u}^\beta \\
&= g_{\gamma\beta}\ddot{u}^\beta + g_{\alpha\gamma}\ddot{u}^\alpha + g_{\gamma\beta,\alpha}\dot{u}^\alpha\dot{u}^\beta + g_{\alpha\gamma,\beta}\dot{u}^\beta\dot{u}^\alpha - g_{\alpha\beta,\gamma}\dot{u}^\alpha\dot{u}^\beta \\
&= 2g_{\gamma\beta}\ddot{u}^\beta + [g_{\gamma\beta,\alpha} + g_{\alpha\gamma,\beta} - g_{\alpha\beta,\gamma}]\dot{u}^\alpha\dot{u}^\beta \\
&= \delta^\sigma_\beta\ddot{u}^\beta + \tfrac{1}{2}g^{\gamma\sigma}[g_{\gamma\beta,\alpha} + g_{\alpha\gamma,\beta} - g_{\alpha\beta,\gamma}]\dot{u}^\alpha\dot{u}^\beta
\end{aligned}$$

where the last equation was obtained contracting with $\frac{1}{2}g^{\gamma\sigma}$ to raise indices. Comparing with the expression for the Christoffel symbols found in equation 4.76, we get

$$\ddot{u}^\sigma + \Gamma^\sigma_{\alpha\beta}\dot{u}^\alpha\dot{u}^\beta = 0$$

which are exactly the equations of geodesics 6.65.

6.6.1 Example Geodesics of sphere

Let S^2 be a sphere of radius a so that the metric is given by

$$ds^2 = a^2 d\theta^2 + a^2\sin^2\theta\, d\phi^2.$$

Then the Lagrangian is

$$L = a^2\dot{\theta}^2 + a^2\sin^2\theta\,\dot{\phi}^2.$$

The Euler-Lagrange equation for the ϕ coordinate is

$$\begin{aligned}
\frac{d}{ds}\left(\frac{\partial L}{\partial \dot{\phi}}\right) - \frac{\partial L}{\partial \phi} &= 0, \\
\frac{d}{ds}(2a^2\sin^2\theta\dot{\phi}) &= 0,
\end{aligned}$$

and therefore the equation integrates to a constant

$$\sin^2\theta\,\dot{\phi} = k.$$

Rather than trying to solve the second Euler-Lagrange equation for θ, we evoke a standard trick that involves reusing the metric. It goes as follows:

$$\begin{aligned}
\sin^2\theta\,\frac{d\phi}{ds} &= k,\\
\sin^2\theta\; d\phi &= k\;ds,\\
\sin^4\theta\; d\phi^2 &= k^2 ds^2,\\
\sin^4\theta\; d\phi^2 &= k^2(a^2 d\theta^2 + a^2\sin^2\theta\; d\phi^2),\\
(\sin^4\theta - k^2a^2\sin^2\theta)\,d\phi^2 &= a^2k^2 d\theta^2.
\end{aligned}$$

The last equation above is separable and it can be integrated using the substitution $u = \cot\theta$.

$$\begin{aligned}
d\phi &= \frac{ak}{\sin\theta\sqrt{\sin^2\theta - a^2k^2}}\,d\theta,\\
&= \frac{ak}{\sin^2\theta\sqrt{1 - a^2k^2\csc^2\theta}}\,d\theta,\\
&= \frac{ak}{\sin^2\theta\sqrt{1 - a^2k^2(1+\cot^2\theta)}}\,d\theta,\\
&= \frac{ak\csc^2\theta}{\sqrt{1 - a^2k^2(1+\cot^2\theta)}}\,d\theta,\\
&= \frac{ak\csc^2\theta}{\sqrt{(1 - a^2k^2) - a^2k^2\cot^2\theta}}\,d\theta,\\
&= \frac{\csc^2\theta}{\sqrt{\frac{1-a^2k^2}{a^2k^2} - \cot^2\theta}}\,d\theta,\\
&= \frac{-1}{\sqrt{c^2-u^2}}\,du, \quad \text{where}(\ c^2 = \tfrac{1-a^2k^2}{a^2k^2}).\\
\phi &= -\sin^{-1}(\tfrac{1}{c}\cot\theta) + \phi_0.
\end{aligned}$$

Here, ϕ_0 is the constant of integration. To get a geometrical sense of the geodesics equations we have just derived, we rewrite the equations as follows:

$$\begin{aligned}
\cot\theta &= c\sin(\phi_0 - \phi),\\
\cos\theta &= c\sin\theta(\sin\phi_0\cos\phi - \cos\phi_0\sin\phi),\\
a\cos\theta &= (c\sin\phi_0)(a\sin\theta\cos\phi) - (c\cos\phi_0)(a\sin\theta\sin\phi.)\\
z &= Ax - By, \quad \text{where } A = c\sin\phi_0,\ B = c\cos\phi_0.
\end{aligned}$$

We conclude that the geodesics of the sphere are great circles determined by the intersections with planes through the origin.

6.6.2 Example Geodesics in orthogonal coordinates.
In a parametrization of a surface in which the coordinate lines are orthogonal, $F = 0$. Then first fundamental form is,

$$ds^2 = E\,du^2 + G\,dv^2,$$

and we have the Lagrangian,

$$L = E\dot{u}^2 + G\dot{v}^2.$$

The Euler-Lagrange equations for the variable u are:

$$\begin{aligned}
\frac{d}{ds}(2E\dot{u}) - E_u\dot{u}^2 - G_u\dot{v}^2 &= 0,\\
2E\ddot{u} + (2E_u\dot{u} + 2E_v\dot{v})\dot{u} - E_u\dot{u}^2 - G_u\dot{v}^2 &= 0,\\
2E\ddot{u} + E_u\dot{u}^2 + 2E_v\dot{u}\dot{v} - G_u\dot{v}^2 &= 0.
\end{aligned}$$

Similarly for the variable v,

$$\begin{aligned}
\frac{d}{ds}(2G\dot{u}) - E_v\dot{u}^2 - G_v\dot{v}^2 &= 0,\\
2G\ddot{v} + (2G_u\dot{u} + 2G_v\dot{v})\dot{v} - E_v\dot{u}^2 - G_v\dot{v}^2 &= 0,\\
2G\ddot{v} - E_v\dot{u}^2 + 2G_u\dot{u}\dot{v} + G_v\dot{v}^2 &= 0.
\end{aligned}$$

So, the equations of geodesics can be written neatly as,

$$\begin{aligned}
\ddot{u} + \frac{1}{2E}[E_u\dot{u}^2 + 2E_v\dot{u}\dot{v} - G_u\dot{v}^2] &= 0,\\
\ddot{v} + \frac{1}{2G}[G_v\dot{v}^2 + 2G_u\dot{u}\dot{v} - E_v\dot{u}^2] &= 0. \qquad (6.87)
\end{aligned}$$

6.6.3 Example Geodesics of surface of revolution

The first fundamental form a surface of revolution $z = f(r)$ in cylindrical coordinates as in 4.7, is

$$ds^2 = (1 + f'^2)\, dr^2 + r^2\, d\phi^2, \qquad (6.88)$$

Of course, we could use the expressions for the equations of geodesics we just derived above, but since the coefficients are functions of r only, it is just a easy to start from the Lagrangian,

$$L = (1 + f'^2)\, \dot{r}^2 + r^2\dot{\phi}^2.$$

Since there is no dependance on ϕ, the Euler-Lagrange equation on ϕ gives rise to a conserved quantity.

$$\begin{aligned}
\frac{d}{ds}(2r^2\dot{\phi}) &= 0,\\
r^2\dot{\phi} &= c \qquad (6.89)
\end{aligned}$$

where c is a constant of integration. If the geodesic $\alpha(s) = \alpha(r(s), \phi(s))$ represents the path of a free particle constrained to move on the surface, this conserved quantity is essentially the angular momentum. A neat result can be obtained by considering the angle σ that the tangent vector $V = \alpha'$ makes with a meridian. Recall that the length of V along the geodesic is constant, so let's set $\|V\| = k$. From the chain rule we have

$$\alpha'(t) = \mathbf{x}_r \frac{dr}{ds} + \mathbf{x}_\phi \frac{d\phi}{ds}.$$

Then

$$\begin{aligned}\cos\sigma &= \frac{< \alpha', \mathbf{x}_\phi >}{\|\alpha'\| \cdot \|\mathbf{x}_\phi\|} = \frac{G\frac{d\phi}{ds}}{k\sqrt{G}}, \\ &= \frac{1}{k}\sqrt{G}\frac{d\phi}{ds} = \frac{1}{k} r\dot{\phi}.\end{aligned}$$

We conclude from 6.89, that for a surface of revolution, the geodesics make an angle σ with meridians that satisfies the equation

$$r\cos\sigma = \text{constant}. \tag{6.90}$$

This result is called *Clairaut's relation.* Writing equation 6.89 in terms of differentials, and reusing the metric as we did in the computation of the geodesics for a sphere, we get

$$\begin{aligned} r^2\, d\phi &= c\, ds, \\ r^4\, d\phi^2 &= c^2\, ds^2, \\ &= c^2[(1+f'^2)\, dr^2 + r^2\, d\phi^2], \\ (r^4 - c^2r^2)\, d\phi^2 &= c^2[(1+f'^2)\, dr^2, \\ r\sqrt{r^2-c^2}\, d\phi &= c\sqrt{1+f'^2}\, dr, \end{aligned}$$

so

$$\phi = \pm c \int \frac{\sqrt{1+f'^2}}{r\sqrt{r^2-c^2}}\, dr. \tag{6.91}$$

If $c = 0$, then the first equation above gives ϕ =constant, so the meridians are geodesics. The parallels r =constant are geodesics when $f'(r) = \infty$ in which case the tangent bundle restricted to the parallel is a cylinder with a vertical generator.

In the particular case of a cone of revolution with a generator that makes an angle α with the z-axis, $f(r) = \cot(\alpha)r$, equation 6.91 becomes:

$$\phi = \pm c \int \frac{\sqrt{1+\cot^2\alpha}}{r\sqrt{r^2-c^2}}\, dr$$

which can be immediately integrated to yield

$$\phi = \pm \csc\alpha \sec^{-1}(r/c) \tag{6.92}$$

As shown in figure 6.4, a ribbon laid flatly around a cone follows the path of a geodesic. None of the parallels, which in this case are the generators of the cone, are geodesics.

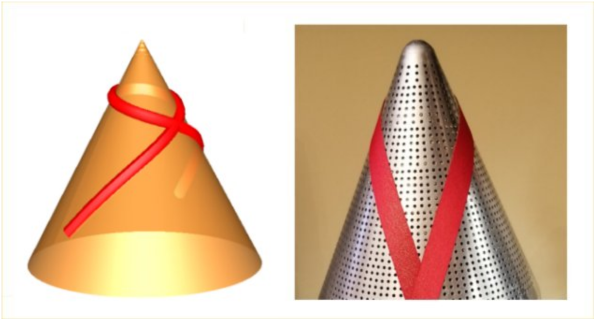


Fig. 6.4: Geodesics on a Cone.

6.7 Geodesics in GR

6.7.1 Example Morris-Thorne (MT) wormhole
In 1987, Michael Morris and Kip Thorne from the California Institute of Technology proposed a tantalizing simple model for teaching general relativity, by alluding to interspace travel in a geometry of traversable wormhole. We constraint the discussion purely to geometrical aspects of the model and not the physics of stress and strains of a "traveler" traversing the wormhole. The MT metric for this spherically symmetric geometry is

$$ds^2 = -c^2 dt^2 + dl^2 + (b_0^2 + l^2)\,(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{6.93}$$

where b_0 is a constant. The obvious choice for a coframe is

$$\begin{aligned}
\theta^0 &= c\, dt, & \theta^2 &= \sqrt{b_o^2 + l^2}\, d\theta,\\
\theta^1 &= dl, & \theta^3 &= \sqrt{b_o^2 + l^2}\, \sin\theta\, d\phi.
\end{aligned}$$

We have $d\theta^0 = d\theta^1 = 0$. To find the connection forms we compute $d\theta^2$ and $d\theta^3$, and rewrite in terms of the coframe. We get

$$\begin{aligned}
d\theta^2 &= \frac{l}{\sqrt{b_o^2 + l^2}}\, dl \wedge d\theta = -\frac{l}{\sqrt{b_o^2 + l^2}}\, d\theta \wedge dl,\\
&= -\frac{l}{b_o^2 + l^2}\, \theta^2 \wedge \theta^1,\\
d\theta^3 &= \frac{l}{\sqrt{b_o^2 + l^2}} \sin\theta\, dl \wedge d\phi + \cos\theta \sqrt{b_o^2 + l^2} d\theta \wedge d\phi,\\
&= -\frac{l}{b_o^2 + l^2}\, \theta^3 \wedge \theta^1 - \frac{\cot\theta}{\sqrt{b_o^2 + l^2}}\, \theta^3 \wedge \theta^2.
\end{aligned}$$

Comparing with the first equation of structure, we start with simplest guess for the connection forms ω's. That is, we set

$$\begin{aligned}
\omega^2{}_1 &= \frac{l}{b_o^2 + l^2}\,\theta^2, \\
\omega^3{}_1 &= \frac{l}{b_o^2 + l^2}\,\theta^3, \\
\omega^3{}_2 &= \frac{\cot\theta}{\sqrt{b_o^2 + l^2}}\,\theta^3.
\end{aligned}$$

Using the antisymmetry of the ω's and the diagonal metric, we have $\omega^2{}_1 = -\omega^1{}_2$, $\omega^1{}_3 = -\omega^3{}_1$, and $\omega^2{}_3 = -\omega^3{}_2$. This choice of connection coefficients turns out to be completely compatible with the entire set of Cartan's first equation of structure, so, these are the connection forms, all other ω's are zero. We can then proceed to evaluate the curvature forms. A straightforward calculus computation which results in some pleasing cancellations, yields

$$\begin{aligned}
\Omega^1{}_2 &= d\omega^1{}_2 + \omega^2{}_1 \wedge \omega^2{}_1 = -\frac{b_o^2}{(b_o^2 + l^2)^2}\,\theta^1 \wedge \theta^2, \\
\Omega^1{}_3 &= d\omega^1{}_3 + \omega^1{}_2 \wedge \omega^2{}_3 = -\frac{b_o^2}{(b_o^2 + l^2)^2}\,\theta^1 \wedge \theta^3, \\
\Omega^2{}_3 &= d\omega^2{}_3 + \omega^2{}_1 \wedge \omega^1{}_3 = \frac{b_o^2}{(b_o^2 + l^2)^2}\,\theta^2 \wedge \theta^3.
\end{aligned}$$

Thus, from equation 6.36, other than permutations of the indices, the only independent components of the Riemann tensor are

$$R_{2323} = -R_{1212} = R_{1313} = \frac{b_o^2}{(b_o^2 + l^2)^2},$$

and the only non-zero component of the Ricci tensor is

$$R_{11} = -2\frac{b_o^2}{(b_o^2 + l^2)^2}.$$

Of course, this space is a 4-dimensional continuum, but since the space is spherically symmetric, we may get a good sense of the geometry by taking a slice with $\theta = \pi/2$ at a fixed value of time. The resulting metric ds_2 for the surface is

$$ds_2{}^2 = dl^2 + (b_o^2 + l^2)\,d\phi^2. \tag{6.94}$$

Let $r^2 = b_o^2 + l^2$. Then $dl^2 = (r^2/l^2)\,dr^2$ and the metric becomes

$$\begin{aligned}
ds_2{}^2 &= \frac{r^2}{r^2 - b_o^2}\,dr^2 + r^2\,d\phi^2, && (6.95) \\
&= \frac{1}{1 - \frac{b_o^2}{r^2}}\,dr^2 + r^2\,d\phi^2. && (6.96)
\end{aligned}$$

Comparing to 4.26 we recognize this to be a catenoid of revolution, so the equations of geodesics are given by 6.91 with $f(r) = b_0 \cosh^{-1}(r/b_0)$. Substituting this value of f into the geodesic equation, we get

$$\phi = \pm c \int \frac{1}{\sqrt{r^2 - b_o^2}\sqrt{r^2 - c^2}} dr. \tag{6.97}$$

There are three cases. If $c = b_o$, the integral gives immediately

$$\phi = \pm(c/b_0)\tanh^{-1}(r/b_0).$$

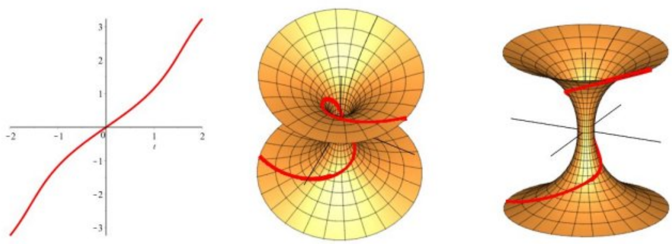


Fig. 6.5: Geodesics on Catenoid.

We consider the case $c > b_0$. The remaining case can be treated in a similar fashion. Let $r = c/\sin\beta$. Then $\sqrt{r^2 - c^2} = r\cos\beta$ and $dr = -r\cot\beta\, d\beta$, so, assuming the initial condition $\phi(0) = 0$, the substitution leads to the integral

$$\begin{aligned}
\phi &= \pm c \int_0^s \frac{1}{r\cos\beta\sqrt{\frac{c^2}{\sin^2\beta} - b_o^2}} \frac{(-r\cos\beta)}{\sin\beta}\, d\beta,\\
&= \pm c \int_0^s \frac{1}{\sqrt{c^2 - b_o^2\sin^2\beta}}\, d\beta,\\
&= \pm \int_0^s \frac{1}{\sqrt{1 - k^2\sin^2\beta}}\, d\beta, \quad (k = b_o/c) && (6.98)\\
&= F(s,k), && (6.99)
\end{aligned}$$

where $F(s,k)$ is the well-known incomplete elliptic integral of the first kind.

Elliptic integrals are standard functions implemented in computer algebra systems, so it is easy to render some geodesics as shown in figure 6.5. The plot of the elliptic integral shown here is for $k = 0.9$. The plot shows clearly that this is a 1-1, so if one wishes to express r in terms of ϕ one just finds the inverse of the elliptic integral which yields a Jacobi elliptic function. Thomas Muller has created a neat Wolfram-Demonstration that allows the user to play with MT wormhole geodesics with parameters controlled by sliders.

6.7.2 Example Schwarzschild Metric

In this section we look at the geodesic equations in a Schwarzschild gravitational field, with particular emphasis on the bounded orbits. We write the metric in the form

$$ds^2 = -h(r)\,dt^2 + \frac{1}{h(r)}\,dr^2 + r^2(d\theta^2 + \sin\theta\,d\phi^2), \tag{6.100}$$

where

$$h(r) = 1 - \frac{2GM}{r}. \tag{6.101}$$

Thus, the Lagrangian is

$$\mathscr{L} = -h\,\dot{t}^2 + \frac{1}{h}\,\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin\theta\dot{\phi}^2. \tag{6.102}$$

The Euler-Lagrange equations for g_{00}, g_{22} and g_{33} yield

$$\begin{aligned}
\frac{d}{ds}\left[-2h\frac{dt}{ds}\right] &= 0,\\
\frac{d}{ds}\left[r^2\frac{d\theta}{ds}\right] - r^2\sin\theta\cos\theta\left[\frac{d\phi}{ds}\right]^2 &= 0,\\
\frac{d}{ds}\left[2r^2\frac{d\phi}{ds}\right] &= 0
\end{aligned}$$

If in the equation for g_{22}, one chooses initial conditions $\theta(0) = \pi/2$, $\dot{\theta}(0) = 0$, we get $\theta(s) = \pi/2$ along the geodesic. We infer from rotation invariance that the motion takes place on a plane. Hereafter, we assume we have taken these initial conditions. From the other two equations we obtain

$$\begin{aligned}
h\frac{dt}{ds} &= E,\\
r^2\frac{d\phi}{ds} &= L.
\end{aligned}$$

for some constants E and L. We recognize the conserved quantities as the “energy” and the angular momentum. Along the geodesic of a massive particle, with unit time-like tangent vector, we have

$$-1 = g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} \tag{6.103}$$

The equations of motion then reduce to

$$\begin{aligned}
-1 &= -h\left[\frac{dt}{ds}\right]^2 + \frac{1}{h}\left[\frac{dr}{ds}\right]^2 + r^2\left[\frac{d\phi}{ds}\right]^2,\\
-1 &= -\frac{E^2}{h} + \frac{1}{h}\left[\frac{dr}{ds}\right]^2 + \frac{L^2}{r^2},\\
E^2 &= \left[\frac{dr}{ds}\right]^2 + h\left[1 + \frac{L^2}{r^2}\right].
\end{aligned}$$

Hence, we obtain the neat equation,

$$E^2 = \left[\frac{dr}{ds}\right]^2 + V(r), \tag{6.104}$$

where $V(r)$ represents the effective potential.

$$\begin{aligned} V(r) &= \left[1 - \frac{2GM}{r}\right]\left[1 + \frac{L^2}{r^2}\right], \\ &= 1 - \frac{2GM}{r} + \frac{L^2}{r^2} - \frac{2MGL^2}{r^3}. \end{aligned} \tag{6.105}$$

If we let $\hat{V} = V/2$ in this expression we recognize the classical $1/r$ potential, and the $1/r^2$ term corresponding to the Coriolis contribution associated with the angular momentum. The $1/r^3$ term is a new term arising from general relativity. Clearly we must have $E^2 < V(r)$. There are multiple cases depending

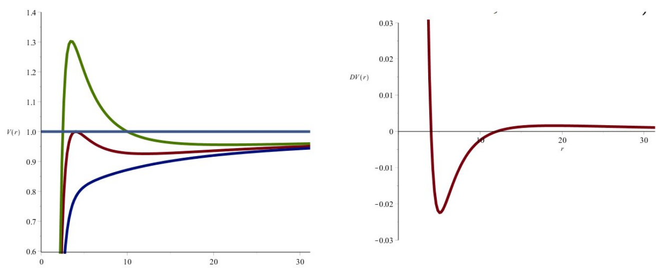


Fig. 6.6: Effective Potential for $L = 3, 4, 5$

on the values of E and L and the nature of the equilibrium points. Here we are primarily concerned with bounded orbits, so we seek conditions for the particle to be in a potential well. This presents us with a nice calculus problem. We compute $V'(r)$ and set equal to zero to find the critical points

$$V'(r) = \frac{2}{r^4}(GMr^2 - L^2 r + 3GML^2) = 0.$$

The discriminant of the quadratic is

$$D = L^2 - 12G^2M^2.$$

If $D < 0$ there are no critical points. In this case, $V(r)$ is a monotonically increasing function on the interval $(2MG, \infty)$, as shown in the bottom left graph in figure 6.6. The maple plots in this figure are in units with $GM = 1$. In the case $D < 0$, all trajectories either fall toward the event horizon or escape to infinity.

If $D > 0$, there are two critical points

$$r_1 = \frac{L^2 - L\sqrt{L^2 - 12G^2M^2}}{2GM},$$
$$r_2 = \frac{L^2 + L\sqrt{L^2 - 12G^2M^2}}{2GM}.$$

The critical point r_1 is a local maximum associated with an unstable circular orbit. The critical point $r_2 > r_1$ gives a stable circular orbit. Using the standard calculus trick of multiplying by the conjugate of the radical in the first term, we see that

$$r_1 \to 3GM,$$
$$r_2 \to \frac{L^2}{GM},$$

as $L \to \infty$. For any L, the properties of the roots of the quadratic imply that $r_1 r_2 = 3L^2$. As shown in the graph 6.6, as L gets larger, the inner radius approaches $3GM$ and the height of the bump increases, whereas the outer radius recedes to infinity. As the value of D approaches 0, the two orbits coalesce at $L^2 = 12G^2M^2$, which corresponds to $r = 6GM$, so this is the smallest value of r at which a stable circular orbit can exist. Since $V(r) \to 1$ as $r \to \infty$, to get bounded orbits we want a potential well with $V(r_1) < 1$. We can easily verify that when $L = 4GM$ the local maximum occurs at $r_1 = 4GM$, which results in a value of $V(r_1) = 1$. This case is the one depicted in the middle graph in figure 6.6, with the graph of $V'(r)$ on the right showing the two critical points at $r_1 = 4GM, \;\; r_2 = 12GM$. Hence the condition to get a bounded orbit is

$$2\sqrt{3}\,GM < L < 4GM,$$
$$E^2 < V(r_1), \;\; r > r_1,$$

so that the energy results in the particle trapped in the potential well to the right of r_1. This is the case that applies to the modification of the Kepler orbits of planets. If we rewrite

$$\frac{dr}{ds} = \frac{dr}{d\phi}\frac{d\phi}{ds} = \frac{L}{r^2}\frac{dr}{d\phi}$$

and substitute into equation 6.104, we get

$$\frac{L^2}{r^4}\left[\frac{dr}{d\phi}\right]^2 = E^2 - \left[1 + \frac{L^2}{r^2}\right]\left[1 - \frac{2GM}{r}\right].$$

If now we change variables to $u = 1/r$, we obtain

$$\frac{du}{d\phi} = -\frac{1}{r^2}\frac{dr}{d\phi} = -u^2\frac{dr}{d\phi},$$

and the orbit equation becomes

$$\left[\frac{du}{d\phi}\right]^2 = \frac{1}{L^2}[E^2 - (1 + L^2u^2)(1 - 2GMu)],$$
$$\phi = \int \frac{L du}{\sqrt{E^2 - (1 + L^2u^2)(1 - 2GML^2u)}} + \phi_0.$$

The solution of the orbit equation is therefore reduced to an elliptic integral. If we expand the denominator

$$\phi = \int \frac{L du}{\sqrt{(E^2-1) + 2GMu - L^2u^2 + 2GML^2u^3}} + \phi_0,$$

and neglect the cubic term, we can complete the squares of the remaining quadratic. The integral becomes one of standard inverse cosine type; hence, the solution gives the equation of an ellipse in polar coordinates

$$u = \frac{1}{r} = C(1 + e\cos(\phi - \phi_0)),$$

for appropriate constants C, shift ϕ_0 and eccentricity e. The solution is automatically expressed in terms of the energy and the angular momentum of the system. More careful analysis of the integral shows that the inclusion of the cubic term perturbs the orbit by a precession of the ellipse. While this approach is slicker, we prefer to use the more elementary procedure of differential equations. Differentiating with respect to ϕ the equation

$$L^2\left[\frac{du}{d\phi}\right]^2 = (E^2-1) + 2GMu - L^2u^2 + 2GML^2u^3,$$

and cancelling out the common chain rule factor $du/d\phi$, we get

$$\frac{d^2u}{d\phi^2} = \frac{GM}{L^2} - u + 3GMu^2$$

Introducing a dimensionless parameter

$$\epsilon = \frac{3G^2M^2}{L^2},$$

we can rewrite the equation of motion as

$$\frac{d^2u}{d\phi^2} + u = \frac{GM}{L^2} + \frac{L^2}{GM}u^2\epsilon. \tag{6.106}$$

The linear part of the equation corresponds precisely to Newtonian motion, and ϵ is small, so we can treat the quadratic term as a perturbation

$$u = u_0 + u_1\epsilon + u_2\epsilon^2 + \dots.$$

Substituting u into equation 6.106, the first approximation is the linear approximation given by

$$u_0'' + u = \frac{GM}{L^2}.$$

The homogenous solution is of the form $u = A\cos(\phi - \phi_0)$, where A and ϕ_0 are the arbitrary constants, and the particular solution is a constant. So the general solution is

$$\begin{aligned} u_0 &= \frac{GM}{L^2} + A\cos(\phi - \phi_0), \\ &= \frac{GM}{L^2}[1 + e\cos(\phi - \phi_0)], \quad e = \frac{AL^2}{GM}. \end{aligned}$$

Without loss of generality, we can align the axes and set $\phi_0 = 0$. In the Newtonian orbit, we would write $u_0 = 1/r$, thus getting the equation of a polar conic.

$$u_0 = \frac{GM}{L^2}(1 + e\cos\phi) \tag{6.107}$$

In the case of the planets, the eccentricity $e < 1$, so the conics are ellipses. Having found u_0 we reinsert u into the differential equation 6.106 and keeping only the terms of order ϵ. We get

$$\begin{aligned} (u_0 + u_1\epsilon)'' + (u_0 + u_1\epsilon) &= \frac{GM}{L^2} + \frac{L^2}{GM}\epsilon(u_0 + u_1\epsilon)^2, \\ (u_0'' + u_0 - \frac{GM}{L^2}) + (u_1'' + u_1)\epsilon &= \frac{L^2}{GM}u_0^2\epsilon. \end{aligned}$$

Thus, the result is a new differential equation for u_1,

$$\begin{aligned} u_1'' + u_1' &= \frac{L^2}{GM}u_0^2, \\ &= \frac{L^2}{GM}[(1 + \tfrac{1}{2}e^2) + 2e\cos\phi + \tfrac{1}{2}e^2\cos 2\phi]. \end{aligned}$$

The equation is again a linear inhomogeneous equation with constant coefficients, so it is easily solved by elementary methods. We do have to be a bit careful since we have a resonant term on the right hand side. The solution is

$$u_1 = \frac{L^2}{GM}[(1 + \tfrac{1}{2}e^2) + 2e\phi\cos\phi - \tfrac{1}{6}e^2\cos 2\phi].$$

The resonant term $\phi\cos\phi$ makes the solution non-periodic, so this is the term responsible for the precession of the elliptical orbits. The precession is obtained by looking at the perihelion, that is, the point in the elliptical orbit at which the planet is closest to the sun. This happens when

$$\begin{aligned} \frac{du}{d\phi} \approx \frac{d}{d\phi}(u_0 + u_1) &= 0, \\ -\sin\phi + (\sin\phi + e\phi\cos\phi + \tfrac{1}{3}e\sin\phi) &= 0. \end{aligned}$$

Starting with the solution $\phi = 0$, after on revolution, the perihelion drifts to $\phi = 2\pi + \delta$. By the perturbation assumptions, we assume δ is small, so to lowest order, the perihelion advance in one revolution is

$$\delta = 2\pi\epsilon = \frac{6\pi G^2M^2}{L^2}. \tag{6.108}$$

From equation 6.107 for the Newtonian elliptical orbit, the mean distance a to the sun is given by the average of the aphelion and perihelion distances, that is

$$a = \frac{1}{2}\left[\frac{L^2/GM}{1+e} + \frac{L^2/GM}{1-e}\right] = \frac{L^2}{GM}\frac{1}{1-e^2}.$$

Thus, if we divide by the period T, the rate of perihelion advance can be written in more geometric terms as

$$\delta = \frac{6\pi GM}{a(1-e^2)T}.$$

The famous computation by Einstein of a precession of 43.1" of an arc per century for the perihelion advance of the orbit of Mercury, still stands as one of the major achievements in modern physics.

For null geodesics, equation 6.103 is replaced by

$$0 = g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds},$$

so the orbit given by the simpler equation

$$E^2 = \left[\frac{dr}{ds}\right]^2 + \frac{L^2}{r^2}h.$$

Performing the change of variables $u = 1/r$, we get

$$\frac{d^2u}{d\phi^2} + u = 3GMu^2.$$

Consider the problem of light rays from a distant star grazing the sun as they approach the earth. Since the space is asymptotically flat, we expect the geodesics to be asymptotically straight. The quantity $3GM$ is of the order of 2km, so it is very small compared to the radius of the sun, so again we can use perturbation methods. We let $\epsilon = 3GM$ and consider solutions of equation

$$u'' + u = \epsilon u^2,$$

of the form

$$u = u_0 + u_1\epsilon.$$

To lowest order the solutions are indeed straight lines

$$\begin{aligned} u_0 &= A\cos\phi + B\sin\phi, \\ 1 &= Ar\cos\phi + Br\sin\phi, \\ 1 &= Ax + By \end{aligned}$$

Without loss of generality, we can align the vertical axis parallel to the incoming light with impact parameter b (distance of closest approach)

$$u_0 = \frac{1}{b}\cos\phi.$$

As above, we reinsert the u into the differential equation and compare the coefficients of terms of order ϵ. We get an equation for u_1,

$$u_1'' + u_1 = \frac{1}{b^2}\cos^2\phi = \frac{1}{2b^2}(1+\cos 2\phi).$$

We solve the differential equation by the method of undetermined coefficients and thus we arrive at the perturbation solution to order ϵ

$$u = \frac{1}{b}\cos\phi + \frac{2\epsilon}{3b^2} - \frac{\epsilon}{3b^2}\cos^2\phi.$$

To find the the asymptotic angle of the outgoing photons, we let $r \to \infty$ or $u \to 0$. Thus we get a quadratic equation for $\cos\phi$.

$$\cos\phi = -\frac{2\epsilon}{3b} = -\frac{2GM}{b}$$

Set $\phi = \frac{\pi}{2} + \delta$. Since δ is small, we have $\sin\delta \approx \delta$, and we see that $\delta = 2GM/b$ is the approximation of the deflection angle of one of the asymptotes. The total deflection is twice that angle

$$2\delta = \frac{4GM}{b}.$$

The computation results in a deflection by the sun of light rays from a distant star of about 1.75". This was corroborated in an experiment lead by Eddington during the total solar eclipse of 1919. The part of the expedition in Brazil was featured in the 2005 movie, The House of Sand. For more details and more careful analysis of the geodesics, see for example, Misner Thorne and Wheeler [21].

6.8 Gauss-Bonnet Theorem

This section is dedicated to the memory of Professor S.-S. Chern. I prelude the section with a short anecdote that I often narrate to my students. In June 1979, an international symposium on differential geometry was held at the Berkeley campus in honor of the retirement of Professor Chern. The invited speakers included an impressive list of the most famous differential geometers at the time, At the end of the symposium, Chern walked on the stage of the packed auditorium to give thanks and to answer some questions. After a few short remarks, a member of the audience asked Chern what he thought was the most important theorem in differential geometry. Without any hesitation he answered, “there is only one theorem in differential geometry, and that is Stokes’ theorem.” This was followed immediately by a question about the most important theorem in analysis. Chern gave the same answer: “there is only one theorem in analysis, Stokes’ theorem. A third person then asked Chern what was the most important theorem in Complex Variables. To the amusement of the crowd, Chern responded, “There is only one theorem in complex variables, and that that is Cauchy’s theorem. But if one assumes the derivative of the function is continuous, then this is just Stokes’ theorem.” Now, of course it is well known that Goursat proved that the hypothesis of continuity of the derivative is automatically satisfied when the function is holomorphic. But the genius of Chern was always his uncanny ability to extract the essential of what makes things work, in the simplest terms.

The Gauss-Bonnet theorem is rooted on the theorem of Gauss (4.72), which combined with Stokes’ theorem, provides a beautiful geometrical interpretation of the equation. This is undoubtedly part of what Chern had in mind at the symposium, and also when wrote in his Euclidean Differential Geometry Notes (Berkeley 1975) [4] that the theorem has “profound consequences and is perhaps one of the most important theorems in mathematics.”

Let $\beta(s)$ by a unit speed curve on an orientable surface M, and let T be the unit tangent vector. There is Frenet frame formalism for M, but if we think of the surface intrinsically as 2-dimensional manifold, then there is no binormal. However, we can define a “geodesic normal” taking $G = J(T)$, where J is the symplectic form 5.50, Then the geodesic curvature is given by the Frenet formula

$$T' = \kappa_g G. \tag{6.109}$$

6.8.1 Proposition Let $\{e_1, e_2\}$ be an orthonormal on M, and let $\beta(s)$ be a unit speed curve as above, with unit tangent T. If ϕ is the angle that T makes with e_1, then

$$\kappa_g = \frac{\partial \phi}{\partial s} - \omega^1{}_2(T). \tag{6.110}$$

Proof Since $\{T, G\}$ and $\{e_1, e_2\}$ are both orthonormal basis of the tangent space, they must be related by a rotation by an angle ϕ, that is

$$\begin{bmatrix} T \\ G \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \tag{6.111}$$

that is,

$$\begin{aligned} T =& \quad (\cos\phi)e_1 + (\sin\phi)e_2, \\ G =& -(\sin\phi)e_1 + (\cos\phi)e_2. \end{aligned} \tag{6.112}$$

Since $T = \beta'$, and $\beta'' = \nabla_t T$ we have

$$\begin{aligned} \beta'' &= -(\sin\phi)\frac{\partial \phi}{\partial s}e_1 + \cos\phi \nabla_T e_1 + (\cos\phi)\frac{\partial \phi}{\partial s}e_2 + \sin\phi \nabla_T e_2, \\ &= -(\sin\phi)\frac{\partial \phi}{\partial s}e_1 + (\cos\phi)\omega^2{}_1(T)e_2 + (\cos\phi)\frac{\partial \phi}{\partial s}e_2 + (\sin\phi)\omega^1{}_2(T)e_1, \\ &= [\frac{\partial \phi}{\partial s} - \omega^1{}_2(T)][-(\sin\phi)e_1] + [\frac{\partial \phi}{\partial s} - \omega^1{}_2(T)][(\cos\phi)e_2], \\ &= [\frac{\partial \phi}{\partial s} - \omega^1{}_2(T)][-(\sin\phi)e_1 + (\cos\phi)e_2], \\ &= [\frac{\partial \phi}{\partial s} - \omega^1{}_2(T)]G, \\ &= \kappa_g G. \end{aligned}$$

comparing the last two equations, we get the desired result.

This theorem is related to the notion discussed in figure 6.2 to the effect that in a space with curvature, the parallel transport of a tangent vector around a closed curve, does not necessarily result on the same vector with which one started. The difference in angle $\Delta\phi$ between a vector and the parallel transport of the vector around a closed curve C is called the *holonomy* of the curve. The holonomy of the curve is given by the integral

$$\Delta\phi = \int_C \omega^1{}_2(T)\, ds. \tag{6.113}$$

6.8.2 Definition Let C be a smooth closed curve on M parametrized by arc length with geodesic curvature κ_g. The line integral $\oint_C \kappa_g\, ds$ is called the *total geodesic curvature*. If the curve is piecewise smooth, the total geodesic curvature is the sum of the integrals of each piece.

A circle of radius R gives an elementary example. The geodesic curvature is the constant $1/R$, so the total geodesic curvature is $(1/R)2\pi R = 2\pi$.

If we integrate formula 6.110 around a smooth simple closed curve C which is the boundary of a region R and use Stokes' Theorem, we get

$$\begin{aligned}\oint_C \kappa_g\, ds &= \oint_C d\phi - \oint_C \omega^1{}_2\, ds,\\ &= \oint_C d\phi - \int\int_R d\omega^1{}_2.\end{aligned}$$

For a smooth simple closed curve, $\int_C d\phi = 2\pi$. Using the Cartan-form version of the theorema egregium 4.106 we get immediately

$$\int\int_R K\, dS + \int_C \kappa_g\, ds = 2\pi. \tag{6.114}$$

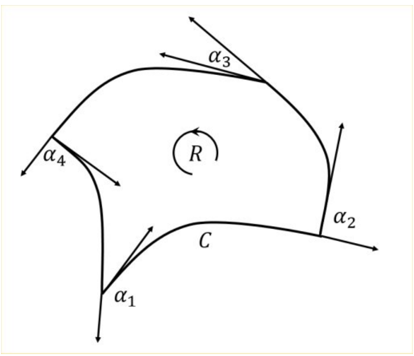


Fig. 6.7: Turning Angles

If the boundary of the region consists of k piecewise continuous functions as illustrated in figure 6.7, the change of the angle ϕ along C is still 2π, but the total change needs to be modified by adding the exterior angles α_k. Thus, we obtain a fundamental result called the *Gauss-Bonnet formula*,

6.8.3 Theorem

$$\int\int_R K\, dS + \int_C \kappa_g\, ds + \sum_k \alpha_k = 2\pi. \tag{6.115}$$

Every interior ι_k angle is the supplement of the corresponding exterior α_k angle, so the Gauss Bonnet formula can also be written as

$$\int\int_R K\, dS + \int_C \kappa_g\, ds + \sum_k (\pi - \iota_k) = 2\pi. \tag{6.116}$$

The simplest manifestation of the Gauss-Bonnet formula is for a triangle in the plane. Planes are flat surfaces, so $K = 0$ and the straight edges are geodesics, so $\kappa_g = 0$ on each of the three edges. The interior angle version of the formula then just reads $3\pi - \iota_1 - \iota_2 - \iota_3 = 2\pi$, which just says that the interior angles of a flat triangle add up to π. Since a sphere has constant positive curvature, the sum of the interior angles of a spherical triangle is larger than π. That amount of this sum over 2π is called the spherical excess. For example, the sum of the interior angles of a spherical triangle that is the boundary of one octant of a sphere is $3\pi/2$, so the spherical excess is $\pi/2$.

6.8.4 Definition The quantity $\int\int K\, dS$ is called the *total curvature*

6.8.5 Example A sphere of radius R has constant Gaussian Curvature $1/R^2$. The surface area of the sphere is $4\pi R^2$, so the total Gaussian curvature for the sphere is 4π.

6.8.6 Example For a torus generated by a circle of radius a rotating about an axis with radius b as in example (4.40), the differential of surface is $dS = a(b + a\cos\theta)\, d\theta d\phi$, and the Gaussian curvature is $K = \cos\theta/[a(b + a\cos\theta)]$, so the total Gaussian curvature is

$$\int_0^{2\pi}\int_0^{2\pi} \cos\theta\, d\theta d\phi = 0.$$

We now relate the Gauss-Bonnet formula to a topological entity.

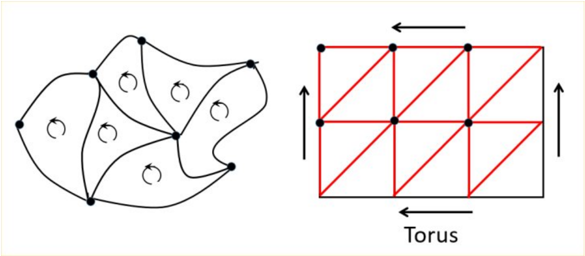


Fig. 6.8: Triangulation

6.8.7 Definition Let M be a 2-dimensional manifold. A *triangulation* of the surface is subdivision of the surface into triangular regions $\{\triangle_k\}$ which are the images of regular triangles under a coordinate patch, such that:

1) $M = \bigcup_k \triangle_k$.
2) $\triangle_i \bigcap \triangle_j$ is either empty, or a single vertex or an entire edge.
3) All the triangles are oriented in the same direction,

For an intuitive visualization of the triangulation of a sphere, think of inflating a tetrahedron or an octahedron into a spherical balloon. We state without proof:

6.8.8 Theorem Any compact surface can be triangulated.

6.8.9 Theorem Given a triangulation of a compact surface M, let V be the number of vertices, E the number of edges and F the number of faces. Then the quantity

$$\chi(M) = V - E + F, \tag{6.117}$$

is independent of the triangulation. In fact the quantity is independent of any "polyhedral" subdivision. This quantity is a topological invariant called the *Euler characteristic*.

6.8.10 Example

1. A balloon-inflated tetrahedron has $V = 4,\ E = 6,\ F = 4$, so the Euler characteristic of a sphere is 2.

2. A balloon-inflated octahedron has $V = 6,\ E = 12,\ F = 8$, so we get the same number 2.

3. The diagram on the right of gigure 6.8 represents a topological torus. In the given rectangle, opposites sides are identified in the same direction. The number of edges without double counting are shown in red, and the number of vertices not double counted are shown in black dots. We have $V = 6,\ E = 18\ F = 12$. So the Euler characteristic of a torus is 0.

4. In one has a compact surface, one can add a "handle", that is, a torus, by the following procedure. We excise a triangle in each of the two surfaces and glue the edges. We lose two faces and the number of edges and vertices cancel out, so the Euler characteristic of the new surface decreases by 2. The Euler characteristic of a pretzel is -4.

5. The Euler characteristic of an orientable surface of genus g, that is, a surface with g holes is given by $\chi(M) = 2 - 2g$.

6.8.11 Theorem Gauss-Bonnet
Let M be a compact, orientable surface. Then

$$\frac{1}{2\pi}\int_M K\, dS = \chi(M). \tag{6.118}$$

Proof Triangulate the surface so that $M = \bigcup_{k=1}^{F} \triangle_k$. We start the Gauss-Bonnet formula

$$\int\int_M K\, dS = \sum_{k=1}^{F} \int\int_{\triangle_k} K\, dS = -\sum_{k=1}^{F} \left[\oint_{\delta\triangle_k} \kappa_g\, ds + \pi + (\iota_{k1} + \iota_{k2} + \iota_{k2})) \right],$$

where F is the number of triangles and the ι_k's are the interior angles of triangle $\triangle_k$. The line integrals of the geodesic curvatures all cancel out since each edge in every triangle is traversed twice, each in opposite directions. Rewriting the equation, we get

$$\int\int_M K\, dS = -\pi F + \mathcal{S}$$

where $\mathcal{S}$ is the sum of all interior angles. Since the manifold is locally Euclidean, the sum of all interior angles at a vertex is 2π, so we have

$$\int\int_M K\, dS = -\pi F + 2\pi V$$

There are F faces. Each face has three edges, but each edge is counted twice, so $3F = 2E$, and we have $F = 2E - 2F$ Substituting in the equation above, we get,

$$\int\int_M K\, dS = -\pi(2E - 2F) + 2\pi V = 2\pi(V - E + F) = \chi(M).$$

This is a remarkable theorem because it relates the bending invariant Gaussian curvature to a topological invariant. Theorems such as this one which cut across disciplines, are the most significant in mathematics. Not surprisingly, it was Chern who proved a generalization of the Gauss-Bonnet theorem to general orientable Riemannian manifolds of even dimensions [5].

Chapter 7

Groups of Transformations

7.1 Lie Groups

At the IX International Colloquium on Group Theoretical Methods in Physics held in 1980 at Cocoyoc, Mexico, one of the invited addresses was delivered by the famous mathematician Bertram Konstant, who years later would be awarded the Wigner Medal. In his opening remarks, Konstant made the following intriguing statement, "In the 1800's, Felix Klein and Sophus Lie decided to divide mathematics among themselves. Klein took the discrete and Lie took the continuous. I am here to tell you that they were both working on the same thing."

In this chapter we present an elementary introduction to Lie groups and corresponding Lie algebras. A simple example of a Lie group is the circle group $U(1) = \{z \in \mathbf{C} : |z|^2 = 1\}$ which, as a manifold, corresponds to the unit circle S^1. In polar form, an element of this group can be written in the form $e^{i\theta}$. The group multiplication $z \to e^{i\theta}z$ corresponds to a rotation of the vector z by an angle θ. Using the matrix representation 5.2.2 for complex numbers and Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$, we see that the group is isomorphic to $SO(2, \mathbf{R})$ as shown in example 3.4. This example captures the essence of what we seek, that is, a group that is also a manifold and that can be associated with some matrix group. A $U(1)$ bundle consists of a base manifold M with a structure that locally looks like a cross product of an open set in M with $U(1)$. Generalizations in which the fibers are Lie groups leads to a structure called a principal fiber bundle. Lie Groups, Lie algebras and principal fiber bundles provide the mathematical foundation in modelling symmetries in classical and quantum physics. The subject is much too rich to give a comprehensive treatment here, but we hope the material will serve as a starting block for further study.

7.1.1 Definition A *Lie group* G is a group that is also a smooth manifold.

It is assumed that the usual group multiplication and inverse operations,

$$\begin{aligned} \mu : G \times G &\to G, \\ (g_1, g_2) &\mapsto g_1 g_2, \\ \iota : G &\to G, \\ g &\mapsto g^{-1}, \end{aligned}$$

are C^∞.

7.1.2 Definition A *Lie subgroup* is a subset $H \subset G$ of a Lie group G, that is itself a Lie group.

7.1.3 Examples

1. The (real) *general linear group* is the set of $n \times n$ matrices,

$$GL(n, \mathbf{R}) = \{A \in M_{n\times n}(\mathbf{R}) : \det(A) \neq 0\}. \tag{7.1}$$

   Topologically, $GL(n, \mathbf{R})$ is equivalent to $\mathbf{R}^{n^2}$ and has the structure of an n^2-dimensional differentiable manifold. The map $\det : GL(n, \mathbf{R}) \mapsto \mathbf{R}$ is continuous. The inverse image of 0 under this map is a closed space and $GL(n, \mathbf{R})$ is the complement, so $GL(n, \mathbf{R})$ is an open subset of $\mathbf{R}^{n^2}$ and thus it is not compact. $GL(n, \mathbf{R})$ is not connected being the union of two disjoint open sets defined by whether $\det(A) > 0$ or $\det(A) < 0$. The connected component $GL^+(n, \mathbf{R})$ corresponding to $\det(A) > 0$, contains the identity. Furthermore, if $\det(A) > 0$ and $\det B > 0$, we have $\det(AB) > 0$ and $\det(A^{-1}) > 0$, so $GL^+(n, \mathbf{R})$ is a (non-compact) Lie subgroup.

2. The subset of $GL^+(n, \mathbf{R})$ with the restriction $\det(A) = 1$ is called the *special linear group* $SL(n, \mathbf{R})$. This is also a non-compact subgroup.

3. The *complex general linear group* is the set of $n \times n$ matrices,

$$GL(n, \mathbf{C}) = \{A \in M_{n\times n}(\mathbf{C}) : \det(A) \neq 0\}. \tag{7.2}$$

   The subgroup of matrices $A \in GL(n, \mathbf{C})$ with $\det(A) = 1$ is called the special linear group $SL(n, \mathbf{C})$.

4. The real *orthogonal group* is the set of $n \times n$ real matrices

$$O(n, \mathbf{R}) = \{A \in M_{n\times n}(\mathbf{R}) : A^{-1} = A^T\}. \tag{7.3}$$

   The condition $A^{-1} = A^T$ is equivalent to $AA^T = A^T A = I$. We have the following

   a) $I^T = I^{-1} = I$, so $I \in O(n, \mathbf{R})$.

   b) If $A, B \in O(n, \mathbf{R})$, then $(AB)(AB)^T = ABB^T A^T = AA^T = I$, so $AB \in O(n, \mathbf{R})$.

c) If $A \in O(n, \mathbf{R})$, then $A^{-1}(A^{-1})^T = A^T(A^T)^T = A^T A = I$.

Hence $O(n, \mathbf{R})$ is a Lie subgroup of the $GL(n, \mathbf{R})$. The map $T(A) = AA^T$ is continuous and $O(n, \mathbf{R}) = T^{-1}(I)$, so $O(n, \mathbf{R})$ is closed.

If we denote by e_k the kth column vector of A, then the matrix element of $A^T A$ in the jth row and kth column is given by

$$(e_j)^T e_k =< e_j, e_k >= \delta_{jk}.$$

Thus, the columns (and rows) of an orthogonal matrix constitute a set of orthonormal vectors. There are n column vectors, so under the Euclidean norm of $\mathbf{R}^{n^2}$, we have $\|A\|^2 = n$. That is, elements of $O(n, \mathbf{R})$ lie on a sphere S^{n^2-1}, so the set is bounded. By the Heine-Borel theorem, $O(n, \mathbf{R})$ is compact. By equation 1.57 and the subsequent Theorem, we can characterize the orthogonal group as the set of linear transformations that preserves the standard metric $g = \text{diag}(+1, +1, \cdots + 1)$ in $\mathbf{R}^n$.

5. If a matrix A is orthogonal, then the condition $AA^T = I$ implies that

$$\begin{aligned}\det(AA^T) &= \det(A)\det(A^T),\\ &= \det(A)^2 = \det(I),\end{aligned}$$

so $\det(A) = \pm 1$. We define the (real) *special orthogonal group* $SO(n, \mathbf{R})$ to be the subset of $O(n, \mathbf{R})$ of orthogonal matrices A, with $\det A = 1$. $SO(n, \mathbf{R})$ is a compact Lie subgroup of dimension $\frac{1}{2}n(n-1)$.

6. The *unitary group* is the set if $n \times n$ matrices

$$U(n) = \{A \in M_{n\times n}(\mathbf{C}) : A^{-1} = A^\dagger\}, \tag{7.4}$$

where $A^\dagger$ is the Hermitian adjoint. This is the complex analog of the orthogonal group. The condition $A^{-1} = A^\dagger$ is equivalent to $AA^\dagger = I$, which implies that $\det(A) = \pm 1$. The subgroup of unitary matrices A with $\det(A) = 1$ is called the Special Unitary group $SU(n)$; it is a compact group of dimension $n^2 - 1$.

7. Let $\{M, g\}$ be the pseudo-Riemannian manifold $\mathbf{R}^n$ with a type (p, q) metric with signature $g = \text{diag}(1, 1, \cdots - 1, -1 \ldots)$. The group of transformations preserving this metric is called $O(p, q)$. If in addition, the matrices $A \in O(p, q)$ are required to have $\det(A) = 1$, the group is called $SO(p, q)$. These groups are not compact. The special case $L = O(1, 3,)$ is the group of transformations preserving the Minkowski metric. This group is called the *Lorentz* group which, is central to relativistic physics.

8. In a completely analogous manner, let $\{M, g\}$ be the pseudo-Riemannian complex manifold $\mathbf{C}^n$ with a hermitian metric $g = \text{diag}(1, 1, \cdots - 1, -1)$ of type (p, q). The group of transformations preserving this metric is called $U(p, q)$. If in addition, the matrices $A \in U(p, q)$ are required to have $\det(A) = 1$, the group is called $SU(p, q)$. These groups are also not compact. The special case $SU(2, 2)$ is isomorphic to the *Poincaré group* and is of interest in twistor theory.

9. Consider the space $\mathbf{F}^{2n}$, where $\mathbf{F}$ stands for the reals $\mathbf{R}$, the complex $\mathbf{C}$, or the quaternion $\mathbf{H}$ algebras. Let $(q^1, \ldots q^n, p_1, \ldots p_n)$ denote the local coordinates. In the case $\mathbf{F} = \mathbf{R}$, we may think of the local coordinates as representing position and momenta. Let Ω be the non-degenerate skew-symmetric two-form

$$\Omega = dq^i \wedge dp_i, \; i = 1 \ldots n. \tag{7.5}$$

The *symplectic group* $Sp(2n, \mathbf{F})$ is the group of transformations preserving the symplectic form Ω. The tensor components of Ω in standard basis are given by

$$\Omega = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, \tag{7.6}$$

where I_n is the identity $n \times n$ matrix. Then the symplectic group is defined by

$$Sp(2n, \mathbf{F}) = \{A \in M_{2n \times 2n}(\mathbf{F}) : A^T \Omega A = \Omega\}. \tag{7.7}$$

The symplectic group is an essential structure in the differential geometry description of Lagrangian and Hamiltonian mechanics. In the simplest case in which $\mathbf{F} = \mathbf{R}$, and $n = 1$, the components of the canonical symplectic form is the complex structure introduced in equation 5.50 in the context of conformal maps. It is immediately clear that $Sp(2, \mathbf{R})$ consists of all 2×2 matrices A with $\det(A) = 1$, so $Sp(2, \mathbf{R}) \cong SL(2, \mathbf{R})$. The symplectic groups are simply connected but not compact. We define the compact group,

$$Sp(n) = Sp(2n, \mathbf{C}) \cap SU(2n),$$

that is, the space of all complex symplectic matrices which are also elements of the special unitary group. $Sp(n)$ can be identified with the quaternionic unitary group $U(n, \mathbf{H})$. In particular, $Sp(1)$ is the set of unit quaternions and $Sp(1) \simeq SU(2)$ is topologically a three sphere S^3. More details on this topic appear later in the discussion of quaternions, starting with equation 8.15

7.1.1 One-Parameter Groups of Transformations

In this section we formalize the notion of flows of vector fields mentioned in definition 1.1.13. The concept of flows of vector fields permeates all of physics. The classical description of magnetic fields illustrates this well. Consider for example the Earth's magnetic field. At any point around the planet, one associates a vector and a direction for the magnetic field at that point. If one picks any such point and follows in an infinitesimal trajectory along the earth magnetic field vector, one arrives at new point with corresponding field vector. Iterating the process, one obtains an integral curve on which the vector field restricted to that curve is tangential to the curve. Doing this at all points in a neighborhood of the point then gives rise to a family non-intersecting integral curves that we usually call the magnetic field lines. Magnetic field lines traced by iron filings around a laboratory-grade magnetic sphere, give a geometrical rendition of the direction of the magnetic field at any point. The converse notion of flows is also intuitively clear as shown in figure 7.1. If one has a non-intersecting family of curves on a neighborhood of a point on a manifold, one would expect that the tangent vectors to the family of curves would constitute a vector field. Students acquainted with more advanced classical physics will know that state space of a dynamical systems is equipped with a real value function H called the Hamiltonian, which effectively represents the energy at each point. Hamiltonian mechanics is then formulated in terms of a symplectic structure that associates with H, a Hamiltonian vector field X_H whose integral curves correspond the solutions of the equations of motion. For a rigorous and elegant treatment of this subject, see Abraham-Marsden [20].

7.1.4 *One-Parameter group of diffeomorphisms*

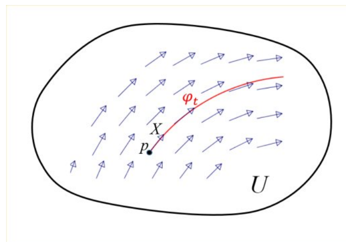


Fig. 7.1: Integral Curve

Let $U \subset M$ be an open subset of an n-dimensional manifold, $p \in U$, and let $I_\epsilon = (-\epsilon, \epsilon)$ with $\epsilon > 0$ be an open interval in $\mathbf{R}$. A one-parameter group of diffeomorphisms is a smooth map,

$$\begin{aligned} \varphi_t : I_\epsilon \times U &\to M \\ (t,p) &\mapsto \varphi_t(p), \ |t| < \epsilon, \end{aligned}$$

with the following properties. Suppose that $t, s \in \mathbf{R}$, with $|t|, |s| < \epsilon$, $|s+t| < \epsilon$, and $\phi_s(p), \phi_t(p), \phi_{s+t}(p) \in U$, then

a) $\varphi_s \circ \varphi_t = \varphi_{s+t}$,

b) $\varphi_0(p) = p$ for all $p \in U$.

The map φ_t is clearly a local diffeomorphism with inverse function given by $(\varphi_t)^{-1} = \varphi_{-t}$. Now consider a vector field X such that $X_p = \varphi'_t(p)$ at each point $p = \varphi_0(p)$, as shown in figure 7.1. If $f : M \to \mathbf{R}$ is a smooth function, then the action X on f as a linear derivation is given by the push-forward formula 1.25

$$\begin{aligned} X_p(f) &= \varphi'_t(p)(f), \\ &= (\varphi_t)_*(\tfrac{d}{dt})(f)|_p, \\ &= \frac{d}{dt}(f \circ \varphi_t(p))|_{t=0} \end{aligned}$$

Conversely, if X is vector field given in local coordinates x^μ in a neighborhood of a point p, given by,

$$X = v^\mu \frac{\partial}{\partial x^\mu}, \quad \mu = 1 \dots n,$$

then, for a curve $\varphi_t(p)$ with initial condition that $\varphi_0(p) = p$, to have X_p as a tangent vector, it must be the case that,

$$\varphi'_t = \frac{dx^\mu}{dt} \frac{\partial}{\partial x^\mu} = X.$$

Thus, we are led to a system of first order ordinary differential equations,

$$\frac{dx^\mu}{dt} = v^\mu(x^1 \dots x^\mu).$$

subject to the condition $x^\mu \circ \varphi_0(p) = v^\mu(p)$ By the existence and uniqueness theorem of solutions of such systems, for sufficiently small ϵ, there exists a unique *integral curve* that satisfies the equation on $I_\epsilon = (-\epsilon, \epsilon)$. We conclude that smooth vector fields can be viewed as generators of infinitesimal groups of diffeomorphisms. The one-parameter group of diffeomorphisms φ_t is also called the *flow* of the vector field X.

At this point it is worthwhile to review the notion of the push-forward of a vector field first introduced in 1.11. Let $\varphi : M \to N$ be a smooth manifold mapping and $X \in \mathscr{X}(M)$. If $g : N \to \mathbf{R}$ is a smooth function, the push-forward of X is a vector field $Y = \varphi_* X \in \mathscr{X}(N)$ defined by

$$(\varphi_* X)(g) = X(g \circ \varphi).$$

More precisely, if $p \in M$,

$$\begin{aligned} Y(g)(\varphi(p)) &= [X(g \circ \varphi)](p), \\ &= X_p(g \circ \varphi)). \end{aligned}$$

so that,

$$Y(g) \circ \varphi = X(g \circ \varphi).. \tag{7.8}$$

Suppose that in addition, φ is a diffeomorphism, let ξ_t be a one-parameter group associated with a vector field $X \in \mathscr{X}(M)$. Then we can push-forward to a one-parameter subgroup ψ_t associated with $\varphi_* X$ in N given by

$$\psi_t = \varphi \circ \xi_t \circ \varphi^{-1}, \tag{7.9}$$

as illustrated in the commuting diagram,

$$\begin{array}{ccc} M & \xrightarrow{\varphi} & N \\ \xi_t \downarrow & & \psi_t \downarrow \\ M & \xrightarrow{\varphi} & N. \end{array}$$

In other words, under a diffeomorphism φ, the integral curves of a vector field X are mapped to the integral curves of $\varphi_* X$.

Diffeomorphisms also allow us to use the inverse of the push-forward to pullback vectors, and push-forward functions with the inverse of the pullback. Specifically, if $f \in \mathscr{F}(M)$ and $Y \in \mathscr{X}(N)$,

$$\varphi^* Y(f) = (\varphi^{-1})_* Y(f) = Y(f \circ \varphi^{-1}), \tag{7.10}$$

$$\varphi_* f = f \circ \varphi^{-1} \tag{7.11}$$

which we can write,

$$\varphi^* Y(f) = Y(\varphi_* f). \tag{7.12}$$

When φ is a diffeomorphism, one may extend the pullback to tensor fields. Suppose M is an n-dimensional manifold. Since M locally looks like $\mathbf{R}^n$, on a coordinate patch on a neighborhood of point p we can pick an orthonormal basis $\{e_1 \ldots, e_n\}$ for the tangent space, with dual basis $\{\theta^1, \ldots, \theta^n\}$. Referring back to section 2.2.1, a tensor field $T \in \mathscr{T}^r_s(M)$ is a section of the bundle of tensor products of the tangent and cotangent spaces. In the given basis the tensor is an expression of the form,

$$T = T^{i_1 \ldots i_r}_{j_1 \ldots j_s}(e_{i_1} \otimes, \cdots \otimes e_{i_r} \otimes \theta^{j_1} \otimes \cdots \otimes \theta^{j_s}) \tag{7.13}$$

The tensor components are defined as,

$$T^{i_1 \ldots i_r}_{j_1 \ldots j_s} = T(\theta^{i_1}, \ldots, \theta^{i_r}, e_{j_1}, \ldots e_{j_s}). \tag{7.14}$$

Let t be a tensor field on N. The generalization of the pull-back 2.68 is given by,

$$(\varphi^* t)_p(\theta^{i_1}, \ldots, \theta^{i_r}, e_{j_1}, \ldots e_{j_s}) = t_{\varphi(p)}(\varphi_*^{-1}\theta^{i_1}, \ldots, \varphi_*^{-1}\theta^{i_r}, \varphi_* e_{j_1}, \ldots \varphi_* e_{j_s}) \tag{7.15}$$

Once again, the fancy equation can be demystified as just a generalized version of the chain rule. If the diffeomorphism is given in local coordinates by $y^\mu = f^\mu(x^\nu)$, so that $e_k = \partial/\partial x^k$ and $\theta^k = dx^k$, equation 7.15 is the classical transformation law for tensors,

$$(\varphi^* t)^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \frac{\partial y^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial y^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial y^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial y^{j_s}} \, t^{k_1 \ldots k_r}_{l_1 \ldots l_s}. \tag{7.16}$$

The matrices $\partial x^\nu/\partial y^\mu$ are allowed because φ is a diffeomorphism, so the Jacobians are invertible. Of course, we can pull-back (or push-forward) vectors and tensors if ϕ is a local diffeomorphism of M into itself.

Perhaps this is an appropriate time to generalize the coordinate-free exterior derivative formula 6.28 to arbitrary forms. Let $\Lambda^k(M)$ be the bundle of alternating covariant tensors of rank k and denote the sections of the bundle by $\Omega^k(M)$. As in 2.3, sections of this bundle are called k-forms on M.

7.1.5 Definition Let ω be a k-form on M, and let $\{X_1, \ldots X_{k+1}\} \in \mathscr{X}(M)$. The *exterior derivative* of ω is the $(k+1)$-form given by, (See, for example, Spivak [34])

$$\begin{aligned} d\omega(X_1, \ldots X_{k+1}) = &\sum_{i=1}^{k+1} (-1)^{i+1} X_i(\omega(X_1, \ldots, \widehat{X_i}, \ldots, X_{k+1}) \\ &\sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_{k+1}). \end{aligned} \tag{7.17}$$

where the "hats" mean that these vectors are excluded. Thus, for a 2-form ω, the formula gives

$$\begin{aligned} d\omega(X, Y, Z) = &X(\omega(Y, Z)) - Y(\omega(X, Z)) + Z(\omega(X, Y)) \\ &- \omega([X, Y], Z) + \omega([X, Z], Y) - \omega([Y, Z], X). \end{aligned}$$

If one chooses local coordinates as in section 2.3, we find that the operator is consistent with 2.65, and satisfies the properties in 2.66 and 2.69. In particular, the following holds:

7.1.6 Theorem Let M, N be a manifolds, $\alpha \in \Omega^k(M)$, $\beta \in \Omega^l(M)$, and $\varphi : M \to N$ be a diffeomorphism as above. Then

a. $d \circ d = 0$,
b. $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta$,
c. $\varphi^*(\alpha \wedge \beta) = \varphi^*\alpha \wedge \varphi^*\beta$,
d. $\varphi^*(d\,\omega) = d(\varphi^*\omega)$.

7.1.2 Lie Derivatives

Let φ_t be the one-parameter group of diffeomorphisms generated by the integral curves of a vector field X. We can use the pullback and the push-forward of φ_t to define a rate of change of functions, forms and vector fields in the direction of X. If $f \in \mathscr{F}(M)$ is a smooth function on M, and $p \in M$ is a point on M, we define the Lie derivative of f with respect to X by

$$\begin{aligned} \pounds_{X_p} f &\equiv \frac{d}{dt}(\varphi_t^* f)\bigg|_{t=0} \\ &= \lim_{t\to 0} \frac{f \circ \varphi_t(p) - f(p)}{t}, \\ &= X_p(f), && (7.18) \\ &= df(X)|_p. && (7.19) \end{aligned}$$

which as expected, is just the directional derivative. If $\omega \in \Omega^1(M)$ is a one form on M, we define the Lie derivative of ω at p by

$$\begin{aligned} (\pounds_X \omega)(p) &\equiv \frac{d}{dt}(\varphi_t^* \omega)\bigg|_{t=0}, \\ &= \lim_{t\to 0} \frac{(\varphi_t^*\omega)(p) - \omega(p)}{t}. \end{aligned}$$

This certainly has the right flavor of a derivative, namely, we pullback the form from a nearby point, compute the difference quotient, and then measure the infinitesimal change by evaluating the limit as t goes to 0. We will compute a formula for $\pounds_X\omega$ a little later in this section. In a similar manner, if $Y \in \mathscr{X}(M)$ is another vector field in M, we define its Lie derivative along X by using the push-forward

$$\pounds_X Y = \lim_{t\to 0} \frac{Y_p - (\varphi_{t*}Y)_p}{t}$$

where $(\varphi_{t*}Y)_p = \varphi_{t*}(Y_{\phi_t^{-1}(p)})$. That is, take the vector Y at $\varphi_t^{-1}(p) = \varphi_{-t}(p)$, push it forward to p and compare the infinitesimal change with Y_p, as shown in figure 7.2. Since φ_t is a diffeomorphism, we can pullback vectors by the inverse of the push-forward, so we could equivalently define $\pounds_X Y$ in manner that looks

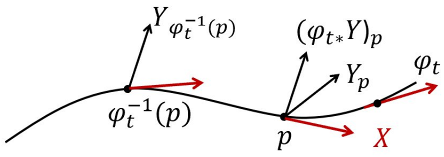


Fig. 7.2: Lie Derivative

more like the definition for functions and forms

$$\mathcal{L}_{X_p} Y = \frac{d}{dt}(\varphi_t^* Y)\bigg|_{t=0} \tag{7.20}$$
$$= \frac{d}{dt}(\varphi_t^{-1})_* Y\bigg|_{t=0} \tag{7.21}$$
$$= \lim_{t\to 0} \frac{(\varphi_t^{-1})_* Y_{\varphi_t(p)} - Y_p}{t} \tag{7.22}$$

In fact, since formula 7.15 shows that one can pull-back tensors in the case of a diffeomorphism, the Lie derivative can be extended to a linear derivation on the full tensor algebra.

7.1.7 Definition Let X be a vector field in M, and $\varphi_t(p)$ be the one-parameter family of diffeomorphisms generated by X, let $T \in \mathscr{T}_s^r$ be a tensor field. The *Lie derivative of the tensor* T with respect to X at p is defined as,

$$\mathcal{L}_{X_p} T = \frac{d}{dt}(\varphi_t^* T)\bigg|_{t=0} \qquad \text{or,}$$
$$\frac{d}{dt}\phi_t^* T = \phi_t^* \mathcal{L}_X T. \tag{7.23}$$

That the second version of the definition follows from the first, can easily be established by a quick computation

$$\frac{d}{dt}\phi_t^* T = \frac{d}{ds}(\phi_{s+t}^* T)\bigg|_{s=0} = \frac{d}{ds}(\phi_t^* \phi_s^* T)\bigg|_{s=0} = \phi_t^* \mathcal{L}_X T$$

The operator $\mathcal{L}_X : \mathscr{T}_s^r \to \mathscr{T}_s^r$ is clearly linear and satisfies Leibnitz rule. We have the following important theorem,

7.1.8 Theorem $\mathcal{L}_X Y = [X, Y]$.
Proof Consider the function $f \circ \varphi_t$. The Taylor expansion about the point p gives

$$f \circ \varphi_t = (f \circ \varphi_t(p))(0) + t\, \frac{d}{dt}[f \circ \varphi_t(p)]\bigg|_{t=0} + \mathscr{O}(t^2),$$
$$= f(p) + tX_p + \mathscr{O}(t^2).$$

Let $f \in \mathscr{F}(M)$. From the definition of the Lie derivative, we have

$$\pounds_X Y = \lim_{t\to 0} \frac{Y_p - (\varphi_t)_* Y}{t}$$

Applying this to f at p, we get,

$$\begin{aligned}
\pounds_{X_p} Y(f) &= \lim_{t\to 0} \left.\frac{Y - (\varphi_t)_* Y}{t}\right|_p (f), \\
&= \lim_{t\to 0} \frac{Y_p(f) - Y_{\varphi_t^{-1}(p)}(f \circ \varphi_t)}{t}, \\
&= \lim_{t\to 0} \frac{Y_p(f) - Y_{\varphi_t^{-1}(p)}(f(p) + tX_p)}{t}, \\
&= \lim_{t\to 0} \left[\frac{Y_p(f) - Y_{\varphi_t^{-1}(p)}(f)}{t} - Y_{\varphi_t^{-1}(p)}(X)(f)\right], \\
&= X_p(Y(f)) - Y_p(X(f)), \\
&= [X,Y]_p(f)
\end{aligned}$$

7.1.9 Theorem If $\varphi : M \to N$ is a diffeomorphism and $f \in \mathscr{F}(M)$ is any smooth function on M, then,

$$\pounds_{\varphi_* X}(\varphi_* f) = \varphi_*(\pounds_X f) \tag{7.24}$$

Proof

$$\begin{aligned}
\pounds_{\varphi_* X}(\varphi_* f)|_{\varphi(p)} &= \varphi_* X|_{\varphi(p)}(f \circ \varphi^{-1}), \\
&= X(f \circ \varphi^{-1} \circ \varphi)(p), \\
&= X(f)(p), \\
&= X(f) \circ \varphi^{-1}|_{\varphi(p)}, \\
\pounds_{\varphi_* X}(\varphi_* f) &= \varphi_*(\pounds_X f).
\end{aligned}$$

7.1.10 Theorem If $\varphi : M \to N$ is a diffeomorphism and $X, Y \in \mathscr{X}(M)$ are vector fields on M, then,

$$\begin{aligned}
\pounds_{\varphi_* X}(\varphi_* Y) &= \varphi_*(\pounds_X Y); \quad \text{that is,} \\
[\varphi_* X, \varphi_* Y] &= \varphi_*[X,Y].
\end{aligned} \tag{7.25}$$

Proof Let $g \in \mathscr{F}(N)$. By equation 7.8, we have to show that $[\varphi_* X, \varphi_* Y](g) \circ \varphi = [X,Y](g \circ \varphi)$. We have,

$$\begin{aligned}
[\varphi_* X, \varphi_* Y](g) \circ \varphi &= [\varphi_* X, \varphi_* Y], \\
&= (\varphi_* X(\varphi_* Y(g)) \circ \varphi - (\varphi_* Y(\varphi_* X(g)) \circ \varphi, \\
&= X(\varphi_* Y(g) \circ \varphi) - Y(\varphi_* X(g) \circ \varphi), \\
&= X(Y(g \circ \varphi)) - Y(X(g \circ \varphi)), \\
&= [X,Y](g \circ \varphi).
\end{aligned}$$

The Lie derivative satisfies the following,

7.1.11 Properties. Let $f \in \mathscr{F}$, $X, Y \in \mathscr{X}$, and T_1, T_2 be tensors. Then
a) $\pounds_X f = X(f)$,
b) $\pounds_X Y = [X, Y]$,
c) $\pounds_X(fY) = X(f)Y + f\pounds_X Y$,
d) $\pounds_X(T_1 \otimes T_2) = \pounds_X T_1 \otimes T_2 + T_1 \otimes \pounds_X T_2$.
e) $\pounds_X(C(T)) = C(\pounds_X T)$, where $C : \mathscr{T}^r_s \to \mathscr{T}^{r-1}_{s-1}$ is a contraction (See **??**).
So, if ω is a one-form, we have,

$$\pounds_X \langle\omega|Y\rangle = \langle \pounds_X \omega|Y\rangle + \langle\omega|\pounds_X Y\rangle.$$

Consequently, the Lie derivative of a one form is given by,

$$\begin{aligned} \langle \pounds_X \omega|Y\rangle &= \pounds_X\langle\omega|Y\rangle - \langle\omega, \pounds_X Y\rangle, \\ \pounds_X\omega(Y) &= \pounds_X(\omega(Y)) - \omega(\pounds_X Y), \\ &= X(\omega(Y)) - \omega([X, Y]) \end{aligned} \tag{7.26}$$

Now that we know the Lie derivative of functions, vector fields, and one forms, it is a straight-forward exercise to use induction on tensor products, to find the formula for the Lie derivative of any tensor field $T \in \mathscr{T}^r_s$. The formula is,

$$\begin{aligned} \pounds_X[T(\omega^1, \ldots, \omega^r, X_1, \ldots X_s)] = {} & \pounds_X T(\omega^1, \ldots \omega^r, X_1, \ldots X_s,) \\ & + \sum_{i=1}^{r} T(\omega^1, \ldots, \pounds_X\omega^i, \ldots, \omega^r, X_1, \ldots, X_s) \\ & + \sum_{i=1}^{s} T(\omega^1, \ldots, \omega^r, X_1, \ldots, \pounds_X X_i, \ldots X_s,). \end{aligned} \tag{7.27}$$

If we set $X = X^k\partial_k$, the formula in component form reads

$$\begin{aligned} \pounds_X T^{i_1\ldots i_r}_{j_1\ldots j_s} = {} & X^k\partial_k T^{i_1\ldots i_r}_{j_1\ldots j_s} \\ & + (\partial_k X^{i_1}) T^{k\, i_2\ldots i_r}_{j_1 j_2\ldots j_s} + (\partial_k X^{i_2}) T^{i_1 k\, i_3\ldots i_r}_{j_1 j_2 j_3\ldots j_s} + \ldots \\ & - (\partial_{j_1} X^k) T^{i_1 i_2\ldots i_r}_{k\, j_2\ldots j_s} - (\partial_{j_2} X^k) T^{i_1 i_2 i_3\ldots i_r}_{j_1 k\, j_3\ldots j_s} - \ldots . \end{aligned} \tag{7.28}$$

In a Riemannian manifold $\{M, g\}$ with Levi-Civita connection ∇, the formula above for the components of the Lie derivative is not manifestly covariant, but it becomes so by replacing the ∂_k's by the covariant derivative ∇_k. That is,

$$\begin{aligned} \pounds_X T^{i_1\ldots i_r}_{j_1\ldots j_s} = {} & X^k\nabla_k T^{i_1\ldots i_r}_{j_1\ldots j_s} \\ & + (\nabla_k X^{i_1}) T^{k\, i_2\ldots i_r}_{j_1 j_2\ldots j_s} + (\nabla_k X^{i_2}) T^{i_1 k\, i_3\ldots i_r}_{j_1 j_2 j_3\ldots j_s} + \ldots \\ & - (\nabla_{j_1} X^k) T^{i_1 i_2\ldots i_r}_{k\, j_2\ldots j_s} - (\nabla_{j_2} X^k) T^{i_1 i_2 i_3\ldots i_r}_{j_1 k\, j_3\ldots j_s} - \ldots . \end{aligned} \tag{7.29}$$

One can verify directly, that all the extra terms with connection coefficients cancel out, and the formula reduces to the previous one. If the components of the Riemannian metric are given by $g_{\mu\nu} = g(\partial_\mu, \partial_\nu)$, and $X = X^\sigma\partial_\sigma$, we have

$$\begin{aligned}\pounds_X g_{\mu\nu} &= X^\sigma\nabla_\sigma g_{\mu\nu} + (\nabla_\mu X^\sigma)g_{\sigma\nu} + (\nabla_\nu X^\sigma)g_{\nu\sigma},\\ &= \nabla_\mu X_\nu + \nabla_\nu X_\mu.\end{aligned}$$

A vector field X that satisfies the equation

$$\pounds_X g = 0 \tag{7.30}$$

is called a *Killing vector.* If φ_t is the one-parameter subgroup corresponding to the flow of a Killing vector, the solutions of the Killing equation

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0 \tag{7.31}$$

represent isometries of the manifold. In Minkowski space, the Killing vector fields correspond to the generators of the Lorentz group discussed in section 8.2. Let $\alpha(t)$ be a geodesic in the manifold with velocity vector V given in local coordinates by $V = V^\mu\partial_\mu = \dot{x}^\mu(t)\partial_\mu$, and suppose that $X = X^\nu\partial_\nu$ is a Killing vector, then,

$$\begin{aligned}\nabla_V(V^\mu X_\mu) &= V^\nu\nabla_\nu(V^\mu X_\mu),\\ &= V^\nu V^\mu\nabla_\nu X_\mu + V^\nu X^\mu\nabla_\nu V_\mu,\\ &= \tfrac{1}{2}V^\nu V^\mu(\nabla_\nu X_\mu + \nabla_\mu X_\nu) + X^\mu V^\nu\nabla_\nu V_\mu,\\ &= 0\end{aligned}$$

The first term vanishes because X is a Killing vector and the second because V is geodesic, so that $\nabla_V V = 0$. Thus, the metric $< V, X >= V^\mu X_\mu$ is a conserved quantity along the geodesic. Roughly speaking, the conserved quantity associated with the local isometry is the momentum, which makes sense, since a free particle travelling along a geodesic is not subjected to external forces.

7.1.12 Theorem If $X \in \mathscr{X}(M)$ and $\omega \in \Omega^k(M)$, $\eta \in \Omega^l(M)$, then,

$$d\pounds_X\omega = \pounds_X d\omega, \tag{7.32}$$

$$\pounds_X(\omega\wedge\eta) = \pounds_X\omega \wedge \pounds_X\eta. \tag{7.33}$$

Proof Let φ_t be the one-parameter flow of X. By definition,

$$\pounds_X\omega(p) = \frac{d}{dt}(\varphi_t^*\omega(p))|_{t=0}.$$

The theorem follows immediately from the fact that d is linear and so, it commutes with d/dt, plus the already established formulas $\varphi_t^* d\omega = d\varphi_t^*\omega$, and $\varphi_t^*(\omega\wedge\eta) = \varphi_t^*\omega\wedge\varphi_t^*\eta$.

We recall the definition if the interior product 2.23. Let M be a manifold, $X, X_1 \ldots, K_k \in \mathscr{X}(M)$ and $\omega \in \Omega^{k+1}(M)$, then,

$$i_X\omega(X_1, \ldots, X_k) = \omega(X, X_1, \ldots, X_k). \tag{7.34}$$

By convention, we set $i_X f = 0$. If ω is a one-form, $i_X\omega = \omega(X)$.

7.1.13 Theorem Let $\varphi : M \to N$ be a diffeomorphism. Then,

$$\varphi^*(i_X\omega) = i_{\varphi^* X}\varphi^*\omega. \tag{7.35}$$

Proof Let $\{Y = \varphi^* X,\ Y_i = \varphi^* X_i, i = 1 \ldots k\} \in \mathscr{X}(M)$. We have

$$\begin{aligned}
i_Y\varphi^*\omega(Y_1,\ldots,Y_k) &= \varphi^*\omega(Y, Y_1,\ldots,Y_k),\\
&= \omega(\varphi_* Y, \varphi_* Y_1,\ldots,\varphi_* XY_k),\\
&= \omega(X, X_1,\ldots X_k),\\
&= i_X\omega(X_1,\ldots,X_k)\\
i_{\varphi * X}\varphi^*\omega(Y_1,\ldots,Y_k) &= \varphi^* i_X\omega(Y_1,\ldots,Y_k),
\end{aligned}$$

If φ_i is a local diffeomorphism on M the theorem can be restated as,

$$\varphi_t^* \circ i_X = i_{\varphi_t^* X} \circ \varphi_t^*. \tag{7.36}$$

7.1.14 Theorem Let $i_X : \Omega^k(M) \to \Omega^{k-1}(M)$, be the interior product and let $\alpha \in \Omega^k(M),\ \beta \in \Omega^l(M),\ f \in \mathscr{F}(M)$. Then

a) $i_X(\alpha\wedge\beta) = i_X\alpha\wedge\beta + (-1)^k\alpha\wedge i_X\beta$,

b) $i_{fX}\alpha = f i_X\alpha$,

c) $i_X df = \pounds_X f = X(f)$

Proof The proof of part (a) involves some combinatorics arising from the definition of the wedge product 2.61. We leave out the somewhat messy details. Part (b) follows immediately from the multilinearity of α and the definition of the interior product. Part (c) is trivial. We have $i_X df = df(X) = X(f)$.

The interior product 7.34, the intrinsic exterior derivative 7.17, and the Lie derivative 7.27 are related by the following formula.

7.1.15 Theorem (H. Cartan)

$$d\circ i_X + i_X\circ d = \pounds_X. \tag{7.37}$$

Proof The proof is by induction. First, we verify that the formula is true for zero-form f and a 1-form ω.

$$\begin{aligned}
(d\,i_X + i_X d)f &= d\,i_X f + i_X df,\\
&= i_X df,\\
&= df(X) = X(f) = \pounds_X f.
\end{aligned}$$

Let ω be a one form and $X, Y \in \mathscr{X}(M)$.

$$\begin{aligned}
(d\, i_X + i_X d)\, \omega(Y) &= (d\,(i_X\omega))(Y)) + i_X(d\omega)(Y), \\
&= d(\omega(X))(Y) + d\omega(X,Y), \\
&= Y(\omega(X)) + X(\omega(y)) - Y(\omega(X)) - \omega([X,Y]), \\
&= X(\omega(Y)) - \omega(\pounds_X Y), \\
&= \pounds_X \omega(Y).
\end{aligned}$$

Suppose the proposition is true for k-forms ω_i. A $k+1$-form can be written as $df_i \wedge \omega_i$ that for simplicity we write as $df \wedge \omega$. By induction hypothesis,

$$d\, i_X \omega + i_X d\, \omega = \pounds_X \omega$$

Using Leibnitz rule, and the properties $d \circ d = 0$, $i_X f = 0$, $i_X df = \pounds_X f$, we compute,

$$\begin{aligned}
(d\, i_X + i_X d)(df \wedge \omega) &= d\, i_X(df \wedge \omega) + i_X d(df \wedge \omega)), \\
&= d(i_X df \wedge \omega - df \wedge i_X \omega) - i_X(df \wedge d\omega), \\
&= d\, \pounds_X f \wedge \omega + \pounds_X f \wedge d\omega - (-df \wedge d\, i_X \omega)) \\
&\qquad - (i_X df \wedge d\omega - df \wedge i_X d\omega), \\
&= d\, \pounds_X f \wedge \omega + \pounds_X f \wedge d\omega - \pounds_X f \wedge d\omega + df \wedge (d\, i_X \omega + i_X\, d\omega) \\
&= \pounds_X df \wedge \omega + df \wedge \pounds_X \omega, \quad \text{(by induction and } \pounds_X\, d = d\, \pounds_X\text{)}, \\
&= \pounds_X (df \wedge \omega),
\end{aligned}$$

which is what we wanted to establish. The diagram in figure 7.3, which is reminiscent of chain-complexes in singular homology, helps to visualize this most elegant result, sometimes called Cartan's magic formula.

$$\begin{tikzcd}
\cdots \arrow[r, "d"] & \Omega^{k-1} \arrow[r, "d"] \arrow[d, "\pounds_X"'] & \Omega^{k} \arrow[r, "d"] \arrow[d, "\pounds_X"'] \arrow[dl, "i_X"'] & \Omega^{k+1} \arrow[r, "d"] \arrow[d, "\pounds_X"'] \arrow[dl, "i_X"'] & \cdots \\
\cdots \arrow[r, "d"] & \Omega^{k-1} \arrow[r, "d"] & \Omega^{k} \arrow[r, "d"] & \Omega^{k+1} \arrow[r, "d"] & \cdots
\end{tikzcd}$$

Fig. 7.3: Cartan's Magic Formula

Cartan's magic formula is useful in establishing the Poincaré lemma. We recall from example 2.82, that a closed form need not be exact. In more general spaces, such as spheres, there are topological considerations. In fact, Consider the de-Rham complex,

$$\cdots \xrightarrow{\;d\;} \Omega^{k-1}(M) \xrightarrow{\;d\;} \Omega^{k}(M) \xrightarrow{\;d\;} \Omega^{k+1}(M) \ldots \tag{7.38}$$

In algebraic topology, a sequence such as this one, for which $d \circ d = 0$, is called a long exact sequence. If one lets $Z^k(M)$ be the set of closed k-forms on a manifold M, and $B^k(M)$ be the set of exact forms, the quotient

$$H^k(M) = Z^k(M)/B^k(M)$$

is called the k-th *cohomology* group $H^k(M)$ of that space. Two closed k-forms ω and ω' are in the same cohomology class if their difference is exact; that is, there exists a $(k-1)$-form ϕ, such that

$$\omega' = \omega + d\phi$$

The de-Rham cohomology groups have deep connections to the topology of the space. A key topological concept we need is that of a homotopy which we define as follows

7.1.16 Definition Lef M' and M be smooth manifolds. Two smooth maps $f, g : M' \to M$ are called *homotopic* if there exists a map $\phi : M' \times [0,1] \to M$, such that

a) $\phi(p', 0) = f(p)$,

b) $\phi(p', 1) = g(p)$.

If we let $\phi_t(p') = \phi(p', t)$, $t \in [0,1]$ then the homotopy describes a smooth deformation of $g = \phi_1$ to $f = \phi_0$. A manifold M is *contractible*, to a point $p_0 \in M$, if there exists a homotopy

$$\phi : M \times [0,1] \to M$$

for which $\phi_1(p) = g(p) = p$ is the identity map, and $\phi_0(p) = f(p) = p_0$ is the constant map.

7.1.17 *Poincaré Lemma*

If M is a manifold which is smoothly contractible to a point, a closed form ω is exact.

Proof We present a proof in the special case in which the manifold is a ball $B^n \in \mathbf{R}^{n+1}$ centered at the origin. This is the alternative proof that appears in Abraham-Marsden [20]. For all $p \in B^n$, and $0 < t \leq 1$, let ϕ_t be the one-parameter group of diffeomorphisms defined by $\phi_t(p) = tp$. This kind of homotopy map is an example of a *deformation retract.* When $t = 1$, the map is the identity map and as $t \to 0$, the ball continuously shrinks to a point. Let X_t be the vector field $X_t(p) = p/t$. We have,

$$\frac{d}{dt}\phi_t(p) = X_t(\phi(p)) = p,$$

so X_t is the tangent vector field of the one-parameter family of curves. Let ω be a closed k-form in B^n. By the definition of the Lie derivative and Cartan's magic formula 7.37, we have

$$\begin{aligned}\frac{d}{dt}(\phi_t^*\omega) &= \phi_t^*(\pounds_{X_t}\omega),\\ &= \phi_t^*(d i_{X_t}\omega + i_{X_t} d\omega),\\ &= \phi_t^*(d i_{X_t}\omega),\end{aligned}$$

Integrating from a small value $\epsilon > 0$ to 1, we get,

$$\begin{aligned}\phi_t^*\omega|_\epsilon^1 &= \int_\epsilon^1 \phi_t^*(di_{X_t}\omega)\,dt,\\ &= d\int_\epsilon^1 \phi_t^*(i_{X_t}\omega)\,dt,\end{aligned}$$

Taking the limit as $\epsilon \to 0$ and recalling that at $t = 1$, ϕ_t is the identity map, we get

$$\omega = d\beta, \qquad \beta = \int_0^1 \phi_t^*(i_{X_t}\omega)\,dt. \tag{7.39}$$

Although the theorem is proved, it is helpful to find a more explicit formula for the $(k-1)$-form β, using the definitions of the push-forward and the interior product. Let $\{e_i, \ldots, e_{k-1}\}$ be part of a set of basis vectors and p the position vector at the point p. Then

$$\begin{aligned}\beta_p(e_1,\ldots,e_{k-1}) &= \int_0^1 \phi_t^* i_p\omega(e_i,\ldots e_{k-1})\,dt,\\ &= \int_0^1 \omega_{\phi(p)}(p,\phi_*e_i\ldots,\phi_*e_{k-1})\,dt,\\ &= \int_0^1 \omega_{tp}(p,te_1,\ldots,te_{k-1})\,dt,\end{aligned}$$

so,

$$\beta_p = \int_0^1 t^{k-1}\omega_{tp}(p,e_1,\ldots e_{k-1})\,dt. \tag{7.40}$$

The more general theorem is computationally more complicated, but the idea is essentially the same. There is a natural one parameter group of diffeomorphisms $\phi_t : M \times [0,1]$ such that ϕ_1 is the identity and ϕ_0 is a constant map. One then seeks a linear map $h : \Omega^k \to \Omega^{k-1}$ such that [1],

$$\phi_1^*\omega - \phi_0^*\omega = d(h\omega) + h(d\omega). \tag{7.41}$$

This property can be represented by the diagram 7.4 which is an example of a *chain homotopy*. The linear map we seek can be obtained by defining $h\omega(p) = \beta_p$. By choice of the homotopy map, the left-hand-side of equation 7.41 is the identity map on forms. By a direct, but non-trivial computation (see [20], [34]), one can verify that the right-hand-side of equation is also equal to ω. Thus, if $d\omega = 0$ we have

$$\omega = d(h\omega)$$

which is what we wanted to establish.

This general form of the theorem is rough for the novice. In Abraham-Marsden, the form $h\omega = \beta$ just pops out of nowhere, so without the alternative

[1] It is common to index the maps with the order of the forms. For example, one would write $h^k : \Omega^k \to \Omega^{k-1}$, and $d^k : \Omega^k \to \Omega^{k+1}$.

$$\begin{tikzcd}
\cdots \arrow[r, "d"] & \Omega^{k-1} \arrow[r, "d"] \arrow[d] & \Omega^{k} \arrow[r, "d"] \arrow[d, "\phi_1^*-\phi_0^*"'] \arrow[dl, "h"'] & \Omega^{k+1} \arrow[r, "d"] \arrow[d] \arrow[dl, "h"'] & \cdots \\
\cdots \arrow[r, "d"] & \Omega^{k-1} \arrow[r, "d"] & \Omega^{k} \arrow[r, "d"] & \Omega^{k+1} \arrow[r, "d"] & \cdots
\end{tikzcd}$$

Fig. 7.4: Chain Homotopy

proof, the whole process appears rather mysterious. Spivak provides some motivation for finding $h\omega$ by first treating the case where ω is a one form. In his book Calculus on Manifolds he carries out the full computation when M is a star-shaped region in $\mathbf{R}^n$. Either way, the computation is more difficult. The Poincaré lemma is often stated by saying that in a manifold M, a closed form is locally exact; that is, given a closed form on an open set $U \subset M$, then for each point $p \in U$, there exists a ball $B \subset M$ centered at p in which the form is exact. Perhaps an explicit construction of the form $h\omega$ in $\mathbf{R}^3$ might help the reader understand the nature of the constructive proof. Let $\mathbf{B}$ be a vector field with $\nabla \cdot \mathbf{B} = 0$ and $p = (x, y, x)$. We seek a vector potential $\mathbf{A}$, such that $\nabla \times \mathbf{A} = \mathbf{B}$. Write p as a tangent vector $p = x^k \partial_k$, and map $\mathbf{B}$ into the 2-form

$$\omega = B_1(p)\, dx^2 \wedge dx^3 - B_2(p)\, dx^1 \wedge dx^3 + B_3(p)\, dx^1 \wedge dx^2.$$

The components of the 1-form $\alpha = A'_k\, dx^k$ constructed in the proof of the Poincaré lemma are

$$\begin{aligned}
A'_i &= h\omega(p)(\partial x^i), \\
&= \int_0^1 t\omega_{tp}(x^k\partial_k, \partial_i)\, dt, \\
&= \int_0^1 t[B_1(tp)\, dx^2 \wedge dx^3 - B_2(tp)\, dx^1 \wedge dx^3 + B_3(tp)\, dx^1 \wedge dx^2](x^k\partial_k, \partial_i)\, dt.
\end{aligned}$$

Since $dx^k(\partial_i) = \delta^k_i$, we get

$$\begin{aligned}
A'_1 &= \int_0^1 t[x^3 B_2(tp) - x^2 B_3(tp)]\, dt, \\
A'_2 &= \int_0^1 t[x^1 B_3(tp) - x^3 B_1(tp)]\, dt, \\
A'_3 &= \int_0^1 t[x^2 B_1(tp) - x^1 B_2(tp)]\, dt,
\end{aligned}$$

For an example let $\mathbf{B} = (y, -z^2, -x)$ and see if we can recover the vector potential $\mathbf{A} = (xy, -yz, xz^2)$ which we used secretly to produce the field. The computation yields

$$
\begin{aligned}
A'_1 &= \int_0^1 [tz(-t^2z^2) - ty(-tx)]\,dt = -\tfrac{1}{4}z^3 + \tfrac{1}{3}xy,\\
A'_2 &= \int_0^1 [ty(ty) - tx(-t^2z^2)]\,dt = -\tfrac{1}{3}y^2 + \tfrac{1}{4}xx^2,\\
A'_3 &= \int_0^1 [ty(ty) - tx(-t^2z^2)]\,dt = \tfrac{1}{3}y^2 + \tfrac{1}{4}xz^2.
\end{aligned}
$$

One can easily verify the curl of the resulting vector potential

$$\mathbf{A}' = (-\tfrac{1}{4}z^3 + \tfrac{1}{3}xy, -\tfrac{1}{3}y^2 + \tfrac{1}{4}xx^2, \tfrac{1}{3}y^2 + \tfrac{1}{4}xz^2)$$

does indeed give the same field $\mathbf{B}$, but we failed to recover the original potential. On the other hand, the difference

$$\mathbf{A} - \mathbf{A}' = \nabla f, \quad \text{where } \ f = \tfrac{1}{3}x^2y + \tfrac{1}{4}xz^3 - \tfrac{1}{3}z^2z,$$

so the two potentials are cohomologous, as expected. This academic example shows the cleverness of the definition of $h\omega$ but it also squelches the hope of an easy way out of solving Maxwell equations for magnetic fields. To obtain the vector potentials in the right gauge on problems of physical significance, students are much better off reading the Feynman Lectures on Physics.

7.1.18 Theorem

$$
\begin{aligned}
[\pounds_X, i_Y] &\equiv \pounds_X \circ\ i_Y - i_Y \circ \pounds_X = \ \ i_{[X,Y]}, &(7.42)\\
[\pounds_X, \pounds_Y] &\equiv \pounds_X \circ \pounds_Y - \pounds_Y \circ \pounds_X = \pounds_{[X,Y]}. &(7.43)
\end{aligned}
$$

Proof Let ω be a k-form and $Y, X, X_1, \ldots, X_{k-1}$ be vector fields. The proof is by direct computation using the formula for Lie derivatives 7.27 and the definition of the interior product 7.34.

$$
\begin{aligned}
&(i_Y \pounds_X \omega)(X_1, \ldots, X_{k-1})\\
&\quad = (\pounds_X\omega)(Y, X_1, \ldots, X_{k-1}),\\
&\quad = \pounds_X(\omega(Y, X_1, \ldots, X_{k-1})) - \sum_{i=1}^{k-1} \omega(Y, X_1, \ldots, [X, X_i], \ldots X_{k-1})\\
&\qquad\qquad - \omega([X,Y], X_1, \ldots x_{k-1}),\ \text{(term not included in sum)}\\
&\quad = \pounds_X i_Y(X_1, \ldots, X_{k-1}) + \sum_{i=1}^{k-1} i_Y(X_1, \ldots, [X, X_i], \ldots X_{k-1})\\
&\qquad\qquad - i_{[X,Y]}(X_1, \ldots, X_{k-1}),\\
&\quad = (\pounds_X i_Y - i_{[X,Y]})\omega(X_1, \ldots, X_{k-1}).
\end{aligned}
$$

We leave the second part as an exercise, using the formula of Cartan 7.37.

A direct consequence of equations 7.43 is that if X and Y are Killing vector fields in a Riemannian manifold $\{M, g\}$, so that $\pounds_X g = \pounds_Y g = 0$, then $[X, Y]$

is also a Killing vector field. Thus, the set of Killing vector fields forms a Lie subalgebra of the Lie algebra of vector fields, in the sense described in the section that follows.

7.2 Lie Algebras

7.2.1 Definition A vector space $\mathfrak{g}$ over a field F (here, $F = \mathbf{R}$ or $\mathbf{C}$) is called a *Lie algebra* if there exists an operation $[\ ,\ ] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ (called the Lie Bracket), such that,

a) $[\ ,\ ]$ is F-bilinear,

b) $[X, Y] = -[Y, X]$, for all $X, Y \in \mathfrak{g}$,

c) $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X.Y]] = 0$, *Jacobi identity*

7.2.2 Definition A *Lie subalgebra* $\mathfrak{h}$ of a Lie algebra $\mathfrak{g}$ is a subspace that is closed under the Lie Bracket.

7.2.3 Theorem The vector space $\mathscr{X}(M)$ together with the operation $\pounds_X Y = [X, Y]$ gives the space the structure of a Lie Algebra.
The Jacobi identity follows directly from equation 7.43 applied to vector fields, but a much more elementary proof follows directly from the definition of the bracket of two vector fields and the property of vector fields being linear derivations on the space of functions. The details of the proof have already appeared in theorem 4.4.2.

Lie subalgebras are intricately connected with the theory of submanifolds $N \subset M$. They can be used to generalized the idea of integral curves. Let M be an n dimensional manifold, and p be a point $p \in M$. A k-dimensional *distribution* at p is a subset $D_p \subset T_pM$. Let $\{X_1, X_2, \ldots, X_k\}$ be a set of linearly independent vector fields in a neighborhood U of p which constitutes a basis for $D_q,\ q \in U$. If these can be chosen in a smooth way, D is called a C^∞ distribution.

7.2.4 Definition Let $\{X_1, X_2, \ldots, X_k\} \mid p \in U\}$ span a distribution D. The distribution is *integrable* if the vectors form a subalgebra of the Lie algebra of vectors fields in M. That is, there exist C^∞ functions $C^l{}_{ij}$, such that

$$[X_i, X_j] = C^k{}_{ij} X_k.$$

7.2.5 Definition A distribution D arises from a *foliation* of M, if for each point $p \in M$, there exists a k-dimensional local submanifold N of M, containing p with

$$i_*(T_pN) = D_p, \text{ for al } p \in U$$

where $i : N \to M$ is the inclusion map.

Locally, foliations look like layers of local submanifolds, called the leaves of the foliation. Perhaps the most famous foliation is the Reeb foliation of S^3

that locally looks like onion layers formed a sock pushed infinitely into itself. A standard method to treat the field equations in general relativity is to think of a 4-dimensional Lorentzian manifold as a 3+1 manifold, foliated by 3-dimensional spatial surfaces evolving in time.

The main result on distributions is,

7.2.6 Theorem (Frobenius) A distribution D is integrable if and only if, it arises from a foliation.

An alternative formulation of the Frobenius integrability theorem can be stated in terms of differential forms. Let $\Omega^*(M)$ be the graded ring on smooth differential forms. The subring $\mathcal{I}(D)$ of all the forms ω that annihilate D, namely, for all $X_1, \dots X_k$ in D

$$\omega(X_1, \dots X_k) = 0, \tag{7.44}$$

generate an ideal $\mathcal{I}(D)$. If $\{e_1, \dots, e_k\}$ is an orthonormal frame in D, then the dual forms $\{\theta^1, \dots, \theta^k\}$ span $\mathcal{I}(D)$. The differential forms version of the Frobenius says that D is integrable if for every $\omega \in \mathcal{I}(D)$, we have $d\omega \in \mathcal{I}(D)$; that is, the differential ideal is closed under exterior derivatives. More specifically, there exist forms $\alpha^j{}_k$, such that

$$d\theta^j = \alpha^j{}_k \wedge \theta^k \in \mathcal{I}(D).$$

The connection between the two versions of the theorem is achieved through the structure constant formula,

$$d\theta^i = \frac{1}{2} C^i{}_{jk} \theta^j \wedge \theta^K$$

which we prove in equation 7.67.

If we extend the orthonormal frame spanning D to an orthonormal frame $\{e_1, \dots, e_k, e_{k+1}, \dots, e_n\}$, with dual forms $\{\theta^1, \dots, \theta^k, \theta^{k+1}, \dots, \theta^n\}$ then the integral submanifolds are defined by the (Pfaffian) system,

$$\theta^m = 0, \quad m = k+1, \dots n.$$

The theorem guarantees the existence of local coordinates $\{x^1, \dots, x^n\}$ about $p \in U$, with tangent vectors $\{\partial/\partial x^1, \dots \partial/\partial x^k\}|_p$ spanning D_p, and with the dual forms

$$\{dx^{k+1}, \dots, dx^n\}|_p$$

annihilating D. The forms $\{\theta^{k+1}, \dots, \theta^n\}$ can then be written as linear combinations of the coordinate one-forms above. The sets

$$N_p = \{x^{k+1}|_p = a^{k+1}, \dots, x^n|_p = a^n\},$$

where the a's are constants, are integral submanifolds of the distribution D_p.

We do not present a proof of this very important theorem in this rather perfunctory treatment of the topic. Instead, we refer the reader to classic textbooks such as [20] or [34]. The theorem is the starting point to the deep subject of Lie groups of symmetries of partial differential equations and prolongation theory.

Given any Lie group G, we can construct an associated Lie algebra $\mathfrak{g}$. Given a group element $g \in G$, we define the *left* and the *right translation* maps $L_g, R_g : G \to G$ by,

$$L_g(g_0) = gg_0, \tag{7.45}$$
$$R_g(g_0) = g_0 g, \quad \text{for all } g_0 \in G. \tag{7.46}$$

Hereafter, for every statement we make about left translation, there is a corresponding statement about right translation. The map L_g is a diffeomorphism with inverse given by $(L_g)^{-1} = L_{g^{-1}}$.

$$\begin{tikzcd}
\mathscr{X}(G) \arrow[r, "L_{g*}"] & \mathscr{X}(G) \\
G \arrow[u, "X"] \arrow[r, "L_g"] & G \arrow[u, "X"']
\end{tikzcd}$$

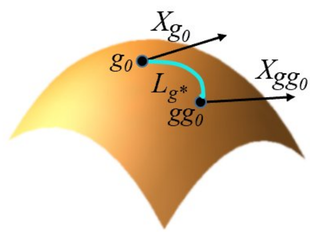


Fig. 7.5: Left Invariant Vector Field.

7.2.7 Definition A vector field $X \in \mathscr{X}(G)$ is called *left invariant* if for all $g \in G$, we have,

$$L_{g*}X = X. \tag{7.47}$$

More specifically, if X_{g_0} is the tangent vector at g_0, then,

$$L_{g*}X_{g_0} = X_{L_g(g_0)} = X_{gg_0},$$

is the tangent vector at gg_0.

7.2.8 Definition Let $\mathfrak{g} = L(G)$ be the set of all left-invariant vector fields in G. Then $\mathfrak{g}$ has the structure of a Lie algebra.
Proof Let $X, Y \in \mathfrak{g}$. Then by equation 7.25 we have,

$$\begin{aligned} L_{g*}[X,Y] &= [L_{g*}X, L_{g*}Y], \\ &= [X,Y], \end{aligned}$$

so $[X,Y] \in \mathfrak{g}$. Thus, the set of left-invariant vector fields is closed under Lie brackets and hence it is a Lie subalgebra of the Lie algebra of all vector fields $\mathscr{X}(G)$.

Let T_eG be the tangent space at the identity of a Lie Group. For every tangent vector $X_e \in T_eG$ we can generate a vector field X by simply defining the value of the vector field at any point $g \in G$, to be the tangent vector,

$$X_g = L_{g*} X_e. \tag{7.48}$$

The vector field X so defined is almost tautologically left invariant. Indeed, if $g_0 \in G$,

$$\begin{aligned} L_{g*} X_{g_0} &= L_{g*}(L_{g_0 *}) X_e, \\ &= L_{(g g_0)*} X_e \end{aligned}$$

For each left invariant vector field X with value X_g, there is a unique tangent vector $X_e = L_{(g^{-1})*} X_g$ at the identity, so we have the following theorem,

7.2.9 Theorem The tangent vector space at the identity T_eG of a lie group is isomorphic to the Lie algebra of left-invariant vector fields $\mathfrak{g} = L(G)$. The isomorphism is obtained by assigning to any left-invariant vector field, its value at the identity.

We now consider the behavior of left invariant vector fields under mappings. Let $\phi : G \to H$ be a homomorphism between two Lie groups G and H with identity elements e and e' respectively, and push-forward map $\phi_* : T_eG \to T_{e'}H$. Consider a tangent vector $X_e \in T_eG$ generating a left invariant vector field X.

$$\begin{tikzcd} X_e \in T_eG \arrow[r, "\phi_*"] \arrow[d] & T_{e'}H \arrow[d] \\ g \in G \arrow[r, "\phi"] \arrow[d, "L_g"] & H \arrow[d, "L_{\phi(g)}"] \\ G \arrow[r, "\phi"] & H \end{tikzcd}$$

Fig. 7.6: Lie Algebra Homomorphism

So, if $g \in G$, then $L_{g*}X = X$. Denote by Y the left invariant vector field in H whose value at the identity is $Y_{e'} = \phi_* X_e$. As shown in figure 7.6, we have,

$$\phi \circ L_g = L_{\phi(g)} \circ \phi.$$

Then, for the push-forward of the vector field X at g, we have,

$$\begin{aligned} \phi_* X_g &= \phi_* L_{g*} X_e, \\ &= (\phi \circ L_g)_* X_e, \\ &= (L_{\phi(g)} \circ \phi)_* X_e, \\ &= L_{\phi g *} \phi_* X_e, \\ &= L_{\phi g *} Y_{e'}, \\ &= Y_{\phi(g)}. \end{aligned}$$

Therefore, the push-forward of a left-invariant vector field X is a left-invariant vector field Y. Since the push-forward preserves brackets, the map ϕ_* is a *Lie algebra homomorphism*, that is,

$$\begin{aligned} \phi_*(aX^1 + bX^2) &= a\phi_* X^1 + b\phi_* X^2, \\ \phi_*[X^1, X^2] &= [\phi_* X^1, \phi_* X^2]. \end{aligned} \tag{7.49}$$

7.2.1 The Exponential Map

Let $\phi : \mathbf{R} \to G$ be a smooth Lie group homomorphism, and let $A = \phi'(0) \in T_eG$ be a tangent vector at the identity. Such homomorphism is called a *one-parameter subgroup* of G. Let $G = GL(n, \mathbf{R})$. Consider the case where $G = GL(n, \mathbf{R})$. This is a matrix group, so a tangent vector at the identity is a matrix; this is the reason why we changed the notation from X to A. Since

$$\phi(s+t) = \phi(s) \circ \phi(t),$$

evaluating the derivative at $s = 0$ gives,

$$\begin{aligned} \frac{d}{dt}\phi(t) &= \phi'(0) \cdot \phi(t), \\ &= A\phi(t), \quad \text{with,} \\ \phi(0) &= I. \end{aligned} \tag{7.50}$$

Here, the dot is matrix multiplication. By analogy to the one-dimensional case, the solution of this differential equation is,

$$\phi(t) = e^{At}, \tag{7.51}$$

where the exponential of a matrix is defined as the power series,

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots + \frac{A^n}{n!} + \ldots.. \tag{7.52}$$

We see that the one-parameter family of matrices $\phi(t) = e^{At}$ is the integral curve of the vector A. This leads to the following definition,

7.2.10 Definition For any Lie group G, we define the *exponential map*

$$\exp : \mathfrak{g} \to G \tag{7.53}$$

as follows. Let $A \in T_eG$, and let $\phi : \mathbf{R} \to G$ be the unique homomorphism such that $\phi'(0) = A$. Then,

$$\exp(A) \equiv e^A = \phi(1)$$

Clearly,

$$\begin{aligned} e^{(s+t)A} &= e^{sA}e^{tA}, \\ e^{-tA} &= (e^{tA})^{-1} \end{aligned}$$

The map $t \mapsto e^{tA}$ is a local diffeomorphism from T_eG to G. The maximal extension of the integral curve is the one-parameter subgroup of G indicated at the beginning of this subsection. The converse is also true. Any one-parameter subgroup of G is generated by a map $t \mapsto e^{tA}$, for some $A \in T_eG$. Since $\mathfrak{g} \cong L(G)$, there is a one-to-one correspondence between the Lie algebra of a Lie group and the one-parameter subgroups. Roughly speaking, the exponential map yields a neighborhood of $e \in G$, which is filled by one-parameter subgroups emanating from e by the integral curves of tangent vectors $A \in T_eG$. In fact if the diagram in figure 7.6 and the result in equation 7.49 are applied to the homomorphism $\phi : \mathbf{R} \to T$ with the condition that,

$$\phi_*(\tfrac{d}{dt}|_0) = A,$$

then ϕ is a Lie algebra homomorphism. The left-invariant vector field generated by A is the vector field tangent to the unique integral curve given by the map $t \mapsto e^{tA}$. We define,

$$\ln(e^A) = A, \quad e^{\ln(1+A)} = 1 + A,$$

wherever the formal power series converges. If A and B are two matrices near 0, we can study the behavior of

$$\ln(e^A e^B) = \ln[(1 + A + \frac{A^2}{2!} + \dots)(1 + B + \frac{B^2}{2!} + \dots)].$$

If we only retain the first order terms, we get,

$$\ln(e^A e^B) \doteq \ln(1 + A + B) = A + B.$$

If we wish to compute the quadratic terms, we formally multiply the power series for the exponentials, and then use the formal series expansion for $\ln(1 + X) = X - \frac{1}{2}X^2 + \dots$. However we need to be careful with the fact that matrix multiplication does not commute. The result is,

$$\begin{aligned}
\ln(e^A e^B) &= \ln\left(1 + A + B + \frac{A^2}{2} + AB + \frac{B^2}{2} + ...\right), \\
&= \left(A + B + \frac{A^2}{2} + AB + \frac{B^2}{2}\right) - \frac{1}{2}\left(A + B + \frac{A^2}{2} + AB + \frac{B^2}{2}\right)^2 + \dots \\
&= A + B + \frac{A^2}{2} + AB + \frac{B^2}{2} - \frac{1}{2}(A + B + \dots)^2 + \dots, \\
&= A + B + \frac{1}{2}[A, B] + \dots.
\end{aligned}$$

The full expansion is called the Campbell-Baker-Hausdorff (CBH) formula. The terms up to third order are,

$$\ln(e^A e^B) = A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B] - \frac{1}{12}[B, [A, B]] + \dots \tag{7.54}$$

All the terms of order two or higher are expressible in terms of brackets, so

$$[A, B] = 0, \Rightarrow e^A e^B = e^{A+B}.$$

Exponentiating the CBH formula 7.54 and keeping only the terms up to order 2, we can establish the following formulas,

7.2.11 Theorem

$$
\begin{aligned}
e^{tA}e^{tB} &= \exp\left\{t(A+B) + \tfrac{1}{2}t^2[A,B] + \mathscr{O}(t^3)\right\}, \\
e^{-tA}e^{-tB}e^{tA}e^{tB} &= \exp\left\{t^2[A,B] + \mathscr{O}(t^3)\right\}, \\
e^{tA}e^{tB}e^{-tA} &= \exp\left\{tB + t^2[A,B] + \mathscr{O}(t^3)\right\}. \qquad (7.55)
\end{aligned}
$$

The first of these formulas follows immediately from the CBH formula. The other two require more manipulation of exponential and logarithmic expansions along the same lines as in the computations leading to 7.54. A complete proof of the this theorem can be found in [34]. Next, we prove that the exponential map is natural with respect to the push-forward.

7.2.12 Theorem Let $\phi : G \to H$ be a smooth homomorphism between two Lie groups G and H. Then the exponential loop in following diagram commutes,

$$
\begin{tikzcd}
T\mathbf{R} \arrow[r, "\alpha_*"] \arrow[d] & T_eG \arrow[r, "\phi_*"] \arrow[d, "\exp"'] & T_{e'}H \arrow[d, "\exp"] \\
\mathbf{R} \arrow[r, "\alpha"] \arrow[rr, bend right, "\psi"'] & G \arrow[r, "\phi"] & H
\end{tikzcd}
$$

That is,

$$\phi \circ \exp = \exp \circ \phi_*. \qquad (7.56)$$

Proof Let $A \in T_eG$ and let $\alpha : \mathbf{R} \to G$ be a one parameter subgroup of G given by $\alpha(t) = e^{tA}$. Then, $A = \alpha'(0) = \alpha_*(\frac{d}{dt}|_0)$. Define $\psi : \mathbf{R} \to H$ by the composition $\psi = \phi \circ \alpha$. That is,

$$\psi(t) = \phi(e^{tA}).$$

We have,

$$
\begin{aligned}
\phi(s+t) &= \phi(e^{(s+t)A}), \\
&= \phi(e^{sA}e^{tA}), \\
&= \phi(e^{sA})\phi(e^{tA}), \\
&= \phi(s)\phi(t),
\end{aligned}
$$

so ϕ is a one parameter subgroup of H. The rest of the proof amounts to untangling the definition of the push-forward as shown in equation 1.25. Suppose that $B \in T_{e'}H$ such that $\psi(t) = e^{tB}$. That means that $B = \psi'(0)$. We consider the action of this tangent vector on an arbitrary smooth function $f : H \to \mathbf{R}$. We have,

$$
\begin{aligned}
B(f) &= \psi'(0)(f) = \psi_* \tfrac{d}{dt}(f)|_0, \\
&= \tfrac{d}{dt}(f \circ \psi)|_0, \\
&= \tfrac{d}{dt}(f \circ \phi \circ \alpha)|_0, \\
&= \alpha'(0)(f \circ \phi) = A(f \circ \phi), \\
&= \phi_* A(f).
\end{aligned}
$$

We conclude that $B = \phi_* A$. Thus, setting $t = 1$, $\psi(t) = e^{tB}$ can be rewritten

$$\phi(e^A) = e^{\phi_* A}, \quad \text{or} \quad \phi(\exp(A)) = \exp(\phi_* A)$$

which is what we wanted to prove.

7.2.2 The Adjoint Map

7.2.13 Definition Let G be a Lie group and let $g_0 \in G$. For each $g \in G$, consider the conjugation automorphism $C_g = L_g R_g^{-1} : G \to G$ given by $g_0 \mapsto g g_0 g^{-1} = L_g R_g^{-1}(g_0)$. We define the *adjoint map* $Ad_g : \mathfrak{g} \to \mathfrak{g}$ by the linear transformation,

$$Ad_g = C_{g*} = (L_g R_g^{-1})_*. \tag{7.57}$$

7.2.14 Definition Denote by $Aut(V)$ automorphism group of all invertible linear transformations of some vector space V over $\mathbf{R}$ (or $\mathbf{C}$). If V has dimension n, then $Aut(V)$ is isomorphic to the matrix group $GL(n, \mathbf{R})$. A homomorphism,

$$\phi : G \to GL(n, \mathbf{R})$$

from a Lie group to $GL(n, \mathbf{R})$ is called a (real) *representation* of order n. If the homomorphism is 1-1, the representation is called *faithful.* If W is subspace of V, and $\phi(g)v \in W$ for all $v \in W$ and $g \in G$, we say that W is an *invariant* subspace. A representation with no non-trivial invariant subspace is called *irreducible.* Same idea applies to Lie algebras. Since $Ad_{gg'} = Ad_g \circ Ad_{g'}$, the adjoint is a homomorphism, $Ad : G \to Aut(\mathfrak{g})$ is called the *adjoint representation.* The kernel is the center of the group G.

By equation 7.56, if $X \in \mathfrak{g}$ we have $\exp(C_* X) = C(\exp(X))$, that is

$$e^{Ad_g(X)} = g(e^X)g^{-1}. \tag{7.58}$$

Consider the case $G = GL(n, \mathbf{R})$. We evaluate the adjoint on a one-parameter subgroup of G. Let $Y \in \mathfrak{g} = \mathfrak{gl}(n, \mathbf{R})$ and $t \mapsto e^{tY}$ be one such one parameter subgroup. Then for a matrix $g \in GL(n, \mathbf{R})$, the conjugation map gives

$$C_g(e^{tY}) = g e^{tY} g^{-1}.$$

Taking the derivative with respect to t and evaluating at $t = 0$, we get

$$Ad_g(Y) = gYg^{-1}. \tag{7.59}$$

Now, we evaluate along the derivative of the adjoint map at $t = 0$, along another one-parameter subgroup $t \mapsto e^{tX}$,

$$\begin{aligned}
Ad_{e^{tX}}Y &= e^{tX}Ye^{-tX},\\
&= (1 + tX + \tfrac{1}{2!}t^2X^2 + \dots)Y(1 - tX + \tfrac{1}{2!}t^2X^2 + \dots),\\
&= 1 + t[X,Y] + \mathscr{O}(t^2),\\
\frac{d}{dt}Ad_{e^{tx}}Y|_{t=0} &= [X,Y]
\end{aligned}$$

We denote the quantity on the left hand side above by the notation ad_XY

$$ad_XY = [X,Y]. \tag{7.60}$$

Equivalently, $ad_X \in End(\mathfrak{g})$ is an endomorphism given by the map $Y \mapsto [X,Y]$. The reemergence of the of an operator yielding the Lie bracket is indicative that there is a Lie derivative floating around. The details are easy to clarify. The map $\varphi(t) = g(t) = e^{tX}$ is a local diffeomorphism generated by the flow of $X \in T_eG$. The Lie derivative of a vector field Y is given by

$$\pounds_{X_p}T = \frac{d}{dt}(\varphi_t^{-1})_*Y\bigg|_{t=0}.$$

The push-forward of the conjugation map $C_g = R_g^{-1}L_g$ action on Y gives,

$$\begin{aligned}
Ad_{g(t)}(Y) &= (R_g^{-1})_*L_{g*}Y\\
&= (R_g^{-1})_*Y, \quad \text{since} Y \text{is left invariant},\\
Ad_{e^{tX}}Y &= (R_g^{-1})_*Y,\\
\frac{d}{dt}Ad_{e^{tX}}Y|_{t=0} &= \frac{d}{dt}(R_g^{-1})_*Y|_{t=0},
\end{aligned}$$

which gives,

$$ad_XY = \pounds_XY.$$

We conclude that,

$$\begin{aligned}
Ad_{\exp X} &= \exp(ad_X),\\
Ad_{e^X} &= e^{ad_X} = 1 + ad_X + \tfrac{1}{2!}(ad_X)^2 + \dots,
\end{aligned} \tag{7.61}$$

where the first term represents of course the identity element in the algebra. If the Lie algebra is finite dimensional, we can define the *Killing form* as the form B given by,

$$B(X,Y) = \mathrm{Tr}(ad_X \circ ad_Y) = \mathrm{Tr}(ad_Xad_Y). \tag{7.62}$$

The Killing form is not a differential form, but rather, a symmetric bilinear entity that plays the role of a metric in the Lie algebra. In the adjoint representation, the Killing form is adjoint invariant, meaning,

$$B(ad_X Y, Z) + B(Y, ad_X Z) = 0$$

In terms of ad_X, the Jacobi identity in definition 7.2 becomes,

$$[ad_X, ad_Y]Z = ad_{[X,Y]}Z, \tag{7.63}$$

which shows explicitly that ad is a Lie algebra homomorphism.

The adjoint map can be used to prove an interesting formulation of the CBH formula first proved in 1899 by Poincaré [13]. Let,

$$\beta(w) = \frac{w}{1-e^{-w}} = \sum_{n=0}^{\infty} \frac{B_n^+ w^n}{n!}$$

be the generating function for the Bernoulli numbers. Define,

$$g(z) = \beta(\ln z) = \frac{z \ln z}{z-1}.$$

Then, the Campbell-Baker-Hausdorff formula can be written in the form,

$$\ln(e^A e^B) = A + \int_0^1 g(e^{ad_A} e^{t\, ad_B}) B\, dt. \tag{7.64}$$

The formula is complicated to use for explicit evaluation of the terms in the expansion, but nonetheless is a neat result because it makes it manifestly clear that the expansion depends only on the brackets. Following Hall [13], we illustrate how to get the first three terms. Set $z = v + 1$ and expand $g(v+1)$ in a Maclaurin series,

$$\begin{aligned} g(v+1) &= \frac{v+1}{v} \ln(v+1), \\ &= \frac{v+1}{v}(v - \tfrac{1}{2}v^2 + \tfrac{1}{3}v^3 + \ldots), \\ &= (v+1)(1 - \tfrac{1}{2}v + \tfrac{1}{3}v^2 + \ldots), \\ &= 1 + \tfrac{1}{2}v - \tfrac{1}{6}v^2 + \ldots. \end{aligned}$$

Next, we set $e^{ad_A} e^{t\, ad_B} - 1 = v$ and evaluate g up to second order in ad_A, ad_B. We ignore terms that contain B on the right of ad_B, since $ad_B B = 0$.

$$\begin{aligned} v &= [I + ad_A + \tfrac{1}{2}(ad_A)^2 + \ldots][I + t\, ad_B + \tfrac{1}{2}t^2(ad_B)^2 + \ldots] - I, \\ &= ad_A + \tfrac{1}{2}(ad_A)^2 + t\, ad_B + \tfrac{1}{2}t^2(ad_B)^2 + \ldots, \\ v^2 &= (ad_A)^2 + t\, ad_A\, ad_B + \ldots, \\ g(v+1) &= I + \tfrac{1}{2}\left(ad_A + \tfrac{1}{2}(ad_A)^2\right) - \tfrac{1}{6}\left((ad_A)^2 + t\, ad_B\, ad_A\right) + \ldots. \\ \int_0^1 g(v+1)\, dt &= I + \tfrac{1}{2}ad_A + \tfrac{1}{12}(ad_A)^2 - \tfrac{1}{12}ad_B\, ad_A + \ldots \end{aligned}$$

Hence,

$$\ln(e^A e^B) = A + B + \tfrac{1}{2}[A, B] + \tfrac{1}{12}[A, [A, B]] - \tfrac{1}{12}[B, [A, B]] + \dots$$

7.2.3 The Maurer-Cartan Form

Let G be a Lie group. A differential form θ in G is called left-invariant if

$$L_g^* \theta = \theta. \tag{7.65}$$

That is, if the form at point g_0 is given by θ_{g_0}, then,

$$L_g^* \theta_{g_0} = \theta_{g^{-1} g_0}.$$

The vector space $\mathfrak{g}^*$ of left-invariant one forms is the dual space of the Lie algebra of $\mathfrak{g}$ left-invariant vector fields. Recall that the Lie algebra of left-invariant vector fields is isomorphic to the tangent space at the identity $T_e G$. If X is a left-invariant vector field and θ is a left invariant one-form the $\theta(X)$ is constant. If θ is left invariant, then for each $g \in G$

$$L_g^*(d\theta) = d(L_g^* \theta) = d\theta,$$

so $d\theta$ is also left-invariant. The canonical form or *Maurer-Cartan* form of G is the form ω is the form that assigns to a left-invariant vector field, its value at the identity, that is,

$$\begin{aligned} \omega(X_g) &= L_{g^{-1}*} X_g, && \text{or equivalently,} \\ \omega(X) &= X, && \text{for all } \; X \in T_e G. \end{aligned}$$

The Maurer-Cartan form ω is left-invariant. On the other hand (the right), we have,

7.2.15 Theorem

$$R_g^* \omega = Ad_{g^{-1}} \omega. \tag{7.66}$$

Proof Let $X \in \mathfrak{g}$ generate a left invariant vector field in G via the flow of the exponential map $X \mapsto e^{tX}$. Let $g_0 \in G$. By definition, $\omega_{g_0}(X_{g_0}) = X$. Then

$$\begin{aligned} (R_g^* \omega_{g_0})(X_{g_0}) &= \omega_{g_0 g}(R_{g*} X_{g_0}), \\ &= \frac{d}{dt}[\omega_{g_0 g}(p e^{tX} g)]_{t=0}, \\ &= \frac{d}{dt}[(g_0 g)^{-1}(g_0 e^{tX} g)]_{t=0}, \\ &= \frac{d}{dt}[g^{-1} e^{tX} g]_{t=0}, \\ &= Ad_{g^{-1}} X, \\ &= Ad_{g^{-1}} \omega_{g_0}(X_{g_0}) \end{aligned}$$

Suppose $\{e_\alpha\}$ is a basis for the Lie algebra of left-invariant vector fields $\mathfrak{g}$. The bracket of two vectors in the Lie algebra must be expressible as a linear combination of the basis vectors, so there exist constants $C^\gamma{}_{\alpha\beta}$ such that

$$[e_\alpha, e_\beta] = C^\gamma{}_{\alpha\beta} e_\gamma. \tag{7.67}$$

The quantities $C^\gamma{}_{\alpha\beta}$ are called the *structure constants*. Since $[e_\alpha, e_\beta] = -[e_\beta, e_\alpha]$, the structure constants are antisymmetric on the lower indices. The frame $\{e_\alpha\}$ is called a Maurer-Cartan frame. Let $\{\omega^\alpha\}$ be the dual basis, so that $\omega^\alpha(e_\beta) = \delta^\alpha_\beta$. By definition, applying the Maurer-Cartan form to e_α returns the value of e_α at the identity. This gives an almost tautological expression for the components of the Maurer-Cartan form in terms of the Maurer-Cartan coframe,

$$\omega = e_\alpha(e) \otimes \omega^\alpha.$$

Applying the definition for the differential of a one-form 6.28,

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

we get,

$$\begin{aligned} d\omega^\alpha(e_\beta, e_\gamma) &= e_\beta(\omega^\alpha(e_\gamma)) - e_\gamma(\omega^\alpha(e_\beta)) - \omega^\alpha([e_\beta, e_\gamma]), \\ &= e_\beta(\delta^\alpha_\gamma) - e_\gamma(\delta^\alpha_\beta) - C^\gamma{}_{\alpha\beta}\omega^\alpha(e_\gamma), \\ &= -C^\alpha{}_{\beta\gamma}. \end{aligned}$$

Using the antisymmetry of the wedge product and the antisymmetry of the structure constants in the lower indices, we can rewrite the last equation as,

$$d\omega^\alpha = -\frac{1}{2} C^\alpha{}_{\beta\gamma}\omega^\beta \wedge \omega^\gamma. \tag{7.68}$$

This equation of structure is called the *Maurer-Cartan equation*. Let X, Y be left-invariant. Since $\omega(X)$ and $\omega(Y)$ are constant, we have $X(\omega(Y)) = Y(\omega(X)) = 0$, so using the definition 6.28 for the differential of a one form, we can also write the Maurer-Cartan equation as

$$d\omega(X, Y) = -\omega([X, Y]). \tag{7.69}$$

There is an annoying factor of 1/2 which makes the notation inconsistent in the literature. Some authors include such a factor in the equation of structure 7.69, but typically those authors also include a 1/2 their definition of the differential of a one form $d\omega(X, Y)$. Other authors restrict the sum in equation 7.68 to values $i < j$, so the factor 1/2 does not appear there. Yet some others avoid the bracket notation altogether, or they invent a new hybrid wedge/bracket $[\omega, \omega]$ of forms that may account for the 1/2. In this latter case, the 2-form represented by the wedge/bracket[2] is usually interpreted as a section of $\Lambda^2 \otimes \mathfrak{g} \otimes \mathfrak{g}$.

[2] If $\alpha, \beta \in \Omega^1 \otimes \mathfrak{g}$ are Lie algebra valued one-forms, the usual definition of the bracket is $[\alpha, \beta](X, Y) = [\alpha(X), \beta(Y)] - [\alpha(Y), \beta(X)]$. Thus $[\alpha, \alpha](X, Y) = 2[\alpha(X), \alpha(Y)]$.

If G is a matrix group, real or complex, the Maurer-Cartan form 3.41 can be written as [3],

$$\omega = A^{-1}dA, \tag{7.70}$$

where, $g = A \in G$ plays the role of the attitude matrix introduced in section 3.4. The form $\omega = g^{-1}dg$ is clearly left-invariant, because if g_0 is another constant matrix, then

$$(g_0 g)d(g_0 g) = g^{-1}dg.$$

We can express the components of the Killing form in terms of the structure constants. First, we compute

$$\begin{aligned}(ad_{e_\alpha} ad_{e_\beta})e_\gamma &= ad_{e_\alpha}([e_\beta, e_\gamma]),\\ &= [e_\alpha, [e_\beta, e_\gamma]],\\ &= [e_\alpha, C^\sigma{}_{\beta\gamma} e_\sigma],\\ &= C^\rho{}_{\alpha\sigma} C^\sigma{}_{\beta\gamma} e_\rho.\end{aligned}$$

Taking the trace means we perform a contraction with the dual form ω^γ which is the same as setting $\gamma = \rho$ on the coefficients on the right. We get

$$B_{\alpha\beta} = C^\rho{}_{\alpha\sigma} C^\sigma{}_{\beta\rho}. \tag{7.71}$$

Notice that the components of the Killing form result from the contraction of the antisymmetric indices of the structure constants; this yields the simplest symmetric tensor that can be constructed from the structure constants. Cartan used the Killing form to characterize an important attribute of Lie algebras called semisimple. A non-Abelian Lie algebra is *simple* if it has no, non-trivial proper ideals. A *semisimple* Lie algebra can be decomposed as the direct sum of simple Lie algebras. The Cartan criterion states states that the Lie algebra is semisimple if and only if the Killing form is non-degenerate. Of course, the structure constants depend on the choice of the basis. But, since the form is symmetric, it can be diagonalized by an orthogonal matrix, so it can be classified by the eigenvalues. If the Lie group is compact and semisimple, the Killing form is positive definite and in diagonal form, all the entries are positive. The most salient achievement of E. Cartan was to provide a complete classification of semisimple Lie algebras. Included in this classification are all the Lie algebras associated with the special Lie groups mentioned above in this chapter.

It is also easy to express the adjoint representation in terms of the structure constants. All is really needed is to define matrices T_α whose components are,

$$[T_\alpha]^\gamma{}_\sigma = C^\gamma{}_{\alpha\sigma}.$$

We then get yet another manifestation of the Jacobi identity 7.2 in the form,

$$[T_\alpha, T_\beta] = C^\gamma{}_{\alpha\beta} T_\gamma. \tag{7.72}$$

This is the component version of equation 7.63, and it shows that the structure constants themselves, generate the adjoint representation.

[3]For a matrix group, $\omega(X_g) = L_{g^{-1}*}(X_g) = dL_g^{-1}(X_g)$. So, the logarithmic differential maps the vector field X_g to its value at the identity e.

7.2.4 Cartan Subalgebra

To get at the physics applications of Lie algebras we need to dip our toes into representation theory. Actually, what we should say is that it is the physics that lead to the development of representation theory by giants like Cartan and Weyl.

7.2.16 Definition A *Cartan subalgebra* $\mathfrak{h} \in \mathfrak{g}$ is a nilpotent subalgebra that is its own normalizer. The dimension of the Cartan subalgebra is called the rank of $\mathfrak{g}$.

Every pair of elements of $\mathfrak{h}$ commute, and every element of $\mathfrak{g}$ that commutes with all elements of $\mathfrak{h}$ is in $\mathfrak{h}$. In this sense, a Cartan subalgebra is the maximum number of commuting generators of the Lie algebra. The idea is to simultaneously diagonalize the basis of $\mathfrak{h}$. The eigenvalues are used to label the states of the system.

Using the Killing form, one finds an orthonormal basis $\{h_i\}$ for $\mathfrak{h}$ and extends to a basis of $\mathfrak{g}$

$$\{h_1, h_2 \ldots, h_k, , g_1, g_{-1}, g_2, g_{-2}, \ldots, g_{\frac{n-k}{2}}, g_{-\frac{n-k}{2}}\}$$

with the following properties:

1. $[h_1, h_j] = 0$.
2. $[h, g] = \lambda(h)g$ for all $h \in \mathfrak{h}$ and $0 \neq g \in \mathfrak{g}$. Or, in terms of the basis, $[h_i, g_j] = \lambda_i^{(j)} g_j$,
3. $[g_j, g_{-j}] \in \mathfrak{h}$.

The first property is a re-statement that each pair basis vectors in $\mathfrak{h}$ commute. The second property is a kind of generalized eigenvector equation for ad_h. For each g_j we associate a position vector $r^{(j)} = (\lambda_1^{(j)}, \lambda_2^{(j)}, \ldots \lambda_k^{(j)})$. These are called the *roots* and the set of all roots is called the root space. The plot of the root vectors in $\mathbf{R}^k$ is a set of arrows that exhibits certain reflection symmetries; the set is called the *root diagram.* Root spaces lead to Cartan's classification of semisimple Lie algebras. The classification can also be visualized by a scheme called Dynkin diagrams. The basis elements g_j and g_{-j} are called the *raising* and *lowering* operators. We will show in the next chapter how this abstract machinery, leads to real concrete results in physics. Lie symmetries of physical systems are interconnected with the deep subject of representation theory.

7.3 Transformation Groups

The importance of this chapter for the purpose of applications to physics is that in many physical models, Lie groups are manifested as transformation groups that act on the system. In the simplest case we have linear transformations in $\mathbf{R}^n$ which in a particular basis, can be represented by matrix multiplication of an element of the general linear group $GL(n, R)$ with a vector.

The group of rotations in $\mathbf{R}^3$ constitutes a symmetry group in the dynamical system of rigid body motion, and the Lorentz group is the essential symmetry group of space-time. The Lie algebra of a Lie group is basically a first order symmetry approximation. If one thinks of Lie group such as the rotation group as a symmetry group acting on a manifold, then an element of the Lie algebra represents an infinitesimal transformation near the identity of the group. Bases vectors of the Lie algebras of the corresponding Lie group transformation are then interpreted as generators of infinitesimal transformations. The exponential map provides a bridge between full elements of the group and infinitesimal transformations represented by elements of the Lie algebra and elements of the Lie group. The Lie algebra is determined by the structure constants and the Maurer-Cartan equations.

7.3.1 Definition Let M be an n-dimensional manifold and G a Lie group. G is called a *Lie transformation group* on (the right of) M, if there exists a smooth map $\mu : M \times G \to M$

$$(p, g) \xrightarrow{\mu} p \cdot g = R_g(p), \quad \text{for } \ (p, g) \in M \times G,$$

such that for each $p \in M$,
1) $p \cdot e = p$
2) $(p \cdot g_1) \cdot (g_2) = p \cdot (g_1 \cdot g_2), \quad$ for every $g_1, g_2 \in G$.

Of course, a one-parameter group of diffeomorphisms in the sense of definition 7.1.1, is a special case of a Lie transformation group, with $G = \mathbf{R}$. Another example would be the action of the rotation group $SO(3, \mathbf{R})$ on a sphere S^2. We are often interested in linear representations of the group acting on some vector space, in which the action respects the linear structure. An example of this, would be adjoint representation of the action of a Lie group into itself.

7.3.2 Definition Let G be a transformation group on M with identity e, and let $p \in M$ be any point in the manifold. We say the transformation is,

1. *Effective*, if the Kernel $K = \{g \in G : p \cdot g = p\} = \{e\}$. In other words, if $g \neq e$, there exists a point p, such that $p \cdot g \neq p$. The Kernel of a group is a normal subgroup. If the Kernel is not trivial, the action of G on M is not effective, but the action of G/K is.

2. *Free*, if $g \neq e$, then $p \cdot g \neq p$. In other words, if $g \neq e$, then there is no fixed point for g. If there is a fixed point p, we define the *isotropy subgroup* of p as, $Iso(p) = \{g \in G : p \cdot g = p\}$.

3. *Transitive*, if $p \neq q$, then there exists a g such that $p \cdot g = q$. The set of all q such that $p \cdot g = p$ for some g, is called the *orbit* of the point p, and it is denoted by Gp. The map $p \mapsto Gp$ which sends p to $p \cdot g$ defines a diffeomorphism $G/Iso(p) \xrightarrow{\cong} Gp$.

Unless otherwise stated, we assume that when we say that a Lie group acts on a manifold, the action is on the right, and the action is transitive and effective. Let X be an element of the Lie algebra $\mathfrak{g}$ of a Lie group G acting on M, and let $p \in M$. The one-parameter subgroup given by the exponential map

$$X \mapsto e^{tX}$$

generates a curve on M given by

$$\varphi_t(p) = p\, e^{tX} = R_{e^{tX}}(p), \tag{7.73}$$

with $\varphi_0(p) = p$ with tangent vector $X^* = \sigma(X)$ at p given by,

$$\sigma(X)|_p = \frac{d}{dt}(p\, e^{tX})|_{t=0} \tag{7.74}$$

If U is a neighborhood of p, the map $\varphi_t(p)$ above constitutes a local one-

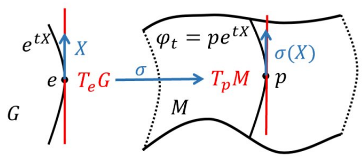


Fig. 7.7: Fundamental Vector

parameter group of transformations of M associated with the vector field $X^* = \sigma(X)$, called the *fundamental vector* field. It is not really possible to draw a good picture of a fundamental vector field, since for starters, all but the most trivial principal fiber bundles, either live in higher dimensions, or have complicated topologies. Nevertheless, figure 7.7 may be of some help in visualizing these vector fields.

7.3.3 Theorem

$$(R_g)_*(\sigma(X)) = \sigma(Ad_{g^{-1}}X). \tag{7.75}$$

Proof Let φ_t but the one-parameter group of diffeomorphisms associated with $\sigma(X)$ at p and ψ_t be the one-parameter group of diffeomorphisms associated with $(R_g)_*(\sigma(X))$ at pg. The map $R_g : M \to M$ is a local diffeomorphism, so as in equation 7.9 we have the commuting diagram,

$$\begin{array}{ccc} M & \xrightarrow{R_g} & M \\ \varphi_t \downarrow & & \psi_t \downarrow \\ M & \xrightarrow{R_g} & M. \end{array}$$

Thus, we get,

$$\begin{aligned} \psi_t &= R_g \circ \varphi_t \circ R_{g^{-1}}, \\ &= R_g \circ R_{e^{tX}} \circ R_{g^{-1}}, \\ &= R_{g^{-1}e^{tX}g}, \end{aligned}$$

The one-parameter group of diffeomorphisms $\{g^{-1}e^{tX}g\}$ is generated by $Ad_{g^{-1}}X$, so the vector field associated with ψ_t is $\sigma(Ad_{g^-1}X)$. We summarize this in the following diagram,

$$\begin{array}{ccc} \mathscr{X}(M) & \xrightarrow{(R_g)_*} & \mathscr{X}(M) \\ \sigma\uparrow & & \sigma\uparrow \\ \mathfrak{g} & \xrightarrow{Ad_{g^-1}} & \mathfrak{g}. \end{array}$$

It might be instructive to present a second proof in the style of theorem 7.66

$$\begin{aligned} (R_{g*})\sigma(X) &= \frac{d}{dt}[R_g(pe^{tX})]_{t=0}, \\ &= \frac{d}{dt}[pe^{tX}g]_{t=0}, \\ &= \frac{d}{dt}[p(gg^{-1})e^{tX}g]_{t=0}, \\ &= \frac{d}{dt}[(pg)g^{-1}e^{tX}g]_{t=0}, \\ &= \frac{d}{dt}[(pg)e^{tAd_{g^-1}X}]_{t=0}, \\ &= \sigma_{pg}Ad_{g^{-1}}X. \end{aligned}$$

Here, we have used equation 7.58 in the next to last step. This formula is consistent with the formula for the pull-back of the Maurer-Cartan form 7.66 by the following computation,

$$\begin{aligned} R_g^*\omega(\sigma(X)) &= \omega(R_{g*}\sigma(X)), \\ &= \omega(\sigma(Ad_{g^-1})X), \\ &= Ad_{g^{-1}}\omega(\sigma(X)). \end{aligned}$$

The map

$$\sigma : \mathfrak{g} \to \mathscr{X}(M)$$

given by

$$X \mapsto X^* = \sigma(X)$$

can also be viewed in alternative way by considering the map

$$\begin{aligned} \sigma_p : G &\to M, \\ g &\mapsto \sigma_p(g) = pg. \end{aligned}$$

Then

$$\sigma(X)(p) = \sigma_{p*}(X)_e$$

This small variation of the definition of a fundamental vector is helpful in establishing the following,

7.3.4 Theorem The map σ is a Lie algebra homomorphism, that is,

$$\sigma([X,Y]) = [\sigma(X), \sigma(Y)]. \tag{7.76}$$

Proof Aside from a small change in notation, the proof here is the same as in Spivak [34], and in Kobayashi-Nomizu [18]. Let $\xi_t(p) = e^{tX} = R_{e^{tX}}p$ be the one-parameter group of diffeomorphisms associated with $X \in \mathfrak{g}$, and let $Y \in \mathfrak{g}$. We extend X and Y to left-invariant vector fields in G. By theorem 7.1.8, we have

$$\begin{aligned}
[X, Y] &= \pounds_X Y, \\
&= \lim_{t \to 0} \frac{Y_e - (\xi_{t*} Y)_e}{t}, \\
&= \lim_{t \to 0} \frac{Y_e - (R_{e^{tX}*} Y)_e}{t}, \\
&= \lim_{t \to 0} \frac{Y_e - (Ad_{e^{-tX}} Y)_e}{t}, \quad \text{as in the proof of theorem 7.1.8.}
\end{aligned}$$

Denote by $R_g : M \to M$, the map $R_g(p) = pg$, then

$$(R_{e^{tX}} \circ \sigma_{pe^{-tX}})(g) = p(e^{-tX})ge^{tX}.$$

Thus, once again by theorem 7.1.8, the Lie bracket of the fundamental vectors gives

$$\begin{aligned}
[\sigma(X), \sigma(Y)] &= \lim_{t \to 0} \frac{\sigma(Y)_p - [R_{e^{tX}*}\sigma(Y)]_p}{t}, \\
&= \lim_{t \to 0} \frac{\sigma_{p*} Y_e - \sigma(Ad_{e^{-tX}}(Y))_p}{t}, \quad \text{as in theorem 7.3.3} \\
&= \lim_{t \to 0} \frac{\sigma_{p*} Y_e - \sigma_{p*}(Ad_{e^{-tX}}(Y))_e}{t}, \\
&= \sigma_{p*} \lim_{t \to 0} \frac{Y_e - (Ad_{e^{-tX}} Y)_e}{t}, \\
&= \sigma([X, Y])
\end{aligned}$$

For the time being, the results in theorems 7.3.3 and 7.3.4 might appear as a pure formality, but as we will see later, they are instrumental in the treatment of connections on principal fiber bundles. The first of these two formulas tell us how to push forward fundamental vectors along the orbit of right-translations. The second theorem states that the fundamental vectors constitute a Lie algebra that is completely determined by the lie algebra of the group. The two results are used in interpreting the meaning of connections on principal fiber bundles, as later defined in 9.3.2.

Chapter 8

Classical Groups in Physics

In this section we present a pedestrian view of some of the common Lie algebras and classical Lie groups that appear in mathematical physics.

8.1 Orthogonal Groups

8.1.1 Rotations in $\mathbf{R}^2$

Let $z = x + iy$ be a complex number, and consider the map introduced in section 5.2.2

$$(x+iy) \xrightarrow{\alpha} \begin{bmatrix} x & y \\ -y & x \end{bmatrix}.$$

The map is clearly a vector space isomorphism between the complex numbers and a subset of the set 2×2 matrices. The map can be written as

$$(x+iy) \xrightarrow{\alpha} xI + yJ,$$

where I is the identity matrix and J is the symplectic matrix 5.50, with $J^2 = -I$. Define

$$U(1) = \{z \in \mathbf{C} : |z| = 1\}$$

to be the group of unimodular complex numbers. If $z \in U(1)$, then we can write z in the form $z = e^{i\theta}$. The map $\theta \to e^{i\theta}$ is not 1-1 because replacing θ by $\theta + 2\pi$ gives the same number. The Kernel of the map is the set $\{2\pi\mathbf{Z}\}$, that is, the integer multiples of 2π. $U(1)$ acts on $\mathbf{C}$ by multiplication, which results on a rotation by θ. The action passes to the circle $S^1 \cong \mathbf{R}/(2\pi\mathbf{Z})$. By Euler's formula, $e^{i\theta} = \cos\theta + i\sin\theta$, so α restricts to a map from $U(1)$ to the special orthogonal group $SO(2,\mathbf{R})$ consisting of 2×2 rotation matrices,

$$e^{i\theta} \xrightarrow{\alpha} R_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

It is an elementary exercise to verify that

$$e^{i\theta_1} \cdot e^{i\theta_2} \xrightarrow{\alpha} R_{\theta_1} \cdot R_{\theta_1}, \qquad \text{that is,}$$

$$e^{i(\theta_1+\theta_2)} \xrightarrow{\alpha} R_\theta = \begin{bmatrix} \cos(\theta_1+\theta_2) & \sin(\theta_1+\theta_2) \\ -\sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2). \end{bmatrix}.$$

The exercise amounts to multiplying the rotation matrices and recognizing the summation formulæ for sine and cosine. Thus, the map is a smooth Lie group homomorphism. Since the map is also a diffeomorphism, we have a Lie group isomorphism $U(1) \cong SO(2,\mathbf{R})$. For reasons that will become apparent in comparing later with the discussion of rotations in $\mathbf{R}^3$ by quaternions, we show the expression for the matrix rotation R_θ as the product of two consecutive rotations by $\theta/2$, by means of the double angle formulas

$$\begin{aligned} R_\theta &= \begin{bmatrix} \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} & 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ -2\sin\frac{\theta}{2}\cos\frac{\theta}{2} & \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} \end{bmatrix}, \\ &= \begin{bmatrix} q_0^2 - q_1^2 & 2q_0q_1 \\ -2q_0q_1 & q_0^2 - q_1^2 \end{bmatrix}, \end{aligned} \tag{8.1}$$

where $q_0 = \cos\frac{\theta}{2}$ and $q_1 = \sin\frac{\theta}{2}$.

Now, consider the exponential map $\phi : \mathbf{R} \to SO(2,\mathbf{R})$ given by,

$$t \xrightarrow{\phi} e^{tA} = \begin{bmatrix} \cos t\theta & \sin t\theta \\ -\sin t\theta & \cos t\theta \end{bmatrix}. \tag{8.2}$$

A matrix e^{tA} is orthogonal if $(e^{tA})^T = (e^{tA})^{-1} = e^{-tA}$, so this implies that $A^T = -A$. Per our previous discussion, the matrix A is a representative of the Lie algebra, so the Lie algebra of the orthogonal group consists of antisymmetric matrices. For the special orthogonal group with matrices with $\det e^A = 1$, the formula 5.45,

$$\det e^A = e^{\operatorname{Tr} A}$$

implies that elements A of the Lie algebra also have 0 trace. $A = \phi'(0)$, with $\phi(0) = I$. so,

$$A = \frac{d}{dt}\begin{bmatrix} \cos t\theta & \sin t\theta \\ -\sin t\theta & \cos t\theta \end{bmatrix}_{t=0}. \tag{8.3}$$

$$= \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix} = \theta \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{8.4}$$

This means that the matrix J is a basis for the Lie algebra $\mathfrak{so}(2,\mathbf{R})$. This example is as simple as it gets, but there are some good lessons to be learned. As it was discussed earlier, the significance of the Lie algebra element A and the basis J is that they constitute the generators of an infinitesimal rotation near the identity. This becomes clear if one looks at the rotation matrix with θ small. Then, $\cos\theta \simeq 1$ and $\sin\theta \simeq \theta$. To first order, the rotation matrix is given by $R_\theta \simeq I + \theta J = I + A$. We verify that the exponential map gives the group element from the infinitesimal generator. The computation is almost identical to the proof of Euler's formula by Maclaurin series. The key is that $J^2 = -I, J^3 = -J$, etc., so that

$$\begin{aligned} e^A &= I + A + \tfrac{1}{2!}A^2 + \tfrac{1}{3!}A^3 + \tfrac{1}{4!}A^4 + \tfrac{1}{5!}A^5 + \ldots \\ &= I + \theta J + \tfrac{1}{2!}(\theta J)^2 + \tfrac{1}{3!}(\theta J)^3 + \tfrac{1}{4!}(\theta J)^4 + \tfrac{1}{5!}(\theta J)^5 \ldots \\ &= (I - \tfrac{1}{2!}\theta^2 + \tfrac{1}{4!}\theta^4 + \cdots +)I + (\theta - \tfrac{1}{3!}\theta^3 + \tfrac{1}{5!}\theta^5 + \ldots)J. \end{aligned}$$

Hence

$$e^{\theta\, J} = (\cos\theta)I + (\sin\theta)J, \tag{8.5}$$

which is the original rotation matrix. The irreducible representations of $U(1)$ are the trivial representation and the 2×2 matrix representations given by the maps,

$$\phi_n(e^{i\theta}) = \begin{bmatrix} \cos 2n\theta & \sin 2n\theta \\ -\sin 2n\theta & \cos 2n\theta \end{bmatrix}, \qquad n \in \mathbf{Z}^+,$$

so basically, the irreducible representations are the homomorphisms of the group into itself.

8.1.2 Rotations in $\mathbf{R}^3$

The Lie group $SO(3, \mathbf{R})$ consists of 3×3 orthogonal matrices with determinant equal to 1. The Lie algebra $\mathfrak{so}(3, \mathbf{R})$ is the set of 3×3 antisymmetric matrices with zero trace. The zero trace condition is superfluous since the diagonal elements of an antisymmetric matrix are zero. In consideration of the case above for $\mathfrak{so}(2, \mathbf{R})$, we choose as basis for $\mathfrak{so}(3, \mathbf{R})$, the matrices.

$$\alpha_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \quad \alpha_y = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \alpha_x = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The exponentials $R_x = e^{\alpha_x \theta_1}, R_y = e^{\alpha_y \theta_2}, R_z = e^{\alpha_z \theta_3}$ represent rotations about the x, y and z axes respectively. In explicit form, these matrices are,

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & \sin\theta_1 \\ 0 & -\sin\theta_1 & \cos\theta_1 \end{bmatrix}, \quad R_y(\theta) = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix}, \quad R_z(\theta) = \begin{bmatrix} \cos\theta_3 & \sin\theta_3 & 0 \\ -\sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Any rotation in $\mathbf{R}^3$ can be obtained by a composition of rotations about the x, y and z axes. However, the standard in physics is to utilize the *Euler angles* $\{\phi, \theta, \psi\}$ introduced by Euler to study the motion of rigid bodies. The general Euler angle rotation is obtained by the composition of three rotations carried as follows (See figure 8.1).

1. Perform a rotation $R_z(\phi)$ by an angle ϕ around the z axis. We label the new axes as $\{\xi, \eta, z\}$.

2. Follow by a rotation $R_\xi(\theta)$ by an angle θ around the new x axis, which in step (1) we labelled ξ. We label the new axes as $\{\xi, \eta', z'\}$.

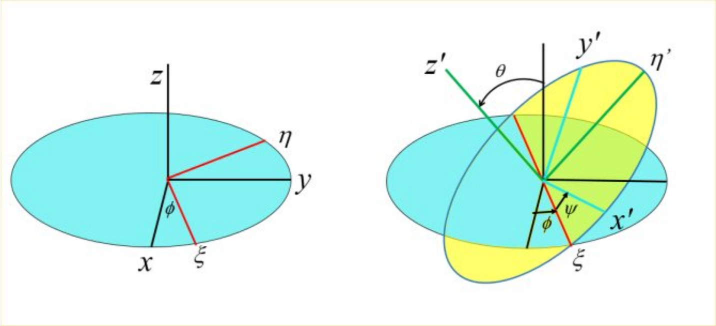


Fig. 8.1: Euler Angles

3. Finish with a rotation $R_{z'}(\psi)$ by an angle ψ around the new z-axis, which in step (2) we labelled z'. The final axes are labelled $\{x', y', z'\}$

The rotation matrices are

$$R_z(\phi) = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, R_\xi(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}, R_{z'}(\psi) = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8.6)$$

A straight-forward matrix multiplication yields the full rotation $R = R_z(\phi) \cdot R_\xi(\theta) \cdot R_{z'}(\psi)$,

$$R = \begin{bmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{bmatrix}. \quad (8.7)$$

Since R is the product of orthogonal matrices, the matrix is also orthogonal and $R^{-1} = R^T$. If we consider the unit 2-sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$, the rotation gives a map $R : S^2 \to S^2$. Any rotation of S^2 can be viewed as a composition R of the three Euler angle rotations, or as a single rotation around an axis pointing towards the image of the north pole along the axis z'. In principle, we should be able to prescribe a direction and an angle as the data to find a matrix representing a rotation by the given angle around the given direction. Finding this data in terms of the Euler angles requires a bit of work.

8.1.3 SU(2)

In this subsection, we develop a representation of rotations in terms of 2×2 complex matrices. As discussed in section 1.4, orthogonal transformations in $\mathbf{R}^n$ are isometries, so they can be described as the group of transformations that preserve length. In $\mathbf{R}^3$ the length is given by $g(X, X) = x^2 + y^2 + y^2$ under the standard metric. Getting a little ahead of ourselves, let's consider the metric $\eta = \text{diag}(+ - --)$ for Minkowski's space as in 2.35. Let $x^\mu = (t, x, y, z)$ be the components of a vector in $M_{1,3}$. Consider the map from $M_{1,3}$ to a 2×2 Hermitian matrix given by,

$$x^\mu = (t, x, y, z) \mapsto x^{A\dot{B}} = \begin{bmatrix} t + z & x - iy \\ x + iy & t - z \end{bmatrix}, \quad \mu = 0, 1, 2, 3 \quad (8.8)$$

The index notation for the matrix $X = (X^{A\dot{B}})$ is meant to elucidate the property of Hermitian matrices for which, the complex conjugate $\overline{X}^{A\dot{B}}$ equals the transpose $X^{B\dot{A}}$. The bar index notation was used in early work on spinors by Veblen and Taub. Bar indices were later changed to prime indices $X = (X^{AA'})$ in some seminal work by R. Penrose in the context of twistors. When convenient, we will invoke the Penrose notation. The map is chosen so that,

$$\|x^\mu\|^2 = \det X = \det(x^{A\dot{B}}).$$

We can write the matrix in terms of a basis,

$$\begin{aligned} x^{A\dot{B}} &= \begin{bmatrix} t+z & x-iy \\ x+iy & t-z \end{bmatrix}, \\ &= t\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + y\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} + z\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ &= x^\mu \sigma_\mu^{A\dot{B}}, \end{aligned} \tag{8.9}$$

where,

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{8.10}$$

Here, $\sigma_0 = I$ and $\{\sigma_i, i = 1,2,3\}$ are the *Pauli matrices.* For now, we constrain to the spatial part of the matrix by restricting to indices $i, j \ldots = 1,2,3$. Since $\det(X)$ is equal to the metric we wish to preserve, we seek unitary matrices

$$Q = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in SU(2),$$

such that $\det(X)$ is invariant under a similarity transformation. specializing to $\mathbf{R}^3$ by setting the coordinate $t = 0$, the similarity transformation reads

$$\begin{aligned} \tilde{X} &= QXQ^\dagger, \\ \begin{bmatrix} \tilde{z} & \tilde{x} - i\tilde{y} \\ \tilde{x} + i\tilde{y} & -\tilde{z} \end{bmatrix} &= \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} z & x - iy \\ x + iy & -z \end{bmatrix} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}. \end{aligned} \tag{8.11}$$

The quantities $\{\alpha, \beta, \gamma, \delta\}$ are called the *Cayley-Klein* parameters. The structure of the Lie algebra $\mathfrak{su}(2)$ can be obtained in a completely analogous manner as was done for the orthogonal groups. If $t \mapsto e^{tA}$ is a one-parameter subgroup of $SU(2)$, then the inverse of e^{tA} is equal to its Hermitian adjoint, that is,

$$\begin{aligned} (e^{tA})^{-1} &= (e^{tA})^\dagger, \\ e^{-tA} &= e^{tA^\dagger}. \end{aligned}$$

Taking the derivative at $t = 0$, we find that $A^\dagger = -A$. The formula $\det(e^A) = e^{\operatorname{Tr} A}$ shows that if $\det(e^A) = 1$ then $\operatorname{Tr} A = 0$. We conclude that the Lie algebra,

$$\mathfrak{su}(2) = \{A \in GL(2, \mathbf{C}) : A^\dagger = -A, \ \operatorname{Tr}(A) = 0\}, \tag{8.12}$$

consists of all traceless anti-Hermitian 2×2 matrices. This means that the Cayley-Klein parameters are not all independent, as they must satisfy the conditions

$$\gamma = -\overline{\beta}, \quad \delta = \overline{\alpha}.$$

The rotation matrix in $\mathbf{R}^3$ represented by the Cayley-Klein parameters can be obtained by direct computation of the matrix multiplication 8.11 and picking out the coefficients of the transformed vectors. One may use the trick of setting

$$x_+ = x + iy, \qquad x_- = x - iy,$$

as done in Goldstein [11], or one can apply the transformation to the basis vectors given by the Pauli matrices, or these days, one can simply insert into computer algebra system. The resulting matrix A is given by

$$A = \begin{bmatrix} \frac{1}{2}(\alpha^2 - \gamma^2 + \delta^2 - \beta^2) & \frac{i}{2}(\gamma^2 - \alpha^2 + \delta^2 - \beta^2) & \gamma\delta - \alpha\beta \\ \frac{i}{2}(\alpha^2 + \gamma^2 - \delta^2 - \beta^2) & \frac{1}{2}(\alpha^2 + \gamma^2 + \beta^2 + \delta^2) & -i(\alpha\beta + \gamma\delta) \\ \beta\delta - \alpha\gamma & i(\alpha\gamma + \beta\delta) & \alpha\delta + \beta\gamma \end{bmatrix}. \tag{8.13}$$

By inspection, the set of Pauli matrices $\sigma = \{\sigma_1, \sigma_2, \sigma_3\}$ is a basis for the Lie algebra. A quick computation gives,

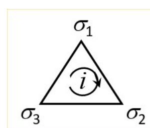
$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I, \;\; \det(\sigma_i) = -1.$$

and structure constants,

$$[\sigma_i, \sigma_j] = 2i\epsilon^k{}_{ij}\sigma_k, \tag{8.14}$$

where $\epsilon^k{}_{ij}$ is the Levi-Civita permutation symbol 2.41.

The factor $2i$ in the formula for the structure constants creates a minor conflict between physicists and mathematicians, but this is historically unavoidable. For example, from the second permutation symbol identity in 2.46, it follows immediately that, because of the i factor, the components of the Killing form are $-4\delta_{ij}$. Thus, for a physicist, a Lie algebra is compact if the Killing form is negative definite, which is the opposite of what was stated earlier. In quantum mechanics, it is customary to denote the set of Pauli matrices by a vector-like notation $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, in which case, the spin operator in the spin 1/2 representation, is written as $\mathbf{J} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\sigma}$. The multiplication table of Pauli matrices exhibits a cyclic permutation feature as shown in the adjacent figure. We can immediately verify that $\sigma_1\sigma_2 = i\sigma_3$, $\sigma_2\sigma_3 = i\sigma_1$ and $\sigma_3\sigma_1 = i\sigma_2$. Thus, as shown in the diagram, the product of two Pauli matrices gives i times the third matrix if the product is taken clockwise, and $-i$ times the third matrix if traversed counterclockwise. At the center if the triangular diagram there is an i as part of a reminder in the pneumonic, not to forget this factor. Since the squares of the Pauli matrices give the identity, to get an analog of Euler's formula in matrix form as in equation 8.5, we use the set $\{i\sigma_1, i\sigma_2, i\sigma_3\}$. Like J, these matrices all have squares equal to $-I$.



At this point, we inject the observation that the algebra of Pauli matrices is very closely related to the set $\mathbf{H}$ of quaternions. A *quaternion* is an entity of the form

$$q = q_0\mathbf{1} + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}, \tag{8.15}$$

where the basis elements satisfy,

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1.$$

As a vector space, the space H of quaternions is isomorphic to $\mathbf{R}^4$ The components (q_1, q_2, q_3) are in 1-1 correspondence with $\mathbf{R}^3$ vectors. It quickly follows that, $\mathbf{ij} = \mathbf{k}$, $\mathbf{jk} = \mathbf{i}$, and $\mathbf{ki} = \mathbf{j}$. A 2×2 matrix representation of the quaternion basis is obtained by the identity matrix, together with setting $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\} = \{-i\alpha_1, -i\alpha_2, -i\alpha_3\}$. Another way of saying this, is to set

$$\mathbf{i} = \sigma_3\sigma_2, \quad \mathbf{j} = \sigma_1\sigma_3, \quad \mathbf{k} = \sigma_2\sigma_1.$$

If one interprets the Pauli matrices as representations of linear transformations in the complex plane, and since multiplication by i represents a rotation by 90^o, we see that in some sense the vector basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ of quaternion space corresponds more to something more like dual planes to the basis vectors given by the Pauli matrices. This is one of those places where the factor of i in the structure constants for Pauli matrices causes differences with mathematical standards, the latter following more the elegant algebraic construction by Hamilton. The triplet (q_1, q_2, q_3) of the quaternion is called the vector component. If one defines the quaternion conjugate by

$$\bar{q} = q_0\mathbf{1} - q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k}$$

it follows that

$$\|q\|^2 = q\bar{q} = q_0^2 + q_1^2 + q_2^2 + q_3^2.$$

If we set

$$z_0 = q_0 + q_2\,\mathbf{i}, \quad \text{and} \quad z_1 = q_2 + q_3\,\mathbf{i},$$

we can identify $\mathbf{H}$ with $\mathbf{C} + \mathbf{C}\,\mathbf{j}$, by writing,

$$q = z_0 + z_1\,\mathbf{j}.$$

In this notation, the conjugate of the quaternion is given by

$$\bar{q} = \bar{z}_0 - z_1\,\mathbf{j} = \bar{z}_0 - \mathbf{j}\,\bar{z}_1.$$

The complex conjugate on the last term above comes from the minus sign introduced by the anti-commutation of $\mathbf{i}$ and $\mathbf{j}$, the price for transposing $\mathbf{j}$ to the front. If $q' = w_0 + w_1\,\mathbf{j}$ is another quaternion, the right action of q on q' by quaternion multiplication gives,

$$\begin{aligned} q'q &= (w_0 + w_1\,\mathbf{j})(z_0 + z_1\,\mathbf{j}), \\ &= w_0z_0 + w_0z_1\,\mathbf{j} + w_1\,\mathbf{j}\,z_0 + w_1\,\mathbf{j}\,z_1\,\mathbf{j}, \\ &= (w_0z_0 - w_1\bar{z}_1) + (w_0z_1 + w_1\bar{z}_0)\,\mathbf{j}. \end{aligned}$$

In matrix form the right action of q can be rewritten as

$$[w_0\ w_1] \xrightarrow{q} [w_0\ w_1] \begin{bmatrix} z_0 & z_1 \\ -\bar{z}_1 & \bar{z}_0 \end{bmatrix}.$$

Thus, if q is a unit quaternion, that is, one with $\|q\| = 1$, the matrix on the right above is in generic form of an element of $SU(2)$. The set of all unit quaternions can be identified with a three-sphere $S^3 \in \mathbf{R}^4$. The quaternions form a division algebra. If $q \neq 0$, then, similar to complex numbers, the inverse is given by

$$q^{-1} = \frac{1}{\|q\|^2}\bar{q}$$

Back to the Lie algebra, the quantities $\{i\sigma_1, i\sigma_2, i\sigma_3\}$ represent the infinitesimal transformation that generate the elements of the group. Thus, for example, to generate a rotation by an angle ϕ around the z-axis, we set,

$$Q_\phi = e^{\frac{i}{2}\phi\sigma_3}. \tag{8.16}$$

Proceeding exactly as in the computation leading to equation 8.5, we find

$$\begin{aligned} Q_\phi &= \cos\tfrac{\phi}{2}\, I + i \sin\tfrac{\phi}{2}\, \sigma_3, & (8.17)\\ &= \begin{bmatrix} e^{i\phi/2} & 0 \\ 0) & e^{-i\phi/2} \end{bmatrix}. & (8.18) \end{aligned}$$

The result is a diagonal matrix since σ_3 is diagonal, and hence, so is any power of σ_3. Yet another computation of the similarity transformation $\widetilde{X} = Q_\phi X Q_\phi^\dagger$, where,

$$X = \begin{bmatrix} z & x - iy \\ x + iy & -z \end{bmatrix}, \quad \widetilde{X} = \begin{bmatrix} \tilde{z} & \tilde{x} - i\tilde{y} \\ \tilde{x} + i\tilde{y} & -\tilde{z} \end{bmatrix},$$

yields,

$$\begin{aligned} \tilde{x} &= \quad x\cos\phi + y\sin\phi, \\ \tilde{y} &= -x\sin\phi + y\cos\phi, \\ \tilde{z} &= z. \end{aligned}$$

These are of course the correct equations for the rotation. It should be noted that in the computation, which we leave as an exercise, we find a natural appearance of double angle formulas for sine and cosine; this is how the $\phi/2$ converts to a ϕ in the final equation. The next Euler angle rotation is given by,

$$\begin{aligned} Q_\theta &= e^{\frac{i}{2}\theta\sigma_1}, & (8.19)\\ &= \cos\frac{\theta}{2}\, I + i\sin\frac{\theta}{2}\, \sigma_1, & (8.20)\\ &= \begin{bmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}. & (8.21) \end{aligned}$$

The third Euler angle rotation looks just like the one in equation 8.18 with ϕ replaced by ψ,

$$Q_\psi = \cos\tfrac{\psi}{2}\, I + i\sin\tfrac{\psi}{2}\,\sigma_3, \tag{8.22}$$
$$= \begin{bmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{bmatrix}. \tag{8.23}$$

The composite rotation is given by

$$Q = Q_\psi Q_\theta Q_\phi = \begin{bmatrix} e^{i(\psi+\phi)/2}\cos\frac{\theta}{2} & ie^{i(\psi-\phi)/2}\sin\frac{\theta}{2} \\ ie^{-i(\psi-\phi)/2}\sin\frac{\theta}{2} & e^{-i(\psi+\phi)/2}\cos\frac{\theta}{2} \end{bmatrix}, \tag{8.24}$$

which gives the Cayley-Klein parameters in terms of Euler angles. It should be noted that if Q represents a rotation, then $-Q$ represents exactly the same rotation, since the minus sign cancels out in the similarity transformation. In this sense, $SU(2)$ is called a double cover of $SO(3,\mathbf{R})$.
Let $\mathbf{n} = (n^1, n^2, n^3)$ be a unit vector, and as before, set $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. We call a unit *Pauli-Bloch vector*, denoted by $\mathbf{n}\cdot\boldsymbol{\sigma}$, the expression given by the matrix,

$$\mathbf{n}\cdot\boldsymbol{\sigma} = n^k\sigma_k^{A\dot{B}} = \begin{bmatrix} n^3 & n^1 - in^2 \\ n^1 + in^3 & -n^3 \end{bmatrix}.$$

It is very easy to verify that given two vectors $\mathbf{a}$ and $\mathbf{b}$, we have

$$(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = (\mathbf{a}\cdot\mathbf{b})\, I + i(\mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}$$

Although we do not need the result above here, it is very neat that the formula which is in essence indicates that the product of quaternions, incorporates both, the dot and the cross products. The formula is helpful in establishing identities for products of quaternions. With the notation above, the equation,

$$e^{\frac{i}{2}\theta(\mathbf{n}\cdot\boldsymbol{\sigma})} = \cos\tfrac{\theta}{2}\, I + i(\mathbf{n}\cdot\boldsymbol{\sigma})\sin\tfrac{\theta}{2} \tag{8.25}$$

gives a generalization of Euler's formula extended to quaternions via Pauli matrices. This is a beautiful result which was the goal that led Hamilton to introduce quaternions in 1843. The formula represents a rotation by an angle θ about an axis in the direction of the unit vector $\mathbf{n}$. In terms of Hamilton quaternions, the rotation matrix in $\mathbf{R}^3$ is obtained by conjugation with a quaternion q,

$$\tilde{X} = qXq^{-1},$$

where,

$$R(q) \equiv R(\theta,\mathbf{n}) = \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2q_1q_2-2q_0q_3 & 2q_1q_3+2q_0q_2 \\ 2q_1q_2+2q_0q_3 & q_0^2-q_1^2+q_2^2-q_3^2 & 2q_2q_3-2q_0q_1 \\ 2q_1q_3-2q_0q_2 & 2q_2q_3-2q_0q_1 & q_0^2-q_1^2-q_2^2+q_3^2 \end{bmatrix}, \tag{8.26}$$

with,

$$q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k},$$
$$= \cos\tfrac{\theta}{2} + [n_1\mathbf{i} + n_2\mathbf{j} + n_3\mathbf{k}]\sin\tfrac{\theta}{2}.$$

This is clearly a generalization corresponding rotation matrix 8.1 in $\mathbf{R}^2$ in terms of half-angle parameters. The formulation is not cluttered by factors of i; this is the preferred form for computer scientists, computer game developers, and an increasing number of engineers, to code rotations in numerical computations.
The Maurer-Cartan form ω of $SU(2)$ can be computed directly from the Cayley-Klein parameters,

$$\omega = Q^{-1} dQ.$$

In the computation we write the Lie algebra valued form as

$$\omega = \begin{bmatrix} \omega^3 & \omega^1 - i\,\omega^2 \\ \omega^1 + i\,\omega^2 & -\omega^3 \end{bmatrix}.$$

We then compute $Q^{-1}dQ$ and read the forms. The computation is actually easier by hand than using Maple, but we recommend a pen with an extra fine tip, and working on a sheet of paper in landscape orientation. The computation is facilitated by noting that we only have to compute the first column to read the components of the form. The result is

$$\begin{aligned} \omega^1 &= \cos\phi\, d\theta + \sin\phi \sin\theta\, d\psi, \\ \omega^2 &= \sin\phi\, d\theta - \cos\phi \sin\theta\, d\psi, \\ \omega^3 &= d\phi + \cos\theta\, d\psi. \end{aligned} \tag{8.27}$$

One can then verify that

$$d\omega^i = \frac{1}{2}\epsilon^i{}_{jk}\omega^j \wedge \omega^k,$$

from which we get the metric associated with the Killing form

$$\begin{aligned} ds^2 &= \delta_{ij}\, \omega^i \omega^j, \\ &= (\omega^1)^2 + (\omega^2)^2 + (\omega^3)^2, \\ &= d\theta^2 + \sin^2\theta\, d\psi^2 + (d\phi + \cos\theta\, d\psi)^2. \end{aligned} \tag{8.28}$$

For a discussion of the dynamics of rigid bodies using Euler angles, see the book Classical Mechanics by Goldstein [11].

8.1.4 Hopf Fibration

In this section we discuss fibration structures over the projective spaces $\mathbf{FP}^1$, where $\mathbf{F}$ stands for one of the division algebras $\{\mathbf{R}, \mathbf{C}, \mathbf{H}, \mathbf{O}\}$, that is the real, the complex, the quaternion, and the octonion algebras. The classical Hopf fibration is the one associated with $\mathbf{CP}^1$, but for pedagogical reasons, it might be more instructive to start with the simpler case of the projective line.

Hopf map on $\mathbf{RP}^1$

Let (q_0, q_1) be coordinates in $\mathbf{R}^2$ and consider the equivalence relation $(q_0, q_1) \sim (\lambda q_0, \lambda q_1)$. The projective line $\mathbf{RP}^1$ is defined by the quotient of the plane with this equivalence relation,

$$\mathbf{RP}^1 = (\mathbf{R}^2 - \{0\})/\sim$$

Geometrically, $\mathbf{RP}^1$ consists of the space of lines through the origin in $\mathbf{R}^2$. The coordinates (q_0, q_1) are called homogenous coordinates of $\mathbf{RP}^1$. Keeping in mind that what really determines a point on the projective space are the ratios of the coordinates, we can cover the manifold with two patches corresponding to $\{q_0/q_1,\ q_1 \neq 0\}$ and $\{q_1/q_0,\ q_0 \neq 0\}$. We subject the coordinates to the restriction

$$q_0^2 + q_1^2 = 1.$$

The equation represents a unit circle S^1 centered at the origin in $\mathbf{R}^2$. The circle can be parametrized by

$$q_0 = \cos\theta, \quad q_1 = \sin\theta.$$

The restriction implies that $|\lambda| = 1$. Topologically, the set of such λ's is the 0-sphere $S^0 = \{1, -1\}$, and has the structure of the group $\mathbf{Z}_2$. Every line through the origin intersects the unit circle in exactly two antipodal points which determine the same line, so we have a fibration

$$\mathbf{Z}_2 \hookrightarrow S^1 \xrightarrow{\pi} S^1.$$

The Hopf map π for this fibration is defined by

$$\pi(q_0, q_1) = (2q_0q_1, |q_0|^2 - |q_1|^2)$$

Of course, the absolute values in the equation above are redundant, but they are included here for motivation for the other Hopf fibrations. If we associate (q_0, q_1) with a rotation matrix in $SO(2)$, the reader will recognize this map as the representation of rotations by half-angles 8.1,

$$\begin{bmatrix} q_0 & q_1 \\ -q_1 & q_0 \end{bmatrix} \xrightarrow{\pi} \begin{bmatrix} q_0^2 - q_1^2 & 2q_0q_1 \\ -2q_0q_1 & q_0^2 - q_1^2 \end{bmatrix}.$$

If we were to try to define a rotation by quaternions in two dimensions, this would be it. Let ζ be the coordinate in $\mathbf{R}$ representing the stereographic projection of a point $(x, y) \in S^1 \xrightarrow{\pi_S} \mathbf{R}$. We recall from equation 5.66 that

$$(x, y) = \left(\frac{2\zeta}{\zeta^2 + 1}), \frac{\zeta^2 - 1}{\zeta^2 + 1}\right)$$

If we now let $\zeta = q_0/q_1$ and simplify the expression above, we get

$$x = 2q_0q_1, \quad y = q_0^2 - q_1^2,$$

which are precisely the coordinates of the image of the Hopf map. In other words, we have a remarkable relation between the Hopf map and the stereographic projection of the base space given by,

$$\pi(q_0, q_1) = \pi_S^{-1}(q_0/q_1),\ q_1 \neq 0.$$

If ζ_1 is the coordinate patch associated with the stereographic projection from the north pole and ζ_2 the patch associate with the south pole, then in the overlap region the transition functions that glue the fibration are given by

$$\phi_{12} = \frac{\zeta_1}{\zeta_2}$$

Hopf map on $\mathbf{CP}^1$

The matrices that represent elements of $SU(2)$ can be written in terms of the Cayley-Klein parameters

$$Q = \begin{bmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{bmatrix},$$

where,

$$\alpha = e^{i(\psi+\phi)/2} \cos \tfrac{\theta}{2}, \tag{8.29}$$

$$\beta = e^{i(\psi-\phi)/2} \sin \tfrac{\theta}{2}. \tag{8.30}$$

With apologies to the purists, we leave out a factor of i in the β parameter in this section. This is done for the sake of better consistency with other formalism that we need for this discussion. We present the Hopf map in terms of Euler angle rotations, but we could just as easily use Hamilton quaternion variables. Since $\det Q = 1$, we have,

$$|\alpha|^2 + |\beta|^2 = 1 \tag{8.31}$$

This can be corroborated immediately since,

$$|\alpha|^2 + |\beta|^2 = \cos^2 \tfrac{\theta}{2} + \sin^2 \tfrac{\theta}{2} = 1$$

We write $(\alpha, \beta) \in \mathbf{C} \times \mathbf{C}$ in the form,

$$\alpha = x^1 + ix^2, \quad \beta = x^3 + ix^4.$$

As a vector space, $\mathbf{C}^2 \cong \mathbf{R}^4$, so equation 8.31 gives parametric equations for a unit sphere $S^3 \in \mathbf{R^4}$.

$$(x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 = 1.$$

In other words, the set of unit quaternions $U(1, \mathbf{H})$ is a sphere S^3 in analogy to $U(1, \mathbf{C})$ which describes a circle S^1. The *classical Hopf fibration* (or Hopf bundle) is the map, $\pi : S^3 \to S^2$ given by,

$$\pi(\alpha, \beta) = (2\alpha\bar{\beta}, |\alpha|^2 - |\beta|^2) \subset \mathbf{C} \times \mathbf{R} \cong \mathbf{R}^3, \tag{8.32}$$

or, in matrix form

$$\begin{bmatrix} z & x-iy \\ x+iy & -z \end{bmatrix} = \begin{bmatrix} |\alpha|^2-|\beta|^2 & 2\alpha\bar{\beta} \\ 2\bar{\alpha}\beta & |\beta|^2-|\alpha|^2 \end{bmatrix}.$$

Indeed, for all α and β, the image of this map is in S^2 because

$$\begin{aligned} |\pi(\alpha,\beta)|^2 &= 4|\alpha|^2|\beta|^2 + (|\alpha|^2-|\beta|^2)^2, \\ &= (|\alpha|^2+|\beta|^2)^2, \\ &= 1 \end{aligned}$$

Any other point (α',β') that maps to the same point $\pi(\alpha,\beta)$ must satisfy $(\alpha',\beta') = (\lambda\alpha,\lambda\beta)$ for some complex number with $|\lambda|^2 = 1$. When this happens, we say that these points are in the same equivalence class. Then, the projective space defined as,

$$\mathbf{CP}^1 = (\mathbf{C}^2 - \{0\})/\sim$$

represents the space of complex lines through the origin of $\mathbf{C^2}$. The complex projective plane $\mathbf{CP}^1$ has the structure of a compact complex manifold of complex dimension 1. Geometrically, it can be viewed as sphere S^2 in which antipodal points are identified. In quantum physics, and quantum computing, this is called the *Bloch* sphere. Points in $\mathbf{CP}^1$ can be described by *homogeneous* coordinates (α,β) as representative of the equivalence classes, or by *inhomogeneous coordinates*

$$\zeta_1 = \frac{\alpha}{\beta},\ \beta \neq 0, \quad \text{or} \quad \zeta_2 = \frac{\beta}{\alpha},\ \alpha \neq 0.$$

Figure 8.2 depicts a somewhat misleading but still useful visualization of

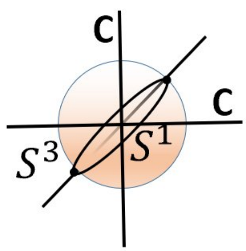


Fig. 8.2: $\mathbf{CP}_1$. Intersection of complex lines with S^3 are S^1's.

the construction of $\mathbf{CP}^1$. The horizontal and vertical axes are copies of $\mathbf{C}$ parametrized by α and β. Since a complex line is really a plane, the cross product is 4-dimensional. The unit sphere centered at the origin is given by the equation $|\alpha|^2 + |\beta|^2 = 1$, so this is a three sphere S^3. The intersection of a complex line with S^3 is a circle S^1. The only point common to two different lines through the origin is the origin, so the corresponding circles of intersection are disjoint. The collection of all these circles is parametrized by a two sphere

S^2. In other words, if $p \in \mathbf{CP}^1 \cong S^2$ then $\pi^{-1}(p)$ is a circle $S^1 \in S^3$. The circles S^1 are called the fibers of the fibration.

$$\pi : U(2)/U(1) \simeq S^3 \xrightarrow{S^1} S^2. \tag{8.33}$$

What is not at all obvious is that any of these circles is linked exactly once with any other circle of the fibration. To "unpack" this fibration in familiar terms, we first compute the map in Cayley-Klein coordinates,

$$\begin{aligned}
2\alpha\bar{\beta} &= 2[e^{i(\psi+\phi)/2}\cos\tfrac{\theta}{2}][e^{-i(\psi-\phi)/2}\sin\tfrac{\theta}{2}],\\
&= 2e^{i\phi}\cos\tfrac{\theta}{2}\sin\tfrac{\theta}{2},\\
&= e^{i\phi}\sin\theta,\\
|\alpha|^2-|\beta|^2 &= \cos^2\tfrac{\theta}{2}-\sin^2\tfrac{\theta}{2},\\
&= \cos\theta.
\end{aligned}$$

Let $\omega = 2\alpha\bar{\beta}$. Identifying $\mathbf{C}\times\mathbf{R}$ with $\mathbf{R}^3$, that is, taking $x = \Re(\omega)$, $y = \Im(\omega)$, we get a point in S^2 in spherical coordinates

$$\begin{aligned}
x &= \cos\phi\sin\theta,\\
y &= \sin\phi\sin\theta,\\
z &= \cos\theta.
\end{aligned}$$

Let ζ be the complex number in equation 5.61, whose inverse image under the stereographic projection π_s gives the coordinates on the sphere,

$$(x,y,z) = \left(\frac{\zeta+\bar{\zeta}}{\zeta\bar{\zeta}+1}, \frac{\zeta-\bar{\zeta}}{i(\zeta\bar{\zeta}+1)}, \frac{\zeta\bar{\zeta}-1}{\zeta\bar{\zeta}+1}\right)$$

Setting $\zeta = \alpha/\beta$ to be the inhomogeneous coordinates of $\mathbf{CP}^1$ and simplifying the double fraction using the fact that $|\alpha|^2+|\beta|^2=1$, we get the Hopf map 8.32. That is, a point $\pi(\alpha,\beta)$ in S^2 given by the image of the Hopf map is just the point in S^2 obtained by the inverse stereographic projection

$$\pi_s^{-1}\left(\frac{\alpha}{\beta}\right).$$

As stated above, the fiber of a point in S^2 is a circle S^1 in S^3. The fibers of points on a circle in S^2 parallel to the equator, are linked circles that lie on a torus - these are called *Villarceau circles.* Geometrically, the Villarceau circles are obtained by the intersection of a torus and a plane tangential to antipodal images of the generating circle. Hopf discovered the fibration in 1931, but I only learned about Hopf fibrations in 1975 from studying Taub's solution to Einstein's equation. Taub's metric has topology $\mathbf{R}\times S^3$ and spatial $SU(2)$ symmetry. The Taub metric is of the form

$$ds^2 = -d\tau^2 + \eta_{ij}(\tau)\omega^i\otimes\omega^j \tag{8.34}$$

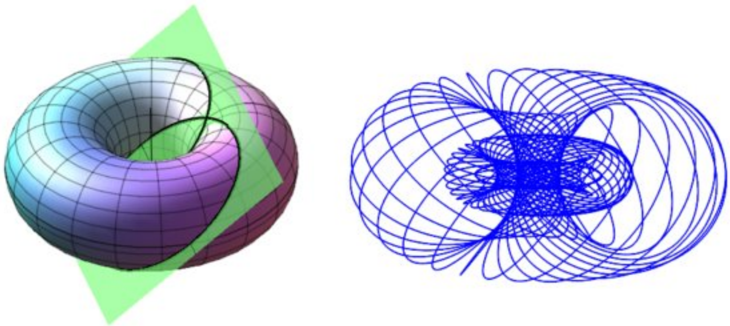


Fig. 8.3: Hopf Fibration

where ω is essentially the Maurer-Cartan form of $\mathfrak{su}(2)$. This is an example of a gravitational instanton. However, the first time I saw a pictorial representation of the fibration was a magnificent hand drawing made by Roger Penrose discussing Robinson Congruences in the context of twistor theory; a reproduction of this drawing appears in [29], for example. One has to marvel at the earlier masters who were able to visualize this complex structure. For us lesser humans, nowadays it takes little effort to render the images with a computer algebra system, by lifting a parallel circle in S^2 to S^3, followed by a stereographic projection from S^3 to $\mathbf{R}^3$. The nested toroidal S^1 links in figure 8.3 are the fibers of three circular parallels in the base space S^2. The reader will find a beautiful explanation of Villarceau circles in a paper by Hirsch [15].

There is a generalization of Hopf fibrations with S^1 fibers to all complex projective spaces, $\mathbf{CP}^n$. Consider the space $\mathbf{C}^n$ with coordinates $Z = (z_1, z_2, \ldots, z_n)$. The equation $|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 = 1$ describes a sphere $S^{2n+1} \subset \mathbf{C}^n$. The space $\mathbf{CP}^n$ of complex lines through the origin. Let $a, b \in S^{2n+1}$. Define an equivalence relation $a \sim b$, if there exists $c \in S^1$ such that $a = bc$. The idea is that a complex line through the origin is a 2-real dimensional plane, which intersects the sphere on a circle. All points in such circles are in the same equivalence class. $\mathbf{CP}^n$ can be identified with the space $\mathbf{C}^n/\sim$. The generalized fibration is,

$$S^{2n+1} \xrightarrow{S^1} \mathbf{CP}^n. \tag{8.35}$$

The corresponding fibration in the real case is

$$S^n \xrightarrow{Z_2} \mathbf{RP}^n, \tag{8.36}$$

since a line through the origin intersects the sphere S^n in two antipodal points. The group elements are the identity and the antipodal map.

Hopf map on $\mathbf{HP}^1$

The construction of the Hopf fibration over quaternion space, follows along the same lines. Let (q_1, q_2) be quaternion coordinates in $\mathbf{H}^2 \simeq \mathbf{R}^8$, with

$$
\begin{aligned}
q_1 &= x^1 + x^2 + \mathbf{i} + x^3\mathbf{j} + x^4\mathbf{k}, \\
q_2 &= x^5 + x^6 + \mathbf{i} + x^7\mathbf{j} + x^8\mathbf{k},
\end{aligned}
$$

Introduce complex coordinates,

$$
\begin{aligned}
z_1 &= x^1 + x^2\mathbf{i}, \quad & z_3 &= x^5 + x^6\mathbf{i}, \\
z_2 &= x^3 + x^4\mathbf{i}, \quad & z_4 &= x^7 + x^8\mathbf{i},
\end{aligned}
$$

so that

$$
\begin{aligned}
q_1 &= z_1 + z_2\mathbf{j}, \\
q_2 &= z_3 + z_4\mathbf{j}.
\end{aligned} \tag{8.37}
$$

Consider the equivalence relation $(q_1, q_2) \sim (\lambda q_1, \lambda q_2), ;\ \lambda \in \mathbf{H}$, and define the quaternionic projective space

$$
\mathbf{HP}^1 = (\mathbf{H}^2 - 0)/\sim .
$$

As before, (q_1, q_2) are homogeneous coordinates representing equivalence classes of quaternionic lines. The space can be covered by two inhomogeneous coordinate charts

$$
\zeta_1 = \frac{q_1}{q_2},\ q_2 \neq 0, \quad \text{or} \quad \zeta_2 = \frac{q_2}{q_1},\ q_1 \neq 0.
$$

We impose the condition,

$$
|q_1|^2 + |q_2|^2 = \sum_{k=1}^{8} |x_k|^2 = 1,
$$

which represents a sphere $S^7 \in \mathbf{H}^2$. This implies that on the sphere S^7, the λ's are unit quaternions, that is $|\lambda|^2 = 1$. Thus, the fibers are 3-spheres, and we have a fibration

$$
\begin{aligned}
&S^3 \hookrightarrow S^7 \xrightarrow{\pi} S^4, \\
&S^3 \hookrightarrow Sp(2)/Sp(1) \xrightarrow{\pi} Sp(1).
\end{aligned}
$$

The Hopf map $\pi : S^7 \to S^4$ is defined by

$$
\pi(q_1, q_2) = (2q_1\bar{q}_2, |q_1|^2 - |q_2|^2) \in \mathbf{H} \times \mathbf{R} \simeq \mathbf{R^5}
$$

These look like the usual suspects. Let $\xi, \eta \in \mathbf{C}$ so that

$$
\xi + \eta\, \mathbf{j} \in \mathbf{H},
$$

and set

$$\begin{aligned} \xi + \eta\,\mathbf{j} &= 2q_1\bar{q}_2, \\ z &= |q_1|^2 - |q_2|^2 \end{aligned}$$

We can then arrange the Hopf map in familiar (quaternionic) matrix form

$$\begin{bmatrix} z & \xi - \eta\,\mathbf{j} \\ \xi + \eta\,\mathbf{j} & -z \end{bmatrix} = \begin{bmatrix} |q_1|^2 - |q_2|^2 & 2q_1\bar{q}_2 \\ 2\bar{q}_1 q_2 & |q_2|^2 - |q_1|^2 \end{bmatrix}$$

We can easily corroborate that (ξ, η, z) represent points in S^4, again by the familiar process

$$\begin{aligned} |\xi|^2 + |\eta|^2 + |z|^2 &= |2q_1\bar{q}_2|^2 + (|q_1|^2 - |q_2|^2)^2, \\ &= |q_1|^4 + 2|q_1|^2|q_2|^2 + |q_2|^4, \\ &= (|q_1|^2 + |q_2|^2)^2 = 1. \end{aligned}$$

One can be a bit more explicit, inserting the complex coordinates 8.37 and carrying out the short computation. We get

$$\begin{aligned} \xi &= 2(z_1\bar{z}_3 + z_2\bar{z}_2), \\ \eta &= 2(z_2 z_3 - z_1 z_4), \\ z &= |z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 \end{aligned}$$

At this point it should not be surprising that if one denotes by ζ_1 the quaternion representing a point on S^4 under the stereographic projection π_S from the north pole to $\mathbf{H} \simeq \mathbf{R}^4$, then the quaternionic Hopf map is related to this projection by the

$$\pi(q_0, q_1) = \pi_S^{-1}\left(\frac{q_0}{q_1}\right)$$

If ζ_1 and ζ_2 represent charts overlapping over a narrow band around the "equator" under the projective maps from the north and south pole respectively, then on the overlap the transition functions are

$$\phi_{12} = \frac{\zeta_1}{\zeta_2}$$

and its inverse on the other direction.

For now, we will stay away from the octonion algebra because it is not associative, however, there is also an projective octonion line version of the Hopf map. The results are summarized in the following list,

$$\begin{aligned} S^0 &\hookrightarrow S^1 \longrightarrow S^1 \cong \mathbf{RP}^1, \\ S^1 &\hookrightarrow S^3 \longrightarrow S^2 \cong \mathbf{CP}^1, \\ S^3 &\hookrightarrow S^7 \longrightarrow S^4 \cong \mathbf{HP}^1, \\ S^7 &\hookrightarrow S^{15} \rightarrow S^8 \cong \mathbf{OP}^1. \end{aligned}$$

The Hopf fibration is a seminal discovery in algebraic topology because, through a formalism called the long, exact, homotopy sequence of a fibration, it became possible to establish the existence of the first, non-vanishing, high dimension homotopy groups of spheres. The long exact homotopy sequence applied to $S^3 \xrightarrow{\pi} S^2$ yields the result

$$\pi^3(S^2) \cong \pi^2(S^2) = \mathbf{Z}.$$

The Hopf fibration associated with $\mathbf{CP}^1 = \mathbf{P}^1(\mathbf{C})$ describes a singly-charged Dirac monopole, and the fibration associated with $\mathbf{HP}^1$ enters in the description of a Yang-Mills instanton. These are explored in chapter 9 after properly introducing connections on principal fiber bundles.

8.1.5 Angular Momentum

As indicated in the preface to this book, we present a simplified and limited version of basic quantum mechanics for the benefit of those mathematics students who have no formal training on the subject. Quantum Mechanics was developed in 1926 by Schrödinger and Heisenberg. The axiomatic description here is a summary of the framework as envisioned by Dirac and von-Neumann. The axioms really are axioms in the sense of Euclid; they cannot be proved. The axioms are not self-evident, but they are founded on experience and physical intuition.

- *Postulate QM1* The state of a particle is described by a wave function $\psi(t, x, y, z)$ in some complex Hilbert space $\mathcal{H}$ with inner product $\langle \mid \rangle$. The quantity,

  $$\psi^* \psi d^3\mathbf{r}$$

  represents a probability density of finding the particle within a volume element $d^3\mathbf{r}$. The total probability is,

  $$P = \int_{\mathcal{H}} \psi^* \psi d^3\mathbf{r} = 1.$$

- *Postulate QM2* Measurable (or observable) quantities such as energy and momentum, are represented by a linear Hermitian operator L acting on ψ. The measurement of the state is given by the expectation value,

  $$\begin{aligned} \langle \psi \rangle &= \langle \psi^* | L | \psi \rangle, \\ &= \int_{\mathcal{H}} \psi^* L \psi d^3\mathbf{r} \end{aligned}$$

- *Postulate QM3* From the spectral theorem for Hermitian operators, the possible outcomes of the observables are the eigenvalues of the operator. The eigenvalues of real and eigenstates corresponding to different eigenvalues are orthogonal.

The position operator is multiplication. The linear momentum and energy operators are obtained intuitively by starting with a classical solution to the wave equation. In dimension one, the quantity,

$$\psi = Ae^{i(kx-\omega t)}$$

is such a solution with speed $v = k/\omega$. The energy E and momentum p are related to the wave number k and the angular frequency ω by the equations,

$$p = \hbar k, \quad E = \hbar\omega,$$

so that,

$$\psi = Ae^{\frac{i}{\hbar}(px-Et)}.$$

Taking partial derivatives with respect to x and t respectively, we get,

$$\begin{aligned}\frac{\partial}{\partial x}\psi &= \frac{i}{\hbar}p\,\psi,\\ \frac{\partial}{\partial t}\psi &= -\frac{i}{\hbar}E\psi,\end{aligned}$$

By ansatz, the choice for the operators is,

$$\begin{aligned}\hat{p_x} &= \frac{\hbar}{i}\frac{\partial}{\partial x},\\ \hat{E} &= i\hbar\frac{\partial}{\partial t},\end{aligned}$$

Generalizing momentum to dimension 3, the operators become,

$$\begin{aligned}\mathbf{p} &= \frac{\hbar}{i}\nabla,\\ \hat{E} &= i\hbar\frac{\partial}{\partial t}.\end{aligned}$$

The operator for total energy operator called the Hamiltonian H, is the Kinetic energy $KE = \mathbf{p}^2/(2m)$ plus the potential energy $PE = V$. Thus, Schrödinger was led to the quantum mechanics equation,

$$\begin{aligned}H\psi &= \hat{E}\psi,\\ -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi &= i\hbar\frac{\partial\psi}{\partial t}.\end{aligned} \tag{8.38}$$

For a free particle for which the energy does not depend on time, the stationary states are described by the discrete set of eigenfunctions and energy eigenvalues of the equation,

$$-\frac{\hbar^2}{2m}\nabla^2\Psi_n = E_n\Psi_n,$$

where

$$\Psi_n = e^{-(i/\hbar)E_n t}\psi_n(\mathbf{r}),$$

and ψ_n depends only one the spatial coordinates $\mathbf{r} = (x_1, x_2, x_3) = (x, y, z)$. Let $\mathbf{p} = (p_x, p_y, p_y)$ be the components of the momentum operator. The following basic commutation relations hold,

$$\begin{aligned} [x_i, x_j] &= 0, \\ [p_i, p_j] &= 0, \\ [p_i, x_j] &= i\hbar\delta_{ij}. \end{aligned} \tag{8.39}$$

The first commutation relation above is trivial since in number multiplication, the order of the factors does not alter the product. The commutation relation for two momenta follows from the symmetry of indices of second order partial derivatives. The third commutation relation can done for each pair of indices. In most elementary quantum mechanics books, one example is worked out and the rest are taken on faith or left as an exercise. Here is one example. Let f be an arbitrary function; compute,

$$\begin{aligned} [p_x, x](f) &= -i\hbar\left(\frac{\partial}{\partial x}(xf) - x\frac{\partial}{\partial x}f\right), \\ &= -i\hbar(f + xf_x - xf_x), \\ &= -i\hbar f, \\ [p_x, x] &= -i\hbar \end{aligned}$$

At this stage, having gained experience in manipulating indices, it is just as easy to do all the cases at once,

$$\begin{aligned} [p_i, x_j]f &= -i\hbar\left(\frac{\partial}{\partial x_i}(x_j f) - x_j\frac{\partial}{\partial x_i}f\right), \\ &= -i\hbar\left(\frac{\partial x_j}{\partial x_i}f + x_j\frac{\partial f}{\partial x_i} - x_j\frac{\partial f}{\partial x_i}\right), \\ &= -i\hbar\delta_{ij}f \\ [p_i, x_j] &= -i\hbar\delta_{ij}. \end{aligned}$$

The definition of the angular momentum operator in quantum mechanics is given by simple extension of the classical formula,

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}. \tag{8.40}$$

The explicit components of angular momentum are,

$$\begin{aligned} L_x &= yp_z - zp_y, \\ L_y &= zp_x - xp_z, \\ L_z &= xp_y - yp_x. \end{aligned} \tag{8.41}$$

In index notation,

$$L_i = \epsilon_i{}^{jk} x_j p_k. \tag{8.42}$$

Let us derive the following commutation relations involving angular momentum,

$$
\begin{aligned}
[L_i, x_j] &= i\hbar \epsilon^k{}_{ij} x_k, \\
[L_i, p_j] &= i\hbar \epsilon^k{}_{ij} p_k,
\end{aligned}
\tag{8.43}
$$

For the first equation, we can demonstrate an instance,

$$
\begin{aligned}
[L_x, y]f &= (yp_z - zp_y)(yf) - y(yp_z - zp_y)f, \\
&= y^2 p_z f - zp_y(yf) - y^2 p_z f + yzp_y f \\
&= -z[p_y, y]f, \\
[L_x, y] &= i\hbar z.
\end{aligned}
$$

But, since we have already introduced the Levi-Civita symbol 2.41, we can use the momentum commutators 8.41 and have fun doing all the cases at once. Here is the computation,

$$
\begin{aligned}
[L_i, x_j]f &= [\epsilon_i{}^{km} x_k p_m, x_j]f, \\
&= \epsilon_i{}^{km} x_k p_m(x_j f) - x_j(\epsilon_i{}^{km} x_k p_m)f, \\
&= x_k \epsilon_i{}^{km} p_m(x_j f) - x_j p_m(f) \\
&= x_k \epsilon_i{}^{km} [p_m, x_j]f \\
&= -i\hbar x_k \epsilon_i{}^{km} \delta_{mj}, \\
&= -i\hbar x_k \epsilon_i{}^k{}_j, \\
&= i\hbar x_k \epsilon^k{}_{ij}
\end{aligned}
$$

The formula for the commutator $[L_i, p_j]$ is very similar and we leave it as an exercise. Instead, we go for the gold of the commutators.

8.1.1 Proposition

$$
[L_i, L_j] = i\hbar \epsilon^k{}_{ij} L_k. \tag{8.44}
$$

As above, we show that the concept is easy by doing the following case

$$
\begin{aligned}
[L_x, L_y] &= L_x(zP_x - xp_z) - (zP_x - xp_z)L_x, \\
&= (L_x z p_x - zP_x L_x) - (L_x x p_z - x p_z L_x), \\
&= (L_x z - zL_x)p_x - x(L_x p_z - p_z L_x), \\
&= -i\hbar y p_x + i\hbar x p_y, \\
&= i\hbar(x p_y - y p_x), \\
&= i\hbar L_z.
\end{aligned}
$$

In the third line above we used $p_x L_x = L_x p_x, \; L_x x = x L_x$. To do all cases at once, one needs the product of permutation symbol identities 2.46. This is a great exercise in index gymnastics but hides the simplicity of the two other independent cases that can be done as above, Equation 8.44 is the primary reason why the topic of angular momentum is included in this section. We will work in units of $\hbar$, that is, we set $\hbar = 1$. Then comparing with the commutator relations 8.14 for the Pauli matrices, we see that apart from a factor of 2, we see that the components of $\mathbf{L}$ are generator of the lie algebra $\mathfrak{su}(2)$. Following the QM postulates, we seek a Hilbert space on which angular momentum acts as a linear operator, to find a function space that provides a representation for the algebra. Naturally, we seek such functions over a two sphere as the base space. The standard process one finds in most classic books on quantum mechanics might look a bit mysterious to those who see it for the first time, but as we will demostrate, it is just an implementation of the Cartan subalgebra for the angular momentum representation of the algebra.

In the case of $\mathfrak{su}(2)$ there is only one generator, in the Cartan subalgebra, which we choose to be L_z, so, the rank of the algebra is one. We look for a representation in which L_z is diagonal. We also seek a *Casimir operator*, namely, an operator that commutes with all basis elements of $\mathfrak{g}$ The number of Casimir operators in a semisimple Lie algebra is equal to the rank of the algebra. The candidate for the Casimir operator for $\mathfrak{su}(2)$ is,

$$\mathbf{L}^2 = L_x^2 + L_y^2 + L_z^2. \tag{8.45}$$

We show that the operator commutes will all three generators, that is.

$$[\mathbf{L}^2, L_x] = [\mathbf{L}^2, L_y] = [\mathbf{L}^2, L_z] = 0$$

Let's show for instance, that $[\mathbf{L}^2, L_z] = 0$. We need to establish that

$$[L_x^2 + L_y^2 + L_z^2, L_z] = 0.$$

First, it is easy to verify by direct computation of both sides that in general,

$$[A^2, B] = A[A, B] + [A, B]A$$

Applying this identity to the square of the components of $\mathbf{L}$, we get,

$$\begin{aligned}
[L_x^2, L_z] &= L_x[L_x, L_z] + [L_x, L_z]L_x, \\
&= -iL_xL_y - iL_yL_x, \\
[L_y^2, L_z] &= L_y[L_y, L_z] + [L_y, L_z]L_y, \\
&= iL_yL_x + iL_xL_y. \\
[L_z^2, L_z] &= 0.
\end{aligned}$$

Adding the last three equations yields,

$$\begin{aligned}
[\mathbf{L}^2, L_z] &= [L_x^2 + L_y^2 + L_z^2, L_z], \\
&= -iL_xL_y - iL_yL_x + iL_yL_x + iL_xL_y, \\
&= 0.
\end{aligned}$$

We introduce the *ladder* operators,

$$\begin{aligned}
L_+ &= L_x + iL_y, \\
L_- &= L_x - iL_y,
\end{aligned} \tag{8.46}$$

We will see below that the ladder operators are the raising and lowering operators of the algebra. The process of finding the irreducible representations starts with establishing a number of commutation relationships. First, it is easy to verify that,

$$\begin{aligned} L_+L_- &= L_x^2 + L_y^2 + L_z, \\ L_-L_+ &= L_x^2 + L_y^2 - L_z. \end{aligned} \tag{8.47}$$

Indeed, we have,

$$\begin{aligned} L_+L_- &= (L_x + iL_y)(L_x - iL_y), \\ &= L_x^2 + L_y^2 + i(L_yL_x - L_xL_y), \\ &= L_x^2 + L_y^2 + i[L_y, L_x], \\ &= L_x^2 + L_y^2 + L_z, \end{aligned}$$

and similarly for L_+L_-. It is also easy to verify that,

$$\begin{aligned} [L_+, L_-] &= 2L_z, \\ [L_z, L_+] &= L_+, \\ [L_z, L_-] &= -L_-. \end{aligned} \tag{8.48}$$

These commutators define the Cartan subalgebra. Since in this case, the subalgebra is one-dimensional, the roots are the vectors in $\mathbf{R}$ given by $r^{(1)} = (1)$ and $r^{(-1)} = (-1)$. We associate the 0 root with L_z. The root diagram called A_1 is a simple as it gets; it consists of two unit vectors at the origin in $\mathbf{R}$. Putting the commutator results above together leads to the following formula,

$$\begin{aligned} \mathbf{L^2} &= L_+L_- + L_z^2 - L_z, \\ &= L_-L_+ + L_z^2 + L_z. \end{aligned} \tag{8.49}$$

The result follows directly from equation 8.47. We show the steps for the first of these.

$$\begin{aligned} L_+L_- &= L_x^2 + L_y^2 + L_z, \\ &= \mathbf{L}^2 - L_z^2 + L_z, \\ \mathbf{L^2} &= L_+L_- + L_z^2 - L_z, \end{aligned}$$

The formalism can then be used to obtain the ubiquitous expression for the momentum operator of a single particle in spherical coordinates. The computation starts with inverting the Jacobian matrix in 2.30,

$$\frac{\partial}{\partial x} = \sin\theta\cos\phi\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\cos\phi\frac{\partial}{\partial\theta} - \frac{\sin\phi}{r\sin\theta}\frac{\partial}{\partial\phi}, \tag{8.50}$$

$$\frac{\partial}{\partial y} = \sin\theta\sin\phi\frac{\partial}{\partial r} + \frac{1}{r}\cos\theta\sin\phi\frac{\partial}{\partial\theta} + \frac{\cos\phi}{r\sin\theta}\frac{\partial}{\partial\phi}, \tag{8.51}$$

$$\frac{\partial}{\partial z} = \cos\theta\frac{\partial}{\partial r} - \frac{1}{r}\sin\theta\frac{\partial}{\partial\theta}. \tag{8.52}$$

After some algebraic manipulations, we get,

$$
\begin{aligned}
L_z &= -i\tfrac{\partial}{\partial\phi}, \\
L_\pm &= e^{\pm i\phi}\left(\pm\tfrac{\partial}{\partial\theta} + i\cot\theta\tfrac{\partial}{\partial\phi}\right), \\
\mathbf{L}^2 &= -\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial f}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}\right]
\end{aligned}
\tag{8.53}
$$

Comparing with equation 3.9 we recognize $\mathbf{L^2}$ as the angular part of the spherical Laplacian. Perhaps the most direct way to get the function space is to solve Laplace's equation in spherical coordinates by separation of variables, that is, assuming that the solution is of the form $\psi = \Theta(\theta)\Phi(\phi)$. The complete process of obtaining the solutions is best suited for a course on partial differential equations or a course in electrodynamics, so we just present the result. The equation is manifestly self-adjoint, hence the eigenvalues are real and eigenvectors corresponding to different eigenvalues are orthogonal. The eigenvalue equations are,

$$
\begin{aligned}
L_z Y_{l,m}(\theta,\phi) &= mY_{l,m}(\theta,\phi) \\
\mathbf{L}^2 Y_{l,m}(\theta,\phi) &= l(l+1)Y_{l,m}(\theta,\phi).
\end{aligned}
\tag{8.54}
$$

Here, eigenvalues $l(l+1)$ are the orbital quantum numbers l are positive integers. The solution of Laplace's equation in spherical coordinates are the well-known spherical harmonics Y_{lm},

$$
Y_{l,m} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}}P_{lm}(\cos\theta)e^{im\phi}, \tag{8.55}
$$

where P_{lm} are the associated Legendre polynomials given by Rodrigues' formula,

$$
P_{lm}(\cos\theta) = \frac{(-1)^l}{2^l l!}\sin^m\theta\left(\frac{d}{d(\cos\theta)}\right)^{m+l}(\sin^{2l}\theta). \tag{8.56}
$$

For each l there are $2l+1$ possible values for$m$ given by the integers $m = -l\ldots l$.

The eigenfunctions ψ can also be obtained by a method which is almost entirely algebraic. First, the eigenvalue equation for the z-component of angular momentum is,

$$
\begin{aligned}
L_z\psi &= \lambda\psi, \\
-i\frac{\partial}{\partial\psi} &= \lambda\psi.
\end{aligned}
$$

The solution is

$$
\psi = f(\theta)e^{-\lambda\theta}.
$$

For the function to be periodic on a sphere, we must have $\lambda = m \in \mathbf{Z}$. The normalized eigenfunctions of the ϕ component are,

$$
\Phi(\phi) = \frac{1}{\sqrt{2\pi}}e^{im\phi}, \quad m = 0, \pm1, \pm2, \ldots.
$$

Since $\mathbf{L}^2 - L_z^2 = L_x^2 + L_y^2$ is physically a positive operator, the absolute value of the eigenvalues of L_z for a given $\mathbf{L}^2$ are bounded. Let l be the integer corresponding to the largest value of L_z for a particular value of $\mathbf{L}^2$. We switch to the bracket notation and denote the eigenstates by

$$\psi = |l, m\rangle.$$

Let ψ_m be an eigenstate of L_z and recall from equation 8.48 that $[L_z, L_\pm] = L_\pm$. Apply L_z to the state $L_\pm\psi$. We get,

$$\begin{aligned} L_z(L_\pm|l,m\rangle) &= (L_\pm L_z + [L_z, L_\pm])|l,m\rangle, \\ &= (L_\pm L_z + L_\pm)|l,m\rangle, \\ &= (m \pm 1)(L_\pm|l,m\rangle). \end{aligned}$$

Thus, if ψ_m is an eigenfunction with eigengvalue m, $L_\pm\psi_m$ are eigenfunctions with eigenvalues $m \pm 1$. The ladder $L_\pm$ operators lower or raise the m quantum number without changing the eigenvalue of $\mathbf{L}^2$. Hence, if $m = l$ is the maximum value for a particular state of $\mathbf{L}^2$, we have.

$$L_+\psi_l = 0$$

Now, apply L_- to this state. Using formula 8.49 yields,

$$L_+L_-\psi_l = (\mathbf{L}^2 - L_z^2 - L_z)\psi_l = 0.$$

But we are seeking states which are simultaneous eigenfunctions of the commuting operators $\mathbf{L}^2$ and L_z, so the eigenvalue of $\mathbf{L}^2$ must be $l(l+1)$. For those who have studied the solution Laplace's equations by separation of variables and infinite series, this would correspond to the step where one sets the eigenvalue of Legendre's differential equation to $l(l+1)$ to cause the infinite series to terminate, and thus yield polynomial solutions. In this manner, we have recovered the eigenvalues in 8.54 almost entirely algebraically,

$$\begin{aligned} L_z|l,m\rangle &= m|l,m\rangle, \\ \mathbf{L}^2|l,m\rangle &= l(l+1)|l,m\rangle. \end{aligned} \tag{8.57}$$

Eigenstates are basis vectors for the Hilbert space, so they should be normalized. Thus, we require

$$\langle l,m|l,m\rangle = 1.$$

Suppose we have constants $C^\pm_{l,m}$ such that,

$$\psi = L_\pm|l,m\rangle = C^\pm_{l,m}|l, m \pm 1\rangle.$$

Then, $\langle\psi|\psi\rangle = |C^\pm_{l,m}|^2$. On the other hand, we have,

$$\begin{aligned} L_\pm|l,m\rangle &= (\mathbf{L}^2 - L_z^2 \mp L_z)|l,m\rangle, \\ &= (l(l+1) - m(m \pm 1))|l,m\rangle, \end{aligned}$$

so we choose,

$$C^{\pm}_{l,m} = \sqrt{l(l+1) - m(m \pm 1)}.$$

We conclude that the effect of applying the ladder operators to normalized eigenstates is

$$L_{\pm}|l,m\rangle = \sqrt{l(l+1) - m(m \pm 1)}|l,m\rangle. \tag{8.58}$$

Recalling that L_+ is a linear first order differential operator, and noting that $L_+|l,l\rangle = 0$, we get a linear first order differential equation for $\Theta(\theta)$ that is very easy to solve. Then, carefully banging the solution with lowering operators leads to Rodrigues' formula for the associated Legendre polynomials. The matrix elements of the representation are complicated. They are described by unitary $(2l+1)$-dimensional unitary matrices called *Wigner D-matrices.* If $R(\psi,\theta,\phi)$ is a rotation by Euler angles, and $|l,m\rangle$ are spherical harmonic eigenstates, the matrix elements are given by,

$$\begin{aligned} D^l{}_{mm'}(\psi,\theta,\phi) &= \langle l,m'|R(\psi,\theta,\phi)|l,m\rangle, && (8.59)\\ &= e^{-im'\phi} d^l{}_{mm'}(\theta) e^{-im\phi}, && (8.60)\end{aligned}$$

where,

$$d^l{}_{mm'}(\theta) = \langle l,m'|e^{i\theta L_x}|l,m\rangle = D^l{}_{mm'}(0,\theta,0) \tag{8.61}$$

We content ourselves in these notes in wetting the appetite of the reader to dig into more details in any senior/first-year graduate level text in quantum mechanics.

8.2 Lorentz Group

The appropriate symmetry group in special relativity is the Lorentz group. This is the group of transformations that leaves invariant the metric $\eta = \mathrm{diag}(+---)$ in Minkowski's space $M_{1,3}$. We will denote Lorentz transformations by the notation,

$$x^{\mu\prime} = L^{\mu}{}_{\nu} x^{\nu}. \tag{8.62}$$

The metric is invariant if, $< Lx, Lx > = < x, x >$, that is,

$$\eta_{\mu\nu} L^{\mu}{}_{\alpha} L^{\nu}{}_{\beta} = \eta_{\alpha\beta}.$$

Transformations for which $|L^{\mu}{}_{\nu}| = 1$, and $L^0{}_0 > 0$, so that past and future are not interchanged constitute the proper, orthochronous Lorentz group $SO^+(1,3)$. The Lorentz transformation laws for tensors T is the same as in the Riemannian metric case as shown in equation 7.16

$$T'^{\beta_1,\dots,\beta_r}_{\alpha_1,\dots,\alpha_s} = \frac{\partial x'^{\beta_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x'^{\beta_r}}{\partial x^{\mu_r}} \frac{\partial x^{\nu_1}}{\partial x'^{\alpha_1}} \cdots \frac{\partial x^{\nu_s}}{\partial x'^{\alpha_s}} T^{\mu_1,\dots,\mu_r}_{\nu_1,\dots,\nu_s}. \tag{8.63}$$

As usual, the metric is used to raise and lower indices and thus convert between covariant and contravariant tensors. Another way to obtain a covariant tensor from a contravariant one, is by use of the permutation symbol, but we need to be a little careful. As noted in the paragraph elaborating on the Hodge star operator 2.87, the Levi-Civita symbol does not transform like a tensor, but rather, like a tensor density of weight (-1). Instead, if g is any Riemannian or pseudo-Riemannian metric such as η we define $e_{\mu\nu\kappa\lambda} = \sqrt{\det g}\, \epsilon_{\mu\nu\kappa\lambda}$, which does transform like a tensor called the Levi-Civita tensor. There are general formulas similar to 2.46 for contractions of the Levi-Civita symbol for any dimension. The pattern can be inferred from the explicit formulas for dimension four listed below,

$$\begin{aligned}
\epsilon^{\mu\nu\kappa\lambda}\epsilon_{\alpha\beta\gamma\delta} &= \delta^{\mu\nu\kappa\lambda}_{\alpha\beta\gamma\delta}, \\
\epsilon^{\mu\nu\kappa\delta}\epsilon_{\alpha\beta\gamma\delta} &= \delta^{\mu\nu\kappa}_{\alpha\beta\gamma}, \\
\epsilon^{\mu\nu\gamma\delta}\epsilon_{\alpha\beta\gamma\delta} &= 2!\delta^{\mu\nu}_{\alpha\beta} = 2(\delta^{\mu}_{\alpha}\delta^{\nu}_{\beta} - \delta^{\mu}_{\beta}\delta^{\nu}_{\alpha}), \\
\epsilon^{\mu\beta\gamma\delta}\epsilon_{\alpha\beta\gamma\delta} &= 3!\delta^{\mu}_{\alpha}.
\end{aligned} \tag{8.64}$$

An appropriate contraction of a tensor T with the Levi-Civita tensor gives another tensor called the dual tensor $\check{T}$. Thus, for example, the dual of an antisymmetric tensor $T_{\mu\nu}$ is

$$\check{T}^{\alpha\beta} = \frac{1}{2\sqrt{\det g}}\epsilon^{\alpha\beta\mu\nu}T_{\mu\nu}. \tag{8.65}$$

A tensor as above for which $T_{\mu\nu} = \check{T}_{\mu\nu}$ is called self-dual. Such self-dual tensors play a special role in the representation of the Lorentz group.

8.2.1 Infinitesimal Transformations

There are 6 infinitesimal generators for the Lie algebra $\mathfrak{so}(1,3)$. We can take these generators to be,

$$j_1 = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{array}\right],\; j_2 = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{array}\right],\; j_3 = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

$$k_1 = \left[\begin{array}{c|ccc} 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],\; k_2 = \left[\begin{array}{c|ccc} 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],\; k_3 = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right].$$

The three generators $\{j_1, j_2, j_3\}$ correspond to the subgroup $SO(3)$ of spatial rotations and thus span the subalgebra $\mathfrak{so}(3)$. The exponential map of these generators yield matrices in which the spatial 3×3 blocks are the same rotation matrices as in 8.1.2. There are three other generators which we call $\{k_1, k_2, k_3\}$ involving the time parameter These generate the boosts. The k generators are not manifestly antisymmetric, but that is because the signature of the metric. The exponential map of the boost generators yield hyperbolic blocks. For example, the k_1 infinitesimal transformation in the t and x coordinates,

$$\begin{bmatrix} t \\ x \end{bmatrix}' = \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \theta \\ \theta & 0 \end{bmatrix}\right]\begin{bmatrix} t \\ x \end{bmatrix}$$

yields a Lorentz transformation of the form,

$$\begin{bmatrix} \cosh\theta & \sinh\theta & 0 & 0 \\ \sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The transformation does represent a rotation, but it is imaginary. Here

$$\cosh\theta = \frac{1}{\sqrt{1-\beta^2}}, \quad \sinh\theta = -\frac{\beta}{\sqrt{1-\beta^2}},$$

with $\beta = v/c$. This is the way these transformations appear in a first course in special relativity. Any infinitesimal Lorentz transformation $x^{\mu\prime} = L^{\mu}{}_{\nu}x^{\nu}$ has the form,

$$\begin{aligned} L^{\mu}{}_{\nu} &= \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu}, \qquad \text{or}, \\ L_{\mu\nu} &= \delta_{\mu\nu} + \omega_{\mu\nu}. \end{aligned}$$

To preserve the metric, we must have,

$$\begin{aligned} \eta_{\mu\nu}L^{\mu}{}_{\alpha}L^{\nu}{}_{\beta} &= \eta_{\mu\nu}(\delta^{\mu}{}_{\alpha} + \omega^{\mu}{}_{\alpha})(\delta^{\mu}{}_{\beta} + \omega^{\mu}{}_{\beta}), \\ &= \eta_{\alpha\beta} + \omega_{\alpha\beta} + \omega_{\beta\alpha}, \\ &= \eta_{\alpha\beta}. \end{aligned}$$

This means that $\omega_{\alpha\beta}$ is antisymmetric, as expected from the analysis of the similar situation with the exponential map of the rotation group. In any representation of the Lorentz group, the infinitesimal transformations take the form,

$$I + \tfrac{1}{2}\omega^{\mu\nu}M_{\mu\nu}, \tag{8.66}$$

where $M_{\mu\nu}$ are the antisymmetric matrices representing the six infinitesimal transformations. These matrices obey the integrability commutation relations,

$$\begin{aligned} [M_{\mu\nu}, M_{\sigma\tau}] &= M_{\mu\nu}M_{\sigma\tau} - M_{\sigma\tau}M_{\mu\nu}, \\ &= g_{\nu\sigma}M_{\mu\tau} - g_{\mu\sigma}M_{\nu\tau} + g_{\mu\tau}M_{\nu\sigma} - g_{\nu\tau}M_{\mu\sigma}. \end{aligned} \tag{8.67}$$

Apart from a factor of $i\hbar$, this is the relativistic *angular momentum tensor*, symbolically written, $\mathbf{r}\wedge\mathbf{p}$, where $\mathbf{r}$ and $\mathbf{p}$ are the position, and the 4-momentum respectively. The notation means that

$$M^{\mu\nu} = x^{\mu}p^{\nu} - p^{\mu}x^{\nu}$$

The spatial components $J_k = \epsilon_k{}^{ij}M_{ij}$ are the generators of the subgroup $SO(3)$, and the spatial-temporal components $K_i = M_{0i}$ generate the boosts. The Lie algebra commutator relations are given by

$$\begin{aligned} [J_i, J_j] &= i\epsilon^{k}{}_{ij}J_k, \\ [K_i, K_j] &= -i\epsilon^{k}{}_{ij}J_k, \\ [J_i, K_j] &= i\epsilon^{k}{}_{ij}J_k, \end{aligned} \tag{8.68}$$

where,

$$\mathbf{J} = \tfrac{1}{2}\boldsymbol{\sigma}.$$

For a spin $\frac{1}{2}$, the boosts are generated by,

$$\mathbf{K} = \pm\tfrac{\mathbf{i}}{\mathbf{2}}\boldsymbol{\sigma},$$

giving two inequivalent representations.

These matrices constitute a representation of the Lie algebra of the Lorentz group called the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ spin representation. The group elements are given by the exponential map,

$$L^{\mu}{}_{\nu} = \exp\left[\tfrac{i}{2}\omega^{\alpha\beta}(M_{\alpha\beta})^{\mu}{}_{\nu}\right]. \tag{8.69}$$

8.2.2 Spinors

In a manner analogous to the construction of the 2-1 isomorphism between $SU(2)$ and $SO(3)$, starting with the map 8.8, we seek a representation of Lorentz transformations of a vector $x^{\mu} \in M_{1,3}$ in terms of transformations of the 2×2 Hermitian matrix $X = X^{A\dot{B}}$. Since $\det X$ is equal to the norm of a vector that we wish to preserve, the condition is equivalent to invariance under unimodular transformations

$$X' = QXQ^{\dagger}, \tag{8.70}$$

where $Q \in SL(2, \mathbf{C})$.

We introduce *spin space* to as a pair $\{S_2, \epsilon_{AB}\}$, where S_2 is a 2-dimensional complex vector space and ϵ_{AB} is the symplectic form with components,

$$\epsilon_{AB} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{8.71}$$

The matrix elements of the symplectic form are the same as the Levi-Civita symbol in dimension 2. It is assumed that spinors obey the transformation law,

$$\phi'_A = \phi_B Q^B{}_A, \tag{8.72}$$

where $Q \in SL(2, \mathbf{C})$. An element $\phi_A \in S_2$ is called a *covariant 2-spinor* of rank 1. Associated with S_2 there are three other spaces, the dual S_2^*, the complex conjugate $\overline{S}_2$, and the complex conjugate dual $\overline{S}_2{}^*$. We will use the following index convention,

$$\begin{aligned} \phi_A &\in S_2\,, & \phi^A &\in S_2^*\,, \\ \phi_{\dot{A}} &\in \overline{S}_2\,, & \phi^{\dot{a}} &\in \overline{S}_2{}^*. \end{aligned}$$

We introduce dual and conjugate versions of the symplectic form $\epsilon_{A\dot{B}}$, ϵ^{AB} etc,. all of which have the same matrix values. We use these to manipulate spinor indices according to the rules,

$$\phi_A = \epsilon_{AB}\phi^B, \qquad \phi^B = \phi_A\epsilon^{AB}, \tag{8.73}$$

$$\phi_A = \epsilon_{A\dot{B}}\phi^{\dot{B}}, \qquad \phi^{\dot{B}} = \phi_A\epsilon^{A\dot{B}} \tag{8.74}$$

namely, we lower on the left and raise on the right. The two operations are inverse of each other as we can easily verify by lowering , then raising and index,

$$\begin{aligned}
\phi^A &= (\epsilon_{BC}\phi^C)\epsilon^{BA}, \\
&= \epsilon_{BC}\epsilon^{BA}\phi^C, \\
&= \delta^A{}_C\phi^C = \phi^A.
\end{aligned}$$

Here we have used the permutation symbol identity which is the same as for the Levi-Civita symbol,

$$\epsilon_{BC}\epsilon^{BA} = \delta^A{}_C.$$

Higher rank spinors can be constructed as elements of tensor products of spin spaces. Because of the antisymmetry of the symplectic form, one needs to be more careful when raising and lowering spinor indices. For instance, Consider, $\phi^A\psi_A \in S_2 \otimes S_2^*$. Then,

$$\begin{aligned}
\phi^A\psi_A &= \phi^A\epsilon_{AB}\psi^B, \\
&= \epsilon_{AB}\phi^A\psi^B, \\
&= -\epsilon_{BA}\phi^A\psi^B, \\
&= -\phi_B\psi^B, \\
&= -\phi_A\psi^A.
\end{aligned}$$

We conclude that exchanging the position of a repeated spinor index introduces a minus sign. In particular, for any rank one spinor ϕ^A, we have

$$\phi^A\phi_A = \phi_A\phi^A = 0.$$

It follows that the full contraction such as $\phi^{ABC}\phi_{ABC}$ of a spinor with odd number of indices with itself is zero. If a spinor is symmetric on any two indices, then contracting on those two indices gives 0. Contraction on two indices reduces the rank of a spinor by two. If a spinor is completely symmetric, it is not possible to reduce the rank by contraction since any such contraction is zero. The only completely antisymmetric spinor must be of rank two since the indices can only attain the values 1 or 2 and there are only to permutations possible. In fact, a completely antisymmetric spinor must be a multiple of ϵ_{AB}. Thus, for example, for any spinor ϕ_A, we have

$$\epsilon_{AB}\phi_C + \epsilon_{CA}\phi_B + \epsilon_{BC}\phi_A = 0$$

because this combination of spinor is antisymmetric and of rank 3. Another good example is the relation,

$$\phi_{AC}\phi^C{}_B = -\phi_A{}^C\phi_{CB}$$

Which is true since the position index C was exchanged. Hence the quantity on the left is an antisymmetric spinor of rank two and it must be a multiple of ϵ_{AB}. It is easy to check that the correct multiplicative factor is given by,

$$\phi_{AC}\phi^C{}_B = -\tfrac{1}{2}\phi_{CD}\phi^{CD}\epsilon_{AB}.$$

We can also have spin-tensors, the main example being the "connecting spinor" $\sigma_\mu^{A\dot{B}}$ in 8.10, which has one covariant tensor index and two spinor indices. For convenience, we list the components here again,

$$\sigma_\mu{}^{A\dot{B}} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

It is elementary to verify that the inverse matrices $\sigma^\mu{}_{A\dot{B}}$, that is, the matrices such that,

$$\sigma^\mu{}_{A\dot{B}}\sigma_\nu{}^{A\dot{B}} = \delta^\mu_\nu,$$

are given by,

$$\sigma^\mu{}_{A\dot{B}} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

The result is consistent with rasing the tensor index with the metric and lowering the spinor indices with the symplectic form. With these matrices, the reverse of equation 8.9 is,

$$x^\mu = \frac{1}{2}\sigma^\mu{}_{A\dot{B}}x^{A\dot{B}}. \tag{8.75}$$

A few words about index conventions are in order. The convention used here most closely resembles that of the early developers Veblen and Taub (See for example [36]). The main difference here is in choosing $\sigma_\mu{}^{A\dot{B}}$ to be consistent with Pauli matrices; this is closer to the choice in the Penrose prime notation. In following established protocol, I have reluctantly adapted the notation with both indices up, which is inconsistent with the index summation convention for matrix multiplication of index-free expressions such as $\sigma_1\sigma_2$. My preference would have been to choose $\sigma_\mu{}^A{}_{\dot{B}}$ to correspond to the Pauli matrices. Instead, lowering the second index results on the following matrices,

$$\sigma_\mu{}^A{}_{\dot{B}} = \left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \right\}$$

It is straight-forward to verify the following spinor identities,

$$\begin{aligned} \sigma^{\mu A\dot{B}}\sigma_{\mu C\dot{D}} &= 2\delta^A_C\delta^{\dot{B}}_{\dot{D}}, \\ \sigma_\mu{}^{A\dot{B}}\sigma^{\mu C\dot{D}} &= 2\epsilon^{AC}\epsilon^{\dot{B}\dot{D}}, \\ \sigma_\mu{}^{A\dot{B}}\sigma^\nu{}_{C\dot{B}} + \sigma_\nu{}^{A\dot{B}}\sigma_{\mu C\dot{B}} &= 2g_{\mu\nu}\delta^A_C. \end{aligned} \tag{8.76}$$

Remembering that σ matrices are Hermitian, and switching the position of summation index B, the last equation above can be rewritten as,

$$\sigma_\mu{}^A{}_{\dot{B}}\overline{\sigma}_\nu{}^B{}_{\dot{C}} + \sigma_\nu{}^A{}_{\dot{B}}\overline{\sigma}_\mu{}^B{}_{\dot{C}} = -2g_{\mu\nu}\delta^A_C.$$

This equation is in the right summation index format for matrix multiplication, so it can be written in an index-free form as,

$$\sigma^\mu\overline{\sigma}^\nu + \sigma^\mu\overline{\sigma}^\mu = -2g^{\mu\nu}I. \tag{8.77}$$

In view of the remarks above, the reader should be cautioned that the matrices in this neat equation are not the standard Pauli matrices in the coordinate representation we have chosen. Equation 8.77 is most important in formulating Dirac's equation, as shown below.
Analogous to the situation for $SO(3)$, For every proper Lorentz transformation L^μ_ν, there are two unimodular matrices Q and $-Q$ such that

$$\sigma^\mu = Q\sigma^\nu Q^{-1} L^\mu_\nu. \tag{8.78}$$

A similar statement holds for improper Lorentz transformations. In this case, we have

$$\sigma^\mu = Q\bar{\sigma}^\nu \bar{Q}^{-1} L^\mu_\nu. \tag{8.79}$$

As discovered by Dirac, to obtain a relativistic extension of Schrödinger's equation in the spin 1/2 representation, which is invariant under Lorentz transformations, one must introduce a second spinor field ψ^A. The field equations are,

$$\begin{aligned} \sigma^{\mu A}{}_{\dot{B}} \left(\frac{\hbar}{i}\frac{\partial}{\partial x^\mu}\right) \bar{\psi}^B &= -imc\phi^A, \\ \sigma^{\mu A}{}_{\dot{B}} \left(\frac{\hbar}{i}\frac{\partial}{\partial x^\mu}\right) \bar{\phi}^B &= -imc\psi^A, \end{aligned} \tag{8.80}$$

where the σ spin-tensors satisfy equation 8.77. The more familiar 4-spinor Dirac equation is obtained by rewriting 8.80 in matrix form,

$$\frac{\hbar}{i}\frac{\partial}{\partial x^\mu} \begin{bmatrix} 0 & -i\sigma^{\mu A}{}_{\dot{B}} \\ i\overline{\sigma}^{\mu A}{}_{\dot{B}} & 0 \end{bmatrix} \begin{bmatrix} \phi^B \\ \bar{\psi}^B \end{bmatrix} = mc \begin{bmatrix} \phi^A \\ \bar{\psi}^A \end{bmatrix},$$

which is the standard *Dirac equation* for a free particle,

$$(\gamma^\mu p_\mu - mc)\Psi = 0. \tag{8.81}$$

Here,

$$\Psi = \begin{bmatrix} \phi^A \\ \bar{\psi}^A \end{bmatrix} \in S_2 \times S_2^*$$

is a Dirac 4-spinor, and γ has the 4×4 matrix representation,

$$\gamma = \begin{bmatrix} 0 & -i\sigma^{\mu A}{}_{\dot{B}} \\ i\overline{\sigma}^{\mu A}{}_{\dot{B}} & 0 \end{bmatrix}.$$

Comparing with equation 8.77, we see that the γ's satisfy the so-called *Clifford algebra* relation,

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu} I. \tag{8.82}$$

We would like to establish some interesting connection between some spinor and tensorial quantities. Consider for instance a null vector $l_\mu \in M_{1,3}$. Since the length of the vector is zero, the corresponding matrix $l_{A\dot{B}}$ has determinant equal to zero; so the first row, is a multiple of the second and hence the matrix is of the form,

$$l^{\mu} = \sigma_{\mu}^{A\dot{B}} \phi_A \bar{\psi}_{\dot{B}}.$$

If we choose,

$$\phi_A = \begin{bmatrix} \zeta \\ 1 \end{bmatrix}.$$

then, up to a constant, the components of the null vector are given by,

$$l_{\mu} = \sigma_{\mu}{}^{A\dot{B}} \phi_A \bar{\phi}_{\dot{B}}.$$

Using the matrix components of the sigma matrices as in 8.10, a short computation gives,

$$l_{\mu} = \tfrac{1}{2}(\zeta\bar{\zeta} + 1, \zeta + \bar{\zeta}, -i(\zeta + \bar{\zeta}), \zeta\bar{\zeta} - 1).$$

Normalizing, the vector becomes,

$$l_{\mu} = \left(1, \frac{\zeta + \bar{\zeta}}{\zeta\bar{\zeta} + 1}, \frac{(\zeta + \bar{\zeta})}{i(\zeta\bar{\zeta} + 1)}, \frac{\zeta\bar{\zeta} - 1}{\zeta\bar{\zeta} + 1}\right). \tag{8.83}$$

The spatial part precisely the inverse image of complex numbers ζ under the stereographic projection 5.61, viewed as inhomogeneous coordinates $\zeta = \phi^1/\phi^2$ on the Riemann sphere $S^2 \cong \mathbf{CP}^1$. The norm of the spatial part is one, so the norm in $M_{1,3}$ is zero, as it should be. One may view the sphere as the intersection of the null cone with the hyperplane $t = 1$ (or -1), so this is essentially the celestial sphere. Spinors transform by elements of $SL(2, \mathbf{C})$ which is the universal covering group of the group of Möbius transformations. Möbius transformations are conformal maps, so this gives a connection with minimal surfaces.

Next, we note that self-dual tensors of rank 2 are associated with symmetric spinors. The connection is made by first defining the spinor,

$$\begin{aligned} \sigma_{\mu\nu}{}^{A}{}_{B} &= \sigma_{[\mu}{}^{A\dot{C}} \sigma_{\nu]B\dot{C}}\,, \\ &= \frac{1}{2}(\sigma_{\mu}{}^{A\dot{C}} \sigma_{\nu B\dot{C}} - \sigma_{\nu}{}^{A\dot{C}} \sigma_{\mu B\dot{C}}), \end{aligned} \tag{8.84}$$

where the contraction $\sigma_{\mu\nu}{}^{A}{}_{A} = 0$ is clearly 0. By direct computation in the coordinate system we have chosen. The values of the traceless matrices are,

$$\sigma_{01}{}^{A}{}_{B} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad \sigma_{02}{}^{A}{}_{B} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_{03}{}^{A}{}_{B} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$\sigma_{12}{}^{A}{}_{B} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \quad \sigma_{23}{}^{A}{}_{B} = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}, \quad \sigma_{13}{}^{A}{}_{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

It follows that $\sigma_{\mu\nu AB}$ is symmetric. The values are,

$$\sigma_{01AB} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_{02AB} = \begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix}, \quad \sigma_{03AB} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
$$\sigma_{12AB} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad \sigma_{23AB} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \quad \sigma_{13AB} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We find from the last set of equations that,

$$\begin{aligned}\sigma_{12} &= \sigma^{12} = -i\sigma_{03},\\ \sigma_{13} &= \sigma^{13} = i\sigma_{02},\\ \sigma_{01} &= -\sigma^{01} = i\sigma_{23},\end{aligned}$$

and hence

$$\check{\sigma}^{\mu\nu} = \frac{1}{2\sqrt{\det g}}\epsilon^{\mu\nu\sigma\tau}\sigma_{\sigma\tau}. \tag{8.85}$$

is self-dual. From this, it follows that if F^{AB} is a symmetric spinor, then

$$F_{\mu\nu} = \sigma_{\mu\nu AB}F^{AB} \tag{8.86}$$

is a self-dual tensor. It can also be verified by computation in the chosen coordinate system that.

$$\sigma_{\mu\nu AB}\sigma^{\sigma\tau AB} = -2\delta^{\sigma}_{\mu}\delta^{\tau}_{\nu.} \tag{8.87}$$

From this, the equation above can be inverted,

$$F^{CD} = \tfrac{1}{8}F_{\sigma\tau}\sigma^{\sigma\tau CD} \tag{8.88}$$

yielding a symmetric spinor from a self-dual tensor. This establishes a 1-1 correspondence between self-dual tensors of rank two, and symmetric spinors of rank two. Tensors that transform like symmetric spinors of rank $2n$ are said to be irreducible under a spin n representation. This is particularly relevant for self-dual Maxwell tensors. More specifically, if,

$$F = F_{\mu\nu}dx^{\mu} \wedge dx^{\nu}$$

is the Maxwell 2-form, the corresponding Maxwell spinor is,

$$F_{A\dot{A}B\dot{B}} = \sigma^{A\dot{A}}_{\mu}\sigma^{B\dot{B}}_{\nu}F_{\mu\nu}$$

Here, (almost) following Penrose, we have used undotted and dotted letters with the same character, to avoid the proliferation of different letters. The Maxwell spinor can be written as,

$$\begin{aligned}F_{A\dot{A}B\dot{B}} &= \frac{1}{2}(F_{A\dot{A}B\dot{B}} - F_{B\dot{B}A\dot{A}}),\\ &= \phi_{AB}\bar{\epsilon}_{\dot{A}\dot{B}} + \bar{\phi}_{\dot{A}\dot{B}}\epsilon_{AB},\end{aligned}$$

where, ϕ_{AB} and $\bar{\phi}_{\dot{A}\dot{B}}$ are symmetric spinors given by,

$$\begin{aligned}\phi_{AB} &= \frac{1}{2}F_{A\dot{C}B}{}^{\dot{C}},\\ \bar{\phi}_{\dot{A}\dot{B}} &= \bar{\phi}_{AB}\end{aligned}$$

This gives a spinor decomposition of the tensor into self-dual and anti-self-dual parts.

The matrices ${\sigma_{\mu\nu}}^A{}_B$ also serve to connect infinitesimal Lorentz and spinor transformations. Given an infinitesimal Lorentz transformation as in equation 8.66, the corresponding infinitesimal spinor transformation is.

$$(I + \tfrac{1}{2}\omega^{\mu\nu} M_{\mu\nu})^A{}_B = \delta^A{}_B + \tfrac{1}{4}\omega^{\mu\nu}(\sigma_{\mu\nu})^A{}_B.$$

In other words, the spinor representation of the angular momentum tensor is,

$$(M_{\mu\nu})^A{}_B = \frac{1}{2}{\sigma_{\mu\nu}}^A{}_B. \tag{8.89}$$

Index manipulation of Dirac 4-spinors Ψ^μ is done by an extension $\epsilon_{\mu\nu}$ of the symplectic form to a 4×4 matrix which is also antisymmetric. In an accepted abuse of language, some authors refer to $\epsilon_{\mu\nu}$ as the metric spinor. As with 2-spinors, we lower on the left and raise on the right,

$$\begin{aligned}\Psi_\mu &= \epsilon_{\mu\nu}\Psi^\nu, \\ \Psi^\mu &= \Psi^\nu \epsilon_{\nu\mu}\end{aligned}$$

In the coordinate basis we have chosen, $\epsilon_{\mu\nu}$ has components

$$\epsilon_{\mu\nu} = \begin{bmatrix} i\epsilon^{AB} & 0 \\ 0 & -i\bar{\epsilon}_{AB} \end{bmatrix} \tag{8.90}$$

The spinor-index version of the Clifford algebra relation 8.82 is

$$\gamma^{\mu\alpha}{}_\beta \gamma^{\nu\beta}{}_\gamma + \gamma^{\nu\alpha}{}_\beta \gamma^{\mu\beta}{}_\gamma = 2g^{\mu\nu}\delta^\alpha{}_\gamma. \tag{8.91}$$

In a completely analogous manner equation 8.84, we construct self-dual 4-spinor ${\gamma_{\mu\nu}}^\alpha{}_\beta$ which we write in matrix form as,

$$\gamma_{\mu\nu} = \gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu = [\gamma_\mu, \gamma_\nu] \tag{8.92}$$

Then, the infinitesimal momentum tensor 8.89 in the 4-spinor representation becomes,

$$(M_{\mu\nu})^\alpha{}_\beta = \tfrac{1}{2}(\gamma_{\mu\nu})^\alpha{}_\beta. \tag{8.93}$$

The matrices $\{I, \gamma^\mu, \gamma^{\mu\nu}\}$ and their duals, span the Clifford algebra. The dual of I is the celebrated matrix

$$\gamma^5 = \frac{i}{4!}\epsilon_{\mu\nu\sigma\tau}\gamma^\mu\gamma^\nu\gamma^\sigma\gamma^\tau. \tag{8.94}$$

The commutation relations 8.67 insure that

$$[\gamma^{\mu\nu}, \gamma^\alpha] = \eta^{\mu\alpha}\gamma^\nu - \eta^{\nu\alpha}\gamma^\mu. \tag{8.95}$$

The dual of the six $\gamma^{\mu\nu}$ matrices do not yield independent matrices, and the dual of γ^μ is essentially $\gamma^\mu\gamma^5$, so the algebra is spanned by $\{I, \gamma^\mu, \gamma^{\mu\nu}, \gamma^\mu\gamma^5, \gamma^5\}$, and it has $1 + 4 + 6 + 4 + 1 = 16$ dimensions,. We have defined the gamma matrices in a particular matrix representation, but the general relations that describe the algebra are basis-independent.

8.3 N-P Formalism

Spinors provide an interesting formalism introduced in 1962 by Newman and Penrose [24] for the study of general relativity. Let $\{M, g_{\mu\nu}\}$ be a space-time with Lorentzian metric of signature $(+ - --)$. Introduce a null tetrad $h^a{}_\mu = \{n_\mu, l_\mu, -m_\mu, -\bar{m}_\mu\}$, and dual frame forms,

$$\theta^a = h^a{}_\mu dx^\mu, \tag{8.96}$$

associated with the frame,

$$e_a = h^\mu{}_a \partial_\mu, \qquad h^\mu{}_a = \{l^\mu, n^\mu, m^\mu, \bar{m}^\mu\}. \tag{8.97}$$

In terms of the tetrad, the space-time metric is given by,

$$ds^2 = \eta_{ab}\theta^a\theta^b, \tag{8.98}$$

where η is the quasi-orthonormal flat metric,

$$\eta_{ab} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \tag{8.99}$$

used to manipulate tetrad indices. Here, the metric tensor has the form

$$g_{\mu\nu} = 2l_{(\mu}n_{\nu)} - 2m_{(\mu}\bar{m}_{\nu)}. \tag{8.100}$$

The Cartan structure equations are,

$$\begin{aligned} d\,\theta^a + \omega^a{}_b \wedge \theta^b &= 0, & \omega_{ab} &= -\omega_{ba}, \\ d\,\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b &= \Omega^a{}_b, & \Omega_{ab} &= -\Omega_{ba}, \\ d\,\Omega^a{}_b - \Omega^a{}_c \wedge \omega^c{}_b + \omega^a{}_c \wedge \Omega^c{}_b &= 0. \end{aligned} \tag{8.101}$$

In this formalism the connection components are called the *Ricci rotation coefficients*, which are defined by,

$$\gamma^a{}_{bc} = h^\nu{}_{b;\mu} h^\mu{}_c h^a{}_\nu. \tag{8.102}$$

The Riemann tetrad components satisfy,

$$\begin{aligned} \omega^a{}_b &= \gamma^a{}_{bc}\theta^c, \\ \Omega^a{}_b &= \tfrac{1}{2} R^a{}_{bcd}\theta^c \wedge \theta^d. \end{aligned} \tag{8.103}$$

Einstein's equations in tetrad form are,

$$R_{ab} - \tfrac{1}{2}\eta_{ab}R = T_{ab}. \tag{8.104}$$

As is well known in the literature, the Riemann tensor admits the decomposition,

$$R_{abcd} = C_{abcd} + \tfrac{R}{12} g_{abcd} + E_{abcd}, \tag{8.105}$$

where, C_{abcd} is the conformal Weyl tensor, and,

$$\begin{aligned} g_{abcd} &= \eta_{ac}\eta_{bd} - \eta_{ad}\eta_{bc}, \\ E_{abcd} &= g_{abf(d}S^f{}_{c)}, \\ S_{ab} &= R_{ab} - \tfrac{1}{4}\eta_{ab}R. \end{aligned} \tag{8.106}$$

The transition between tensor and spinor quantities is made by contraction with the connecting spinors $\tau^\mu{}_{A\dot{B}}$ which satisfy,

$$g_{\mu\nu}\tau^\mu{}_{A\dot{C}}\tau^\nu{}_{B\dot{C}} = \epsilon_{AB}\bar{\epsilon}_{\dot{C}\dot{D}}.$$

The spinor dual one-forms $\theta^{A\dot{B}}$ are defined by,

$$\theta^{A\dot{B}} = \tau_\mu{}^{A\dot{B}}\,dx^\mu = \theta^a\tau_a{}^{A\dot{B}},$$

where the tetrad connecting spinors are chosen such that,

$$\begin{aligned} ds^2 &= \det\theta^{A\dot{B}}, \\ &= \begin{bmatrix} \theta^0 & \theta^2 \\ \theta^3 & \theta^1 \end{bmatrix}. \end{aligned}$$

We now introduce a spin dyad $\zeta^A{}_a = (\phi^A, \psi^A)$, with $\phi^A\psi_A = 1$. The null tetrad can be written as,

$$\begin{aligned} l_\mu &= \tau_\mu{}^{A\dot{B}}\phi_A\bar{\phi}_{\dot{B}}, & m_\mu &= \tau_\mu{}^{A\dot{B}}\phi_A\bar{\psi}_B, \\ n_\mu &= \tau_\mu{}^{A\dot{B}}\psi_A\bar{\psi}_{\dot{B}}, & \bar{m}_\mu &= \tau_\mu{}^{A\dot{B}}\psi_A\bar{\phi}_{\dot{B}}, \end{aligned} \tag{8.107}$$

The spin coefficients corresponding the 24 Ricci rotation coefficients are given by,

$$\Gamma_{ab\mu} = \zeta_{Aa;\mu}\zeta^A{}_b. \tag{8.108}$$

The 12 complex spin coefficients may be arranged in three groups of four, according to the scheme,

$$\begin{aligned} A_\mu &= \Gamma_{00\mu} = \{\kappa, \rho, \sigma, \tau\}, \\ B_\mu &= \Gamma_{01\mu} = \{\epsilon, \alpha, \beta, \gamma\} = \Gamma_{10\mu}, \\ C_\mu &= \Gamma_{11\mu} = \{\pi, \lambda, \mu, \nu\}, \end{aligned} \tag{8.109}$$

The *spin connection* $\Gamma_{AB\mu}$ which gives rise to the covariant derivative of spinors, is related to the spin coefficients by,

$$\Gamma_{ab\mu} = \Gamma_{AB\mu}\zeta^A{}_a\zeta^B{}_b.$$

Following equation 8.84, we define,

$$\sigma_{ab}{}^A{}_B = \sigma_{[a}{}^{A\dot{C}}\sigma_{b]B\dot{C}}, \tag{8.110}$$

which we use to construct the connection and curvature spinors,

$$\begin{aligned} \Gamma^A{}_B &= \omega^{ab}\sigma_{ab}{}^A B, \\ \Omega^A{}_B &= \Omega^{ab}\sigma_{ab}{}^A B. \end{aligned} \tag{8.111}$$

We can now write the spinor version of the equations of structure, which not surprisingly, have a very similar appearance,

$$\begin{aligned} d\,\theta^{A\dot{B}} + \Gamma^A{}_C \wedge \theta^{C\dot{B}} + \overline{\Gamma}^{\dot{B}}{}_{\dot{C}} \wedge \theta^{A\dot{C}} &= 0, \\ d\,\Gamma^A{}_B + \Gamma^A{}_C \wedge \Gamma^C{}_B &= \Omega_{AB}, \\ d\Omega^A{}_B - \Omega^A{}_C \wedge \Gamma^C{}_B + \Gamma^A{}_C \wedge \Omega^C{}_B &= 0. \end{aligned} \tag{8.112}$$

As already discussed, if the coframes θ^a undergo a Lorentz transformation $\hat{\theta}^a = L^a{}_b\theta^b$, the spinor coframe undergoes a similarity transformation $Q \times \bar{Q} \in SL(2, \mathbf{C}) \times \overline{SL}(2\mathbf{C})$,

$$\hat{\theta}^{A\dot{B}} = Q^A{}_C\,\theta^{C\dot{D}}\bar{Q}^{\dot{B}}{}_{\dot{D}}\,.$$

In matrix notation, the connection and curvature spinors transform according to,

$$\begin{aligned} \hat{\Gamma} &= Q\Gamma Q^{-1} + QdQ^{-1}, \\ \hat{\Omega} &= Q\Omega Q^{-1}. \end{aligned} \tag{8.113}$$

The spin connection $\Gamma^A{}_B$ and the curvature spinor $R^A{}_{Bcd}$ satisfy equations analogous to 8.103

$$\begin{aligned} \Gamma^A{}_B &= \Gamma^A{}_{Bc}\theta^c = \Gamma^A{}_{CB\dot{D}}\theta^{C\dot{D}}, \\ \Omega^A{}_B &= \tfrac{1}{2}R^A{}_{Bcd}\,\theta^c \wedge \theta^d, \\ &= \tfrac{1}{2}R^{A\dot{A}}{}_{B\dot{B}C\dot{C}D\dot{D}}\,\theta^{C\dot{C}} \wedge \theta^{D\dot{D}}. \end{aligned} \tag{8.114}$$

We also have a decomposition of curvature spinor into irreducible components

$$\begin{aligned} R_{A\dot{A}B\dot{B}C\dot{C}D\dot{D}} = {}& \Psi_{ABCD}\bar{\epsilon}_{\dot{A}\dot{B}}\bar{\epsilon}_{\dot{C}\dot{D}} + \bar{\Psi}_{\dot{A}\dot{B}\dot{C}\dot{D}}\epsilon_{AB}\epsilon_{CD} \\ & + \tfrac{1}{12}R[\epsilon_{AC}\epsilon_{BD}\bar{\epsilon}_{\dot{A}\dot{C}}\bar{\epsilon}_{\dot{B}\dot{D}} + \epsilon_{AB}\epsilon_{CD}\bar{\epsilon}_{\dot{A}\dot{D}}\bar{\epsilon}_{\dot{B}\dot{C}}] \\ & + \epsilon_{AB}\Phi_{CD\dot{A}\dot{B}}\epsilon_{\dot{C}\dot{D}} + \epsilon_{CD}\Phi_{AB\dot{C}\dot{D}}\epsilon_{\dot{A}\dot{B}}, \end{aligned} \tag{8.115}$$

where,

$$\Phi_{AB\dot{C}\dot{D}} = \Phi_{(AB)(\dot{C}\dot{D})} = \overline{\Phi}_{AB\dot{C}\dot{D}}, \tag{8.116}$$

is the traceless Ricci spinor, and,

$$\Psi_{ABCD} = \Psi_{(ABCD)} \tag{8.117}$$

is the completely symmetric Weyl conformal spinor. Finally, the spinorial version of Einstein's equation takes the form,

$$\Phi_{AB\dot{C}\dot{D}} = \tfrac{1}{4}(T_{A\dot{C}B\dot{D}} + T_{B\dot{C}A\dot{D}}), \qquad T = T^a{}_a = R. \tag{8.118}$$

Equations 8.118 and 8.101 are called the Newman-Penrose (N-P) equations. When written in detail in terms of the 12 spin coefficients, the (N-P) formalism results in a formidable set of systems of coupled first order differential equations consisting of 4 metric equations, 18 spin coefficient equations, and 8 Bianchi identities. Fortunately, the spin coefficients have geometric interpretations that motivate imposing conditions on the N-P equations that makes them tractable. The Weyl spinor leads an elegant classification of so-called, *algebraically special* space-times. The classification was originally done by Petrov, but it is now commonly known as the Cartan-Petrov-Pirani-Penrose classification. One starts by writing the completely symmetric conformal spinor as,

$$\Psi_{ABCD} = \alpha_{(A}\beta_B\gamma_C\delta_{D)}. \tag{8.119}$$

Each of the rank one spinors is associated with a null vector. The classification is as follows,

1. Type I. Algebraically general - 4 distinct null directions.
2. Type II. Two null directions coincide.
3. Type D. There are two pairs of null directions that coincide.
4. Type III. Three principal directions coincide.
5. Type N. All principal directions coincide - also called Type Null.

The 1962 seminal paper by Newman and Penrose [24] is noted by the elegant proof in terms of the spin coefficients, of the Goldberg-Sachs theorem. The theorem states that a non-flat, vacuum space-time is algebraically special, if, and only if, it contains a null geodesic congruence that is shear-free; that is, there is a null vector with $\kappa = 0$ and $\sigma = 0$.

The literature on applications of the N-P formalism is huge. A Google search on "Newman-Penrose Spin Coefficients" yields over 54,000 results. We provide here a very simple example.

8.3.1 Example Consider the Vaidya metric in Eddington-Finkelstein coordinates 6.75. A null tetrad can be adapted to this metric by choosing

$$\begin{aligned} l_\mu &= \frac{\partial}{\partial r}, \\ n_\mu &= \frac{\partial}{\partial u} - \frac{1}{2}(1 - \tfrac{2m}{r})\frac{\partial}{\partial r}, \end{aligned} \qquad \text{and} \qquad \begin{aligned} m_\mu &= \frac{1}{r\sqrt{2}}\left[\frac{\partial}{\partial \theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial \phi}\right], \\ \overline{m}_\mu &= \frac{1}{r\sqrt{2}}\left[\frac{\partial}{\partial \theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial \phi}\right]. \end{aligned}$$

Thus, we have an associated spin dyad as in equation 8.107

$$l_\mu \to \phi_A\bar{\phi}_{\dot{B}}, \quad n_\mu \to \psi_A\bar{\psi}_{\dot{B}}, \quad m_\mu \to \phi_A\bar{\psi}_{\dot{B}}.$$

The only non-zero component of the curvature spinor is

$$\Psi_2 = \Psi_{ABCD}\,\phi^A\phi^B\psi^C\psi^D = -\frac{m(u)}{r^3}, \tag{8.120}$$

which is consistent with the space being of Petrov type D. By a clever idea of allowing r to assume complex valued, and then performing a complex rotation, Newnman and Janis were able to obtain the Kerr metric [25].

8.3.1 The Kerr Metric

When Einstein first introduced the theory of general relativity in 1915, it took but two months for Schwarzschild to develop a solution to the vacuum field equations. It took another 45 years to find a Ricci-flat solution describing an axially symmetric, rotating, space-time. The solution was found by R. Kerr in 1963. A space-time is said to be in *Kerr-Schild* form, if the line element can be written as

$$g_{\mu\nu} = \eta_{\mu\nu} + H\, l_\mu l_\nu, \tag{8.121}$$

where η is the flat metric, H is a scalar function, and l_μ a null vector with respect to g and η. It is easy to show that the Schwarzschild metric can be written in Kerr-Schild form. Start with the Eddington-Finkelstein coordinates,

$$ds^2 = 2dr du + [1 - \tfrac{2m}{r}]\, du^2 - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2.$$

Since

$$(du + dr)^2 = du^2 + 2du dr + dr^2.$$

we can solve for $2du\, dr$ and substitute into the metric. We get

$$ds^2 = -\tfrac{2m}{r} du^2 - dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2 - (du + dr)^2.$$

Now, we let $x^0 = u + r$. The transformation yields

$$ds^2 = -(dx^0)^2 - dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2 - \tfrac{2m}{r}(dx^0 - dr)^2,$$

which is the desired Kerr-Schild form with the Minkowski metric written in spherical coordinates. In Boyer-Lindquist coordinates the Kerr metric is given by (See [21])

$$ds^2 = \frac{\Delta}{\rho^2}(dt - a\sin^2\theta\, d\phi)^2 - \frac{\rho^2}{\Delta} dr^2 - \rho^2\, d\theta - \frac{\sin^2\theta}{\rho^2}[(r^2 + a^2)\, d\phi - a\, dt]^2, \tag{8.122}$$

where,

$$\begin{aligned} \Delta &= r^2 - 2mr + a^2; \quad a^2 < m^2, \\ \rho^2 &= r^2 + a^2\cos^2\theta. \end{aligned} \tag{8.123}$$

Since the metric coefficients do not depend on t and ϕ, the quantities $\partial_t = \frac{\partial}{\partial\phi}$ and $\partial_\phi = \frac{\partial}{\partial\phi}$ are Killing vector fields, and we get associated conserved quantities E and L as in the case of the Schwarzschild space-time. When $a = 0$ the line element immediately reduces to the Schwarzschild metric. When $m = 0$, the cross terms with $(dt\, d\phi)$ cancel out and one gets

$$ds^2 = dt^2 - \frac{r^2 + a^2\cos^2\theta}{r^2 + a^2} dr^2 - (r^2 + a^2\cos^2)\, d\theta^2 - (r^2 + a^2)\, d\phi^2.$$

This one is not obvious, but it is the flat Minkowski metric in oblate spheroidal coordintates, that is, one for which the spatial part is the $\mathbf{R}^3$ metric based on ellipsoids

$$\frac{x^2}{r^2+a^2}+\frac{y^2}{r^2+a^2}+\frac{z^2}{r^2}=1, \tag{8.124}$$

parametrized by the transformation

$$\begin{aligned}
x &= \sqrt{r^2+a^2}\,\cos\phi\sin\theta,\\
y &= \sqrt{r^2+a^2}\,\sin\phi\sin\theta,\\
z &= r\cos\theta.
\end{aligned}$$

The metric blows up at $\rho^2=0$ and $\Delta=0$. When $\rho^2=0$ we have

$$r=0, \quad \theta=\frac{\pi}{2}.$$

This constitutes a real singularity of the curvature. On the other hand, it can be shown that the singularity $\Delta=0$ can be be removed by a change of coordinates, so this singularity is more like the apparent singularity of the Schwarzschild metric at $r=2m$. The quadratic equation $\Delta=0$ has solutions

$$r_{\pm}=m\pm\sqrt{r^2-a^2}.$$

Thus the space-time is divided into three regions,

$$\begin{aligned}
R_1: &\quad \text{where} \quad r_+<r,\\
R_2: &\quad \text{where} \quad r_-<r<r_+,\\
R_3: &\quad \text{where} \quad r<r_-,
\end{aligned}$$

We live in region R_1. The boundaries at $r_{\pm}$ represent an outer and an inner event horizon respectively. A time-like particle can cross from region R_1 to R_2 and from R_2 to R_3, but not the other way around. As such, r_+ is the real event horizon. A new feature that is not present in the Schwarzschild metric is a region called the *ergosphere* which lies outside the event horizon in inside the oblate region defined by $g_{00}=0$. Particles entering the ergosphere from region R_1 are subjected to frame-dragging by the rotation of the black hole. By inspection of the metric 8.122, we see that,

$$\begin{aligned}
g_{00} &= \frac{\Delta}{\rho^2}-\frac{a^2\sin^2\theta}{\rho^2},\\
&= \frac{r^2-2rm+a^2\cos^2\theta}{\rho^2}.
\end{aligned}$$

Thus, the outer boundary of the ergosphere is given by the root

$$r_e^+=m+\sqrt{m^2-a^2\cos^2\theta}, \tag{8.125}$$

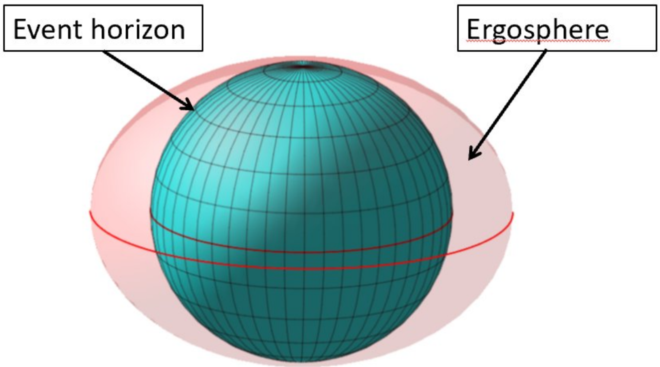


Fig. 8.4: Ergosphere

The rotational frame-dragging is caused by the change at this boundary of the Killing vector field ∂_t from being time-like to space-like. In Boyer-Lindquist coordinates it is relatively easy to compute the connection forms. We choose an orthonormal coframe

$$\begin{aligned}
\theta^0 &= \frac{\sqrt{\Delta}}{\rho}(dt - a\sin^2\theta\, d\phi),\\
\theta^1 &= \frac{\rho}{\sqrt{\Delta}}\, dr,\\
\theta^2 &= \rho\, d\theta,\\
\theta^3 &= \frac{\sin\theta}{\rho}[(r^2 + a^2)d\phi - a\, dt],
\end{aligned} \tag{8.126}$$

so that

$$ds^2 = (\theta^0)^2 - (\theta^1)^2 - (\theta^2)^2 - (\theta^3)^2.$$

The idea is the same as in previous connection computations, noting that the connection forms are antisymmetric. We take the exterior derivatives of the coframe forms and express the results in terms of the basis. We read the connection coefficients from the first Cartan structure equation, and check for consistency for possible missing terms. The presence of cross terms makes the calculation a bit more challenging and requires some finesse. We compute the quantities we need as we go along, starting with,

$$\begin{aligned}
\rho\, d\rho &= r\, dr - a^2\cos\theta\sin\theta\, d\theta,\\
d\Delta &= 2(r-m)\, dr.
\end{aligned}$$

The order of the computation is not really important, so we might as well begin with θ^1 and θ^2, that are the easiest. We have

$$\begin{aligned} d\,\theta^1 &= d\,(\tfrac{\rho}{\sqrt{\Delta}}) \wedge dr, \\ &= \tfrac{\sqrt{\Delta}\,d\rho - \rho\,d\sqrt{\Delta}}{\Delta} \wedge dr, \\ &= \tfrac{1}{\Delta}[\tfrac{\sqrt{\Delta}}{\rho}(r\,dr - a^2\cos\theta\sin\theta\,d\,\theta)] \wedge dr, \\ &= -\tfrac{1}{\rho\sqrt{\Delta}}\,a^2\cos\theta\sin\theta\,\frac{\theta^2}{\rho} \wedge \tfrac{\sqrt{\Delta}}{\rho}\,\theta^1, \\ &= \frac{a^2\cos\theta\sin\theta}{\rho^3}\,\theta^1 \wedge \theta^0. \end{aligned}$$

Continuing,

$$\begin{aligned} d\,\theta^2 &= d\rho \wedge d\theta, \\ &= \frac{1}{\rho} r\,dr \wedge d\theta, \\ &= \frac{r}{\rho}\frac{\sqrt{\Delta}}{\rho}\,\theta^1 \wedge \frac{1}{\rho}\,\theta^2, \\ &= -\frac{r\sqrt{\Delta}}{\rho^3}\,\theta^2 \wedge \theta^1. \end{aligned}$$

Comparing these differentials with the structure equations

$$d\,\theta^1 = -\omega^1{}_j \wedge \theta^j, \quad d\,\theta^2 = -\omega^2{}_j \wedge \theta^j,$$

we infer that

$$\omega^1{}_2 = -\frac{a^2\cos\theta\sin\theta}{\rho^3}\,\theta^1 - \frac{r\sqrt{\Delta}}{\rho^3}\,\theta^2.$$

However, we should not be surprised if the expressions above for $d\,\theta^1$ and $d\,\theta^2$ have other terms that either add to zero or wedge to zero. The other two structure equations are more elaborate. We have

$$\begin{aligned} d\theta^0 &= d\,(\tfrac{\sqrt{\Delta}}{\rho}) \wedge (dt - a\sin^2\theta\,d\phi) + \tfrac{\sqrt{\Delta}}{\rho}(-2a\sin\theta\cos\theta\,d\theta \wedge d\,\phi), \\ &= \tfrac{\rho\,d\sqrt{\Delta} - \sqrt{\Delta}\,d\,\rho}{\rho^2} \wedge \tfrac{\rho}{\sqrt{\Delta}}\,\theta^0 - \tfrac{2a\sqrt{\Delta}}{\rho}\sin\theta\cos\theta\,d\theta \wedge d\,\phi, \\ &= \frac{1}{\rho^2}[\rho\tfrac{(r-m)}{\sqrt{\Delta}}dr - \tfrac{\sqrt{\Delta}}{\rho}(r\,dr - a^2\cos\theta\sin\theta)] \wedge \tfrac{\rho}{\sqrt{\Delta}}\,\theta^0 - \tfrac{2a\sqrt{\Delta}}{\rho}\sin\theta\cos\theta\,d\theta \wedge d\,\phi, \\ &= \frac{1}{\rho^2}[\tfrac{\rho(r-m)}{\sqrt{\Delta}} - \tfrac{\sqrt{\Delta}r}{\rho}]\,dr \wedge \tfrac{\rho}{\sqrt{\Delta}}\,\theta^0 + \frac{1}{\rho^3}(a^2\cos\theta\sin\theta)\,\theta^2 \wedge \theta^0 - \tfrac{2a\sqrt{\Delta}}{\rho}\sin\theta\cos\theta\,d\theta \wedge d\,\phi, \\ &= [\tfrac{\rho^2(r-m) - r\Delta}{\rho^3\sqrt{\Delta}}]\,\theta^1 \wedge \theta^0 + \frac{1}{\rho^3}(a^2\cos\theta\sin\theta)\,\theta^2 \wedge \theta^0 - \tfrac{2a\sqrt{\Delta}}{\rho}\sin\theta\cos\theta\,d\theta \wedge d\,\phi. \end{aligned}$$

For the last term in the right, we will need to express $d\phi$ in terms of the coframe. This is easily done by eliminating dt and solving for $d\phi$ from the equations for

θ^0 and θ^3.

$$
\begin{aligned}
dt &= \tfrac{\rho}{\sqrt{\Delta}}\,\theta^0 + a\sin^2\theta\,d\phi,\\
\theta^3 &= \tfrac{\sin\theta}{\rho}[(r^2+a^2)\,d\phi - \tfrac{a\rho}{\sqrt{\Delta}}\,\theta^0 - a^2\sin^2\theta\,d\phi],\\
&= \tfrac{\sin\theta}{\rho}[\rho^2\,d\phi - \tfrac{a\rho}{\sqrt{\Delta}}\,\theta^0],\\
d\phi &= \tfrac{1}{\rho\sin\theta}\,\theta^2 + \tfrac{a}{\rho\sqrt{\Delta}}\,\theta^0.
\end{aligned}
$$

Substituting for $d\phi$ into the last equation for $d\theta^0$, we get after some simplification

$$
d\theta^0 = \left[\frac{\rho^2(r-m) - r\Delta}{\rho^3\sqrt{\Delta}}\right]\theta^1\wedge\theta^0 - \frac{a^2\cos\theta\sin\theta}{\rho^3}\,\theta^2\wedge\theta^0 - \frac{2a}{\rho^3}\sqrt{\Delta}\cos\theta\,(\theta^2\wedge\theta^3). \quad (8.127)
$$

We proceed to compute $d\theta^3$ in a similar manner.

$$
\begin{aligned}
d\theta^3 &= d\,(\frac{\sin\theta}{\rho})\wedge[(r^2+a^2)\,d\phi - a\,dt] + \frac{\sin\theta}{\rho}\,2r\,dr\wedge d\phi,\\
&= \frac{\rho\cos\theta\,d\theta - \sin\theta\,d\rho}{\rho^2}\wedge\frac{\rho}{\sin\theta}\,\theta^3 + \frac{2r\sin\theta}{\rho}\,dr\wedge d\phi,\\
&= \frac{1}{\rho\sin\theta}[\rho\cos\theta\,d\theta - \sin\theta\,d\rho]\wedge\theta^3 + \frac{2r\sin\theta}{\rho}\,dr\wedge d\phi.
\end{aligned}
$$

The rest of the computation is completely straight-forward. We substitute the differentials $d\theta, d\rho, dr$ and $d\phi$ in terms of the coframe, and simplify. Notice that we do no need to solve for dt. We leave it to the reader to verify the result,

$$
d\theta^3 = \frac{2ar\sin\theta}{\rho^3}\,\theta^1\wedge\theta^0 + \frac{r\sqrt{\Delta}}{\rho^3}\,\theta^1\wedge\theta^3 + \frac{\cos\theta}{\rho^3\sin\theta}(r^2+a^2)\,\theta^2\wedge\theta^3. \quad (8.128)
$$

Anticipating possible missing terms required for consistency, we split the $\theta^2\wedge\theta^3$ in the equation 8.127 for $d\theta^0$ as

$$
-\frac{2a\cos\theta}{\rho^3}\,\theta^2\wedge\theta^3 = -\frac{a\cos\theta}{\rho^3}\,\theta^2\wedge\theta^3 + \frac{a\cos\theta}{\rho^3}\,\theta^3\wedge\theta^2,
$$

and do the same for the $\theta^1 \wedge \theta^0$ in the expression 8.128. Together with 8.3.1, we can now read all the independent connection forms

$$
\begin{aligned}
\omega^0{}_1 &= [\tfrac{\rho^2(r-m)-r\Delta}{\rho^3\sqrt{\Delta}}]\,\theta^0 - \frac{ar\sin\theta}{\rho^3}\,\theta^3,\\
\omega^0{}_2 &= -\frac{(a^2\cos\theta\sin\theta)}{\rho^3}\,\theta^0 + \frac{a\cos\theta\sqrt{\Delta}}{\rho^3}\,\theta^3,\\
\omega^0{}_3 &= \frac{a\cos\theta\sqrt{\Delta}}{\rho^3}\,\theta^2 - \frac{ar\sin\theta}{\rho^3}\,\theta^1,\\
\omega^1{}_2 &= -\frac{a^2\cos\theta\sin\theta}{\rho^3}\,\theta^1 - \frac{r\sqrt{\Delta}}{\rho^3}\,\theta^2,\\
\omega^1{}_3 &= -\frac{ar\sin\theta}{\rho^3}\,\theta^0 - \frac{r\sqrt{\Delta}}{\rho^3}\,\theta^3,\\
\omega^2{}_3 &= -\frac{\cos\theta}{\rho^3\sin\theta}(r^2+a^2)\,\theta^3 - \frac{a\cos\theta\sqrt{\Delta}}{\rho^3}\,\theta^0
\end{aligned}
\tag{8.129}
$$

The only term above that could not be read immediately from the computed differentials is the second term in $\omega^2{}_3$. Here, the θ^0 term comes from a modification of the formula for $d\theta^2$ that is required for consistency with $\omega^2{}_0 = -\omega^0{}_2$. The computation of the curvature form requires no finesse; it is just a lengthy "plug and chug" calculation that presently is not a task for human beings. With the use of a computer algebra system such as Maple or Mathematica, one can verify that the curvature is Ricci-flat. For a full discussion of physical implications of the Kerr geometry, see Misner, Thorne and Wheeler [21].

8.3.2 Eth Operator

One of the most pulchritudinous results arising from the N-P formalism, was the serendipitous discovery by Newman and Penrose of a characterization of asymptotically half-flat space-times in terms of an operator called eth. The operator acts on a space consisting of spin-weighted functions on a sphere. A function η on a sphere has spin weight s if it transforms as,

$$\eta' = e^{is\psi}\eta, \tag{8.130}$$

under a rotation about the north pole. The spin-weight is constrained to be a $\frac{1}{2}$-integer. Here we are only interested in the nature of the operator in its relation to representations, but to provide some historical context, we say a few words about half-flat space times. The simplest way to introduce the notion of half-flat is to consider the *good cut* differential equation,

$$\eth^2 Z(\zeta,\tilde{\zeta})Z(\zeta,\tilde{\zeta}) = \sigma^0(Z,\zeta,\tilde{\zeta}), \tag{8.131}$$

where the *eth* operator is defined by,

$$
\begin{aligned}
\eth\eta &= 2P^{1-s}\frac{\partial}{\partial\zeta}(P^s\eta),\\
\bar{\eth}\eta &= 2P^{1+s}\frac{\partial}{\partial\bar{\zeta}}(P^{-s}\eta).
\end{aligned}
\tag{8.132}
$$

Here, P is the conformal factor in the Fubini-Study metric 5.63,

$$P = \tfrac{1}{2}(1+\zeta\bar{\zeta}),$$

and $\sigma^0(Z,\zeta,\tilde{\zeta})$ is some complex valued function (which in context represents the asymptotic value of the shear spin index σ). The idea in the N-P formalism is that if one could solve this equation, then one could construct an (asymptotically) half-flat space-time. The definition of a half-flat space-time starts with space-time analytically continued into the complex. Then the components of the spinor image of the Weyl tensor,

$$C_{abcd} \mapsto \Psi_{ABCD}\bar{\epsilon}_{\dot{A}\dot{B}}\bar{\epsilon}_{\dot{C}\dot{D}} + \tilde{\Psi}_{\dot{A}\dot{B}\dot{C}\dot{D}}\epsilon_{AB}\epsilon_{CD} \tag{8.133}$$

are now independent; here indicated by replacing $\bar{\Psi}$ by $\tilde{\Psi}$. The two components are the self-dual, and the anti-self-dual parts of the spinor. The space is half-flat, if it is Ricci flat, and,

$$\tilde{\Psi}_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0 \tag{8.134}$$

The reason a sphere S^2 enters into the picture can be motivated by a simple geometrical argument. If space time were spherical, then null rays would converge to a single point at infinity. Conformal null infinity in this case can be viewed as the intersection of a hyperplane with the 4-sphere at the north pole. However, if the space is Lorentzian and asymptotically flat, then conformal infinity looks like a hyperplane intersecting with a hyperbolic surface, which is a cone with topology $S^2 \times R$.

In spherical coordinates, the eth operator acting on a function η of spin weight, takes the tantalizing form

$$\begin{aligned}
\eth\eta &= -(\sin\theta)^s\left[\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right](\sin\theta)^{-s}\eta,\\
\bar{\eth}\eta &= -(\sin\theta)^{-s}\left[\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right](\sin\theta)^{s}\eta.
\end{aligned} \tag{8.135}$$

We have,

$$\begin{aligned}
(\eth\eta)' &= e^{i(s+1)\phi}\eta,\\
(\bar{\eth}\eta)' &= e^{i(s-1)\phi}\eta,
\end{aligned} \tag{8.136}$$

so these act as raising and lowering operators of spin weight. One can also verify that,

$$(\bar{\eth}\eth - \eth\bar{\eth})\eta = 2s\eta. \tag{8.137}$$

The eigenfunctions of $\bar{\eth}\eth\eta = \eth^2\eta = 0$ are called *spin-weighted spherical harmonics* and are denoted by ${}_sY_{lm}(\theta,\phi)$, where $|s| < l$. Some authors have pointed out that these entities were previously known to Gelfand. In the case $s = 0$, the operator $\eth^2$ is just the Laplacian, so the eigenfunctions are spherical harmonics. Since $\eth$ and $\bar{\eth}$ raise and lower the spin weight respectively, the spin-weighted spherical harmonics can be obtained by successive applications of the operators to spherical harmonics. The elegant formulas were derived by Goldberg, et-al [10].

$$\begin{aligned}
\eth^2\,{}_sY_{lm} &= -(l-s)(l+s+1)\,{}_sY_{lm},\\
\eth\,{}_sY_{lm} &= \sqrt{(l-s)(l+s+1)}\,{}_{s+1}Y_{lm},\\
\bar{\eth}\,{}_sY_{lm} &= -\sqrt{(l+s)(l-s+1)}\,{}_{s+1}Y_{lm}.
\end{aligned}$$

Applying these equations iteratively, one can show that

$$ {}_sY_{lm} = \begin{cases} \sqrt{\frac{(l-s)!}{(l+s)!}}\eth^s Y_{lm} & \text{if } 0 \le s \le l, \\ \sqrt{\frac{(l+s)!}{(l-s)!}}(-1)^s\bar{\eth}^{-s} Y_{lm} & \text{if } -l \le s \le 0 \end{cases}. \tag{8.138}$$

In the context of representation theory of the rotation group, the main result is that Wigner $D^l{}_{mm'}$ matrix elements can be neatly expressed after a messy computation, by the neat formula [10],

$$D^l{}_{-ms}(\phi,\theta,-\psi) = (-1)^m\sqrt{\frac{4\pi}{2l+1}}\,{}_sY_{lm}(\theta,\phi)e^{is\phi} \tag{8.139}$$

In 1985, T. Dray [7] proved that with a appropriate choice of spin gauge, spin weighted spherical harmonics were the same as the monopole harmonics introduced by Wu and Yang as solutions of a semiclassical electron in the field of a Dirac monopole. This is not surprising since, as we will see in the next chapter, the Dirac monopole is associated with a connection on a $U(1)$ Hopf bundle over S^2, whereas the transformation law for spin- weighted function is basically a gauge transformation in such a bundle.

8.4 SU(3)

The $SU(3)$ group was in introduced by Gell-Mann in 1961, as a candidate for a symmetry gauge group to accommodate quark "flavors". In the language a particle physics in this theory, Hadrons are made up of three quarks with flavors called: up, down and strange (u, d, s), at a time when particles with "color" attributes of charm, top, or bottom (c, t, b) were unknown. If one denotes a flavor state by,

$$|\psi_f\rangle = \begin{bmatrix} u \\ d \\ s \end{bmatrix},$$

The isospin action by $g \in SU(3)$ is simply given by $|\psi_f\rangle \mapsto g|\psi_f\rangle$. The Lie algebra $\mathfrak{su}(3)$ consists of 3×3 traceless, Hermitian matrices. The dimension of the special unitary group $SU(n)$ is n^2-1, so $\mathfrak{su}(3)$ has 8 generators. Gell-Mann chose for these generators, the closest extension of Pauli matrices. The

8 Gell-Mann matrices are, (see Georgi [9])

$$\lambda_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \lambda_2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \lambda_3 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\lambda_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \lambda_5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix},$$
$$\lambda_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \lambda_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & 0i & 0 \end{bmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \tag{8.140}$$

Anticipating a factor of $2i$ as in the case of Pauli matrices, we use the standard convention of denoting the structure constants by,

$$[\lambda_j, \lambda_k] = 2i f^i{}_{jk} \lambda_i. \tag{8.141}$$

Effectively, the matrices are normalized so that,

$$\mathrm{Tr}(\lambda_j \lambda_k) = \delta_{jk}$$

It is customary to set

$$T_j = \tfrac{1}{2}\lambda_j$$

The structure constants turn out to be completely antisymmetric, so, modulo permutations, there are only 8 independent ones. The upper 2×2 block of $\{\lambda_1, \lambda_2, \lambda_3\}$ are just the Pauli matrices, so this set constitutes an $\mathfrak{su}(2)$ subalgebra, and we have

$$f^i{}_{jk} = \epsilon^i{}_{jk}$$

whenever all indices are less than or equal to 3. The isotopic spin $SU(2)$ algebra generated by the exponential map of these generators, result in rotations of the flavor state $|\psi_f\rangle$ that leave s invariant. Defining,

$$\begin{aligned} \tau_+ &= \sqrt{3}\lambda_8 + \lambda_3, \\ \tau_- &= \sqrt{3}\lambda_8 - \lambda_3, \end{aligned}$$

one can also identify two more $\mathfrak{su}(2)$ subalgebras generated by $\{\lambda_4, \lambda_5, \tau_+\}$ and $\{\lambda_6, \lambda_7, \tau_-\}$ respectively. The rest of the non-zero structure constants are easily computed using symbolic manipulation software. The results are,

$$\begin{aligned} f_{147} &= f_{246} = f_{257} = f_{345} = \tfrac{1}{2}, \\ f_{156} &= f_{367} = -\tfrac{1}{2}, \\ f_{458} &= f_{687} = \tfrac{1}{2}\sqrt{3}. \end{aligned}$$

A concise formula for the structure constants is given by,

$$f_{ijk} = -\tfrac{i}{4}\,\mathrm{Tr}(\lambda_i[\lambda_j, \lambda_k]) \tag{8.142}$$

Another useful fact is the formula for the anti-commutators,

$$\{\lambda_j, \lambda_k\} = \tfrac{4}{3}\delta_{jk} I + 2{d^i}_{jk}\lambda_k, \qquad \text{where,}$$
$$d_{ijk} = \tfrac{1}{4}\,\mathrm{Tr}(\lambda_i\{\lambda_j, \lambda_k\}).$$

The Killing form,

$$g_{jk} = {f^i}_{jm}{f^m}_{ki} = 3\delta jk,$$

modulo the usual problematic factor of $2i$, is non-degenerate and positive definite, as expected from a semi-simple, compact group. The Cartan subalgebra is spanned by $h_3 = T_3$ and $h_8 = T_8$ which commute with all the other generators. Thus, the subalgebra is of rank 2 and there are 2 Casimir operators. Since the generators of the Cartan subalgebra are already in diagonal form, is is easy to find the eigenvectors and corresponding weights [9],

$$\begin{bmatrix}1\\0\\0\end{bmatrix} \mapsto (\tfrac{1}{2}, \tfrac{1}{2\sqrt{3}}), \quad \begin{bmatrix}0\\1\\0\end{bmatrix} \mapsto (-\tfrac{1}{2}, \tfrac{1}{2\sqrt{3}}), \quad \begin{bmatrix}0\\0\\1\end{bmatrix} \mapsto (0, -\tfrac{1}{\sqrt{3}}).$$

To find the roots we need generators that take one weight to another. From

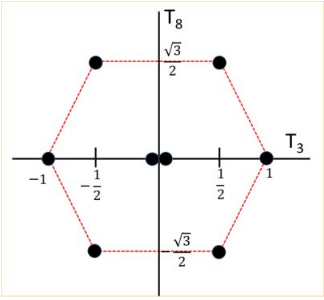


Fig. 8.5: Root Diagram $A_2 = \mathfrak{SU}(3)$

lessons learned from $\mathfrak{su}(2)$ we take as raising and lowering operators,

$$\frac{1}{\sqrt{2}}(T_1 \pm iT_2), \quad \frac{1}{\sqrt{2}}(T_4 \pm iT_5), \quad \frac{1}{\sqrt{2}}(T_6 \pm iT_7).$$

A straight-forward computation of the commutators with the generators of the Cartan subalgebra gives,

$$[T_3, (T_1 \pm iT_2)] = \pm((T_1 \pm iT_2),$$
$$[T_8, (T_1 \pm iT_2)] = 0,$$

So, we have found two roots, $(\pm 1, 0)$. Continuing the computation for the next set of ladder operators, we get,

$$[T_3, (T_4 \pm iT_5)] = \pm\frac{1}{2}((T_4 \pm iT_5),$$
$$[T_8, (T_4 \pm iT_5)] = \pm\frac{\sqrt{3}}{2}((T_4 \pm iT_5),$$

so the next pair of roots are $(\pm\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$. Finally, the commutators for the third set of ladder operators, leads to,

$$[T_3, (T_6 \pm iT_6)] = \mp\frac{1}{2}((T_6 \pm iT_7),$$
$$[T_8, (T_6 \pm iT_6)] = \pm\frac{\sqrt{3}}{2}((T_6 \pm iT_6),$$

giving the last pair of roots $(\mp\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$. The complexification of $\mathfrak{su}(3)$ is $\mathfrak{sl}(3, \mathbf{C})$. In Cartan's classification of semisimple Lie algebras, the root system for $\mathfrak{sl}(n+1, \mathbf{C})$ is called A_n.

In the root diagram $A_2 = \mathfrak{su}(3)$, the roots form a regular hexagon with two roots at the center, as shown in figure 8.5. We refer the reader to the famous book by Georgi [9] for a full discussion of representations of groups in Physics.

Chapter 9

Bundles and Applications

9.1 Fiber Bundles

The late 1970's was an exciting time to be a graduate student at Berkeley. At the time, the University had a powerhouse of some of the top, world-class mathematicians in differential geometry and related fields, including, Chern, Kobayashi, Wolf, Gilkey and Weinstein; a number of renowned general relativity researchers such as Taub, Marsden, and Sachs; as well as a battery of visiting faculty and invited speakers at the frontiers of research. Prior to 1975, with the exception of Professor Sachs, who had a dual appointment in the physics department, I don't think I ever saw, either as an undergraduate or as graduate student, a physics professor enter the math building or a math professor walk the hallways of the physics building. It just so happened, that on 1975, Belavin, Polyakov, Schwartz, and Tyupkin, published a paper on pseudoparticle solutions to the Yang-Mills equations [3]. This so-called BPST instanton, drew widespread attention in the physics community. The instantons are extremals under a variational principle of the Yang-Mills Lagrangian, which generalizes the electromagnetic Lagrangian 2.123, to non-Abelian Lie algebras. The paper included a provocative discussion of topological properties such as homotopy classes and a footnote referring to a particular equation as a Pontrjagin class. A. Trautman [37] is credited as the first mathematician to observe that the BPST instanton (and Dirac's monopole) corresponded to a connection on a Hopf bundle. The details will be presented in this chapter. In 1977, Schwarz (apparently the correct spelling) used the Atiyah-Singer index theorem to show that the number of instantons and zero fermion modes is given by some topological invariant $(8k - 3)$ [32]. Perhaps the inclusion of such heavy-duty mathematics made some of the particle physicists a bit uncomfortable. I say this because in 1977, when I. M. Singer was offered a position at Berkeley, his seminars on the Penrose twistor programme and gauge theory got flooded with non-Abelian gauge physicists including Mandelstam. The relevance of twistor theory to gauge fields was first established by R. Ward in a brilliant short paper [39] in which he showed that certain complex vector bundles related to $\mathbf{CP}^3$, in twistor theory, could be used to generate self-dual gauge fields. This also drew the attention of algebraic geometers such as R. Hartshorne, working on moduli spaces of vector bundles, The presence of Singer at Berkeley attracted a slew of prominent visitors such as M. Atiyah, S. Yau, and later, A. Lichnerowicks. Atiyah and Singer became major contributors to the mathematical formulation of Yang-Mills Theory; in particular, in a paper published in 1978, [2], Atiyah, Hitchin and Singer computed the dimension of the moduli space of irreducible, self-dual connections for Yang-Mills equations in 4-space for all compact gauge groups. The dimension of this space for $SU(2)$ is the topological invariant $(8k-3)$ derived by Schwarz.

Among the physicists attending Singer's lectures was a young researcher named A. Hanson with whom I partnered to become the note-takers for the seminar series. The section on Yang-Mills in these notes, is partly distilled from my 1977-78 notes with Hanson on the lectures by Singer. A couple of years later, Hanson, along with T. Eguchi and P. Gilkey, who introduced me to algebraic topology, published a comprehensive work, in which among other things, they announced the discovery of a new metric for a gravitational instanton [8]

We have already encountered several examples of fiber bundles of interest in physics. The most fundamental of these are the tangent and cotangent bundles and tensor product thereof. We have also discussed the Hopf bundle which is endowed with a non-trivial topology. We present now a more formal approach to fiber bundles, still keeping in mind that to try to make the material more accessible to physicists, we occasionally might sacrifice a bit of rigor in favor of simplicity.

9.1.1 Definition A smooth *fiber bundle* is a set $\xi = \{E, M, \pi, F\}$ consisting of manifolds E, M and F and a C^∞ map $\pi : E \to M$ from E onto M satisfying the following properties:

1. The map is *locally trivial*. This means that for every point $p \in M$, there exists an open set U containing p, together with a diffeomorphism $\phi : \pi^{-1}(U) \to U \times F$. We call $F_p = \pi^{-1}(\{p\})$ the fiber at the point p. We allow the simpler notation $F_p = \pi^{-1}(p)$ for the fiber at p and

$$F|_U = \pi^{-1}(U) = \bigcup_{p \in U} F_p$$

   for the fiber space over the set U. This part of the definition says that locally, the bundle looks like a cross-product; that is, $\pi^{-1}(U) = U \times F$. The pair $\{U, \phi\}$ is called a coordinate neighborhood or a coordinate patch, or a local trivialization of the bundle.

2. The fiber space is glued in a smooth manner. More specifically, if $\{U_i, \phi_i\}$ and $\{U_j, \phi_j\}$ are two coordinate neighborhoods with nonempty intersection and $p \in U_i \bigcap U_j$, then

$$\phi_{ij} = \phi_i^{-1}\phi_j : U_i \times F \to U_j \times F$$

   is a diffeomorphism. The quantities ϕ_{ij} are called transition functions. The sets $\{U_i\}$ constitute an open cover of M. (See figure 9.1).

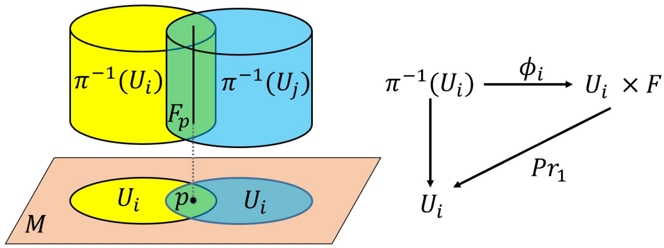


Fig. 9.1: Fiber Bundle

The manifold M is called the base space, E is called the total space or the bundle space, and π the projection map. The notation

$$F \to E \xrightarrow{\pi} M \tag{9.1}$$

or simply

$$\pi : E \to M$$

is also used, probably because it is easier to typeset. There is a common abuse of language in calling E the bundle space, since the bundle is really the set ξ. A *trivial bundle* is one that is a simple cross product, $E = M \times F$. In that sense, we could view $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ as a bundle with base space $M = \mathbf{R}$ and fiber $F = \mathbf{R}$. There would be no advantage to view $\mathbf{R}^2$ this way, but it is a bundle anyway. The total space is the union of all the fibers, which locally does look like a cross product, but globally it might have a non-trivial topology, as in the case of the Hopf bundle 8.1.4. It is possible to start the treatment of fiber bundles with topological spaces, in which case, the projection map would be a continuous function, the coordinate maps homeomorphisms, and the base space would have the quotient topology. These topological bundles can then be given differentiable structures as above. In fact, it would be more natural to treat this subject in terms of categories, morphisms, and functors, but we will resist the temptation.

9.1.2 Example Let $M = S^1$, $I = [-1, 1]$ and $E = S^1 \times I$. This trivial bundle is just a cylinder with a circular base space and each fiber a copy of the interval $[-1, 1]$. Topologically, we can construct the bundle by looking at the base space

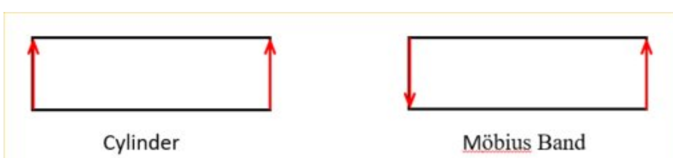


Fig. 9.2: Möbius Band

as an interval, say $[-\pi, \pi]$ with the end points identified. We glue the bundle by identifying the fibers at the endpoints, as suggested by the arrows pointing in the same direction in the picture in the left, on figure 9.2. On the other hand, if we identify the vertical edges in opposite direction as shown in rectangle on the right, we get a strip with a twist, which is topologically equivalent to a Möbius band. A smooth parametrization of this surface and a picture rendering the surface is given in 5.1.

9.1.3 Definition A *section* s of a bundle $\xi = \{E, M, \pi, F\}$ is a smooth map $s : M \to E$ such that

$$\pi \circ s = id_M$$

This is the same concept introduced in the context of the tangent bundle for which the sections are called vector fields (See figure 1.1). We use the same notation $\Gamma(E)$ for the set of all smooth sections. For the tangent bundle, the fibers are copies of $\mathbf{R}^n$, so they are vector spaces of dimension n. In such a case we call the bundle a *vector bundle*. The tangent bundle $T\mathbf{R}^n$ is a trivial bundle. If $p \in U \subset \mathbf{R}^n$, then $\pi^{-1}U \cong U \times \mathbf{R}^n$ and the coordinate patch maps are,

$$\phi : \left(p, a^1\frac{\partial}{\partial x^1}, \ldots, a^n\frac{\partial}{\partial x^n}\right) \mapsto (p, a^1, \ldots, a^n).$$

This slightly more formal point of view is consistent with our earlier definition of a tangent vector. A vector bundle has more structure that a run-of-the-mill fiber bundle. If (p, f) is an point in the bundle in the intersection of two coordinate patches $\{U_i, \phi_i\}$ and $\{U_j, \phi_j\}$, the transition functions satisfy,

$$\phi_{ij} = (p, \varphi_{ij}(p) f),$$

where $\varphi_{ij}(p) \in GL(n, \mathbf{R})$[1] gives a linear isomorphism on the fibers. For the tangent bundle, this is yet another way of saying that if we change coordinates near p, the components of a tangent vector at p change by an action of an element of $GL(n, \mathbf{R})$ represented by the Jacobian. If the fibers are k-dimensional with a $GL(k, \mathbf{R})$ action on the fibers, we just say that the vector bundle is k-dimensional, even though the real total dimension is $(n + k)$. A vector bundle of dimension 1 is called a *line bundle*. The normal bundle of the sphere S^2 in $\mathbf{R}^3$ would be an example of a line bundle. Since the fibers of vector bundles are vector spaces and every vector space has a 0 vector, there is a special section s such that $s(p) = (p, 0)$; this is called the *zero section*. This is a trivial global section in all vector bundles. The set of all sections of a vector bundle has a natural structure of a vector space, in which the zero-section is the zero-vector. There is no problem finding non-singular global sections of the tangent bundle $T(\mathbf{R}^n)$. However, for submanifolds M of $\mathbf{R}^n$, there might be topological obstructions to the existence of vector fields that are non-zero everywhere. For example, the reader may be acquainted with the theorem that one can not "comb" a hairy sphere S^2. As proved by Poincaré, the obstruction in this case is the Euler characteristic which would need to vanish, but for the sphere, $\chi(S^2) = 2$.

[1] Again, we adopt this notation with reluctance, although is common in the literature. The quantities $\varphi_{ij}(p)$ are matrix-valued, so they really should be written as ${\varphi^i}_j(p)$

9.1.4 Definition A *covering space* of a space M is a bundle $\xi = \{E, M, \pi, F\}$ in which every point $p \in M$, has a neighborhood $U \subset M$, such that $\pi^{-1}(U)$, is the disjoint union of sets V_k, each homeomorphic to U. In other words, the fibers of U are discrete. Here, π is a local homeomorphism and M has the quotient topology.

9.1.5 Example Most likely, the first examples of covering spaces students encounter in a course in algebraic topology are associated with the circle S^1. The map $\pi : S^1 \to S^1$ given by

$$\pi(e^{i\theta}) = e^{in\theta}, \qquad n \in \mathbf{Z}^+$$

gives an n-sheet covering of S^1 for each positive integer n. One can envision this covering as a rubber band folded into n loops around a cylinder. The continuous homomorphism $\pi : \mathbf{R} \to S^1$ given by

$$\pi(t) = e^{it}$$

is a covering of S^1 by the real line. A covering that is simply connected as it is the case here, is called a *universal covering space.* This map is the starting point in establishing that the fundamental group of S^1, also called the first homotopy group π_1, is given by

$$\pi_1(S^1) \cong \mathbf{Z}$$

A differentiable manifold structure is not required in this example. All that is needed is that $S^1 = \mathbf{R}/\mathbf{Z}$ is a topological group, $\mathbf{R}$ is simply connected, and $\mathbf{Z}$ is a discrete subgroup of $\mathbf{R}$. The homotopy equivalence classes are the loops that have the same winding number. The additive group structure is given adding the number of loops of two group elements.

9.1.6 Example The projective space $\mathbf{RP}^n$ of lines in $\mathbf{R}^{n+1}$ is defined by the quotient space of S^n obtained by identifying antipodal points; that is, the two antipodal points of intersection of the sphere with a line through the origin. The covering space of $\mathbf{RP}^n$ is the real version of the Hopf bundle

$$\pi : S^n \to \mathbf{RP}^n,$$

with fiber $\mathbf{Z}_2$. The group covering transformations consist of the identity and the antipodal map. In this example

$$\pi_1(\mathbf{RP}^n) \cong \mathbf{Z}_2.$$

I have to thank professor P. Gilkey for motivating me to overcome the fear of algebraic topology machinery, with his wonderful illustration of the above, for the case $n = 3$. He showed up the first day of classes with a toy consisting of two equilateral triangles, one large, one small. The triangles were connected at corresponding vertices with three separate untangled strings. He set the large triangle on the table and flipped the small triangle in the air by half revolution. He asked for a volunteer in a class of 5 students to untangle the strings only allowing parallel translations of the small triangle. Clearly, it was not doable. He reset the toy and then flipped the small triangle a full revolution. I got lucky and untangled the strings almost immediately. He proceeded to illustrate by combinations of half-turns and full-turns, until it was almost self-evident, that there were only two possible outcomes. Of course, the key question of the day was, how does one prove that? He explained that the deformations of the shape of the strings by parallel motion were examples of homotopies; that a rotation in space left two antipodal points on a sphere fixed; defined projective space; and concluded that what we had here was a manifestation that the first homotopy group of the projective space had two generators. Four weeks later we had learned enough tools to prove the assertion.

9.1.7 Definition Let $\xi = \{E, M, \pi, F\}$ and $\xi' = \{E', M', \pi', F'\}$ be smooth vector bundles. A *vector bundle map* is a pair of smooth maps $f_M : M \to M'$ and $f : E \to E'$, such that

1. The diagram 9.3 commutes, ,

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[d, "\pi"] & E' \arrow[d, "\pi'"'] \\ M \arrow[r, "f_M"] & M' \end{tikzcd}$$

Fig. 9.3: Bundle Map: $f_M \circ \pi = \pi' \circ f$.

2. The map induced by f on the fibers is a linear map.

The meaning of this commuting diagram is that fibers are mapped to fibers. Indeed, if $F_p = \pi^{-1}(p)$ is a fiber at $p \in M$, then,

$$\begin{aligned} (\pi' \circ f)(F_p) &= (f_M \circ \pi)(F_p), \\ &= (f_M \circ \pi)(\pi^{-1}(p)), \\ &= f_M(p), \end{aligned}$$

so that a point in the fiber of $p \in U \subset M$ lands on a point in the fiber of $f_M(p)$. Thus if makes sense to say that the map induced on the fibers needs to be a linear map of vector spaces. More specifically, if $(p, v) \in E$, then $f(p, v) = (f_M(p), T(p) \cdot v)$, where $T : U \to\in \mathbf{L}(F, F')$ gives a linear map $T(p)$ from the fiber F_p to the fiber $F'_{f_M(p)}$. If these linear maps are vector space isomorphisms and f_M is a diffeomorphism, the bundle map is called a *vector bundle isomorphism.* In this case, the two bundles are essentially the same.

9.1.8 Definition *Pull-back bundle*

Let $\pi : E \to M$ be a vector bundle with fiber F and let $f : M' \to M$ be a smooth map. We can define a pull-back bundle denoted by $f^*\pi : f^*E \to M'$ by assigning the fiber $F_{p'}$ to each point $p' \in M'$, the corresponding fiber $F_{f(p')}$ at $p = f(p')$. More precisely, if $v \in E$, then

$$f^*E = \{(p', v) \in M' \times E \mid f(p') = \pi(v)\}, \quad f^*\pi(p', v) = p' \tag{9.2}$$

It clear that $f^*E \subset M' \times E$ is locally trivial, as it should be. If $\{U_i, \phi_i\}$ is a cover of $\pi : E \to M$ by coordinate patches, then $\{f^{-1}(U_i)\}$ is a cover of M' with transition functions $f^*\phi_{ij}(p') = \phi_{ij}(p)$. if,

$$M'' \xrightarrow{g} M' \xrightarrow{f} M$$

then,

$$(f \circ g)^*E \cong f^*(g^*E)$$

This should elicit memories of the properties of the pull-back of differential forms 2.68.
One of the main application of pull-back bundles is the interesting relation to homotopy. Recall from definition 7.1.16, that two maps $f, g : M' \to M$ are called *homotopic* if there exist a map $\phi : M' \times [0,1] \to M$, such that

a) $\phi(p', 0) = f(p)$,
b) $\phi(p', 1) = g(p)$.

The main theorem in this regard is that if f and g are homotopic, then the pull-back bundles $f^*E \cong g^*E$ are isomorphic.

9.1.9 Corollary Let $M' = M =$ be a contractible space and suppose the homotopy is a deformation retract (see 7.1.17),

$$\phi_t(p) = \phi(p, t) = tp.$$

Then the theorem says that $E = f^*E \cong g^*E = M \times F$, which proves that if M is contractible, then E is trivial.

9.2 Principal Fiber Bundles

A smooth principal fiber bundle (PBF) is essentially a fiber bundle in which the fibers are Lie groups, or manifolds on which there is a free and transitive action by a Lie group . In that sense, one can view the base space as the parameter space for a family of fibers F where at any point p in the base space, the fiber F_p is diffeomorphic to a Lie group. More formally, we have the following definition

9.2.1 Definition *Principal fiber bundle*
A smooth principal fiber bundle is a set $\xi = \{E, M, \pi, F, G\}$, where $F \to E \xrightarrow{\pi} M$ is a fiber bundle, and G is a Lie group acting on E freely along the fibers.

As a reminder, G acts on E on the right, if there exists a smooth map $\mu : E \times G \to E$ given by $(b, g) \mapsto bg$. The action is free along the fibers if the only element that acts as an identity is the identity and the action is transitive if any two points on the same fiber are connected by an element of the group. Thus, if the action is free and transitive, the fibers are diffeomorphic to the orbits of the group. This means that the bundles

$$\begin{array}{ccc} E & \xrightarrow{id} & E, \\ \pi \downarrow & & pr \downarrow \\ M & \xrightarrow{\simeq} & E/G \end{array}$$

are isomorphic. Here, pr is the natural projection of E onto its cosets.

9.2.2 Example Bundle of Frames

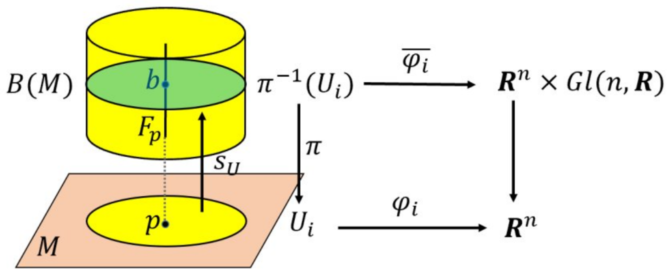


Fig. 9.4: Bundle of Frames

,

One of the most important principal fiber bundles is the bundle of frames. We hope that the following discussion does not obscure simplicity for the sake of formalism. The bundle of frames is the structure one gets by attaching to each point on a manifold, the space of all possible frame fields of tangent vectors. The structure group is $Gl(n, \mathbf{R})$, which acts on the fibers at each point on the manifold, by matrix multiplication that changes one frame into another. Parallel transport of vectors and frames on the manifold, correspond to choosing a horizontal subspace of the tangent space of the bundle. Choosing a horizontal subspace in the bundle is then equivalent to choosing a connection.

9.2.3 Definition Let M be a smooth manifold of dimension n, and let $\{(\phi_i, U_i)\}$ be an atlas of coordinate charts covering M. As usual, we label the coordinates of $p \in U_i$ as $(x^1, \ldots, x^n)$. The *frame bundle* $B(M)$ is defined as the bundle $\pi : B(M) \to M$, where

$$B(M) = \{(p, e_1, e_2, \ldots, e_n) : p \in M, \beta = (e_1, e_2, \ldots) \text{ is a basis for } T_p(M).\}$$

The projection map π is given by

$$\begin{aligned} \pi : F(M) &\to M, \\ (p, e_1, \ldots e_n) &\xrightarrow{\pi} p. \end{aligned}$$

Let (ϕ, U) be a coordinate chart with coordinates $(x^1, \ldots, x^n)$, with $\overline{U} = \pi^{-1}U$. As shown in figure 9.4, we can lift the chart in M

$$\phi : U \to \mathbf{R}^n$$

to a chart in $B(M)$,

$$\overline{\phi} : \overline{U} \to \mathbf{R}^n \times Gl(n, R) \simeq \mathbf{R}^{n+n^2},$$

as follows. Let $\{\partial_i = \frac{\partial}{\partial x^i}\}$ be the standard basis for the tangent space T_pM, associated with the coordinates $p = (x^1, \ldots, x^n)$. Let $b \in B(M)$ be a point with coordinates

$$b = (p, e_1, \ldots, e_n)$$

on the fiber F_p. Then, as shown in equation 3.1, there exists a matrix $A \in Gl(n, \mathbf{R})$, with

$$e_i = A^j{}_i \partial_j$$

Matrix multiplication by A is on the right, but, as in 3.1, we choose to write the equation as above, to make it clear that ∂_i is not acting as a differential operator on the components of A. Thus, $b \in F_p$ can be written as

$$b = (p, A^i{}_j \frac{\partial}{\partial x^i}).$$

The bundle coordinate patch on $\overline{U}$ is defined by

$$\overline{\phi}_U (p, e_1, \ldots, e_n) = (p, A^i{}_j)$$

Here, we identify the matrix $A \in Gl(n, \mathbf{R})$ with a point in $\mathbf{R}^{n^2}$, where the coordinates in $\mathbf{R}^{n^2}$ are given by the entries in the column vectors of A. The standard coordinates of $Gl(n, \mathbf{R})$ are given by the matrices $X^i{}_j$ that have a 1 entry in the i^{th} row, j^{th} column, and a 0 entry everywhere else.
The right action $\mu : B(M) \times Gl(n, \mathbf{R}) \to B(M)$ along the fibers is given by

$$\mu : (b, g) \to bg = (p, f_i = e_j \, g^j{}_i); \quad g \in Gl(n, \mathbf{R}).$$

Given two overlapping charts in $B(M)$

$$\overline{\phi}_i : \pi^{-1}(U_i) \to U_i \times Gl(n, \mathbf{R}),$$
$$\overline{\phi}_j : \pi^{-1}(U_j) \to U_i \times Gl(n, \mathbf{R}).$$

with transition functions $\overline{\phi}_{ij}$ on $\overline{U}_i \cap \overline{U}_j$, given by $\overline{\phi}_{ij} = (\overline{\phi})_i^{-1} \overline{\phi}_j$. The group action on the fibers satisfies,

$$(bg_1)g_2 = b(g_1 g_2).$$

The atlas $(\overline{\phi}_i, \overline{U}_i)$ thus gives $B(M)$ the structure of a differentiable manifold. A local section of the bundle $s_U \in \Gamma(B(m)))$ represents a smooth choice of a family of frames at points in U, with $\pi \circ s_U = id$.

9.2.4 Example Let G be Lie group and $H \subset G$ a compact subgroup. Then $\pi : G \to G/H$, where π is the projection map onto the orbit space, is an H-bundle. This is one of the early results in the theory of fiber bundles, first proved by H. Samelson in 1941.

9.2.5 Example The projective space fibration
Recall that the complex projective space $\mathbf{CP}^n$ is the quotient space $\mathbf{C}^n/\sim$, where $a, b \in \mathbf{C}^n$ are equivalent, $a \sim b$, if there exists a $c \in S^1$ such that $a = bc$. That is, $\mathbf{CP}^n$ is the space of lines through the origin. Then.

$$S^{2n+1} \xrightarrow{S^1} \mathbf{CP}^n$$

is a principal $U(1)$ bundle. The special case for $n = 1$ is the ubiquitous Hopf map 8.32.

9.2.6 Example The $SO(n)$ bundle
The group of special orthogonal matrices

$$SO(n) = \{A \in M_{n\times n}(\mathbf{R}) : A^{-1} = A^T, \det A = 1\}$$

acts transitively on the unit sphere $S^{n-1} \subset \mathbf{R}^n$. The subgroup that leaves the north pole fixed, that is, the isotropy subgroup of the point $e_1 = (1, 0, \ldots, 0)$, is the set of matrices of the form

$$A = \begin{bmatrix} 1 & 0 \\ 0 & B \end{bmatrix}, \qquad B \in SO(n-1).$$

Thus, $SO(n)/SO(n-1)$ is diffeomorphic to S^{n-1}. This, with the projection map

$$\pi : SO(n) \xrightarrow{SO(n-1)} S^{n-1}$$

constitutes a $SO(n-1)$ bundle.

9.2.7 Example The $U(n)$ bundle
The group of unitary matrices

$$U(n) = \{A \in M_{n\times n}(\mathbf{C}) : A^{-1} = A^{\dagger}\}$$

acts transitively on the unit sphere $S^{2n-1} \subset \mathbf{C}^n$. The subgroup that leaves the north pole fixed, that is, the isotropy subgroup of $e_1 = (1, 0, 0, \ldots, 0)$, is the set of matrices of the form,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & B \end{bmatrix}, \qquad B \in U(n-1).$$

Thus, $U(n)/U(n-1)$ is diffeomorphic to S^{2n-1}. This, with the projection map

$$\pi : U(n) \xrightarrow{U(n-1)} S^{2n-1}$$

constitutes a $U(n-1)$ bundle. The bundle structure above is also true if the unitary groups is replaced by special unitary groups, which corresponds to requiring the matrices A to have $\det A = 1$, or equivalently, to picking an orientation of the frames.

9.2.8 Example The Grassmannian
Let $\mathbf{F}$ denote, either of the vector spaces $\mathbf{R}$ or $\mathbf{C}$. The space of orthonormal k-frames in $\mathbf{F}^n$ is called the Stiefel manifold $V_k(\mathbf{F}^n)$. We can characterize the Stiefel manifolds as the set of $n \times k$ matrices,

$$V_k(\mathbf{F}^n) = \{A \in M_{n\times k} :\ AA^\dagger = I_k\}, \tag{9.3}$$

where I_k is the $k \times k$ identity matrix. We interpret I_k as a matrix $I_k = [e_1, e_2, \ldots, e_k]$ of orthonormal basis vectors. Viewed as representing linear transformations $L(\mathbf{F}^k, \mathbf{F}^n)$, the matrices A have rank k.

Analogous to the construction of projective spaces, we say that two matrices $A, B, \in V_k(\mathbf{F}^n)$ are equivalent, $A \sim B$, if there exists a $k \times k$ orthogonal (or unitary in the complex case) matrix C such that $A = BC$. The *Grassmannian* is defined as

$$Gr(k, \mathbf{F}^n) = V_k(\mathbf{F}^n)/\sim . \tag{9.4}$$

The Grassmannian is the space of k-planes in $\mathbf{F}^n$. In $\mathbf{R}^n$ the projection map

$$\pi : Gr(k, \mathbf{R}^n) \xrightarrow{O(k)} V_k(\mathbf{R}^n), \tag{9.5}$$

is a principal bundle with fiber group $O(k)$. The group $O(n)$ acts transitively on $V_k(\mathbf{R})$, and the isotropy subgroup of the $n \times k$ matrix

$$A = \begin{bmatrix} I_k \\ 0 \end{bmatrix}.$$

consists of the matrices in $O(n)$ of the form

$$\begin{bmatrix} I_k & 0 \\ 0 & B \end{bmatrix},$$

where $B \in O(n-k)$. Thus, we have a diffeomorphism,

$$O(n)/O(n-k) \xrightarrow{\cong} V_k(\mathbf{R}^n),$$

and we can write the Grassmannian bundle as,

$$\pi : Gr(k, \mathbf{R}^n) \xrightarrow{O(k)} O(n)/O(n-k). \tag{9.6}$$

Equivalently, we have

$$Gr(k, \mathbf{R}^n) \cong \frac{O(n)}{O(k) \times O(n-k)} \tag{9.7}$$

In the complex and quaternionic vector spaces, we have

$$
\begin{aligned}
Gr(k, \mathbf{C}^n) &\cong \frac{U(n)}{U(k) \times U(n-k)}, \\
Gr(k, \mathbf{H}^n) &\cong \frac{Sp(n)}{Sp(k) \times Sp(n-k)},
\end{aligned} \tag{9.8}
$$

The Grassmannian of lines in $\mathbf{R}^3$ is the projective space $\mathbf{RP}^2$ and the Grassmannian of planes is the same since to every line through the origin, there corresponds a unique orthogonal plane. Hence, the simplest Grassmannian that is not a projective space is the space $Gr(2, \mathbf{R}^4)$ of planes in $\mathbf{R}^4$, or equivalently, the space of projective lines in $\mathbf{CP}^3$. The space $\mathbf{CP}^3$ of projective lines in $\mathbf{C}^4$ with a Hermitian metric of signature $(++--)$ is the base space for twistor theory; in this case the symmetry group preserving the metric is $SU(2,2)$ which is the double cover of the conformal group. It has been cited by R. Penrose [29], that the geometrical foundation for the theory can be traced back to the work of Plücker and Klein on subspaces of planes. In view of this historical setting, we present a brief summary of the parametrization of two dimensional subspaces of $\mathbf{R}^4$ used by Plücker. Given two vectors $\mathbf{v} = (v^1, v^2, v^3, v^4)$ and $\mathbf{w} = (w^1, w^2, w^3, w^4)$, we can determine a plane by a linear map $L(\mathbf{R}^2, \mathbf{R}^4)$ with matrix representation given by a 4×2 matrix A of rank 2, whose column vectors are $\mathbf{v}^T$, and $\mathbf{w}^T$. That is,

$$
A = \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \\ v_3 & w_3 \\ v_4 & w_4 \end{bmatrix},
$$

The Grassmannian $Gr(2, \mathbf{R}^4)$ is the space of equivalence classes of such matrices. *Plücker coordinates* p_{ij} are defined by the determinants of pairs of rows i and j.

$$
\begin{aligned}
p_{ij} &= v_i w_j - v_j w_i, \\
&= \begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} \\ -p_{12} & 0 & p_{23} & p_{24}, \\ -p_{13} & -p_{23} & 0 & p_{34}, \\ -p_{14} & -p_{24} & -p_{34} & 0 \end{bmatrix}.
\end{aligned}
$$

We may view the antisymmetric matrix $P = (p_{ij})$ as a element of the Lie algebra $\mathfrak{so}_4$, and the quantities $(p_{12} : p_{13} : p_{14} : p_{34} : p_{24} : p_{23})$ as homogenous coordinates in $\mathbf{RP}^5$. A short computation using the antisymmetry property in the definition of p_{ij}, yields,

$$
p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0
$$

The square of this entity is equal to $\det(P)$. This equation, up to a constant, represents a quadric hypersurface Q in $\mathbf{RP}^5$. The permutation signs of the equation give a hint that there is a duality lurking somewhere, namely the orthogonal subspaces. Introducing independent coordinates

$$
\begin{aligned}
p_{12} &= X + R, & p_{34} &= X - R, \\
p_{13} &= S - Y, & p_{24} &= S + Y, \\
p_{14} &= Z + T, & p_{23} &= Z - T,
\end{aligned}
$$

the equation becomes,

$$(X^2 + Y^2 + Z^2) - (R^2 + S^2 + T^2) = 0$$

By re-scaling the homogenous coordinates we can write the quadric Q as,

$$X^2 + Y^2 + Z^2 = 1, \qquad R^2 + S^2 + T^2 = 1.$$

Thus, Q has the topology of a torus $S^2 \times S^2$.

9.2.9 Example S^3
The $U(n)$ bundle with $n = 2$ gives $U(2)/U(1) \cong S^3$.
The Grassmannian $Gr(1, \mathbf{C}^2) \cong \mathbf{CP}^1 \cong S^2$, gives the bundle

$$
\begin{gathered}
U(2)/U(1) \xrightarrow{U(1)} \frac{U(2)}{U(1) \times U(1)}, \\
S^3 \xrightarrow{S^1} S^2.
\end{gathered}
$$

This view of the Hopf bundle is the foundation for the argument that a three-sphere S^3 is homeomorphic to the union of two solid tori whose intersection are their common boundaries with topology $S^1 \times S^1$. As a static picture, figures such as in 8.3 are as good as it gets in trying to visualize the union of the two tori.

9.2.10 Definition *Associated vector bundle*
Let $\xi = \{E, M, \pi, F, G\}$ be a PFB. Suppose there is a vector space V on which G acts on the left. Let $(e, v) \in E \times V$ and $g \in G$. We can define an action on the cross product $E \times V$ by

$$(e, v)g \to (eg, g^{-1}v),$$

Denote $(E \times V)/G$ by $E \times_G V$. Then the natural projection map,

$$p : E \times_G V \xrightarrow{V} M$$

defines a vector bundle called the associated vector bundle.

9.2.11 Example If $E = B(M)$ is the bundle of frames, and $V = \mathbf{R}^n$, the associated vector bundle is the tangent bundle.

9.3 Connections on PFB's

We have noted earlier, that in a Riemannian manifold, there exists a unique, torsion-free connection (see theorem 6.9). The metric on the manifold allows us to define orthonormal frames, and the connection gives a prescription on how to parallel transport tangent vectors and frames. We now present the bundle viewpoint of connections. We will do this quite generally, but the reader should keep in mind the bundle of frames as the model space. There is a learning curve for the mathematical formalism, but the idea is very intuitive. We illustrate this with bundle of frames. Given a point p on a manifold M, the fiber of the bundle of frames $F(M)$ consists of the point and all the frames at that point. The action of the general linear group along the fibers, transforms one frame at p onto another frame at p. Since the right action of the group is transitive and effective, there is a natural way to identify a vertical direction in the tangent space at a point b on the bundle, namely, a vertical tangent vector at b on the frame bundle, corresponds to a frame at the point p in the base manifold. If the frames are restricted to be orthonormal, the picture is the same, but one has to reduce the group to the orthogonal group. Thus, the action of the group tells us how move frames along the fibers, but it does not tell us how to move a frame to the fiber of a nearby point on the manifold. This requires more structure, namely a connection. In the case of Riemannian manifold, the natural structure is provided by the Levi-Civita connection, which quantifies how to parallel transport a frame along any particular curve. Lifting the curve and the moving the frames along that curve in the bundle, would then yield a section of the bundle that we could identify as a horizontal direction. Thus, the basic idea of connection on a principal fiber bundle amounts to choosing horizontal direction for the tangent space of the bundle.

Let $\xi = \{E, M, \pi, F, G\}$ be a principal fiber bundle and let the coordinate chart U_i give a local trivialization of the bundle

$$\begin{tikzcd} \pi^{-1}(U_i) \arrow[r, "\phi_i"] & U_i \times G. \\ U_i \arrow[u, "s_i"] \arrow[ur] & \end{tikzcd}$$

Let $p \in U_i$ and $b \in F_p \subset \pi^{-1}(U_i)$, so that $\pi(b) = p$. Here we use the notation

$$\begin{aligned} \phi_i : \pi^{-1}(U_i) &\to U_i \times G, \\ \phi_i(b) &= (\pi(b), \varphi_i(p)). \end{aligned}$$

On the overlap $U_i \cap U_j$ of two coordinate charts with

$$\begin{aligned} \phi_i(b) &= (\pi(b), \varphi_i(p)), \\ \phi_j(b) &= (\pi(b), \varphi_j(p)), \end{aligned}$$

the transition functions $\varphi_{ij} = \varphi_i^{-1} \cdot \varphi_j$ give a map

$$\begin{aligned} \varphi_{ij} : U_i \cap U_j &\longrightarrow G, \\ p :&\xrightarrow{\varphi_{ij}} \varphi_{ij}(p) \end{aligned}$$

If

$$s_i : U_i \to \pi^{-1}(U_i),$$
$$s_j : U_j \to \pi^{-1}(U_j),$$

are sections of the bundle over sets with $U_i \cap U_j \neq \varnothing$, then on the overlap the sections are related by

$$s_j(p) = s_i(p)\varphi_{ij}(p). \tag{9.9}$$

9.3.1 Definition Let $\xi = \{E, M, \pi, F, G\}$ be a principal fiber bundle and let $b \in F_p$ be a point on the bundle over the fiber F_p. A tangent vector $Y \in T_bE$ is called *vertical* if

$$\pi_* Y = 0.$$

The vector space V_b of all such vectors is called the *vertical subspace* of T_bE. If $X \in \mathfrak{g}$, that action of the group G on E induces a fundamental vector field $Y = \sigma(X)$ as defined by equation 7.74. Such a vector field would then yield a vertical vector at any point $b \in F_p$.

9.3.1 Ehresmann Connection

We now introduce the following,

9.3.2 Definition Ehresmann connection
A *connection* Γ on a principal fiber bundle is a choice of a subspace H_b of T_bE, such that

a) For each $b \in E$, we have $T_bE = V_b \oplus H_b$,
b) $R_{g*}H_b = H_{bg}$,
c) H_b is a C^∞ distribution.

The vector space H_b is called the *horizontal subspace* of T_bE at b, and tangent vectors in this space are called *horizontal.* Condition (a) says that any tangent vector $Y \in T_bE$ can be split as a sum of a vertical and a horizontal component that we denote as,

$$Y = vY + hY.$$

Condition (b) says that the distributions $b \to H_b$ is right-invariant under the action of G . A connection on a principal fiber bundle as defined above is called an *Ehresmann connection.*
Given a connection Γ we define a Lie-Algebra valued one form

$$\omega_b : T_bE \to \mathfrak{g}$$

that for each tangent vector $Y \in T_bE$, it assigns the unique vector vector $X \in \mathfrak{g}$ whose fundamental vector field $\sigma(X)$ is the vertical component of Y. In the language of distributions, H_b is the kernel of the map,

$$H_b = \{Y \in T_bE \ : \ \omega(Y) = 0.\}$$

Since the map ω is onto for every $b \in E$, the kernel H_b is a linear subspace of T_bE with dimension equal to the dimension of M with $\omega(Y_H) = 0$. To be clear, we are saying,

$$\omega(Y) = \begin{cases} X & \text{if } Y = \sigma(X), \\ 0 & \text{if } Y \text{ is horizontal.} \end{cases}$$

Here we assume that the space H_b annihilated by ω is a C^∞ distribution in the sense of Frobenius 7.44.

Motivated by the formula for the pull-back of the Maurer-Cartan form 7.66 and by equation 7.75, we can equivalently characterize an Ehresmann connection, by the conditions stated in the following,

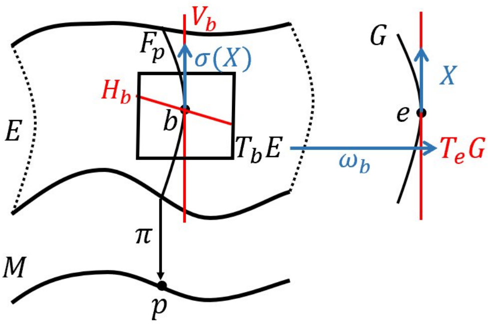


Fig. 9.5: Ehresmann Connection

9.3.3 Theorem An Ehresmann Connection on a principal fiber bundle $\xi = \{E, M, \pi, F, G\}$, is a smooth, Lie Algebra valued form $\omega \in \Omega(E)$ that satisfies the following conditions:

a) $\omega(\sigma(X)) = X$, for all $X \in \mathfrak{g}$,

b) $R_g^*\omega(Y) = Ad_{g^{-1}}\omega(Y)$, for all $g \in G$, and all tangent vectors Y on E.

Part (a) is immediate from the definition. Part (b) essentially follows from the fact that the fundamental vector field associated with $R_{g*}\omega$ is $Ad_{g^{-1}}\omega$. Y can be split uniquely into a horizontal and a vertical component. If Y is horizontal, then both sides of the equation are zero. If Y is vertical, it is the fundamental vector field $\sigma(X)$ for some $X \in \mathfrak{g}$. Then

$$\begin{aligned} (R_g^*\omega)_b(Y) &= \omega_{bg}(R_{g*}Y), \\ &= \omega_{bg}(\sigma(Ad_{g^{-1}}X)) \quad \text{by equation 7.75} \\ &= Ad_{g^{-1}}X \\ (R_g^*\omega)_b(Y) &= Ad_{g^{-1}}(\omega_b(Y)). \end{aligned} \tag{9.10}$$

We now show how to pull down a connection ω on a principal fiber bundle E down to a family of local connections on the manifold. Here we adapt the procedure from Kobayashi and Nomizu [18], with apologies to the authors for diminishing the elegance of their proof, for the sake of clarity provided by adding more details. Let $\{U_i\}$ be an open cover of M with coordinate charts $\{(\phi_i, U_i)\}$. For each i let $s_i(p)$ be the section over U_i defined by

$$s_i(p) = \phi_i^{-1}(p, e), \quad p \in U_i, \quad e = id \in G$$

Given a connection ω in the bundle E, define local connections on M using the pullback of the section maps. Thus, over each pair of overlapping charts $U_i \cap U_j \neq \varnothing$, we define,

$$\begin{aligned} \omega_i &= s_i^* \omega, \\ \omega_j &= s_j^* \omega. \end{aligned}$$

Since the transition functions map

$$\varphi_{ij} : U_i \cap U_j \to G,$$

we can pull-back forms. In particular, let θ be the left invariant Maurer-Cartan form on G. We define a $\mathfrak{g}$-valued form on $U_i \cap U_j$ by

$$\begin{aligned} \varphi_{ij}^* &: TG \to T(U_i \cap U_j), \\ \theta_{ij} &= \varphi_{ij}^* \theta. \end{aligned}$$

For applications to gauge theory, following result is very important,

9.3.4 Theorem On $U_i \cap U_j \neq \varnothing$, the local forms ω_i and θ_{ij} on M satisfy the condition

$$\omega_j = (Ad_{\varphi_{ij}^{-1}})\omega_i + \theta_{ij}. \tag{9.11}$$

Proof Given a point $p \in U_i \cap U_j$, let $X_p \in T_p(U_i \cap U_j)$ be a tangent vector to a curve $x(t)$, with $x(0) = p$ and $X = x'(t)$. Then the transition equation for the sections 9.9 reads

$$s_j(x(t)) = s_i(x(t))\varphi_{ij}(x(t)).$$

The push-forward

$$s_{j*}(X_p) : T_p(U_i \cap U_j) \to T_{s_j(p)E},$$

is the image of $(s_{i*}(X), \varphi_{ij*}(X))$ under the isomorphism

$$T_{s_i(p)}E \otimes T_{\varphi_{ij}(p)}G \cong T_{s_j(p)}E.$$

More specifically, taking the derivative d/dt and evaluating a $t = 0$ as done with the product rule formula 6.2, we get

$$\begin{aligned} \frac{d}{dt}[s_j(x(t))]_{t=0} &= \frac{d}{dt}[s_i(x(t))\ \varphi_{ij}(x(t))]_{t=0}, \\ &= \frac{d}{dt}[s_i(x(t))\ \varphi_{ij}(p))]_{t=0} + \frac{d}{dt}[s_i(p))\ \varphi_{ij}(x(t))]_{t=0}, \\ s_{j*}(X) &= \frac{d}{dt}[R_{\varphi_{ij}(p))}\ s_i(x(t))]_{t=0} + \frac{d}{dt}[s_i(p))\varphi_{ij}(x(t))]_{t=0}, \\ &= R_{\varphi_{ij}(p)*}\ s_{i*}(X) + s_{i*}(p)\ \varphi_{ij*}(X)). \end{aligned}$$

We now apply ω to both side remembering the general definition of the pullback $s^*\omega(X) = \omega(s_* X)$. We get

$$\begin{aligned} \omega(s_{j*}(X)) &= \omega(R_{\varphi_{ij}(p)*}\ s_{i*}(X)) + \omega(s_{i*}(p)\ \varphi_{ij*}(X)), \\ s_j^*\omega(X) &= R^*_{\varphi_{ij}(p)}\omega(s_{i*}(X)) + \omega(s_{i*}(p)\ \varphi_{ij*}(X)), \\ \omega_j(X) &= R^*_{\varphi_{ij}(p)}\omega_i(X)) + \omega(s_{i*}(p)\ \varphi_{ij*}(X)), \\ \omega_j(X) &= (Ad_{\varphi_{ij}^{-1}})\omega_i(X)) + \omega(s_{i*}(p)\ \varphi_{ij*}(X)), \end{aligned}$$

where in the first term on the right, we have used the condition 9.10 for an Ehresmann connection. The second term on the right is a bit trickier. We see from the diagram below

$$\begin{array}{ccc} T_p\ (U_i \cap U_j) & \xrightarrow{\varphi_{ij*}} & T_{\varphi_{ij}(p)}G \\ \downarrow & & \downarrow \\ p \in (U_i \cap U_j) & \xrightarrow{\varphi_{ij}} & \varphi_{ij}(p) \in G, \end{array}$$

that $\varphi_{ij}(x(t)$ is a curve in G, whose differential map sends the tangent vector X at p to a tangent vector in G at $\varphi_{ij}(p)$

$$X_p \xrightarrow{\varphi_{ij*}} \varphi_{ij*}(X)|_{\varphi_{ij}(p)}$$

On the other hand, one can think of $s_i(p)$ as a map from G to E given by

$$\begin{aligned} s_i(p) &: G \to E, \\ g &\xrightarrow{s_i(p)} s_i(p)g. \end{aligned}$$

Thus $s_{i*}(p)\ \varphi_{ij*}(X)$ is the push-forward of $\varphi_{ij*}(X)$ to TE by the Jacobian map $s_{i*}(p) : TG \to TE$. If $Y \in \mathfrak{g}$ is the left-invariant vector [2] in G such that $Y = \varphi_{ij*}(X)$ at $\varphi_{ij}(p)$, then, we have

$$\theta(Y) = \theta(\varphi_{ij*}X) = Y$$

The image of Y under $s_{i*}(p)$ corresponds to a fundamental vector $\sigma(Y)$, therefore, by the definition of an Ehresmann connection

$$\begin{aligned} \omega(s_{i*}(p)\ \varphi_{ij*}(X) &= \omega(\sigma(Y)), \\ &= Y, \\ &= \theta(\varphi_{ij*}(X)), \\ &= \varphi_{ij}^*\theta(X), \\ &= \theta_{ij}(X) \end{aligned}$$

[2]Specifically, $Y_e = L^{-1}_{\varphi_{ij}(p)*}(\varphi_{ij*}(X)_{\varphi_{ij}(p)})$. The notation is a bit cluttered but the concept is rather simple. The vector Y generates a one parameter subgroup of G whose tangent vector at $\varphi_{ij}(p)$ concides with $\varphi_{ij*}(X)$. The integral curve of Y induces a fundamental vector field $\sigma(Y)$ on the fiber F_p

The converse of the theorem is obtained by reversing the argument above. This concludes the proof.

If G is a matrix group, and on the overlap of the charts $U_i \cap U_j$ we denote the transition functions $\varphi_{ij}(p)$ by a matrix transformation B, then equation 9.11 reads

$$\omega_j = B^{-1}\omega_i B + B^{-1}dB,$$

which we immediately recognize as the transformation law for an affine connection in the manifold, (or a local gauge transformation in the language of physics.) Since this holds for any pair of overlapping patches, we see that the connection in the bundle gives rise to a family of connections on M which piece together as we desire on any overlap.

9.3.2 Horizontal Lift

Given a principal fiber bundle $\xi = \{E, \pi, M, G\}$, and $b \in E$, we define the *horizontal lift* $X_b^\sharp$ of a vector field $X \in \mathscr{X}(M)$ to be the a horizontal vector at p that projects to X; that is

a) $v(X_b^\sharp) = 0$,

b) $\pi_*(X_b^\sharp) = X_{\pi(b)}$.

The horizontal lift is right translation invariant meaning

c) $R_{g*}X_b^\sharp = X_{bg}^\sharp$ for all $b \in E$ and $g \in G$.

We have the following,

9.3.5 Proposition Let $X^\sharp$ and $Y^\sharp$ be horizontal lifts of X and Y respectively, and let $f \in \mathscr{F}(M)$. Denote by $f^\sharp$ the composition $f^\sharp : E \xrightarrow{\pi} M \xrightarrow{f} \mathbf{R}$. Then

a) $X^\sharp + Y^\sharp = (X+Y)^\sharp$,

b) $f^\sharp X^\sharp = (fX)^\sharp$,

c) $h[X^\sharp, Y^\sharp] = [X, Y]^\sharp$.

Proof Only part (c) requires a little thinking. The proof rests on the fact that the push-forward of the Lie bracket is equal the bracket of the push-forwards, as shown in equation 7.25. We have

$$\begin{aligned}
\pi_*(h[X^\sharp, Y^\sharp]) &= \pi_*([hX^\sharp, hY^\sharp], \\
&= \pi_*([X^\sharp, Y^\sharp]), \\
&= [\pi_* X^\sharp, \pi_* Y^\sharp], \\
&= [X, Y].
\end{aligned}$$

To give a better illustration of the horizontal lift of vector fields, consider the bundle of frames $E = B(M)$.

If ∇ is a connection on M, using the notion of parallelism defined by 6.64, we can parallel transport the tangent space along curves in M. Let $\{x^1, \ldots, x^n\}$ be coordinates in a coordinate chart about a point $p \in M$ and let $\{\partial_i = \frac{\partial}{\partial x^i}\}$ be the standard basis for T_pM. The horizontal lifts $\{(\partial_i)^\sharp\}$ then constitute a basis for the distribution $b \mapsto H_b$. Let $b = (p, e_1, \ldots e_n) \in B(M)$, with

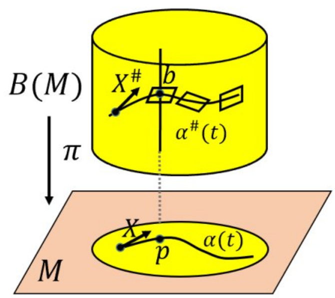
$e_j = A^i{}_j\partial_i,\ A \in Gl(n, \mathbf{R})$, and $\alpha(t)$ be a curve in M with $\alpha(0) = p$. By parallel translation $\{e_i(t) = e_i|_{\alpha(t)}\}$ of the frame $\{e_i\}|_p$, we define a curve

$$\alpha^\sharp(t) = (\alpha(t), e_1(t), \dots e_n(t)).$$

Since $(\pi \circ \alpha^\sharp)(t) = \alpha(t)$, we get a horizontal lift $\alpha^\sharp(t)$ of $\alpha(t)$. Thus a connection on M allows us to get unique horizontal lifts of curves in M. A connection on the frame bundle, together with the notion of horizontal lifts of vector fields, provide a way to naturally lift curves on M to the bundle. Let $\alpha(t)$ be a curve in M with tangent vector field T in a neighborhood U where α is injective. Lift T horizontally to $T^\sharp$, and let $\alpha^\sharp$ be the integral curve in $B(M)$. If $T^\sharp_b$ is horizontal, that is $T^\sharp_b \in H_b$, then by the properties of the bundle connection, $R_{g*}T^\sharp_b = T^\sharp_{pg}$ is also horizontal. Thus, The curve $\alpha^\sharp$ is horizontal independent of the point b a $t = 0$. The horizontal lift defines a parallel transport on the manifold. The idea extends to any principal fiber bundle. For a more careful treatment, see for example, Kobayashi and Nomizu [18] or Spivak [34].

9.3.3 Curvature Form

Returning the concept of a connection on a principal fiber bundle $\xi = \{E, \pi, M, G\}$, and $\xi = \{E, \pi, M, G\}$, with $b \in E$, we introduce the following .



9.3.6 Definition A form ϕ of degree k in E is called a *tensorial form of adjoint type* if

$$R^*_g\phi = Ad_{g^{-1}} \cdot \phi, \quad \text{for all } g \in G,$$

and $\phi(X_1, \dots, X_k) = 0$ if at least one of the tangent vectors X_i in E os vertical. The $k + 1$ form $D\phi = (d\phi)h$, that is

$$D\phi(X_1, \dots, X_{k+1}) = d\phi(hX_1, \dots, hX_{k+1}),$$

is a tensorial form $D\phi$ called the *exterior covariant derivative* of ϕ.

Now we come the main result of this section. First, we will need,

9.3.7 Lemma If Y_1 is horizontal and $Y_2 = \sigma(X_2)$ is a fundamental vertical vector generated by X_2, then $[Y_1, Y_2]$ is horizontal.
Proof Let $\varphi_t = e^{tX_2}$ be the one-parameter subgroup in G generating Y_2 by right translation R_{ϖ_t} . Then

$$\begin{aligned}[Y_1, Y_2] &= -[Y_2, Y_1] = -\pounds_{Y_2} Y_1, \\ &= -\lim_{t \to 0} \frac{Y_1 - R_{\varphi_t *} Y_1}{t}.\end{aligned}$$

If Y_1 is horizontal, so is $R_{\varphi_t *}Y_1$, so the left hand side $[Y_1, Y_2]$ is also horizontal.

9.3.8 Theorem Let ω be a connection on E. Then the curvature form defined as $D\omega$ satisfies the structure equation

$$D\omega(Y_1, Y_2) = d\omega(Y_1, Y_2) + [\omega(Y_1), \omega(Y_2)]. \tag{9.12}$$

Proof Every vector in $Y \in T_E$ can be split into the sum of a vertical and a horizontal vector. Both sides of the structure equation are skew-symmetric and bilinear, so it suffices to treat the following three cases
Case 1. Y_1 and Y_2 are horizontal. Then $\omega(Y_1) = \omega(Y_2) = 0$, and $hY_1 = Y_1,\ hY_2 = y_2$. Inserting into equation 9.12, we get

$$D\omega(Y_1, Y_2) = d\omega Y_1, Y_2 = d\omega(hY_1, hY_2),$$

which is precisely the definition of $D\omega$.
Case 2. Y_1 and Y_2 are vertical. By definition, $\Omega(Y_1, Y_2) = 0$, thus we have to prove that the right hand side is also 0. Since Y_1, Y_2 are fundamental vector fields, there exist vectors $X_1, X_2 \in \mathfrak{g}$ such that $Y_1 = \sigma(X_1)$ and $Y_2 = \sigma(X_2)$. So $\omega(Y_1) = X_1$ and $\omega(Y_2) = X_2$ are constant. From the definition of the differential of a one-form 6.28, we have,

$$\begin{aligned} d\omega(Y_1, Y_2) &= Y_1(\omega(Y_2)) - Y_2(\omega(Y_1)) - \omega([Y_1, Y_2]), \\ &= -\omega([Y_1, Y_2]) = -\omega([\sigma(X_1), \sigma(X_2)]), \\ &= -\omega(\sigma[X_1, X_2]), \quad \text{by theorem 7.3.4}, \\ &= -[X_1, X_2] = -[\omega(Y_1), \omega(Y_2)]. \end{aligned}$$

Thus,

$$d\omega(Y_1, Y_2) + [\omega(Y_1), \omega(Y_2)] = 0.$$

Case 3. Y_1 is horizontal and Y_2 is vertical. By definition $\Omega(Y_1, Y_2) = 0$, so we have to show that right hand side is also 0. Extend Y_1 to a horizontal vector field, and let $X_2 \in \mathfrak{g}$ be the vector generating $Y_2 = \sigma(X_2)$. Then as in case 2, $\omega(Y_2) = \omega(\sigma(X_2))$ is constant, so $Y_1(\omega(Y_2)) = 0$ and $[\omega(Y_1), \omega(Y_2)] = 0$. It remains to show that $d\omega(Y_1, Y_2) = 0$, We have,

$$\begin{aligned} d\omega(Y_1, Y_2) &= Y_1(\omega(Y_2)) - Y_2(\omega(Y_1)) - \omega([Y_1, Y_2]), \\ &= -\omega([Y_1, Y_2]), \\ &= 0. \qquad \text{by lemma 9.3.7}. \end{aligned}$$

9.4 Gauge Fields

As described in the historical notes earlier, physicists and mathematicians developed the notion of a connection on a principal fiber bundle independently, and it wasn't until the 1970's that they realized that they were talking about the same objects. Here is short lexicon of the corresponding terms used in the two disciplines

Mathematics	Physics
Principal fiber bundle	Gauge space
G structure group	Gauge group (such as $SU(2)$)
Connection form	Gauge potential (such as $\mathbf{A}$)
Curvature form	Field strength (such as $\mathbf{E}$ and $\mathbf{B}$)
Local trivialization	Choice of gauge
Transition function	Change of gauge

To get to the physical significance of the principal fiber bundle formalism, let ω be a connection on the PFB, with curvature $D\omega$. Assuming the structure group has dimension k, let $\{e_1, \ldots, e_k\}$ be a basis for the Lie algebra $\mathfrak{g}$. Then we can write the components of the connection as $\omega = \omega^\alpha e_\alpha$, and the structure equation 9.12 reads,

$$\Omega^\alpha = d\omega^\alpha + \frac{1}{2} C^\alpha_{\beta\gamma} \omega^\beta \wedge \omega^{\gamma\cdot}$$

The α's in the forms Ω^α and ω^α are Lie algebra indices which reflect the fact that the forms are Lie algebra valued. The reader will of course note the similarity to the Maurer-Cartan equations 7.68. If we pick a local trivialization $\{U, \varphi\}$ with local section $s : U \to E$, and label the local forms

$$A = s^*\omega, \quad F = s^*\Omega,$$

we get the expression

$$F^\alpha = dA^\alpha + \frac{1}{2} C^\alpha_{\beta\gamma} A^\beta \wedge A^{\gamma\cdot}$$

Better yet, if the local coordinates of the manifold be denoted by $\{x^\mu\}$, we can write the equation above to include the tensor indices,

$$F^\alpha_{\mu\nu} = \frac{A^\alpha_\mu}{\partial x^\nu} - \frac{A^\alpha_\nu}{\partial x^\mu} + \frac{1}{2} C^\alpha_{\beta\gamma} A^\beta_\mu \wedge A^\gamma_\nu, \tag{9.13}$$

which is the familiar form encountered in the physics of Yang-Mills fields. On the non-empty overlap of two coordinate charts, ω and Ω transform as connection and a tensorial form should.

9.4.1 Electrodynamics

We take a closer look at the special case of electrodynamics. In the classical theory of electromagnetism, we find the simplest example of a gauge theory. If F is the Maxwell 2-form, then as in 2.115, we have $dF = 0$. Therefore, by the Poincaré lemma, in a simply connected region, there exist a one form A such that $F = dA$. In tensor components, this reads

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu.$$

The one form $A = A_\mu \, dx^\mu$ is not unique because the transformation

$$A \mapsto A' = A + d\varphi$$

leaves the strength field F invariant. In vector notation, $A_\mu = (\phi, \mathbf{A})$, the gauge freedom reads

$$\mathbf{A}' = \mathbf{A} + \nabla\varphi,$$
$$\varphi = \phi - \frac{\partial\phi}{\partial t},$$

and the corresponding fields $\mathbf{E}$ and $\mathbf{B}$ remain invariant. Thus, one can solve the dynamic equations working with A, knowing that the observables are gauge independent. The gauge freedom of the electromagnetic field is an asset, rather than a liability, for it allows one to adjust the potentials to have properties that do not affect the fields. Of these, perhaps the most useful is the Lorentz gauge $\partial_\mu A^\mu = 0$ that leads to the wave equation

$$\Box A^\mu = J^\mu$$

for the potential, and the polarization states of electromagnetic waves. For time dependent fields, the solutions are called the retarded potentials [17]

$$\phi(t, \mathbf{r}) = \frac{1}{4\pi}\int \frac{\rho(t_r, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d^3\mathbf{r}',$$
$$\mathbf{A}(t, \mathbf{r}) = \frac{1}{4\pi}\int \frac{\mathbf{J}(t_r, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d^3\mathbf{r}',$$

where,

$$t_r = t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}.$$

The gauge group is probably more evident in quantum electrodynamics (QED). From equation 2.123, the classical electromagnetic Lagrangian is

$$L_{EM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + J^\mu A_\mu.$$

As shown there, the Euler-Lagrange equations lead to Maxwell equations. The Dirac equation 8.81 (with $\hbar = c = 1$) for electron/positron fields

$$(i\gamma^\mu\partial_\mu - m)\Psi = 0$$

is generated by the Dirac Lagrangian for fermion fields of spin $\frac{1}{2}$ and mass m,

$$L_D = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi. \tag{9.14}$$

The Dirac Lagrangian is invariant under the phase transformations

$$\psi(x) \mapsto \psi'(x) = e^{ie\lambda}\psi(x), \quad \overline{\psi}(x) \mapsto \overline{\psi}'(x) = e^{-ie\lambda}\psi(x).$$

This is called an internal global symmetry, where global refers to the symmetry being independent of the position, and internal to the symmetry not changing the location. Since $e^{ie\lambda}$ is unimodular, the gauge group is $U(1)$. On the other hand, if we want to impose a local symmetry by letting $\lambda(x)$ depend on x, the Dirac Lagrangian transforms as

$$L_D \mapsto L'_D = \overline{\psi}[i\gamma^\mu(\partial_\mu - ie\partial_\mu\lambda) - m]\psi.$$

To make the new Lagrangian invariant requires the introduction of a covariant derivative operator

$$\nabla_\mu = \partial_\mu + ieA_\mu, \tag{9.15}$$

with a corresponding modification to the Lagrangian

$$L_D = \overline{\psi}(i\gamma^\mu\nabla_\mu - m)\psi. \tag{9.16}$$

Finally, if we add the electromagnetic Lagrangian, we get the full QED Lagrangian

$$L_{QED} = \overline{\psi}(i\gamma^\mu\nabla_\mu - m)\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{9.17}$$

Thus, local invariance leads to a coupling with the electromagnetic potential A^μ which now can evidently be interpreted as connection on a $U(1)$ bundle, thus providing a mechanism for covariant derivative along the sections of the bundle. The QED lagrangian can also be written to elucidate better the coupling to E& M, as

$$L_{QED} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + A_\mu J^\mu, \tag{9.18}$$

where $J^\mu = e\overline{\psi}\gamma^\mu\psi$. If the structure group is replaced by $G = U(1) \times SU(2)$ we get the Weinberg-Salam standard model; if the group is enlarged to $G = SU(3)$, we get Quantum Chromodynamics (QCD). In either case the Lagrangian requires only a modification for the curvature form F to have an extra index to indicate that it is Lie algebra valued. Thus, the QCD Lagrangian is

$$L_{QED} = \overline{\psi}(i\gamma^\mu\nabla_\mu - m)\psi - \tfrac{1}{4}F^\alpha_{\mu\nu}F_\alpha^{\mu\nu}.. \tag{9.19}$$

As before, the field strength F is the curvature of the connection

$$F = DA = dA + +ie\,A \wedge A,$$

but compared electromagnetism, the wedge/bracket makes this a non-Abelian gauge theory.

9.4.2 Dirac Monopole

There are no magnetic monopoles, but if there were, we would like the fields to satisfy an extended Maxwell equation

$$\nabla \cdot \mathbf{B} = 4\pi\rho_m,$$

where $\rho_m = g\delta(\mathbf{r})$ is the point density of magnetic charge. Then, the solution is a $1/r^2$ law

$$\mathbf{B} = g\frac{\mathbf{r}}{r^3}$$

Let F be the electromagnetic 2-form for a pure magnetic field. By Stokes' theorem the flux over a closed surface R bounding a volume V, is

$$\Phi_R = \int_R F = \int_R \mathbf{B} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{B}\, dV,$$

where,

$$F = \mathbf{B} \cdot d\mathbf{S} = \frac{g}{r^3}(x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy).$$

If we constrain F to a 2-sphere centered at the origin and convert to spherical coordinates, the form F simplifies to

$$F = g \sin\theta\ d\theta \wedge d\phi.$$

There is of course no globally defined potential for F because that would imply that $dF = 0$ and that is no longer true. Still, we seek local forms A with $dA = F$. Since up to the constant factor g, F is the curvature form for a sphere, as shown in example 4.5.9, the natural candidate are the components of the Cartan connection form

$$A_{trial} = -g \cos\theta\ d\phi$$

Unfortunately the potential has singularities that might not be apparent in spherical coordinates, but become evident in Cartesian coordinates

$$A_{trial} = -g\frac{z}{r^3}\frac{x\,dy - y\,dx}{x^2 + y^2}$$

This is the same problematic form 2.82 which we noted as a standard counterexample in the discussion of the Poincaré lemma. The form is singular along $x^2 + y^2 = 0$. In Dirac's original construction of the monopole solution, he allowed for the singularities by essentially cutting of the lower z-axis, a set usually called a Dirac string. The modern approach circumvents the singularities by constructing a connection on a Hopf bundle. Let (z^1, z^2) be coordinates on $\mathbf{C}^2$, with

$$z^1 = x^1 + ix^2, \quad z^2 = x^3 + ix^4.$$

and define $\mathbf{CP}^1$ as in section 8.1.4 by the quotient $\mathbf{C}/\sim$ with the equivalence class

$$(z^1, z^2) \sim (\lambda z^1, \lambda z^2),\ \lambda \in \mathbf{C},$$

and constraining to the three sphere $S^3 : |z^1|^2 + |z^2|^2 = 1$. Following the convention in equation 5.60, let

$$\zeta_1 = \frac{x + iy}{1 - z}$$

be the complex number associated with point $p(x, y, z) \in S^2$ under the stereographic projection from the north pole. This gives a coordinate chart for S^2, but the chart misses the north pole. To cover the sphere, we create another chart by a stereographic projection from the south pole. By the same process of ratio and proportions for similar right triangles, we find that the complex number ζ_2 that represents the same point in the sphere is given by

$$\zeta_2 = \frac{x - iy}{1 + z}$$

The minus sign in the y coordinate is needed to preserve the orientation of the coordinate axes. The chart based on the south pole maps the north pole to 0 and the south pole to ∞, so we should expect the change of variables to behave like the conformal inversion $f(z) = 1/z$ in the complex plane. Not surprisingly this is exactly what we get,

$$\begin{aligned}\frac{1}{\zeta_1} &= \frac{\bar{\zeta}_1}{\zeta_1\bar{\zeta}_1},\\ &= (x - iy)\frac{x^2 + y^2}{(1 - z)^2},\\ &= (x - iy)\frac{1 - z^2}{(1 - z)^2},\\ &= \frac{x - iy}{1 + z} = \zeta_2.\end{aligned}$$

Thus, we can form a cover of S^2 by two charts $\{U_1, \zeta_1\}$ and $\{U_2, \zeta_2\}$ that overlap on an infinitesimal neighborhood of the equator,

$$\begin{aligned}U_1 &= \{(\theta, \phi) : \tfrac{\pi}{2} - \epsilon < \theta \leq \pi\},\\ U_2 &= \{(\theta, \phi) : 0 \leq \theta < \tfrac{\pi}{2} + \epsilon.\},\end{aligned}$$

where in the overlap, $\zeta_2 = 1/\zeta_1$. If one sets $\zeta_1 = z^1/z^2$, the Hopf fibration $S^3 \xrightarrow{\pi} S^2$ is obtained by associating a point $\pi(z^1, z^2)$ on S^2 with the inverse image of the stereographic projection $\pi_s^{-1}(z^1/z^2)$ of S^2 to $\mathbf{C}$. It is unavoidable to have a minor index inconsistency in the chart labels in the sense that ζ_1 is associated with the projection from the north pole, but the coordinate chart U_1 is about the south pole. The bundle charts of $S^3 \xrightarrow{\pi} S^2$ are given by

$$\begin{aligned}\phi_1 &: \pi^{-1}(U_1) \to U_1 \times U(1), \quad \phi(z^1, z^2) = (\zeta_1, \tfrac{z^1}{|z^1|}),\\ \phi_2 &: \pi^{-1}(U_2) \to U_2 \times U(1), \quad \phi(z^1, z^2) = (\zeta_2, \tfrac{z^2}{|z^2|}),\end{aligned}$$

We can now perform the flux integrals on the two overlapping hemispheres from the poles to an angle θ

$$\begin{aligned}\Phi_1 &= \int\!\!\int F = -2\pi g(1 + \cos\theta),\\ \Phi_2 &= \int\!\!\int F = \ \ 2\pi g(1 - \cos\theta).\end{aligned}$$

Using Stoke's theorem, $\int \mathbf{A} \cdot d\mathbf{r} = \int\int F$ and using the symmetry around a parallel circle C on S^2 at fixed angle θ, we can set $\mathbf{A} = A_\phi e_\phi$. The line integrals yield the components of the two vector field potentials with corresponding local connections

$$\begin{aligned}
(A_1)_\phi &= -\frac{g(1+\cos\theta)}{r\sin\theta}, \quad \text{or} \quad A_1 = -g(1+\cos\theta)\, d\phi, \\
(A_2)_\phi &= \frac{g(1-\cos\theta)}{r\sin\theta}, \quad \text{or} \quad A_2 = \ \ g(1-\cos\theta)\, d\phi.
\end{aligned} \tag{9.20}$$

As before, the singularity structure of the connections is more evident in Cartesian coordinates. Multiplying and dividing equations 9.20 by r, we get

$$\begin{aligned}
A_1 &= -g\frac{1}{r}(r + r\cos\theta)d\phi, \\
&= -g\frac{1}{r}(r+z)\frac{xdy - ydx}{x^2+y^2}, \\
&= -g\frac{r+z}{r(r^2-z^2)}(xdy - ydx), \\
&= -g\frac{1}{r(r-z)}(xdy - ydx)
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
A_2 &= g\frac{1}{r}(r - r\cos\theta)d\phi, \\
&= g\frac{1}{r}(r-z)\frac{xdy - ydx}{x^2+y^2}, \\
&= g\frac{r-z}{r(r^2-z^2)}(xdy - ydx), \\
&= g\frac{1}{r(r+z)}(xdy - ydx)
\end{aligned}$$

Now we construct a twisted $U(1)$ principal fiber bundle with local trivializations $\pi^{-1}(U_1)$ and $\pi^{-1}(U_2)$ with transition functions ϕ_{12} on the overlap given by

$$\phi_{12} : \zeta_1 \to \left[\frac{\zeta_2}{\zeta_1}\right]^n, \ n \in \mathbf{Z}$$

Here, the transition functions are unimodular, so they can be written as $e^{in\phi} \in U(1)$. The number n is required to be an integer to insure smooth gluing on the overlap. If $n = 0$ we get a trivial bundle $S^2 \times S^1$. If $n = 1$ we get the standard Hopf bundle. If s_1 and s_2 are sections over U_1 and U_2 respectively, we get an Ehresmann connecton on the bundle with

$$A_1 = s_1^*\omega, \quad \text{and} \quad A_2 = s_2^*\omega.$$

On the overlap, the connection transformation

$$A_2 = \phi_{12}^{-1} A_1 \phi_{12} + \phi_{12}^{-1} d\phi_{12},$$

just reads

$$A_2 - A_1 = 2gd\phi.$$

As ϕ goes around the equator, we must have

$$\int_0^{2\pi} 2g\phi\, d\phi = 2n\pi,$$

which means that

$$2g = n \in \mathbf{Z}$$

This is Dirac's quantization condition for magnetic monopoles. The integer n corresponds to the first Chern c_1 class of the bundle. If in addition the particle has an electric charge, the wave function must satisfy Shcrödinger's equation

$$\frac{1}{2m}(\mathbf{p}-\frac{\mathbf{e}}{\mathbf{c}}\mathbf{A})^{\mathbf{2}}|\psi\rangle = \mathbf{E}|\psi\rangle.$$

Under the gauge invariance $\mathbf{A} \to \mathbf{A} + \nabla\lambda$, given wave functions $|\psi_1\rangle$ and $|\psi_2\rangle$ on U_1 and U_2, they must transform as

$$|\psi_1\rangle = e^{\frac{ie\lambda}{\hbar c}}|\psi_2\rangle, \quad \lambda = 2g\phi$$

where we have restored $\hbar$ and c to standard units. For a fixed value of θ, as the wave functions go from 0 to 2π, the requirement that $|\psi\rangle$ be single-valued implies that

$$\frac{2eg}{\hbar c} = n.$$

That is, for a singly charged monopole ($n = 1$), the ratio of the magnetic to the electric charge is given in terms of the fine structure constant

$$\frac{g}{e} = \frac{1}{2}\frac{\hbar c}{e^2} = \frac{137}{2} \simeq 69; \tag{9.21}$$

an amusing result which I first learned from professor Raymond Sachs in a problem assigned in my upper division course in E &M. The magnetic monopole is often called a topological charge, for it essentially arises from the classification of the transition functions with are basically continuous functions from the equator S^1 to $U(1)$ and hence, they correspond the fundamental homotopy group $\pi_1(U_1) = \mathbf{Z}$

It is worthwhile to view the monopole connection via the complex structure of the base space. Observe that other than a "Lie algebra" factor of i, the Dirac monopole 2-form is given by the Kähler form 5.65

$$F = ig\,\frac{d\zeta \wedge d\bar{\zeta}}{(1+\zeta\bar{\zeta})^2}.$$

We thus expect to have a complex gauge complex potential that leads to this 2-form. First, we use of Euler angles

$$z_1 = e^{i(\psi+\phi)/2}\cos\tfrac{\theta}{2}, \tag{9.22}$$

$$z_2 = e^{i(\psi-\phi)/2}\sin\tfrac{\theta}{2}, \tag{9.23}$$

to write the equation $|z_1|^2 + |z_2|^2 = 1$ of the three sphere S^3. The induced Riemannian metric on S^3 is given

$$\begin{aligned} ds^2 &= 4(d\bar{z}_1 dz_1 + d\bar{z}_2 dz_2),\\ &= d\theta^2 + \sin^2\theta\, d\phi^2 + (d\psi + \cos\theta\, d\phi)^2. \end{aligned}$$

The form

$$\omega = d\psi + \cos\theta\, d\phi$$

defines a connection on S^3 viewed as an S^1 bundle over S^2. This natural a connection in complex terms, is induced by the restriction of the form

$$\bar{z}^1 dz^1 + \bar{z}^2 dz^2$$

from $\mathbf{C}^2$ to S^3. The curvature of this connection

$$\begin{aligned}\Omega &= \sin\theta \; d\phi \wedge d\theta, \\ &= i\frac{d\zeta \wedge d\bar{\zeta}}{(1+\zeta\bar{\zeta})^2}.\end{aligned}$$

extended to Minkowski space, corresponds to a magnetic monopole of field strength 1. By a short computation, we obtain the real and imaginary parts of this form. We get

$$\begin{aligned}\Re\,(\bar{z}^1 dz^1 + \bar{z}^2 dz^2) &= x^1 dx^1 + x^2 dx^2 + x^3 dx^3 + x^4 dx^4, \\ &= \tfrac{1}{2} d((x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2) \\ &= 0. \qquad\qquad \text{on} \quad S^3, \\ \Im\,(\bar{z}^1 dz^1 + \bar{z}^2 dz^2) &= -x^2 dx^1 + x^1 dx^2 - x^4 dx^3 + x^4 dx^3.\end{aligned}$$

Thus we set

$$\omega = i(-x^2 dx^1 + x^1 dx^2 - x^4 dx^3 + x^4 dx^3).$$

We verify our assertion by looking at the pullback of ω for the sections s_1 and s_2 under de bundle trivialization constructed of hemispheres of S^2 overlapping over an infinitesimal band around the equator. We leave to the reader to verify that

$$A_1 = s_1^*(\omega) = i\,\Im\left(\frac{\bar{\zeta}\,d\zeta}{1+\zeta\bar{\zeta}}\right).$$

Recall that

$$\zeta_1 = \cot(\tfrac{\theta}{2})\, e^{i\phi}.$$

Then, as in the computation leading to the Fubini-Study metric 5.63, we have

$$\begin{aligned}\bar{\zeta} &= \cot(\tfrac{\theta}{2})\, e^{-i\phi}, \\ d\zeta &= -\tfrac{1}{2}\csc^2(\tfrac{\theta}{2})\, e^{i\theta}\, d\theta + i\cot(\tfrac{\theta}{2})\, e^{i\phi}\, d\phi \\ \Im\,(\bar{\zeta}\, d\zeta) &= \cot^2(\tfrac{\theta}{2})\, d\phi, \\ &= \frac{1+\cos\theta}{1-\cos\theta}\, d\phi.\end{aligned}$$

On the other hand

$$1 + \zeta\bar{\zeta} = 1 + \cot^2(\tfrac{\theta}{2}) = \csc^2(\tfrac{\theta}{2}) = \frac{2}{1-\cos\theta}$$

Combining these results together, we get

$$A_1 = ig(1+\cos\theta)\; d\phi$$

For the chart based on the stereographic projection from the South pole, we use

$$\zeta_2 = \tan(\tfrac{\theta}{2})\, e^{-i\phi}.$$

A completely analogous computation gives the local gauge potential

$$A_2 = -ig(1 - \cos\theta)\, d\phi$$

9.4.3 BPST Instanton

The complex version of the Dirac monopole introduced in last part of the section above can be extrapolated to the quaternionic projective space $\mathbf{HP}^1$. Let $(q_1, q_2) \in \mathbf{H}^2$, and define the equivalence relation

$$(q_1, q_2) \sim (\lambda q_1, \lambda q_2), \quad \lambda \in \mathbf{H}$$

The quaternionic projective space $\mathbf{HP}^1$ is defined by the quotient $\mathbf{H}^2/\sim$ with this equivalence relation. Effectively, the projective space is the space of quaternionic lines through the origin. The restriction

$$|q_1|^2 + |q_2|^2 = 1$$

defines a unit sphere S^7 centered at the origin. Extending the crude visualization shown in 8.2 for the complex projective space, the intersection of quaternionic lines with S^7 yield three spheres S^3. The restriction implies that $\lambda \in \mathbf{H}$ is a unit quaternion, and the set of unit quaternions is the unitary group

$$Sp(1) = U(1, \mathbf{H}) \simeq SU(2)$$

This leads to the quaternionic Hopf bundle

$$S^3 \hookrightarrow S^7 \xrightarrow{\pi} S^4 \cong \mathbf{HP}^1.$$

There is a natural $\mathfrak{sp}(1)$-valued connection on the bundle defined by

$$\omega = \Im\,(\bar{q}^1 dq^1 + \bar{q}^2 dq^2)$$

where again, we neglect the real part, since that vanishes on S^7. Using the stereographic projection from the North and South poles respectively, cover S^4 by two quaternionic charts $\{U_1, \zeta_1\}$, and $\{U_2, \zeta_2\}$, which overlap on a narrow band around the S^3 equator. Of course, in physics the preferred parametrization of this S^3 is by using Euler angles, and the action of $SU(2)$ on the bundle charts are by right multiplication with Euler angle matrices Q. On the overlap in the base space, the transition functions are

$$\phi_{12} = \left(\frac{\zeta_1}{\zeta_2}\right)^n.$$

The BPST instanton bundle corresponds to $n = 1$ If $s : U \to \pi^{-1}(U)$ is a section of the bundle over one of these charts, we get a connection

$$A = s^*\omega = \Im\left(\frac{\bar{q}\, dq}{1 + q\bar{q}}\right) \tag{9.24}$$

This is the gauge potential for the famous *BPST instanton.* The pullback F

$$F = s^*\Omega = dA + A \wedge A$$

of the curvature Ω in the bundle, represents the field strength. Since the chart is locally Euclidean through the stereographic projection, we may think F as a connection on $\mathbf{R}^4$. On $\mathbf{R}^4$, the connection is anti-self-dual in the sense of the Hodge $\star$ map,

$$\star F = -F$$

References

[1] H. Abrams 1971 *The World of M. C. Escher*, Meulenhoff International, Netherlands (1971)

[2] Atiyah, M., Hitchin, N., and Singer, I. 1978 *Self-duality in four-dimensional Riemannian geometry*, Proc. R. Soc. Lond. A. 362, (1978) pp 425-461.

[3] Belavin, A., Polyakov, A., Schwartz, A., and Tyupkin, Y. 1975 *Pseudoparticle Solutions of the Yang-Mills Equations*, Physics Letters B, Vol. 59, No, 1, (1975) pp 85-87.

[4] Chern, S. S. 1975: *Introduction to Differential Geometry.* Notes, University of California, Berkeley, CA.

[5] Chern, S. S. 1944 *A Simple Intrinsic Proof of the Gauss-Bonnet Formula for Closed Riemannian Manifolds*, Ann. Math. 45 (4), 1944, 747-752.

[6] Chern, S. S. 1955 *An Elementary Proof of the Existence of Isoterhmal Parameters on a Surface*, Proceedings of AMS, Vol. 6, No, 4, (1955), 771-782.

[7] Dray, T. 1984 1966 *The Relationship between Monopole Harmonics and Spin-weighted Spherical Harmonics*, Journal of Mathematical Physics, 26, 1030 (1985). https://doi.org/10.1063/1.526533

[8] Eguchi, T., Gilkey, P., and Hanson, A. 1980 *Gravitation, Gauge Theories and Differential Geometry*, Physics Reports, 66, No, 6 (1980). pp 213-293

[9] Georgi, H. 1999: *Lie Algebras in Particle Physics.* Frontiers in Physics V54, Westview (1999).

[10] Goldberg, J., Mcfarland, A., Newman, E., Rohrlich, F. and Sudarshan, C. 1966 *Spin-s Spherical Harmonics and* ð, Proceedings of AMS, Vol. 6, No, 4, (1955), 771-782.

[11] Goldstein, Herbert, 1950: *Classical Mechanics.* Addison-Wesley, 398pp.

[12] Grey, Alfred, 2006: *Modern Differential Geometry of Curves and Surfaces with Mathematica.* 3rd ed. Chapman & Hall, CRC, 977pp.

[13] Hall, B. 2015: *Lie Groups, Lie Algebras, and Representations.* 2nd ed. Springer, Graduate Texts in Mathematics 222, Switzerland, (2015).

[14] Hicks, Noel, 1965: *Notes on Differential Geometry.* Van Nostrand Reinhold, Princeton, NJ.

[15] Hirsh, A. 2002 *Extension of 'Villarceau-Section' to Surfaces of Revolution with a Generating Conic* , J. of Geometry and Graphics. Vol 6, (2002) No, 2, pp121-132.

[16] Hoffman, K. and Kunze, R., 1971: *Linear Algebra.* 2nd ed. Prentice-Hall, 407pp.

[17] J. D. Jackson, 1962: *Classical Electrodynamics.*

[18] Kobayashi, S and Nomizu, K 1963: *Foundations of Differential Geometry.* Wiley & Sons, New York, London, 1963.

[19] Eisenhart, L. P. 1960: *A Treatise on the Differential Geometry of Curves and Surfaces.* Ginn and Company, Boston, 1909. Reprint by Dover Publications, New York, 1960.

[20] Abraham, R. and Marsden, J. 1978: *Foundations of Mechanics.* 2nd ed. Addison-Wesley, 838pp.

[21] Misner, C. W., Thorne, K. S., and Wheeler, J. A. 1973: *Gravitation.* W.H. Freeman and Company, 1279pp.

[22] Morris, M.S. and Thorne, K.S., 1987: Wormholes in Spacetime and their use for Interstellar Travel: A Tool for Teaching General Relativity. *Am. J. Phys.*, **56**(5), pp395-412.

[23] Müller, Thomas, 2008: Exact geometric optics in a Morris-Thorne wormhole spacetime. *Phys. Rev. D*, **77**, 044043-1-044043-11.

[24] Newman, E. and Penrose, R. 1962. *An Approach to Gravitational Radiation by a Method of Spin Coefficients*, Journal of Mathematical Physics 3, 566 (1962), pp 556-578. https://doi.org/10.1063/1.1724257

[25] Newman, E. and Janis, R. 1962. *Note on the Kerr Spinning-Particle Metric*, Journal of Mathematical Physics 6, 915 (1965), pp 915-917. https://doi.org/10.1063/1.1704350

[26] O'Neill, Barrett, 2006: *Elementary Differential Geometry.* 2nd ed. Academic Press, 503pp.

[27] Oprea, J., 1997: *Differential Geometry and Its Applications.* Prentice-Hall, Englewood Cliffs, NJ.

[28] Oprea, J., 2000: *The Mathematics of Soap Films.* Student Mathematical, Library, Volume 10, AMS, Providence, RI.

[29] Penrose, R. 1987. *On the Origins of Twistor Theory.* Gravitation and Geometry, a volume on honour of I. Robinson, Biblipolis, Naples, 25pp.

[30] Penrose, R. 1976. *Nonlinear Gravitons and Curved Twistor Theory*, General Relativity and Gravitation 7 (1976), No. 1 pp 31-52. and Geometry, a volume on honour of I. Robinson, Biblipolis, Naples, 25pp.

[31] Rogers, C and Schief, W. K., 2002. *On Regular Solutions of Euclidean Yang-Mills Equations.* Physic Cambridge University Press.

[32] Schwarz, A, 1977. *Bäcklund and Darboux Transformations : Geometry and Modern Applications in Soliton Theory.* Physics Letters B, Vol. 67, No, 2, (1977) pp 172-174.

[33] Spivak, M. 1965: *Calculus on Manifolds.* Addison Wesley (1965), 159pp

[34] Spivak, M. 1979: *A comprehensive introduction to Differential Geometry 5 volume set.* 2nd ed. Publish or Perish, Texas.

[35] Struik, Dirk J., 1961: *Lectures on Classical Differential Geometry.* 2nd ed. Dover Publications, Inc., 232pp.

[36] Taub, A. H. 1939: *Tensor Equations Equivalent to the Dirac Equations.* Annals of Mathematics, Vol 40, Number 4, pp.937-947.

[37] Trautman, A. 1977: *Solutions of the Maxwell and Yang-Mills Equations Associated with Hopf Fibrings.* International Journal of Theoretical Physics, Vol 16, Number 8 (1977), pp.561-565.

[38] Terng,C and Uhlenbeck, K, 2000: *Geometry of Solitons.* Notices of the AMS, Vol 47, Number 1.

[39] Ward, R. S., 1977: *On Self-dual Gauge Fields.* Physics Letters A, Vol. 61, No, 2, (1977) pp 81-82.

[40] Weisstein, E. W. From Mathworld- A Wolfram Web Resource: [Available online at `http://mathworld.wolfram.com/CostaMinimalSurface.html`.]

Index

Differential Geometry in Physics is a treatment of the mathematical foundations of the theory of general relativity and gauge theory of quantum fields. The material is intended to help bridge the gap that often exists between theoretical physics and applied mathematics.

The approach is to carve an optimal path to learning this challenging field by appealing to the much more accessible theory of curves and surfaces. The transition from classical differential geometry as developed by Gauss, Riemann and other giants, to the modern approach, is facilitated by a very intuitive approach that sacrifices some mathematical rigor for the sake of understanding the physics. The book features numerous examples of beautiful curves and surfaces often reflected in nature, plus more advanced computations of trajectory of particles in black holes. Also embedded in the later chapters is a detailed description of the famous Dirac monopole and instantons.

Features of this book:

- Chapters 1-4 and chapter 5 comprise the content of a one-semester course taught by the author for many years.
- The material in the other chapters has served as the foundation for many master's thesis at University of North Carolina Wilmington for students seeking doctoral degrees.
- An open access ebook is available at Open UNC (openunc.org).
- The book contains over 80 illustrations, including a large array of surfaces related to the theory of soliton waves that does not commonly appear in standard mathematical texts on differential geometry.

GABRIEL LUGO is an associate professor in the Department of Mathematics and Statistics at the University of North Carolina Wilmington. He earned his undergraduate and Ph. D. degrees from the University of California, Berkeley. Lugo has a strong interest in instructional technology in collegiate education, and has done extensive work in that area with colleagues in the math, chemistry, and computer science departments.

UNCW®

UNIVERSITY *of* NORTH CAROLINA WILMINGTON

WILLIAM MADISON RANDALL LIBRARY

Distributed by UNC Press
www.uncpress.org