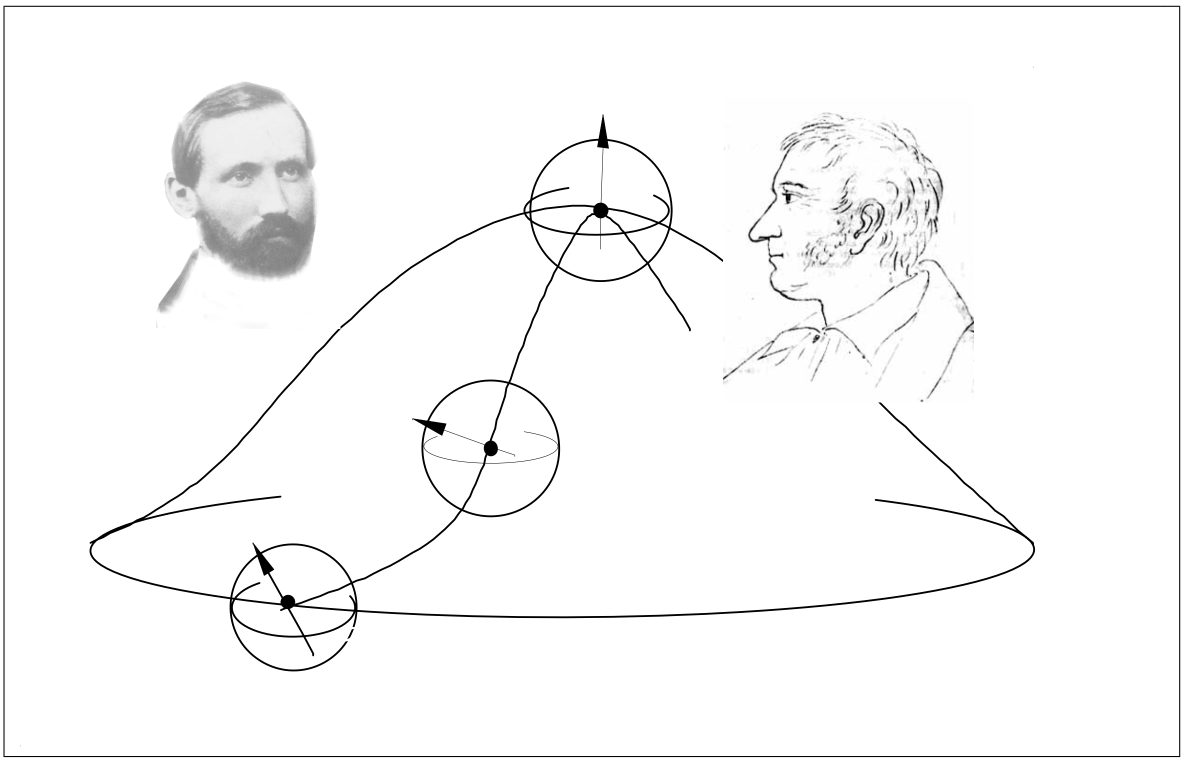
# Lectures on Differential Geometry

Wulf Rossmann



(Updated October 2003)

# To the student

This is a collection of lecture notes which I put together while teaching courses on manifolds, tensor analysis, and differential geometry. I offer them to you in the hope that they may help you, and to complement the lectures. The style is uneven, sometimes pedantic, sometimes sloppy, sometimes telegram style, sometimes long–winded, etc., depending on my mood when I was writing those particular lines. At least this set of notes is visibly finite. There are a great many meticulous and voluminous books written on the subject of these notes and there is no point of writing another one of that kind. After all, we are talking about some fairly old mathematics, still useful, even essential, as a tool and still fun, I think, at least some parts of it.

A comment about the nature of the subject (elementary differential geometry and tensor calculus) as presented in these notes. I see it as a natural continuation of analytic geometry and calculus. It provides some basic equipment, which is indispensable in many areas of mathematics (e.g. analysis, topology, differential equations, Lie groups) and physics (e.g. classical mechanics, general relativity, all kinds of field theories).

If you want to have another view of the subject you should by all means look around, but I suggest that you don't attempt to use other sources to straighten out problems you might have with the material here. It would probably take you much longer to familiarize yourself sufficiently with another book to get your question answered than to work out your problem on your own. Even though these notes are brief, they should be understandable to anybody who knows calculus and linear algebra to the extent usually seen in second-year courses. There are no difficult theorems here; it is rather a matter of providing a framework for various known concepts and theorems in a more general and more natural setting. Unfortunately, this requires a large number of definitions and constructions which may be hard to swallow and even harder to digest. (In this subject the definitions are much harder than the theorems.) In any case, just by randomly leafing through the notes you will see many complicated looking expressions. Don't be intimidated: this stuff is easy. When you looked at a calculus text for the first time in your life it probably looked complicated as well. Perhaps it will help to contemplate this piece of advice by Hermann Weyl from his classic *Raum–Zeit–Materie of* 1918 *(my translation)*. *Many will be horrified by the flood of formulas and indices which here drown the main idea* *of differential geometry (in spite of the author's honest effort for conceptual clarity). It is certainly regrettable that we have to enter into purely formal matters in such detail and give them so much space; but this cannot be avoided. Just as we have to spend laborious hours learning language and writing to freely express our thoughts, so the only way that we can lessen the burden of formulas here is to master the tool of tensor analysis to such a degree that we can turn to the real problems that concern us without being bothered by formal matters.*

W. R.

# Flow chart



Chapter 3 is independent of chapter 2 and is used only in section 4.3.

# Contents

# Chapter 1

# Manifolds

## 1.1 Review of linear algebra and calculus

**A. Linear algebra**. A *(real) vector space* is a set $V$ together with two operations, vector addition $u + v$ $(u, v \in V)$ and scalar multiplication $\alpha v$ $(\alpha \in \mathbb{R}, v \in V)$. These operations have to satisfy those axioms you know (and can find spelled out in your linear algebra text). Example: $\mathbb{R}^n$ is the vector space of real $n$–tuples $(x^1, \cdots, x^n)$, $x^i \in \mathbb{R}$ with componentwise vector addition and scalar multiplication. The basic fact is that *every vector space has a basis*, meaning a set of vectors $\{v_i\}$ so that any other vector $v$ can be written *uniquely* as a linear combination $\sum \alpha^i v_i$ of $v_i$'s. We shall always assume that our space $V$ is finite–dimensional, which means that it admits a finite basis, consisting of say $n$ elements. It that case any other basis has also $n$ elements and $n$ is called the *dimension* of $V$. For example, $\mathbb{R}^n$ comes equipped with a *standard basis* $e_1, \cdots, e_n$ characterized by the property that $(x^1, \cdots, x^n) = x^1 e_1 + \cdots + x^n e_n$. We may say that we can "identify" $V$ with $\mathbb{R}^n$ after we fix an ordered basis $\{v_1, \cdots, v_n\}$, since the $x \in V$ correspond one–to–one to their $n$–tuples of components $(x^1, \cdots, x^n) \in \mathbb{R}$. But note(!): this identification depends on the choice of the basis $\{v_i\}$ which "becomes" the standard basis $\{e_i\}$ of $\mathbb{R}^n$. The indices on $\alpha^i$ and $v_i$ are placed the way they are with the following rule in mind.

1.1.**1 Summation convention**. Any index occurring twice, once up, once down, is summed over. For example $x^i e_i = \sum_i x^i e_i = x^1 e_1 + \cdots + x^n e_n$. We may still keep the $\sum$'s if we want to remind ourselves of the summation.

A *linear transformation* (or *linear map*) $A : V \to W$ between two vector spaces is a map which respects the two operations, i.e. $A(u+v) = Au + Av$ and $A(\alpha v) = \alpha A v$. One often writes $Av$ instead of $A(v)$ for linear maps. In terms of a basis $\{v_1, \cdots, v_n\}$ for $V$ and $\{w_1, \cdots, w_m\}$ for $W$ this implies that $Av = \sum_{ij} \alpha^i a_i^j v_j$ if $v = \sum_i \alpha^i v_i$ for some indexed system of scalars $(a_i^j)$ called the *matrix of $A$ with respect to the bases* $\{v_i\}$, $\{w_j\}$. With the summation convention the equation $w = Av$ becomes $\beta^j = a_i^j \alpha^i$. Example: the matrix of the identity transformation $1 : V \to V$ (with respect to any basis) is the Kronecker delta $\delta^i_j$ defined by

$$\delta^i_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The *inverse* (if it exists) of a linear transformation $A$ with matrix $(a^j_i)$ is the linear transformation $B$ whose matrix $(b^j_i)$ satisfies

$$a^i_k b^k_j = \delta^i_j$$

If $A : V \to V$ is a linear transformation of $V$ into itself, then the *determinant* of $A$ is defined by the formula

$$\det(A) = \sum \epsilon_{i_1 \cdots i_n} a^{i_1}_1 \cdots a^{i_n}_n \tag{1}$$

where $\epsilon_{i_1 \cdots i_n} = \pm 1$ is the sign of the permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$. This seems to depend on the basis we use to write $A$ as a matrix $(a^i_j)$, but in fact it doesn't. Recall the following theorem.

1.1.**2 Theorem**. *$A : V \to V$ is invertible if and only if* $\det(A) \neq 0$. $\square$

There is a formula for $A^{-1}$ (Cramer's Formula) which says that

$$A^{-1} = \frac{1}{\det(A)}\, \tilde{A}, \qquad \tilde{a}^j_i = (-1)^{i+j} \det[a^l_k | kl \neq ji] \ .$$

The $ij$ entry $\tilde{a}^j_i$ of $\tilde{A}$ is called the $ji$ cofactor of $A$, as you can look up in your linear algebra text. This formula is rarely practical for the actual calculation of $A^{-1}$ for a particular $A$, but it is sometimes useful for theoretical considerations or for matrices with variable entries.

The *rank* of a linear transformation $A : V \to W$ is the dimension of the image of $A$, i.e. of

$$\mathrm{im}(A) = \{w \in W : w = Av \text{ for some } v \in V\}.$$

This rank is equal to maximal number of linearly independent columns of the matrix $(a^i_j)$, and equals maximal number of linearly independent rows as well. The linear map $A : V \to W$ is *surjective* (i.e. onto) iff[1] $\mathrm{rank}(A) = m$ and *injective* (i.e. one–to–one) iff $\mathrm{rank}(A) = n$. A *linear functional* on $V$ is a scalar–valued linear function $\varphi : V \to \mathbb{R}$. In terms of components with respect to a basis $\{v_1, \cdots, v_n\}$ we can write $v = \sum x^i v_i$ and $\varphi(v) = \sum \xi_i x^i$. For example, if we take $(\xi_1, \xi_2, \cdots, \xi_n) = (0, \cdots, 1 \cdots 0)$ with the 1 in the $i$–th position, then we get the linear functional $v \to x^i$ which picks out the $i$–th component of $v$ relative to the basis $\{v_1, \cdots, v_n\}$. This functional is denoted $v^i$ (index upstairs). Thus $v^i(\sum x^j v_j) = x^i$. This means that $v^i(v_j) = \delta^i_j$. The set of all linear functionals on $V$ is again a vector space, called the *dual space* to $V$ and denoted $V^*$. If $\{v_i\}$ is a basis for $V$, then $\{v^i\}$ is a basis for $V^*$, called the *dual basis*. If we write $v = \sum x^i v_i$ and $\varphi = \sum \xi_i v^i$ the $\varphi(v) = \sum \xi_i x^i$ as above. If $A : V \to W$ is a linear transformation we get a linear transformation $A^* : W^* \to V^*$ of the dual spaces in the reverse direction defined by$(A^*\psi)(v) = \psi(Av)$.$A^*$ is called the *transpose* of $A$. In terms of components with respect to bases $\{v_i\}$ for $V$

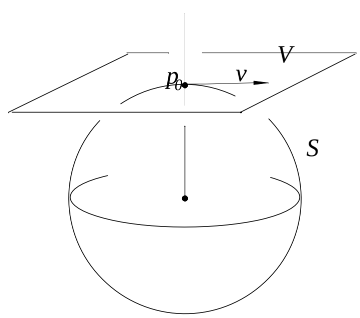
[1] "iff" means "if and only if".

and $\{w_j\}$ for $W$ we write $\psi = \eta_j w^j$, $v = x^i v_i$, $Av_i = a_i^j w_j$, $A^* w^j = (a^*)_i^j v^i$ and then the above equation reads

$$\sum (a^*)_i^j \eta_j x^i = \sum \eta_j a_i^j x^i.$$

From this it follows that $(a^*)_i^j = a_i^j$. So as long as you write everything in terms of components you never need to mention $(a^*)_i^j$ at all. (This may surprise you: in your linear algebra book you will see a definition which says that the transpose $[(a^*)_{ij})]$ of a matrix $[a_{ij}]$ satisfies $(a^*)_{ij} = a_{ji}$. What happened?)

1.1.**3 Examples**. The only example of a vector space we have seen so far is the space of $n$–tuples $\mathbb{R}^n$. In a way, this *is* the only example there is, since any $n$–dimensional vector space can be identified with $\mathbb{R}^n$ by means of a basis. But one must remember that this identification depends on the choice of a basis!
(a) The set $\mathrm{Hom}(V, W)$ consisting of all linear maps $A : V \to W$ between two vector spaces is again a vector space with the natural operations of addition and scalar multiplication. If we choose bases for $V$ and $W$ when we can identify $A$ with its matrix $(a_i^j)$ and so $\mathrm{Hom}(V, W)$ gets identified with the matrix space $\mathbb{R}^{m\times n}$.
(b) A function $U \times V \to \mathbb{R}$, $(u, v) \to f(u, v)$ is *bilinea*r is $f(u, v)$ is linear in $u$ and in $v$ separately. These functions form again a vector space. In terms of bases we can write $u = \xi^i u_i$, $v = \eta^j v_j$ and then $f(u, v) = c_{ij}\xi^i\eta^j$. Thus after choice of bases we may identify the space of such $B$'s again with $\mathbb{R}^{n\times m}$. We can also consider *multilinear* functions $U \times \cdots \times V \to \mathbb{R}$ of any finite number of vector variables, which are linear in each variable separately. Then we get $f(u, \cdots, v) = c_{i\cdots j}\xi^i \cdots \eta^j$. A similar construction applies to maps $U \times \cdots \times V \to W$ with values in another vector space.
(c) Let $S$ be a sphere in 3–space. The tangent space $V$ to $S$ at a point $p_o \in S$ consists of all vectors $v$ which are orthogonal to the radial line through $p_o$.




*

Fig. 1

It does not matter how you think of these vectors: whether geometrically, as arrows, which you can imagine as attached to $p_o$, or algebraically, as 3–tuples $(\xi, \eta, \zeta)$ in $\mathbb{R}^3$ satisfying $x_o\xi + y_o\eta + z_o\zeta = 0$; in any case, this tangent space is a 2–dimensional vector space associated to $S$ and $p_o$. But note: if you think of $V$ as the points of $\mathbb{R}^3$ lying in the plane $x_o\xi + y_o\eta + z_o\zeta$ =const. through $p_o$

tangential to $S$, then $V$ is not a vector space with the operations of addition and scalar multiplication in the surrounding space $\mathbb{R}^3$. One should rather think of $V$ as vector space of its own, and its embedding in $\mathbb{R}^3$ as something secondary, a point of view which one may also take of the sphere $S$ itself. You may remember a similar construction of the tangent space $V$ to any surface $S = \{p = (x, y, z) \mid f(x, y, z) = 0\}$ at a point $p_o$: it consists of all vectors orthogonal to the gradient of $f$ at $p_o$ and is defined only it this gradient is not zero. We shall return to this example in a more general context later.

**B. Differential calculus**. The essence of calculus is *local linear approximation.* What it means is this. Consider a map $f : \mathbb{R}^n \to \mathbb{R}^m$ and a point $x_o \in \mathbb{R}^n$ in its domain. $f$ admits a local linear approximation at $x_o$ if there is a *linear* map $A : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$f(x_o + v) = f(x_o) + Av + o(v), \tag{2}$$

with an "error term" $o(v)$ satisfying $o(v)/\|v\| \to 0$ as $v \to 0$. If so, then $f$ is said to be *differentiable* at $x_o$ and the linear transformation $A$ (which depends on $f$ and on $x_o$) is called the *differential* of $f$ at $x_o$, denoted $df_{x_o}$. Thus $df_{x_o} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation depending on $f$ and on $x_o$. It is evidently given by the formula

$df_{x_o}(v) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}[f(x_o + \epsilon v) - f(x_o)]$.

In this definition $f(x)$ does not need to be defined for all $x \in \mathbb{R}^n$, only in a neighbourhood of $x_o$, i.e. for all $x$ satisfying $\|x - x_o\| < \epsilon$ for some $\epsilon > 0$. We say simply $f$ is differentiable if is differentiable at every $x_o$ in its domain. The definition of "differentiable" requires that the domain of $f$ must be an open set. i.e. it contain a neighbourhood of each of its points, so that $f(x_o+v)$ is defined for all $v$ near 0. We sometimes write $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ to indicate that $f$ need only be partially defined, on some *open* subset.

1.1.**4 Example**. Suppose $f(x)$ is itself linear. Then

$$f(x_o + v) = f(x_o) + f(v) \tag{3}$$

so (2) holds with $A(v) = f(v)$ and $o(v) \equiv 0$. Thus for a linear map $df_x(v) = f(v)$ for all $x$ and $v$.

(b) Suppose $f(x, y)$ is bilinear, linear in $x$ and $y$ separately. Then

$$f(x_o + v, y_o + w) = f(x_o, y_o) + f(x_o, w) + f(v, y_o) + f(v, w).$$

*Claim.* $f(v, w) = o((v, w))$. *Check.* In terms of components, we have $v = (\xi^i)$, $w = (\eta^j)$, and $f(v, w) = \sum c_{ij}\xi^i\eta^j$, so

$$\begin{aligned} |f(v, w)| &\le \sum |c_{ij}\xi^i\eta^j| \\ &\le C\max\{|c_{ij}|\}\max\{|\xi^i\eta^j|\} \le C \left(\max\{|\xi^i|, |\eta^j|\}\right)^2 \\ &\le C\sum(\xi^i)^2 + (\eta^j)^2 = C\|(v, w)\|^2, \end{aligned}$$

hence $|f(v, w)|/\|(v, w)\| \le C\|(v, w)\|$and$\to 0$ as $(v, w) \to (0, 0)$. □

It follows that (2) holds with $x$ replaced by $(x, y)$ and $v$ by $(v, w)$ if we take $A(v, w) = f(x_o, w) + f(v, y_o)$ and $o((v, w)) = f(v, w)$. Thus for all $x, y$ and $v, w$

$$df_{(x,y)}(v, w) = f(x, w) + f(v, y). \tag{4}$$

The following theorem is a basic criterion for deciding when a map $f$ is differentiable.

1.1.**5 Theorem**. *Let $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ be a map. If the partial derivatives $\partial f/\partial x^i$ exist and are continuous in a neighbourhood of $x_o$, then $f$ is differentiable at $x_o$. Furthermore, the matrix of its differential $df_{x_o} : \mathbb{R}^n \to \mathbb{R}^m$ is the Jacobian matrix $(\partial f/\partial x^i)_{x_o}$.*
(Proof omitted) □

Thus if we write $y = f(x)$ as $y^j = f^j(x)$ where $x = (x^i) \in \mathbb{R}^n$ and $y = (y^i) \in \mathbb{R}^m$, then we have

$$df_{x_o}(v) = \left(\frac{\partial f^i}{\partial x^j}\right)_{x_o} v^j.$$

We shall often suppress the subscript $x_o$, when the point $x_o$ is understood or unimportant, and then simply write $df$. A function $f$ as in the theorem is said to be of class $C^1$; more generally, $f$ is of class $C^k$ if all partials of order $\leq k$ exist and are continuous; $f$ is of class $C^\infty$ if it has continuous partials of all orders. We shall assume throughout that all functions are differentiable as often as needed, even if not stated explicitly each time. We call such functions *smooth* and to be definite we take this to mean of class $C^\infty$.

1.1.**6 Example**. (a) Consider a function $f : \mathbb{R}^n \cdots \to \mathbb{R}$, $y = f(x^1, \cdots, x^n)$ of $n$ variables $x^i$ with scalar values. The formula for $df(v)$ becomes

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

This can be written as a matrix product

$$df(v) = \begin{bmatrix} \frac{\partial f}{\partial x^1} & \cdots & \frac{\partial f}{\partial x^n} \end{bmatrix} \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}$$

Thus $df$ can be represented by the *row* vector with components $\partial f/\partial x^i$.
(b) Consider a function $f : \mathbb{R} \cdots \to \mathbb{R}^n$ , $(x^1, \cdots, x^n) = (f^1(t), \cdots, f^n(t))$ of one variable $t$ with values in $\mathbb{R}^n$. (The $n$ here plays the role to the $m$ above). The formula for $df(v)$ becomes

$$\begin{bmatrix} df^1(v) \\ \vdots \\ df^n(v) \end{bmatrix} = \begin{bmatrix} \frac{df^1}{dt} \\ \vdots \\ \frac{df^n}{dt} \end{bmatrix}.$$

In matrix notation we could represent $df$ by the *column* vector with components $df^i/dt$. (
$v$ is now a scalar .) Geometrically such a function represents a parametrized curve in $\mathbb{R}^n$ and we can think of think of $p = f(t)$ think as the position of a moving point at time $t$. We shall usually omit the function symbol $f$ and simply write $p = p(t)$ or $x^i = x^i(t)$. Instead $df^i/dt$ we write $\dot{x}^i(t)$ instead of $df$ we write

$\dot{p}(t)$, which we think of as a vector with components $\dot{x}^i(t)$, the velocity vector of the curve.
(c) As a particular case of a scalar-valued function on $\mathbb{R}^n$ we can take the $i$-th coordinate function

$$f(x^1, \cdots, x^n) = x^i.$$

As is customary in calculus we shall simply write $f = x^i$ for this function. Of course, this is a bit objectionable, because we are using the same symbol $x^i$ for the function and for its general value. But this notation has been around for hundreds of years and is often very convenient. Rather than introducing yet another notation it is best to think this through once and for all, so that one can always interpret what is meant in a logically correct manner. Take a symbol like $dx^i$, for example. This is the differential of the function $f = x^i$ we are considering. Obviously for this function $f = x^i$ we have $\frac{\partial x^i}{\partial x^j} = \delta^i_j$. So the general formula $df(v) = \frac{\partial f}{\partial x^j} v^j$ becomes

$$dx^i(v) = v^i.$$

This means that $df = dx^i$ picks out the $i$-th component of the vector $v$, just like $f = x^i$ picks out the $i$-the coordinate of the point $p$. As you can see, there is nothing mysterious about the differentials $dx^i$, in spite of the often confusing explanations given in calculus texts. For example, the equation

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

for the differential of a scalar-valued function $f$ is literally correct, being just another way of saying that

$$df(v) = \frac{\partial f}{\partial x^i} v^i$$

for all $v$. As a further example, the approximation rule

$$f(x_o + \Delta x) \approx f(x_o) + \frac{\partial f}{\partial x^i} \Delta x$$

is just another way of writing

$$f(x_o + v) = f(x_o) + \left(\frac{\partial f}{\partial x^i}\right)_{x_o} v^i + o(v),$$

which is the definition of the differential.
(d) Consider the equations defining polar coordinates

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Think of these equations as a map between two different copies of $\mathbb{R}^2$, say $f : \mathbb{R}^2_{r\theta} \to \mathbb{R}^2_{xy}$, $(r, \theta) \to (x, y)$. The differential of this map is given by the equations

$$dx = \frac{\partial x}{\partial r} dr + \frac{\partial x}{\partial \theta} d\theta = \cos\theta \, dr - \sin\theta \, d\theta$$
$$dy = \frac{\partial y}{\partial r} dr + \frac{\partial y}{\partial \theta} d\theta = \sin\theta \, dr + \cos\theta \, d\theta$$

This is just the general formula $df^i = \frac{\partial f^i}{\partial x^j} dx^j$ but this time with $(x^1, x^2) = (r, \theta)$ the coordinates in the $r\theta$ plane $\mathbb{R}^2_{r\theta}$, $(y^1, y^2) = (x, y)$ the coordinates in the $xy$ plane $\mathbb{R}^2_{xy}$, and $(x, y) = (f^1(r, \theta), f^2(r, \theta))$ the functions defined above. Again, there is nothing mysterious about these differentials.

1.1.**7 Theorem (Chain Rule: general case).** *Let $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ and $g : \mathbb{R}^m \cdots \to \mathbb{R}^l$ be differentiable maps. Then $g \circ f : \mathbb{R}^n \cdots \to \mathbb{R}^l$ is also differentiable (where defined) and*

$$d(g \circ f)_x = (dg)_{f(x)} \circ df_x$$

*(whenever defined).*

The proof needs some preliminary remarks. Define the norm $\|A\|$ of a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ by the formula $\|A\| = \max_{\|x\|=1} \|A(x)\|$. This max is finite since the sphere $\|x\| = 1$ is compact (closed and bounded). Then $\|A(x)\| \le \|A\|\|x\|$ for any $x \in \mathbb{R}^n$: this is clear from the definition if $\|x\| = 1$ and follows for any $x$ since $A(x) = \|x\|A(x/\|x\|)$ if $x \ne 0$.

**Proof** . Fix $x$ and set $k(v) = f(x + v) - f(x)$. Then by the differentiability of $g$,

$$g(f(x + v)) - g(f(x)) = dg_{f(x)}k(v) + o(k(v))$$

and by the differentiability of $f$,

$$k(v) = f(x + v) - f(x) = df_x(v) + o(v)$$

We use the notation $O(v)$ for any function satisfying $O(v) \to 0$ as $v \to 0$. (The letter big $O$ can stand for a different such function each time it occurs.) Then $o(v) = \|v\|O(v)$ and similarly $o(k(v)) = \|k(v)\|O(k(v)) = \|k(v)\|O(v)$. Substitution the expression for $k(v)$ into the preceding equation gives

$$g(f(x + v)) - g(f(x)) = dg_{f(x)}(df_x(v)) + \|v\|dg_{f(x)}(O(v)) + \|k(v)\|O(v).$$

We have $\|k(v)\| \le \|df_x(v)\| + o(v) \le \|df_x\|\|v\| + C\|v\| \le C'\|v\|$. It follows that

$$g(f(x + v)) - g(f(x)) = dg_{f(x)}(df_x(v)) + \|v\|O(v)$$

or yet another $O$. □

Of particular importance is the special case of the Chain Rule for curves. A curve in $\mathbb{R}^n$ is a function $t \to x(t)$, $\mathbb{R} \cdots \to \mathbb{R}^n$, defined on some interval. In that case we also write $\dot{x}(t)$ or $dx(t)/dt$ for its derivative vector $(dx^i(t)/dt)$ (which is just the differential of $t \to x(t)$ considered as a vector).

1.1.**8 Corollary (Chain Rule: special case).** *Let $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ be a differentiable map and $p(t)$ a differentiable curve in $\mathbb{R}^n$. Then $f(p(t))$ is also differentiable (where defined) and*

$$\frac{df(p(t))}{dt} = df_{p(t)}\Big(\frac{dp(t)}{dt}\Big).$$

*Geometrically, this equation says the says that the tangent vector $df(p(t))/dt$ to the image $f(p(t))$ of the curve $p(t)$ under the map $f$ is the image of its tangent vector $dp(t)/dt$ under the differential $df_{p(t)}$.*

1.1.**9 Corollary**. *Let $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ be a differentiable map $p_o \in \mathbb{R}^n$ a point and $v \in \mathbb{R}^n$ any vector. Then*

$$df_{p_o}(v) = \frac{d}{dt}\, f(p(t))\Big|_{t=t_o}$$

*for any differentiable curve $p(t)$ in $\mathbb{R}^n$ with $p(t_o) = p_o$ and $\dot{p}(t_o) = v$.*

The corollary gives a recipe for calculating the differential $df_{x_o}(v)$ as a derivative of a function of one variable $t$: given $x_0$ and $v$, we can choose any curve $x(t)$ with $x(0) = x_o$ and $\dot{x}(0) = v$ and apply the formula. This freedom of choice of the curve $x(t)$ and in particular its independence of the coordinates $x^i$ makes this recipe often much more convenient that the calculation of the Jacobian matrix $\partial f^i/\partial x^j$ in Theorem 1.1.5.

1.1.**10 Example**. Return to example 1.1.4, p 12

a) If $f(x)$ is linear in $x$, then

$$\frac{d}{dt}\, f(x(t)) = f(\frac{dx(t)}{dt})$$

b) If $f(x,\, y)$ is bilinear in $(x,\, y)$ then

$$\frac{d}{dt}\, f(x(t),\, y(t)) = f\Big(\frac{dx(t)}{dt},\, y(t)\Big) + f\Big(x(t),\, \frac{dy(t)}{dt}\Big)$$

i.e. $df_{(x,y)}(u,v) = f(u,x) + f(x,v)$. We call this the product rule for bilinear maps $f(x,y)$. For instance, let $\mathbb{R}^{n\times n}$ be the space of all real $n\times n$ matrices and

$$f : \mathbb{R}^{n\times n}\times\, \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},\; f(X,Y) = XY$$

the multiplication map. This map is bilinear, so the product rule for bilinear maps applies and gives

$$\frac{d}{dt} f(X(t),Y(t)) = \frac{d}{dt}(X(t)Y(t)) = \dot{X}(t)Y(t) + X(t)\dot{Y}(t).$$

The Chain Rule says that this equals $df_{(X(t),Y(t)}(\dot{X}(t),\dot{Y}(t))$. Thus $df_{(X,Y)}(U,V) = UX + YV$. This formula for the differential of the matrix product $XY$ is more simply written as the Product Rule

$$d(XY) = (dX)Y + X(dY).$$

You should think about this formula until you see that it perfectly correct: the differentials in it have a precise meaning, the products are defined, and the equation is true. The method of proof gives a way of disentangling the meaning of many formulas involving differentials: just think of $X$ and $Y$ as depending on a parameter $t$ and rewrite the formula in terms of derivatives:

$$\frac{d(XY)}{dt} = \frac{dX}{dt}Y + X\frac{dY}{dt}.$$

1.1.**11 Coordinate version of the Chain Rule**. In terms of coordinates the above theorems read like this (using the summation convention).

(a) *Special case.* In coordinates write $y = f(x)$ and $x = x(t)$ as

$$y^j = f^j(x^1, \cdots, x^n),\; x^i = x^i(t)$$

respectively. Then

$$\frac{dy^j}{dt} = \frac{\partial y^j}{\partial x^i}\frac{dx^i}{dt}.$$

(b) *General case.* In coordinates, write $z = g(y)$, $y = f(x)$ as

$$z^k = g^k(y^1, \cdots, y^m),\; y^j = f^j(x^1, \cdots, x^n)$$

respectively. Then

$$\frac{\partial z^k}{\partial x^i} = \frac{\partial z^k}{\partial y^j}\frac{\partial y^j}{\partial x^i}.$$

In the last equation we consider some of the variables are considered as functions of others by means of the preceding equation, without indicating so explicitly. For example, the $z^k$ on the left is considered a function of the $x^i$ via $z = g(f(x))$, on the right $z^k$ is first considered as a function of $y^j$ via $z = g(y)$ and after the differentiation the partial $\partial z^k/\partial y^j$ is considered a function of $x$ via $y = g(x)$. In older terminology one would say that some of the variables are dependent, some are independent. This kind of notation is perhaps not entirely logical, but is very convenient for calculation, because the notation itself tells you what to do mechanically, as you know very well from elementary calculus. This is just what a good notation should do, but one should always be able to go beyond the mechanics and understand what is going on.

1.1.12 Examples.
*(a)*Consider the tangent vector to curves $p(t)$ on the sphere $x^2+y^2+z^2=R^2$ in $\mathbb{R}^3$.
The function $f(x,y,z)=x^2+y^2+z^2$ is constant on such a curve $p(t)$, i.e. $f(p(t))$ =const, and, hence

$0=\frac{d}{dt}f(p(t))|_{t=t_o}=df_{p(t_o)}(\dot p(t_o))=x_o\xi+y_o\eta+z_o\zeta$

where $p(t_o)=(x_o,y_o,z_o)$ and $\dot p(t_o)=(\xi,\eta,\zeta)$. Thus we see that tangent vectors at $p_o$ to curves on the sphere do indeed satisfy the equation $x_o\xi+y_o\eta+z_o\zeta=0$, as we already mentioned.

(b) Let $r=r(t)$, $\theta=\theta(t)$ be the parametric equation of a curve in polar coordinates $x=r\cos\theta$, $y=r\sin\theta$. The tangent vector of this curve has components $\dot x,\dot y$ given by

$$\begin{aligned}\frac{dx}{dt}&=\frac{\partial x}{\partial r}\frac{dr}{dt}+\frac{\partial x}{\partial\theta}\frac{d\theta}{dt}=\cos\theta\frac{dr}{dt}-r\sin\theta\frac{d\theta}{dt}\\ \frac{dy}{dt}&=\frac{\partial y}{\partial r}\frac{dr}{dt}+\frac{\partial y}{\partial\theta}\frac{d\theta}{dt}=\sin\theta\frac{dr}{dt}+r\cos\theta\frac{d\theta}{dt}\end{aligned}$$

*In particular, consider a $\theta$-coordinate line $r=r_o$ (constant), $\theta=\theta_o+t$ (variable). As a curve in the $r\theta$ plane this is indeed a straight line through $(r_o,\theta_o)$; its image in the $xy$ plane is a circle through the corresponding point $(x_o,y_o)$. The equations above become*

$$\begin{aligned}\frac{dx}{dt}&=\frac{\partial x}{\partial r}\frac{dr}{dt}+\frac{\partial x}{\partial\theta}\frac{d\theta}{dt}=(\cos\theta)\,0-(r\sin\theta)\,1\\ \frac{dy}{dt}&=\frac{\partial y}{\partial r}\frac{dr}{dt}+\frac{\partial y}{\partial\theta}\frac{d\theta}{dt}=(\sin\theta)\,0+(r\cos\theta)\,1\end{aligned}$$

They show that the tangent vector to image in the $xy$ plane of the $\theta$-coordinate line the image of the $\theta$-basis vector $e_\theta=(1,0)$ under the differential of the polar coordinate map $x=r\cos\theta$, $y=r\sin\theta$. The tangent vector to the image in the $xy$ plane of the $\theta$-coordinate line is the image of the $\theta$-basis vector $e_\theta=(0,1)$ under the differential of the polar coordinate map $x=r\cos\theta$, $y=r\sin\theta$. The tangent vector to the image of the $r$-coordinate line is similarly the image of the $r$-basis vector $e_r=(1,0)$.

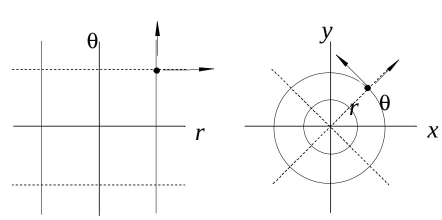



Fig. 2

(c) Let $z = f(r, \theta)$ be a function given in polar coordinates. Then $\partial z/\partial x$ and $\partial z/\partial y$ are found by solving the equations

$$\begin{aligned}\frac{\partial z}{\partial r} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial z}{\partial x}\cos\theta + \frac{\partial z}{\partial y}\sin\theta \\ \frac{\partial z}{\partial \theta} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \theta} = \frac{\partial z}{\partial x}(-r\sin\theta) + \frac{\partial z}{\partial y}r\cos\theta\end{aligned}$$

which gives (e.g. using Cramer's Rule)

$$\begin{aligned}\frac{\partial z}{\partial x} &= \cos\theta\frac{\partial z}{\partial r} - \frac{\sin\theta}{r}\frac{\partial z}{\partial \theta} \\ \frac{\partial z}{\partial y} &= \sin\theta\frac{\partial z}{\partial r} + \frac{\cos\theta}{r}\frac{\partial z}{\partial \theta}\end{aligned}$$

1.1.**13 Corollary**. Suppose $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ and $g : \mathbb{R}^m \cdots \to \mathbb{R}^n$ are inverse functions, i.e. $g \circ f(x) = x$ and $f \circ g(y) = y$ whenever defined. Then $df_x$ and $dg_y$ are inverse linear transformations when $y = f(x)$, i.e. $x = g(y)$:

$$dg_{f(x)} \circ df_x = 1,\ df_{g(y)} \circ dg_y = 1.$$

This implies in particular that $df_x : \mathbb{R}^n \to \mathbb{R}^m$ is an invertible linear transformation, and in particular $n = m$. If we write in coordinates $y = f(x)$ as $y^j = y^j(x)$ and $x = g(y)$ as $x^i = x^i(y)$ this means that we have

$$\frac{\partial y^i}{\partial x^j}\frac{\partial x^j}{\partial y^k} = \delta^i_k \text{ and } \frac{\partial x^i}{\partial y^j}\frac{\partial y^j}{\partial x^k} = \delta^i_k$$

For us, the most important theorem of differential calculus is the following.

1.1.14 **Theorem (Inverse Function Theorem)**. *Let $f : \mathbb{R}^n \cdots \to \mathbb{R}^n$ be a $C^k (k \geq 1)$ map and $x_o$ a point in its domain. If $df_{x_o} : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then $f$ maps some open neighbourhood $U$ of $x_o$ on–to$-$one onto an open neighbourhood $V$ of $f(x_o)$ and the inverse map $g : V \to U$ is also $C^k$.*

We paraphrase the conclusion by saying that $f$ *is locally $C^k$–invertible* at $x_o$. If we write again $y = f(x)$ as $y^j = y^j(x)$ then the condition that $df_{x_o}$ be invertible means that $\det(\partial y^i/\partial x^j)_{x_o} \neq 0$. We shall give the proof of the Inverse Function Theorem after we have looked at an example.

1.1.**15 Example: polar coordinates in the plane**. Consider the map $F(r, \theta) = (x, y)$ form the $r\theta$–plane to the $xy$–plane given by the equations $x = r\cos\theta$, $y = r\sin\theta$. The Jacobian determinant of this map is

$$\det\begin{bmatrix}\cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{bmatrix} = r$$

Thus $\det[\partial(x, y)/\partial(r, \theta)] \neq 0$ except when $r = 0$, i.e. $(x, y) = (0, 0)$. Hence for any point $(r_o, \theta_o)$ with $r_o \neq 0$ one can find a neighbourhood $U$ of $(r_o, \theta_o)$ in the $r\theta$-plane and a neighbourhood $V$ of $(x_o, y_o) = (r_o\cos\theta_o, r_o\sin\theta_o)$ in the $xy$-plane so that $F$ maps $U$ one-to-one onto $V$ and $F : U \to V$ has a smooth inverse $G : V \to U$. It is obtained by solving $x = r\cos\theta$, $y = r\sin\theta$ for $(r, \theta)$, subject to the restriction $(r, \theta) \in U, (x, y) \in V$. For $x \neq 0$ the solution can be written as $r = \pm\sqrt{x^2 + y^2}$, $\theta = \arctan(y/x)$ where the sign $\pm$ must be chosen so that $(r, \theta)$ lies in $U$. But this formula does not work when $x = 0$, even though a local inverse exists as long as $(x, y) \neq (0, 0)$. For example, for any point off the positive $x$-axis one can take for $U$ the region in the $r\theta$-plane described by $r > 0$, $-\pi < \theta < \pi$. $V$ is then the corresponding region in the $xy$-plane, which is just the points off the positive $x$-axis. It is geometrically clear that the map $(r, \theta) \to (x, y)$ map $V$ one-to-one onto $U$. For any point off the negative $x$-axis one can take for $U$ the region in the $r\theta$-plane described by $r > 0$, $0 < \theta < \pi$. $V$ is then the corresponding region in the $xy$-plane, which is just the points off the negative $x$-axis. It is again geometrically clear that the map $(r, \theta) \to (x, y)$ map $V$ one-to-one onto $U$.

For the proof of the Inverse Function Theorem we need a lemma, which is itself very useful.

1.1.16 **Contraction Lemma**. *Let $U$ be an open subset of $\mathbb{R}^n$ and $F : U \to U$ be any map of $U$ into itself which is a contraction, i.e. there is $K$ is a positive constant $\leq 1$ so that for all $x, y \in U$, $\|F(y) - F(x)\| \leq K\|y - x\|$. Then $F$ has a unique fixed point $x_o$ in $U$, i.e. a point $x_o \in U$ so that $F(x_o) = x_o$. Moreover, for any $x \in U$,*

$$x_o = \lim_{n\to\infty} F^n(x) \tag{5}$$

*where $F^n = F \circ \cdots \circ F$ is the $n$–th iterate of $F$.*

**Proof**. Let $x \in U$. We show that $\{F^n(x)\}$ converges. By Cauchy's convergence criterion, it suffices to show that for any $\epsilon > 0$ there is an $N$ so that $\|F^n(x) - F^m(x)\| \leq \epsilon$ for $n, m \geq N$. For this we may assume $n \geq m$ so that we can write $n - m = r \geq 0$. Then

$$\begin{aligned} \|F^n(x) - F^m(x)\| &\leq K\|F^{n-1}(x) - F^{m-1}(x)\| \\ &\leq \cdots \leq K^m\|F^r(x) - x\|. \end{aligned} \tag{6}$$

From this and the Triangle Inequality we get

$$\begin{aligned} \|F^r(x) - x\| &\leq \|F^r(x) - F^{r-1}(x)\| + \|F^{r-1}(x) - F^{r-2}(x)\| + \cdots + \|F(x) - x\| \\ &\leq (K^{r-1} + K^{r-2} + \cdots + 1)\|F(x) - x\| \\ &= \tfrac{K^r - 1}{K - 1}\|F(x) - x\| \end{aligned} \tag{7}$$

Since $0 < K < 1$ the fraction is bounded by $1/(K - 1)$ and the RHS of (6) tends to 0 as $m \to \infty$, as desired. Thus the limit $x_o$ in (5) exists for any given $x \in U$. To see that $x_o$ is a fixed point, consider

$$\|F(F^n(x)) - F(x_o)\| \leq K\|F^n(x) - x_o\|.$$

As $n \to \infty$, the LHS becomes $\|x_o - F(x_o)\|$ while the RHS becomes $K\|x_o - x_o\| = 0$, leaving us with $x_o - F(x_o) = 0$. This proves the existence of a fixed point. To prove uniqueness, suppose we also have $F(x_1) = x_1$. Then

$$\|x_o - x_1\| = \|F(x_o) - F(x_1)\| \leq K\|x_o - x_1\|$$

which implies $\|x_o - x_1\| = 0$, since $K \neq 1$. □

**Proof of the Inverse Function Theorem**. The idea of the proof is this. To find an inverse for $f$ we have to solve the equation $y = f(x)$ for $x$ if $y$ is given:

$x = f^{-1}(y)$. We rewrite this equation as $y + x - f(x) = x$, which says that $x$ is a fixed point of the map $h(x) := y + x - f(x)$, $y$ being given. So we'll try to show that this $h$ is a contraction map. The proof itself takes several steps.
(1) *Preliminaries.* We may assume $x_0 = 0$, by composing $f$ with the map $x \to x - x_o$. Next we may assume that $df_0 = 1$, the identity map, by replacing $f$ by $(df_0)^{-1} \circ f$ and using 1.1.12, p 1.1 We set $g(x) = x - f(x)$. Then $dg_0 = 1 - df_0 = 0$.
(2) *Claim.* There is $r > 0$ so that for $\|x\| \le 2r$ we have

$$\|dg_x(v)\| \le \frac{1}{2}\ \|v\| \tag{8}$$

for all $v$. *Check.* $\|dg_x(v)\| \le \|dg_x\|\|v\|$. For $x = 0$ we have $\|dg_x\| = 0$, so by continuity we can make $\|dg_x\|$arbitrarily small, and in particular $\le 1/2$, by making $\|x\|$sufficiently small, which we may write as $\|x\| \le 2r$.
(3) *Claim.* $g(x) = x - f(x)$ maps the ball $B_r = \{x \mid \|x\| \le r\}$ into $B_{r/2}$. Check. By the Fundamental Theorem of Calculus we have

$$g(x) = \int_0^1 \frac{d}{dt} g(tx)dt = \int_0^1 dg_{tx}(x)dt$$

hence from the triangle inequality for integrals and (8),

$$\|g(x)\| \le \int_0^1 \|dg_{tx}(x)\|dt \le \frac{1}{2}.$$

(4) *Claim.* Given $y \in B_{r/2}$ there is a unique $x \in B_r$ so that $f(x) = y$. Check. Fix $y \in B_{r/2}$ and consider$h(x) = y + g(x)$with $x \in B_r$. Since $\|y\| \le r/2$ and $\|x\| \le r$ we have $\|h(x)\| \le \|y\| + \|g(x)\| \le r$. So $h$ maps $B_r$ into itself. Let $x(t) = x_1 + t(x_2 - x_1)$. By the Fundamental Theorem of Calculus we have

$$h(x_2) - h(x_1) = \int_0^1 \frac{d}{dt} g(x(t))dt = \int_0^1 dg_{x(t)}(x_2 - x_1)dt.$$

Together with (8) this gives

$$\|h(x_2) - h(x_1)\| \le \int_0^1 \|dg_{x(t)}(x_2 - x_1)\|dt \le \int_0^1 \frac{1}{2}\|x_2 - x_1\|dt = \frac{1}{2}\|x_2 - x_1\|.$$

Thus $h : B_r \to B_r$ is a contraction map with $K = 1/2$ in the Contraction Lemma. It follows that $h$ has a unique fixed point, i.e for any $y \in B_{r/2}$ there is a unique $x \in B_r$ so that $y + g(x) = x$, which amounts to $y = f(x)$. This proves the claim. Given $y \in B_{r/2}$ write $x = f^{-1}(y)$ for the unique $x \in B_r$ satisfying $f(x) = y$. Thus $f^{-1} : B_{r/2} \to f^{-1}(B_{r/2}) \subset B_r$ is an inverse of $f : f^{-1}(B_{r/2}) \to B_{r/2}$.
(5) *Claim.* $f^{-1}$ is continuous. Check. Since $g(x) = x - f(x)$ we have $x = g(x) + f(x)$. Thus

$$\|x_1 - x_2\| = \|f(x_1) - f(x_2) + g(x_1) - g(x_2)\|$$

$\leq \|f(x_1) - f(x_2)\| + \|g(x_1) - g(x_2)\|$
$\leq \|f(x_1) - f(x_2)\| + \frac{1}{2}\|x_1 - x_2\|$
so $\|x_1 - x_2\| \leq 2\|f(x_1) - f(x_2)\|$, i.e. $\|f^{-1}(y_1) - f^{-1}(y_2)\| \leq 2\|y_1 - y_2\|$. Hence $f^{-1}$ is continuous.
(6) Claim. $f^{-1}$ is differentiable on the open ball $U = \{y \mid \|y\| < \frac{1}{2}r\}$. Check. Let $y, y_1 \in U$. Then $y = f(x), y_1 = f(x_1)$ with $x, x_1 \in B_r$. We have

$$f^{-1}(y) - f^{-1}(y_1) - (df_{x_1})^{-1}(y - y_1) = x - x_1 - (df_{x_1})^{-1}(f(x) - f(x_1)). \quad (9)$$

Since $f$ is differentiable,

$$f(x) = f(x_1) + df_{x_1}(x - x_1) + o(x - x_1).$$

Substituting this into the RHS of (9) we find

$$(df_{x_1})^{-1}(o(x - x_1)). \quad (10)$$

Since the linear map $df_{x_1}$depends continuously on $x_1 \in B_r$ and $\det df_{x_1} \neq 0$ there, its inverse $(df_{x_1})^{-1}$ is continuous there as well. (Think of Cramer's formula for $A^{-1}$.) Hence we can apply the argument of step (2) to (10) and find that

$$\|(df_{x_1})^{-1}(o(x - x_1)\| \leq Co(x - x_1). \quad (11)$$

for $x, x_1 \in B_r$. We also know that $\|x - x_1\| \leq 2\|y - y_1\|$. Putting these inequalities together we find from (9) that

$$f^{-1}(y) - f^{-1}(y_1) - df_{x_1}^{-1}(y - y_1) = o(y - y_1)$$

for another little $o$. This proves that $f^{-1}$ is differentiable at $y_1$ with $(df^{-1})_{y_1} = (df_{x_1})^{-1}$.
(7) *Conclusion of the proof.* Since $f$ is of class $C^1$ the matrix entries $\partial f^i/\partial x^j$ of $df_x$ are continuous functions of $x$, hence the matrix entries of

$$(df^{-1})_y = (df_x)^{-1} \quad (12)$$

are continuous functions of $y$. (Think again of Cramer's formula for $A^{-1}$ and recall that $x = f^{-1}(y)$ is a continuous function of $y$.) This proves the theorem for $k = 1$. For $k = 2$, we have to show that (12) is still $C^1$ as a function of $y$. As a function of $y$, the RHS of (11) is a composite of the maps $y \to x = f^{-1}(y)$, $x \to df_x$, and $A \to A^{-1}$. This first we just proved to be $C^1$; the second is $C^1$ because $f$ is $C^2$; the third is $C^1$ by Cramer's formula for $A^{-1}$. This proves the theorem for $k = 2$, and the proof for any $k$ follows from the same argument by induction. □

**EXERCISES 1.1**

1. Let $V$ be a vector space, $\{v_i\}$ and $\{\tilde{v}_i\}$ two basis for $V$. Prove (summation convention throughout):
(a) There is an *invertible* matrix $(a_i^j)$ so that $\tilde{v}_j = a_j^i v_i$ .

(b) The components $(x^i)$ and $(\tilde{x}^j)$ of a vector $v \in V$ are related by the equation $x^i = a^i_j \tilde{x}^j$. [Make sure you proof is complete: you must prove that the $x^i$ have to satisfy this relation.]
(c) The components $(\xi_i)$ and $(\tilde{\xi}_j)$ of a linear functional vector $\varphi \in V^*$ with respect to the dual bases are related by the equation $\tilde{\xi}_j = a^i_j \xi_i$ . [Same advice.]

2. Suppose $A$, $B : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformations of $\mathbb{R}^n$ which are inverses of each other. Prove from the definitions that their matrices satisfy $a^j_k b^k_i = \delta^j_i$.

3. Let $A : V \to W$ be a linear map. (a) Suppose $A$ is injective. Show that there is a basis of $V$ and a basis of $W$ so that in terms of components $x \to y := A(x)$ becomes $(x^1, \cdots, x^n) \to (y^1, \cdots, y^n, \cdots, y^m) := (x^1, \cdots, x^n, 0, \cdots, 0)$.
(b) Suppose $A$ is surjective. Show that there is a basis of $V$ and a basis of $W$ so that in terms of components $x \to y := A(x)$ becomes $(x^1, \cdots, x^m, \cdots, x^n) \to (y^1, \cdots, y^m) := (x^1, \cdots, x^m)$.
*Note*. Use these bases to identify $V \approx \mathbb{R}^n$ and $W \approx \mathbb{R}^m$. Then (a) says that $A$ becomes the inclusion $\mathbb{R}^n \to \mathbb{R}^m$ $(n \leq m)$ and (b) says that $A$ becomes the projection $\mathbb{R}^n \to \mathbb{R}^m$ $(n \geq m)$. Suggestion. In (a) choose a basis $\{v_1, \cdots, v_m\}$ for $V$ and $\{w_1, \cdots, w_m, w_{m+1}, \cdots, w_n\}$ for $W$ so that $w_j = A(v_j)$ for $j \leq m$. Explain how. In (b) choose a basis $\{v_1, \cdots, v_m, v_{m+1}, \cdots, v_n\}$ for $V$ so that $A(v_j) = 0$ iff $j \geq m + 1$. Show that $\{w_1 = A(v_1), \cdots, w_m = A(v_m)\}$ is a basis for $W$. You may wish to use the fact that *any linearly independent subset of a vector space can be completed to a basis*, which you can look up in your linear algebra text.

4. Give a detailed proof of the equation $(a^*)^j_i = a^j_i$ stated in the text. Explain carefully how it is related to the equation $(a^*)_{ij} = a_{ji}$ mentioned there. Start by writing out carefully the definitions of the four quantities $(a^*)^j_i$, $a^j_i$, $(a^*)_{ij}$, $a_{ji}$.

5. Deduce the corollary 1.1.8 from the Chain Rule.

6. Let $f : (r, \theta, \phi) \to (x, y, z)$ be the spherical coordinate map:

$$x = \rho \cos\theta \sin\phi,\ y = \rho \sin\theta \sin\phi,\ z = \rho\ \cos\phi.$$

(a) Find $dx, dy, dz$ in terms of $d\rho,\ d\theta, d\phi$ . (b) Find $\frac{\partial f}{\partial \rho}, \frac{\partial f}{\partial \theta}, \frac{\partial f}{\partial \phi}$ in terms of $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$.

7. Let $f : (\rho, \theta, \phi) \to (x, y, z)$ be the spherical coordinate map as above. (a)Determine all points $(\rho, \theta, \phi)$ at which $f$ is locally invertible. (b)For each such point specify an open neighbourhood, as large as possible, on which $f$ is one-to-one. (c) Prove that $f$ is not locally invertible at the remaining point(s).

8. (a) Show that the volume $V(u, v, w)$ of the parallelepiped spanned by three vectors $u, v, w$ in $\mathbb{R}^3$ is the absolute value of $u \cdot (v \times w)$. [Use a sketch and the geometric properties of the dot product and the cross product you know from your first encounter with vectors as little arrows.]
(b) Show that $u \cdot (v \times w) = |u, v, v|$the determinant of the matrix of the components of $u, v, w$ with respect to the standard basis $e_1, e_2, e_3$ of $\mathbb{R}^3$. Deduce that for any linear transformation $A$ of $\mathbb{R}^3$, $V(Au, Av, Aw) = |\det A| V(u, v, w)$.

Show that this formula remains valid for the matrix of components with respect to any right–handed orthonormal basis $u_1, u_2, u_3$ of $\mathbb{R}^3$. [Right–handed means $\det[u_1, u_2, u_3] = +1$.]

9. Let $f : (\rho, \theta, \phi) \to (x, y, z)$ be the spherical coordinate map. [See problem 6]. For a given point $(\rho_o, \theta_o, \phi_o)$ in $\rho\theta\phi$-space, let vectors $v_\rho, v_\theta, v_\phi$ at the point $(x_o, y_o, z_o)$ in $xyz$-space which correspond to the three standard basis vectors $e_\rho, e_\theta, e_\phi = (1, 0, 0), (0, 1, 0), (0, 0, 1)$ in $\rho\theta\phi$-space under the differential $df$ of $f$.
(a) Show that $v_\rho$ is the tangent vector of the $\rho$–coordinate curve $\mathbb{R}^3$, i.e. the curve with parametric equation $\rho =$arbitrary (parameter), $\theta = \theta_o, \phi = \phi_o$ in spherical coordinates. Similarly for $v_\theta$ and $v_\phi$. Sketch.
(b) Find the volume of the parallelepiped spanned by the three vectors $v_\rho, v_\theta, v_\phi$ at $(x, y, z)$. [See problem 8.]

10. Let $f, g$ be two differentiable functions $\mathbb{R}^n \to \mathbb{R}$. Show from the definition (1.4) of $df$ that

$$d(fg) = (df)g + f(dg).$$

11. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathrm{C}^2$. Show that the second partials are symmetric:
$\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j} f = \frac{\partial}{\partial x^j}\frac{\partial}{\partial x^i} f$ for all $ij$.
[You may consult your calculus text.]

12. Use the definition of $df$ to calculate the differential $df_x(v)$ for the following functions $f(x)$. [Notation: $x = (x^i)$, $v = (v^i)$.]
(a) $\sum_i c_i x^i$ ($c_i$ =constant) (b) $x^1 x^2 \cdots x^n$ (c)$\sum_i c_i (x^i)^2$.

13. Consider $\det(X)$ as function of $X = (X_i^j) \in \mathbb{R}^{n\times n}$.
a) Show that the differential of det at $X = 1$ (identity matrix) is the trace, i.e. $d\det_1(A) = \mathrm{tr}A$. [The trace of $A$ is defined as $\mathrm{tr}A = \sum_i A_i^i$. Suggestion. Expand $\det(1 + tA)$ using the definition (1) of det to first order in $t$, i.e. omitting terms involving higher powers of $t$. Explain why this solves the problem. If you can't do it in general, work it out at least for $n = 2$, 3.]
b) Show that for any invertible matrix $X \in \mathbb{R}^{n\times n}$, $d\det_X(A) = \det(X)\,\mathrm{tr}(X^{-1}A)$. [Suggestion. Consider a curve $X(t)$ and write $\det(X(t)) = \det(X(t_o))\det(X(t_o)^{-1}X(t))$. Use the Chain Rule 1.1.7.]

14. Let $f(x^1, \cdots, x^n)$ be a differentiable function satisfying $f(tx^1, \cdots, tx^n) = t^N f(x^1, \cdots, x^n)$ for all $t > 0$ and some (fixed) N. Show that $\sum_i x^i \frac{\partial f}{\partial x^i} = Nf$. [Suggestion. Chain Rule.]

15. Calculate the Jacobian matrices for the polar coordinate map $(r, \theta) \to (x, y)$ and its inverse $(x, y) \to (r, \theta)$, given by $x = r\cos\theta$, $y = r\sin\theta$ and $r = \sqrt{x^2 + y^2}$, $\theta = \arctan\frac{y}{x}$. Verify by direct calculation that these matrices are inverses of each other, as asserted in corollary (1.1.13).

16. Consider an equation of the form $f(x, y) = 0$ where $f(x, y)$ is a smooth function of two variables. Let $(x_o, y_o)$ be a particular solution of this equation for which $(\partial f/\partial y)_{(x_o, y_o)} \neq 0$. Show that the given equation has a unique solution $y = y(x)$ for $y$ in terms of $x$ and $z$, provided $(x, y)$ is restricted to lie in a suitable

neighbourhood of $(x_o, y_o)$. [Suggestion. Apply the Inverse Function Theorem to $F(x,y) := (x, f(x,y))$].

17. Consider a system of $k$ equations in $n = m + k$ variables, $F^1(x^1, \cdots, x^n) = 0$, $\cdots$, $F^k(x^1, \cdots, x^n) = 0$ where the $F^i$ are smooth functions. Suppose $\det[\partial F^i/\partial x^{m+j} |\ 1 \leq i,j \leq k] \neq 0$ at a point $p_o \in \mathbb{R}^n$. Show that there is a neighbourhood $U$ of $p_o$ in $\mathbb{R}^n$ so that if $(x^1, \cdots, x^n)$ is restricted to lie in $U$ then the equations admit a unique solution for $x_{m+1}, \cdots, x_n$ in terms of $x_1, \cdots, x_m$. [This is called the *Implicit Function Theorem.*]

18. (a) Let $t \to X(t)$ be a differentiable map from $\mathbb{R}$ into the space $\mathbb{R}^{n\times n}$ of $n \times n$ matrices. Suppose that $X(t)$ is invertible for all $t$. Show that $t \to X(t)^{-1}$ is also differentiable and that $\frac{dX^{-1}}{dt} = -X^{-1}\left(\frac{dX}{dt}\right)X^{-1}$ where we omitted the variable $t$ from the notation. [Suggestion. Consider the equation $XX^{-1} = 1$.]
(b) Let $f$ be the inversion map of $\mathbb{R}^{n\times n}$ itself, given by $f(X) = X^{-1}$. Show that $f$ is differentiable and that $df_X(Y) = -X^{-1}YX^{-1}$.

## 1.2 Manifolds: definitions and examples

The notion of "manifold" took a long time to crystallize, certainly hundreds of years. What a manifold *should* be is clear: a kind of space to which one can apply the methods of differential calculus, based on local linear approximations for functions. Riemann struggled with the concept in his lecture of 1854. Weyl is credited with its final formulation, worked out in his book on Riemann surfaces of 1913, in a form adapted to that context. In his *Leçons* of 1925-1926 Cartan sill finds that *la notion générale de variété est assez difficile à définir avec précison* and goes on to give essentially the definition offered below. As is often the case, all seems obvious once the problem is solved. We give the definition in three axioms, à la Euclid.

1.2.**1 Definition**. An $n$–dimensional *manifold* $M$ consists of *points* together with *coordinate systems*. These are subject to the following axioms.

MAN **1**. Each coordinate system provides a one-to-one correspondence between the points $p$ in a certain subset $U$ of $M$ and the points $x = x(p)$ in an *open subse*t of $\mathbb{R}^n$.
MAN **2**. The *coordinate transformation* $\tilde{x}(p) = f(x(p))$ relating the coordinates of the same point $p \in U \cap \tilde{U}$ in two different coordinate systems $(x^i)$ and $(\tilde{x}^i)$ is given by a smooth map

$$\tilde{x}^i = f^i(x^1, \cdots, x^n),\ i = 1, \cdots, n.$$

The domain $\{x(p) \mid p \in U \cap \tilde{U}\}$ of $f$ is required to be open in $\mathbb{R}^n$.
MAN **3**. Every point of $M$ lies in the domain of some coordinate system.

Fig. 1 shows the maps $p \to x(p)$, $p \to \tilde{x}(p)$, and $x \to \tilde{x} = f(x)$

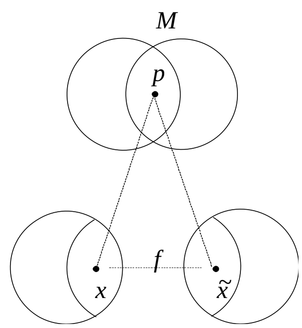



Fig. 1

In this axiomatic definition of "manifold", points and coordinate systems function as primitive notions, about which we assume nothing but the axioms MAN 1–3. The axiom MAN 1 says that a coordinate system is entirely specified by a certain partially defined map $M \cdots \to \mathbb{R}^n$,

$$p \mapsto x(p) = (x^1(p), \cdots, x^n(p))$$

and we shall take "coordinate system" to mean this map, in order to have something definite. But one should be a bit flexible with this interpretation: for example, in another convention one could equally well take "coordinate system" to mean the inverse map $\mathbb{R}^n \cdots \to M, x \to p(x)$ and when convenient we shall also use the notation $p(x)$ for the point of $M$ corresponding to the coordinate point $x$.
It will be noted that we use the same symbol $x = (x^i)$ for both the coordinate map $p \to x(p)$ and a general coordinate point $x = (x^1, \cdots, x^n)$, just as one does with the $xyz$-coordinates in calculus. This is premeditated confusion: it leads to a notation which tells one what to do (just like the $\partial y^i/\partial x^j$ in calculus) and suppresses extra symbols for the coordinate map and its domain, which are usually not needed. If it is necessary to have a name for the coordinate domain we can write $(U, x)$ for the coordinate system and if there is any danger of confusion because the of double meaning of the symbol $x$ (coordinate map and coordinate point) we can write $(U, \phi)$ instead. With the notation $(U, \phi)$ comes the term chart for coordinate system and the term *atlas* for any collection $\{(U_\alpha, \phi_\alpha)\}$ of charts satisfying MAN 1–3. In any case, one should always keep in mind that a manifold is not just a set of points, but a set of points together with an atlas, so that one should strictly speaking consider a manifold as a pair $(M, \{(U_\alpha, \phi_\alpha)\})$. But we usually call $M$ itself a manifold, and speak of its manifold structure when it is necessary to remind ourselves of the atlas $\{(U_\alpha, \phi_\alpha)\}$. Here are some examples to illustrate the definition.

1.2.**2 Example: the Cartesian plane**. As the set of points we take $\mathbb{R}^2 = \{p = (x, y) \mid x, y \in \mathbb{R}\}$. As it stands, this set of points is not a manifold: we have to specify a collection of the coordinate systems satisfying the axioms MAN 1–3. Each coordinate system will be a map $p \to (x_1(p), x_2(p))$, which maps a domain of points $p$, to be specified, one–to–one onto an open set of pairs $(x_1, x_2)$. We list some possibilities.

*Cartesian coordinates* $(x, y)$: $x(p) = x$, $y(p) = y$ if $p = (x, y)$. *Domain*: all $p$. Note again that $x$,$y$ are used to denote the coordinate of a general point $p$ as well as to denote the coordinate functions $p \to x(p), y(p)$

*Polar coordinates* $(r, \theta)$: $x = r\cos\theta$, $y = r\sin\theta$. *Domain*: $(r,\theta)$ may be used as coordinates in a neighbourhood of any point where $\partial(x, y)/\partial(r, \theta) = r \neq 0$, i.e. in a neighbourhood of any point except the origin. A possible domain is the set of points $p$ for which $r > 0, 0 < \theta < 2\pi$, the $p = (x, y)$ given by the above equations for these values of $r, \theta$; these are just the $p$'s off the positive $x$-axis. Other domains are possible and sometimes more convenient.

*Hyperbolic coordinates* $(u,\psi)$: $x = u\,\cosh\psi$, $y = u\sinh\psi$. *Domain*: The pairs $(x, y)$ representable in this form are those satisfying $x^2 - y^2 = u^2 \geq 0.(u, \psi)$ may be used as coordinates in a neighbourhood of any point $(x, y)$ in this region for which $\det \partial(x,y)/\partial(u,\psi) = u \neq 0$. The whole region $\{(x, y) \mid x^2 - y^2 > 0\}$ corresponds one-to-one to $\{(u, \psi) \mid u \neq 0\}$ and can serve as coordinate domain. In the region $\{(x, y) \mid x^2 - y^2 < 0\}$ one can use $x = u\sinh\psi, y = u\cosh\psi$ as coordinates.

*Linear coordinates* $(u, v)$: $u = ax + by$, $v = cx + dy$, where $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is any invertible $2 \times 2$ matrix (i.e. $ab - cd \neq 0$). A special case are orthogonal linear coordinates, when the matrix is orthogonal. Domain: all $p$.

*Affine coordinates* $(u, v)$: $u = x_o + ax + by$, $v = y_o + cx + dy$, where $x_o, y_o$ are arbitrary and $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is any invertible 2×2 matrix (i.e. $ad - bc \neq 0$). A special case are *Euclidean coordinate*s, when the matrix is orthogonal. Domain: all $p$.

As we mentioned, these are some possible choices for coordinate systems on $\mathbb{R}^2$. The question is which to choose for the collection of coordinate systems required in MAN 1–3. For example, the Cartesian coordinate system $\{(x, y)\}$ by itself evidently satisfies the axioms MAN 1–3, hence this single coordinate system suffices to make $\mathbb{R}^2$ into a manifold. On the other hand, we can add the polar coordinates $(r, \theta)$, with domain $r > 0$, $0 < \theta < 2\pi$, say. So that we take as our collection of coordinates $\{(x, y), (r, \theta)\}$. We now have to check that the axioms are satisfied. The only thing which is not entirely clear is that the coordinate transformation $(x, y) \to (r, \theta) = (f^1(x, y), f^2(x, y))$ is smooth on its domain, as specified by MAN 2. First of all, the domain of this map consists of the $(x, y)$ off the positive $x$–axis $\{(x, y) \mid y = 0, x \geq 0\}$, which corresponds to $\{(r, \theta) \mid r > 0, 0 < \theta < 2\pi\}$. Thus this domain is open, and it is clear geometrically that $(x, y) \to (r, \theta)$ is a one–to–one correspondence between these sets. However, the map $(x, y) \to (r, \theta)$ this map is not given explicitly, but is defined implicitly as the inverse of the map $(\theta, r) \to (x, y) = (r\cos\theta, r\sin\theta)$. We might try to simply write down the inverse map as

$$r = f^1(x, y) = (x^2 + y^2), \quad \theta = f^2(x, y) = \arctan(y/x).$$

But this is not sufficient: as it stands, $f^2(x,y)$ is not defined when $x=0$, but some of these points belong to the domain of $(x,y) \to (r,\theta)$. It is better to argue like this. As already mentioned, we know that the inverse map $f$ exists on the domain indicated. Since the inverse of a map between two given sets is unique (if it exists) this map $f$ must coincide with the local inverse $F$ guaranteed by the Inverse Function Theorem whenever wherever are both are defined as maps between the same open sets. But the Inverse Function Theorem says also that $F$ is smooth where defined. Hence $f$ is then smooth as well at points where some such $F$ can be found, i.e. at all points of its domain where the Jacobian determinant $\partial(x,y)/\partial(r,\theta) \neq 0$ i.e. $r \neq 0$, and this is includes all points in the domain of $f$. Hence all three axioms are satisfied for $M = \mathbb{R}^2$ together with $\{(x,y),(r,\theta)\}$.

Perhaps I belabored too much the point of verifying MAN 2 for the implicitly defined map $(x,y) \to (r,\theta)$, but it illustrates a typical situation and a typical argument. In the future I shall omit these details. There is another important point. Instead of specifying the domain of the coordinates $(r,\theta)$ by $r>0, 0<\theta<2\pi$, it is usually sufficient to know that the equations $x = r\cos\theta$, $y = r\sin\theta$ can be used to specify coordinates in some neighbourhood of any point $(x,y)$ where $r \neq 0$, as guaranteed by the Inverse Function Theorem. For example, we could admit all of these local specifications of $(r,\theta)$ among our collection of coordinates (without specifying explicitly the domains), and then all of the axioms are evidently satisfied. This procedure is actually most appropriate, since it brings out the flexibility in the choice of the coordinates. It is also the procedure implicitly followed in calculus, where one ignores the restrictions $r>0$, $0<\theta<2\pi$ as soon as one discusses a curve like $r = \sin 2\theta$.

We could add some or all of the other coordinates defined above and check again that the axioms still hold and we face again the question which to choose. One naturally has the feeling that it should not matter, but strictly speaking, according to the definition, we would get a different manifold structure on the point set $\mathbb{R}^2$ for each choice. We come back to this in a moment, but first I briefly look at some more examples, without going through all the detailed verifications, however.

1.2.**3** **Example: Cartesian 3-space** $\mathbb{R}^3 = \{(x,y,z) \mid x,y,z \in \mathbb{R}\}$. I only write down the formulas and leave the discussion to you, in particular the determination of the coordinate domains.

*Cartesian coordinates* $(x,y,z)$: $x(p) = x, y(p) = y, z(p) = z$.
*Cylindrical coordinates* $(r,\theta,z)$: $x = r\cos\theta$, $y = r\sin\theta$, $z = z$.
*Spherical coordinates* $(\rho,\theta,\phi)$: $x = \rho\cos\theta\,\sin\phi$, $y = \rho\sin\theta\sin\phi$, $z = \rho\cos\phi$.
*Hyperbolic coordinates* $(u,\theta,\psi)$: $x = u\cos\theta\,\sinh\psi$, $y = u\sin\theta\sinh\psi$, $z = u\cosh\psi$.
*Linear coordinates* $(u,\mathrm{v},w)$: $(u,\mathrm{v},w) = (x,y,z)A$, $A$ any invertible $3\times 3$ matrix. Special case: *orthogonal linear coordinates*: $AA^T = I$.
*Affine coordinates*: $(u,\mathrm{v},w) = (x_o,y_o,z_o) + (x,y,z)A$, $(x_o,y_o,z_o)$ any point, $A$ any invertible $3\times 3$ matrix. Special case: *Euclidean coordinates*: $AA^T = I$.

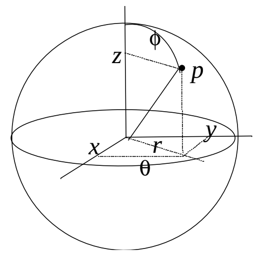



Fig. 2. Spherical coordinates

1.2.**4 Cartesian n-space** $\mathbb{R}^n = \{p = (x^1, \cdots, x^n) \mid x^i \in \mathbb{R}\}$.
*Linear coordinates* $(y^1, \cdots, y^n)$: $y^j = a^j_i x^i$, $\det(a^j_i) \neq 0$.
*Affine coordinates*: $(y^1, \cdots, y^n)$: $y^j = x^j_o + a^j_i x^i$, $\det(a^j_i) \neq 0$.
*Euclidean coordinates*: $(y^1, \cdots, y^n)$: $y^j = x^j_o + a^j_i x^i$, $\sum_k a^i_k a^j_k = \delta^{ij}$

It can be shown that the transformations $(x^i) \to (y^i)$ defining the Euclidean coordinates are the only ones which preserve Euclidean distance. They are called *Euclidean transformations*. Instead of $\mathbb{R}^n$ one can start out with any $n$–dimensional real vector space $V$. Given a basis $e_1, \cdots, e_n$ of $V$, define *linear coordinates* $(x^i)$ on $V$ by using the components $x^i = x^i(p)$ as coordinates$p = x^1 e_1 + \cdots x^n e_n$.This single coordinate system makes $V$ into a manifold. This manifold structure on $V$ does not depend on the basis in a sense to be explained after definition 1.2.8.

1.2.**5 The 2-sphere** $S^2 = \{p = (x, y, z) \mid x^2 + y^2 + z^2 = 1\}$. Note first of all that the triple $(x, y, z)$ itself cannot be used as a coordinate system on $S^2$. So we have to do something else.
*Parallel projection coordinates* $(x$,$y)$:

$$(x, y): \quad x = x,\ y = y,\ z = \sqrt{1 - x^2 - y^2}, \quad \text{Domain: } z > 0$$

Similar with the negative root for $z < 0$ and with $(x, y)$ replaced by $(x, z)$ or by $(y, z)$. This gives 6 coordinate systems, corresponding to the parallel projections onto the 3 coordinate planes and the 2 possible choices of sign $\pm$ in each case.
*Geographical coordinates* $(\theta, \phi)$: $x = \cos\theta \sin\phi$, $y = \sin\theta \sin\phi$, $z = \cos\phi$. *Domain*: $0 < \theta < 2\pi$, $0 < \phi < \pi$. (Other domains are possible.)
*Central projection coordinates* $(u, v)$

$$x = \frac{u}{\sqrt{1 + u^2 + v^2}},\ y = \frac{v}{\sqrt{1 + u^2 + v^2}},\ z = \frac{1}{\sqrt{1 + u^2 + v^2}}.\ \textit{Domain: } z > 0..$$

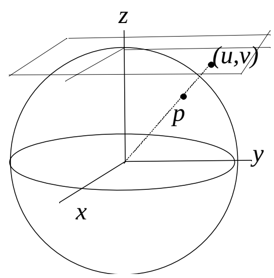



Fig. 3. Central projection

This is the central projection of the upper hemisphere $z > 0$ onto the plane $z = 1$. One could also take the lower hemisphere or replace the plane $z = 1$ by $x = 1$ or by $y = 1$. This gives again 6 coordinate systems. The 6 parallel projection coordinates by themselves suffice to make $S^2$ into a manifold, as do the 6 central projection coordinates. The geographical coordinates do not, even if one takes all possible domains for $(r, \theta)$: the north pole $(0, 0, 1)$ never lies in a coordinate domain. However, if one defines another geographical coordinate system on $S^2$ with a different north pole (e.g. by replacing $(x, y, z)$ by $(z, y, x)$ in the formula) one obtains enough geographical coordinates to cover all of $S^2$. All of the above coordinates could be defined with $(x, y, z)$ replaced by any orthogonal linear coordinate system $(u, v, w)$ on $\mathbb{R}^3$.

Let's now go back to the definition of manifold and try understand what it is trying to say. Compare the situation in analytic geometry. There one starts with some set of points which comes equipped with a distinguished *Cartesian coordinate system* $(x^1, \cdots, x^n)$, which associates to each point $p$ of $n$–space a coordinate $n$–tuple $(x^1(p), \cdots, x^n(p))$. Such a coordinate system is assumed fixed once and for all, so that we may as well the points to be the $n$–tuples, which means that our "Cartesian space" becomes $\mathbb{R}^n$, or perhaps some subset of $\mathbb{R}^n$, which better be open if we want to make sense out differentiable functions. Nevertheless, but one may introduce other *curvilinear coordinates* (e.g. polar coordinates in the plane) by giving the *coordinate transformation* to the Cartesian coordinates (e.g $x$ =rcos $\theta$, $y$ =rsin $\theta$). Curvilinear coordinates are not necessarily defined for all points (e.g. polar coordinates are not defined at the origin) and need not take on all possible values (e.g. polar coordinates are restricted by r$\geq 0$, $0 \leq \theta < 2\pi$). The requirement of a distinguished Cartesian coordinate system is of course often undesirable or physically unrealistic, and the notion of a *manifold* is designed to do away with this requirement. However, the axioms still require that there be *some* collection of coordinates. We shall return in a moment to the question in how far this collection of coordinates should be thought of as intrinsically distinguished.

The essential axiom is MAN 2, that any two coordinate systems should be related by a *differentiable* coordinate transformation (in fact even smooth, but this is a comparatively minor, technical point). This means that manifolds must locally look like $\mathbb{R}^n$ as far as a differentiable map can "see": all local properties of

$\mathbb{R}^n$ which are preserved by differentiable maps must apply to manifolds as well. For example, some sets which one should expect to turn out to be manifolds are the following. First and foremost, any open subset of some $\mathbb{R}^n$, of course; smooth curves and surfaces (e.g. a circle or a sphere); the set of all rotations of a Euclidean space (e.g. in two dimensions, a rotation is just by an angle and this set looks just like a circle). Some sets one should not expect to be manifolds are: a half–line (with an endpoint) or a closed disk in the plane (because of the boundary points); a figure–8 type of curve or a cone with a vertex; the set of the possible rotations of a steering wheel (because of the "singularity" when it doesn't turn any further. But if we omit the offending singular points from these non–manifolds, the remaining sets should still be manifolds.) These example should give some idea of what a manifold is supposed to be, but they may be misleading, because they carry additional structure, in addition to their manifold structure. For example, for a smooth surface in space, such as a sphere, we may consider the length of curves on it, or the variation of its normal direction, but these concepts (length or normal direction) do not come from its manifold structure, and do not make sense on an arbitrary manifold. A manifold (without further structure) is an amorphous thing, not really a space with a geometry, like Euclid's. One should also keep in mind that a manifold is not completely specified by naming a set of points: one also has to specify the coordinate systems one considers. There may be many natural ways of doing this (and some unnatural ones as well), but to start out with, the coordinates have to be specified in some way.

Let's think a bit more about MAN 2 by recalling the meaning of "differentiable": a map is differentiable if it can be approximated by a linear map to first order around a given point. We shall see later that this imposes a certain kind of structure on the set of points that make up the manifolds, a structure which captures the idea that a manifold can in some sense be approximated to first order by a linear space. "Manifolds are linear in infinitesimal regions" as classical geometers would have said.

One should remember that the definition of "differentiable" requires that the function in question be defined in a whole neighbourhood of the point in question, so that one may take limits from any direction: the domain of the function must be open. As a consequence, the axioms involve "openness" conditions, which are not always in agreement with the conventions of analytic geometry. (E.g. polar coordinates must be restricted by $r > 0$, $0 < \theta < 2\pi$; −strict inequalities!). I hope that this lengthy discussion will clarify the definition, although I realize that it may do just the opposite.

As you can see, the definition of "manifold" is really very simple, much shorter than the lengthy discussion around it; but I think you will be surprised at the amount of structure hidden in the axioms. The first item is this.

1.2.**6 Definition**. A *neighbourhood* of a point $p_o$ in $M$ is any subset of $M$ containing all $p$ whose coordinate points $x(p)$ in some coordinate system satisfy $\|x(p) - x(p_o)\| < \epsilon$ for some $\epsilon > 0$. A subset of $M$ is open if it contains a neighbourhood of each of its points, and closed if its complement is open.

This definition makes a manifold into what is called a topological space: there is a notion of "neighbourhood" or (equivalently) of "open set". An open subset $U$ of $M$, together with the coordinate systems obtained by restriction to $U$, is again an $n$–dimensional manifold.

1.2.**7 Definition**. A map $F : M \to N$ between manifolds is *of class* $\mathrm{C}^k$, $0 \leq k \leq \infty$, if $F$ maps any sufficiently small neighbourhood of a point of $M$ into the domain of a coordinate system on $N$ and the equation $q = F(p)$ defines a $\mathrm{C}^k$ map when $p$ and $q$ are expressed in terms of coordinates:

$$y^j = F^j(x^1, \cdots, x^n),\ j = 1, \cdots, m.$$

We can summarize the situation in a *commutative diagram* like this:

$$\begin{array}{ccc} M & \xrightarrow{F} & N \\ (x^i)\downarrow & & \downarrow(y^j) \\ \mathbb{R}^n & \underset{(F^j)}{\longrightarrow} & \mathbb{R}^m \end{array}$$

This definition is independent of the coordinates chosen (by axiom MAN 2) and applies equally to a map $F : M \cdots \to N$ defined only on an open subset of $N$. A smooth map $F : M \to N$ which is bijective and whose inverse is also smooth is called a *diffeomorphism* of $M$ onto $N$.

Think again about the question if the axioms capture the idea of a manifold as discussed above. For example, an open subset of $\mathbb{R}^n$ together with its single Cartesian coordinate system is a manifold in the sense of the definition above, and if we use another coordinate system on it, e.g. polar coordinates in the plane, we would strictly speaking have another manifold. But this is not what is intended. So we extend the notion of "coordinate system" as follows.

1.2.**8 Definition**. A *general coordinate system* on $M$ is *any* diffeomorphism from an open subset $U$ of $M$ onto an open subset of $\mathbb{R}^n$.

These *general* coordinate systems are admitted on the same basis as the coordinate systems with which $M$ comes equipped in virtue of the axioms and will just be called "coordinate systems" as well. In fact we shall identify manifold structures on a set $M$ which give the same *general* coordinates, even if the collections of coordinates used to define them via the axioms MAN 1–3 were different. Equivalently, we can define a manifold to be a set $M$ together with all *general* coordinate systems corresponding to *some* collection of coordinates satisfying MAN 1–3. These *general* coordinate systems form an atlas which is *maximal*, in the sense that it is not contained in any strictly bigger atlas. Thus we may say that a manifold is a set together with a maximal atlas; but since any atlas can always be uniquely extended to a maximal one, consisting of the *general* coordinate systems, any atlas determines the manifold structure. That there are always plenty of coordinate systems follows from the inverse function theorem:

1.2.**9 Theorem**. *Let $x$ be a coordinate system in a neighbourhood of $p_o$ and $f : \mathbb{R}^n \cdots \to \mathbb{R}^n$ $\tilde{x}^i = f^i(x^1, \cdots, x^n)$, a* smooth *map defined in a neighbourhood* *of* $x_o = x(p_o)$ *with* $\det(\partial\tilde{x}^i/\partial x^j)_{x_o} \neq 0$. *Then the equation* $\tilde{x}(p) = f(x(p))$ *defines another coordinate system in a neighbourhood of* $p_o$.

In the future we shall use the expression "a coordinate system *around* $p_o$" for "a coordinate system defined in some neighbourhood of $p_o$".
A map $F : M \to N$ is a *local diffeomorphism* at a point $p_o \in M$ if $p_o$ has an open neighbourhood $U$ so that $F|U$ is a diffeomorphism of $U$ onto an open neighbourhood of $F(p_o)$. The inverse function theorem says that this is the case if and only if $\det(\partial F^j/\partial x^i) \neq 0$ at $p_o$. The term local diffeomorphism by itself means "local diffeomorphism at all points".

1.2.**10** **Examples**. Consider the map $F$ which wraps the real line $\mathbb{R}^1$ around the unit circle $S^1$. If we realize $S^1$ as the unit circle in the complex plane, $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, then this map is given by $F(x) = e^{ix}$. In a neighbourhood of any point $z_o = e^{ix_o}$ of $S^1$ we can write $z \in S^1$ uniquely as $z = e^{ix}$ with $x$ sufficiently close to $x_o$, namely $|x - x_o| < 2\pi$. We can make $S^1$ as a manifold by introducing these locally defined maps $z \to x$ as coordinate systems. Thus for each $z_o \in S^1$ fix an $x_o \in \mathbb{R}$ with $z_o = e^{ix_o}$ and then define $x = x(z)$ by $z = e^{ix}$, $|x - x_o| < 2\pi$, on the domain of $z$'s of this form. (Of course one can cover $S^1$ with already two of such coordinate domains.) In these coordinates the map $x \to F(x)$ is simply given by the formula $x \to x$. But this does not mean that $F$ is one–to–one; obviously it is not. The point is that the formula holds only *locally*, on the coordinate domains. The map $F : \mathbb{R}^1 \to S^1$ is not a diffeomorphism, only a local diffeomorphism.

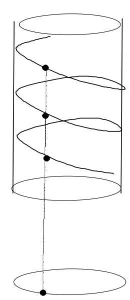


Fig.4. The map $\mathbb{R}^1 \to S^1$

1.2.**11** **Definition**. If $M$ and $N$ are manifolds, the *Cartesian product* $M \times N = \{p, q) \mid p \in M, q \in N\}$ becomes a manifold with the coordinate systems $(x^1, \cdots, x^n, y^1, \cdots, y^m)$ where $(x^1, \cdots, x^n)$ is a coordinate system of $M$, $(y^1, \cdots, y^m)$ for $N$.

The definition extends in an obvious way for products of more than two manifolds. For example, the product $\mathbb{R}^1 \times \cdots \times \mathbb{R}^1$ of $n$ copies of $\mathbb{R}^1$ is $\mathbb{R}^n$; the product $S^1 \times \cdots \times S^1$ of $n$ copies of $S^1$ is denoted $\mathrm{T}^n$ and is called the $n$–torus. For $n = 2$ it can be realized as the doughnut–shaped surface by this name. The maps $\mathbb{R}^1 \to S^1$ combine to give a local diffeomorphism of the product manifolds $F : \mathbb{R}^1 \times \cdots \times \mathbb{R}^1 \to S^1 \times \cdots \times S^1$.

1.2.**12 Example: Nelson's car.** A rich source of examples of manifolds are configuration spaces from mechanics. Take the configuration space of a car, for example. According to Nelson, (Tensor Analysis, 1967, p.33), "*the configuration space of a car is the four dimensional manifold* $M = \mathbb{R}^2 \times T^{\mathbf{2}}$ *parametrized by* $(x, y, \theta, \phi)$*, where* $(x, y)$ *are the Cartesian coordinates of the center of the front axle, the angle* $\theta$ *measures the direction in which the car is headed, and* $\phi$ *is the angle made by the front wheels with the car. (More realistically, the configuration space is the open subset* $-\theta_{\max} < \theta < \theta_{\max}$*.)*" (Note the original design of the steering mechanism.)

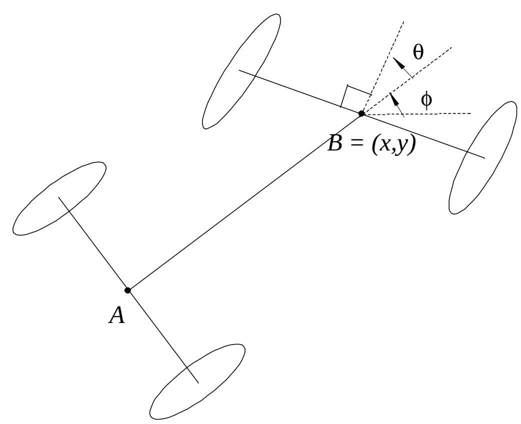



Fig.5. Nelson's car

A *curve* in $M$ is an $M$−valued function $\mathbb{R} \cdots \to M$, written $p = p(t))$, defined on some interval. Exceptionally, the domain of a curve is not always required to be open. The curve is of class $\mathrm{C}^k$ if it extends to such a function also in a neighbourhood of any endpoint of the interval that belongs to its domain. In coordinates $(x^i)$ a curve $p = p(t)$ is given by equations $x^i = x^i(t)$. A manifold is *connected* if any two of its point can be joined by a continuous curve.

1.2.**13 Lemma and definition**. *Any* $n$*-dimensional manifold is the disjoint union of* $n$*-dimensional connected subsets of* $M$*, called the* connected components *of* $M$*.*

**Proof**. Fix a point $p \in M$. Let $M_p$ be the set of all points which can be joined to $p$ by a continuous curve. Then $M_p$ is an open subset of $M$ (exercise). Two such $M_p$'s are either disjoint or identical. Hence $M$ is the disjoint union of the distinct $M_p$'s. □

In general, a topological space is called *connected* if it cannot be decomposed into a disjoint union of two non–empty open subsets. The lemma implies that for manifolds this is the same as the notion of connectedness defined in terms of curves, as above.

1.2.**14 Example**. The sphere in $\mathbb{R}^3$ is connected. Two parallel lines in $\mathbb{R}^2$ constitute a manifold in a natural way, the disjoint union of two copies of $\mathbb{R}$, which are its connected components.
Generally, given any collection of manifolds of the same dimension $n$ one can make their disjoint union again into an $n$–dimensional manifold in an obvious way. If the individual manifolds happen not to be disjoint as sets, one must keep them disjoint, e.g. by imagining the point $p$ in the $i$–th manifold to come with a label telling to which set it belongs, like $(p, i)$ for example. If the individual manifolds are connected, then they form the connected components of the disjoint union so constructed. Every manifold is obtained from its connected components by this process. The restriction that the individual manifolds have the same dimension $n$ is required only because the definition of "$n$–dimensional manifold " requires a definite dimension.

We conclude this section with some remarks on the definition of manifolds. First of all, instead of $C^\infty$ functions one could take $C^k$ functions for any $k \geq 1$, or real analytic functions (convergent Taylor series) without significant changes. One could also take $C^0$ functions (continuous functions) or holomorphic function (complex analytic functions), but then there are significant changes in the further development of the theory.
There are many other equivalent definitions of manifolds. For example, the definition is often given in two steps, first one defines (or assumes known) the notion of topological space, and then defines manifolds. One can then require the charts to be homeomorphisms from the outset, which simplifies the statement of the axioms slightly. But this advantage is more than offset by the inconvenience of having to specify separately topology and charts, and verify the homeomorphism property, whenever one wants to define a particular manifold. It is also inappropriate logically, since the atlas automatically determines the topology. Two more technical axioms are usually added to the definition of manifold.

MAN **4** (Hausdorff Axiom). Any two points of $M$ have disjoint neighbourhoods.
MAN **5** (Countability Axiom). $M$ can be covered by *countably many* coordinate balls.

The purpose of these axioms is to exclude some pathologies; all examples we have seen or shall see satisfy these axioms (although it is easy to artificially construct examples which do not). We shall not need them for most of what we do, and if they are required we shall say so explicitly.
As a minor notational point, some insist on specifying explicitly the domains of functions, e.g. write $(U, \phi)$ for a coordinate system or $f : U \to N$ for a partially defined map; but the domain is rarely needed and $f : M \cdots \to N$ is usually more informative.

**EXERCISES 1.2**

1. (a) Specify domains for the coordinates on $\mathbb{R}^3$ listed in example 1.2.3.
(b) Specify the domain of the coordinate transformation $(x, y, z) \to (\rho, \theta, \phi)$ and prove that this map is smooth.

2. Is it possible to specify domains for the following coordinates on $\mathrm{R}^3$ so that the axioms MAN 1–3 are satisfied?

(a) The coordinates are cylindrical and spherical coordinates as defined in example 1.2.3. You may use several domains with the same formula, e.g use several restrictions on r and/or $\theta$ to define several cylindrical coordinates $(r, \theta, z)$ with different domains.

(b) The coordinates are spherical coordinates $(\rho, \theta, \phi)$ but with the origin translated to an arbitrary point $(x_o, y_o, z_o)$. (The transformation equations between $(x, y, z)$ and $(\rho, \theta, \phi)$ are obtained from those of example 1.2.3 by replacing $x, y, z$ by $x - x_o, y - y_o, z - z_o$, respectively.) You may use several such coordinate systems with different $(x_o, y_o, z_o)$.

3. Verify that the parallel projection coordinates on $S^2$ satisfy the axioms MAN 1–3. [See example 1.2.5. Use the notation $(U, \phi)$, $(\tilde{U}, \tilde{\phi})$ the coordinate systems in order not to get confused with the $x$–coordinate in $\mathbb{R}^3$. Specify the domains $U, \tilde{U}, U \bigcap \tilde{U}$ etc. as needed to verify the axioms.]

4. (a)Find a map $F : \mathrm{T}^2 \to \mathbb{R}^3$ which realizes the 2–torus $\mathrm{T}^2 = S^1 \times S^1$ as the doughnut–shaped surface in 3–space $\mathrm{R}^3$ referred to in example 1.2.11. Prove that $F$ is a smooth map. (b) Is $\mathrm{T}^2$ is connected? (*Prove* your answer.) (c)What is the *minimum* number of charts in an atlas for $\mathrm{T}^2$? (Prove your answer and exhibit such an atlas. You may use the fact that $\mathrm{T}^2$ is compact and that the image of a compact set under a continuous map is again compact; see your several–variables calculus text, e.g. Marsden–Tromba.)

5. Show that as $(x^i)$ and $(y^j)$ run over a collections of coordinate systems for $M$ and $N$ satisfying MAN 1–3, the $(x^i, y^j)$ do the same for $M \times N$.

6. (a) Imagine two sticks flexibly joined so that each can rotate freely about the about the joint. (Like the *nunchucks* you know from the Japanese martial arts.) The whole contraption can move about in space. (These ethereal sticks are not bothered by hits: they can pass through each other, and anything else. Not useful as a weapon.) Describe the configuration space of this mechanical system as a direct product of some of the manifolds defined in the text. Describe the subset of configurations in which the sticks are not in collision and show that it is open in the configuration space.

(b) Same problem if the joint admits only rotations at right angles to the "grip stick" to whose tip the "hit stick" is attached. (Wind–mill type joint; good for lateral blows. Each stick now has a tip and a tail.)

[Translating all of this into something precise is part of the problem. Is the configuration space a manifold at all, in a reasonable way? If so, is it connected or what are its connected components? If you find something unclear, add precision as you find necessary; just explain what you are doing. Use sketches.]

7. Let $M$ be a set with two manifold structures and let $M'$, $M''$ denote the corresponding manifolds.
(a) Show that $M' = M''$ *as manifolds* if and only if the identity map on $M$ is smooth as a map $M' \to M''$ and as a map $M'' \to M'$. [Start by stating the definition what it means that "$M' = M''$ *as manifolds*".]
(b) Suppose $M'$, $M''$ have the same open sets and each open set $U$ carries the same collection of smooth functions $f : U \to \mathbb{R}$ for $M'$ and for $M''$. Is then $M' = M''$ as manifold? (Proof or counterexample.)

8. Specify a collection $\{(U_\alpha, \phi_\alpha)\}$ of partially defined maps $\phi_\alpha : U_\alpha \subset \mathbb{R} \to \mathbb{R}$ as follows.
(a) $U = \mathbb{R}$, $\phi(t) = t^3$, (b) $U_1 = \mathbb{R} - \{0\}$, $\phi_1(t) = t^3$, $U_2 = (-1, 1)$, $\phi_2(t) = 1/1-t$.
Determine if $\{(U_\alpha, \phi_\alpha)\}$ is an atlas (i.e. satisfies MAN $1-3$). If so, determine if the corresponding manifold structure on $\mathbb{R}$ is the same as the usual one.

9. (a) Let $S$ be a subset of $\mathbb{R}^n$ for some $n$. Show that there is at most one manifold structure on $S$ so that a partially defined map $\mathbb{R}^k \cdots \to S$ (any $k$) is smooth if and only if $\mathbb{R}^k \cdots \to S \xrightarrow{\subset} \mathbb{R}^n$ is smooth. [Suggestion. Write $S', S''$ for $S$ equipped with two manifold structures. Show that the identity map $S' \to S''$ is smooth in both directions. ]
(b) Show that (a) holds for the usual manifold structure on $S = \mathbb{R}^m$ considered as subset of $\mathbb{R}^n$ $(m \leq n)$.

10. (a) Let $\mathrm{P}^1$ be the set of all one–dimensional subspaces of $\mathbb{R}^2$ (lines through the origin). Write $\langle x \rangle = \langle x_1, x_2 \rangle$ for the line through $x = (x_1, x_2)$. Let $U_1 = \{\langle x_1, x_2 \rangle : x_1 \neq 0\}$ and define $\phi_1 : U_1 \to \mathbb{R}$, $\langle x \rangle \to x_2/x_1$. Define $(U_2, \phi_2)$ similarly by interchanging $x_1$ and $x_2$ and prove that $\{(U_1, \phi_1), (U_2, \phi_2)\}$ is an atlas for $\mathrm{P}^1$ (i.e. MAN 1–3 are satisfied.) Explain why the map $\phi_1$ can be viewed as taking the intersection with the line $x_1 = 1$. Sketch.
(b) Generalize part (a) for the set of one–dimensional subspaces $\mathrm{P}^n$ of $\mathbb{R}^{n+1}$. [Suggestion. Proceed as in part (a): the line $x_1 = 1$ in $\mathbb{R}^2$ is now replaced by the $n$–plane in $\mathbb{R}^{n+1}$ given by this equation. Consider the other coordinate $n$–planes $x_i = 0$ as well. Sketching will be difficult for $n > 2$. The manifold $\mathrm{P}^n$ is called (real) projective $n$–space.]

11. Let $\mathrm{P}^n$ be the set of all one–dimensional subspaces of $\mathbb{R}^{n+1}$ (lines through the origin). Let $F : S^n \to \mathbb{R}^n$ be the map which associates to each $p \in S^n$ the line $\langle p \rangle = \mathbb{R}p$.
(a) Show that $\mathrm{P}^n$ admits a unique manifold structure so that the map $F$ is a local diffeomorphism.
(b)Show that the manifold structure on $\mathrm{P}^n$ defined in problem 10 is the same as the one defined in part (a).

12. Generalize problem 10 to the set $\mathrm{G}_{k,n}$ of $k$–dimensional subspaces of $\mathbb{R}^n$. [Suggestion. Consider the map $\phi$ which intersects a $k$–dimensional subspace $P \in \mathrm{G}_{k,n}$ with the *coordinate* $(n-k)$*–plane* with equation $x_1 = x_2 = \cdots = x_k = 1$ and similar maps using other coordinate $(n-k)$–planes of this type. The manifolds $\mathrm{G}_{k,n}$ are called *(real) Grassmannian manifolds*.) ]

13. Let $M$ be a manifold, $p \in M$ a point of $M$. Let $M_p$ be the set of all points which can be joined to $p$ by a continuous curve. Show that $M_p$ is an open subset of $M$.

14. Define an $n$–dimensional *linear manifold* to be a set $L$ together with a maximal atlas of charts $L \to \mathbb{R}^n$, $p \to x(p)$ which are everywhere defined bijections and any two of which are related by an invertible *linear* transformation $\tilde{x}^j = a^j_i x^i$. Show that every vector space is in a natural way a linear manifold and vice versa. [Thus a linear manifold is really the same thing as a vector space. The point is that vector spaces could be defined in a way analogous to manifolds.]

15. (a) Define a notion of *affine manifold* in analogy with (a) using affine coordinate transformations as defined in 1.2.7. Is every vector space an affine manifold in a natural way? How about the other way around?
(b) Define an affine space axiomatically as follows.
**Definition**. An $n$–dimensional *affine space* consists of set $A$ of *points* together with a set of *vectors* which form an $n$–dimensional vector space $V$.
**ASP 1**. Any two points $p, q$ in $A$ determine a vector $\vec{pq}$ in $V$.
**ASP 2**. Given a point $p \in A$ and a vector $v \in V$ there is a unique point $q \in A$ so that $v = \vec{pq}$.
**ASP 3**. For any three points $a, b, c \in A$, $\vec{ab} + \vec{bc} = \vec{ac}$.
Show that every affine manifold space is in a natural way an affine space and vice versa. [Thus an affine manifold is really the same thing as an affine space.]

16. Give further examples of some types of "spaces" one can define in analogy with smooth manifolds, like those in the previous two problems. In each case, discuss if this type of space is or is not a manifold in a natural way. [Give at least one example that is and one that isn't. You *do* know very natural examples of such "spaces".]

## 1.3 Vectors and differentials

The notion of a "vector" on a manifold presents some conceptual difficulties. We start with some preliminary remarks. We may paraphrase the definition of "manifold" by saying that a manifold can locally be identified with $\mathbb{R}^n$, but this identification is determined only up to diffeomorphism. Put this way, it is evident that anything which can be defined for $\mathbb{R}^n$ by a definition which is local and preserved under diffeomorphism can also be defined for manifolds. The definition of "smooth function" is a first instance of this principle, but there are many others, as we shall see. One of them is a notion of "vector", but there

are already several notions of "vector" connected with $\mathbb{R}^n$, and we should first make clear which one we want to generalize to manifolds: it is the notion of a tangent vector to a differentiable curve at a given point.
Consider a differentiable curve $p(t)$ on a manifold $M$. Fix momentarily a coordinate system $(x^i)$ defined around the points $p(t)$ for $t$ in some open interval around some $t = t_o$. For each $t$ in this interval the coordinates $x^i(t)$ of $p(t)$ are differentiable functions of $t$ and we can consider the $n$–tuple of derivatives $(\dot{x}^i(t))$. It depends not only on the curve $p(t)$ but also on the coordinate system $(x^i)$ and to bring this out we call the $n$–tuple $(\xi^i) = (\dot{x}^i(t_o))$ the *coordinate tangent vector* of $p(t)$ at $p_o$ (more accurately one should say "at $t_o$") relative to the coordinate system $(x^i)$. If $\tilde{x}^j = \tilde{x}^j(p)$ is another coordinate system defined around $p(t_o)$, then we can write $\tilde{x}(p) = f(x(p))$ by the coordinate transformation as in MAN 2, p 24 and by the chain rule the two coordinate tangent vectors $(\tilde{\xi}^i)$ and $(\xi^i)$ are related by

$$\tilde{\xi}^j = \Big(\frac{\partial \tilde{x}^j}{\partial x^i}\Big)_{p_o} \xi^i.$$

From a logical point of view, this transformation law is the only thing that matters and we shall just take it as an axiomatic definition of "vector at $p_o$". But one might still wonder, what a vector "really" is: this is a bit similar to the question what a number (say a real number or even a natural number) "really" is. One can answer this question by giving a constructive definition, which defines the concept in terms of others (ultimately in terms of concepts defined axiomatically, in set theory, for example), but they are artificial in the sense that these definitions are virtually never used directly. What matters are certain basic properties, for vectors just as for numbers; the construction is of value only in so far as it guarantees that the properties postulated are consistent, which is not in question in the simple situation we are dealing with here. Thus we ignore the metaphysics and simply give an axiomatic definition.

1.3.**1 Definition**. A *tangent vector* (or simply *vector*) $v$ at $p \in M$ is a quantity which relative to a coordinate system $(x^i)$ around $p$ is represented by an $n$–tuple $(\xi^1, \cdots, \xi^n)$. The $n$–tuples $(\xi^i)$ and $(\tilde{\xi}^j)$ representing $v$ relative to two coordinate different systems $(x^i)$ and $(\tilde{x}^j)$ are related by the transformation law

$$\tilde{\xi}^j = \Big(\frac{\partial \tilde{x}^j}{\partial x^i}\Big)_{p} \xi^i. \tag{1}$$

**Explanation**. To say that a vector is *represented* by an $n$–tuple relative to a coordinate system means that a vector associates to a coordinate system an $n$–tuple and two vectors are considered identical if they associate the same $n$–tuple to any one coordinate system. If one still feels uneasy about the word "quantity" (which just stands for "something") and prefers that the objects one deals with should be defined constructively in terms of other objects, then one can simply take a vector at $p_o$ to be a rule (or "function") which to $(p_o, (x^i))$ associates $(\xi^i)$ subject to (1). This is certainly acceptable logically, but awkward conceptually. There are several other constructive definitions designed to produce more palatable substitutes, but the axiomatic definition by the transformation rule, though the oldest, exhibits most clearly the general principle about manifolds and $\mathbb{R}^n$ mentioned and is really quite intuitive because of its relation to tangent vectors of curves. In any case, it is essential to remember that the $n$–tuple $(\xi^i)$ depends not only on the vector $v$ at $p$, but also on the coordinate system $(x^i)$. Furthermore, by definition, a vector is "tied to a point": vectors at different points which happen to be represented by the same $n$–tuple in one coordinate system will generally be represented by different $n$–tuples in another coordinate system, because the partials in (1) depend on the point $p_o$.

Any differentiable curve $p(t)$ in $M$ has a tangent vector $\dot{p}(t_o)$ at any of its points $p_o = p(t_o)$: this is the vector represented by its coordinate tangent vector $(\dot{x}^i(t_o))$ relative to the coordinate system $(x^i)$. In fact, every vector $v$ at a point $p_o$ is the tangent vector $\dot{p}(t_o)$ to some curve; for example, if $v$ is represented by $(\xi^i)$ relative to the coordinate system $(x^i)$ then we can take for $p(t)$ the curve defined by $x^i(t) = x^i_o + t\xi^i$ and $t_o = 0$. There are of course many curves with the same tangent vector $v$ at a given point $p_o$, since we only require $x^i(t) = x^i_o + t\xi^i + o(t)$.

**Summary**. Let $p_o \in M$. Any differentiable curve $p(t)$ through $p_o = p(t_o)$ determines a vector $v = \dot{p}(t_o)$ at $p_o$ and every vector at $p_o$ is of this form. Two such curves determine the same vector if and only if they have the same coordinate tangent vector at $p_o$ in some (and hence in any) coordinate system around $p_o$.

One could evidently *define* the notion of a vector at $p_o$ in terms of differentiable curves: a vector at $p_o$ can be taken as an equivalence class of differentiable curves through $p_o$, two curves being considered "equivalent" if they have the same coordinate tangent vector at $p_o$ in any coordinate system.

Vectors *at the same point* $p$ may be added and scalar multiplied by adding and scalar multiplying their coordinate vectors relative to coordinate systems. In this way the vectors at $p \in M$ form a vector space, called the tangent space of $M$ at $p$, denoted $T_pM$. The fact that the set of all tangent vectors of differentiable curves through $p_o$ forms a vector space gives a precise meaning to the idea that manifolds are linear in "infinitesimal regions".

Relative to a coordinate system $(x^i)$ around $p \in M$, vectors $v$ at $p$ are represented by their coordinate vectors $(\xi^i)$. The vector with $\xi^k = 1$ and all other components $= 0$ relative to the coordinate system $(x^i)$ is denoted $(\partial/\partial x^k)_p$ and will be called the $k$–th *coordinate basis vector* at $p$. It is the tangent vector $\partial p(x)/\partial x^k$ of the k–th c*oordinate line* through $p$, i.e. of the curve $p(x^1, x^2, \cdots, x^n)$ parametrized by $x^k = t$, the remaining coordinates being given their value at $p$. The $(\partial/\partial x^k)_p$ form a basis for $T_pM$ since every vector $v$ at $p$ can be uniquely written as a linear combination of the $(\partial/\partial x^k)_p$: $v = \xi^k(\partial/\partial x^k)_p$.The components $\xi^k$ of a vector at $p$ with respect to this basis are just the $\xi^k$ representing $v$ relative to the coordinate system $(x^k)$ according to the definition of "vector at $p$" and are sometimes called the components of $v$ relative to the coordinate system $(x^i)$. It is important to remember that $\partial/\partial x^1$ (for example) depends on *all* of the coordinates $(x^1, \cdots, x^n)$, not just on the coordinate function $x^1$.

1.3.**2 Example**: vectors on $\mathbb{R}^2$. Let $M = \mathbb{R}^2$ with Cartesian coordinates $x, y$. We can use this coordinate system to represent any vector $v \in T_p\mathbb{R}^2$ at any point $p \in \mathbb{R}^2$ by its pair of components $(a, b)$ with respect to this coordinate system: $v = a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}$. The distinguished Cartesian coordinate system $(x, y)$ on $\mathbb{R}^2$ therefore allows us to identify all tangent spaces $T_p\mathbb{R}^2$ again with $\mathbb{R}^2$, which explains the somewhat confusing interpretation of pairs of numbers as points or as vectors which one encounters in analytic geometry.
Now consider polar coordinates $r, \theta$, defined by $x = r\cos\theta$, $y = r\sin\theta$. The coordinate basis vectors $\partial/\partial r$, $\partial/\partial\theta$ are $\partial p/\partial r$, $\partial p/\partial\theta$ where we think of $p = p(r, \theta)$ as a function of $r, \theta$ and use the partial to denote the tangent vectors to the coordinate lines $r \to p(r, \theta)$ and $\theta \to p(r, \theta)$. We find

$$\begin{aligned}\frac{\partial}{\partial r} &= \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y} = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y},\\ \frac{\partial}{\partial\theta} &= \frac{\partial x}{\partial\theta}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\theta}\frac{\partial}{\partial y} = -r\sin\theta\frac{\partial}{\partial x} + r\cos\theta\frac{\partial}{\partial y}.\end{aligned}$$

The components of the vectors on the right may be expressed in terms of $x, y$ as well, e.g. using $r = \sqrt{x^2 + y^2}$ and $\theta = \arctan(y/x)$, on the domain where these formulas are valid.

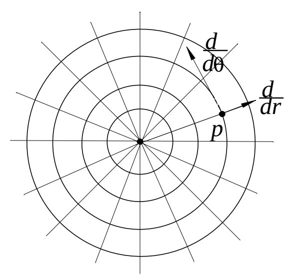



Fig. 1. Coordinate vectors in polar coordinates.

1.3.**3 Example: tangent vectors to a vector space as manifold**. In $\mathbb{R}^n$ can one identify vectors at different points, namely those vectors that have the same components in the Cartesian coordinate system. Such vectors will also have the same components in any affine coordinate system (exercise). Thus we can identify $T_p\mathbb{R}^n = \mathbb{R}^n$ for any $p \in \mathbb{R}^n$. More generally, let $V$ be any $n$–dimensional real vector space. Choose a basis $e_1, \cdots, e_n$ for $V$ and introduce linear coordinates $(x^i)$ in $V$ by using the components $x^i$ of $p \in V$ as coordinates:

$$p = x^1 e_1 + \cdots + x^n e_n.$$

With any tangent vector $v \in T_pV$ to $V$ as manifold, represented by an $n$–tuple $(\xi^i)$ relative to the coordinate system $(x^i)$ according to definition 1.3.1, associate an element $\vec{v} \in V$ by the formula

$$\vec{v} = \xi^1 e_1 + \cdots + \xi^n e_n. \tag{2}$$

Thus we have a map $T_pV \to V$, $v \to \vec{v}$, defined in terms of the linear coordinates $(x^i)$. This map is in fact independent of these coordinates (exercise). So for any $n$–dimensional real vector space $V$, considered as a manifold, we may identify $T_pV = V$ for any $p \in V$.

1.3.**4 Example: tangent vectors to a sphere**. Consider 2–sphere $S^2$ as a manifold, as in §2. Let $p = p(t)$ be a curve on $S^2$. It may also be considered as a curve in $\mathbb{R}^3$ which happens to lie on $S^2$. Thus the tangent vector $\dot{p}(t_o)$ at $p_o = p(t_o)$ can be considered as an element in $T_{p_o}S^2$ as well as of $T_{p_o}\mathbb{R}^3$. This means that we can think of $T_{p_o}S^2$ as a subspace of $T_{p_o}\mathbb{R}^3 \approx \mathbb{R}^3$. If we use geographical coordinates $(\theta, \phi)$ on $S^2$, so that the inclusion $S^2 \to \mathbb{R}^3$, is given $(\theta, \phi) \to (x, y, z)$,

$$(\theta, \phi) \to (x, y, z) = (\cos\theta \sin\phi,\ \sin\theta\sin\phi,\ \cos\phi) \tag{3}$$

then the inclusion map $T_pS^2 \to T_p\mathbb{R}^3$ at a general point $p = p(\theta, \phi) \in S^2$, is found by differentiation along the $\theta\phi$–coordinate lines (exercise)

$$\begin{aligned}\frac{\partial}{\partial\theta} &\to \frac{\partial x}{\partial\theta}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\theta}\frac{\partial}{\partial y} + \frac{\partial z}{\partial\theta}\frac{\partial}{\partial z} = -\sin\theta\sin\phi\ \frac{\partial}{\partial x} + \cos\theta\sin\phi\ \frac{\partial}{\partial y}\\ \frac{\partial}{\partial\phi} &\to \frac{\partial x}{\partial\phi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\phi}\frac{\partial}{\partial y} + \frac{\partial z}{\partial\phi}\frac{\partial}{\partial z} = \cos\theta\cos\phi\ \frac{\partial}{\partial x} + \sin\theta\cos\phi\ \frac{\partial}{\partial y} - \sin\phi\ \frac{\partial}{\partial z}\end{aligned} \tag{4}$$

In this way $T_pS^2$ is identified with the subspace of $T_p\mathbb{R}^3 \approx \mathbb{R}^3$ given by

$$T_pS^2 \approx \{v \in T_p\mathbb{R}^3 = \mathbb{R}^3 : v \cdot p = 0\}. \tag{5}$$

(exercise).

1.3.**5 Definition**. A *vector field* $X$ on $M$ associates to each point $p$ in $M$ a vector $X(p) \in T_pM$. We also admit vector fields defined only on open subsets of M.

Examples are the *basis vector fields* $\partial/\partial x^k$ of a coordinate system: by definition, $\partial/\partial x^k$ has components $v^i = \delta^i_k$ relative to the coordinate system $(x^i)$. On the coordinate domain, every vector field can be written as

$$X = X^k \frac{\partial}{\partial x^k} \tag{6}$$

for certain scalar functions $X^k$. $X$ is said to be of class $\mathrm{C}^k$ if the $X^k$ have this property. This notation for vector fields is especially appropriate, because $X$ gives an operator on smooth functions $f : M \to \mathbb{R}$, denoted $f \to Xf$ and defined by $Xf := df(X)$. In coordinates $(x^i)$ this says $Xf = X^i(\partial f/\partial x^i)$, i.e. the operator $f \to Xf$ in (6) is interpreted as differential operator.

**A note on notation**. The mysterious symbol $\delta x$, favoured by classical geometers like Cartan and Weyl, can find its place here: $\delta$ stands for a vector field $X$, $x$ for a coordinates system $(x^k)$, and $\delta x$ quite literally and correctly for the coordinate vector field $(\delta x^k) = (X^k)$ representing $\delta = X$ in the coordinate system. If several vector fields are needed one can introduce several deltas like $\delta x, \delta' x, \cdots$. The notation $dx$ instead of $\delta x$ or $\delta' x$ can also be found in the literature, but requires an exceptionally cool head to keep from causing havoc. The delta notation is convenient because it introduces both a vector field and (if necessary) a coordinate system without ado and correctly suggests manipulations like $\delta(fg) = (\delta f)g + f(\delta g)$; it is commonly used in physics and in the Calculus of Variations, where the vector fields typically enter as "infinitesimal variations" of some quantity or other. One should be wary, however, of attempts to interpret $\delta x$ as a *single vector* at a point $x$ rather than a vector *field* relative to a coordinate system; this will not do in constructs like $\delta'\delta x$, which are perfectly meaningful and frequently called for in situations where the delta notation is wielded according to tradition[2].

We now consider a differentiable map $f : M \to N$ between two manifolds. It should be clear how to defined the differential of $f$ in accordance with the principle that any notion for $\mathbb{R}^n$ which is local and preserved under diffeomorphism will make sense on a manifold. To spell this out in detail, we need a lemma. Relative to a coordinate system $(x^1, \cdots, x^n)$ on $M$ and $(y^1, \cdots, y^m)$ on $N$, we may write $f(p)$ as $y^j = f^j(x^1, \cdots, x^n)$ and consider the partials $\partial y^j / \partial x^i$.

1.3.**6 Lemma**. *Let $f\colon M \to N$ be a differentiable map between manifolds, $p_o$ a point in $M$.*
*In coordinates $(x^i)$ around $p_o$ and $(y^j)$ around $f(p_o)$ consider $q = f(p)$ as a map $y^j = f^j(x^1, \cdots x^n)$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ with differential given by $\eta^j = \left(\frac{\partial y^j}{\partial x^i}\right)_{p_o} \xi^i$.*
*The map from vectors at $p_o$ with components $(\xi^i)$ relative to $(x^i)$ to vectors at $f(p_o)$ with components $(\eta^j)$ relative to $(y^j)$ defined by this equation is independent of these coordinate systems.*

**Proof**. Fix a point $p$ on $M$ and a vector $v$ at $p$. Suppose we take two pairs of coordinate systems $(x^i)$, $(y^j)$ and $(\tilde{x}^i), (\tilde{y}^j)$ to construct the quantities $(\eta^i)$ and $(\tilde{\eta}^i)$ from the components $(\xi^i)$and $(\tilde{\xi^i})$ of $v$ at $p$ with respect to $(x^i)$ and $(\tilde{x}^i)$:

$$\eta^j := \frac{\partial y^j}{\partial x^i}\xi^i, \qquad \tilde{\eta}^j := \frac{\partial \tilde{y}^j}{\partial \tilde{x}^i}\tilde{\xi}^i.$$

We have to show that these $(\eta^i)$ and $(\tilde{\eta}^i)$ are the components of the same vector $w$ at $f(p)$ with respect to $(y^j)$ and $(\tilde{y}^j)$ i.e. that

$$\tilde{\eta}^j \stackrel{?}{=} \frac{\partial \tilde{y}^j}{\partial y^k} y^i.$$

the partials being taken at $f(p)$. But this is just the Chain Rule:

$$\begin{aligned}
\tilde{\eta}^j &= \frac{\partial \tilde{y}^j}{\partial \tilde{x}^i}\tilde{\xi}^i \qquad [\text{definition of } \tilde{\eta}^j] \\
&= \frac{\partial \tilde{y}^j}{\partial \tilde{x}^i}\frac{\partial \tilde{x}^i}{\partial x^k}\xi^k \qquad [(\xi^k) \text{ is a vector}] \\
&= \frac{\partial \tilde{y}^j}{\partial x^k}\xi^k \qquad [\text{Chain Rule}] \\
&= \frac{\partial \tilde{y}^j}{\partial y^i}\frac{\partial y^i}{\partial x^k}\xi^k \qquad [\text{Chain Rule}] \\
&= \frac{\partial \tilde{y}^j}{\partial y^i}\eta^i \qquad [\text{definition of } \eta^i]
\end{aligned}$$

□

[2]E.g. Cartan "Leçons" (1925-1926) §159, Weyl "Raum-Zeit-Materie" (1922 edition) §16

The lemma allows us to define the differential of a map $f : M \to N$ as the linear map $df_p : T_p \to T_{f(p)}N$ which corresponds to the differential of the map $y^j = f^j(x)$ representing $f$ with respect to coordinate systems.

1.3.**7 Definition**. With the notation of the previous lemma, the vector at $f(p)$ with components $\eta^j = \left(\frac{\partial y^j}{\partial x^i}\right)_p \xi^i$ is denoted $df_p(v)$. The linear transformation $df_p : T_pM \to T_{f(p)}N$, $v \to df_p(v)$, is called the *differential* of $f$ at $p$.

This definition applies as well to functions defined only in a neighbourhood of the point $p$. As a special case, consider the differential of $df_p$ of a scalar valued functions $f : M \to \mathbb{R}$. In that case $df_p : M \to T_{f(p)}\mathbb{R} = \mathbb{R}$ is a linear functional on $T_pM$. According to the definition, it is given by the formula $df_p(v) = (\partial f/\partial x^i)_p \xi^i$ in coordinates $(x^i)$ around $p$. Thus $df_p$ looks like a gradient in these coordinates, but the $n$–tuple $(\partial f/\partial x^i)_p$ does *not* represent a vector at $p$ : $df_p$ does not belong to $T_pM$ but to the dual space $T_p^*M$ of linear functionals on $T_pM$.
The differentials $dx_p^k$ of the coordinate functions $p \to x^i(p)$ of a coordinate system $M \cdots \to \mathbb{R}^n, p \to (x^i(p))$ are of particular importance. It follows from the definition of $df_p$ that $(dx^k)_p$ has $k$–th component $= 1$, all others $= 0$. This implies that

$$dx^k(v) = \xi^k, \tag{7}$$

the k-th component of v. We can therefore think of the $dx^k$ as the components of a general vector $v$ at a general point $p$, just we can think of the $x^k$ as the coordinates of a general point $p$: the $dx^k$ pick out the components of a vector just like the $x^i$ pick out the coordinates of point.

1.3.**8 Example: coordinate differentials for polar coordinates**. We have $x = r\cos\theta$, $y = r\sin\theta$,

$dx = \frac{\partial x}{\partial r}dr + \frac{\partial x}{\partial \theta}d\theta = \cos\theta dr - r\sin\theta d\theta$,
$dy = \frac{\partial y}{\partial r}dr + \frac{\partial y}{\partial \theta}d\theta = \sin\theta dr + r\cos\theta d\theta$.

In terms of the coordinate differentials the differential $df$ of a scalar–valued function $f$ can be expressed as

$$df = \frac{\partial f}{\partial x^k}dx^k; \tag{8}$$

the subscripts "$p$" have been omitted to simplify the notation. In general, a map $f : M \to N$ between manifolds, written in coordinates as $y^j = f^j(x^1, \cdots, x^n)$ has its differential $df_p : T_pM \to T_{f(p)}N$ given by the formula $dy^j = (\partial f^j/\partial x^i)dx^i$, as is evident if we think of $dx^i$ as the components $\xi^i = dx^i(v)$ of a general vector at $p$, and think of $dy^j$ similarly. The transformation property of vectors is built into this notation.
The differential has the following geometric interpretation in terms of tangent vectors of curves.

1.3.**9 Lemma**. *Let $f : M \cdots \to N$ be a $C^1$ map between manifolds, $p = p(t)$ a $C^1$ curve in $M$. Then $df$ maps the tangent vector of $p(t)$ to the tangent vector* *of* $f(p(t))$*:*

$$\frac{df(p(t))}{dt} = df_{p(t)}\Big(\frac{dp(t)}{dt}\Big).$$

**Proof**. Write $p = p(t)$ in coordinates as $x^i = x^i(t)$ and $q = f(p(t))$ as $y^i = y^i(t)$. Then

$$\frac{dy^j}{dt} = \frac{\partial y^j}{\partial x^i}\frac{dx^i}{dt} \text{ says } \frac{df(p(t))}{dt} = df_{p(t)}\Big(\frac{dp(t)}{dt}\Big)$$

as required. □

Note that the lemma gives a way of calculating differentials:

$$df_{p_o}(v) = \frac{d}{dt} f(p(t))|_{t=t_o}$$

for any differentiable curve $p(t)$ with $p(t_o) = p_o$ and $\dot{p}(t_o) = v$.

1.3.**10 Example**. Let $f : S^2 \to \mathbb{R}^3$, $(\theta, \phi) \to (x, y, z)$ be the inclusion map, given by

$$x = \cos\theta \sin\phi,\ y = \sin\theta\sin\phi,\ z = \cos\phi \tag{9}$$

Then $df_p : T_pS^2 \to T_p\mathbb{R}^3 \approx \mathbb{R}^3$ is given by (4). If you think of $d\theta, d\phi$ as the $\theta\phi$–components of a general vector in $T_pS^2$ and of $dx, dy, dz$ as the $xyz$–components of a general vector in $T_p\mathbb{R}^3$ (see (7)), then this map $df_p : T_pS^2 \to T_p\mathbb{R}^3$, is given by differentiation of (9), i.e.

$dx = \frac{\partial x}{\partial \theta}d\theta + \frac{\partial x}{\partial \phi}d\phi = -\sin\theta\sin\phi d\theta + \cos\theta\cos\phi d\phi,$

$dy = \frac{\partial y}{\partial \theta}d\theta + \frac{\partial y}{\partial \phi}d\phi = \cos\theta\sin\phi d\theta + \sin\theta\cos\phi d\phi,$

$dz = \frac{\partial z}{\partial \theta}d\theta + \frac{\partial z}{\partial \phi}d\phi = -\sin\phi d\phi.$

The Inverse Function Theorem, its statement being local and invariant under diffeomorphisms, generalizes immediately from $\mathbb{R}^n$ to manifolds:

1.3.**11 Inverse Function Theorem**. *Let* $f : M \to N$ *be a* smooth *map of manifolds. Suppose* $p_o$ *is a point of* $M$ *at which the differential* $df_{p_o} : T_{p_o}M \to T_{f(p_o)}N$ *is bijective. Then* $f$ *is a diffeomorphism of a neighbourhood of* $p_o$ *onto a neighbourhood of* $f(p_o)$.

**Remarks**. (a) To say that the linear $df_{p_o}$ is bijective means that $\mathrm{rank}(df_{p_o}) = \dim M = \dim N$

(b) The theorem implies that one can find coordinate systems around $p_o$ and $f(p_o)$ so that on the coordinate domain $f : M \to N$ becomes the identity map $\mathbb{R}^n \to \mathbb{R}^n$.

The Inverse Function Theorem has two important corollaries which we list themselves as theorems.

1.3.**12 Immersion Theorem**. *Let* $f : M \to N$ *be a* smooth *map of manifolds,* $n = \dim M$, $m = \dim N$. *Suppose* $p_o$ *is a point of* $M$ *at which the differential* $df_{p_o} : T_{p_o}M \to T_{f(p_o)}N$ *is injective. Let* $(x^1, \cdots, x^n)$ *be a coordinate system around* $p_o$. *There is a coordinate system* $(y^1, \cdots, y^m)$ *around* $f(p_o)$ *so that* $p \to q := f(p)$ *becomes*

$$(x^1, \cdots, x^n) \to (y^1, \cdots, y^n, \cdots, y^m) := (x^1, \cdots, x^n, 0, \cdots, 0).$$

**Remarks**. (a) To say that the linear map $df_{p_o}$ is injective (i.e. one–to–one) means that $\operatorname{rank}(df_{p_o}) = \dim M \leq \dim N$. We then call $f$ an *immersion* at $p_o$.
(b) The theorem says that one can find coordinate systems around $p_o$ and $f(p_o)$ so that on the coordinate domain $f : M \to N$ becomes the inclusion map $\mathbb{R}^n \to \mathbb{R}^m$ $(n \leq m)$.
**Proof**. Using coordinates, it suffices to consider a (partially defined) map $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ $(n \leq m)$. Suppose we can a find local diffeomorphism $\varphi : \mathbb{R}^m \cdots \to \mathbb{R}^m$ so that $f = \varphi \circ i$ where $i : \mathbb{R}^n \to \mathbb{R}^m$ is the inclusion.

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{i} & \mathbb{R}^m \\ \mathrm{id} \downarrow & & \downarrow \varphi \\ \mathbb{R}^n & \underset{f}{\rightarrow} & \mathbb{R}^m \end{array}$$

Then the equation $f(x) = \varphi(i(x))$, says that $f$ "becomes" $i$ if we use $\varphi$ as coordinates on $\mathbb{R}^m$ and the identity on $\mathbb{R}^n$.
Now assume $\operatorname{rank}(df_{p_o}) = n \leq m$. Write $q = f(p)$ as $y^j = f^j(x^1, \cdots, x^n), j \leq m$. At $p_o$, the $m \times n$ matrix $[\partial f^j/\partial x^i]$,$j \leq m$, $i \leq n$, has $n$ linearly independent rows (indexed by some $j$'s). Relabeling the coordinates $(y^j)$ we may assume that the $n \times n$ matrix $[\partial f^j/\partial x^i]$, $i, j \leq n$, has rank $n$, hence is invertible. Define $\varphi$ by

$$\varphi(x^1, \cdots x^n, \cdots, x^m) = f(x^1, \cdots, x^n) + (0, \cdots, 0,\ x^{n+1}, \cdots, x^m).$$

i.e.

$$\varphi = (f^1, \cdots, f^n,\ f^{n+1} + x^{n+1}, \cdots, f^m + x^m),$$

Since $i(x^1, \cdots, x^n) = (x^1, \cdots, x^n, 0, \cdots, 0)$ we evidently have $f = \varphi \circ i$. The determinant of the matrix $(\partial \varphi^j/\partial x^i)$ has the form

$$\det \begin{bmatrix} \partial f^j/\partial x^i & 0 \\ * & 1 \end{bmatrix} = \det[\partial f^j/\partial x^i],\ i, j \leq n$$

hence is nonzero at $i(p_o)$. By the Inverse Function Theorem $\varphi$ is a local diffeomorphism at $p_o$, as required. □

**Example**. a)Take for $f : M \to N$ the inclusion map $i : S^2 \to \mathbb{R}^3$. If we use coordinates $(\theta, \phi)$ on $S^2 = \{r = 1\}$ and $(x, y, z)$ on $\mathbb{R}^3$ then $i$ is given by the familiar formulas

$$i : (\theta, \phi) \to (x, y, z) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi).$$

A coordinate system $(y^1, y^2, y^2)$ on $\mathbb{R}^3$ as in the theorem is given by the slightly modified spherical coordinates $(\theta, \phi, \rho - 1)$ on $\mathbb{R}^3$, for example: in these coordinates the inclusion $i\colon S^2 \to \mathbb{R}^3$ becomes the standard inclusion map $(\theta, \phi) \to (\theta, \phi, \rho - 1) = (\theta, \phi, 0)$ on the coordinate domains. (That the same labels $\theta$ and $\phi$ stand for coordinate functions both on $S^2$ and $\mathbb{R}^3$ should cause no confusion.)
b)Take for $f : M \to N$ a curve $\mathbb{R} \cdots \to M, t \to p(t)$. This is an immersion at $t = t_o$ if $\dot{p}(t_o) \neq 0$. The immersion theorem asserts that near such a point $p(t_o)$ there are coordinates $(x^i)$ so that the curve $p(t)$ becomes a coordinate line, say $x^1 = t$, $x^2 = 0, \cdots, x^n = 0$.

1.3.**13 Submersion Theorem**. Let $f : M \to N$ be a smooth map of manifolds, $n = \dim M$, $m = \dim N$. Suppose $p_o$ is a point of $M$ at which the differential $df_{p_o} : T_{p_o}M \to T_{f(p_o)}N$ is surjective. Let $(y^1, \cdots, y^m)$ be a coordinate system around $f(p_o)$. There is a coordinate system $(x^1, \cdots, x^n)$ around $p_o$ so that

$$p \to q := f(p)$$

becomes

$$(x^1, \cdots, x^m, \cdots, x^n) \to (y^1, \cdots, y^m) := (x^1, \cdots, x^m)$$

**Remarks**. (a) To say that the linear map $df_{p_o}$ is surjective (i.e. onto) means that

$$\operatorname{rank}(df_{p_o}) = \dim N \leq \dim M.$$

We then call $f$ a submersion at $p_o$.
(b) The theorem says that one can find coordinates systems around $p_o$ and $f(p_o)$ so that on the coordinate domain $f : M \to N$ becomes the projection map $\mathbb{R}^n \to \mathbb{R}^m$ $(n \geq m)$.
**Proof**. Using coordinates, it suffices to consider a partially defined map $f : \mathbb{R}^n \cdots \to \mathbb{R}^m$ $(n \geq m)$. Suppose we can find a local diffeomorphism $\varphi : \mathbb{R}^n \cdots \to \mathbb{R}^n$ so that $p \circ \varphi = f$ where $p : \mathbb{R}^m \to \mathbb{R}^n$ is the projection.

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{f} & \mathbb{R}^m \\ \varphi \downarrow & & \downarrow \mathrm{id} \\ \mathbb{R}^n & \xrightarrow[p]{} & \mathbb{R}^m \end{array}$$

Then we can use $\varphi$ as a coordinate system and the equation $f(x) = p(\varphi(x))$, says that $f$ "becomes" $i$ if we use $\varphi$ as coordinates on $\mathbb{R}^n$ and the identity on $\mathbb{R}^n$.
Now assume $\operatorname{rank}(df_{p_o}) = m \leq n$. Write $q = f(p)$ as $y^j = f^j(x^1, \cdots, x^n)$. At $p_o$, the $m \times n$ matrix $[\partial f^j/\partial x^i]$,$j \leq m$, $i \leq n$, has $m$ linearly independent columns (indexed by some $i$'s). Relabeling the coordinates $(x^i)$ we may assume that the $m \times m$ matrix $[\partial f^j/\partial x^i]$, $i, j \leq m$, has rank $m$, hence is invertible. Define $\varphi$ by

$$\varphi(x^1, \cdots, x^n) = (f(x^1, \cdots, x^n),\, x^{m+1}, \cdots, x^n).$$

Since $p(x^1, \cdots, x^m, \cdots, x^n) = (x^1, \cdots, x^m)$, we have $f = p \circ \varphi$. The determinant of the matrix $(\partial \varphi^j/\partial x^i)$ has the form

$$\det \begin{bmatrix} \partial f^j/\partial x^i & 0 \\ * & 1 \end{bmatrix} = \det[\partial f^j/\partial x^i],\ ij \leq n$$

hence is nonzero at $p_o$. By the Inverse Function Theorem $\varphi$ is a local diffeomorphism at $p_o$, as required. □

**Example**. a)Take for $f : M \to N$ the radial projection map $\pi : \mathbb{R}^3 - \{0\} \to S^2$, $p \to p/\|p\|$. If we use the spherical coordinates $(\theta, \phi, \rho)$ on $\mathbb{R}^3 - \{0\}$ and $(\theta, \phi)$ on $S^2$ then $\pi$ becomes the standard projection map $(\theta, \phi, \rho) \to (\theta, \phi)$ on the coordinate domains .

b)Take for $f : M \to N$ a scalar function $f : M \to \mathbb{R}$. This is a submersion at $p_o$ if $df_{p_o} \neq 0$. The submersion theorem asserts that near such a point there are coordinates $(x^i)$ so that $f$ becomes a coordinate function, say $f(p) = x^1(p)$.

**Remark**. The immersion theorem says that an immersion $f : M \to N$ locally becomes a linear map $\mathbb{R}^n \to \mathbb{R}^m$, just like its tangent map $df_{p_o} : T_{p_o}M \to T_{f(p_o)}N$. The submersion theorem can be interpreted similarly. These a special cases of a more general theorem (*Rank Theorem*), which says that $f : M \to N$ becomes a linear map $\mathbb{R}^n \to \mathbb{R}^m$, hence also like its tangent map $df_{p_o} : T_{p_o}M \to T_{f(p_o)}N$, provided $df_p$ has constant rank for all $p$ in a neighbourhood of $p_o$. This is automatic for immersions and submersions (exercise). A proof of this theorem can be found in Spivak, p.52, for example.
There remain a few things to be taken care of in connection with vectors and differentials.

1.3.**14 Covectors and 1–forms**. We already remarked that the differential $df_p$ of a scalar function $f : M \to \mathbb{R}$ is a linear functional on $T_pM$. The space of all linear functionals on $T_pM$ is called the *cotangent space* at $p$, denoted $T_p^*M$. The coordinate differentials $dx^i$ at $p$ satisfy $dx^i(\partial/\partial x^j) = \delta^i_j$, which means that the $dx^i$ form the dual basis to the $\partial/\partial x^j$. Any element $w \in T_p^*M$ is therefore of the form $w = \eta_i dx^i$; its components $\eta_i$ satisfy the transformation rule

$$\tilde{\eta}_j = \Big(\frac{\partial x^i}{\partial \tilde{x}^j}\Big)_p \eta_i \tag{10}$$

which is not the same as for vectors, since the upper indices are being summed over. (Memory aid: the tilde on the right goes downstairs like the index $j$ on the left.) Elements of $T_p^*M$ are also called *covectors* at $p$, and this transformation rule could be used to define "covector" in a way analogous to the definition of vector. Any covector at $p$ can be realized as the differential of some smooth defined near $p$, just like any vector can be realized as the tangent vector to some smooth curve through $p$. In spite of the similarity of their transformation laws, one should carefully distinguish between vectors and covectors.
A *differential 1-form* (or covector field) $\varphi$ on an open subset of $M$ associates to each point $p$ in its domain a covector $\varphi_p$. Examples are the *coordinate differentials* $dx^k$: by definition, $dx^k$ has components $\eta_i = \delta^k_i$ relative to the coordinate system $(x^i)$. On the coordinate domain, every differential 1-form $\varphi$ can be written as

$$\varphi = \varphi_k dx^k$$

for certain scalar functions $\varphi_k$. $\varphi$ is said to be of class $C^k$ if the $\varphi_k$ have this property.

1.3.**15 Definition. Tangent bundle and cotangent bundle**. The set of all vectors on $M$ is denoted $TM$ and called the *tangent bundle* of $M$; we make it into a manifold by using as coordinates $(x^i, \xi^i)$ of a vector $v$ at $p$ the coordinates $x^i(p)$ of $p$ together with the components $dx^i(v) = \xi^i$ of $v$. (Thus $\xi^i = dx^i$ as function on $TM$, but the notation $dx^i$ as coordinate on $TM$ gets to be confusing in combinations such as $\partial/\partial\xi^i$.) As $(x^i)$ runs over a collection of coordinate systems of $M$ satisfying MAN 1-3, the $(x^i, \xi^i)$ do the same for $TM$. The tangent bundle comes equipped with a *projection* map $\pi : TM \to M$ which sends a vector $v \in T_pM$ to the point $\pi(v) = p$ to which it is attached.

1.3.**16 Example**. From (1.3.3) and (1.3.4), p 41 we get the identifications
a) $T\mathbb{R}^n = \{(p, v) : p, v \in \mathbb{R}^n\} = \mathbb{R}^n \times \mathbb{R}^n$
b) $TS^2 = \{(p, v) \in \mathbb{R}^3 \times \mathbb{R}^3 : p \cdot v = 0 \text{ (dot product)}\}$
The set of all covectors on $M$ is denoted $T^*M$ and called the *cotangent bundle* of $M$; we make it into a manifold by using as coordinates $(x^i, \xi_i)$ of a covector $w$ at $p$ the coordinates $(x^i)$ of $p$ together with the components $(\xi_i)$ of $w$. (If one identifies $v \in T_pM$ with the function $\varphi \to \varphi(v)$ on $T_p^*M$, then $\xi_i = \partial/\partial x^i$ as a function on cotangent vectors.) As $(x^i)$ runs over a collection of coordinate systems of $M$ satisfying MAN 1-3, p 24 the $(x^i, \xi_i)$ do the same for $T^*M$. There is again a projection map $\pi : T^*M \to M$ which sends a covector $w \in T_p^*M$ to the point $\pi(w) = p$ to which it is attached.

## EXERCISES 1.3

1. Show that addition and scalar multiplication of vectors at a point $p$ on a manifold, as defined in the text, does indeed produce vectors.

2. Prove the two assertions left as exercises in 1.3.3.

3. Prove the two assertions left as exercises in 1.3.4.

4. (a) Prove the transformation rule for the components of a covector $w \in T_p^*M$:

$$\tilde{\eta}_j = \left(\frac{\partial x^i}{\partial \tilde{x}^j}\right)_p \eta_i. \tag{*}$$

(b) Prove that covectors can be defined be defined in analogy with vectors in the following way. Let $w$ be a quantity which relative to a coordinate system $(x^i)$ around $p$ is represented by an $n$–tuple $(\eta^i)$ subject to the transformation rule (*). Then the scalar $\eta_i\xi^i$ depending on the components of $w$ and of a vector $v$ at $p$ is independent of the coordinate system and defines a linear functional on $T_pM$ (i.e. a covector at $p$).
(c)Show that any covector at $p$ can be realized as the differential $df_p$ of some smooth function $f$ defined in a neighbourhood of $p$.

5. Justify the following rules from the definitions and the usual rules of differentiation.

$$\text{(a) } d\tilde{x}^k = \frac{\partial \tilde{x}^k}{\partial x^i} dx^i \qquad \text{(b) } \frac{\partial}{\partial \tilde{x}^k} = \frac{\partial x^i}{\partial \tilde{x}^k}\frac{\partial}{\partial x^i}$$

6. Let $(\rho, \theta, \phi)$ be spherical coordinates on $\mathbb{R}^3$. Calculate the coordinate vector fields $\partial/\partial\rho$, $\partial/\partial\theta$, $\partial/\partial\phi$ in terms of $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$, and the coordinate differentials $d\rho, d\theta, d\phi$ in terms of $dx, dy, dz$. (You can leave their coefficients in terms of $\rho, \theta, \phi$). Sketch some coordinate lines and the coordinate vector fields at

some point. (Start by drawing a sphere $\rho$ =constant and some $\theta$, $\phi$–coordinate lines on it.)

7. Define coordinates $(u, v)$ on $\mathbb{R}^2$ by the formulas

$$x = \cosh u \cos v, \qquad y = \sinh u \sin v.$$

a) Determine all points $(x, y)$ in a neighbourhood of which $(u, v)$ may be used as coordinates.
b) Sketch the coordinate lines $u = 0,\ 1,\ 3/2,\ 2$ and $v = 0, \pi/6, \pi/4, \pi/3, \pi/2, 2\pi/3,\ 3\pi/4,\ 5\pi/6$.
c) Find the coordinate vector fields $\partial/\partial u$, $\partial/\partial v$ in terms of $\partial/\partial x$, $\partial/\partial y$.

8. Define coordinates $(u, v)$ on $\mathbb{R}^2$ by the formulas

$$x = \frac{1}{2}(u^2 - v^2), \qquad y = uv.$$

a) Determine all points $(x, y)$ in a neighbourhood of which $(u, v)$ may be used as coordinates.
b) Sketch the coordinate lines $u$,$v = 0,\ 1/2,\ 1,\ 3/2,\ 2$ .
c) Find the coordinate vector fields $\partial/\partial u$, $\partial/\partial v$ in terms of $\partial/\partial x$, $\partial/\partial y$.

9. Let $(u, \theta, \psi)$ hyperbolic coordinates be on $\mathbb{R}^3$, defined by

$$x = u \cos\theta\ \sinh\psi,\ y = u \sin\theta \sinh\psi,\ z = u \cosh\psi.$$

a) Sketch some surfaces $u$ =constant (just enough to show the general shape of these surfaces).
b) Determine all points $(x, y, z)$ in a neighbourhood of which $(u, \theta, \psi)$ may be used as coordinates.
c) Find the coordinate vector fields $\partial/\partial u, \partial/\partial\theta, \partial/\partial\psi$ in terms of $\partial/\partial x, \partial/\partial y, \partial/\partial z$.

10. Show that as $(x^i)$ runs over a collection of coordinate systems for $M$ satisfying MAN 1-3, the $(x^i, \xi^i)$ defined in the text do the same for $TM$.

11. Show that as $(x^i)$ runs over a collection of coordinate systems for $M$ satisfying MAN 1-3, the $(x^i, \xi_i)$ defined in the text do the same for $T^*M$.

12. a) Prove that $T(M\times M)$ is diffeomorphic to $(TM)\times(TM)$ for every manifold $M$.
b) Is $T(TM)$ diffeomorphic to $TM \times TM$ for every manifold $M$? (Explain. Give some examples. Prove your answer if you can. Use the notation $M \approx N$ to abbreviate "$M$ is diffeomorphic with $N$".)

13. Let $M$ be a manifold and let $W = T^*M$. Let $\pi : W \to M$ and $\rho : TW \to W$ be the projection maps. For any $z \in TW$ define $\theta(z) \in \mathbb{R}$ by

$$\theta(z) := \rho(z)(d\pi_{\rho(z)}(z)).$$

($\theta$ is called the *canonical* 1*–form* on the cotangent bundle $T^*M$.)

a) Explain why this makes sense and defines a 1–form $\theta$ on $W$.
b) Let $(x^i)$ be a coordinate system on $M$, $(x^i, \xi_i)$ the corresponding coordinates on $W$, as in 3.17. Show that $\theta = \xi_i dx^i$. [Suggestion. Show first that $d\pi(z) = dx^i(z)\partial/\partial x^i$.]

14. Let $M$ be a manifold, $p \in M$ a point. For any $v \in T_pM$ define a linear functional $f \to D_v(f)$ on smooth functions $f : M \cdots \to \mathbb{R}$ defined near $p$ by the rule $D_v(f) = df_p(v)$. Then $D = D_v$ evidently satisfies

$$D(fg) = D(f)g(p) + f(p)D(g). \tag{$*$}$$

Show that, conversely, any linear functional $f \to D(f)$ on smooth functions $f : M \cdots \to \mathbb{R}$ defined near $p$ which satisfies (*) is of the form $D = D_v$ for a unique vector $v \in T_pM$.
[Remark. Such linear functionals $D$ are called *point derivation*s at $p$. Hence there is a one–to–one correspondence between vectors at $p$ and point–derivations at $p$. ]

15. Let $M$ be a manifold, $p \in M$ a point. Consider the set $C_p$ of all smooth curves $p(t)$ through $p$ in $M$, with $p(0) = p$. Call two such curves $p_1(t)$ and $p_2(t)$ *tangential* at $p$ if they have the same coordinate tangent vector at $t = 0$ relative to some coordinate system. Prove in detail that there is a one–to–one correspondence between equivalence classes for the relation "being tangential at $p$" and tangent vectors at $p$. (Show first that being *tangential* at $p$ is an equivalence relation on $C_p$.)

16. Suppose $f : M \to N$ is and $f$ a *immersion* at $p_o \in M$. Prove that $f$ is an immersion at all points $p$ in a neighbourhood of $p_o$. Prove the same result for submersions. [Suggestion. Consider rank($df$). Review the proof of the immersion (submersion) theorem.]

17. Let $M$ be a manifold, $TM$ its tangent bundle, and $\pi : TM \to M$ the natural projection map. Let $(x^i)$ be a coordinate system on $M$, $(x^i, \xi^i)$ the corresponding coordinate system on $TM$ (defined by $\xi^i = dx^i$ as function on $TM$). Prove that the differential of $\pi$ is given by

$$d\pi(z) = dx^i(z)\left(\frac{\partial}{\partial x^i}\right). \tag{*}$$

Determine if $\pi$ is an immersion, a submersion, a diffeomorphism, or none of these. [The formula (*) needs explanation: specify how it is to be interpreted. What exactly is the symbol $dx^i$ on the right?]

*18. Consider the following quotation from Élie Cartan's Leçons* of 1925–1926 (p33). "*To each point $p$ with coordinates $(x^1, \cdots, x^n)$ one can attach a system of Cartesian coordinates with the point $p$ as origin for which the basis vectors $e_1, \cdots, e_n$ are chosen in such a way the coordinates of the infinitesimally close point $p'(x^1 + dx^1, \cdots, x^n + dx^n)$ are exactly $(dx^1, \cdots, dx^n)$. For this it suffices that the vector $e_i$ be tangent to the ith coordinate curve (obtained by varying*

*only the coordinate $x^i$) and which, to be precise, represents the velocity of a point traversing this curve when on considers the variable coordinate $x^i$ as time.*"
(a)What is our notation for $e_i$?
(b)Find page and line in this section which corresponds to the sentence "*For this it suffices*$\cdots$".
(c)Is it strictly true that $dx^1, \cdots, dx^n$ are Cartesian coordinates? Explain.
(d)What *must* be meant by the "the infinitesimally close point $p'(x^1+dx^1, \cdots, x^n+dx^n)$" for the sentence to be correct?
(e)Be picky and explain why the sentence "*To each point $p$ with coordinates $(x^1, \cdots, x^n)$ one can attach* $\cdots$" might be misleading to less astute readers. Rephrase it correctly with minimal amount of change.
*(f)Rewrite the whole quotation in our language, again with minimal amount of change.*

## 1.4 Submanifolds

1.4.1 **Definition**.Let $M$ be an n−dimensional manifold, $S$ a subset of $M$. A point $p \in S$ is called a *regular* point of $S$ if $p$ has an open neighbourhood $U$ in $M$ that lies in the domain of some coordinate system $x^1, \cdots, x^n$ on $M$ with the property that the points of $S$ in $U$ are precisely those points in $U$ whose coordinates satisfy $x^{m+1} = 0, \cdots, x^n = 0$ for some $m$. This $m$ is called the *dimension of $S$ at $p$*. Otherwise $p$ is called a *singular* point of $S$. $S$ is called an m−dimensional (*regular*) submanifold of $M$ if every point of $S$ is regular of the same dimension $m$.

**Remarks**. a) We shall summarize the definition of "$p$ is regular point of $S$" by saying that $S$ is given by the equations $x^{m+1} = \cdots,\ x^n = 0$ locally around $p$. The number $m$ is independent of the choice of the $(x^i)$: if $S$ is also given by $\tilde{x}^{\tilde{m}+1} = \cdots = \tilde{x}^n = 0$, then maps $(x^1, \cdots, x^m) \leftrightarrow (\tilde{x}^1, \cdots, \tilde{x}^{\tilde{m}})$ which relate the coordinates of the points of $S$ in $U \bigcap \tilde{U}$ are inverses of each other and smooth. Hence their Jacobian matrices are inverses of each other, in particular $m = \tilde{m}$.
b) We admit the possibility that $m = n$ or $m = 0$. An $n$-dimensional submanifold $S$ must be open in $M$ i.e. every point of $S$ has neighbourhood in $M$ which is contained in $S$. At the other extreme, a 0-dimensional submanifold is discrete, i.e. every point in $S$ has a neighbourhood in $M$ which contains only this one point of $S$.
By definition, a coordinate system on a submanifold $S$ consists of the restrictions $t^1 = x^1|_S, \cdots, t^m = x^m|_S$ to $S$ of the first m coordinates of a coordinate system $x^1, \cdots, x^n$ on $M$ of the type mentioned above (for some $p$ in $S$); as their domain we take $S \bigcap U$. We have to verify that coordinate systems satisfy MAN1–3.
MAN 1. $t(S \bigcap U)$ consists of the $(x^i)$ in the open subset $x(U) \bigcap \mathbb{R}^m$ of $\mathbb{R}^n$. (Here we identify $\mathbb{R}^m = \{x \in \mathbb{R}^n\colon\ x^{m+1} = \cdots = x^n = 0\}$.)
MAN 2. Then map $(x^1, \cdots, x^m) \to (\tilde{x}^1, \cdots, \tilde{x}^m)$ which relates the first $m$ coordinates of the points of $S$ in $U \bigcap \tilde{U}$ is smooth with open domain $x(U \bigcap \tilde{U}) \bigcap \mathbb{R}^m$.
MAN 3. Every $p \in S$ lies in some domain $U \bigcap S$, by definition.

Let $S$ be a submanifold of $M$, and $i : S \to M$ the inclusion map. The differential $di_p : T_pS \to T_pM$ maps the tangent vector of a curve $p(t)$ in $S$ into the tangent vector of the same curve $p(t) = i(p(t))$, considered as curve in $M$. We shall use the following lemma to identify $T_pS$ with a subspace of $T_pM$.

1.4.2 **Lemma**. L*et $S$ be a submanifold of $M$. Suppose $S$ is given by the equations*

$$x^{m+1} = 0, \cdots , x^n = 0$$

*locally around $p$. The differential of the inclusion $S \overset{\subset}{\to} M$ at $p \in S$ is a bijection of $T_pS$ with the subspace of $T_pM$ given by the linear equations*

$$(dx^{m+1})_p = 0, \cdots , (dx^n)_p = 0.$$

*This subspace consists of tangent vectors at $p$ of differentiable curves in $M$ that lie in $S$.*

**Proof**. In the coordinates $(x^1, \cdots , x^n)$ on $M$ and $(t^1, \cdots , t^m|_S)$ on $S$ the inclusion map $S \to M$ is given by

$$x^1 = t^1, \cdots , x^m = t^m, x^{m+1} = 0, \cdots , x^n = 0.$$

In the corresponding linear coordinates $(dx^1, \cdots , dx^n)$ on $T_pM$ and $(dt^1, \cdots , dt^m)$ on $T_pS$ its differential is given by

$$dx^1 = dt^1, \cdots , dx^m = dt^m, dx^{m+1} = 0, \cdots , dx^n = 0.$$

This implies the assertion. $\square$.
This lemma may be generalized:

1.4.3 **Theorem**. *Let $M$ be a manifold, $f^1, \cdots , f^k$* smooth *functions on $M$. Let $S$ be the set of points $p \in M$ satisfying $f^1(p) = 0, \cdots , f^k(p) = 0$. Suppose the differentials $df^1, \cdots , df^k$ are linearly independent at every point of $S$.*
*(a)$S$ is a submanifold of $M$ of dimension $m = n - k$ and its tangent space at $p \in S$ is the subspace of vectors $v \in T_pM$ satisfying the linear equations*

$$(df^1)_p(v) = 0, \cdots , (df^k)_p(v) = 0.$$

*(b)If $f^{k+1}, \cdots , f^n$ are any $m = n - k$ additional* smooth *functions on $M$ so that all $n$ differentials $df^1, \cdots , df^n$ are linearly independent at $p$, then their restrictions to $S$, denoted*

$$t^1 := f^{k+1}|_S, \cdots , t^m := f^{k+m}|_S,$$

*form a coordinate system around $p$ on $S$.*

**Proof 1**. Let $f = (f^1, \cdots , f^k)$. Since $\text{rank}(df_p) = k$, the Submersion Theorem implies there are coordinates $(x^1, \cdots , x^n)$ on $M$ so that $f$ locally becomes $f(x^1, \cdots , x^n) = (x^{m+1}, \cdots , x^{m+k})$, where $m + k = n$. So $f$ has component functions $f^1 = x^{m+1}, \cdots , f^{m+k} = x^{m+k}$. Thus we are back in the situation of the lemma, proving part (a). For part (b) we only have to note that the $n$-tuple $F = (f^1, \cdots, f^n)$ can be used as a coordinate system on $M$ around $p$ (Inverse Function Theorem), which is just of the kind required by the definition of "submanifold".

**Proof 2**. Supplement the $k$ functions $f^1, \cdots, f^k$, whose differentials are linearly independent at $p$, by $m = n - k$ additional functions $f^{k+1}, \cdots, f^n$, defined and smooth near $p$, so that all $n$ differentials $df^1, \cdots, df^n$ are linearly independent at $p$. (This is possible, because $df^1, \cdots, df^m$ are linearly independent at $p$, by hypothesis.) By the Inverse Function Theorem, t*he equation* $(x^1, \cdots, x^n) = (f^1(p), \cdots, f^n(n))$ *defines* a coordinate system $x^1, \cdots, x^n$ around $p_o = f(x_o)$ on $M$ so that $S$ is locally given by the equations $x^{m+1} = \cdots = x^n = 0$, and the theorem follows from the lemma. □

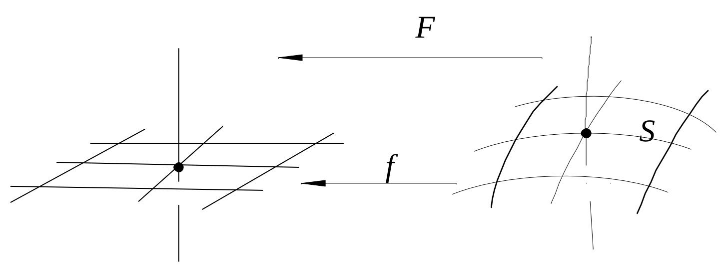



Fig 1. $F(p) = (x^1, \cdots, x^n)$.

**Remarks**. (1)The second proof is really the same as the first, with the proof of the Submersion Theorem relegated to a parenthetical comment; it brings out the nature that theorem.

(2)The theorem is local; even when the differentials of the $f^j$ are linearly dependent at some points of $S$, the subset of $S$ where they are linearly independent is still a submanifold.

1.4.4 **Examples.** (a)T*he sphere* $S = \{p \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$. Let $f = x^2 + y^2 + z^2 - 1$. Then $df = 2(xdx + ydy + zdz)$ and $df_p = 0$ iff $x = y = z = 0$ i.e. $p = (0, 0, 0)$. In particular, $df$ is everywhere non–zero on $S$, hence $S$ is a submanifold, and its tangent space at a general point $p = (x, y, z)$ is given by $xdx + ydy + zdz = 0$. This means that, as a subspace of $\mathbb{R}^3$, the tangent space at the point $p = (x, y, z)$ on $S$ consists of all vectors $v = (a, b, c)$ satisfying $xa + yb + zc = 0$, as one would expect.

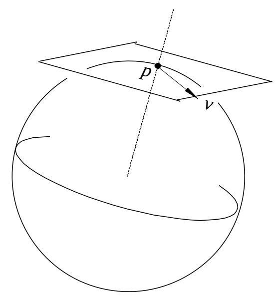



Fig 2. $v \in T_pS$

In spherical coordinates $\rho, \theta, \phi$ on $\mathbb{R}^3$ the sphere $S^2$ is given by $\rho = 1$ on the coordinate domain, as required by the definition of "submanifold". According to the definition, $\theta$ and $\phi$ provide coordinates on $S^2$ (where defined). If one identifies the tangent spaces to $S^2$ with subspaces of $\mathbb{R}^3$ then the coordinate vector fields $\partial/\partial\theta$ and $\partial/\partial\phi$ are given by the formulas of (1.3.4).
(b)T*he circle* $C = \{\, p \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1,\ ax+by+cz = d\}$ with fixed $a, b, c$ not all zero. Let $f = x^2+y^2-1, g = ax+by+cz-d$. Then $df = 2(xdx+ydy+zdz)$ and $dg = adx + bdx + cdz$ are linearly independent unless $(x, y, z)$ is a multiple of $(a, b, c)$. This can happen only if the plane $ax + by + cz = d$ is parallel to the tangent plane of the sphere at $p(x, y, z)$ in which case $C$ is either empty or reduces to the point of tangency. Apart for that case, $C$ is a submanifold of $\mathbb{R}^3$ and its tangent space $T_pC$ at any of its points $p(x, y, z)$ is the subspace of $T_p\mathbb{R}^3 = \mathbb{R}^3$ given by the two independent linear equations $df = 0, dg = 0$, hence has dimension $3 - 2 = 1$. Of course, $C$ is also the submanifold of the sphere $S = \{p \in \mathbb{R}^3 \mid f = 0\}$ by $g = 0$ and $T_pC$ the subspace of $T_pS$ given $dg = 0$.

1.4.5 **Example.** T*he cone* $S = \{p \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 0\}$. We shall show the following.
($a$) The cone $S - \{(0, 0, 0)\}$ with the origin excluded is a 2-dimensional submanifold of $\mathbb{R}^3$.
(b) The cone $S$ with the origin $(0, 0, 0)$ is not a 2-dimensional submanifold of $\mathbb{R}^3$.

**Proof of (a)**. Let $f = x^2 + y^2 - z^2$. Then $df = 2(xdx + ydy - zdz)$ and $df_p = 0$ iff $x = y = z = 0$ i.e. $p = (0, 0, 0)$. Hence the differential $d(x^2 + y^2 - z^2) = 2(xdx + ydy - zdz)$ is everywhere non$-$zero on $S - \{(0, 0, 0)\}$ which is therefore is a $(3 - 1)$-dimensional submanifold.

**Proof of (b)** (by contradiction). Suppose (!) $S$ were a 2-dimensional submanifold of $\mathbb{R}^3$. Then the tangent vectors at $(0, 0, 0)$ of curves in $\mathbb{R}^3$ which lie on $S$ would form a 2-dimensional subspace of $\mathbb{R}^3$. But this is not the case: for example, we can find three curves of the form $p(t) = (ta, tb, tc)$ on $S$ whose tangent vectors at $t = 0$ are linearly independent.

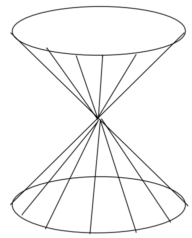


Fig. 3. The cone

1.4.6 **Theorem**. *Let $M$ be an $n$-dimensional manifold, $g : \mathbb{R}^m \to M$, $(t^1, \cdots, t^m) \to g(t^1, \cdots, t^m)$, a* smooth *map. Suppose the differential $dg_{u_o}$ has rank $m$ at some point $u_o$.*
*a) There is a neighbourhood $U$ of $u_o$ in $\mathbb{R}^m$ whose image $S = g(U)$ is an $m$-dimensional submanifold of $M$.*
*b) The equation $p = g(t^1, \cdots, t^m)$ defines a coordinate system $(t^1, \cdots, t^m)$ around $p_o = g(u_o)$ on $S$.*
*c) The tangent space of $S$ at $p_o = g(u_o)$ is the image of $dg_{u_o}$.*

**Proof 1**. Since $\text{rank}(dg_{u_o}) = m$, the Immersion Theorem implies there are coordinates $(x^1, \cdots, x^n)$ on $M$ so that $g$ locally becomes $g(x^1, \cdots, x^m) = (x^1, \cdots, x^m, 0, \cdots, 0)$. In these coordinates, the image of a neighbourhood of $u_o$ in $\mathbb{R}^m$ is given by $x^{m+1} = 0, \cdots, x^n = 0$, as required for a submanifold. This proves (a) and (b). Part (c) is also clear, since in the above coordinates the image of $dg_{u_o}$ is given by $dx^{m+1} = \cdots = dx^n = 0$.
**Proof 2**. Extend the map $p = g(u^1, \cdots, t^m)$ to a map $p = G(x^1, \cdots, x^m, x^{m+1}, \cdots, x^n)$ *of $n$ variables, defined and* smooth *near the given point $x_o = (u_o, 0)$, whose differential $dG$ has rank $n$ at $x_o$. (This is possible, because $dg$ has rank $m$ at $u_o$ by hypothesis.)* By the Inverse Function Theorem, t*he equation $p = G(x^1, \cdots, x^m, x^{m+1}, \cdots, x^n)$ then defines* a coordinate system $x^1, \cdots, x^n$ around $p_o = G(x_o) = g(u_o)$ on $M$ so that $S$ is locally given by the equations $x^{m+1} = \cdots = x^n = 0$, and the theorem follows from the lemma. □

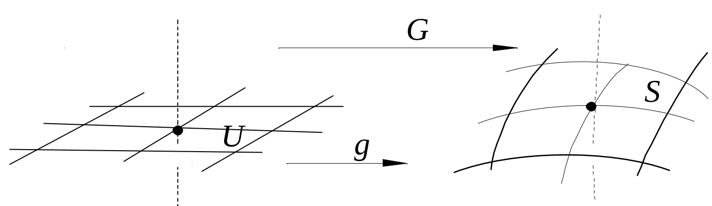



Fig.4. $p = G(x^1, \cdots, x^m)$

**Remarks**. The remarks after the preceding theorem apply again, but with an important modification. The theorem is again local in that $g$ need not be defined on all of $\mathbb{R}^m$, only on some open set $D$ containing $x_o$. But *even if $dg$ has rank $m$ everywhere on $D$, the image $g(D)$ need not be a submanifold of $M$*, as one can see from curves in $\mathbb{R}^2$ with self-intersections.

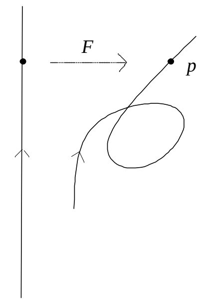



Fig.5. A curve with a self–intersection

1.4.7 **Example**. a) Parametrized curves in $\mathbb{R}^3$. Let $p = p(t)$ be a parametrized curve in $\mathbb{R}^3$. Then the differential of $g : t \to p(t)$ has rank 1 at $t_o$ iff $dp/dt \neq 0$ at $t = t_{o.}$

b) Parametrized surfaces in $\mathbb{R}^3$. Let $p = p(u, v)$ be a parametrized curve in $\mathbb{R}^3$. Then differential of $g : (u, v) \to p(u, v)$ has rank 2 at $(u_o, v_o)$ iff the $3{\times}2$ matrix $[\partial p/\partial u,\ \partial p/\partial v]$ has rank 2 at $(u, v) = (u_o, v_o)$ iff the cross product $\partial p/\partial u \times \partial p/\partial v \neq 0$ at $(u, v) = (u_o, v_o)$.

1.4.8 **Example**: *lemniscate of Bernoulli.* Let $C$ be the curve in $\mathbb{R}^2$ given by $r^2 = \cos 2\theta$ in polar coordinates. It is a figure 8 with the origin as point of self-intersection.

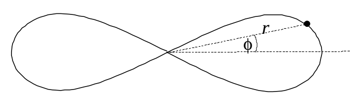



Fig. 6. The lemniscate of Bernoulli

We claim: (a) $C$ is a one-dimensional submanifold of $\mathbb{R}^2$ if the origin is excluded.
(b) The equations

$$x = \frac{\cos t}{1 + \sin^2 t} \quad y = \frac{\sin t \cos t}{1 + \sin^2 t}$$

define a map $\mathbb{R} \to \mathbb{R}^2$, $t \to p(t)$, of $\mathbb{R}$ onto $C$ with $\dot{p}(t) \neq 0$ everywhere.
(c) The above equations define a coordinate $t$ around every point of $C$ except the origin.
(d) $C$ is not a one-dimensional submanifold of $\mathbb{R}^2$ if the origin is included.
**Verification**. a) Let $f = r^2 - \cos 2\theta$. Then $df = 2rdr - 2\sin\theta d\theta \neq 0$ unless $r = 0$ and $\sin\theta = 0$. Since $(r, \theta)$ can be used as coordinate in a neighbourhood of every point except the origin, the assertion follows from theorem 1.4.3.
b) In Cartesian coordinates the equation $r^2 = \cos 2\theta = \cos^2\theta - \sin^2\theta$ becomes

$$(x^2 + y^2)^2 = x^2 - y^2.$$

By substitution on sees that $x = x(t)$ and $y = y(t)$ satisfy this equation and by differentiation one sees that $dx/dt$ and $dy/dt$ never vanish simultaneously. That

$g$ maps $\mathbb{R}$ onto $C$ is seen as follows. The origin corresponds to $t = (2n+1)\pi/2$, $n \in \mathbb{Z}$; as $t$ runs through an interval between two successive such points, $p(t)$ runs over one loop of the figure 8.
c) This follows from (b) and Theorem 1.4.6. Note that it follows from the discussion above that we can in fact choose as coordinate domain all of $C - \{\text{origin}\}$ if we restrict $t$ to the union of two successive open intervals, e.g. $(-\pi/2, \pi/2) \bigcup (\pi/2, 3\pi/2)$.
d) One computes that there are two linearly independent tangent vectors $\dot{p}(t)$ for two successive values of $t$ of the form $t = (2n+1)\pi/2$ (as expected). Hence $C$ cannot be submanifold of $\mathbb{R}^2$ of dimension one, which is the only possible dimension by (a).

## EXERCISES 1.4

1. Show that the submanifold coordinates on $S^2$ are also coordinates on $S^2$ when the manifold structure on $S^2$ is defined by taking the orthogonal projection coordinates in axioms MAN 1–3.

**2.** Let $S = \{p = (x, y, x^2) \in \mathbb{R}^3 \mid x^2 - y^2 - z^2 = 1\}$.
(a) Prove that $S$ is a submanifold of $\mathbb{R}^3$.
(b) Show that for any $(\psi, \theta)$ the point $p = (x, y, z)$ given by

$$x = \cosh\psi, \;\; y = \sinh\psi\cos\theta, \;\; z = \sinh\psi\sin\theta$$

lies on $S$ and prove that $(\psi, \theta)$ defines a coordinate system on $S$. [Suggestion. Use Theorem 1.4.6.]
(c) Describe $T_pS$ as a subspace of $\mathbb{R}^3$. (Compare Example 1.4.4.)
(d) Find the formulas for the coordinate vector fields $\partial/\partial\psi$ and $\partial/\partial\theta$ in terms of the coordinate vector fields $\partial/\partial x, \partial/\partial y, \partial/\partial z$, on $\mathbb{R}^3$. (Compare Example 1.4.4.)

3. Let $C$ be the curve $\mathbb{R}^3$ with equations $y - 2x^2 = 0$, $z - x^3 = 0$.
a) Show that $C$ is a one-dimensional submanifold of $\mathbb{R}^3$.
b) Show that $t = x$ can be used as coordinate on $C$ with coordinate domain all of $C$.
c) In a neighbourhood of which points of $C$ can one use $t = z$ as coordinate?
[Suggestion. For (a) use Theorem 1.4.3. For (b) use the description of $T_pC$ in Theorem 1.4.3 to show that the Inverse Function Theorem applies to $C \to \mathbb{R}$, $p \to x$.]

4. Let $M$ $=\mathrm{M}_3(\mathbb{R}) = \mathbb{R}^{3\times 3}$ be the set of all real $3 \times 3$ matrices regarded as a manifold. (We can take the matrix entries $X_{ij}$ of $X \in M$ as a coordinate system defined on all of $M$.) Let O(3) be the set of orthogonal $3 \times 3$ matrices: $O(3) = \{X \in \mathrm{M}_3(\mathbb{R}) \mid X^* = X^{-1}\}$, where $X^*$ is the transpose of $X$. Show that $O(3)$ is a 3-dimensional submanifold of $\mathrm{M}_3(\mathbb{R})$ with tangent space at $X \in O(3)$ given by

$$T_XO(3) = \{V \in \mathrm{M}_3(\mathbb{R}) \mid (X^{-1}V)^* = -X^{-1}V \text{ i.e. } X^{-1}V \text{ is skew-symmetric}\}.$$

[Suggestion. Consider the map $F$ from $\mathrm{M}_3(\mathbb{R})$ to the space $\mathrm{Sym}_3(\mathbb{R})$ of symmetric $3 \times 3$ matrices defined by $F(X) = X^*X$. Show that the differential $dF_X$ of this map is given by

$$dF_X(V) = X^*V + X^*V.$$

Conclude that $dF_X$ surjective for all $X \in O(3)$. Apply theorem 1.4.3.]

5. Continue with the setup of the previous problem. Let exp: $\mathrm{M}_3(\mathbb{R}) \to \mathrm{M}_3(\mathbb{R})$ be the matrix exponential defined by $\exp V = \sum_{k=0}^{\infty} V^k/k!$. If $V \in \mathrm{Skew}_3(\mathbb{R}) = \{V \in \mathrm{M}_3(\mathbb{R}) \mid V^* = -V\}$, then $\exp V \in O(3)$ because $(\exp V)^*(\exp V) = (\exp -V)(\exp V) = I$. Show that the equation

$$X = \exp V, \quad V = (V_{ij}) \in \mathrm{Skew}_3(\mathbb{R})$$

defines a coordinate system $(V_{ij})$ in a neighbourhood of $I$ in $O(3)$. [Suggestion. Use theorem 1.4.6]

6. Let $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 1,\ x = c\}$. For which values of $c$ is $S$ a submanifold of $\mathbb{R}^3$? For those $c$ for which it is not, which points have to be excluded so that the remaining set is a submanifold? Sketch.

7. a) Let $S$ be the subset of $\mathbb{R}^3$ given by an equation $z = f(x, y)$ where $f$ is a smooth function. (I.e. $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid z = f(x, y)\}$ is the *graph* of a smooth function $f$ of two variables.) Show that $S$ is a 2-dimensional submanifold of $\mathbb{R}^3$.
b) Let $S$ be a 2-dimensional submanifold of $\mathbb{R}^3$. Show that $S$ can locally be given by an equation $z = f(x, y)$ where $(x, y, z)$ the Cartesian coordinates *in a suitable order*. ( I.e. for every point $p_o \in S$ one can find a smooth function $f$ defined in some neighbourhood $U$ of $p_o$ in $\mathbb{R}^3$ so that $S \bigcap U$ consists of all points $p = (x, y, z)$ in $U$ satisfying $z = f(x, y)$.)[Suggestion. Use the *Implicit Function Theorem.* See §1.1, Exercise 16 for the statement. ]

8. State and prove a generalization of the previous problem, parts (a) and (b), for a $m$-dimensional submanifolds $S$ of $\mathbb{R}^n$.

9. Let $S$ be the *helicoid* in $\mathbb{R}^3$ with equation $z = \theta$ in cylindrical coordinates, i.e.

$$S = \{p = (x, y, z) \in \mathbb{R}^3 \mid z = \theta,\ x = r\cos\theta,\ y = r\sin\theta,\ r > 0, \theta < \infty\}.$$

a) Sketch $S$. Find an equation for $S$ in Cartesian coordinates.
b) Show that $S$ is a 2-dimensional submanifold of $\mathbb{R}^3$.
c) Find all points (if any) in a neighbourhood of which one can use $(r, \theta)$ as coordinates on $S$.
d) Same for $(r, z)$.
e) Same for $(x, y)$.
f) Same for $(\theta, z)$.

10. Let $C$: $p = p(t)$, $-\infty < t < \infty$, be a smooth curve in $\mathbb{R}^3$. The *tangential developable* is the surface $S$ in $\mathbb{R}^3$ swept out by the tangent line of $C$, i.e.

$$S = \{p = p(u) + v\dot{p}(u) \mid -\infty < u, v < \infty\}.$$

a) Sketch $S$ for some $C$ of your choice illustrating the general idea.
b) Which parameter points $(u_o, v_o)$,if any, have a neighbourhood $U$ so that the part $S_U$ of $S$ parametrized by $U$ is a submanifold of $\mathbb{R}^3$? Explain what this means geometrically. Give an example.
c) Determine the tangent space to $S_U$ at a point $p_o$ with parameters $(u_o, v_o)$ as in (b) as a subspace of $\mathbb{R}^3$. [Use Theorem 1.4.6 for parts (b) and (c)]

11. Let $C$ be the helix in $\mathbb{R}^3$ with parametric equations

$$x = \cos t,\ y = \sin t,\ z = t.$$

Let $S$ be the surface swept out by the tangent line of $C$. [See previous problem].
a) Find a parametric equations $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ for $S$.
b) Which points of $S$ have to be omitted (if any) so that the rest is a submanifold of $\mathbb{R}^3$? Prove your answer. [This is not a special case of the previous problem. Explain why not.]
*In exercises 11−16 a set $S$ in $\mathbb{R}^2$ or in $\mathbb{R}^3$ is given.*
*a) Find the regular points of $S$ and specify a coordinate system around each regular point.*
*b) Find all singular points of $S$, if any. (Prove that these points are singular.) If $S$ depends parameters $a, b, \cdots$ you may have to consider various cases, depending on the values of the parameters. Try to sketch $S$.*

12. The surface with parametric equations

$$x = \frac{2au^2}{1+u^2}, \quad y = \frac{au(u^2-1)}{1+u^2}, \quad z = v.$$

13. The set of all points $P$ in a plane for which the product of the distances to two given points $F_1, F_2$ has a constant value $a^2$. (Denote the distance between $F_1$ and $F_2$ by $b > 0$. The set of these points $P$ is called the *ovals of Cassini.*)

14. The curve with equation $r = 2a\cos\theta + 2b$ in polar coordinates $(a, b \geq 0)$.

15. The curve in the plane with parametric equations $x = \cos^3 t$, $y = \sin^3 t$.

16. The endpoints of a mobile line segment $AB$ of constant length $2a$ are constrained to glide on the coordinate axis in the plane. Let $OP$ be the perpendicular to $AB$ from the origin $O$ to a point $P$ on $AB$. $S$ is the set of all possible positions of $P$. [Suggestion. Show that $r = a\sin 2\theta$.]

17. The surface obtained by rotating the curve $z$ =sin$y$ in the $yz$-plane about the $z$-axis.

18. Show that $TS^2$ is diffeomorphic with the submanifold $V$ of $\mathbb{R}^6$ defined by

$$V = \{(x, y, z; v, \eta, \zeta) : x^2 + y^2 + z^2 = 1,\ vx + \eta y + \zeta z = 0\}.$$

Use the following steps.
(a) Prove that $V$ is a submanifold of $\mathbb{R}^3 \times \mathbb{R}^3 = \mathbb{R}^6$.

(b) Let $i : S^2 \to \mathbb{R}^3$ be the inclusion map and $F = di : TS^2 \to T\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3$ its differential. Show that $F(TS^2) \subset V$ and that $F\colon\ TS^2 \to V$ is a diffeomorphism. [Suggestion. Start by specifying coordinates on $S^2$ and on $\mathbb{R}^3$ and write out a formula for $i : S^2 \to \mathbb{R}^3$ in these coordinates.)

19. Let $M$ be a manifold, $S$ a submanifold of $M$.
a) Show that a subset $V$ of $S$ is open in $S$ if and only if $V = S \bigcap U$ for some open subset $V$ of $M$.
b) Show that a partially defined function $g : S \cdots \to \mathbb{R}$ defined on an open subset of $S$ is smooth (for the manifold structure on $S$) if an only if it extends to a smooth function $f : M \cdots \to \mathbb{R}$ locally, i.e. in a neighbourhood in $M$ of each point where $g$ is defined.
c) Show that the properties (a) and (b) characterize the manifold structure of $S$ uniquely.

20. Let $M$ be an $n$–dimensional manifold, $S$ an $m$–dimensional submanifold of $M$. For $p \in S$, let

$$T_p^{\perp}S = \{w \in T_p^*M : w(T_pS) = 0\}$$

and let $T^{\perp}S \subset T^*M$ be the union of the $T_p^{\perp}S$. ($T^{\perp}S$ is called the *conormal bundle* of $S$ in $M$).
a) Show that $T^{\perp}S$ is an $n$–dimensional submanifold of $T^*M$.
b) Show that the canonical 1–form $\theta$ on $T^*M$ (previous problem) is zero on tangent vectors to $T^{\perp}S$.

21. Let $S$ be the surface in $\mathbb{R}^3$ obtained by rotating the circle $(y-a)^2 + z^2 = b^2$ in the $yz$–plane about the $z$–axis. Find the regular points of $S$. Explain why the remaining points (if any) are not regular. Determine a subset of $S$, as large as possible, on which one can use $(x, y)$ as coordinates. [You may have to distinguish various cases, depending on the values of $a$ and $b$. Cover all possibilities.]

22. Let $M$ be a manifold, $S$ a submanifold of $M$, $i : S \to M$ the inclusion map, $di : TS \to TM$ its differential. Prove that $di(TS)$ is a submanifold of $TM$. What is its dimension? [Suggestion. Use the definition of submanifold.]

23. Let $S$ be the surface in $\mathbb{R}^3$ obtained by rotating the circle $(y-a)^2 + z^2 = b^2$ $(a > b > 0)$ in the $yz$–plane about the $z$–axis.
(a) Find an equation for $S$ and show that $S$ is a submanifold of $\mathbb{R}^3$.
(b) Find a diffeomorphism $F : S^1 \times S^1 \to S$, expressed in the form

$$F\colon\ x = x(\phi, \theta), y = y(\phi, \theta),\ z = z(\phi, \theta)$$

where $\phi$ and $\theta$ are the usual angular coordinates on the circles. [Explain why $F$ it is smooth. Argue geometrically that it is one–to–one and onto. Then prove its inverse is also smooth.]

24. Let $S$ be the surface in $\mathbb{R}^3$ obtained by rotating the circle $(y-a)^2 + z^2 = b^2$ $(a > b > 0)$ in the $yz$–plane about the $z$–axis and let $C_s$ be the intersection of

$S$ with the plane $z = s$. Sketch. Determine all values of $s \in \mathbb{R}$ for which $C_s$ a submanifold of $\mathbb{R}^3$ and specify its dimension. (Prove your answer.)

25. Under the hypothesis of part (a) of Theorem 1.4.3, prove that in part (b) the linear independence of all $n$ differentials $(df^1, \cdots, df^n)$ at $p$ is necessary for $f^1|_S, \cdots f^m|_S$ to form a coordinate system around $p$ on $S$.

26. (a)Prove the parenthetical assertion in proof 1 of theorem 1.4.3. (b)Same for 1.4.6.

## 1.5 Riemann metrics

A manifold does not come with equipped with a notion of "metric" just in virtue of its definition, and there is no natural way to define such a notion using only the manifold structure. The reason is that the definition of "metric" on $\mathbb{R}^n$ (in terms of Euclidean distance) is neither local nor invariant under diffeomorphisms, hence cannot be transferred to manifolds, in contrast to notions like "differentiable function" or "tangent vector". To introduce a metric on a manifold one has to add a new piece of structure, in addition to the manifold structure with which it comes equipped by virtue of the axioms. Nevertheless, a notion of metric on a manifold can be introduced by a generalization of a notion of metric of the type one has in a Euclidean space and on surfaces therein. So we look at some examples of these first.

1.5.1 Some preliminary examples. The Euclidean metric in $\mathbb{R}^3$ is characterized by the distance function

$$\sqrt{(x_1 - x_0)^2 - (y_1 - y_0)^2 - (z_1 - z_0)^2}.$$

One then uses this metric to define the length of a curve $p(t) = (x(t), y(t)z(t))$ between $t = a$ and $t = b$ by a limiting procedure, which leads to the expression

$$\int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt.$$

As a piece of notation, this expression often is written as

$$\int_a^b \sqrt{(dx)^2 + (dy)^2 + (dz)^2}.$$

The integrand is called the element of arc and denoted by $ds$, but its meaning remains somewhat mysterious if introduced in this formal, symbolic way. It actually has a perfectly precise meaning: it is a function on the set of all tangent vectors on $\mathbb{R}^3$, since the coordinate differentials $dx, dy, dz$ are functions on tangent vectors. But the notation $ds$ for this function, is truly objectionable: $ds$ is not the differential of any function $s$. The notation is too old to change and besides gives this simple object a pleasantly old fashioned flavour. The function $ds$ on tangent vectors characterizes the Euclidean metric just a well as the distance functions we started out with. It is convenient to get rid of the square root and write $dx^2$ instead of $(dx)^2$ so that the square of $ds$ becomes:

$$ds^2 = dx^2 + dy^2 + dz^2.$$

This is now a quadratic function on tangent vectors which can be used to characterize the Euclidean metric on $\mathbb{R}^3$. For our purposes this $ds^2$ is more a suitable object than the Euclidean distance function, so we shall simply call $ds^2$ itself the metric. One reason why it is more suitable is that $ds^2$ can be easily written down in any coordinate system. For example, in cylindrical coordinates we use $x = r\cos\theta$, $y = r\sin\theta$, $z = z$ to express $dx, dy, dz$ in terms of $dr, d\theta, dz$ and substitute into the above expression for $ds^2$; similarly for spherical coordinates $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$, $z = \rho\,\cos\phi$. This gives

$$\begin{aligned}&\text{cylindrical: } dr^2 + r^2 d\theta^2 + dz^2,\\ &\text{spherical: } d\rho^2 + \rho^2\sin^2\phi\, d\theta^2 + d\phi^2.\end{aligned}$$

Again, $dr^2$ means $(dr)^2$ etc. The same discussion applies to the Euclidean metric in $\mathbb{R}^n$ in any dimension. For polar coordinates in the plane $\mathbb{R}^2$ one can draw a picture to illustrate the $ds^2$ in manner familiar from analytic geometry:

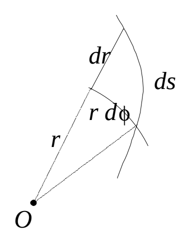



Fig. 1. $ds$ in polar coordinates: $ds^2 = dr^2 + r^2\,d\theta^2$

In Cartesian coordinates $(x^i)$ the metric in $\mathbb{R}^n$ is $ds^2 = \sum(dx^i)^2$. To find its expression in arbitrary coordinates $(y^i)$ one uses the coordinate transformation $x^i = x^i(y^1, \cdots, y^n)$ to express the $dx^i$ in terms of the $dy^i$:

$$dx^i = \sum_k \frac{\partial x^i}{\partial y^k} dy^k \text{ gives } \sum_{\mathrm{I}} (dx^i)^2 = \sum_{\mathrm{I}} \Big(\sum_k \frac{\partial x^i}{\partial y^k} dy^k\Big)^2 = \sum_{\mathrm{I}kl} \frac{\partial x^i}{\partial y^k}\frac{\partial x^i}{\partial y^l} dy^k dy^l.$$

Hence

$$\sum_i (dx^i)^2 = \sum_{kl} g_{kl} dy^k dy^l, \text{ where } g_{kl} = \sum_{\mathrm{i}} \frac{\partial x^i}{\partial y^k}\frac{\partial x^i}{\partial y^l}.$$

The whole discussion applies equally to the case when we start with some other quadratic function instead to the Euclidean metric. A case of importance in physics is the Minkowski metric on $\mathbb{R}^4$, which is given by the formula

$$(dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.$$

More generally, one can consider a pseudo-Euclidean metric on $\mathbb{R}^n$ given by

$$\pm(dx^1)^2 \pm (dx^1)^2 \pm \cdots \pm (dx^n)^2$$

for some choice of the signs.
A generalization into another direction comes form the consideration of surfaces in $\mathbb{R}^3$, a subject which goes back to Gauss (at least) and which motivated Riemann in his investigation of the metrics now named after him. Consider a smooth surface $S$ (2–dimensional submanifold) in $\mathbb{R}^3$. A curve $p(t)$ on $S$ can be considered as a special kind of curve in $\mathbb{R}^3$, so we can defined its length between $t = a$ and $t = b$ by the same formula as before:

$$\int_a^b \sqrt{dx^2 + dy^2 + dz^2}.$$

In this formulas we consider $dx, dy, dz$ as functions of tangent vectors to $S$, namely the differentials of the restrictions to $S$ of the Cartesian coordinate functions $x, y, z$ on $\mathbb{R}^3$. The integral is understood in the same sense as before: we evaluate $dx, dy, dz$ on the tangent vector $\dot{p}(t)$ of $p(t)$ and integrate with respect to $t$. (Equivalently, we may substitute directly $x(t), y(t), z(t)$ for $x, y, z$.)
In terms of a coordinate system $(u, v)$ of $S$, we can write

$$x = x(u, v), y = y(u, v), z = z(u, v))$$

for the Cartesian coordinates of the point $p(u, v)$ on $S$ . Then on $S$ we have

$$dx^2 + dy^2 + dz^2 = \Big(\frac{\partial x}{\partial u}du + \frac{\partial x}{\partial v}dv\Big)^2 + \Big(\frac{\partial y}{\partial u}du + \frac{\partial y}{\partial v}dv\Big)^2 + \Big(\frac{\partial z}{\partial u}du + \frac{\partial z}{\partial v}dv\Big)^2$$

which we can write as

$$ds^2 = g_{uu}du^2 + 2g_{uv}dudv + g_{vv}dv^2$$

where the coefficients $g_{uu}, 2g_{uv}, g_{vv}$ are obtained by expanding the squares in the previous equation and the symbol $ds^2$ denotes the quadratic function on tangent vectors to $S$ defined by this equation. This function can be thought of as defining the metric on $S$, in the sense that it allows one to compute the length of curves on $S$, and hence the distance between two points as the infimum over the length of curves joining them.
As a specific example, take for $S$ the sphere $x^2 + y^2 + z^2 = R^2$ of radius $R$. (We don't take $R = 1$ here in order to see the dependence of the metric on the radius.) By definition, the metric on $S$ is obtained from the metric on $\mathbb{R}^3$ by restriction. Since the coordinates $(\theta, \phi)$ on $S$ are the restrictions of the spherical coordinates $(\rho, \theta, \phi)$ on $\mathbb{R}^3$, we immediately obtain the $ds^2$ on $S$ from that on $\mathbb{R}^3$ :

$$ds^2 = R^2 \sin^2 \phi \, d\theta^2 + d\phi^2.$$

(Since $\rho \equiv R$ on $S$, $d\rho \equiv 0$ on $TS$).

The examples make it clear how a metric on manifold $M$ should be defined: it should be a function on tangent vectors to $M$ which looks like $\sum g_{ij}dx^i dx^j$ in a coordinate system $(x^i)$, perhaps with further properties to be specified. (We don't want $g_{ij} = 0$ for all $ij$, for example.) Such a function is called a quadratic form, a concept which makes sense on any vector space, as we shall now discuss.

1.5.2 **Definitions. Bilinear forms and quadratic forms**. Let $V$ be an $n$–dimensional real vector space. A *bilinear form* on $V$ is a function $g : V \times V \to \mathbb{R}$, $(u, v) \to g(u, v)$, which is linear in each variable separately, i.e. satisfies

$$g(u,v\mathrm{v} + w) = g(u, v) + g(u, w); \quad g(u + v, w) = g(u, w) + g(v, w)$$
$$g(au, v) = ag(u, v); \quad g(u, av) = ag(u, v)$$

for all $u, v \in V$, $a \in \mathbb{R}$. It is *symmetric* if

$$g(u, v) = g(v, \mathrm{u})$$

and it is *non-degenerate* if

$$g(u, v) = 0 \text{ for all } v \in V \text{ implies } u = 0.$$

A bilinear form $g$ on $V$ gives a linear map $V \to V^*$, $v \to g(v, \cdot)$ and to say that $g$ is non–degenerate means that $V \to V^*$ has kernel $\{0\}$, hence is an isomorphism, since $\dim V^* = \dim V$. The *quadratic form* associated to a bilinear form $g$ is the function $Q : V \to \mathbb{R}$ defined by $Q(v) = g(v, v)$. If $g$ is symmetric then $Q$ determines $g$ by the formula

$$Q(u + v) = Q(u) + Q(v) + 2g(u, v).$$

Hence a symmetric bilinear form $g$ is essentially "the same thing" as a quadratic form $Q$. One says that $g$ or $Q$ is *positive definite* if $Q(v) > 0$ for $v \neq 0$. Then $g$ is necessarily non–degenerate. If $(e_i)$ is a basis of $V$ and if one writes $v = v^i e_i$, then one has $g(v, w) = g_{ij}v^i w^j$ where $g_{ij} = g(e_i, e_j)$.

**Remark**. Any non-degenerate quadratic form $Q$ can be expressed as

$$Q = \pm(\xi^1)^2 \pm \cdots \pm (\xi^n)^2$$

where $\xi^i$ are the component functionals with respect to a suitable basis $e_i$, i.e. $v = \xi^i(v)e_i$. Such a basis is called *orthonormal* for $Q$. The number of $\pm$signs is independent of the basis and is called the *signature* of the form.
After this excursion into linear algebra we now return to manifolds.

1.5.3 **Definition**. A *Riemann metric* on $M$ associates to each $p \in M$ a non–degenerate symmetric bilinear form $g_p$ on $T_pM$.

The corresponding quadratic form $Q$ is denoted $ds^2$ and determines $g$ uniquely. Relative to a coordinate system $(x^i)$ we can write

$$ds^2 = \sum g_{ij}dx^i dx^j.$$

The coefficients $g_{ij}$ are given by

$$g_{ij} = g(\partial/\partial x^i, \partial/\partial x^j).$$

As part of the definition we require that the $g_{ij}$ are smooth functions. This is requirement evidently independent of the coordinates. An equivalent condition is that $g(X, Y)$ be a smooth function for any two smooth vector fields $X, Y$ on $M$ or an open subset of $M$.

We add some remarks. (1)We emphasize once more that $ds^2$ is not the differential of some function $s^2$ on $M$, nor the square of the differential of a function $s$ (except when $\dim M = 0, 1$); the notation is rather explained by the examples discussed earlier.
(2) The term "Riemann metric" is sometimes reserved for positive definite metrics and then "pseudo-Riemann" or "semi-Riemann" is used for the possibly indefinite case. A manifold together with a Riemann metric is referred to as a *Riemannian manifold*, and may again be qualified as "pseudo" or "semi".
(3)*At a given point* $p_o \in M$ one can find a coordinate system $(\tilde{x}^i)$ so that the metric $ds^2 = g_{ij}dx^i dx^j$ takes on the pseudo-Euclidean form $\pm\delta_{ij}d\tilde{x}^i d\tilde{x}^j$. [Reason: as remarked above, quadratic form $g_{ij}(p_o)\xi^i\xi^j$ can be written as $\pm\delta_{ij}\tilde{\xi}^i\tilde{\xi}^j$ in a suitable basis, i.e. by a linear transformation $\xi^i = a^i_j\tilde{\xi}^j$, which can be used as a coordinate transformation $x^i = a^i_j\tilde{x}^j$.] But it is generally not possible to do this simultaneously at all points in a coordinate domain, not even in arbitrarily small neighbourhoods of a given point.
(4)As a bilinear form on tangent spaces, a Riemann metric is a conceptually very simple piece of structure on a manifold, and the elaborate discussion of arclength etc. may indeed seem superfluous. But a look at some surfaces should be enough to see the essence of a Riemann metric is not to be found in the algebra of bilinear forms.

For reference we record the transformation rule for the coefficients of the metric, but we omit the verification.

1.5.4 **Lemma**. *The coefficients* $g_{ij}$ *and* $\tilde{g}_{kl}$ *of the metric with respect to two coordinate systems* $(x^i)$ *and* $(\tilde{x}^j)$ *are related by the transformation law*

$$\tilde{g}_{kl} = g_{ij}\frac{\partial x^i}{\partial \tilde{x}^k}\frac{\partial x^j}{\partial \tilde{x}^l}.$$

We now consider a submanifold $S$ of a manifold $M$ equipped with a Riemann metric $g$. Since the tangent spaces of $S$ are subspaces to those of $M$ we can restrict $g$ to a bilinear form $g_S$ on the tangent spaces of $S$. If the metric on $M$ is positive definite, then so is $g_S$. In general, however, it happen that this form $g_S$ is degenerate; but if it is non–degenerate at every point of $S$, then $g_S$ is a Riemann metric on $S$, called the called the Riemann metric on $S$ *induced* by the Riemann metric $g$ on $M$. For example, the metric on a surface $S$ in $\mathbb{R}^3$ discussed earlier is induced from the Euclidean metric in $\mathbb{R}^3$ in this sense.

A Riemann metric on a manifold makes it possible to define a number of geometric concepts familiar from calculus, for example volume.

1.5.5 **Definition**. Let $R$ be a bounded region contained is the domain of a coordinate system $(x^i)$. The *volume* of $R$ (with respect to the Riemann metric $g$) is

$$\int \cdots \int \sqrt{|\det g_{ij}|} dx^1 \cdots dx^n$$

where the integral is over the coordinate-region $\{(x^i(p)) \mid p \in R\}$ corresponding to the points in $R$. If $M$ is two-dimensional one says *area* instead of volume, if $M$ is one-dimensional one says *arclength*.

1.5.6 **Proposition**. *The above definition of volume is independent of the coordinate system.*

**Proof**. Let $(\tilde{x}^i)$ be another coordinate system. Then

$$\tilde{g}_{kl} = g_{ij} \frac{\partial x^i}{\partial \tilde{x}^k} \frac{\partial x^j}{\partial \tilde{x}^l}.$$

So

$$\det \tilde{g}_{kl} = \det(g_{ij} \frac{\partial x^i}{\partial \tilde{x}^k} \frac{\partial x^j}{\partial \tilde{x}^l}) = \det g_{ij} (\det \frac{\partial x^i}{\partial \tilde{x}^k})^2,$$

and

$$\int \cdots \int \sqrt{|\det \tilde{g}_{ij}|} d\tilde{x}^1 \cdots d\tilde{x}^n = \int \cdots \int \sqrt{|\det g_{ij}|} |\det \frac{\partial x^i}{\partial \tilde{x}^k}| d\tilde{x}^1 \cdots d\tilde{x}^n$$

$$= \int \cdots \int \sqrt{|\det g_{ij}|} dx^1 \cdots dx^n,$$

by the change of variables formula. □

1.5.7 **Remarks**.
(a) If $f$ is a real-valued function one can define the *integral* of $f$ over $R$ by the formula

$$\int \cdots \int f \sqrt{|\det g_{ij}|}\, dx^1 \cdots dx^n,$$

provided $f(p)$ is an integrable function of the coordinates $(x^i)$ of $p$. The same proof as above shows that this definition is independent of the coordinate system.
(b) If the region $R$ is not contained in the domain of a single coordinate system the integral over $R$ may defined by subdividing $R$ into smaller regions, just as one does in calculus.

1.5.8 **Example**. Let $S$ be a two-dimensional submanifold of $\mathbb{R}^3$ (smooth surface). The Euclidean metric $dx^2 + dy^2 + dz^2$on $\mathbb{R}^3$ gives a Riemann metric $g = ds^2$ on

$S$ by restriction. Let $u, v$ coordinates on $S$. Write $p = p(u, v)$ for the point on $S$ with coordinates $(u, v)$. Then

$$\sqrt{|\det g_{ij}|} = \left\| \frac{\partial p}{\partial u} \times \frac{\partial p}{\partial v} \right\|.$$

Here $g_{ij}$ is the matrix of the Riemann metric in the coordinate system $u, v$. The right-hand side is the norm of the cross-product of vectors in $\mathbb{R}^3$. This shows that the "volume" defined above agrees with the usual definition of surface area for a surface in $\mathbb{R}^3$.

We now consider the problem of finding the shortest line between two points on a Riemannian manifold with a positive definite metric. The problem is this. Fix two points $A, B$ in $M$ and consider curves $p = p(t)$, $a \leq t \leq b$ from $p(a) = A$ to $p(b) = B$. Given such a curve, consider its *arc–length*

$$\int_a^b ds = \int_a^b \sqrt{Q} dt.$$

as a function of the curve $p(t)$. The integrand $\sqrt{Q}$ is a function on tangent vectors evaluated at the velocity vector $\dot{p}(t)$ of $p(t)$. The integral is independent of the parametrization. We are looking for "the" curve for which this integral is minimal, but we have no guarantee that such a curve is unique or even exists. This is what is called a *variational problem* and it will be best to consider it in a more general setting.

Consider a function $S$ on the set of *paths* $p(t)$, $a \leq t \leq b$ between two given points $A = p(a)$ and $B = p(b)$ of form

$$S = \int_a^b L(p, \dot{p}) dt. \tag{1}$$

We now use the term "path" to emphasize that in general the parametrization is important: $p(t)$ is to be considered as function on a given interval $[a, b]$, which we assume to be differentiable. The integrand $L$ is assumed to be a given function $L = L(p, v)$ on the tangent vectors on $M$, i.e. a function on the tangent bundle $TM$. We here denote elements of $TM$ as pairs $(p, v)$, $p \in M$ being the point at which the vector $v \in T_pM$ is located. The problem is to find the path or paths $p(t)$ for which make $S$ a maximum, a minimum, or more generally *stationary*, in the following sense.

Consider a one–parameter family of paths $p = p(t, \epsilon)$, $a \leq t \leq b$, $-\alpha \leq \epsilon \leq \alpha$, from $A = p(a, \epsilon)$ to $B = p(b, \epsilon)$ which agrees with a given path $p = p(t)$ for $\epsilon = 0$. (We assume $p(t, \epsilon)$ is at least of class $C^2$ in $(t, \epsilon)$.)

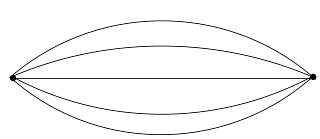


Fig. 2. The paths $p(t, \epsilon)$

Then $S$ evaluated at the path $p(\epsilon, t)$ becomes a function $S$ of $\epsilon$ and if $p(t) = p(0,t)$ makes $S$ a maximum or a minimum then

$$\Big(\frac{dS}{d\epsilon}\Big)_{\epsilon=0} = 0 \tag{2}$$

for all such one–parameter *variations* $p(\epsilon, t)$ of $p(t)$. The converse is not necessarily true, but any path $p(t)$ for which (2) holds for all variations $p(\epsilon, t)$ of $p(t)$ is called a *stationary* (or *critical*) path for the path–function $S$.
We now compute the derivative (2). Choose a coordinate system $x = (x^i)$ on $M$. We assume that the curve $p(t)$ under consideration lies in the coordinate domain, but this is not essential: otherwise we would have to cover the curve by several coordinate systems. We get a coordinate system $(x, \xi)$ on $TM$ by taking as coordinates of $(p, v)$ the coordinates $x^i$ of $p$ together with the components $\xi^i$ of $v$. (Thus $\xi^i = dx^i$ as function on tangent vectors.) Let $x = x(t, \epsilon)$ be the coordinate point of $p(t, \epsilon)$, and in (2) set $L = L(x, \xi)$ evaluated at $\xi = \dot{x}$. Then

$$\frac{dS}{d\epsilon} = \int_a^b \frac{\partial L}{\partial \epsilon} dt = \int_a^b \Big(\frac{\partial L}{\partial x^k}\frac{\partial x^k}{\partial \epsilon} + \frac{\partial L}{\partial \xi^k}\frac{\partial \dot{x}^k}{\partial \epsilon}\Big) dt.$$

Change the order of the differentiation with respect to $t$ and $\epsilon$ in the second term in parentheses and integrate by parts to find that this

$$= \int_a^b \Big(\frac{\partial L}{\partial x^k} - \frac{d}{dt}\frac{\partial L}{\partial \xi^k}\Big)\frac{\partial x^k}{\partial \epsilon} dt + \Big[\frac{\partial L}{\partial \xi^k}\frac{\partial x^k}{\partial \epsilon}\Big]_{t=a}^{t=b}$$

The terms in brackets is zero, because of the boundary conditions $x(\epsilon, a) \equiv x(A)$ and $x(\epsilon, b) \equiv x(B)$. The whole expression has to vanish at $\epsilon = 0$, for all $x = x(t, \epsilon)$ satisfying the boundary conditions. The partial derivative $\partial x^k/\partial \epsilon$ at $\epsilon = 0$ can be any $C^1$ function $w^k(t)$ of $t$ which vanishes at $t = a, b$ since we can take $x^k(t, \epsilon) = x^k(t) + \epsilon w^k(t)$, for example. Thus the initial curve $x = x(t)$ satisfies

$$\int_a^b \Big(\frac{\partial L}{\partial x^k} - \frac{d}{dt}\frac{\partial L}{\partial \xi^k}\Big) w^k(t) dt = 0$$

for all such $w^k(t)$. From this one can conclude that

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\frac{\partial L}{\partial \xi^k} = 0. \tag{3}$$

for all $t$, $a \le t \le b$ and all $k$. For otherwise there is a $k$ so that this expression is non-zero, say positive, at some $t_o$, hence in some interval about $t_o$. For this $k$, choose $w^k$ equal to zero outside such an interval and positive on a subinterval and take all other $w^k$ equal to zero. The integral will be positive as well, contrary to the assumption. The equation (3) is called the *Euler–Lagrange* equation for the variational problem (2).

We add some remarks on what is called the *principle of conservation of energy* connected with the variational problem (2). The energy $E = E(p, v)$ associated to $L = L(p, v)$ is the function defined by

$$E = \frac{\partial L}{\partial \xi^k}\xi^k - L$$

in a coordinate system $(x, \xi)$ on $TM$ as above. (It is actually independent of the choice of the coordinate system $x$ on $M$). If $x = x(t)$ satisfies the Euler–Lagrange equation (3) and we take $\xi = \dot{x}$, then $E = E(x, \xi)$ satisfies

$$\frac{dE}{dt} = \frac{d}{dt}\Big(\frac{\partial L}{\partial \xi^k}\dot{x}^k - L\Big) = \frac{d}{dt}\Big(\frac{\partial L}{\partial \xi^k}\Big)\dot{x}^k + \frac{\partial L}{\partial \xi^k}\ddot{x}^k - \Big(\frac{\partial L}{\partial x^k}\Big)\dot{x}^k - \frac{\partial L}{\partial \xi^k}\ddot{x}^k \tag{4}$$
$$= \Big(\frac{d}{dt}\frac{\partial L}{\partial \xi^k} - \frac{\partial L}{\partial x^k}\Big)\dot{x}^k = 0.$$

Thus $E$ =constant along the curve $p(t)$.

We now return to the particular problem of arc length. Thus we have to take $L = \sqrt{Q}$ in (1). The integral (1) becomes

$$\int_a^b \sqrt{Q}dt$$

and is in this case independent of the parametrization of the path $p(t)$. The derivative (2) becomes

$$\int_a^b \frac{1}{2}\Big(Q^{-1/2}\ \frac{\partial Q}{\partial \epsilon}\Big)_{\epsilon=0}dt.$$

Because of the independence of parametrization, we may assume that the initial curve is parametrized by arclength, i.e. $Q \equiv 1$ for $\epsilon = 0$. Then we get

$$\frac{1}{2}\int_a^b \Big(\frac{\partial Q}{\partial \epsilon}\Big)_{\epsilon=0}dt$$

This means in effect that we can replace $L = \sqrt{Q}$ by $L = Q = g_{ij}\xi^i\xi^j$. For this new $L$ the energy $E$ becomes $E = 2Q - Q = Q$. So the speed $\sqrt{Q}$ is constant along $p(t)$. We now write out (3) explicitly and summarize the result.

1.5.9 **Theorem**. L*et $M$ be Riemannian manifold. A path $p = p(t)$ makes $\int_a^b Qdt$ stationary if and only if its parametric equations $x^i = x^i(t)$ in any coordinate system $(x^i)$ satisfy*

$$\frac{d}{dt}\Big(g_{ik}\frac{dx^i}{dt}\Big) - \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0. \tag{5}$$

This holds also for a curve of minimal arc–length $\int_a^b \sqrt{Q}dt$ between two given points if traversed at constant speed. Any curve satisfying (5) has constant speed, i.e. $g(\dot{p},\dot{p})$ =constant.

1.5.10 **Definition**. *Any* curve on a Riemannian manifold $M$ satisfying (5) is called a *geodesic* (of the Riemann metric $g$).

For this definition the metric need not be positive definite, but even if it is, a geodesic need not be the shortest curve between any two of its points. (Think of a great circle on a sphere; see the example below.) But if the metric is positive definite it can be shown that for $p$ sufficiently close to $q$ the geodesic from $p$ to $q$ is the unique shortest line.

1.5.11 **Proposition**. *Given a point $p_o \in M$ and a vector $v_o$ at $p_o$ there is a unique geodesic $p = p(t)$, defined in some interval about $t = 0$, so that $p(0) = p_o$ and $\dot{p}(0) = v_o$.*

**Proof**. The system of second-order differential equations (5) has a unique solution $(x^i(t))$ with a given initial condition $(x^i(0))$, $(\dot{x}^i(0))$. □

1.5.12 **Examples**.
*A. Geodesics in Euclidean space.* For the Euclidean metric $ds^2 = (dx^1)^2 + \cdots + (dx^n)^2$ the equations (5) read $d^2x^i/dt^2 = 0$, so the geodesics are the straight lines $x^i(t) = c^i t + x^i_o$ traversed at constant speed. (No great surprise, but good to know that (5) works as expected.)
*B. Geodesics on the sphere.* The metric is $ds^2 = \sin^2\phi d\theta^2 + d\phi^2$ (we take $R = 1$). The equations (5) become

$$\frac{d}{dt}\Big(\sin^2\phi\frac{d\theta}{dt}\Big) = 0, \qquad \frac{d^2\phi}{dt^2} - \sin\phi\cos\phi\Big(\frac{d\theta}{dt}\Big)^2 = 0$$

These equations are obviously satisfied if we take $\theta = \alpha$, $\phi = \nu t$. This is a great circle through the north-pole $(x, y, z) = (0, 0, 1)$ traversed with constant speed. Since any vector at $p_o$ is the tangent vector of some such curve, all geodesics starting at this point are of this type (by the above proposition). Furthermore, given any point $p_o$ on the sphere, we can always choose an orthogonal coordinate system in which $p_o$ has coordinates $(0, 0, 1)$. Hence the geodesics starting at any point (hence all geodesics on the sphere) are great circles traversed at constant speed.

1.5.13 **Definition**. A map $f : M \to N$ between Riemannian manifolds is called a *(local) isometry* if it is a (local) diffeomorphism and preserves the metric, i.e. for all $p \in M$, the differential $df_p : T_pM \to T_{f(p)}N$ satisfies

$$ds^2_M(v) = ds^2_N(df_p(v))$$

for all $v \in T_pM$.

1.5.14 **Remarks**. (1) In terms of scalar products, this is equivalent to

$$g_M(v, w) = g_N(df_p(v), df_p(w))$$

for all $v, w \in T_pM$.
(2) Let $(x^1, \cdots, x^n)$ and $(y^1, \cdots, y^m)$ be coordinates on $M$ and $N$ respectively. Write

$$ds_M^2 = g_{ij}dx^i dx^j \quad ds_N^2 = h_{ab}dy^a dy^b$$

for the metrics and

$$y^a = f^a(x^1, \cdots, x^n),\ a = 1, \cdots, m \tag{6}$$

for the map $f$. To say that $f$ is an isometry means that $ds_N^2$ becomes $ds_M^2$ if we substitute for the $y^a$ and $dy^a$ in terms if $x^i$ and $dx^i$ by means of the equation (6).
(5) If $f$ is an isometry then it preserves arclength of curves as well. Conversely, any smooth map preserving arclength of curves is an isometry. (Exercise).

1.5.15 **Examples**

a) *Euclidean space*. The *linear* transformations of $\mathbb{R}^n$ which preserve the Euclidean metric are of the form $x \to Ax$ where $A$ is an orthogonal real $n \times n$ matrix ($AA^* = I$). The set of all such matrices is called the *orthogonal group*, denoted $O(n)$. If the Euclidean metric is replaced by a pseudo-Euclidean metric with $p$ plus signs and $q$ minus signs the corresponding set of "pseudo-orthogonal" matrices is denoted $O(p, q)$. It can be shown that any transformation of $\mathbb{R}^n$ which preserves a pseudo-Euclidean metric is of the form $x \to x_o + Ax$ with $A$ linear orthogonal.

b) *The sphere*. Any orthogonal transformation $A \in O(3)$ of $\mathbb{R}^3$ gives an isometry of $S^2$ onto itself, and it can be shown that all isometries of $S^2$ onto itself are of this form. The same holds in higher dimensions and for pseudo-spheres defined by a pseudo-Euclidean metric.

c) *Curves and surfaces*. Let $C : p = p(\sigma)$, $\sigma \in \mathbb{R}$, be a curve in $\mathbb{R}^3$ parametrized by arclength, i.e. $\dot{p}(\sigma)$ has length $= 1$. If we assume that the curve is nonsingular, i.e. a submanifold of $\mathbb{R}^3$, then $C$ is a 1-dimensional manifold with a Riemann metric and the map $\sigma \to p(\sigma)$ is an isometry of $\mathbb{R}$ onto $C$. In fact it can be shown that for *any* connected 1-dimensional Riemann manifold $C$ there exists an isometry of an interval on the Euclidean line $\mathbb{R}$ onto $C$, which one can still call *parametrization by arclength*. The map $\sigma \to p(\sigma)$ is not necessarily 1–1 on $\mathbb{R}$ (e.g. for a circle), but it is always locally 1–1. Thus any 1-dimensional Riemann manifold is locally isometric with the straight line $\mathbb{R}$. This does not hold in dimension 2 or higher: for example a sphere is not locally isometric with the plane $\mathbb{R}^2$ (exercise 18); only very special surfaces are, e.g. cylinders (exercise 17; such surfaces are called developable.)

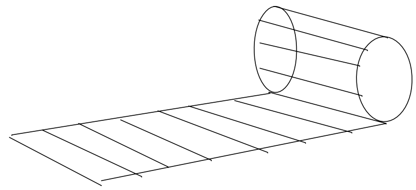


Fig. 3. A cylinder "developed" into a plane

d) *Poincaré disc and Klein upper half-plane.* Let $D = \{z = x + iy \mid x^2 + y^2 < 1\}$ the unit disc and $H = \{w = u + iv \in \mathbb{C} \mid v > 0\}$ be the upper halfplane. The *Poincaré metric* on $D$ and the *Klein metric* on $H$ and are defined by

$$ds_D^2 = \frac{4dzd\bar{z}}{(1 - z\bar{z})^2} = 4\frac{dx^2 + dy^2}{(1 - x^2 - y^2)^2} \qquad ds_H^2 = \frac{-4dwd\bar{w}}{(w - \bar{w})^2} = \frac{du^2 + dv^2}{v^2} \ .$$

The complex coordinates $z$ and $w$ are only used as shorthand for the real coordinates $(x, y)$ and $(u, v)$. The map $f : w \to z$ defined by

$$z = \frac{1 + iw}{1 - iw}$$

sends $H$ onto $D$ and is an isometry for these metrics. To verify the latter use the above formula for $z$ and the formula

$$dz = \frac{2idw}{(1 - iw)^2}$$

for $dz$ to substitute into the equation for $ds_D^2$; the result is $ds_H^2$. (Exercise.)

1.5.16 The set of all invertible isometries of a Riemannian metric $ds^2$ manifold $M$ is called the isometry group of $M$, denoted Isom$(M)$ or Isom$(M, ds^2)$. It is a group in the algebraic sense, which means that the composite of two isometries is an isometry as is the inverse of any one isometry. In contrast to the above example, the isometry group of a general Riemann manifold may well consist of the identity transformation only, as is easy to believe if one thinks of a general surface in space.

1.5.17 On a connected manifold $M$ with a positive-definite Riemann metric one can define the distance between two points as the infimum of the lengths of all curves joining these points. (It can be shown that this makes $M$ into a metric space in the sense of topology.)

**EXERCISES 1.5**

1. Verify the formula for the Euclidean metric in spherical coordinates:

$$ds^2 = d\rho^2 + \rho^2 \sin^2 \phi \, d\theta^2 + d\phi^2.$$

2. Using $(x, y)$ as coordinates on $S^2$, show that the metric on

$$S^2 \;= \{p = (x, y, z) \mid x^2 + y^2 + z^2 = R^2\}$$

is given by

$$dx^2 + dy^2 + \frac{(xdx + ydy)^2}{R^2 - x^2 - y^2}$$

Specify a domain for the coordinates $(x, y)$ on $S^2$.

3. Let $M = \mathbb{R}^3$ . Let $(x^0, x^1, x^2)$ be Cartesian coordinates on $\mathbb{R}^3$. Define a metric $ds^2$ by

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2.$$

Pseudo-spherical coordinates $(\rho, \psi, \theta)$ on $\mathbb{R}^3$ are defined by the formulas

$$x^0 = \rho \cosh \psi, \;\; x^1 = \rho \sinh \psi \cos \theta, \;\; x^2 = \rho \sinh \psi \sin \theta.$$

Show that

$$ds^2 = -d\rho^2 + \rho^2\, d\psi^2 + \rho^2 \sinh^2 \psi\, d\theta^2.$$

4. Let $M = \mathbb{R}^3$ . Let $(x^0, x^1, x^2)$ be Cartesian coordinates on $\mathbb{R}^3$. Define a metric $ds^2$ by

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2.$$

Let $S = \{p = (x^0, x^1, x^2) \in \mathbb{R}^3 \mid (x^0)^2 - (x^1)^2 - (x^2)^2 = 1\}$ with the metric $ds^2$ induced by the metric $ds^2$ on $\mathbb{R}^3$. (With this metric $S$ is called a *pseudo-sphere*).
(a) Prove that $S$ is a submanifold of $\mathbb{R}^3$.
(b) Show that for any $(\psi, \theta)$ the point $p = (x^0, x^1, x^2)$ given by $x^0 = \cosh \psi$, $x^1 = \sinh \psi \cos \theta$, $x^2 = \sinh \psi \sin \theta$ lies on $S$. Use $(\psi, \theta)$ as coordinates on $S$ and show that

$$ds^2 = d\psi^2 + \sinh^2 \psi\, d\theta^2.$$

(This shows that the induced metric $ds^2$ on $S$ is positive definite.)

5. Prove the transformation rule $\tilde{g}_{kl} = g_{ij} \frac{\partial x^i}{\partial \tilde{x}^k} \frac{\partial x^j}{\partial \tilde{x}^l}$.

6. Prove the formula $\sqrt{|\det g_{ij}|} = \left\| \frac{\partial p}{\partial u} \times \frac{\partial p}{\partial v} \right\|$ for surfaces as stated in the text. [Suggestion. Use the fact that $g_{ij} = g\Big(\frac{\partial p}{\partial x^i}, \frac{\partial p}{\partial x^j}\Big)$. Remember that $\|A \times B\|^2 = \|A\|^2 \|B\|^2 - (A \cdot B)^2$ for two vectors $A, B$ in $\mathbb{R}^3$.]

7. Find the expression for the Euclidean metric $dx^2 + dy^2$ on $\mathbb{R}^2$ using the following coordinates $(u, v)$.

$$\text{a)} x = \cosh u \cos v,\; y = \sinh u \sin v \qquad \text{b)}\; x = \frac{1}{2}(u^2 - v^2),\; y = uv.$$

8. Let $S$ be a surface in $\mathbb{R}^3$ with equation $z = f(x, y)$ where $f$ is a smooth function.
a) Show that $S$ is a submanifold of $\mathbb{R}^3$.

b) Describe the tangent space $T_pS$ at a point $p = (x, y, z)$ of $S$ as a subspace of $T_p\mathbb{R}^3 = \mathbb{R}^3$.
c) Show that $(x, y)$ define a coordinate system on $S$ and that the metric $ds^2$ on $S$ induced by the Euclidean metric $dx^2 + dy^2 + dz^2$ on $\mathbb{R}^3$ is given by

$$ds^2 = (1 + f_x^2)dx^2 + 2f_xf_ydxdy + (1 + f_y^2)dy^2.$$

d) Show that the area of a region $R$ on $S$ is given by the integral

$$\iint \sqrt{1 + f_x^2 + f_y^2}\, dxdy$$

over the coordinate region corresponding to $R$. (Use definition 1.5.5.)

9. Let $S$ be the surface in $\mathbb{R}^3$ obtained by rotating the curve $y = f(z)$ about the $z$–axis.
a) Show that $S$ is given by the equation $x^2 + y^2 = f(z)^2$ and prove that $S$ is a submanifold of $\mathbb{R}^3$. (Assume $f(z) \neq 0$ for all $z$.)
b) Show that the equations $x = f(z)\cos\theta$, $y = f(z)\sin\theta$, $z = z$ define a coordinate system $(z, \theta)$ on $S$. [Suggestion. Use Theorem 1.4.6]
c) Find a formula for the Riemann metric $ds^2$ on $S$ induced by the Euclidean metric $dx^2 + dy^2 + dz^2$ on $\mathbb{R}^3$.
d) Prove the coordinate vector fields $\partial/\partial z$ and $\partial/\partial\theta$ on $S$ are everywhere orthogonal with respect to the inner product $g(u, v)$ of the metric on $S$. Sketch.

10. Show that the generating curves $\theta$ =const. on a surface of rotation are geodesics when parametrized by arclength. Is every geodesic of this type? [See the preceding problem. "Parametrized by arclength" means that the tangent vector has length one.]

11. Determine the geodesics on the pseudo-sphere of problem 4. [Suggestion. Imitate the discussion for the sphere.]

12. Let $S = \{p = (x, y, z) \mid x^2 + y^2 = 1\}$ be a right circular cylinder in $\mathbb{R}^3$. Show that the helix $x = \cos t$, $y = \sin t$, $z = ct$ is a geodesic on $S$. Find all geodesics on $S$.

13. Prove remark 1.5.14 (2).

14. Prove that the map $f$ defined in example 1.5.15(d) does map $D$ onto $H$ and complete the verification that it is an isometry.

15. With any complex 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ associate the *linear fractional transformation* $f : w \to z$ defined by

$$z = \frac{aw + b}{cw + d}.$$

a) Show that a composite of two such transformations corresponds to the product of the two matrices. Deduce that such transformation is invertible if $\det A \neq 0$, and if so, one can arrange $\det A = 1$ without changing $f$.

b)Suppose $A = \begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix}$, $\det A = 1$. Show that $f$ gives an isometry of Poincaré disc $D$ onto itself. [The convention of using the same notation $(x^i)$ for a coordinate system as for the coordinates of a general points may cause some confusion here: in the formula for $f$, both $z = x + iy$ and $w = u + iv$ will belong to $D$; you can think of $(x, y)$ and $(u, v)$ as two different coordinate systems on $D$, although they are really the same.]
c) Suppose $A$ is real and $\det A = 1$. Show that $f$ gives an isometry of Klein upper halfplane $H$ onto itself.

16. Let $P$ be the pseudosphere $x^2 - y^2 - z^2 = 1$ in $\mathbb{R}^3$ with the metric $ds_P^2$ induced by the Pseudo-Euclidean metric $-dx^2 + dy^2 + dz^2$ in $\mathbb{R}^3$.
a) For any point $p = (x, y, z)$ on the upper pseudosphere $x > 0$ let $(u, v, 0)$ be the point where the line from $(-1, 0, 0)$ to $(x, y, z)$ intersects the $yz$-plane. Sketch. Show that

$$x = \frac{2u}{1 - u^2 - v^2}, \quad y = \frac{2v}{1 - u^2 - v^2}, \; z = \frac{2}{1 - u^2 - v^2} - 1.$$

b) Show that the map $(x, y, z) \to (u, v)$ is an isometry from the upper pseudosphere onto the Poincaré disc. [Warning: the calculations in this problem are a bit lengthy.]

17. Let $S$ be the cylinder in $\mathbb{R}^3$ with base a curve in the $xy$-plane $x = x(\sigma)$, $y = y(\sigma)$, defined for all $-\infty < \sigma < \infty$ and parametrized by arclength $\sigma$. Assume that $S$ is a submanifold of $\mathbb{R}^3$ with $(\sigma, z)$ as coordinate system. Show that there is an isometry $f : \mathbb{R}^2 \to S$ from the Euclidean plane onto $S$. Use this fact to find the geodesics on $S$.

18. a) Calculate the circumference $L$ and the area $A$ of a disc of radius $\rho$ about a point of the sphere $S^2$, e.g. the points with $0 \le \phi \le \rho$. [Answer: $L = 2\pi \sin \rho$, $A = 2\pi(1 - \cos \rho)$].
b) Prove that the sphere is not locally isometric with the plane $\mathbb{R}^2$.

## 1.6 Tensors

1.6.1 **Definition**. A *tensor* $T$ at a point $p \in M$ is a quantity which, relative to a coordinate system $(x^k)$ around $p$, is represented by an indexed system of real numbers $(T^{ij\cdots}_{lk\cdots})$. The $(T^{ij\cdots}_{kl\cdots})$ and $(\tilde{T}^{ab\cdots}_{cd\cdots})$ representing $T$ in two coordinate systems $(x^i)$ and $(\tilde{x}^j)$ are related by the transformation law

$$\tilde{T}^{ab\cdots}_{cd\cdots} = T^{ij\cdots}_{kl\cdots} \frac{\partial \tilde{x}^a}{\partial x^i} \frac{\partial \tilde{x}^b}{\partial x^j} \cdots \frac{\partial x^k}{\partial \tilde{x}^c} \frac{\partial x^l}{\partial \tilde{x}^d} \cdots . \tag{1}$$

where the partials are taken at $p$.

The $T^{ij\cdots}_{kl\cdots}$ are called the *components* of $T$ relative to the coordinate system $(x^i)$. If there are r upper indices $ij\cdots$ and s lower indices $kl\cdots$ the tensor is said to be of *type* $(r,s)$. It spite of its frightful appearance, the rule (1) is very easily remembered: the indices on the $\partial\tilde{x}^a/\partial x^i$and $\partial x^k/\partial\tilde{x}^c$ etc. on the right "cancel" against those on $T^{ij\cdots}_{kl\cdots}$ to produce the indices on the left and the tildes on the right are placed up or down like the corresponding indices on the left.

If the undefined term "quantity" in this definition is found objectionable, one can simply take it to *be* a rule which associates to every coordinate system $(x^i)$ around $p$ a system of numbers $(T^{ij\cdots}_{kl\cdots})$. But then a statement like "a magnetic field is a $(0,2)$–tensor" requires some conceptual or linguistic acrobatics of another sort.

1.6.2 **Theorem**. *An indexed system of numbers $(T^{ij\cdots}_{kl\cdots})$ depending on a coordinate system $(x^k)$ around $p$ defines a tensor at $p$ if and only if the function*

$$T(\lambda,\mu,\cdots;v,w,\cdots)=T^{ij\cdots}_{kl\cdots}\lambda_i\mu_j\cdots v^k w^l\cdots$$

*of $r$ covariant vectors $\lambda=(\lambda_i),\mu=(\mu_j)\cdots$ and $s$ contravariant vectors $v=(v^k)$,$w$= $(w^l)\cdots$ is independent of the coordinate system $(x^k)$. This establishes a one–to–one correspondence between tensors and functions of the type indicated.*

**Explanation**. A function of this type is called *multilinear*; it is linear in each of the (co)vectors

$$\lambda=(\lambda_i),\mu=(\mu_j),\cdots,v=(v^k),w=(w^l),\cdots$$

separately (i.e. when all but one of them is kept fixed). The theorem says that one can think of a tensor at $p$ as a multilinear function $T(\lambda,\mu,\cdots;v,w,\cdots)$ of $r$ covectors $\lambda,\mu,\cdots$ and $s$ vectors $v,w,\cdots$ at $p$, independent of any coordinate system. Here we write the covectors first. One could of course list the variables $\lambda,\mu,\cdots;v,w\cdots$ in any order convenient. In components, this is indicated by the position of the indices, e.g.

$$T(v,\lambda,w)=T^{j}_{ik}v^j\lambda_j w^k$$

for a tensor of type $(1,2)$.

**Proof**. The function is independent of the coordinate system if and only if it satisfies

$$\tilde{T}^{ab\cdots}_{cd\cdots}\tilde{\lambda}_a\tilde{\mu}_b\cdots\tilde{v}^c\tilde{w}^d\cdots=T^{ij\cdots}_{kl\cdots}\lambda_i\mu_j\cdots v^k w^l\cdots \tag{2}$$

In view of the transformation rule for vectors and covectors this means that

$$\tilde{T}^{ab\cdots}_{cd\cdots}\tilde{\lambda}_a\tilde{\mu}_b\cdots\tilde{v}^c\tilde{w}^d\cdots=T^{ij\cdots}_{kl\cdots}\frac{\partial\tilde{x}^{\mathrm{a}}}{\partial x^i}\frac{\partial\tilde{x}^{\mathrm{b}}}{\partial x^j}\cdots\frac{\partial x^k}{\partial\tilde{x}^{\mathrm{c}}}\frac{\partial x^l}{\partial\tilde{x}^{\mathrm{d}}}\cdots\tilde{\lambda}_a\tilde{\mu}_b\cdots\tilde{v}^c\tilde{w}^d\cdots$$

Fix a set of indices $a,b,\cdots,c,d,\cdots$. Choose the corresponding components $\tilde{\lambda}_a,\ \tilde{\mu}_b,\ \cdots\tilde{v}^c,\ \tilde{w}^d\cdots$ to be $=1$ and all other components $=0$. This gives the transformation law (1). Conversely if (1) holds then this argument shows that (2) holds as well. That the given rule is a one–to–one correspondence is clear, since relative to a coordinate system both the tensor and the multilinear function are uniquely specified by the system of numbers $(T^{ij\cdots}_{kl\cdots})$. $\square$.

As mentioned, one can think of a tensor as being "the same thing" as a multilinear function, and this is in fact often taken as the definition of "tensor", instead of the transformation law (1). But that point of view is sometimes quite awkward. For example, a linear transformation $A$ of $T_pM$ defines a $(1,1)$ tensor $(A_i^j)$ at $p$, namely its matrix with respect to the basis $\partial/\partial x^i$, and it would make little sense to insist that we should rather think of $A$ a bilinear form on $T_pM \times T_p^*M$. It seems more reasonable to consider the multilinear–form interpretation of tensors as just one of many.

1.6.3 **Definition.** A *tensor field* $T$ associates to each point $p$ of $M$ a tensor $T(p)$ at $p$. The tensor field is of class $\mathrm{C}^k$ if its components with respect to any coordinate system are of class $\mathrm{C}^k$.
Tensor fields are often simply called tensors as well, especially in the physics literature.

1.6.4 **Operations on tensors** at a given point $p$.

(1) *Addition and scalar multiplication.* Componentwise. Only tensors of the same type can be added. E.g. If $T$ has components $(T_{ij})$, and $S$ has components $(S_{ij})$, then $S+T$ has components $S_{ij}+T_{ij}$.

(2) *Symmetry operations.* If the upper or lower indices of a tensor are permuted among themselves the result is again a tensor. E.g if $(T_{ij})$ is a tensor, then the equation $S_{ij}=T_{ji}$ defines another tensor $S$. This tensor may also defined by the formula $S(v,w)=T(w,v)$.

(3) *Tensor product.* The product of two tensors $T=(T_j^i)$ and $S=(S_l^k)$ is the tensor $T\otimes S$ (also denoted $ST$) with components $(T\otimes S)_{jl}^{ik}=T_j^iS_l^k$. The indices $i=(i_1,i_2\cdots)$ etc. can be multi-indices, so that this definition applies to tensors of any type.

(4) *Contraction.* This operation consists of setting an upper index equal to a lower index and summing over it (summation convention). For example, from a $(1,2)$ tensor $(T_{ij}^k)$ one can form two $(0,1)$ tensors $U,V$ by contractions $U_j=T_{kj}^k$, $V_i=T_{ik}^k$ (sum over $k$). One has to verify that the above operations do produce tensors. As an example, we consider the contraction of a $(1,1)$ tensor:

$$\tilde{T}_a^a = T_j^i \frac{\partial \tilde{x}^a}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^a} = T_j^i \delta_i^j = T_i^i.$$

The verification for a general tensor is the same, except that one has some more indices around.

1.6.5 **Lemma**. *Denote by $T_p^{r,s}M$ the set of tensors of type $(r,s)$ at $p\in M$.*
*a)$T_p^{r,s}M$ is a vector space under the operations addition and scalar multiplication.*
*b) Let $(x^i)$ be a coordinate system around $p$. The $r\times s$–fold tensor products tensors*

$$\frac{\partial}{\partial x^i}\otimes\frac{\partial}{\partial x^j}\otimes\cdots\otimes dx^k\otimes dx^l\otimes\cdots \qquad (*)$$

*form a basis for* $T_p^{r,s}M$. *If* $T \in T_p^{r,s}M$ *has components* $(T^{ij\cdots}_{kl\cdots})$ *relative to* $(x^i)$, *then*

$$T = T^{ij\cdots}_{kl\cdots} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \otimes \cdots \otimes dx^k \otimes dx^l \otimes \cdots . \qquad (**)$$

**Proof**. a) is clear. For (b), remember that $\partial/\partial x^i$ has $i$–component $= 1$ and all other components $= 0$, and similarly for $dx^k$. The equation (**) is then clear, since both sides have the same components $T^{ij\cdots}_{kl\cdots}$. This equation shows also that every $T \in T_p^{r,s}M$ is uniquely a linear combination of the tensors (*), so that these do indeed form a basis. □

1.6.6 **Tensors on a Riemannian manifold**.

We now assume that $M$ comes equipped with a Riemann metric $g$, a non–degenerate, symmetric bilinear form on the tangent spaces $T_pM$ which, relative to a coordinate system $(x^i)$, is given by $g(v,w) = g_{ij}v^iw^j$.

1.6.7 **Lemma**. For any $v \in T_pM$ there is a unique element $gv \in T_p^*M$ so that $gv(w) = g(v,w)$ for all $w \in T_pM$. The map $v \to gv$ is a linear isomorphism $g : T_pM \to T_p^*M$.

**Proof**. In coordinates, the equation $\lambda(w) = g(v,w)$ says $\lambda_iw^i = g_{ij}v^iw^j$. This holds for all $w$ if and only if $\lambda_i = g_{ij}v^i$ and thus determines $\lambda = gv$ uniquely. The linear map $v \to gv$ is an isomorphism, because its matrix $(g_{ij})$ is invertible. □

In terms of components the map $v \to g(v)$ is usually denoted $(v^i) \to (v_i)$ and is given by $v_i = g_{ij}v^j$. This shows that the operation of lowering indices on a vector is independent of the coordinate system (but dependent on the metric $g$). The inverse map $T_p^*M \to T_pM$, $\lambda \to v = g^{-1}\lambda$, is given by\ $v^i = g^{ij}\lambda_j$ where $(g^{ij})$ is the inverse matrix of $(g_{ij})$, i.e. $g^{ik}g_{kj} = \delta^i_j$. In components the map $\lambda \to g^{-1}\lambda$ is expressed by raising the indices: $(\lambda_i) \to (\lambda^i)$. Since the operation $v^i \to v_i$ is independent of the coordinate system, so is its inverse operation $\lambda_i \to \lambda^i$. We use this fact to prove the following.

1.6.8 **Lemma**. $(g^{ij})$ is a $(2,0)$-tensor.

**Proof**. For any two covectors $(v_i), (w_j)$, the scalar

$$g^{ij}v_iw_j = g^{ij}g_{ia}v^ag_{jb}w^b = \delta^j_a g_{jb}v^aw^b = g_{ab}v^aw^a \qquad (4)$$

is independent of the coordinate system. Hence the theorem applies. □

1.6.9 **Definition**. The *scalar product* of covectors $\lambda = (\lambda_i)$, $\mu = (\mu_j)$ is defined by $g^{-1}(\lambda,\mu) = g^{ij}\lambda_i\mu_j$ where $(g^{ij})$ is the inverse matrix of $(g_{ij})$. This means that we transfer the scalar product in $T_pM$ to $T_p^*M$ by the isomorphism $T_pM \to T_p^*M$, $v \to gv$, i.e. $g^{-1}(\lambda,\mu) = g(g^{-1}\lambda, g^{-1}\mu)$.

The tensors $(g_{ij})$ and $(g^{ij})$ may be used to raise and lower the indices of arbitrary tensors in a Riemannian space, for example $T_i^j = g^{jk}T_{ik}$, $S_{ij} = g_{ik}S^k_j$. The scalar product $(S,T)$ of any two tensors of the same type is defined by raising or lowering the appropriate indices and contracting, e.g $(S,T) = T_{ij}S^{ij} = T^{ij}S_{ij}$ for $T = (T_{ij})$ and $S = (S_{ij})$.

## Appendix 1: Tensors in Euclidean 3-space $\mathbb{R}^3$

We consider $\mathbb{R}^3$ as a Euclidean space, equipped with its standard scalar product $g(v,w) = v \cdot w$. This allows one to first of all to "identify" vectors with covectors. In Cartesian coordinates $(x^1, x^2, x^3)$ this identification $v \leftrightarrow \lambda$ is simply given by $\lambda_i = v^i$, since $g_{ij} = \delta_{ij}$. This extends to arbitrary tensors, as a special case of raising and lowering indices in a Riemannian manifold. There are further special identifications which are peculiar to three dimensions, concerning alternating $(0,s)$–tensors, i.e. tensors whose components change sign when two indices are exchanged. The alternating condition is automatic for $s = 0, 1$; for $s = 2$ it says $T_{ij} = -T_{ji}$; for $s = 3$ it means that under a permutation $(ijk)$ of $(123)$ the components change according to the sign $\epsilon_{ijk}$ of the permutation: $T_{ijk} = \epsilon_{ijk}T_{123}$. If two indices are equal, the component is zero, since it must change sign under the exchange of the equal indices. Thus an alternating $(0,3)$ tensor $T$ on $\mathbb{R}^3$ is of the form $T_{ijk} = \tau\epsilon_{ijk}$ where $\tau = T_{123}$ is a real number and the other components are zero. This means that as alternating 3–linear form $T(u,v,w)$ of three vectors one has

$$T(u,v,w) = \tau \det[u,v,w]$$

where $\det[u,v,w]$ is the determinant of the matrix of components of $u, v, w$ with respect to the standard basis $e_1, e_2, e_3$. Thus one can "identify" alternating $(0,3)$–tensors $T$ on $\mathbb{R}^3$ with scalars $\tau$. This depends only on the "oriented volume element" of the basis $(e_1, e_2, e_3)$, meaning we can replace $(e_1, e_2, e_3)$ by $(Ae_1, Ae_2, Ae_3)$ as long as $\det A = 1$. There are no non–zero alternating $(0,s)$–tensors with $s > 3$ on $\mathbb{R}^3$, since at least two of the $s > 3$ indices $1, 2, 3$ on a component of such a tensor must be equal. Alternating $(0,2)$–tensors on $\mathbb{R}^3$ can be described with the help of the cross–product as follows.

**Lemma**. *In $\mathbb{R}^3$ there is a one-to-one correspondence between alternating $(0,2)$-tensors $F$ and vectors $v$ so that $F \leftrightarrow v$ if*

$$F(a,b) = v \cdot (a \times b) = \det[v,a,b]$$

*as alternating bilinear form of two vector variables $a, b$.*

**Proof**. We note that

$$\det[v,a,b] = F(a,b) \text{ means } \epsilon_{kij}v^k a^i b^j = F_{ij}a^i b^j. \qquad (*)$$

Given $(v^k)$, (*) can be solved for $F_{ij}$, namely $F_{ij} = \epsilon_{kij}v^k$. Conversely, given $(F_{ij})$, (*)can be solved for $v^k$:

$$v^k = \frac{1}{2}\epsilon^{kij}\det[v,e_i,e_j] = \frac{1}{2}\epsilon^{kij}F(e_i,e_j) = \frac{1}{2}\epsilon^{kij}F_{ij}.$$

Here $e_i = (\delta_i^j)$, the $i$-th standard basis vector. □

Geometrically one can think of an alternating $(0,2)$–tensor $F$ on $\mathbb{R}^3$ as a function which associates to the "plane element" of two vectors $a, b$ the scalar $F(a,b)$. The "plane element" of $a, b$ is to be thought of as represented by the cross–product $a \times b$, which determines the plane and oriented area of the parallelogram spanned by $a$ and $b$. We now discuss the physical interpretation of some types of tensors.

*Type* $(0,0)$. A tensor of type $(0,0)$ is a scalar $f(p)$ depending on $p$. E. g. the electric potential of
a stationary electric field.

*Type* $(1,0)$. A tensor of type $(1,0)$ is a (contravariant) vector. E. g. the velocity vector $\dot{p}(t)$ of a curve $p(t)$, or the velocity $\mathrm{v}(p)$ at $p$ of a fluid in motion, the electric current $\mathrm{J}(p)$ at $p$.

*Type* $(0,1)$. A tensor of type $(0,1)$ is a covariant vector. E. g. the differential $df = (\partial f/\partial x^i)dx^i$at $p$ of a scalar function. A force $F = (F_i)$ acting at a point $p$ should be considered a covector: for if $v = (v^i)$ is the velocity vector of an object passing through $p$, then the scalar $F(v) = F_i v^i$ is the work per unit time done by $F$ on the object, hence independent of the coordinate system. This agrees with the fact that a static electric field $E$ can be represented as the differential of a scalar function $\varphi$, its potential: $E = d\varphi$, i.e. $E_i = \partial\varphi/\partial x^i$.

*Type* $(2,0)$. In Cartesian coordinates in $\mathbb{R}^3$, the cross-product $v \times w$ of two vectors $v = (v^i)$ and $w = (w^j)$ has components

$$(v^2w^3 - v^3w^2, -v^1w^3 + v^3w^1, v^1w^2 - v^2w^1).$$

On any manifold, if $v = (v^i)$ and $w = (w^j)$ are vectors at a point $p$ then the quantity $T^{ij} = v^iw^j - v^jw^i$ is a tensor of type $(2,0)$ at $p$. Note that this tensor is *alternating* i.e. satisfies the symmetry condition $T^{ij} = -T^{ji}$. Thus the cross-product $v \times w$ of two vectors in $\mathbb{R}^3$can be thought of as an alternating $(2,0)$ tensor. (But this is not always appropriate: e.g. the velocity vector of a point $p$ performing a uniform rotation about the point o is $r \times w$ were $r$ is the direction vector from o to $p$ and $w$ is a vector along the axis of rotation of length equal to the angular speed. Such a rotation cannot be defined on a general manifold.)
On any manifold, the alternating $(2,0)$-tensor $T^{ij} = v^iw^j - v^jw^i$ can be thought of as specifying the two-dimensional "plane element" spanned by $v = (v^i)$ and $w = (w^j)$ at $p$.

*Type* $(0,2)$. A tensor T of type $(0,2)$ can be thought of as a *bilinear function* $T(v,w) = T_{ij}v^iw^j$ of two vector variables $v = (v^i)$, w= $(w^j)$ at the same point $p$. Such a function can be *symmetric*, i.e. $T(v,w) = T(w,v)$ or $T^{ij} = T^{ji}$, or *alternating*, i.e. $T(v,w) = -T(w,v)$ or $T^{ij} = -T^{ji}$, but need not be either. For example, on any manifold a Riemann metric $g(v,w) = g_{ij}v^iw^j$ is a symmetric $(0,2)$-tensor field (which in addition must be nondegenerate: $\det(g_{ij}) \neq 0$). A magnetic field $B$ on $\mathbb{R}^3$ should be considered an alternating $(0,2)$ tensor field in $\mathbb{R}^3$: for if J is the current in a conductor moving with velocity $V$, then the scalar $(B \times J) \cdot V$ is the work done by the magnetic field on the conductor, hence must be independent of coordinates. But this scalar

$$(B \times J) \cdot V = B \cdot (J \times V)\,\det[B, J, V] = \epsilon_{kij} B^k J^i V^j$$

is bilinear and alternating in the variables $J, V$ hence defines a tensor $(B_{ij})$ of type (0,2) given by $B_{ij} = \epsilon_{kij} B^k$ in Cartesian coordinates. $\epsilon_{kij} = \pm 1$ is the sign of the permutation $(kij)$ of (123). The scalar is $B(J, V) = B_{ij} J^i V^j$. It is clear that in any coordinate system $B(J, v) = -B(V, J)$, i.e. $B_{ij} = -B_{ji}$ since this equation holds in the Cartesian coordinates.

As noted above, in $\mathbb{R}^3$, any alternating $(0, 2)$ tensor $T$ can be written as $T(a, b) = v \cdot (a \times b)$ for a unique vector $v$. Thus $T(v, w)$ depends only on the vector $v \times w$ which characterizes the “2-plane element” spanned by $v$ and $w$. This is true on any manifold: an alternating $(0, 2)$-tensor $T$ can be thought of as a function $T(a, b)$ of two vector variables $a, b$ at the same point $p$ which depends only on the “2-plane element” spanned by $a$ and $b$.

Type $(1, 1)$. A tensor $T$ of type $(1, 1)$ can be thought of a linear transformation v$\rightarrow T(v)$ of the space of vectors at a point $p$: $T$ transforms a vector $v = (v^i)$ at $p$ into the vector $T(v) = (w^j)$ given by $w^j = T_i^j v^i$.

Type $(1, 3)$. The stress tensor $S$ of an elastic medium in $\mathbb{R}^3$ associates to the plane element spanned by two vectors $a, b$ at a point $p$ the force acting on this plane element. If this plane is displaced with velocity $v$, then the work done per unit time is a scalar which does not depend on the coordinate system. This scalar is of the form $S(a, b, v) = S_{ijk} a^i b^j v^k$ if $a = (a^i)$, $b = (b^j)$, $v = (v^k)$. If this formula is to remain true in any coordinate system, S must be a tensor of type $(0, 3)$. Furthermore, since this scalar depends only on the plane-element spanned by a,b, it must be alternating in these variables: $S(a, b, v) = -S(b, a, v)$, i.e. $S_{ijk} = -S_{jik}$. Use the correspondence (b) to write in Cartesian coordinates $S_{ijk} = \epsilon_{ij\mathrm{l}} P^l_k$.Then

$$S(a, b, v) = P_l^k \epsilon_{ijk} a^i a^j v^l \ = \mathrm{P}(a \times b) \cdot v.$$

This formula uses the correspondence (b) and remains true in any coordinate system related to the Cartesian coordinate system by a transformation with Jacobian determinant 1.

## Appendix 2: Algebraic Definition of Tensors

The tensor product $V \otimes W$ of two finite dimensional vector spaces (over any field) may be defined as follows. Take a basis $\{e_j \mid j = 1, \cdots, m\}$ for $V$ and a basis $\{f_k | k = 1, \cdots, n\}$ for $W$. Create a vector space, denoted $V \otimes W$, with a basis consisting of $nm$ symbols $e_j \otimes f_k$. Generally, for $x = \sum_j x_j e_j \in V$ and $y = \sum_k y_k f_k \in W$, let

$$x \otimes y = \sum_{ik} x_j y_k e_j \otimes f_k \in V \otimes W.$$

The space $V \otimes W$ is independent of the bases $\{e_j\}$ and $\{f_k\}$ in the following sense. Any other bases $\{\tilde{e}_j\}$ and $\{\tilde{f}_k\}$ lead to symbols $\tilde{e}_j \tilde{\otimes} \tilde{f}_k$ forming a basis of $V \tilde{\otimes} W$. We may then identify $V \tilde{\otimes} W$ with $V \otimes W$ by sending the basis $\tilde{e}_j \tilde{\otimes} \tilde{f}_k$ of $V \tilde{\otimes} W$ to the vectors $\tilde{e}_j \otimes \tilde{f}_k$ in $V \otimes W$, and vice versa.
The space $V \otimes W$ is called the tensor product of $V$ and $W$. The vector $x \otimes y \in V \otimes W$ is called the tensor product of $x \in V$ and $w \in W$. (One should keep in mind that an element of $V \otimes W$ looks like $\sum \zeta_{ij} e_j \otimes f_k$ and may not be expressible as a single tensor product $x \otimes y = \sum x_j y_k e_j \otimes f_k$.) The triple tensor products $(V_1 \otimes V_2) \otimes V_3$ and $V_1 \otimes (V_2 \otimes V_3)$ may be identified in an obvious way and simply denoted $V_1 \otimes V_2 \otimes V_3$. Similarly one can form the tensor product $V_1 \otimes V_2 \otimes \cdots \otimes V_N$ of any finite number of vector spaces.

This construction applies in particular if one takes for each $V_j$ the tangent space $T_p M$ or the cotangent space $T_p^* M$ at a given point $p$ of a manifold $M$. If one takes $\{\partial/\partial x^i\}$ and $\{dx^i\}$ as basis for these spaces and writes out the identification between the tensor spaces constructed by means of another pair of bases $\{\partial/\partial \tilde{x}^i\}$ and $\{d\tilde{x}^i\}$ in terms of components, one finds exactly the transformation rule for tensors used in definition 1.6.1.

**Another message from Herman Weyl.** For edification and moral support contemplate this message from Hermann Weyl's *Raum-Zeit-Materie* (my translation). *Entering into tensor calculus has certainly its conceptual difficulties, apart of the fear of indices, which has to be overcome. But formally this calculus is of extreme simplicity, much simpler, for example, than elementary vector calculus. Two operations: multiplication and contraction, i.e. juxtaposition of components of tensors with distinct indices and identification of two indices, one up one down, with implicit summation. It has often been attempted to introduce an invariant notation for tensor calculus$\cdots$. But then one needs such a mass of symbols and such an apparatus of rules of calculation (if one does not want to go back to components after all) that the net effect is very much negative. One has to protest strongly against these orgies of formalism, with which one now starts to bother even engineers.* – Elsewhere one can find similar sentiments expressed in similar words with reversed casting of hero and villain.

**EXERCISES** 1.6

1. Let $(T^{ij\cdots}_{kl\cdots})$ be the components of a tensor (field) of type ($r$,s) with respect to the coordinate system $(x^i)$. Show that the quantities $\partial T^{ij\cdots}_{kl\cdots}/\partial x^m$, depending on one more index m, do not transform like the components of a tensor unless ($r$,s) $= (0,0)$. [You may take ($r$,s) $= (1,1)$ to simplify the notation.]

2. Let $(\varphi_{\mathrm{k}})$ be the components of a tensor (field) of type $(0,1)$ (= 1-form) with respect to the coordinate system $(x^i)$. Show that the quantities $(\partial\varphi_i/\partial x^j)$ –

$(\partial\varphi_j/\partial x^i)$ depending on two indices $ij$ form the components of a tensor of type $(0,2)$.

3. Let $(X^i)$, $(Y^j)$ be the components of two tensors (fields) of type $(1,0)$ (=vector fields) with respect to the coordinate system $(x^i)$. Show that the quantities $X^j(\partial Y^i/\partial x^j) - Y^j(\partial X^i/\partial x^j)$ depending on one index $i$ form the components of a tensor of type $(1,0)$.

4. Let $f$ be a $\mathrm{C}^2$ function. (a) Do the quantities $\partial^2 f/\partial x^i\partial x^j$ depending on the indices $ij$ form the components of a tensor field? [Prove your answer.]
(b) Suppose $p$ is a point for which $df_p = 0$. Do the quantities $(\partial^2 f/\partial x^i\partial x^j)_p$ depending on the indices $ij$ form the components of a tensor at $p$? [Prove your answer.]

5. (a) Let $I$ be a quantity represented by $(\delta_i^j)$ relative to every coordinate system. Is $I$ a $(1,1)$-tensor? Prove your answer. [$\delta_i^j$ =Kronecker delta:= 1 if $i$ =j and = 0 otherwise.]
(b) Let J be a quantity represented by $(\delta_{ij})$ relative to every coordinate system. Is J a $(0,2)$-tensor? Prove your answer. [$\delta_{ij}$ =Kronecker delta := 1 if $i$ =j and = 0 otherwise.]

6. (a) Let $(T_j^i)$ be a quantity depending on a coordinate system around the point $p$ with the property that for every $(1,1)$-tensor $(S_i^j)$ at $p$ the scalar $T_j^i S_i^j$ is independent of the coordinate system. Prove that $(T_j^i)$ represents a $(1,1)$-tensor at $p$.
(b) Let $(T_j^i)$ be a quantity depending on a coordinate system around the point $p$ with the property that for every $(0,1)$-tensor $(S_i)$ at $p$ the quantity $T_j^i S_i$, depending on the index j, represents a $(0,1)$-tensor at $p$. Prove that $(T_j^i)$ represents a $(1,1)$-tensor at $p$.
(c) State a general rule for quantities $(T_{kl\cdots}^{ij\cdots})$ which contains both of the rules (a) and (b) as special cases. [You need not prove this general rule: the proof is the same, just some more indices.]

7. Let $p = p(t)$ be a $\mathrm{C}^2$ curve given by $x^i = x^i(t)$ in a coordinate system $(x^i)$. Is the "acceleration" $(\ddot{x}^i(t)) = (d^2x^i(t)/dt^2)$ a vector at $p(t)$? Prove your answer.

8. Let $(S_i)$ and $(T^i)$ be the components in a coordinate system $(x^i)$ of smooth tensor fields of type $(0,1)$ and $(1,0)$ respectively. Which of the following quantities are tensor fields? Prove your answer and indicate the type of those which are tensor fields.
(a) $\partial(S_iT^i)/\partial x^k$ (b) $\partial(S_iT^j)/\partial x^k$
(c)$(\partial S_i/\partial x^j) - (\partial S_j/\partial x^i)$ (d)$(\partial T^i/\partial x^j) - (\partial T^j/\partial x^i)$.

9. Prove that the operations on tensors defined in the text do produce tensors in the following cases.
a) Addition and scalar multiplication. If $T_{ij}$ and $S_{ij}$ are tensors, so are $S_{ij}+T_{ij}$ and $cS_{ij}$ ($c$ any scalar).

b) Symmetry operations. If $T_{ij}$ is a tensor, then the equation $S_{ij} = T_{ji}$ defines another tensor $S$. Prove that this tensor may also defined by the formula $S(v,w) = T(\mathrm{w}, v)$.
c) Tensor product. If $T_i$ and $S^j$ are tensors, so is $T_i S^j$.

10. a) In the equation $A_i^j dx^i \otimes (\partial/\partial x^j) = \tilde{A}_i^j d\tilde{x}^i \otimes (\partial/\partial \tilde{x}^j)$ transform the left side using the transformation rules for $dx^i$ and $\partial/\partial x^j$ to find the transformation rule for $A_i^j$ (thus verifying that the transformation rule is "built into the notation" as asserted in the text).
b) Find the tensor field $y dx \otimes (\partial/\partial y) - x dy \otimes (\partial/\partial x)$ on $\mathbb{R}^2$ in polar coordinates $(r, \phi)$.

11. Let $(x^1, x^2) = (r, \theta)$ be polar coordinates on $\mathbb{R}^2$. Let $T$ be the tensor with components $T_{11} = \tan\theta,\ T_{12} = 0,\ T_{21} = 1 + r,\ T_{22} = e^r$ in these coordinates. Find the components $T^{ij}$ when both indices are raised using the Euclidean metric on $\mathbb{R}^2$.

12. Let $V \otimes W$ be the tensor product of two finite–dimensional vector spaces as defined in Appendix 2.
a) Let $\{e_i\}$, $\{f_j\}$ be bases for $V$ and $W$, respectively. Let $\{\tilde{e}_r\}$ and $\{\tilde{f}_s\}$ be two other bases and write $e_i = a_i^r \tilde{e}_r,\ f_j = b_j^s \tilde{f}_s$. Suppose $T \in V \otimes W$ has components $T^{ij}$ with respect to the basis $\{e_i \otimes f_j\}$ and $\tilde{T}^{rs}$ with respect to $\{\tilde{e}_r \otimes \tilde{f}_s\}$. Show that $\tilde{T}^{rs} = T^{ij} a_i^r b_j^s$. Generalize this transformation rule to the tensor product $V_1 \otimes V_2 \otimes \cdots$ of any finite number of vector spaces. (You need not prove the generalization in detail. Just indicate the modification in the argument.)
b)This construction applies in particular if one takes for each $V_j$ the tangent space $T_p M$ or the cotangent space $T_p^* M$ at a given point $p$ of a manifold $M$. If one takes $\{\partial/\partial x^i\}$ and $\{dx^i\}$ as basis for these spaces and writes out the identification between the tensor spaces constructed by means of another pair of bases $\{\partial/\partial \tilde{x}^i\}$ and $\{d\tilde{x}^i\}$ in terms of components, one finds exactly the transformation rule for tensors used in definition 1.6.1. (Verify this statement.)

# Chapter 2

# Connections and curvature

## 2.1 Connections

For a scalar-valued, differentiable function $f$ on a manifold $M$ defined in neighborhood of a point $p_o \in M$ we can define the derivative $D_v f = df_p(v)$ of $f$ along a tangent vector $v \in T_pM$ by the formula

$$D_v f = \frac{d}{dt} f(p(t))|_{t=0} = \lim_{t\to 0} \frac{1}{t}\Big( f(p(t)) - f(p_o)\Big)$$

where $p(t)$ is any differentiable curve on $M$ with $p(0) = p_o$ and $\dot{p}(0) = v$. One would naturally like to define a directional derivative of an arbitrary differentiable tensor field $F$ on $M$, but this cannot be done in the same way, because one cannot subtract tensors at different points on $M$. In fact, on an arbitrary manifold this cannot be done at all (in a reasonable way) unless one adds another piece of structure to the manifold, called a "covariant derivative". This we now do, starting with a covariant derivative of vector fields, which will later be used to define a covariant derivative of arbitrary tensor fields.

2.1.1 **Definition**. A *covariant derivative* on $M$ is an operation which produces for every "input" consisting of (1) a tangent vector $v \in T_{p_o}M$ at some point $p_o$ and (2) a smooth vector field $X$ defined in a neighbourhood of $p$ a vector in $T_pM$, denoted $\nabla_v X$. This operation is subject to the following axioms:

C$D$1. $\nabla_{(u+v)}X = (\nabla_u X) + (\nabla_v X)$
CD2. $\nabla_{av}X = a\nabla_v X$
CD3. $\nabla_v(X + Y) = (\nabla_v X) + (\nabla_v Y)$
CD4. $\nabla_v(fX) = (D_v f)X + f(p)(\nabla_v X)$

for all u,$v \in T_pM$, all smooth vector fields $X$,$Y$ defined around $p$, all smooth functions $f$ defined in a neighbourhood of $p$, and all scalars $a \in \mathbb{R}$.

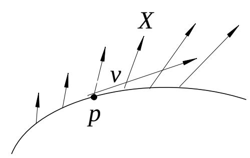



Fig. 1. Input

If $X$,$Y$ are two vector fields, we define another vector field $\nabla_X Y$ by $(\nabla_X Y)_p = \nabla_{X_p} Y$ for all points $p$ where both $X$ and $Y$ are defined. As final axiom we require:

CD5. If $X$ and $Y$ are smooth, so is $\nabla_X Y$

2.1.2 **Example**. Let $V$ be an $n$-dimensional vector space considered as a manifold. Its tangent space $T_pV$ at any point $p$ can be identified with $V$ itself: a basis $(e_i)$ of $V$ gives linear coordinates $(x^i)$ defined by $p = x^i e_i$ and tangent vectors $X = X^i \partial/\partial x^i$ in $T_pV$ may be identified with vectors $X = X^i e_i$ in $V$. A vector field $X$ on $V$ can therefor be indentied with a $V$–valued function on $V$ and one can take for $\nabla_v X$ the usual directional derivative $D_v X$, i.e.

$$D_v X(p_o) = \lim_{t \to 0} \frac{1}{t}\Big(X(p(t)) - X(p_o)\Big) \tag{1}$$

where $p(t)$ is any differentiable curve on $M$ with $p(0) = p$ and $\dot{p}(0) = v$. If one writes $X = X^i e_i$ then this is the componentwise directional derivative

$$D_v X = (D_v X^1)e_1 + \cdots + (D_v X^n)e_n. \tag{1}$$

Now let $S$ be a submanifold of the Euclidean space $E$, i.e a vector space equipped with a with its positive definite inner product. For any $p \in S$ the tangent space $T_pS$ is a subspace of $T_pE$ and vector field $X$ on $S$ is an $E$-valued function on $S$ whose value $X_p$ at $p \in S$ lies in the subspace $T_pS$ of $E$. The componwise directional derivative $D_v X$ in $E$ of a vector field on $S$ along a tangent vector $v \in T_pS$ is in general no longer tangential to $S$, hence cannot be used to define a covariant derivative on $S$. However, any vector in $E$ can be uniquely written as a sum of vector in $T_pS$ (its *tangential component*) and a vector orthogonal to $T_pS$ and the definition

$$\nabla^S_v X := \textit{tangential component of } D_v X. \tag{4}$$

does define a covariant derivative on $S$. By definiton, this definition means that

$$D_v X = \nabla^S_v X + \textit{ a vector orthogonal to } T_pS. \tag{5}$$

$D_v X$ is the covariant derivative on the ambient Euclidean space $E$, i.e. the componentwise derivative defined by equation (2). It is evidently well defined, even though $X$ is only defined on $S$: one chooses for $p(t)$ a curve which lies in $S$. In fact, to define $\nabla_v X$ at $p_o$ it evidently suffices to know $X$ along a curve $p(t)$

with $p(0) = p_o$ and $\dot{p}(0) = p$. (We shall see later that is true for any covariant derivative on a manifold.) $\nabla^S$ is a covariant derivative on the submanifold $S$ of $\mathbb{R}^n$, called the covariant derivative *induced* by the inner product on $\mathbb{R}^n$. (The inner product is needed for the orthogonal splitting.)

2.1.3 **Example**. Let $S$ be the sphere $S = \{p \in \mathbb{R}^3 \mid \|p\| = 1\}$ in $E = \mathbb{R}^3$. Let $x, y, z$ be the Cartesian coordinates on $\mathbb{R}^3$. Define coordinates $\phi, \theta$ on $S$ by $(\theta, \phi)$ *defined by*

$$x = \cos\theta \sin\phi, \quad y = \sin\theta \sin\phi, \quad z = \cos\phi$$

As vectors on $\mathbb{R}^3$ tangential to $S$ the coordinate vector fields are

$$\begin{aligned} \frac{\partial}{\partial\phi} &= \cos\theta\cos\phi\frac{\partial}{\partial x} + \sin\theta\cos\phi\frac{\partial}{\partial y} - \sin\phi\frac{\partial}{\partial z} \\ \frac{\partial}{\partial\theta} &= -\sin\theta\sin\phi\frac{\partial}{\partial x} + \cos\theta\sin\phi\frac{\partial}{\partial y}. \end{aligned}$$

$D_v X$ is the componentwise derivative in the coordinate system $x, y, z$,

$$\begin{aligned} &\text{if } X = X^x\frac{\partial}{\partial x} + \ X^y\frac{\partial}{\partial y} + \ X^z\frac{\partial}{\partial z} \\ &\text{then } D_v X = (D_v X^x)\frac{\partial}{\partial x} + \ (D_v X^y)\frac{\partial}{\partial y} + \ (D_v X^z)\frac{\partial}{\partial z} \end{aligned}$$

where $D_v f = df(v)$ is the directional derivative of a function $f$ as above. For example, if we take $X = \frac{\partial}{\partial\phi}$ as well as $v = \frac{\partial}{\partial\phi}$ and write $\frac{D}{\partial\phi}$ for $D_v$ when $v = \frac{\partial}{\partial\phi}$ then we find

$$\begin{aligned} \frac{D}{\partial\phi}\frac{\partial}{\partial\phi} &= -\sin\phi\ \cos\theta\ \frac{\partial}{\partial x} - \ \sin\phi\ \sin\theta\ \frac{\partial}{\partial y} - \ \cos\phi\ \frac{\partial}{\partial z} \\ &= -x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} - z\frac{\partial}{\partial z}\ . \end{aligned}$$

This vector is obviously orthogonal to $T_pS = \{a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y} + c\frac{\partial}{\partial z} \mid ax + by + cz = 0\}$. Therefore its tangential component is zero: $\frac{\nabla^S}{\partial\phi}\frac{\partial}{\partial\phi} \equiv 0$. Note that we can calculate $D_Y X$ even if $X$ is only defined on $S^2$, provided $Y$ is tangential to $S^2$, e.g. in the calculation of $\frac{D}{\partial\phi}\frac{\partial}{\partial\phi}$ above.

From now on we assume that $M$ has a covariant derivative $\nabla$. Let $x^1, \cdots, x^n$ be a coordinate system on $M$. To simplify notation, write $\partial_j = \partial/\partial x^j$, and $\nabla_j X = \nabla X/\partial x^j$ for the covariant derivative of $X$ along $\partial_j$. On the coordinate domain, vector fields $X$,$Y$ can be written as $X = \sum_j X^j\partial_j$ and $Y = \sum_j Y^j\partial_j$. Calculate:

$$\begin{aligned} \nabla_Y X &= \sum_{ij} \nabla_{Y^j\partial_j}(X^i\partial_i) \quad \text{[by CD1 and CD3]} \\ &= \sum_{ij} Y^j(\partial_j X^i)\partial_i + \ Y^j X^i(\nabla_j\partial_i) \quad \text{[by CD2 and CD4].} \end{aligned}$$

Since $\nabla_j \partial_i$ is itself a smooth vector field [CD5], it can be written as

$$\nabla_j \partial_i = \sum_k \Gamma^k_{ij}\ \partial_k \tag{6}$$

for certain smooth functions $\Gamma^k_{ij}$ on the coordinate domain. Omitting the summation signs we get

$$\nabla_Y X = \ (Y^j \partial_j X^k + \ \Gamma^k_{ij} Y^j X^i)\partial_k\ . \tag{7}$$

One sees from this equation that one can calculate any covariant derivative as soon as one knows the $\Gamma^k_{ij}$. In spite of their appearance, the $\Gamma^k_{ij}$ are *not* the components of a tensor field on $M$. The reason can be traced to the fact that to calculate $\nabla_Y X$ at a point $p \in M$ it is not sufficient to know the vector $X_p$ at $p$.

As just remarked, the value of $\nabla_Y X$ at $p$ depends not only on the value of $X$ at $p$; it depends also on the partial derivatives at $p$ of the components of $X$ relative to a coordinate system. On the other hand, it is not necessary to know *all* of these partial derivatives to calculate $\nabla_v X$ for a given $v \in T_pM$. In fact it suffices to know the directional derivatives of these components alone the vector $v$ only. To see this, let $p(t)$ be a differentiable curve on $M$. Using equation (7) with $Y$ replaced by the tangent vector $\dot{p}(t)$ of $p(t)$ one gets

$$\begin{aligned} \nabla_{\dot{p}} X &= \Big(\partial_j X^k \frac{dx^j}{dt} + \ \Gamma^k_{ij}\ \frac{dx^j}{dt} X^i\Big)\partial_k \\ &= \Big(\frac{dX^k}{dt} + \ \ \Gamma^k_{ij}\ \ \frac{dx^j}{dt} X^i\Big)\partial_k . \end{aligned}$$

For a given value of $t$, this expression depends only on the components $X^i$ of $X$ at $p = p(t)$ and their derivatives $\frac{dX^k}{dt}$ along the tangent vector $v = \dot{p}(t)$ of the curve $p(t)$. If one omits even the curve $p = p(t)$ from the notation the last equation reads

$$\nabla X = \ (dX^k + \ \omega^k_i)\partial_k . \tag{8}$$

where $\omega^k_i = \Gamma^k_{ij}\ dx^j$. This equation evidently gives the covariant derivative $\nabla_v X$ as a linear function of a tangent vector $v$ with value in the tangent space. Its components $dX^k + \ \omega^k_i$ are linear functions of $v$, i.e. differential forms. The differential forms $\omega^k_i = \Gamma^k_{ij}\ dx^j$ are called the *connection forms* with respect to the coordinate frame $\{\partial_k\}$ and the $\Gamma^k_{ij}$ *connection coefficients*. The equation $\nabla_j \partial_i = \Gamma^k_{ij}\ \partial_k$ for the $\Gamma^k_{ij}$ becomes the equation $\nabla \partial_i = \omega^k_i \partial_k$ for the $\omega^k_i$.

*2.1.4* **Example.** Let $S$ be an $m$-dimensional submanifold of the Euclidean space $E$ with the induced covariant derivative $\nabla = \nabla^S$. Let $X$ be a vector field on $S$ and $p = p(t)$ be a curve. Then $X$ may be also considered as a vector field on $E$ defined along $S$, hence as a function on $S$ with values in $E = T_pE$. By definition, the covariant derivative $DX/dt$ in $E$ is the derivative $dX/dt$ of this vector-valued function along $p = p(t)$, the covariant derivative $\nabla X/dt$ on $S$ its tangential component. Thus

$$\frac{DX}{dt} = \frac{\nabla X}{dt} + \text{a vector orthogonal to } S$$

n terms of a coordinate system $(t^1, \cdots, t^m)$ on $S$, write $X = X^i \frac{\partial}{\partial t^i}$ and then

$$\frac{DX}{dt} = (dX^k + \omega^k_j)\frac{\partial}{\partial t^k} + \text{a vector orthogonal to } S$$

In particular take $X$ to be the $i$th coordinate vector field $\partial/\partial t^i$ on $S$ and $p = p(t)$ the $j$th coordinate line through a point with tangent vector $\dot{p} = \partial/\partial t^j$. Then $X = \partial/\partial t^i$ is the $i$th partial $\partial p/\partial t^i$ of the $E$–valued function $p = p(t^1, \cdots, t^m)$ and its covariant derivative $DX/dt$ in $E$ is

$$\begin{aligned}\frac{\partial^2 p}{\partial t^j \partial t^i} &= \frac{\nabla}{\partial t^j}\frac{\partial}{\partial t^i} + \text{a vector orthogonal to } S\\ &= \sum_k \Gamma^k_{ij} \frac{\partial}{\partial t^k} + \text{a vector orthogonal to } S\end{aligned}$$

Take the scalar product of both sides with $\frac{\partial}{\partial t^l}$ in $E$ to find that

$$\frac{\partial^2 p}{\partial t^j \partial t^i} \cdot \frac{\partial}{\partial t^l} = \Gamma^k_{ij} \frac{\partial}{\partial t^k} \cdot \frac{\partial}{\partial t^l}$$

since $\frac{\partial}{\partial t^l}$ is tangential to $S$. This may be written as

$$\frac{\partial^2 p}{\partial t^j \partial t^i} \cdot \frac{\partial}{\partial t^l} = \Gamma^k_{ij}\, g_{kl}$$

where $g_{kl}$ is the coefficient of the induced Riemann metric $ds^2 = g_{ij}dt^i dt^j$ on $S$. Solve for $\Gamma^k_{ij}$ to find that

$$\Gamma^k_{ij} = g^{lk} \frac{\partial^2 p}{\partial t^i \partial t^j} \cdot \frac{\partial p}{\partial t^l}.$$

where $g^{kl} g_{lj} = \delta^k_j$.

As specific case, take $S$ to be the sphere $x^2 + y^2 + z^2 = 1$ in $E = \mathbb{R}^3$ with coordinates $p = p(\theta, \phi)$ defined by $x = \cos\theta \sin\phi$ , $y = \sin\theta\sin\phi$, $z = \cos\phi$. As vector fields in $E$ along $S$ the coordinate vector fields are

$$\begin{aligned}\frac{\partial p}{\partial \theta} &= -\sin\theta\sin\phi\frac{\partial}{\partial x} + \cos\theta\sin\phi\frac{\partial}{\partial y}.\\ \frac{\partial p}{\partial \phi} &= \cos\theta\cos\phi\frac{\partial}{\partial x} + \sin\theta\cos\phi\frac{\partial}{\partial y} - \sin\phi\frac{\partial}{\partial z}\end{aligned}$$

The componentwise derivatives in $E$ are

$$\begin{aligned}
\frac{\partial^2 p}{\partial \theta^2} &= -\cos\theta\sin\phi\frac{\partial}{\partial x} - \sin\theta\sin\phi\frac{\partial}{\partial y} \\
\frac{\partial^2 p}{\partial \phi^2} &= \cos\theta\sin\phi\frac{\partial}{\partial x} - \sin\theta\sin\phi\frac{\partial}{\partial y} - \cos\phi\frac{\partial}{\partial z} \\
\frac{\partial^2 p}{\partial \theta\partial\phi} &= -\sin\theta\cos\phi\frac{\partial}{\partial x} + \cos\theta\cos\phi\frac{\partial}{\partial y}
\end{aligned}$$

The only nonzero scalar products $\frac{\partial^2 p}{\partial t^j \partial t^i} \cdot \frac{\partial}{\partial t^l}$ are

$$\frac{\partial^2 p}{\partial \theta^2}\cdot\frac{\partial p}{\partial \phi} = -\sin\phi\cos\phi, \qquad \frac{\partial^2 p}{\partial\theta\partial\phi}\cdot\frac{\partial p}{\partial\theta} = \sin\phi\cos\phi$$

and the Riemann metric is $ds^2 = \sin^2\theta\, d\theta^2 + d\phi^2$. Since the matrix $g_{ij}$ is diagonal the equations for the $\Gamma$s become $\Gamma^k_{ij} = \frac{1}{g_{kk}}\frac{\partial^2 p}{\partial t^i \partial t^j} \cdot \frac{\partial p}{\partial t^k}$ and the only nonzero $\Gamma$s are

$$\Gamma^{\phi}_{\theta\theta} = -\frac{\sin\phi\cos\phi}{\sin^2\theta}, \quad \Gamma^{\theta}_{\phi\theta} = \Gamma^{\theta}_{\theta\phi} = \frac{\sin\phi\cos\phi}{\sin^2\theta}.$$

2.1.5 **Definition**. Let $p(t)$ be a curve in $M$. A vector field $X(t)$ *along* $p(t)$ is a rule which associates to each value of $t$ for which $p(t)$ is defined a vector $X(t) \in T_{p(t)}M$ at $p(t)$.

If the curve $p(t)$ intersects itself, i.e. if $p(t_1) = p(t_2)$ for some $t_1 \neq t_2$ a vector field $X(t)$ along $p(t)$ will generally attach two different vectors $X(t_1)$ and $X(t_2)$ to the point $p(t_1) = p(t_2)$.

Relative to a coordinate system we can write

$$X(t) = \sum_i X^i(t)\Big(\frac{\partial}{\partial x^i}\Big)_{p(t)} \tag{9}$$

at all points $p(t)$ in the coordinate domain. When the curve is smooth the we say $X$ is smooth if the $X^i$ are smooth functions of $t$ for all coordinate systems. If $p(t)$ is a smooth curve and $X(t)$ a smooth vector field along $p(t)$, we define

$$\frac{\nabla X}{dt} = \nabla_{\dot p} X.$$

The right side makes sense in view of (8) and defines another vector field along $p(t)$.
In coordinates,

$$\frac{\nabla X}{dt} = \sum_k \Big(\frac{dX^k}{dt} + \sum_{ij} \Gamma^k_{ij}\,\frac{dx^j}{dt}\, X^i\Big)\partial_k\ . \tag{10}$$

2.1.6 **Lemma (Chain Rule)**. *Let $X = X(t)$ be a vector field along a curve $p(t)$ and let $X = X(\tilde{t})$ be obtained from $X(t)$ by a change of parameter $t = f(\tilde{t})$. Then*

$$\frac{\nabla X}{d\tilde{t}} = \frac{dt}{d\tilde{t}}\frac{\nabla X}{dt}.$$

**Proof**. A change of parameter $t = f(\tilde{t})$ in a curve $p(t)$ is understood to be a diffeomorphism between the intervals of definition. Now calculate:

$$\begin{aligned}\frac{\nabla X}{d\tilde{t}} &= \sum_k \Big(\frac{dX^k}{d\tilde{t}} + \sum_{ij} \Gamma^k_{ij} \frac{dx^j}{d\tilde{t}} X^i\Big)\frac{\partial}{\partial x^k} \\ &= \sum_k\Big(\frac{dX^k}{dt}\frac{dt}{d\tilde{t}} + \sum_{ij} \Gamma^k_{ij} \frac{dx^j}{dt}\frac{dt}{d\tilde{t}}X^i\Big)\frac{\partial}{\partial x^k} \\ &= \frac{dt}{d\tilde{t}}\frac{\nabla X}{dt}.\end{aligned}$$

2.1.7 **Theorem**. *Let $p(t)$ be a* smooth *curve on $M$, $p_o = p(t_o)$ a point on $p(t)$. Given any vector $v \in T_{p_o}M$ at $p_o$ there is a unique* smooth *vector field $X$ along $p(t)$ so that*

$$\frac{\nabla X}{dt} \equiv 0, X(t_o) = v.$$

**Proof**. In coordinates, the conditions on $X$ are

$$\frac{dX^k}{dt} + \sum_{ij} \Gamma^k_{ij} \frac{dx^j}{dt} X^i = 0, \quad X^k(t_o) = v^k.$$

From the theory of differential equations one knows that these equations do indeed have a unique solution $X^1, \cdots, X^n$ where the $X^k$ are smooth functions of $t$. This proves the theorem when the curve $p(t)$ lies in a single coordinate domain. Otherwise one has to cover the curve with several overlapping coordinate domains and apply this argument successively within each coordinate domain, taking as initial vector $X(t_j)$ for the $j$-th coordinate domain the vector at the point $p(t_j)$ obtained from the previous coordinate domain. □

**Remark**. Actually the existence theorem for systems of differential equations only guarantees the existence of a solution $X^k(t)$ for $t$ in some interval around 0. This should be kept in mind, but we shall not bother to state it explicitly. This caveat applies whenever we deal with solutions of differential equations.

2.1.8 **Supplement to the theorem**. *The vector field $X$ along a curve $C : p = p(t)$ satisfying*

$$\tfrac{\nabla X}{dt} \equiv 0, X(t_o) = v.$$

*is independent of the parametrization.*

This should be understood in the following sense. Consider the curve $p = p(t)$ and the vector field $X = X(t)$ as functions of $\tilde{t}$ by substituting $t = f(\tilde{t})$. Then $\nabla X/d\tilde{t} = f'(\tilde{t})\nabla X/dt = 0$, so $X = X(\tilde{t})$ is the unique vector field along $p = p(\tilde{t})$ satisfying $X(\tilde{t}_o) = X(t_o) = v$.

2.1.9 **Definitions**. Let $C : p = p(t)$ be a smooth curve on $M$.

(a) A vector smooth field $X(t)$ along c satisfying $\nabla X/dt \equiv 0$ is said to be *parallel* along $p(t)$.

(b) Let $p_o = p(t_o)$ and $p_1 = p(t_1)$ two points on $C$. For each vector $v_o \in T_{p_o}M$ define a vector $v_1 \in T_{p_1}M$ by $w = X(t_1)$ where $X$ is the unique vector field along $p(t)$ satisfying $\nabla X/dt \equiv 0, X(t_o) = v_o$. The vector $v_1 \in T_{p_1}M$ is called *parallel transport of* $v_o \in T_{p_o}M$ from $p_o$ to $p_1$ along $p(t)$. It is denoted

$$v_1 = T(t_o \to t_1)v_o. \tag{12}$$

(c) The map $T(t_o \to t_1) : T_{p_o}M \to T_{p_1}M$ is called *parallel transport* from $p_o$ to $p_1$ along $p(t)$.

It is important to remember that $T(t_o \to t_1)$ depends in general on the curve $p(t)$ from $p_o$ to $p_1$, not just on the endpoints $p_o$ and $p_1$. This is important and to bring this out we may write $C\colon p = p(t)$ for the curve and $T_C(t_o \to t_1)$ for the parallel transport along $C$.

2.1.10 **Theorem**. *Let $p(t)$ be a* smooth *curve on $M$, $p_o = p(t_o)$ and $p_1 = p(t_1)$ two points on $p(t)$. The parallel transport along $p(t)$ from $p_o$ to $p_1$,*

$$T(t_o \to t_1) \ : \ T_{p_o}M \to T_{p_1}M$$

*is an invertible linear transformation.*

**Proof**. The linearity of $T(t_o \to t_1)$ comes from the linearity of the covariant derivative along $p(t)$:

$$\frac{\nabla}{dt}(X + \ Y) = (\frac{\nabla}{dt}X) + \ (\frac{\nabla}{dt}Y) \ , \qquad (\frac{\nabla}{dt}aX) = a(\frac{\nabla X}{dt}) \ (a \in \mathbb{R}).$$

The map $T(t_o \to t_1)$ is invertible and its inverse is $T(t_1 \to t_o)$ since the equations

$$\frac{\nabla X}{dt} = 0, \, X(t_o) = v_o, \, X(t_1) = v_1$$

say both $v_1 = T(t_1 \to t_o)v_o$ and $v_o = T(t_o \to t_1)v_1$. □

In coordinates $(x^i)$ the parallel transport equation $\nabla X/dt = 0$ along a curve $p = p(t)$ says that $\nabla X = (dX^k + \omega_i^k)\partial_k = 0$ if one sets $dx^i = \dot{x}^i(t)dt$ in $\omega_i^k = \Gamma_{ij}^k \, dx^j$. In this sense the parallel transport is described by the equations $dX^k = -\omega_i^k$ and one might say that $-\omega_i^k = -\Gamma_{ij}^k \, dx^j$ is the matrix of the "infinitesimal" parallel transport along the vector with components $dx^j$. This matrix should *not* be thought of as a linear transformation of the tangent space at $p(x^j)$ into itself. (That transformation would depend on coordinate system.). Classical differential geometers[1] thought of it as a linear transformation from the tangent space at $p(x^j)$ to the tangent space at the "infinitesimally close" point $p(x^j + dx^j)$, and this is the way this concept was introduced by Levi-Civita in 1917.

A covariant derivative on a manifold is also called an *affine connection*, because it leads to a notion of parallel transport along curves which "connects" tangent vectors at different points, a bit like in the "affine space" $\mathbb{R}^n$, where one may transport tangent vectors from one point to another parallel translation (see the example below). The fundamental difference is that parallel transport for a general covariant derivative depends on a curve connecting the points, while in an affine space it depends on the points only. While the terms "covariant derivatives" and "connection" are logically equivalent, they carry different connotations, which can be gleaned from the words themselves.

2.1.11 **Example** 2.1.2 **(continued)**. Let $M = \mathbb{R}^n$ with the covariant derivative defined earlier. We identify tangent vectors to $M$ at any point with vectors in $\mathbb{R}^n$. Then parallel transport along any curve keeps the components of vectors with respect to the standard coordinates constant, i.e. $T(t_o \to t_1) : T_{p_o}M \to T_{p_1}M$ is the identity mapping if we identify both $T_{p_o}M$ and $T_{p_1}M$ with $\mathbb{R}^n$ as just described.

2.1.12 **Example** 2.1.3 **(continued)**. Let $E = \mathbb{R}^3$ with its positive definite inner product, $S = \{p \in E \mid \|p\| = 1\}$ the unit sphere about the origin. $S$ is a submanifold of $E$ and for any $p \in S$, $T_pS$ is identified with the subspace $\{v \in E \mid v \cdot p = 0\}$ of $E$ orthogonal to $p$. $S$ has a covariant derivative $\nabla$ induced by the componentwise covariant derivative $D$ in $E$: for $v \in T_pS$ and $X$ a vector field on $S$, i.e.

$$\nabla_p X = \text{tangential component of } DX.$$

The tangential component of a vector on can be obtained by subtracting the normal component, i.e.

$$\nabla_p X = \nabla_v^E X - (N \cdot D_v X)N$$

where $N$ is the unit normal vector $N$ at point $p$ on $S$ (in either direction). On the sphere $S$ we simply have $N = p$ in $T_pE = E$. So if $X(t)$ is a vector field along a curve $p(t)$ on $S$, then its covariant derivative on $S$ is

$$\frac{\nabla X}{dt} = \frac{DX}{dt} - (p(t) \cdot \frac{dX}{dt})p(t)$$

where $DX/dt$ is the componentwise derivative on $E = \mathbb{R}^3$. We shall explicitly determine the parallel transport along great circles on $S$. Let $p_o \in S$ be a point of $S$, $e \in T_pS$ a unit vector. Thus $e \cdot p_o = 0, e \cdot e = 1$. The great circle through $p_o$ in the direction $e$ is the curve $p(t)$ on $S$ given by $p(t) = (\cos t)p_o +$ $(\sin t)e$.

[1] Cartan "Leçons" (1925-1926) §85, Weyl "Raum-Zeit-Materie" (1922 edition) §12 and §15.

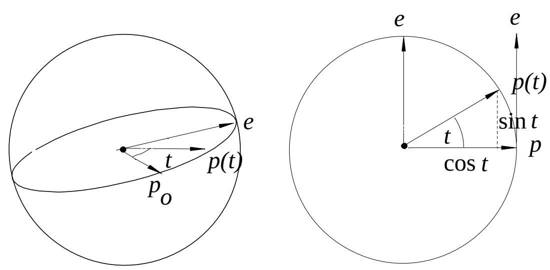



Fig. 2. A great circle on $S$

A vector field $X(t)$ along $p(t)$ is parallel if $\nabla X/dt = 0$. Take in particular for $X$ the vector field $E(t) = \dot{p}(t)$ along $p(t)$, i.e $E(t) = -\sin t\, p_o + \cos t\, e$ . Then $DE/dt = \ddot{p}(t) = -p(t)$, from which it follows that $\nabla E/dt \equiv 0$ by the formula above. Let $f \in T_pS$ be one of the two unit vectors orthogonal to $e$, i.e. $f \cdot p_o = 0, f \cdot e = 0, f \cdot f = 1$. Then $f \cdot p(t) = 0$ for any $t$, so $f \in T_{p(t)}S$ for any $t$. Define a second vector field $F$ along $p(t)$ by setting $F(t) = f$ for all t. Then again $\nabla F/dt \equiv 0$. Thus the two vector fields $E$, $F$, are parallel along the great circle $p(t)$ and form an orthonormal basis of the tangent space to $S$ at any point of $p(t)$.

Any tangent vector $v \in T_{p_o}S$ can be uniquely written as $v = ae + bf$ with $a, b \in \mathbb{R}$. Let $X(t)$ be the vector field along $p(t)$ defined by $X(t) = aE(t)+bF(t)$ for all t. Then $\nabla X/dt \equiv 0$. Thus the parallel transport along $p(t)$ is given by

$$T(0 \to t) : T_{p_o}S \to T_{p(t)}S,\ ae + bf \to aE(t) + bF(t).$$

In other words, *parallel transport along $p(t)$ leaves the components of vectors with respect to the bases $E$, $F$ unchanged.* This may also be described geometrically like this: $w = T(0 \to t)v$ *has the same length as $v$ and makes the same angle with the tangent vector $\dot{p}(t)$ at $p(t)$ as $v$ makes with the tangent vector $\dot{p}(0)$ at $p(0)$.* (This follows from the fact that $E, F$ are orthonormal and $E(t) = \dot{p}(t)$.)

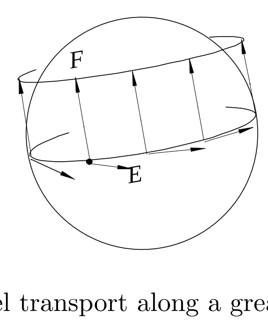



Fig. 3. Parallel transport along a great circle on $S$

2.1.13 **Theorem**. *Let $p = p(t)$ be a* smooth *curve in $M$, $p_o = p(t_o)$ a point on $p(t)$, $v_1, \cdots, v_n$ a basis for the tangent space $T_{p_o}M$ at $p_o$. There are unique parallel vector fields $E_1(t), \cdots, E_n(t)$ along $p(t)$ so that*

$$E_1(t_o) = v_1, \cdots, E_n(t_o) = v_n.$$

*For any value of $t$, the vectors $E_1(t), \cdots, E_n(t)$ form a basis for $T_{p(t)}M$.*

**Proof**. This follows from theorems 2.1.7, 2.1.10, since an invertible linear transformation maps a basis to a basis. □

2.1.14 **Definition**. Let $p = p(t)$ be a smooth curve in $M$. A *parallel frame* along $p(t)$ is an $n$–tuple of *parallel* vector fields $E_1(t), \cdots, E_n(t)$ along $p(t)$ which form a basis for $T_{p(t)}M$ for any value of $t$.

2.1.15 **Theorem**. *Let $p = p(t)$ be a* smooth *curve in $M$, $E_1(t), \cdots, E_n(t)$ a parallel frame along $p(t)$.*
(*a*) *Let $X(t) = \xi^i(t)\ E_i(t)$ be a* smooth *vector field along $p(t)$. Then*

$$\frac{\nabla X}{dt} = \frac{d\xi^i}{dt} E_i(t).$$

(*b*) *Let $v = a^i E_i(t_o) \in T_{p(t_o)}M$ a vector at $p(t_o)$. Then for any value of $t$,*

$$T(t_o \to t)v = a^i E_i(t).$$

**Proof**. (a) Since $\nabla E_i/dt = 0$, by hypothesis, we find

$$\frac{\nabla X}{dt} = \frac{\nabla(\xi^i E_i)}{dt} = \frac{d\xi^i}{dt} E_i + \xi^i \frac{\nabla E_i}{dt} = \frac{d\xi^i}{dt} E_i + 0$$

as required.
(b) Apply $T(t_o \to t)$ to $v = a^i E_i(t_o)$ we get by linearity,

$$T(t_o \to t)v = a^i T(t_o \to t) E_i(t_o) = a^i E_i(t)$$

as required. □

**Remarks**. This theorem says that with respect to a parallel frame along $p(t)$,
(a) the covariant derivative along $p(t)$ equals the componentwise derivative,
(b) the parallel transport along $p(t)$ equals the transport by constant components.

### Covariant differentiation of arbitrary tensor fields

2.1.16 **Theorem**. *There is a unique operation which produces for every "input" consisting of* (1) *a tangent vector $v \in T_pM$ at some point $p$ and* (2) *a* smooth *tensor field defined in a neighbourhood of $p$ a tensor $\nabla_v T$ at $p$ of the same type, denoted $\nabla_v T$, subject to the following conditions.*

0. *If $X$ is a vector field, the $\nabla_v X$ is its covariant derivative with respect to the given covariant-derivative operation on $M$.*
*1.* $\nabla_{(u+v)}T = (\nabla_u T) + (\nabla_v T)$
*2.* $\nabla_{av}T = a(\nabla_v T)$
*3.* $\nabla_v(T + S) = (\nabla_v T) + (\nabla_v S)$
*4.* $\nabla_v(S \cdot T) = (\nabla_v S) \cdot T(p) + S(p) \cdot (\nabla_v T)$
*for all $p \in M$, all $u, v \in T_pM$, all $a \in \mathbb{R}$, and all tensor fields $S, T$ defined in a neighbourhood of $p$. The products of tensors like $S \cdot T$ are tensor products $S \otimes T$ contracted with respect to any collection of indices (possibly not contracted at all).*

**Remark**. For given coordinates $(x^i)$, abbreviate $\partial_i := \partial/\partial x^i$ and $\nabla_i := \nabla_{\partial/\partial x^i}$. If $X$ is a vector field and $T$ a tensor field, then $\nabla_X T$ is a tensor field of the same type as $T$. If $X = X^k \partial_k$, then $\nabla_X T = X^k \nabla_k T$, by rules (1.) and (2.). So one only needs to know the components $(\nabla_k T)^i_j$ of $\nabla_k T$. The explicit formula is easily worked out, but will not be needed.

We shall indicate the proof after looking at some examples.

2.1.17 **Examples**.
(a) *Covariant derivative of a scalar function.* Let $f$ be a smooth covariant derivative of a covector scalar function on $M$. For every vector field $X$ we have

$$\nabla_v(fX) = (\nabla_v f)X(p) + f(p)(\nabla_v X)$$

by (4). On the other hand since $\nabla_v fX$ is the given covariant derivative, by (0.), we have

$$\nabla_v(fX) = (D_v f)X(p) + f(p)(\nabla_v X)$$

where $D_v f = \mathrm{d}f_p(v)$ by the axiom CD4. Thus $\nabla_v f = D_v f$.

(b) *Covariant derivative of a covector field* Let $F = F_i dx^i$ be a smooth covector field (differential 1-form). Let $X = X^i \partial_i$ be any vector field. Take $F \cdot X = F_i X^i$. By 4.

$$\nabla_j(F \cdot X) = (\nabla_j F) \cdot X + F \cdot (\nabla_j X)$$

In terms of components:

$$\nabla_j(F^i X_i) = (\nabla_j F)_i X^i + F_i(\nabla_j X)^i \tag{1}$$

By part (a), the left side is:

$$\partial_j(F_i X^i) = (\partial_j F_i)X^i + F_i(\partial_j X^i) \tag{2}$$

Substitute the right side of (1) for the left side of (2):

$$(\nabla_j F)_i \mathrm{X}^i + F_i(\nabla_j X)^i = (\partial_j F_i)\mathrm{X}^i + F_i(\partial_j X^i). \tag{3}$$

Rearrange terms in (3):

$$[(\nabla_j F)_i - \partial_j F_i]X^i = F_i[-(\nabla_j X)^i + \partial_j X^i]. \tag{4}$$

Since

$$\nabla_j X = (\partial_j X^i)\partial_i + X^i(\nabla_j \partial_i) = (\partial_j X^k)\partial_k + X^i \Gamma^k_{ij}\partial_k$$

we get

$$(\nabla_j X)^k = \partial_j X^k + X^i \Gamma^k_{ij} \tag{5}$$

Substitute (5) into (4) after changing an index $i$ to $k$ in (4):

$$[(\nabla_j F)_i - \partial_j F_i]X^i = F_k[-(\nabla_j X)^k + \partial_j X^k] = F_k[-X^i \Gamma^k_{ij}]$$

Since $X$ is arbitrary, we may compare coefficients of $X^i$:

$$(\nabla_j F)_i - \partial_j F_i = -F_k \Gamma^k_{ij}.$$

We arrive at the formula:

$$(\nabla_j F)_i = \partial_j F_i - F_k \Gamma^k_{ij}.$$

which is also written as $F_{i,j} = \partial_j F_i - F_k \Gamma^k_{ij}$.

In particular, for the coordinate differential $F = dx^a$, i.e. $F_i = \delta_{ia}$ we find

$$\nabla_j (dx^b) = -\Gamma^b_{ij} dx^i$$

which should be compared with

$$\nabla_j \Big(\frac{\partial}{\partial x^a}\Big) = \Gamma^k_{aj} \frac{\partial}{\partial x^k} \ .$$

The last two formulas together with the product rule 4. evidently suffice to compute $\nabla_v T$ for any tensor field $T = T^{a\cdots}_{b\cdots} dx^b \otimes \frac{\partial}{\partial x^a} \otimes \cdots$ . This proves the uniqueness part of the theorem. To prove the existence part one has to show that if one defines $\nabla_v T$ in this unique way, then 1.–4. hold, but we shall omit this verification.

### EXERCISES 2.1

1. (a) Verify that the covariant derivatives defined in example 2.1.2 above do indeed satisfy the axioms CD1 - CD5.
(b) Do the same for example 2.1.3.

2. Complete the proof of Theorem 2.1.10 by showing that $T = T_C(t_o \to t_1) : T_{p(t_o)}M \to T_{p(t_1)}M$ is linear, i.e. $T(u + v) = T(u) + T(v)$ and $T(av) = aT(v)$ for all $u, v \in T_{p(t_o)}M$, $a \in \mathbb{R}$.

3. Verify the assertions concerning parallel transport in $\mathbb{R}^n$ in example 2.1.11.

4. Prove theorem 2.1.16.

5. Let $M$ be the 3-dimensional vector space $\mathbb{R}^3$ with an indefinite inner product $(v, v) = -x^2 - y^2 + z^2$ if $v = xe_1 + ye_2 + ze_3$ and corresponding metric $-dx^2 - dy^2 + dz^2$. Let $S$ be the pseudo-sphere $S = \{p \in \mathbb{R}^3 \mid -x^2 - y^2 + z^2 = 1\}$.

(a) Let $p \in S$. Show that every vector $v \in \mathbb{R}^3$ can be uniquely written as a sum of a vector tangential to $S$ at $p$ an a vector orthogonal to $S$ at $p$:

$$v = a + b,\ a \in T_pS,\ (b \cdot w)q = 0 \text{ for all } w \in T_pS.$$

[Suggestion: Let e, $f$ be an orthonormal basis $e, f$ for $T_pS$: $(e \cdot e) = -1$, $(f \cdot f) = -1$, $(e \cdot f) = 0$. Then any a$\in T_pS$ can be written as a$= \alpha e + \beta f$ with $\alpha, \beta \in \mathbb{R}$. Try to determine $\alpha, \beta$ so that $b = v - a$ is orthogonal to $e, f$.]
(b) Define a covariant derivative on S as in the positive definite case:

$$\nabla_v^S X = \textit{component of } \nabla_v\ X \textit{ tangential to } T_pS.$$

Let $p_o$ be a point of $S$, $e \in T_{p_o}$S a unit vector (i.e. $(e, e) = -1$). Let $p(t)$ be the curve on $S$ defined by $p(t) = (\cosh\ t)p+ (\sinh\ t)e$. Find two vector fields $E(t)$, $F(t)$ along $p(t)$ so that parallel transport along $p(t)$ is given by

$$T(0 \to t)\colon T_{p_o}S \ \to\ T_{p(t)}S,\ aE(0) +\ bF(0)\ \to aE(t) +\ bF(t).$$

6. Let $S$ be the surface (torus) in Euclidean 3–space $\mathbb{R}^3$ with the equation

$$(\sqrt{x^2 + y^2} - a)^2 + z^2 = b^2 \qquad a > b > 0.$$

Let $(\theta, \phi)$ be the coordinates on $S$ defined by

$$x = (a + b\cos\phi)\cos\theta,\ y = (a + b\cos\phi)\sin\theta,\ z = b\sin\phi.$$

Let $C$: $p = p(t)$ be the curve with parametric equations $\phi = t, \theta = 0$ (constant).
a) Show that the vector fields $E = \|\partial p/\partial\phi\|^{-1}\partial p/\partial\phi$ and $F = \|\partial p/\partial\theta\|^{-1}\partial p/\partial\theta$ form a parallel orthonormal frame along $C$.
b) Consider the parallel transport along $C$ from $p_o = (a+b, 0, 0)$ to $p_1 = (a, 0, b)$. Find the parallel transport $v_1 = (\alpha_1, \beta_1, \gamma_1)$ of a vector $v_o = (\alpha_o, \beta_o, \gamma_o) \in \mathbb{R}^3$ tangent to $S$ at $p_o$.

7. Let $(\theta, r)$ be polar coordinates on $\mathbb{R}^2$: $x = r\cos\theta$, $y = r\sin\theta$. Calculate all $\Gamma_{ij}^k$ for the covariant derivative on $\mathbb{R}^2$ defined in example 2.1.2. (Use $r, \theta$ as indices instead of $1, 2$.)

8. Let $(\rho, \theta, \phi)$ be spherical coordinates on $\mathbb{R}^3$: $x = \rho\sin\phi\cos\theta, y = \rho\sin\phi\sin\theta$, $z = \rho\cos\phi$. Calculate $\Gamma_{\phi\rho}^k(k = \rho, \phi, \theta)$ for the covariant derivative on $\mathbb{R}^3$ defined in example 2.1.2. (Use $\rho, \phi, \theta$ as indices instead of $1, 2, 3$.)

9. Define coordinates $(u, v, z)$ on $\mathbb{R}^3$ by the formula $x = \frac{1}{2}(u^2 - v^2), y = uv, z = z$. Calculate $\Gamma_{uz}^k$, $kj = u, v, z$, for the covariant derivative on $\mathbb{R}^3$ defined in example 2.1.2. (Use $u, v, z$ as indices instead of $1, 2, 3$.)

10. Let $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid z = x^2 + y^2\}$ and let $\nabla = \nabla^S$ be the induced covariant derivative. Define a coordinate system $(r, \theta)$ on $S$ by $x = r\cos\theta$, $y = r\text{sin}\theta$, $z = r^2$. Calculate $\nabla_r\partial_r$ and $\nabla_\theta\partial_\theta$ .

11. Let $S$ be a surface of revolution, $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid r = f(z)\}$ where $r^2 = x^2 + y^2$ and $f$ a positive differentiable function. Let $\nabla = \nabla^S$ be

the induced covariant derivative. Define a coordinate system $(z, \theta)$ on $S$ by $x = f(z)\cos\theta$, $y = f(z)\sin\theta$, $z = z$.
a) Show that the coordinate vector fields $\partial_z = \partial p/\partial z$ and $\partial_\theta = \partial p/\partial\theta$ are orthogonal at each point of $S$.
b) Let $C$: $p = p(t)$ be a meridian on $S$, i.e. a curve with parametric equations $z = t$, $\theta = \theta_o$. Show that the vector fields $E = \|\partial_\theta\|^{-1}\partial_\theta$ and $F = \|\partial_z\|^{-1}\partial_z$ form a parallel frame along $C$.
[Suggestion. Show first that $E$ is parallel along $C$. To prove that $F$ is also parallel along $C$ differentiate the equations $(F \cdot F) = 1$ and $(F \cdot E) = 0$ with respect to $t$.]

12. Let $S$ be a surface in $\mathbb{R}^3$ with its induced connection $\nabla$. Let $C$: $p = p(t)$ be a curve on $S$.
a) Let $E(t), F(t)$ be two vector fields along $C$. Show that

$$\frac{d}{dt}(E \cdot F) = \Big(\frac{\nabla E}{dt} \cdot F\Big) + \Big(E \cdot \frac{\nabla F}{dt}\Big)$$

b) Show that parallel transport along $C$ preserves the scalar products vectors, i.e.

$$(T_C(t_o \to t_1)u, T_C(t_o \to t_1)v) = (u, v)$$

for all $u, v \in T_{p(t_o)}S$.
c) Show that there exist orthonormal parallel frames $E_1, E_2$ along $C$.

13. Let $T$ be a tensor-field of type (1,1). Derive a formula for $(\nabla_k T)^i_j$. [Suggestion: Write $T = T^{\mathrm{a}}_{\mathrm{b}}\frac{\partial}{\partial x^{\mathrm{a}}} \otimes dx^{\mathrm{b}}$ and use the appropriate rules 0.– 4.]

## 2.2 Geodesics

2.2.1 **Definition**. Let $M$ be a manifold with a connection $\nabla$. A smooth curve $p = p(t)$ on $M$ is said to be a *geodesic* (of the connection) $\nabla$ if $\nabla\dot{p}/dt = 0$ i.e. the tangent vector field $\dot{p}(t)$ is parallel along $p(t)$. We write this equation also as

$$\frac{\nabla^2 p}{dt^2} = 0. \tag{1}$$

In terms of the coordinates $(x^1(t), \cdots, x^n(t))$ of $p = p(t)$ relative to a coordinate system $x^1, \cdots, x^n$ the equation (1) becomes

$$\frac{d^2x^k}{dt^2} + \sum_{ij} \Gamma^k_{ij}\frac{dx^i}{dt}\,\frac{dx^j}{dt} = 0.$$

2.2.2 **Example**. Let $M = \mathbb{R}^n$ with its usual covariant derivatives. For $p(t) = (x_1(t), \cdots, x_n(t))$ the equation $\nabla^2 p/dt^2 = 0$ says that $\ddot{x}_i(t) = 0$ for all $t$, so $x_i(t) = a_i + tb_i$ and $x(t) = a + tb$ is a straight line.

2.2.3 **Example**. Let $S = \{p \in E \mid \|p\| = 1\}$ be the unit sphere in a Euclidean 3-space $E = \mathbb{R}^3$. Let $C$ be the great circle through $p \,\epsilon\, S$ with direction vector

$e$, i,e. $p(t) = \cos(t)p + \sin(t)e$ where $e \cdot p = 0$, $e \cdot e = 1$. We have shown previously that $\nabla^2 p/dt^2 = 0$. Thus each great circle is a geodesic. The same is true if we change the parametrization by some constant factor, say $\alpha$, so that $p(t) = \cos(\alpha t)p+ \sin(\alpha t)e$. It will follow from the theorem below that every geodesic is of this form

2.2.4 **Theorem**. *Let $M$ be a manifold with a connection $\nabla$. Let $p \in M$ be a point in $M$, $v \in T_pM$ a vector at $p$. Then there is a unique geodesic $p(t)$, defined on some interval around t= 0, so that $p(0) = p, \dot{p}(0) = v$.*

**Proof**. In terms of a coordinate system around $p$, the condition on $p$ is

$$\frac{d^2x^k}{dt^2} + \sum_{ij} \Gamma^k_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0,$$

$$x^k(0) = x_0, \quad \frac{dx^k(0)}{dt} = v^k.$$

This system of differential equations has a unique solution subject to given the initial condition. □

2.2.5 **Definition**. Fix $p_o \in M$. A coordinate system $(\bar{x}^t)$ around $p_o$ is said to be a *local inertial frame* (or *locally geodesic coordinate system*) at $p_o$ if all $\bar{\Gamma}^t_{rs}$ vanish at $p_o$, i.e. if

$$\Big(\frac{\nabla}{\partial \bar{x}^r}\frac{\partial}{\partial \bar{x}^s}\Big)_{p_o} = 0 \text{ for all } rs.$$

2.2.6 **Theorem**. *There exists a local inertial frame $(\bar{x}^t)$ around $p_o$ if and only if the $\Gamma^k_{ij}$ with respect to an arbitrary coordinate system $(x^i)$ satisfy*

$$\Gamma^k_{ij} = \Gamma^k_{ji} \text{ at } p_o \text{ (all } ijk) \quad \textit{i.e.} \quad \frac{\nabla}{\partial x^i}\frac{\partial}{\partial x^j} = \frac{\nabla}{\partial x^j}\frac{\partial}{\partial x^i} \text{ at } p_o \text{ (all } ij).$$

**Proof**. We shall need the transformation formula for the $\Gamma^k_{ij}$

$$\bar{\Gamma}^t_{rs} = \frac{\partial \bar{x}^t}{\partial x^k}\Big(\frac{\partial^2 x^k}{\partial \bar{x}^r \partial \bar{x}^s} + \frac{\partial x^i}{\partial \bar{x}^r}\frac{\partial x^j}{\partial \bar{x}^s}\Gamma^k_{ij}\Big)$$

whose proof is left as an exercise. Start with an arbitrary coordinate system $(x^i)$ around $p_o$ and consider a coordinate transformation expanded into a Taylor to second order:

$$(x^k - x^k_o) = a^k_t(\bar{x}^t - \bar{x}^t_o) + \frac{1}{2}a^k_{rs}(\bar{x}^r - \bar{x}^r_o)(\bar{x}^s - \bar{x}^s_o) + \cdots$$

Thus $\partial x^k/\partial \bar{x}^t = a^k_t$ and $\partial^2 x^k/\partial \bar{x}^r \partial \bar{x}^s = a^k_{rs}$ at $p_o$. We want to determine the coefficients $a^k_t, a^k_{rs}, \cdots$ so that the transformed $\bar{\Gamma}^t_{rs}$ all vanish at $p_o$. The coefficients $a^k_t$ must form an invertible matrix, representing the differential of the coordinate transformation at $p_o$. The coefficients $a^k_{rs}$ must be symmetric in $rs$, representing the second order term in the Taylor series. These coefficients

$a^k_t, a^k_{rs}, \cdots$ must further be chosen so that the term in parentheses in the transformation formula vanishes, i.e. $a^k_{rs} + a^i_r a^j_s \Gamma^k_{ij} = 0$. For each $k$, this equation determines the $a^k_{rs}$ uniquely, but requires that $\Gamma^k_{ij}$ be symmetric in $ij$, since $a^k_{rs}$ has to be symmetric in $rs$. The higher-order terms are left arbitrary. (They could all be chosen equal to zero, to have a specific coordinate transformation.) Thus a coordinate transformation of the required type exists if and only if the $\Gamma^k_{ij}$ are symmetric in $ij$. □

2.2.7 **Remark**. The proof shows that, relative to a given coordinate system $(x^i)$, the Taylor series of a local inertial frame $(\bar{x}^i)$ at $p_o$ is uniquely up to second order if the first order term is fixed arbitrarily. For this one can see that a local inertial frame is unique up to a transformation of the form

$$x^i - x^i_o = a^i_t(\bar{x}^t - \bar{x}^t_o) + o(2)$$

where $o(2)$ denotes terms of degree $> 2$.

2.2.8 **Definition**. The covariant derivative $\nabla$ is said to be symmetric if the condition of the theorem holds for all point $p_o$, i.e. if $\Gamma^k_{ij} \equiv \Gamma^k_{ji}$ for some (hence every) coordinate system $(x^i)$.

2.2.9 **Lemma**. Fix $p_o \in M$. For $v \in T_{p_o}M$, let $p_v(t)$ be the unique geodesic with $p_v(0) = p_o$, $p'_v(0) = v$. Then $p_v(at) = p_{av}(t)$ for all $a, t \in \mathbb{R}$.

**Proof**. Calculate

$$\begin{aligned}
\frac{d}{dt}\Big(p_v(at)\Big) &= a\Big(\frac{dp_v}{dt}\Big)(at)\\
\frac{\nabla}{dt}\frac{d}{dt}\Big(p_v(at)\Big) &= a^2\Big(\frac{\nabla}{dt}\frac{dp_v}{dt}\Big)(at) = 0\\
\frac{d}{dt}p_v(at)\Big|_{t=0} &= a^2\ \frac{dp_v}{dt}(at)\Big|_{t=0} = av
\end{aligned}$$

Hence $p_v(at) = p_{av}(t)$. □

2.2.10 **Corollary**. *Fix $p \in M$. There is a unique map $T_pM \to M$ which sends the line $tv$ in $T_pM$ to the geodesic $p_v(t)$ with initial velocity $v$.*

**Proof**. This map is given by $v \to p_v(1)$: it follows from the lemma that $p_v(t) = p_{tv}(1)$, so this map does send the line $tv$ to the geodesic $p_v(t)$. □

2.2.11 **Definition**. The map $T_pM \to M$ of the previous corollary is called the *geodesic spray or exponential map* of the connection at $p$, denoted $v \to \exp_p v$.

The exponential notation may seem strange; it can be motivated as follows. By definition, the tangent vector to the curve $\exp_p(tv)$ parallel transport of $v$ from 0 to $t$. If this parallel transport is denoted by $T(\exp_p(tv))$, then the definition says that

$$\frac{d\exp_p(tv)}{dt} = T(\exp_p(tv))v,$$

which is a differential equation for $\exp_p(tv)$ formally similar to the equation $d\exp(tv)/dt = \exp(tv)v$ for the scalar exponential. That the map $\exp_p : T_pM \to M$ is smooth follows from a theorem on differential equation, which guarantees that the solution $x^k = x^k(t)$ of (3) depends in a smooth fashion on the initial conditions (4).

2.2.12 **Theorem**. *For any $p \in M$ the geodesic spray $\exp_p : T_pM \to M$ maps a neighbourhood of 0 in $T_pM$ one-to-one onto a neighbourhood of $p$ in $M$.*

**Proof**. We apply the inverse function theorem. The differential of $\exp_p$ at 0 is given by

$$v \to \Big(\frac{d}{dt}\exp_p(tv)\Big)_{t=0} = v$$

hence is the identity on $T_pM$. □

This theorem allows us to introduce a coordinate system $(x^i)$ in a neighbourhood of $p_o$ by using the components of $v \in T_{p_o}M$ with respect to some fixed basis $e_1, \cdots, e_n$ as the coordinates of the point $p = \exp_{p_o} v$. In more detail, this means that the coordinates $(x^i)$ are defined by the equation $p = \exp_{p_o}(x^1e_1 + \cdots + x^ne_n)$.

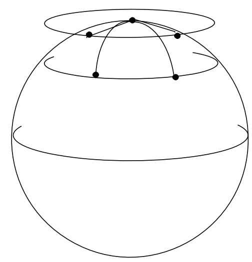


Fig. 1. The geodesic spray on $S^2$

2.2.13 **Theorem**. *Assume $\nabla$ is symmetric. Fix $p_o \in M$, and a basis $(e_1, \cdots, e_n)$ for $T_{p_o}M$. Then the equation $p = \exp_{p_o}(x^1e_1 + \cdots + x^ne_n)$ defines a local inertial frame $(x^i)$ at $p_o$.*

2.2.14 **Definition**. Such a coordinate system $(x^i)$ around $p_o$ is called a *normal inertial frame at $p_o$* (or a *normal geodesic coordinate system at $p_o$*). It is unique up to a linear change of coordinates $\bar{x}^j = a^j_i x^i$, corresponding to a change of basis $\bar{e}_i = a^j_i e_j$.

**Proof of the theorem**. We have to show that all $\Gamma^k_{ij} = 0$ at $p_o$. Fix $v = \xi^i e_i$. The $(x^i)$-coordinates of $p = \exp_{p_o}(tv)$ are $x^i = t\xi^i$. Since these $x^i = x^i(t)$ are the coordinates of the geodesic $p_v(t)$ we get

$$\frac{d^2x^k}{dt^2} + \sum_{ij} \Gamma^k_{ij}\frac{dx^i}{dt}\,\frac{dx^j}{dt} = 0 \text{ (for all } k).$$

This gives

$$\sum_{ij} \Gamma^k_{ij}\xi^i\xi^j = 0.$$

Here $\Gamma^k_{ij} = \Gamma^k_{ij}(t\xi^1, \cdots, t\xi^n)$. Set $t = 0$ in this equation and then take the second-order partial $\partial^2/\partial\xi^i\partial\xi^j$ to find that $\Gamma^k_{ij} + \Gamma^k_{ji} = 0$ at $p_o$ for all $ijk$. Since $\nabla$ is symmetric, this gives $\Gamma^k_{ij} = 0$ at $p_o$ for all $ijk$, as required. $\square$

**EXERCISES 2.2**

1. Use theorem 2.2.4 to prove that every geodesic on the sphere $S$ of example 2.2.3 is a great circle.

2. Referring to exercise 7 of 2.2, show that for each constant $\alpha$ the curve $p(t) = (\cosh \alpha t)p_o +$ $(\sinh \alpha t)e$ is a geodesic on the pseudosphere $S$.

3. Prove the transformation formula for the $\Gamma^k_{ij}$ *from the definition* of the $\Gamma^k_{ij}$.

4. Suppose( $\bar{x}^i$) is a local inertial frame at $p_o$. Let $(x^i)$ be a coordinate system so that

$$x^i - x^i_o = a^i_t(\bar{x}^t - \bar{x}^t_o) + \mathrm{o}(2)$$

as in remark 2.2.7. Prove that $(x^i)$ is also a local inertial frame at $p_o$. (It can be shown that *any* two local inertial frames at $p_o$ are related by a coordinate transformation of this type.)

5. Let $(\bar{x}^i)$ be a coordinate system on $\mathbb{R}^n$ so that $\bar{\Gamma}^t_{rs} = 0$ at all points of $\mathbb{R}^n$. Show that $(\bar{x}^i)$ is an *affine* coordinate system, i.e. $(\bar{x}^i)$ is related to the Cartesian coordinate system $(x^i)$ by an affine transformation:

$$\bar{x}^t = \bar{x}^t_o + a^t_j(x^j - x^j_o).$$

6. Let $S$ be the cylinder $x^2 + y^2 = 1$ in $\mathbb{R}^3$ with the induced connection. Show that for any constants $a$,$b$,$c$,$d$ the curve $p = p(t)$ with parametric equations

$$x = \cos(at + b),\ y = \sin(at + b),\ z = ct + d$$

is a geodesic and that every geodesic is of this form.

7. Let $p = p(t)$ be a geodesic on some manifold $M$ with a connection $\nabla$. Let $p = p(t(s))$ be the curve obtained by a change of parameter $t = t(s)$. Show that $p(t(s))$ is a geodesic if and only if $t = as + b$ for some constants $a, b$.

8. Let $p = p(t)$ be a geodesic on some manifold $M$ with a connection $\nabla$. Suppose there is $t_o \neq t_1$ so that $p(t_o) = p(t_1)$ and $\dot{p}(t_o) = \dot{p}(t_1)$ . Show that the geodesic is periodic, i.e. there is a constant $c$ so that $p(t + c) = p(t)$ for all $t$.

9. Let $S$ be a surface of revolution, $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid r = f(z)\}$ where $r^2 = x^2 + y^2$ and $f$ a positive differentiable function. Let $\nabla = \nabla^S$ be the induced covariant derivative. Define a coordinate system $(z, \theta)$ on $S$ by $x = f(z)\cos\theta$, $y = f(z)\sin\theta$, $z = z$. Let $p = p(t)$ be a curve of the form $\theta = \theta_o$ parametrized by

arclength, i.e. with parametric equations $\theta = \theta_o$, $z = z(t)$ and $(\dot{p}(t) \cdot \dot{p}(t)) \equiv 1$. Show that $p(t)$ is a geodesic. [Suggestion. Show that the vector $\ddot{p}(t)$ in $\mathbb{R}^3$ is orthogonal to both coordinate vector fields $\partial_z$ and $\partial_\theta$ on $S$ at $p(t)$. Note that $\dot{p}(t) = (\partial p/\partial z)\dot{z}(t)$ is parallel to $\partial_z$ while $\partial_z$ and $\partial_\theta$ are orthogonal. Differentiate the equations $(\dot{p}(t) \cdot \dot{p}(t)) \equiv 1$ and $(\partial_z \cdot \partial_\theta) \equiv 0$ with respect to $t$.]

## 2.3 Riemann curvature

Throughout, $M$ is a manifold with a connection $\nabla$.

2.3.1 **Definition**. A *parametrized surface* in $M$ is a smooth function $\mathbb{R}^2 \to M$, $(\sigma, \tau) \mapsto p = p(\sigma, \tau)$, defined (at least) on some rectangle $\{(\sigma, \tau) \mid |\sigma - \sigma_o| \leq a, |\tau - \tau_o| \leq b\}$.

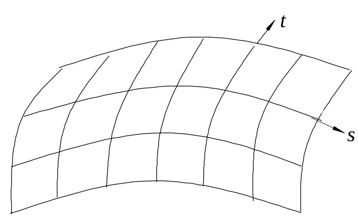



Fig. 1. A parametrized surface

A *vector field along a parametrized surface* assigns to each point $p = p(\sigma, \tau)$ a vector $W = W(\sigma, \tau)$ in $T_pM$. The *parameter vector fields* $U = \partial p(\sigma, \tau)/\partial\sigma$ and $V = \partial p(\sigma, \tau)/\partial\tau$ along $p = p(\sigma, \tau)$ are furthermore tangential to the surface, but in general $W = W(\sigma, \tau)$ need not be. The covariant derivatives along the parameter vector fields are the covariant derivatives along the $\sigma$–lines and $\tau$–lines, denoted $\nabla/\partial\sigma$ and $\nabla/\partial\tau$.

2.3.2 **Theorem**. Let $p = p(\sigma, \tau)$ be a parametrized surface. *The vector*

$$T(U, V) := \frac{\nabla}{\partial\tau}\frac{\partial p}{\partial\sigma} - \frac{\nabla}{\partial\sigma}\frac{\partial p}{\partial\tau}$$

*depends only on the vectors $U$ and $V$* at the point $p(\sigma, \tau)$. *Relative to a coordinate system $(x^i)$ this vector is given by*

$$T(U, V) = T^k_{ij} U^i V^j \frac{\partial}{\partial x^k}$$

*where $T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}$ is a tensor of type $(1, 2)$, called the torsion tensor of the connection $\nabla$.*

**Proof**. In coordinates $(x^i)$ write $p = p(\sigma, \tau)$ as $x^i = x^i(\sigma, \tau)$. Calculate:

$$\begin{aligned}
\frac{\partial p}{\partial \sigma} &= \frac{\partial x^i}{\partial \sigma}\frac{\partial}{\partial x^i} \\
\frac{\nabla}{\partial \tau}\frac{\partial p}{\partial \sigma} &= \frac{\nabla}{\partial \tau}\Big(\frac{\partial x^i}{\partial \sigma}\frac{\partial}{\partial x^i}\Big) \\
&= \frac{\partial^2 x^i}{\partial \tau \partial \sigma}\frac{\partial}{\partial x^i} + \frac{\partial x^i}{\partial \sigma}\frac{\nabla}{\partial \tau}\frac{\partial}{\partial x^i} \\
&= \frac{\partial^2 x^i}{\partial \tau \partial \sigma}\frac{\partial}{\partial x^i} + \frac{\partial x^i}{\partial \sigma}\frac{\partial x^j}{\partial \tau}\frac{\nabla}{\partial x^j}\frac{\partial}{\partial x^i}
\end{aligned}$$

Interchange $\sigma$ and $\tau$ and subtract to get

$$\frac{\nabla}{\partial \tau}\frac{\partial p}{\partial \sigma} - \frac{\nabla}{\partial \sigma}\frac{\partial p}{\partial \tau} = \frac{\partial x^i}{\partial \sigma}\frac{\partial x^j}{\partial \tau}\Big(\frac{\nabla}{\partial x^j}\frac{\partial}{\partial x^i} - \frac{\nabla}{\partial x^i}\frac{\partial}{\partial x^j}\Big).$$

Substitute $\frac{\nabla}{\partial x^j}\frac{\partial}{\partial x^i} = \Gamma^k_{ij}\frac{\partial}{\partial x^k}$ to find

$$\frac{\nabla}{\partial \tau}\frac{\partial p}{\partial \sigma} - \frac{\nabla}{\partial \sigma}\frac{\partial p}{\partial \tau} = \frac{\partial x^i}{\partial \sigma}\frac{\partial x^j}{\partial \tau}(\Gamma^k_{ij} - \Gamma^k_{ji})\frac{\partial}{\partial x^k} = (\Gamma^k_{ij} - \Gamma^k_{ji})U^i V^j\frac{\partial}{\partial x^k}.$$

Read from left to right this equation shows first of all that the left side depends only on $U$ and $V$. Read from right to left the equation shows that the vector on the right is a bilinear function of $U, V$ independent of the coordinates, hence defines a tensor $T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}$. □

2.3.3 *Theorem.* Let $p = p(\sigma, \tau)$ be a parametrized surface and $W = W(\sigma, \tau)$ a vector field along this surface. The vector

$$R(U, V)W := \frac{\nabla}{\partial \tau}\frac{\nabla W}{\partial \sigma} - \frac{\nabla}{\partial \sigma}\frac{\nabla W}{\partial \tau}$$

*depends only on the vectors* $U, V, W$ at the point $p(\sigma, \tau)$. *Relative to a coordinate system* $(x^i)$ *this vector is given by*

$$R(U, V)W = R^i_{mkl}U^k V^l W^m \frac{\partial}{\partial x^i}$$

*where* $\partial_i = \partial/\partial x^i$ *and*

$$R^i_{mkl} = (\partial_l \Gamma^i_{mk} - \partial_k \Gamma^i_{ml} + \Gamma^p_{mk}\Gamma^i_{pl} - \Gamma^p_{ml}\Gamma^i_{pk}) \qquad (R)$$

$R = (R^i_{mkl})$ *is a tensor of type* $(1, 3)$, *called the curvature tensor of the connection* $\nabla$.

The proof requires computation like the one in the preceding theorem, which will be omitted.

2.3.4 **Lemma**. *For any vector field* $X(t)$ *and any curve* $p = p(t)$

$$\frac{\nabla X(t)}{dt} = \lim_{\epsilon \to 0}\frac{1}{\epsilon}\Big(X(t+\epsilon) - T(t \to t+\epsilon)X(t)\Big).$$

**Proof**. Recall the formula for the covariant derivative in terms of a parallel frame $E_1(t), \cdots, E_n(t)$ along $p(t)$ (Theorem 11.15): if $X(t) = \xi^i(t)E_i(t)$, then

$$\begin{aligned}\frac{\nabla X}{dt} &= \frac{d\xi^i}{dt}E_i(t)\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(\xi^i(t+\epsilon)-\xi^i(t)\Big)E_i(t)\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(\xi^i(t+\epsilon)-\xi^i(t)\Big)E_i(t+\epsilon)\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(\xi^i(t+\epsilon)E_i(t+\epsilon)-\xi^i(t)E_i(t+\epsilon)\Big)\\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big(X(t+\epsilon)-T(t\to t+\epsilon)X(t)\Big),\end{aligned}$$

the last equality coming from the fact the components with respect to a parallel frame remain under parallel transport. □

We now give a geometric interpretation of $R$.

2.3.5 **Theorem**. *Let $p_o$ be a point of $M$, $U, V, W \in T_{p_o}M$ three vectors at $p_o$. Let $p = p(\sigma,t)$ be a surface so that $p(\sigma_o, \tau_o) = p_o$, $\partial p/\partial\sigma|_{(\sigma_o\tau_o)} = U$, $\partial p/\partial\tau|_{(\sigma_o\tau_o)} = V$ and $X$ a* smooth *defined in a neighbourhood of $p_o$ with $X(p_o) = W$. Define a linear transformation $T = T(\sigma_o, \tau_o; \Delta\sigma, \Delta\tau)$ of $T_{p_o}M$ by $TW$ =parallel transport of $W \in T_{p_o}M$ along the boundary of the "rectangle" $\{p(\sigma,\tau) \mid \sigma_o \le \sigma \le \sigma_o + \Delta\sigma,\ \tau_o \le \tau \le \Delta\tau\}$. Then*

$$R(U,V)W = \lim_{\Delta\sigma,\Delta\tau\to 0}\frac{1}{\Delta\sigma\Delta\tau}(W - TW) \qquad (1)$$

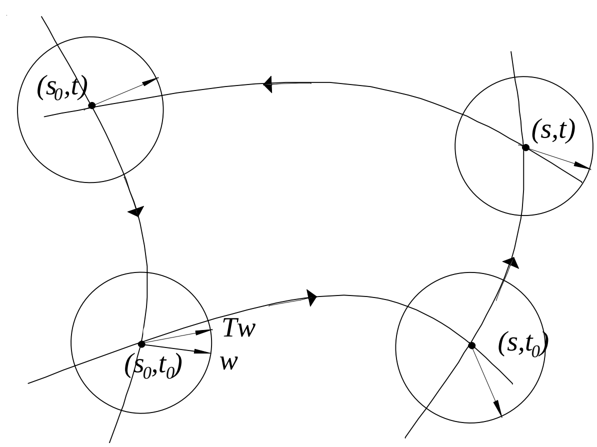



Fig. 2. Parallel transport around a parallelogram.

**Proof**. Set $\sigma = \sigma_o + \Delta\sigma$, $\tau = \tau_o + \Delta\tau$. Decompose $T$ into the parallel transport along the four sides of the "rectangle":

$$TW = T(\sigma_o, \tau \to \tau_o)T(\sigma \to \sigma_o, \tau)T(\sigma, \tau_o \to \tau)T(\sigma_o \to \sigma, \tau_o)W. \qquad (2)$$

Imagine this expression substituted for $T$ in the limit on the right-hand side of (1). Factor out the first two terms of (2) from the parentheses in (1) to get

$$\begin{aligned}
&\lim_{\Delta\sigma,\Delta\tau\to 0}\frac{1}{\Delta\sigma\Delta\tau}(W-TW) \ = \\
&\quad = \lim_{\Delta\sigma,\Delta\tau\to 0}\frac{1}{\Delta\sigma\Delta\tau}\Big\{T(\sigma_o,\tau\to\tau_o)T(\sigma\to\sigma_o,\tau)\Big\}\circ \\
&\qquad \circ\Big\{T(\sigma_o\to\sigma,\tau)T(\sigma_o,\tau_o\to\tau)W - T(\sigma,\tau_o\to\tau)T(\sigma_o\to\sigma,\tau_o)\Big)W\Big\}.
\end{aligned} \tag{3}$$

Here we use $T(\sigma\to\sigma_o,\tau)T(\sigma_o\to\sigma,\tau)=I$ (= identity) etc. The expression in the first braces in (3) approaches $I$. It may be omitted from the limit (3). (This means that we consider the difference between the transports from $p(\sigma_o,\,\tau_o)$ to the opposite vertex $p(\sigma,\tau)$ along "left" and "right" path rather than the change all the way around the rectangle.) To calculate the limit of the expression in the second braces in (3) we write the formula in the lemma as

$$\Big(\frac{\nabla X}{dt}\Big)_{t_o} = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\Big(X(t)-T(t_o\to t)X(t_o)\Big) \tag{4}$$

Choose a smooth vector field $X(\sigma,\tau)$ along the surface $p(\sigma,\tau)$ so that $W = X(\sigma_o,\tau_o)$. With (4) in mind we rewrite (3) by adding and subtracting some terms:

$$\begin{aligned}
\lim_{\Delta\sigma,\Delta\tau\to 0}\frac{1}{\Delta\sigma\Delta\tau}(W-TW) = \lim_{\Delta\sigma,\Delta\tau\to 0}&\frac{1}{\Delta\sigma\Delta\tau} \\
T(\sigma_o\to\sigma,\tau)[T(\sigma_o,\tau_o\to\tau)X(\sigma_o,\tau_o)&-X(\sigma_o,\tau)] \\
+[(T(\sigma_o\to\sigma,\tau)X(\sigma_o,\tau)&-X(\sigma,\tau)] \\
-T(\sigma,\tau_o\to\tau)[T(\sigma_o\to\sigma,\tau_o)X(\sigma_o,\tau_o)&-X(\sigma,\tau_o)] \\
-[T(\sigma,\tau_o\to\tau)X(\sigma,\tau_o)&-X(\sigma,\tau)]
\end{aligned} \tag{5}$$

This means that decompose the difference between the left and right path into changes along successive sides:

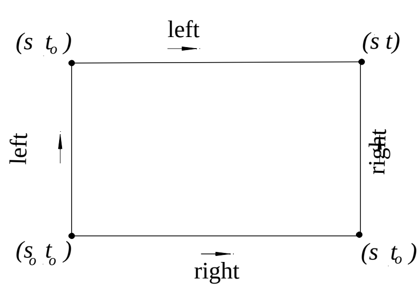



Fig.3. Left and right path.

Multiply through by $1/\Delta\sigma\Delta\tau$ and take the appropriate limits of the expression in the brackets using (4) to get:

$$\begin{aligned}
&\lim_{\Delta\sigma,\Delta\tau\to 0} \frac{1}{\Delta\sigma\Delta\tau}(W-TW)\\
&= \text{limit of } \Big\{ -\frac{1}{\Delta\sigma} T(\sigma_o\to\sigma,\tau)\Big(\frac{\nabla X}{\partial\tau}\Big)_{(\sigma_o,\tau_o)} - \frac{1}{\Delta\tau}\Big(\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau)}\\
&\qquad + \frac{1}{\Delta\tau} T(\sigma,\tau_o\to\tau)\Big(\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau_o)} + \frac{1}{\Delta\sigma}\Big(\frac{\nabla X}{\partial\tau}\Big)_{(\sigma,\tau_o)}\Big\}\\
&= \text{limit of } \Big\{ \frac{1}{\Delta\sigma}\Big\{\Big(\frac{\nabla X}{\partial\tau}\Big)_{(\sigma,\tau_o)} - T(\sigma_o\to\sigma,\tau)\Big(\frac{\nabla X}{\partial\tau}\Big)_{(\sigma_o,\tau_o)}\Big\}\\
&\qquad + \frac{1}{\Delta\tau}\Big\{T(\sigma,\tau_o\to\tau)\Big(\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau_o)} - \Big(\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau)}\Big\}\Big\}\\
&= \Big(\frac{\nabla}{\partial\sigma}\frac{\nabla X}{\partial\tau} - \frac{\nabla}{\partial\tau}\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau_o)} = R(U,V)W.
\end{aligned}$$

□

2.3.6 **Theorem** (Riemann). *If $R=0$, then locally around any point there is a coordinate system $(x^i)$ on $M$ such that $\Gamma^k_{ij}\equiv 0$, i.e. with respect to this coordinate system $(x^i)$, the covariant derivative $\nabla$ is the componentwise derivative:*

$$\nabla_Y\Big(\sum X^i\frac{\partial}{\partial x^i}\Big) = \sum_i (D_Y X^i)\frac{\partial}{\partial x^i}.$$

*The coordinate system is unique up to an affine transformation: any other such coordinate system $(\bar{x}^j)$ is of the form*

$$\bar{x}^j = \bar{x}^j_o + c^j_i x^i$$

*for some constants $\bar{x}^j_o$, $c^j_i$ with* $\det c^j_i \neq 0$.

2.3.7 **Remark**. If $R=0$ the covariant derivative $\nabla$ is said to be flat. A coordinate system $(x^i)$ as in the theorem is called affine. With respect to such an affine coordinate system $(x^i)$ parallel transport $T_{p_o}M\to T_{p_1}M$ along any curve leaves unchanged the components with respect to $(x^i)$. In particular, for a flat connection parallel transport is independent of the path.

**Proof**. Riemann's own proof requires familiarity with the integrability conditions for systems of first-order partial differential equations. Weyl outlines an appealing geometric argument in §16 of "Raum, Zeit, Materie" (1922 edition). It runs like this. Assume $R=0$, i.e. the parallel transport around an infinitesimal rectangle is the identity transformation on vectors. The same is then true for any finite rectangle, since it can be subdivided into arbitrarily small ones. This implies that the parallel transport between any two points is independent of the path, as long as the two points can be realized as opposite vertices of some such rectangle.

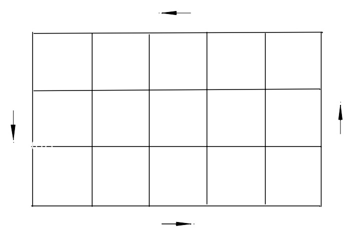


Fig.3. Subdividing a rectangle.

Fix a point $p_o$. A vector $V$ at $p_o$ can be transported to all other points (independently of the path) producing a parallel vector field $X$. As coordinates of the point $p(1)$ on the solution curve $p(t)$ of $dp/dt = X(p)$ through $p(0) = p_o$ take the components $(x^i)$ of $V = x^1e_1 + \cdots + x^ne_n$ with respect to a fixed basis for $T_{p_o}M$. All of this is possible at least in some neighbourhood of $p_o$ and provides a coordinate system there.

One may think of the differential equation $dp/dt = X(p)$ as an infinitesimal transformation of the manifold which generates the finite transformation defined by $p(0) \mapsto p(t)$. The induced transformation on tangent vectors is the parallel transport (proved below). This means that the parallel transport proceeds by keeping constant the components of vectors in the coordinate system $(x^i)$ just constructed. The covariant derivative is then the componentwise derivative. If $(\bar{x}^i)$ is any other coordinate system with this property, then $\partial/\partial x^j = c_i^j\partial/\partial\bar{x}^i$ for certain constants $c_j^i$, because all of these vector fields are parallel. Thus $\bar{x}^j = \bar{x}_o^j + c_i^j x^i$. □

As Weyl says, the construction shows that a connection with $R = 0$ gives rise to a commutative group of transformations which (locally) implements the parallel transport just like the group of translations of an affine space.

To prove the assertion in italics, write $p_o \mapsto \exp(tX)p_o$ for the transformation defined $d\exp(tX)p_o/dt = X(\exp(tX)p_o)$, $p(0) = p_o$. The induced transformation on tangent vectors sends the tangent vector $V = \dot{p}(s_o)$ of a differentiable curve $p(s)$ through $p_o = p(s_o)$ to the vector $d\exp(tX)p(s)/ds|_{s=s_o}$ at $\exp(tX)p_o$. The assertion is that these vectors are parallel along $\exp(tX)p_o$. This follows straight from the definitions:

$$\begin{aligned}
&\frac{\nabla}{dt}\frac{d}{ds}\exp(tX)p(s) = \\
&\quad = \frac{\nabla}{ds}\frac{d}{dt}\exp(tX)p(s) \quad [R = 0] \\
&\quad = \frac{\nabla}{ds}X(p(s)) \quad [\text{definition of } \exp(tX)] \\
&\quad = 0 \quad [X(p(s)) \text{ is parallel, by definition}].
\end{aligned}$$

## EXERCISES 2.3

1. Let $\nabla$ be a covariant derivative on $M$, $T$ its torsion tensor.

(a) Show that the equation

$$'\nabla_V X = \nabla_V X + \frac{1}{2} T(V, X)$$

defines another covariant derivative $'\nabla$ on $M$. [Verify at least the product rule CD4.]
(b) Show that $'\nabla$ is symmetric i.e the torsion tensor $'T$ of $'\nabla$ is $'T = 0$.
(c) Show that $\nabla$ and $'\nabla$ have the same geodesics, i.e. a smooth curve $p = p(t)$ is a geodesic for $\nabla$ if and only if it is a geodesic for $'\nabla$.

2. Prove theorem 2.3.3. [Suggestion: use the proof of theorem 2.3.2 as pattern.]

3. Take for granted the existence of a tensor $R$ satisfying

$$R(U,V)W = \Big(\frac{\nabla}{\partial\sigma}\frac{\nabla X}{\partial\tau} - \frac{\nabla}{\partial\tau}\frac{\nabla X}{\partial\sigma}\Big)_{(\sigma_o,\tau_o)}, \quad R(U,V)W = R^i_{mkl} U^k V^l W^m \Big(\frac{\partial}{\partial x^i}\Big)_{p_o},$$

as asserted in theorem 2.3.3. Prove that

$$(\nabla_k \nabla_l - \nabla_l \nabla_k)\partial_m = -R^i_{mkl}\partial_i.$$

[Do not use formula ($R$).]

4. Suppose parallel transport $T_{p_o}M \to T_{p_1}M$ is independent of the path from $p_o$ to $p_1$ (for all $p_o, p_1 \in M$). Prove that $R = 0$.

5. Consider this statement in the proof of theorem 2.3.6. *Assume $R = 0$, i.e. the parallel transport around an infinitesimal rectangle is the identity transformation on vectors. The same is then true for any finite rectangle, since it can be subdivided into arbitrarily small ones.* Write this statement as an equation for the parallel transport as a limit and explain why that limit is the identity transformation as stated.

## 2.4 Gauss curvature

Let $S$ be a smooth surface in Euclidean three–space $\mathbb{R}^3$, i.e. a two-dimensional submanifold. For $p \in S$, let $N = N(p)$ be one of the two unit normal vectors to $T_pS$. (It does not matter which one, but we do assume that the choice is consistent, so that $N(p)$ depends continuously on $p \in S$. This is always possible locally, i.e. in sufficiently small regions of $S$, but not necessarily globally.) Since $N(p)$ is a unit vector, one can view $p \to N(p)$ as a map from the surface $S$ to the unit sphere $S^2$. This is the *Gauss map* of the surface. To visualize it, imagine you move on the surface holding a copy of $S^2$ which keeps its orientation with respect to the surrounding space (no rotation, pure translation), equipped with a compass needle $N(p)$ always pointing perpendicular of the surface. The variation of the needle on the sphere reflects the curvature of the surface, the unevenness in the terrain (but can only offer spiritual guidance if you are lost on the surface.)

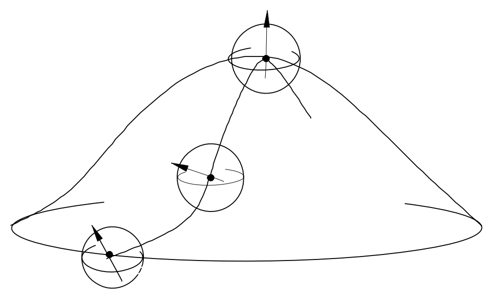


Fig.*1. The Gauss map*

*In the following lemma and throughout this section vectors in $\mathbb{R}^3$ will be denoted by capital letters like $U, V$ and their scalar product by $U \cdot V$. To simplify the notation we write $N = N(p)$ for the unit normal at $p$ and $dNU = dN_p(U)$ for its differential applied to the vector $U$ in subspace $T_pS$ of $\mathbb{R}^3$.*

2.4.1 **Lemma**. *$T_pS = N^\perp = T_N S^2$ as subspace of $\mathbb{R}^3$ and the differential $dN$ is a self-adjoint linear transformation of this 2–dimensional space, i.e. $dNU \cdot V = U \cdot dNV$ for all $U, V \in T_pS$.*

*Proof.* That $T_pS = N^\perp = \{U \in T_p\mathbb{R}^3 = \mathbb{R}^3 \mid U \cdot N = 0\}$ is clear from the definition of $N$, and $T_NS^2 = N^\perp$ is clear as well. Let $p = p(\sigma, \tau)$ be any smooth map $\mathbb{R}^2 \to S$. Then $\partial p/\partial\sigma \cdot N = 0$ and by differentiation

$$\frac{\partial^2 p}{\partial\tau\partial\sigma} \cdot N + \frac{\partial p}{\partial\sigma} \cdot \frac{\partial N}{\partial\tau} = 0 \quad \text{i.e.} \quad \frac{\partial p}{\partial\sigma} \cdot dN\frac{\partial p}{\partial\tau} = -\frac{\partial^2 p}{\partial\tau\partial\sigma} \cdot N \tag{1}$$

Since the right side is symmetric in $\sigma, \tau$, the equation $dNU \cdot V = U \cdot dNV$ holds for $p = p(\sigma, \tau)$, $U = \partial p/\partial\sigma$, $V = \partial p/\partial\tau$ and hence for all $p \in S$ and $U, V \in T_pS$. □

Since $T_pS = T_NS^2$ as subspace of $\mathbb{R}^3$ we can consider $dN : T_pS \to T_NS^2$ also as a linear transformation of the tangent plane $T_pS$. Part (c) shows that this linear transformation is self-adjoint, i.e. the bilinear from $U \cdot dNV$ on $T_pS$ is symmetric. The negative $-U \cdot dNV$ of this bilinear from is called the *second fundamental form* of $S$, a term also applied to the corresponding quadratic form $-dNU \cdot U$. We shall denote it $\Pi$ (mimicking the traditional symbol $II$, which looks strange when equipped with indices):

$$\Pi(U, V) = -U \cdot dNV.$$

Incidentally, *first fundamental form* of $S$ is just another name for scalar product $U \cdot V$ or the corresponding quadratic form $U \cdot U$, which is just the Riemann metric $ds^2$ on $S$ induced by the Euclidean metric on $E$.

The connection $\nabla^S = \nabla$ on $S$ induced by the standard connection $\nabla^E = D$ on $E$ is given by

$$\nabla_Y X = \text{tangential component of } D_Y X,$$
$$\nabla_Y X = D_Y X - (D_Y X \cdot N)N.$$

The formula $A=(A-(A\cdot N)N)+(A\cdot N)N$ for the decomposition of a a vector $A$ into its components perpendicular and parallel to a unit vector $N$ has been used.

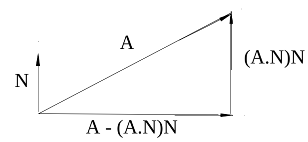



Fig.2. *The Gauss map*

We shall calculate the torsion and the curvature of this connection. For this purpose, let $p=p(\sigma,\tau)$ be a parametrization of $S$, $\partial p/\partial\sigma$ and $\partial p/\partial\tau$ the corresponding tangent vector fields, and $W=W(\sigma,\tau)$ any other vector field on $S$. We have

$$T\Big(\frac{\partial p}{\partial\sigma},\ \frac{\partial p}{\partial\tau}\Big)=\frac{\nabla}{\partial\tau}\frac{\partial p}{\partial\sigma}-\frac{\nabla}{\partial\sigma}\frac{\partial p}{\partial\tau},$$

$$R\Big(\frac{\partial p}{\partial\sigma},\ \frac{\partial p}{\partial\tau}\Big)=\frac{\nabla}{\partial\tau}\frac{\nabla W}{\partial\sigma}-\frac{\nabla}{\partial\sigma}\frac{\nabla W}{\partial\tau}..$$

The first equation gives

$$T\Big(\frac{\partial p}{\partial\sigma},\ \frac{\partial p}{\partial\tau}\Big)=\Big(\textit{tang. comp. of }\ \frac{\partial}{\partial\tau}\frac{\partial p}{\partial\sigma}-\frac{\partial}{\partial\sigma}\frac{\partial p}{\partial\tau}\Big)=0,$$

To work out the second equation, calculate

$$\begin{aligned}\frac{\nabla}{\partial\tau}\frac{\nabla W}{\partial\sigma}&=\frac{\nabla}{\partial\tau}\Big(\frac{\partial W}{\partial\sigma}-(\frac{\partial W}{\partial\sigma}\cdot N)N\Big)\\&=\textit{tang. comp. of }\ \frac{\partial^2W}{\partial\tau\partial\sigma}-\Big(\frac{\partial}{\partial\tau}(\frac{\partial W}{\partial\sigma}\cdot N)\Big)N-(\frac{\partial W}{\partial\sigma}\cdot N)\frac{\partial N}{\partial\tau}\end{aligned}$$

The second term may be omitted because it is orthogonal to $S$. The last term is already tangential to $S$, because $0=\partial(N\cdot N)/\partial\tau=2(\partial N/\partial\tau)\cdot N$ Thus

$$\frac{\nabla}{\partial\tau}\frac{\nabla W}{\partial\sigma}=\Big(\textit{tang. comp. of }\ \frac{\partial^2W}{\partial\tau\partial\sigma}\Big)-(\frac{\partial W}{\partial\sigma}\cdot N)\frac{\partial N}{\partial\tau}$$

In this equation, interchange $\sigma$ an $\tau$ and subtract to find

$$R(\frac{\partial p}{\partial\sigma},\frac{\partial p}{\partial\tau})W=(\frac{\partial W}{\partial\tau}\cdot N)\frac{\partial N}{\partial\sigma}-(\frac{\partial W}{\partial\sigma}\cdot N)\frac{\partial N}{\partial\tau}.$$

These formulas can be rewritten as follows. Since $W\cdot N=0$ one finds by differentiation that

$$\frac{\partial W}{\partial\tau}\cdot N+W\cdot\frac{\partial N}{\partial\tau}=0.$$

and similarly with $\tau$ replaced by $\sigma$. Hence

$$R(\frac{\partial p}{\partial\sigma},\frac{\partial p}{\partial\tau})W=(W\cdot\frac{\partial N}{\partial\sigma})\frac{\partial N}{\partial\tau}-(W\cdot\frac{\partial N}{\partial\tau})\frac{\partial N}{\partial\sigma}.$$

If we consider $p \to N(p)$ as a function on $S$ with values in $S^2 \subset \mathbb{R}^3$, and write $U, V, W$ for three tangent vectors at $p$ on $S$, this equation becomes

$$R(U,V)W = (W \cdot dNU)\, dNV - (W \cdot dNV)\, dNU.$$

Hence

$$(R(U,V)W, Z) = (W \cdot dNU)\,(dNV \cdot Z) - (W \cdot dNV)(dNU \cdot Z). \tag{3}$$

Choose an orthonormal basis of the 2 dimensional vector space $T_pS$ to represents its elements $V$ as column 2–vectors $[V]$ and the linear transformation $dN$ as a $2 \times 2$ matrix $dN$. The scalar product $V \cdot W$ on $T_pS$ is then the matrix product $[V]^*[W]$ with the transpose $[V]^*$. Then (3) can be written as

$$\begin{aligned}(R(U,V)W, Z) &= \det \begin{bmatrix} W \cdot dNU & W \cdot dNV \\ Z \cdot dNU & Z \cdot dNV \end{bmatrix} \\ &= \det\{[W,Z]^*[dN][U,V]\} \\ &= \det\{[W,Z]^* \det[dN] \det[U,V]\} \\ &= \det[dN] \det\{[W,Z]^*[U,V]\} \\ &= \det dN \det \begin{bmatrix} W \cdot U & W \cdot V \\ Z \cdot U & Z \cdot V \end{bmatrix}\end{aligned}$$

Thus

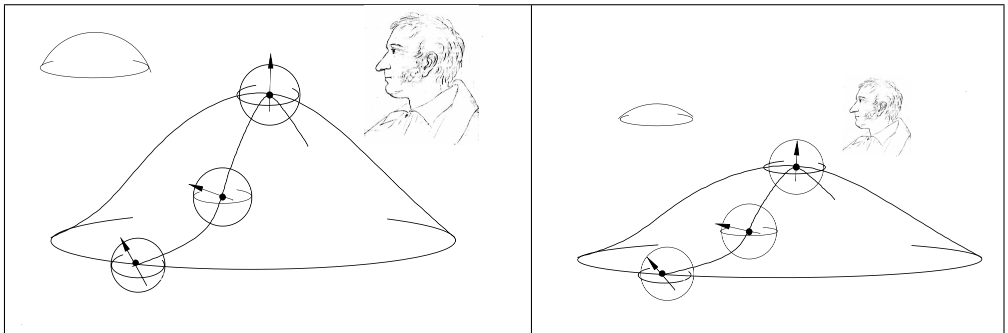
$$(R(U,V)W, Z) = K \det \begin{bmatrix} W \cdot U & W \cdot V \\ Z \cdot U & Z \cdot V \end{bmatrix}, \quad K := \det dN. \tag{5}$$

The scalar function $K = K(p)$ is called the *Gauss curvature* of $S$. The equation (5) shows that for a surface $S$ the curvature tensor $R$ is essentially determined by $K$. In connection with this formula it is important to remember that $N$ is to be considered as a map $N : S \to S^2$ and $dN$ as a linear transformation of the 2–dimensional subspace $T_pS = T_N S^2$of $\mathbb{R}^3$.

One can give a geometric interpretation of $K$ as follows. The multiplication rule for determinants implies that $[dNU, dNV] = \det[dN] \det[U, V]$, i.e.

$$K = \frac{\det[dNU, dNV]}{\det[U,V]}. \tag{6}$$

Up to sign, $\det[U, V]$ is the area of the parallelogram in $T_pS$ spanned by the vectors $U$ and $V$ and $\det[dNU, dNV]$ the area of the parallelogram in $T_N S^2$ spanned by the image vectors $dNU$ and $dNV$ under the differential of the *Gauss map* $N : S \to S^2$. So $K$ the ratio of the areas of two infinitesimal parallelograms, one on $S$, the other its image on $S^2$. One might say that $K$ *is the "amount" of variation of the normal on $S$ per unit area*. The sign $\pm$ in (6) indicates whether the Gauss map preserves or reverses the sense of rotation around the common normal direction. (The normal directions at point on $S$ and at its image on $S^2$ are specified by the same vector $N$, so that the sense of rotation around it can be compared, for example by considering a small loop on $S$ and its image on $S^2$.)



Large curvature, large Gauss image | Small curvature, small Gauss image

Fig. 3

**Remarks on the computation of the Gauss curvature.** For the purpose of computation one can start with the formula

$$K = \det dN = \frac{\det(dNV_i \cdot V_j)}{\det(V_i \cdot V_j)} \tag{8}$$

which holds for any two $V_1, V_2 \in T_pS$ for which the denominator is nonzero. If $M$ is any nonzero normal vector, then we can write $N = \mu M$ where $\mu = \|M\|^{-1}$. Since $dN = \mu(dM) + (d\mu)M$ formula (8) gives

$$K = \frac{\det(dMV_i \cdot V_j)}{(M \cdot M)\det(V_i \cdot V_j)} \tag{9}$$

Suppose $S$ is given in parametrized form $S$: $p = p(t^1, t^2)$. Take for $V_i$ the vector field $p_i = \partial p/\partial t^i$ and write $M_i = \partial M/\partial t^i$ for the componentwise derivative of $M$ in $\mathbb{R}^3$. Then (9) becomes

$$K = \frac{\det(M_i \cdot p_j)}{(M \cdot M)\det(p_i \cdot p_j)} \tag{10}$$

Formula (1) implies $dNp_i \cdot p_j = -p_{ij} \cdot N$ where $p_{ij} = \partial^2 p/\partial t^i \partial t^j$, hence also $dMp_i \cdot p_j = -p_{ij} \cdot M$. Thus (10) can be written as

$$K = \frac{\det(p_{ij} \cdot M)}{(M \cdot M)\det(p_i \cdot p_j)} \tag{11}$$

Take $M = p_1 \times p_2$, the cross product. Since $\|p_1 \times p_2\|^2 = \det(p_i \cdot p_j)$, the equation (11) can be written as

$$K = \frac{\det[p_{ij} \cdot (p_1 \times p_2)]}{(\det(p_i \cdot p_j))^2}. \tag{12}$$

This can be expressed in terms of the fundamental forms as follows. Write

$$ds^2 = g_{ij}dt^i dt^j, \Pi = \Pi_{ij}dt^i dt^j.$$

Then

$$\Pi_{ij} = \Pi(p_i, p_j) = -dNp_i \cdot p_j = \frac{p_{ij} \cdot (p_1 \times p_2)}{\|p_1 \times p_2\|}, \quad g_{ij} = p_i \cdot p_j$$

Hence (11) says

$$K = \frac{\det \Pi_{ij}}{\det g_{ij}}.$$

2.4.3 **Example: the sphere**. Let $S = \{p \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = r^2\}$ be the sphere of radius $r$. (We do not take $r = 1$ in order to see how the curvature depends on $r$.) We may take $N = r^{-1}p$ when $p$ is considered a vector. For any coordinate system on $S^2$, the formula (10) becomes

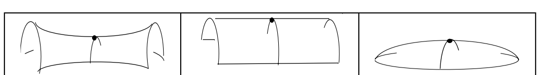
$$K = \frac{\det\left(r^{-1}p_i \cdot p_j\right)}{\det(p_i \cdot p_j)} = r^{-2}$$

So the sphere of radius $r$ has *constant* Gauss-curvature $K = 1/r^2$.

2.4.4 **Example.** Let $S$ be the surface with equation $ax^2+by^2-z = 0$. A normal vector is the gradient of the defining function, i.e. $M = (2ax, 2by, -1)$. As coordinates on $S$ we take $(t^1, t^2) = (x, y)$ and write$p = p(x, y) = (x, y, ax^2+by^2)$. By formula (10),

$$\begin{aligned} K &= \frac{1}{M \cdot M} \frac{(M_x \cdot p_x)(M_y \cdot p_y) - (M_x \cdot p_y)(M_y \cdot p_x)}{(p_x \cdot p_x)(p_y \cdot p_y) - (p_x \cdot p_y)(p_y \cdot p_x)} \\ &= \frac{1}{(4a^2x^2 + 4b^2y^2 + 1)} \frac{4ab - 0}{(1 + (2ax)^2)(1 + (2by)^2) - (2ax2by)^2} \\ &= \frac{4ab}{(4a^2x^2 + 4b^2y^2 + 1)^2} \end{aligned}$$

The picture shows how sign of $K$ reflects the shape of the surface.



| | | |
|---|---|---|
| $K < 0$ | $K = 0$ | $K > 0$ |

Fig 4. The sign of $K$

**Remarks on the curvature plane curves and the curvature of surfaces**.

Gauss's theory of surfaces in Euclidean 3 space has an exact analog for curves in the Euclidean plane, known from the early days of calculus and analytic geometry. Let $C$ be a smooth curve in $\mathbb{R}^2$, i.e. a one-dimensional submanifold. Let $N = N(p)$ be one two unit normals to $T_p$, chosen so as to vary continuously with $p$. It provides a smooth map from $C$ to the unit circle $S^1$, which we call again the *Gauss map* of the curve.

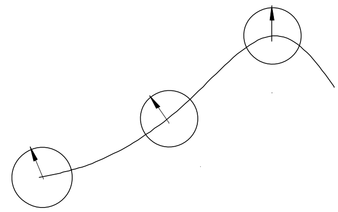


Fig.5. Curvature of a plane curve

We have $T_pC = N^\perp = T_pS^1$ as subspace of $\mathbb{R}^2$ so that $dN$ is a linear transformation of this one-dimensional vector space, i.e. $dNU = \kappa U$ for some scalar $\kappa = \kappa(p)$, which is evidently the exact analog of the Gauss curvature of a surface.

If we take $U = \dot{p}$ for any $p = p(t)$ is any parametrization of $C$, then the equation $dNU = \kappa U$ becomes $dN/dt = \kappa dp/dt$. In particular, if $p = p(s)$ is a parametrization by arclength, so that $\|dp/ds\| = 1$, then $\kappa = \pm\|dN/ds\|$. (The sign depends on the sense of traversal of the curve and is not important here). This formula can be written as $\kappa = \pm d\theta/ds$ in terms of the arclength functions $s$ on $C$ and $\theta$ on $S^1$ taken from some initial points. It shows that $\kappa$ is the amount of variation of the normal $N$ per unit length, again in complete analogy with the Gauss curvature of a surface.

From this point of view it is very surprising that there is a fundamental difference between $K$ and $\kappa$: $K$ is independent of the embedding of the surface $S$ in $\mathbb{R}^3$, but $\kappa$ is very much dependent on the embedding of the curve in $\mathbb{R}^2$. The curvature $\kappa$ of a curve $C$ should therefore not be confused with the Riemann curvature tensor $R$ of the connection on $C$ induced by its embedding in the Euclidean plane $\mathbb{R}^2$, which is identically zero: the connection depends only on the induced Riemann metric, which is locally isometric to the Euclidean line via parametrization by arclength. For emphasis we call $\kappa$ the *relative curvature* of $C$, because of its dependence on the embedding of $C$ in $\mathbb{R}^2$.

Nevertheless, there is a relation between the Riemann or Gauss curvature of a surface $S$ and the relative curvature of certain curves. Return to the surface $S$ in $\mathbb{R}^3$. Fix a point $p$ on $S$ and consider the 2-planes though $p$ containing the normal line of $S$ through $p$.

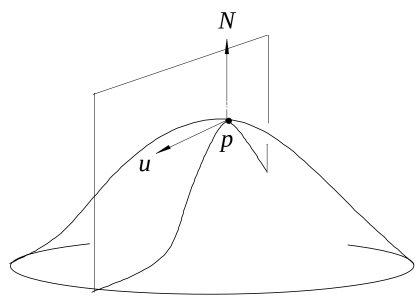



Fig.6. A normal section

Each such plane is specified by a unit vector $U$ in $T_pS$, $p$ and $N$ being fixed. Let $C_U$ be the curve of the intersection of $S$ with such a 2-plane. The normal $N$ of $S$ servers also as normal along $C_U$ and the relative curvature of $C_U$ at $p$ is $\kappa(U) = (dNU \cdot U)$ since $U$ is by construction a unit normal in $T_pC_U$. The curves $C_U$ are called *normal sections* of $S$ and $\kappa(U)$ the *normal curvature* of $S$ in the direction $U \in T_pS$.

Now consider $dN$ again as linear transformation of $T_pS$. It is a self-adjoint linear transformation of the two dimensional vector space $T_pS$, hence has two real eigenvalues, say $\kappa_1, \kappa_2$. These are called the *principal curvatures* of $S$ at $p$. The corresponding eigenvectors $U_1, U_2$ give the *principal directions* of $S$ at $p$. They are always orthogonal to each other, being eigenvectors of a self-adjoint linear transformation (assuming the two eigenvalues are distinct). The Gauss curvature is $K = \det dN = \kappa_1\kappa_2$, since the determinant of a matrix is the product of its eigenvalues.

2.4.2 **Lemma**. *The principal curvatures $\kappa_1, \kappa_2$ are the maximum and the minimum values of $\kappa(U)$ as $U$ varies over all unit-vectors in $T_pS$.*

**Proof**. For fixed $p = p_o$ on $S$, consider $\kappa(U) = dNU \cdot U$ as a function on the unit-circle $U \cdot U = 1$ in $T_{p_o}S$. Suppose $U = U_o$ is a minimum or a maximum of $\kappa(U)$. Then for any parametrization $U = U(t)$ of $U \cdot U = 1$ with $U(t_o) = U_o$ and $\dot{U}(t_o) = V_o$:

$$0 = \frac{d}{dt}(dN(U) \cdot U)\bigg|_{t=t_o} = dNV_o \cdot U_o + dNU_o \cdot V_o = 2\, dNU_o \cdot V_o.$$

Hence $dNU_o$ must be orthogonal to the tangent vector $V_o$ to the circle at $U_o$, i.e. $dNU_o = \kappa_o U_o$ for some scalar $\kappa_o$. Thus the maximum and minimum values of $\kappa(U)$ are eigenvalue of $dN$, hence *the* two eigenvalues unless $\kappa(U)$ is constant and $dN$ a scalar matrix. □

To bring out the analogy with surfaces in space we considered only in the plane rather than curves in space. For a space curve $C$ (one-dimensional submanifold of $\mathbb{R}^3$) the situation is different, because the subspace of $\mathbb{R}^3$ orthogonal to $T_pC$

now has dimension two, so that one can choose two independent normal vectors at each point. A particular choice may be made as follows. Let $p = p(s)$ be a parametrization of $C$ by arclength, so that the tangent vector $T := dp/ds$ is a unit vector. By differentiation of the relations $T \cdot T = 1$ one finds $T \cdot dT/ds = 0$, i,e, $dT/ds$ is orthogonal to $T$, as is its cross-product with $T$ itself. Normalizing these vectors one obtains three orthonormal vectors $T, N, B$ at each point of $C$ defined by the *Frenet formulas*

$$\frac{dT}{ds} = \kappa N, \quad \frac{dN}{ds} = \kappa T + \tau B, \quad \frac{dB}{ds} = -\tau.$$

The scalar functions $\kappa$ and $\tau$ are called curvature and torsion of the space curve (not to be confused with the Riemann curvature and torsion of the induced connection, which are both identically zero).

There is a vast classical literature on the differential geometry of curves and surfaces. Hilbert and Cohn-Vossen (1932) offer a casual tour through some parts of this subject, which can an overview of the area. But we shall go no further here.

EXERCISES 2.4

*In the following problems*[2] *$S$ is a surface in $\mathbb{R}^3$, $(\sigma, \tau)$ a coordinate system on $S$, $N$ a unit-normal.*

1. a)Let $(t^1, t^2)$ be a coordinate system on $S$. Show that all components $R_{imkl} = g_{ij} R^j_{mkl}$ of the Riemann curvature tensor are zero except

$$R_{1212} = -R_{2112} = -R_{1221} = R_{2121}$$

and that this component is given by

$$R_{1212} = K(g_{11} g_{22} - g_{12}^2).$$

in terms of the Gauss curvature $K$ and the metric $g_{ij}$. [Suggestion. Use formula (5)]
b)Find the components $R^j_{mkl}$.
c)Write out the $R^i_{mkl}$ for the sphere $S = \{p = (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = r^2\}$ using the coordinates $(\phi, \theta)$ on $S$ by $x = r \sin\phi \cos\theta$, $y = r \sin\phi \sin\theta$, $z = r \cos\phi$. [You may use Example 2.4.5.]

2. Determine if the following statements are true or false. (Provide a proof or a counterexample).
a) If $\Pi = 0$ everywhere, then $S$ is a plane or a part of a plane.
b) If $K = 0$ everywhere, then $S$ is a plane or a part of a plane.

3. Suppose $S$ is given as a graph $z = f(x, y)$.

[2] Several of the exercises are taken from the *Recueil d'exercises de géométrie différentielle*, by A. Fedenko et al., Editions Mir, Moskow, 1982.

a) Show that the second fundamental form is

$$\text{II} = \frac{1}{\sqrt{1 + f_x^2 + f_y^2}}\,(f_{xx}^2 dx^2 + f_{yy}^2 dy^2 + 2f_{xy}dxdy).$$

b) Show that the Gauss curvature is

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}\,.$$

4. Let $S$ be a surface of revolution about the $z$-axis, with equation $r = f(z)$ in cylindrical coordinates $(r, \theta, z)$. (Assume $f > 0$ everywhere.) Using $(\theta, z)$ as coordinates on $S$, show that the second fundamental form is

$$\text{II} = \frac{1}{\sqrt{f_z^2 + 1}}(f d\theta^2 - f_{zz}dz^2)$$

(The subscripts $z$ indicate derivatives with respect to $z$).

5. Let $S$ be given by an equation $f(x, y, z) = 0$. Show that the Gauss curvature of $S$ is

$$K = \frac{-1}{(f_x^2 + f_y^2 + f_z^2)^2}\ \det \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} & f_x \\ f_{yx} & f_{yy} & f_{yz} & f_y \\ f_{zx} & f_{zy} & f_{zz} & f_z \\ f_x & f_y & f_z & 0 \end{bmatrix}.$$

The subscripts denote partial derivatives. [Suggestion. Let $M = (f_x, f_y, f_z)$, $\mu = (f_x^2 + f_y^2 + f_z^2)^{-1/2}$, $N = \mu M$. Consider $M$ as a column and $dM$ as a $3 \times 3$ matrix. Show that the equation to be proved is equivalent to

$$K = -\det \begin{bmatrix} \mu dM & N \\ N^* & 0 \end{bmatrix}.$$

Split the vectors $V \in \mathbb{R}^3$ on which $\mu dM$ operates as $V = T + bN$ with $T \in T_p\, S$ and decompose $\mu dM$ accordingly as a block matrix. You will need to remember elementary column operations on determinants.]

6. Let $C$:$p = p(\sigma)$ curve in $\mathbb{R}^3$. The *tangential developable* is the surface $S$ in $\mathbb{R}^3$ swept out by the tangent line of $C$, i.e. $S$ consists of the points $p = p(\sigma, \tau)$ given by

$$p = p(\sigma) + \tau\dot{p}(\sigma).$$

Show that the principal curvatures of $S$ are

$$k_1 = 0,\ k_2 = \frac{\tau}{\tau\kappa}$$

where $\kappa$ and $\tau$ are the curvature and the torsion of the curve $C$.

7. Calculate the first fundamental form $ds^2$, the second fundamental form $\Pi$, and the Gauss curvature $K$ for the following surfaces $S$: $p = p(\sigma,\tau)$ using the given parameters $\sigma,\tau$ as coordinates. The letters $a,b,c$ denote positive constants.
a) Ellipsoid of revolution: $x = a\cos(\sigma)\cos(\tau)$, $y = a\cos(\sigma)\sin(\tau)$, $z = c\sin(\sigma)$.
b) Hyperboloid of revolution of one sheet: $x = a\cosh\sigma\cosh\tau$, $y = a\cosh\sigma\sin\tau$, $z = c\sinh\sigma$.
c) Hyperboloid of revolution of two sheets: $x = a\sinh\sigma\cosh\tau$, $y = a\sinh\sigma\sin\tau$, $z = c\cosh\sigma$.
d) Paraboloid of revolution: $x = \sigma\cos\tau$, $y = \sigma\sin\tau$, $z = \sigma^2$.
e) Circular cylinder: $x = R\cos\tau$, $y = R\sin\tau, z = \sigma$.
f) Circular cone without vertex: $x = \sigma\cos\tau$, $y = \sigma\sin\tau, z = a\sigma$ $(\sigma \neq 0)$.
g) Torus: $x = (a + b\cos\sigma)\cos\tau$, $y = (a + b\cos\sigma)\sin\tau$, $z = b\sin\sigma$.
h) Catenoid: $x = \cosh(\sigma/a)\cos\tau$, $y = \cosh(\sigma/a)\sin\tau$, $z = \sigma$.
i) Helicoid: $x = \sigma\cos\tau$, $y = \sigma\sin\tau$, $z = a\tau$.

8. Find the principal curvatures $k_1, k_2$ at the points $(\pm a, 0, 0)$ of the hyperboloid of two sheets

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

## 2.5 Levi-Civita's connection

We have discussed two different pieces of geometric structure a manifold $M$ may have, a Riemann metric $g$ and a connection $\nabla$; they were introduced quite independently of each other, although submanifolds of a Euclidean space turned out come equipped with both in a natural way. If fact, it turns out that any Riemann metric automatically gives rise to a connection in a natural way, as we shall now explain. Throughout, $M$ denotes a manifold with a Riemann metric denoted $g$ or $ds^2$. We shall now often use the term *connection for covariant derivative.*

2.5.1 **Definition**. A connection $\nabla$ is said to be compatible with the Riemann metric $g$ if parallel transport (defined by $\nabla$) along any curve preserves scalar products (defined by $g$): for any curve $p = p(t)$

$$g(T(t_o \to t_1)u, T(t_o \to t_1)v)) = g(u,v) \tag{1}$$

for all $u, v \in T_{p(t_o)}M$, any $t_o, t_1$.
Actually it suffices that parallel transport preserve the quadratic form (square-length) $ds^2$:

2.5.2 **Remark**. A linear transformation $T : T_{p_o}M \to T_{p_1}M$ which preserves the quadratic form $ds^2(u) = g(u,u)$ also preserves the scalar product $g(u,v)$, as follows immediately from the equation $g(u+v, u+v) = g(u,u) + 2g(u,v) + g(v,v))$.

2.5.3 **Theorem**. *A connection $\nabla$ is compatible with the Riemann metric $g$ if and only if for any two vector fields $X(t)$, $Y(t)$ along a curve c: $p = p(t)$*

$$\frac{d}{dt}g(X,Y) = g\Big(\frac{\nabla X}{dt}, Y\Big) + g\Big(X, \frac{\nabla Y}{dt}\Big). \tag{2}$$

**Proof**. We have to show (1) $\Leftrightarrow$ (2).

($\Rightarrow$) Assume (1) holds. Let $E_1(t), \cdots, E_n(t)$ be a parallel frame along $p(t)$. Write $g(E_i(t_o), E_j(t_o)) = c_{ij}$. Since $E_i(t) = T_c(t_o \to t)E_i(t_o)$(1) gives $g(E_i(t), E_j(t)) = c_{ij}$ for all $t$. Writ $X(t) = X^i(t)E_i(t)$, $Y(t) = Y^i(t)E_i(t)$. Then

$$g(X,Y) = X^iY^jg(E_i, E_j) = c_{ij}X^iY^i.$$

and by differentiation

$$\frac{d}{dt}g(X,Y) = c_{ij}\frac{dX^i}{dt}Y^j + c_{ij}X^i\frac{dY^{ji}}{dt} = g\Big(\frac{\nabla X}{dt}, Y\Big) + g\Big(X, \frac{\nabla Y}{dt}\Big).$$

This proves (2).
($\Leftarrow$) Assume (2) holds. Let c: $p = p(t)$ be any curve, $p_o = p(t_o)$, and $u, v \in T_{p(t_o)}M$ two vectors. Let $X(t) = T(t_o \to t)u$ and $Y(t) = T(t_o \to t)v$. Then $\nabla X/dt = 0$ and $\nabla Y/dt = 0$. Thus by (2)

$$\frac{d}{dt}g(X,Y) = g\Big(\frac{\nabla X}{dt}, Y\Big) + g\Big(X, \frac{\nabla Y}{dt}\Big) = 0.$$

Consequently $g(X,Y)$ =constant. If we compare this constant at $t = t_o$ and at $t = t_1$ we get $g(X(t_o), Y(t_o)) = g(X(t_1), Y(t_1))$, which is the desired equation (1). □

2.5.4 **Corollary**. *A connection $\nabla$ is compatible with the Riemann metric $g$ if and only if for any three* smooth *vector fields $X, Y, Z$*

$$D_Zg(X,Y) = g(\nabla_ZX, Y) + g(X, \nabla_ZY) \tag{3}$$

*where generally $D_Zf = df(Z)$.*

**Proof**. At a given point $p_o$, both sides of (3) depend only on the value $Z(p_o)$ of $Z$ at $p_o$ and on values of $X$ and $Y$ along some curve c: $p = p(t)$ with $p(t_o) = p_o$, $p'(t_o) = Z(p_o)$. If we apply (2) to such a curve and set $t = t_o$ we find that (3) holds at $p_o$. Since $p_o$ is arbitrary, (3) holds at all points. □

2.5.5 **Theorem (Levi-Civita)**. *For any Riemannian metric $g$ there is a unique symmetric connection $\nabla$ that is compatible with $g$.*

**Proof**. (Uniqueness) Assume $\nabla$ is symmetric and compatible with $g$. Fix a coordinate system $(x^i)$. From (3), $\partial_ig(\partial_j, \partial_k) = g(\nabla_i\partial_j, \partial_k) + g(\partial_j, \nabla_i\partial_k)$. This may be written as

$$g_{jk,i} = \Gamma_{k,ji} + \Gamma_{j,ki} \tag{4}$$

where $g_{jk,i} = \partial_i g_{jk} = \partial_i g(\partial_j, \partial_k)$ and $\Gamma_{k,ji} = g(\nabla_i \partial_j, \partial_k) = g(\Gamma^l_{ji}\partial_l, \partial_k)$ i.e.

$$\Gamma_{k,ji} = g_{kl}\Gamma^l_{ji}. \tag{5}$$

Since $\nabla$ is symmetric $\Gamma^l_{ji} = \Gamma^l_{ij}$ and therefore

$$\Gamma_{k,ji} = \Gamma_{k,ij}. \tag{6}$$

Equations (4) and (6) may be solved for the $\Gamma$'s as follows. Write out the equations corresponding to (4) with the cyclic permutation of the indices: $(ijk)$, $(kij)$, $(jki)$. Add the last two and subtract the first. This gives (together with (6)):

$$g_{ki,j} + g_{ij,k} - g_{jk,i} = (\Gamma_{i,kj} + \Gamma_{k,ij}) + (\Gamma_{\mathrm{j},ik} + \Gamma_{i,jk}) - (\Gamma_{k,ji} + \Gamma_{j,ki}) = 2\Gamma_{i,jk}.$$

Therefore

$$\Gamma_{i,kj} = \frac{1}{2}(g_{ki,j} + g_{ij,k} - g_{jk,i}). \tag{7}$$

Use the inverse matrix $(g^{ab})$ of $(g_{ij})$ to solve (5) and (7) for $\Gamma^l_{jk}$:

$$\Gamma^l_{jk} = \frac{1}{2}g^{li}(g_{ki,j} + g_{ij,k} - g_{jk,i}). \tag{8}$$

This proves that the $\Gamma$'s, and therefore the connection $\nabla$ are uniquely determined by the $g_{ij}$, i.e. by the metric $g$.
(Existence). Given $g$, choose a coordinate system $(x^i)$ and define $\nabla$ to be the (symmetric) connection which has $\Gamma^l_{jk}$ given by (8) relative to this coordinate system. If one defines $\Gamma_{k,ji}$ by (5) one checks that (4) holds. This implies that (3) holds, since any vector field is a linear combination of the $\partial_i$. So $\nabla$ is compatible with the metric $g$. That $\nabla$ is symmetric is clear from (8). □

2.5.6 **Definitions**. (a) The unique connection $\nabla$ compatible with a given Riemann metric $g$ is called the *Levi-Civita connection*[3] of $g$.
(b) The $\Gamma_{k,ij}$ and the $\Gamma^l_{jk}$ defined by (7) and (8) are called *Christoffel symbols* of the first and second kind, respectively. Sometimes the following notation is used (sometimes with other conventions concerning the positions of the entries in the symbols)

$$\begin{Bmatrix} l \\ jk \end{Bmatrix} = \Gamma^l_{jk} = \frac{1}{2}g^{li}(g_{ki,j} + g_{ij,k} - g_{jk,i}). \tag{9}$$

$$[jk,i] = \Gamma_{i,kj} = g_{il}\Gamma^l_{jk} = \frac{1}{2}(g_{ki,j} + g_{ij,k} - g_{jk,i}), \tag{10}$$

These equations are called *Christoffel's formulas.*

[3]Sometimes called "Riemannian connection". Riemann introduced many wonderful concepts, but "connection" is not one of them. That concept was introduced by Levi-Civita in 1917 as the tangential component connection on a surface in a Euclidian space and by Weyl in 1922 in general.

**Riemann metric and Levi-Civita connection on submanifolds**.

**Riemann metric and curvature**. The basic fact is this.

2.5.7 **Theorem (Riemann)**. *If the curvature tensor $R$ of the Levi-Civita connection of a Riemann metric is $R = 0$, then locally there is a coordinate system $(x^i)$ on $M$ such that $g_{ij} = \pm\delta_{ij}$, i.e. with respect to this coordinate system $(x^i)$, $ds^2$ becomes a pseudo-Euclidean metric*

$$ds^2 = \sum_i \pm(dx^i)^2.$$

*The coordinate system is unique up to a pseudo-Euclidean transformation: any other such coordinate system $(\tilde{x}^j)$ is of the form*

$$\tilde{x}^j = \tilde{x}^j_o + c^j_i x^i$$

*for some constants $\tilde{x}^j_o, c^j_i$ with $\sum_{lk} g_{lk} c^l_i c^k_j = g_{ij}\delta_{ij}$ .*

*This is a slight refinement of Riemann's theorem for connections, as* discussed at the end of §2.4. The transformations $\exp(tX)$ generated by parallel vector fields are now isometries, because of the compatibility of the Levi-Civita connection and the Riemann metric. If the basis $e_1, \cdots, e_n$ of $T_{p_o}M$ used there to introduce the coordinates $(x^i)$ is chosen to be orthonormal, then these coordinates provide a local isometry with Euclidean space. In his "Leçons" of 1925-1926 (§178), Cartan shows further an analogous situation prevails for Riemann metrics whose curvature is "constant", the Euclidean space then being replaced by a sphere or by a pseudo-sphere. –The "constant" curvature condition means that the scalar $K$ defined by Gauss's expression

$$(R(U,V)W,Z) = K \det \begin{bmatrix} W\cdot U & W\cdot V \\ Z\cdot U & Z\cdot V \end{bmatrix}$$

be a constant, i.e. independent of the vectors and of the point.

**Geodesics of the Levi-Civita connection**. We have two notions of *geodesic* on $M$: (1) the geodesics of the Riemann metric $g$, defined as "*shortest lines*", characterized by the differential equation

$$\frac{d}{dt}(g_{br}\frac{dx^b}{dt}) - \frac{1}{2}(g_{cd,r})\frac{dx^c}{dt}\frac{dx^d}{dt} = 0 \quad \text{(for all } r). \tag{11}$$

(2) the geodesics of the Levi-Civita connections $\nabla$, defined as "*straightest lines*", characterized by the differential equation

$$\frac{d^2x^k}{dt^2} + \Gamma^k_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt} = 0 \quad \text{(for all } k). \tag{12}$$

The following theorem says that these two notions of "geodesic" coincide, so that we can simply speak of "geodesics".

2.5.8 **Theorem**. *The geodesics of the Riemann metric are the same as the geodesics of the Levi-Civita connection.*

**Proof**. We have to show that (11) is equivalent to (12) if we set

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl}(g_{jl,i} + g_{li,j} - g_{ij,l}).$$

This follows by direct calculation. □

Incidentally, we know from Proposition 1.5.9 that a geodesics of the metric have constant speed, hence so do the geodesics of the Levi-Civita connection, but it is just as easy to see this directly, since the velocity vector of a geodesic is parallel and parallel transport preserves length.

The comparison of the equations for the geodesics as shortest lines and straightest lines leads to an algorithm for the calculation of the connection coefficients $\Gamma_{ij}^k$ in terms of the metric $ds^2 = g_{ij}dx^i dx^k$ which is often more convenient than Chritoffel's. For this purpose, recall that the Euler-Lagrange equation §1.5 (3) for the variational problem $\int_a^b L dt = \min$, $L := g_{ij}\dot{x}^i\dot{x}^j$ given involves the expression

$$\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^k} - \frac{\partial L}{\partial x^k}) = \frac{d}{dt}\Big(g_{ik}\frac{dx^i}{dt}\Big) - \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^i}{dt}\frac{dx^j}{dt}$$

which is

$$\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^k} - \frac{\partial L}{\partial x^k}) = \frac{1}{2} g_{kl}\frac{d^2x^l}{dt^2} + \Gamma_{k,ij}\dot{x}^i\dot{x}^j$$

according to the calculation omitted in the proof of the theorem. So the rule is this. Starting with $L := g_{ij}\dot{x}^i\dot{x}^j$, calculate $\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^k} - \frac{\partial L}{\partial x^k})$ by formal differentiation, and read off the coefficient in the term $\Gamma_{k,ij}\dot{x}^i\dot{x}^j$ of the resulting expression.

**Example**. The Euclidean metric in spherical coordinates on $\mathbb{R}^3$ is $ds^2 = d\rho^2 + \rho^2 d\phi^2 + \rho^2 \sin^2\phi d\theta^2$. Hence $L = \dot{\rho}^2 + \rho^2\,\dot{\phi}^2 + \rho^2\sin^2\phi\,\dot{\theta}^2$ and

$$\begin{aligned}
\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{\rho}} - \frac{\partial L}{\partial \rho}) &= \ddot{\rho}^2 - \rho\dot{\phi}^2 - \rho\sin^2\phi\,\dot{\theta}^2 \\
\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} - \frac{\partial L}{\partial \phi}) &= \rho^2\ddot{\phi} + 2\rho\dot{\rho}\dot{\phi} - \rho^2\sin\phi\cos\phi\,\dot{\theta}^2 \\
\frac{1}{2}(\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta}) &= \rho^2\sin^2\phi\ddot{\theta} + 2\rho\sin^2\phi\dot{\rho}\dot{\phi} + 2\rho^2\sin\phi\cos\phi\,\dot{\phi}\dot{\theta}
\end{aligned}$$

The non-zero connection coefficients are

$$\begin{aligned}
&\Gamma_{\rho,\phi\phi} = -\rho, \quad \Gamma_{\rho,\phi\theta} = -\rho\sin^2\phi \\
&\Gamma_{\phi,\rho\phi} = \rho, \quad \Gamma_{\phi,\theta\theta} = -\rho^2\sin\phi\cos\phi \\
&\Gamma_{\theta,\rho\phi} = \rho\sin^2\phi, \quad \Gamma_{\theta,\phi\theta} = \rho^2\sin\phi\cos\phi
\end{aligned}$$

Divide these three equations by by the coefficients $1, \rho^2, \rho^2\sin^2\phi$ of the diagonal metric to find

$$\Gamma^{\rho}_{\phi\phi} = -\rho^2, \quad \Gamma^{\rho}_{\phi\theta} = -\rho\sin^2\phi$$
$$\Gamma^{\phi}_{\rho\phi} = \frac{1}{\rho^2}, \quad \Gamma^{\phi}_{\theta\theta} = -\sin\phi\cos\phi$$
$$\Gamma^{\theta}_{\rho\phi} = \frac{1}{\rho}, \quad \Gamma^{\theta}_{\phi\theta} = \frac{\cos\phi}{\sin\phi}$$

**Riemannian submanifolds**.

2.5.9 **Definition**. Let $S$ be an $m$–dimensional submanifold of a manifold $M$ with a Riemann metric $g$. Assume the restriction $g^S$of $g$ to tangent vectors to $S$ is non-degenerate, i.e. if $\{v_1, \cdots, v_{\mathrm{m}}\}$ is a basis for $T_pS \subset T_pM$, then $\det g(v_i, v_j) \neq 0$. Then $g^S$ is called the *induced* Riemannian metric on $S$ and $S$ is called a *Riemannian submanifold* of $M, g$. This is always assumed if we speak of an induced Riemann metric $g_S$ on $S$. It is automatic if the metric $g$ is positive definite.

2.5.10 **Lemma**. *Let $S$ be an $m$-dimensional Riemannian submanifold of $M$ and $p \in S$. Every vector $v \in T_pM$ can be uniquely written in the form $v = v^S + v^N$ where $v^S$ is tangential to $S$ and $v^N$ is orthogonal to $S$, i.e.*

$$T_pM = T_pS \oplus T_p^{\perp}S$$

*where $T_p^{\perp}S = \{v \in T_pM : g(w, v) = 0$ for all $v \in T_pS\}$.*

**Proof**. Let $(e_1, \cdots, e_m, \cdots, e_n)$ be a basis for $T_pM$ whose first $m$ vectors form a basis for $T_pS$. Write $v = x^ie_i$. Then $v \in T_p^{\perp}S$ iff $g(e_1, e_i)x^i = 0, \cdots, g(e_m, e_i)x^i = 0$. This is a system of $m$ equations in the $n$ unknowns $x^i$ which is of rank $m$, because $\mathrm{dett}_{ij\leq m}g(e_i, e_j) \neq 0$. Hence it has $n-m$ linearly independent solutions and no nonzero solution belongs to $\mathrm{span}\{e_i \colon i \leq m\} = T_pS$. Hence i.e. $\dim T_p^{\perp} = n - m$, $\dim T_pS = m$, and $T_p^{\perp}\bigcap T_pS = 0$. This implies the assertion. □

The components in the splitting $v = v^S + v^N$ of the lemma are called the *tangential* and *normal* components of $v$, respectively.

2.5.11 **Theorem**. *Let $S$ be an $m$-dimensional submanifold of a manifold $M$ with a Riemannian metric $g$, $g^S$ the induced Riemann metric on $S$. Let $\nabla$ be the Levi-Civita connection of $g$, $\nabla^S$ the Levi-Civita connection of $g^S$. Then for any two vector-fields $X, Y$ on $S$,*

$$\nabla^S_Y X = \text{tangential component of } \nabla_Y X \tag{9}$$

**Proof.** First *define* a connection on $S$ by (9). To show that this connection is symmetric we use Theorem 2.3.2 and the notation introduced there:

$$\frac{\nabla^S}{\partial\sigma}\frac{p}{\partial\tau} - \frac{\nabla^S}{\partial\tau}\frac{p}{\partial\sigma} = \Big(\frac{\nabla}{\partial\sigma}\frac{p}{\partial\tau} - \frac{\nabla}{\partial\tau}\frac{p}{\partial\sigma}\Big)^S = 0$$

since the Levi-Civita connection $\nabla$ on $M$ is symmetric. It remains to show that $\nabla^S$ is compatible with $g^S$. For this we use Lemma 2.5.4. Let $X, Y, Z$ be three vector fields on $S$. Considered as vector fields on $M$ along $S$ they satisfy

$$D_Z g(X,Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y).$$

Since $X$, $Y$ are already tangential to $S$, this equation may be written as

$$D_Z g^S(X,Y) = g^S(\nabla^S_Z X, Y) + g^S(\mathrm{X}, \nabla^S_Z Y).$$

It follows that $\nabla^S$ is indeed the Levi-Civita connection of $g^S$. $\square$

**Example: surfaces in $\mathbb{R}^3$.** Let $S$ be a surface in $\mathbb{R}^3$. Recall that the Gauss curvature $K$ is defined by $K = \det dN$ where $N$ is a unit normal vector-field and $dN$ is considered a linear transformation of $T_pS$. It is related to the Riemann curvature $R$ of the connection $\nabla^S$ by the formula

$$(R(U,V)W,Z) = K \det \begin{bmatrix} W\cdot U & W\cdot V \\ Z\cdot U & Z\cdot V \end{bmatrix}$$

The theorem implies that $\nabla^S$, and hence $R$, is completely determined by the Riemann metric $g^S$ on $S$, which is Gauss's *theorema egregium.*

We shall now discuss some further properties of geodesics on a Riemann manifold. The main point is what is called *Gauss's Lemma*, which can take on any one of forms to be given below, but we start with some preliminaries.

2.5.12 **Definition**. Let $S$ be a non-degenerate submanifold of $M$. The set $NS \subset TM$ all normal vectors to $S$, is called the **normal bundle** of $S$:

$$NS = \{v \in T_pM : p \in S, v \in T_pS^{\perp}\}$$

where $T_pS^{\perp}$ is the orthogonal complement of $T_pS$ in $T_pM$ with respect to the metric on $M$.

The Riemannian submanifold $S$ will remain fixed from now on and $N$ will denote its normal bundle.

2.5.13 **Lemma** . *$N$ is an $n$-dimensional submanifold of the tangent bundle $TM$.*

**Proof**. In a neighbourhood of any point of $S$ we can find an orthonormal family $E_1, \cdots, E_n$ of vectors fields on $M$, so that at points $p$ in $S$ the first $m = \dim S$ of them form a basis for $T_pS$ (with the help of the Gram–Schmidt process, for example). Set $\xi^i(v) = g(E_i, v)$ as function on $TM$. We may also choose a coordinate system $(x^1, \cdots, x^n)$ so that $S$ is locally given by $x^{m+1} = 0, \cdots, x^n = 0$. The $2n$ function $(x^i, \xi^i)$ form a coordinate system on $TM$ so that $N$ is given by $x^{m+1} = 0, \cdots, x^n = 0$, $\xi^1 = 0, \cdots, \xi^m = 0$, as required by the definition of "submanifold". $\square$

The set $N$ comes equipped with a map $N \to S$ which maps a normal vector $v \in T_pS^{\perp}$ to its base-point $p \in S$. We can think of $S$ as a subset of $N$, the zero-section, which consists of the zero-vectors $0_p$ at points $p$ of $S$.

2.5.14 **Proposition**. *The geodesic spray* $\exp : N \to M$ *is locally bijective around any point of* $S$.

**Proof**. We use the inverse function theorem. For this we have to calculate the differential of exp along the zero section $S$ in $N$. Fix $p_o \in S$. Then $S$ and $T_{p_o}S^\perp \subset T_{p_o}M$ are two submanifolds of $N$ whose tangent spaces $T_{p_o}S$ and $T_{p_o}S^\perp$ are orthogonal complements in $T_{p_o}M$. ($g$ is nondegenerate on $T_{p_o}S$). We have to calculate $d\exp_{p_o}(w)$. We consider two cases.
(1) $w \in T_{p_o}S$ is the tangent vector of a curve $p(t)$ in $S$
(2) $w \in T_{p_o}S^\perp$ is the tangent vector of the straight line $tw$ in $T_{p_o}S^\perp$.
Then we find

$$(1)\ \mathrm{d}\exp_{p_o}(w) = \Big(\frac{d}{dt}\ \exp 0_{p(t)}\Big)_{t=0} = \Big(\frac{d}{dt}\ p(t)\Big)_{t=0} = w \in T_{p_o}S$$

$$(2)\ \mathrm{d}\exp_{p_o}(w) = \Big(\frac{d}{dt}\ \exp tw\Big)_{t=0} = w \in T_{p_o}S^\perp$$

Thus $\mathrm{d}\exp_{p_o} w = w$ in either case, hence $d\exp_{p_o}$ has full rank $n = \dim N = \dim M$. □

For any $c \in \mathbb{R}$, let

$$N_c = \{v \in N \mid g(v, v) = c\}.$$

At a given point $p \in S$, the vectors in $N_c$ at $p$ from a "sphere" (in the sense of the metric $g$) in the normal space $T_pS^\perp$. Let $S_c = \exp N_c$ be the image of $N_c$ under the geodesic spray exp, called a normal tube around $S$ in $M$; it may be thought of as the points at constant distance $\sqrt{c}$ from $S$, at least if the metric $g$ is positive definite.
We shall need the following observation.

2.5.15 **Lemma**. *Let* $t \to p = p(s,t)$ *be a family of geodesics depending on a parameter* $s$. *Assume they all have the same (constant) speed independent of* $s$, *i.e.* $g(\partial p/\partial t, \partial p/\partial t) = c$ *is independent of* $s$. *Then* $g(\partial p/\partial s, \partial p/\partial t)$ *is constant along each geodesic.*

**Proof**. Because of the constant speed,

$$0 = \frac{\partial}{\partial s} g\Big(\frac{\partial p}{\partial t}, \frac{\partial p}{\partial t}\Big) = 2g\Big(\frac{\nabla \partial p}{\partial s \partial t}, \frac{\partial p}{\partial t}\Big)$$

Hence

$$\frac{\partial}{\partial t} g\Big(\frac{\partial p}{\partial s}, \frac{\partial p}{\partial t}\Big) = g\Big(\frac{\nabla \partial p}{\partial t \partial s}, \frac{\partial p}{\partial t}\Big) + g\Big(\frac{\partial p}{\partial s}, \frac{\nabla \partial p}{\partial t \partial t}\Big) = 0 + 0$$

the first 0 because of the symmetry of $\nabla$ and the previous equality, the second 0 because $t \to p(s,t)$ is a geodesic. □
We now return to $S$ and $N$.

2.5.16 **Gauss Lemma (Version 1)**. *The geodesics through a point of* $S$ *with initial velocity orthogonal to* $S$ *meet the normal tubes* $S_c$ *around* $S$ *orthogonally.*

**Proof**. A curve in $S_c = \exp N_c$ is of the form $\exp_{p(s)} v(s)$ where $p(s) \in S$ and $v(s) \in T_{p(s)}S^\perp$. Let $t \to p(s,t) = \exp_{p(s)} tv(s)$ be the geodesic with initial velocity $v(s)$. Since $v(s) \in N_c$ these geodesics have speed independent of $s$: $g(\partial p/\partial t, \partial p/\partial t) = g(v(s), v(s)) = c$. Hence the lemma applies. Since $(\partial p/\partial t)_{t=0} = v(s) \in T_{p(s)}S^{\perp}$ is perpendicular to $(\partial p/\partial s)_{t=0} = dp(s)/ds \in T_{p(s)}S$ at $t = 0$, the lemma says that $\partial p/\partial t$ remains perpendicular to $\partial p/\partial s$ for all $t$. On the other hand, $p(s, 1) = \exp v(s)$ lies in $S_c$, so $\partial p/\partial s$ is tangential to $S_c$ at $t = 1$. Since all tangent vectors to $S_c = \exp N_c$ are of this form we get the assertion. □

2.5.17 **Gauss Lemma (Version 2)**. *Let $p$ be any point of $M$. The geodesics through $p$ meets the spheres $S_c$ in $M$ centered at $p$ orthogonally.*

**Proof**. This is the special case of the previous lemma when $S = \{p\}$ reduces to a single point. □

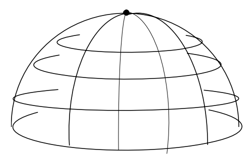


Fig. 1. Gauss's Lemma for a sphere

2.5.18 **Remark**. The *spheres around* $p$ are by definition the images of the "spheres" $g(v, v) = c$ in $T_pM$ under the geodesic spray. If the metric is positive definite, these are really the points at constant distance from $p$, as follows from the definition of the geodesics of a Riemann metric. For example, when $M = S^2$ and $p$ the north pole, the "spheres" centered at $p$ are the circles of latitude $\phi$ =constant and the geodesics through $p$ the circles of longitude $\theta$ =constant.

**EXERCISES 2.5**

1. Complete the proof of *Theorem* 2.5.10.

2. Let $g$be Riemann metric, $\nabla$the Levi-Civita connection of $g$. Consider the Riemann metric $g = (g_{ij})$ as a $(0, 2)$-tensor. Prove that $\nabla g = 0$, i.e. $\nabla_v g = 0$ for any vector v. [Suggestion. Let $X, Y, Z$ be a vector field and consider $D_Z(g(X, Y))$. Use the axioms defining the covariant derivative of arbitrary tensor fields.]

3. Use Christoffel's formulas to calculate the Christoffel symbols $\Gamma_{ij,k}$, $\Gamma^k_{ij}$ in spherical coordinates $(\rho, \theta, \phi)$: $x = \rho \cos\theta\, \sin\phi$, $y = \rho \sin\theta \sin\phi$, $z = \rho\cos\theta$. [Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$ on $\mathbb{R}^3$. Use $\rho, \theta, \phi$ as indices $i, j, k$ rather than $1, 2, 3$.]

4. Use Christoffel's formulas to calculate the $\Gamma^k_{ij}$ for the metric

$$ds^2 = (dx^1)^2 + [(x^2)^2 - (x^1)^2](dx^2)^2.$$

5. Let $S$ be an $m$-dimensional submanifold of a Riemannian manifold $M$. Let $(x^1, \cdots, x^m)$ be a coordinate system on $S$ and write points of $S$ as $p =$

$p(x^1, \cdots, x^m)$. Use Christoffel's formula (10) to show that the Christoffel symbols $\Gamma_{ij}^k$ of the induced Riemann metric $ds^2 = g_{ij}dx^i dx^j$ on $S$ are given by

$$\Gamma_{ij}^k = g^{kl} g\Big(\frac{\nabla}{\partial x^j}\frac{\partial p}{\partial x^k}, \frac{\partial p}{\partial x^l}\Big).$$

[The inner product $g$ and the covariant derivative $\nabla$ on the right are taken on $M$ and $g^{kl}g_{lj} = \delta_{kj}$.]

6. a) Let $M, g$ be a manifold with a Riemann metric, $f : M \to M$ an isometry. Suppose there is a $p_o \in M$ and a basis $(v_1, \cdots, v_n)$ of $T_{p_o}M$ so that $f(p_o) = p_o$ and $df_p(v_i) = v_i$, $i = 1, \cdots, n$. Show that $f(p) = p$ for all points $p$ in a neighbourhood of $p_o$. [Suggestion: consider geodesics.]
b) Show that every isometry of the sphere $S^2$ is given by an orthogonal linear transformation of $\mathbb{R}^3$. [Suggestion. Fix $p_o \in S^2$ and an orthonormal basis $v_1, v_2$ for $T_{p_o}S^2$. Let $A$ be an isometry of $S^2$. Apply (a) to $f = B^{-1}A$ where $B$ is a suitably chosen orthogonal transformation.]

7. Let $S$ be a 2–dimensional manifold with a positive definite Riemannian metric $ds^2$. Show that around any point of $S$ one can introduce an orthogonal coordinate system, i.e. a coordinate system $(\sigma, v)$ so that the metric takes the form

$$ds^2 = a(\sigma, v)d\sigma^2 + b(\sigma, v)dv^2.$$

[Suggestion: use Version 2 of Gauss's Lemma.]

8. Prove that (4) implies (3).

*The following problems deal with some aspects of the question to what extent the connection determines the Riemann metric by the requirement that they be compatible (as in Definition* 2.5.1*).*

9. Show that a Riemann metric $g$ on $\mathbb{R}^n$ is compatible with the standard connection $D$ with zero components $\Gamma_{ij}^k = 0$ in the Cartesian coordinates $(x^i)$, if and only if $g$ has constant components $g_{ij} = c_{ij}$ in the Cartesian coordinates $(x^i)$.

10. Start with a connection $\nabla$ on $M$. Suppose $g$ and $'g$ are two Riemann metrics compatible with $\nabla$. (a) If $g$ and $'g$ agree at a single point $p_o$ then they agree in a whole neighbourhood of $p_o$. [Roughly: the connections determines the metric in terms of its value at a single point, in the following sense.
(b) Suppose $g$ and $'g$ have the same signature, i.e. they have the same number of $\pm$signs if expressed in terms of an orthonormal bases (not necessarily the same for both) at a given point $p_o$ (see remark (3) after definition 5.3). Show that in any coordinate system $(x^i)$, the form $g_{ij}dx^i dx^j$ can be (locally) transformed into $'g_{ij}dx^i dx^j$ by a linear transformation with constant coefficients $x^i \to c^i_j x^j$. [Roughly: the connection determines the metric up to a linear coordinate transformation.]

## 2.6 Curvature identities

Let $M$ be a manifold with Riemannian metric $g$, $\nabla$ its Levi-Civita connection. For brevity, write $(u, v) = g(u, v)$ for the scalar-product.

2.6. 1 **Theorem (Curvature Identities)**. *For any vectors $u, v, w, a, b$ at a point $p \in M$*
*1.* $R(u, v) = -R(v, u)$
*2.* $(R(u, v)a, b) = -(a, R(u, v)b)$ *[$R(u, v)$ is skew-adjoint]*
*3.* $R(u, v)w + R(w, u)v + R(v, w)u = 0$ *[Bianchi's Identity]*
*4.* $(R(u, v)a, b) = (R(a, b)u, v)$

**Proof**. Let $p = p(\sigma, \tau)$ be a surface in $M$, $U = \frac{\partial p}{\partial\sigma}$, $V = \frac{\partial p}{\partial\tau}$ (for any values of $\sigma, \tau$). Let $X = X(\sigma, \tau)$ be a smooth vector field along $p = p(\sigma, \tau)$. Then

$$R(U, V)X = \frac{\nabla}{\partial\tau}\frac{\nabla X}{\partial\sigma} - \frac{\nabla}{\partial\sigma}\frac{\nabla X}{\partial\tau}. \qquad (*)$$

1. This is clear from (*).
2. Let $A = A(\sigma, \tau)$, $B = B(\sigma, \tau)$ be smooth vector fields along $p = p(\sigma, \tau)$. Compute

$$\frac{\partial}{\partial\sigma}(A, B) = (\frac{\nabla A}{\partial\sigma}, B) + (A, \frac{\nabla B}{\partial\sigma}),$$
$$\frac{\partial}{\partial\tau}\frac{\partial}{\partial\sigma}(A, B) = (\frac{\nabla}{\partial\tau}\frac{\nabla}{\partial\sigma}A, B) + (\frac{\nabla A}{\partial\sigma}, \frac{\nabla B}{\partial\tau}) + (\frac{\nabla A}{\partial\tau}, \frac{\nabla B}{\partial\sigma}) + (A, \frac{\partial\nabla}{\partial\tau}\frac{\partial\nabla}{\partial\sigma}B).$$

Interchange $\sigma$ and $\tau$, subtract, and use (*):

$$0 = (R(U, V)A, B) + (A, R(U, V)B).$$

3. This time take $p = p(\sigma, \tau, \rho)$ to be a smooth function of t*hree* scalar-variables $(\sigma, \tau, \rho)$. Let

$$U = \frac{\partial p}{\partial\sigma},\ V = \frac{\partial p}{\partial\tau},\ \mathrm{W} = \frac{\partial p}{\partial\rho}.$$

Then

$$R(U, V)\mathrm{W} = \frac{\nabla}{\partial\tau}\frac{\nabla}{\partial\sigma}\frac{\partial p}{\partial\rho} - \frac{\nabla}{\partial\sigma}\frac{\nabla}{\partial\tau}\frac{\partial p}{\partial\rho}.$$

In this equation, permute $(\sigma, \tau, \rho)$ cyclically and add (*u*sing $\frac{\nabla}{\partial\tau}\frac{\partial p}{\partial\rho} = \frac{\nabla}{\partial\rho}\frac{\partial p}{\partial\tau}$ etc.) to find the desired formula.
4. This follows from 1–3 by a somewhat tedious calculation (omitted). □

By definition

$$R(u, v)w = R^i_{qkl}u^k v^l w^q \partial_i.$$

The components of $R$ with respect to a coordinate system $(x^i)$ are defined by the equation

$$R(\partial_k, \partial_l)\partial_q = R^i_{qkl}\partial_i.$$

We also set

$$(R(\partial_k,\partial_l)\partial_q,\partial_j) = R^i_{qkl}g_{ij} = R_{\mathrm{j}qkl}$$

2.6. 2 **Corollary**. *The components of $R$ with respect to any coordinate system satisfy the following identities.*

1. $R_{bcd}a^a = -R^a_{bdc}$
2. $R_{abcd} = -R_{bacd}$
3. $R^a_{bcd} + R^a_{dbc} + R^a_{cdb} = 0$
4. $R_{abcd} = R_{dcba}$

**Proof.** This follows immediately from the theorem. □

2.6. 3 **Theorem (Jacobi's Equation)**. *Let $p = p_\sigma(\tau)$ be a one-parameter family of geodesics ($\sigma$ =parameter). Then*

$$\frac{\nabla^2}{\partial\tau^2}\frac{\partial p}{\partial\sigma} = R(\frac{\partial p}{\partial\sigma},\frac{\partial p}{\partial\tau})\frac{\partial p}{\partial\tau}.$$

**Proof**.

$$\begin{aligned}\frac{\nabla^2}{\partial\tau^2}\frac{\partial p}{\partial\sigma} &= \frac{\nabla}{\partial\tau}\frac{\nabla}{\partial\sigma}\frac{\partial p}{\partial\tau} && [\text{symmetry of } \nabla]\\ &= \frac{\nabla}{\partial\tau}\frac{\nabla}{\partial\sigma}\frac{\partial p}{\partial\tau} - \frac{\nabla}{\partial\sigma}\frac{\nabla}{\partial\tau}\frac{\partial p}{\partial\tau} && [p_\sigma(\tau) \text{ is a geodesic}]\\ &= R(\frac{\partial p}{\partial\sigma},\frac{\partial p}{\partial\tau})\frac{\partial p}{\partial\tau}\end{aligned}$$

□

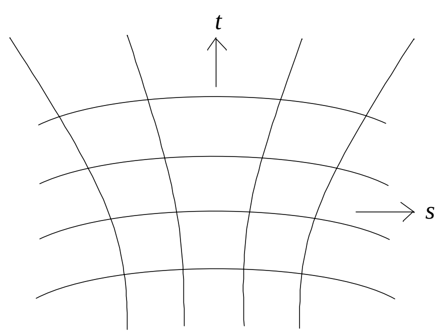



Fig.1 A family of geodesics

2.6. 4 **Definition**. For any two vectors $v, w$, $\mathrm{Ric}(v,w)$ is the trace of the linear transformation $u \to R(u,v)w$. In coordinates:

$$R(u,v)w = R^i_{qkl}u^kv^lw^q\partial_i,$$
$$\mathrm{Ric}(v,w) = R^k_{qkl}v^lw^q.$$

Thus Ric is a tensor obtained by a contraction of $R$: $(\mathrm{Ric})_{ql} = R^k_{qkl}$. One also writes $(\mathrm{Ric})_{ql} = R_{ql}$.

2.6. 5 **Theorem**. *The Ricci tensor is symmetric, i.e.* $\mathrm{Ric}(v,w) = \mathrm{Ric}(w,v)$ *for all* v*ectors* $v, w$.

**Proof**. Fix $v, w$ and take the trace of Bianchi's identity, considered as a linear transformation of $u$:

$$\mathrm{tr}\{u \to \ R(u,v)w\} + \mathrm{tr}\{u \to R(v,w)u\} + \mathrm{tr}\{u \to R(w,u)v\} = 0.$$

Since $u \to R(v,w)u$ is a skew-adjoint transformation, $\mathrm{tr}\{u \to R(v,w)u\} = 0$. Since $R(w,u) = -R(u,w), \mathrm{tr}\{u \to R(w,u)v\} = -\mathrm{tr}\{u \to R(u,w)v\}$. Thus $\mathrm{tr}\{u \to R(u,v)w\} = \mathrm{tr}\{u \to R(u,w)v\}$, i.e. $\mathrm{Ric}(v,w) = \mathrm{Ric}(w,v)$.

□

# Chapter 3

# Calculus on manifolds

## 3.1 Differential forms

By definition, a covariant tensor $T$ at $p \in M$ is a quantity which relative to a coordinate system $(x^i)$ around $p$ is represented by an expression

$$T = \sum T_{ij\cdots}\, dx^i \otimes dx^j \otimes \cdots$$

with real coefficients $T_{ij\cdots}$ and with the differentials $dx^i$ being taken at $p$. The tensor need not be homogeneous, i.e. the sum involves $(0,k)$–tensors for different values of $k$. Such expressions are added and multiplied in the natural way, but one must be careful to observe that multiplication of the $dx^i$ is not commutative, nor satisfies any other relation besides the associative law and the distributive law. Covariant tensors at $p$ can be thought of as purely formal algebraic expressions of this type. Differential forms are obtained by the same sort of construction if one imposes in addition the rule that the $dx^i$ *anticommute*. This leads to the following definition.

3.1.1 **Definition**. A *differential form* $\varpi$ at $p \in M$ is a quantity which relative to a coordinate system $(x^i)$ around $p$ is represented by a formal expression

$$\varpi = \sum f_{ij\cdots}\, dx^i \wedge dx^j \wedge \cdots \tag{1}$$

with real coefficients $f_{ij\cdots}$ and with the differentials $dx^i$ being taken at $p$. Such expressions are added and multiplied in the natural way but subject to the relation

$$dx^i \wedge dx^j = -dx^j \wedge dx^i. \tag{2}$$

If all the wedge products $dx^i \wedge dx^i \wedge \cdots$ in (1) contain exactly $k$ factors, then $\varpi$ said to be *homogeneous of degree* $k$ and is called a $k$–form at $p$. A *differential form on* $M$ (also simply called a form) associates to each $p \in M$ a form at $p$.

The form is smooth if its coefficients $f_{ij\cdots}$ (relative to any coordinate system) are smooth functions of $p$. This will always be assumed to be the case.

**Remarks**. (1) The definition means that we consider identical expressions (1) which can be obtained from each other using the relations (2), possibly repeatedly or in conjunction with the other rules of addition and multiplication. For example, since $dx^i \wedge dx^i = -dx^i \wedge dx^i$ (any $i$) one finds that $2dx^i \wedge dx^i = 0$, so $dx^i \wedge dx^i = 0$. The expressions for $\varpi$ in two coordinate systems $(x^i)$, $(\tilde{x}^i)$ are related by the substitutions $x^i = f^i(\tilde{x}^1, \cdots, \tilde{x}^n)$, $dx^i = (\partial f^i/\partial \tilde{x}^j)d\tilde{x}^j$ on the intersection of the coordinate domains.
(2) By definition, the $k$-fold wedge product $dx^i \wedge dx^j \wedge \cdots$ transforms like the $(0,k)$-tensor $dx^i \otimes dx^j \otimes \cdots$, but is *alternating* in the differentials $dx^i, dx^j$, i.e. changes sign if two *adjacent* differentials are interchanged.
(3) Every differential form is uniquely a sum of homogeneous differential forms. For this reason one can restrict attention to forms of a given degree, except that the wedge-product of a $k$-form and an $l$-form is a $(k+l)$-form.

3.1.2 **Example**. In polar coordinates $x = r\cos\theta$, $y = r\sin\theta$,

$$dx = \cos\theta dr - r\sin\theta d\theta, dy = \sin\theta dr + r\cos\theta d\theta$$
$$dx \wedge dy = (\cos\theta dr - r\sin\theta d\theta) \wedge (\sin\theta dr + r\cos\theta d\theta)$$
$$= \cos\theta\sin\theta dr\wedge dr + r\cos^2\theta dr\wedge d\theta - r\sin^2\theta d\theta\wedge dr - r^2\sin\theta\cos\theta d\theta\wedge d\theta$$
$$= r dr \wedge d\theta$$

3.1.3 **Lemma**. *Let $\varpi$ be a differential $k$-form, $(x^i)$ a coordinate system. Then $\varpi$ can be uniquely represented in the form*

$$f_{ij\cdots}dx^i \wedge dx^j \wedge \cdots \ (i < j < \cdots) \tag{3}$$

*with indices in increasing order.*

**Proof**. It is clear that $\varpi$ can be represented in this form, since the differentials $dx^i$ may reordered at will at the expense of introducing some change of sign. Uniqueness is equally obvious: if two expressions (1) can be transformed into each other using (2), then the terms involving a given set of indices $\{i, j\cdots\}$ must be equal, since no new indices are introduced using (2). Since the indices are ordered $i < j < \cdots$ there is only one such term in either expression, and these must then be equal. □

3.1.4 **Definition.** A $(0,k)$-tensor $T_{ij\cdots}dx^i \otimes dx^j \otimes \cdots$ is alternating if its components change sign when two adjacent indices are interchanged $T_{\cdots ij\cdots} = -T_{\cdots ji\cdots}$.for all k-tuples of indices $(\cdots, i, j, \cdots)$. Equivalently, $T(\cdots, v, w, \cdots) = -T(\cdots, w, v, \cdots)$ for all $k$-tuples of vectors $(\cdots, v, w, \cdots)$.

3.1.5 **Remark**. Under a general permutation $\sigma$ of the indices, the components of an alternating tensor are multiplied by the sign $\pm1$ of the permutation: $T_{\cdots\sigma(\mathrm{i})\sigma(j)\cdots} = \mathrm{sgn}(\sigma)T_{\cdots ji\cdots}$....

3.1.6 **Lemma**. *There is a one-to-one correspondence between differential $k$-forms and alternating $(0,k)$-tensors so that the form $f_{ij\cdots}dx^i \wedge dx^j \wedge \cdots$ $(i < j < \cdots)$ corresponds to the tensor $T_{ij\cdots}dx^i \otimes dx^j \otimes \cdots$ defined by*

(1) $T_{ij\cdots} = f_{ij\cdots}$ *if* $i < j < \cdots$ *and*
(2) $T_{ij\cdots}$ *changes sign when two adjacent indices are interchanged.*

**Proof.** As noted above every $k$ form can be written *uniquely* as

$$f_{ij\cdots}dx^i \wedge dx^j \wedge \cdots \ (i < j < \cdots)$$

where the sum goes over *ordered* k-tuples $i < j < \cdots$. It is also clear that an alternating $(0,k)$-tensor $T_{ij\cdots}dx^i \otimes dx^j \otimes \cdots$ is *uniquely* determined by its components

$$T_{ij\cdots} = f_{ij\cdots} \text{ if } i < j < \cdots .$$

indexed by ordered k-tuples $i < j < \cdots$. Hence the formula

$$T_{ij\cdots} = f_{ij\cdots} \text{ if } i < j < \cdots .$$

does give a one-to-one correspondence. The fact that the $T_{ij\cdots}$ defined in this way transform like a $(0,k)$-tensor follows from the transformation law of the $dx^i$. □.

3.1.7 **Example: forms on $\mathbb{R}^3$.**
(1) The 1-form $Adx + Bdy + Cdz$ is a covector, as we know.
(2) The differential 2-form $Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy$ corresponds to the $(0,2)$-tensor $T$ with components $T_{zy} = -T_{yz} = P$, $T_{zx} = -T_{xz} = Q$, $T_{xy} = -T_{yx} = R$. This is just the tensor

$$P(dy \otimes dz - dz \otimes dy) + Q(dz \otimes dx - dx \otimes dz) + R(dx \otimes dy - dy \otimes dx)$$

If we write $v = (P, Q, R)$, then $T(a, b) = v \cdot (a \times b)$ for all vectors $a, b$..
(3) The differential 3-form $Ddx \wedge dy \wedge dz$ corresponds to the $(0,3)$-tensor $T$ with components $D\epsilon_{ijk}$where $\epsilon_{ijk}$ =*sign of the permutation* $(ijk)$ of $(xyz)$. (We use $(xyz)$ as indices rather than (123).) Thus every 3-form in $\mathbb{R}^3$ is a multiple of the form $dx \wedge dy \wedge dz$. This form we know: for three vectors $u = (v^i), v = (v^j), w = (w^k)$, $\epsilon_{ijk}u^i v^j w^k = \det[\mathrm{u}, v, w]$. Thus $T(u, v, w) = D\det[u, v, w]$.

3.1.8 **Remark**. (a) We shall now *identify* $k$-forms with alternating $(0,k)$-tensors. This identification can be expressed by the formula

$$dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_{\sigma \in S_k} \operatorname{sgn}(\tau) dx^{i_{\sigma(1)}} \otimes \cdots \otimes dx^{i_{\sigma(k)}} \tag{4}$$

The sum runs over the group $S_k$ of all permutations $(1, \cdots, \mathrm{k}) \to (\sigma(1), \cdots, \sigma(\mathrm{k}))$; $\operatorname{sgn}(\sigma) = \pm 1$ is the sign of the permutation $\sigma$. The multiplication of forms then takes on the following form. Write $k = p+q$ and let $S_p \times S_q$ be the subgroup of the group $S_k$ which permutes the indices $\{1, \cdots, p\}$ and $\{p+1, \cdots, p+q\}$ among themselves. Choose a set of coset representatives $[S_{p+q}/S_p \times S_q]$ for the quotient so that every element of $\sigma \in S_{p+q}$ can be uniquely written as $\sigma = \tau\sigma'\sigma''$ with

$\tau \in [S_{p+q}/S_p \times S_q]$ and $(\sigma', \sigma'') \in S_p \times S_q$. If in (4) one first performs the sum over $(\sigma', \sigma'')$ and then over $\tau$ one finds the formidable formula

$$(dx^{i_1} \wedge \cdots \wedge dx^{i_p}) \wedge (dx^{i_{p+1}} \wedge \cdots \wedge dx^{i_{p+q}}) \tag{5}$$
$$\sum_{\tau \in [S_{p+q}/S_p \times S_q]} \operatorname{sgn}(\tau)(dx^{i_{\tau(1)}} \wedge \cdots \wedge dx^{i_{\tau(p)}}) \otimes (dx^{i_{\tau(p+1)}} \wedge \cdots \wedge dx^{i_{\tau(p+q)}})$$

For $[S_{p+q}/S_p \times S_q]$ one can take (for example) the $\tau \in S_{p+q}$ satisfying

$$\tau(1) < \cdots < \tau(p) \text{ and } \tau(p+1) < \cdots < \tau(p+q).$$

(These $\tau$'s are called "*shuffle permutations*"). On the other hand, one can also let the sum in (5) run over all $\tau \in S_{p+q}$ provided one divides by $p!q!$ to compensate for the redundancy. Finally we note that (4) and (5) remain valid if the $dx^i$ are replaced by arbitrary 1–forms $\theta^i$, and are then independent of the coordinates.
The formula (5) gives the multiplication law for differential forms when considered as tensors via (4). Luckily, the formulas (4) and (5) are rarely needed. The whole point of differential forms is that both the alternating property and the transformation law is built into the notation, so that they can be manipulated "mechanically".

(b) We now have (at least) three equivalent ways of thinking about $k$-forms:
 (i) formal expressions $f_{ij\cdots} dx^i \wedge dx^j \wedge \cdots$
 (ii) alternating tensors $T_{ij\cdots} dx^i \otimes dx^j \otimes \cdots$
 (iii) alternating multilinear functions $T(v, w, \cdots)$

3.1.9 **Theorem**. *On an $n$–dimensional manifold, any differential $n$–form is written as*

$$Ddx^1 \wedge dx^2 \wedge \cdots \wedge dx^n$$

*relative to a coordinate system $(x^i)$. Under a change of coordinates $\tilde{x}^i = \tilde{x}^i(x^1, \cdots, x^n)$,*

$$Ddx^1 \wedge dx^2 \wedge \cdots \wedge dx^n = \tilde{D}d\tilde{x}^1 \wedge d\tilde{x}^2 \wedge \cdots \wedge d\tilde{x}^n$$

*where $D = \tilde{D} \det(\partial\tilde{x}^i/\partial x^j)$.*

**Proof**. Any $n$–form is a linear combination of terms of the type $dx^{i_1} \wedge \cdots \wedge dx^{i_n}$. This $= 0$ if two indices occur twice. Hence any $n$–form is a multiple of the form $dx^1 \wedge \cdots \wedge dx^n$. This proves the first assertion. To prove the transformation rule, compute:

$$d\tilde{x}^1 \wedge \cdots \wedge d\tilde{x}^n = \frac{\partial\tilde{x}^1}{\partial x^{i_1}} dx^{i_1} \wedge \cdots \wedge \frac{\partial\tilde{x}^n}{\partial x^{i_n}} dx^{i_n} = \epsilon_{i_1 \cdots i_n} \frac{\partial\tilde{x}^1}{\partial x^{i_1}} \cdots \frac{\partial\tilde{x}^n}{\partial x^{i_n}} dx^1 \wedge \cdots \wedge dx^n$$
$$= \det\Big(\frac{\partial\tilde{x}^i}{\partial x^j}\Big) dx^1 \wedge \cdots \wedge dx^n$$

Hence

$$\tilde{D}d\tilde{x}^1 \wedge \cdots \wedge d\tilde{x}^n = Ddx^1 \wedge \cdots \wedge dx^n$$

where $\tilde{D}\det(\partial\tilde{x}^i/\partial x^j) = D$. $\square$

### Differential forms on a Riemannian manifold

From now on assume that there is given a Riemann metric $g$ on $M$. In coordinates $(x^i)$ we write $ds^2 = g_{ij}dx^i dx^j$ as usual.

3.1.10 **Theorem**. *The $n$–form*

$$\sqrt{|\det(g_{ij})|}\ dx^1 \wedge \cdots \wedge dx^n \tag{6}$$

*is independent of the coordinate system $(x^i)$ up to a $\pm$ sign.*

More precisely, the forms (6) corresponding to two coordinate systems $(x^i)$, $(\tilde{x}^j)$ agree on the intersection of the coordinate domains up to the sign of $\det(\partial x^i/\partial\tilde{x}^j)$.

**Proof**. Exercise.

3.1.11 **Definitions**. The Jacobian $\det(\partial x^i/\partial\tilde{x}^j)$ of a coordinate transformation is always non-zero (where defined), hence cannot change sign along any continuous curve so that it will be either positive or negative on any connected set. If there is a collection of coordinates covering the whole manifold (i.e. an atlas) for which the Jacobian of the coordinate transformation between any two coordinate systems is positive, then the manifold is said to be *orientable.* These coordinates systems are then called *positively oriented*, as is any coordinate system related to them by a coordinate transformation with everywhere positive Jacobian. Singling out one such a collection of positively oriented coordinates determines what is called an orientation on $M$. A connected orientable manifold has $M$ has only two orientations, but if $M$ is has several connected components, the orientation may be chosen independently on each. We shall now assume that $M$ is oriented and only use positively oriented coordinate systems. This eliminates the ambiguous sign in the above $n$-from, which is then called the volume element of the Riemann metric $g$ on $M$, denoted $\mathrm{vol}_g$.

3.1.12 **Example**. In $\mathbb{R}^3$ with the Euclidean metric $dx^2 + dy^2 + dz^2$ the volume element is :
*Cartesian coordinates* $(x, y, z)$: $dx \wedge dy \wedge dz$
*Cylindrical coordinates* $(r, \theta, z)$: $rdr \wedge d\theta \wedge dz$
*Spherical coordinates* $(\rho, \theta, \phi)$: $\rho^2 \sin\phi d\rho \wedge d\theta \wedge d\phi$

3.1.13 **Example**. Let $S$ be a two-dimensional submanifold of $\mathbb{R}^3$ (smooth surface). The Euclidean metric $dx^2 + dy^2 + dz^2$on $\mathbb{R}^3$ gives a Riemann metric $g = ds^2$ on $S$ by restriction. Let u,v coordinates on $S$. Write $p = p(u, v)$ for the point on $S$ with coordinates $(u, v)$. Then

$$\sqrt{|\det g_{ij}|} = \left\| \frac{\partial p}{\partial u} \times \frac{\partial p}{\partial v} \right\|.$$

Here $g_{ij}$ is the matrix of the Riemann metric in the coordinate system u,v. The right-hand side is the norm of the cross-product of vectors in $\mathbb{R}^3$.
Recall that in a Riemannian space indices on tensors can be raised and lowered at will.

3.1.14 **Definition**. The scalar product $g(\alpha,\beta)$ of two $k$-forms

$$\alpha = a_{ij\cdots}dx^i \wedge dx^j \cdots , \quad \beta = b_{ij\cdots}dx^i \wedge dx^j \cdots , \ i < j \cdots ,$$

is

$$g(\alpha,\beta) = a_{ij\cdots}b^{ij\cdots},$$

summed over ordered $k$-tuples $i < j < \cdots$.

3.1.15 **Theorem (Star Operator)**. *For any $k$-form $\alpha$ there is a unique $(n-k)$-form $*\alpha$ so that*

$$\beta \wedge *\alpha = g(\beta,\alpha) vol_g. \tag{7}$$

*for all $k$-forms $\beta$. The explicit formula for this $*$-operator is*

$$*a_{i_{k+l}\cdots i_n} = \sqrt{|\det(g_{ij})|}\epsilon_{i_1\cdots i_n} a^{i_1\cdots i_k}, \tag{8}$$

*where the sum is extended over ordered $k$-tuples. Furthermore*

$$*(*\alpha) = (-1)^{k(n-k)} sgn \det(g_{ij})\alpha. \tag{9}$$

**Proof**. The statement concerns only tensors at a given point $p_o$. So we fix $p_o$ and we choose a coordinate system $(x^i)$ around $p_o$ so that the metric at the point $p_o$ takes on the pseudo-Euclidean form $g_{ij}(p_o) = g_i\delta_{ij}$ with $g_i = \pm 1$. To verify the first assertion it suffices to take for $\alpha$ a fixed basis $k$-form $\alpha = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ with $i_1 < \cdots < i_k$. Consider the equation (7) for all basis forms $\beta = dx_{j_1} \wedge \cdots \wedge dx_{j_k}, j_1 < \cdots < j_k$. Since $g_{ij} = g_i\delta_{ij}$ at $p_o$, distinct basis $k$-forms are orthogonal, and one finds

$$(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) \wedge *(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = g_{i_1} \cdots g_{i_k} dx_1 \wedge \cdots \wedge dx_n$$

for $\beta = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ and zero for all other basis forms $\beta$. One sees that

$$*(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = \epsilon_{i_1\cdots i_n} g_{i_1} \cdots g_{i_k} dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_n} \tag{10}$$

where $i_{k+1}, \cdots , i_n$ are the remaining indices in ascending order. This proves the existence and uniqueness of $*\alpha$ satisfying (7). The formula (10) also implies (8) for the components of the basis forms $\alpha$ in the particular coordinate system $(x^i)$ and hence for the components of any $k$-form $\alpha$ in this coordinate system $(x^i)$. But the $\sqrt{|\det(g_{ij})|}\epsilon_{i_1\cdots i_n}$ are the components of the $(0, n)$ tensor corresponding to the $n$-form in theorem 3.1.10. It follows that (8) is an equation between components of tensors of the same type, hence holds in all coordinate systems as soon as it holds in one. The formula (9) can been seen in the same way: it suffices to take the special $\alpha$ and $(x^i)$ above.

□

The operator $\alpha \to *\alpha$ on differential forms is called the (Hodge) star operator

3.1.16 **Example: Euclidean 3-space**. Metric: $dx^2 + dy^2 + dz^2$
The $*$-operator is given by:
(0) For 0-forms: $*f = f dx \wedge dy \wedge dz$
(1) For 1-forms: $*(Pdx + Qdy + Rdz) = (Pdy \wedge dz - Qdx \wedge dz + Rdx \wedge dy)$
(2) For 2-forms: $*(Ady \wedge dz + Bdx \wedge dz + Cdx \wedge dy) = (Adx - Bdy + Cdz)$.
(3) For 3-forms: $*(Ddx \wedge dy \wedge dz) = D$.

If we identify vectors with covectors by the Euclidean metric, then we have the formulas

$$*(a \wedge b) = a \times b, \quad *(a \wedge b \wedge c) = a \cdot (b \times c).$$

3.1.17 **Example: Minkowski space**. Metric: $(dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$

2-form: $F = F_{ij} dx^i \wedge dx^j$ is written as

$$F = \sum_\alpha E_\alpha dx^0 \wedge dx^\alpha + B^1 dx^2 \wedge dx^3 + B^2 dx^3 dx^1 + B^3 dx^1 \wedge dx^2;$$

where $\alpha = 1, 2,3$. Then

$$*F = -\sum_\alpha \ B^\alpha dx^0 \wedge dx^\alpha + E_1 dx^2 \wedge dx^3 + E_2 dx^3 \wedge dx^1 + E_3 dx^1 \wedge dx^2.$$

## Appendix: algebraic definition of alternating tensors

Let $V$ be a finite–dimensional vector space (over any field). Let $T = T(V)$ be the space of all tensor products of elements of $V$. If we fix a basis $\{e_1, \cdots, e_n\}$ for $V$ then the elements of $T$ can be uniquely written as finite sums $\sum a_{ij\cdots} e_i \otimes e_j \cdots$. Tensor multiplication $a \otimes b$ makes $T$ into an algebra, called the *tensor algebra* of $V$. (It is the algebra freely generated by a basis $e_1, \cdots, e_n$, i.e. without any relations.) Let $I$ be the ideal of $T$ generated by tensors of the form $x \otimes y - y \otimes x$ with $x, y \in V$. Let $\Lambda = \Lambda(V)$ be the quotient algebra $T/I$. The multiplication in $\Lambda$ is denoted $a \wedge b$. In terms of a basis $\{e_i\}$, this means that $\Lambda$ is the algebra generated by the basis elements subject to the relations $e_i \wedge e_j = -e_j \wedge e_i$. It is called the *exterior algebra* of $V$. As a vector space $\Lambda(V)$ is isomorphic with the subspace of all alternating tensors in $T(V)$, as in 3.1.3, 3.1.4. (But the formula analogous to (4), i.e.

$$\sum_\sigma \ \mathrm{sgn}(\sigma) \ e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(k)}} \to e_{i_1} \wedge \cdots \wedge e_{i_k}$$

is not quite the natural map $T \to \Lambda = T/I$ restricted to alternating tensors; the latter would have an additional factor $k!$ on the right.) It is of course possible to realize $\Lambda$ as the space of alternating tensors in the first place, without

introducing the quotient $T/I$; but that makes the multiplication in $\Lambda$ look rather mysterious (as in (5)).
When this construction is applied with $V$ the space $T_p^*M$ of covectors at a given point $p$ of a manifold $M$ we evidently get the algebra $\Lambda(T_p^*M)$ of differential forms *at the point* $p$. A differential form on all of $M$, as defined in 3.1.1, associates to each $p \in M$ an element of $\Lambda(T_p^*M)$ whose coefficients (with respect to a basis $dx^i \wedge \widehat{dx^j} \wedge \cdots$ $(i < j < \cdots)$ coming from a local coordinate system $\{x^i\}$) are smooth functions of the $x^i$.

**EXERCISES 2.2**

1. Prove the formulas in *Example* 3.1.2.

2. Prove the formula in *Example* 3.1.13

3. Prove the formula (0)-(3) in *Example* 3.1.16

4. Prove the formulas $*(\mathrm{a}\wedge\mathrm{b}) = \mathrm{a}\times\mathrm{b}$, and $*(\mathrm{a}\wedge\mathrm{b}\wedge\mathrm{c}) = \mathrm{a}\cdot(\mathrm{b}\times\mathrm{c})$ in Example 3.1.16

5. Prove the formula for $*F$ in *Example* 3.1.16

6. Prove the $T_{ij\cdots}$ defined as in *Lemma* 3.1.5 in terms of a $(0,k)$-form $\mathrm{f}_{ij\cdots}dx^i \wedge dx \cdots d$o transform like the components of a $(0,k)$-tensor, as stated in the proof of the lemma.

7. Verify the formula $*(*F) = (-1)^{\mathrm{k(n-k)}}\mathrm{sgndet}(g_{ij})F$ of *Theorem* 3.1.15 for $(0,k)$–tensors on $\mathbb{R}^3$, k= $0, 1, 2, 3$ directly using the formulas of Example 3.1.16.

8. Let $\varphi$ be the following differential form on $\mathbb{R}^3$:

a) $(x^2+y^2+z^2)dx+ydy+dz$ b) $xdy\wedge dz+ydx\wedge dz$ c) $(x^2+y^2+z^2)dx\wedge dy\wedge dz$.

Find the expression for $\varphi$ in spherical coordinates $(\rho, \theta, \phi)$.

9. Let $\varphi$ be the following differential form on $\mathbb{R}^3$ in spherical coordinates:

a) $\sin^2\theta d\rho + \rho\cos\phi d\theta + \rho^2 d\phi$ b) $\rho\cos\theta d\rho \wedge d\theta - d\rho \wedge d\phi$ c)$\rho^2 d\rho \wedge d\theta \wedge d\phi$

(1) Use the formula (7) of Theorem 3.1.15 to find $*\varphi$.
(2) Use the formula (8) of Theorem 3.1.15 to find $*\varphi$.

10. Give a direct proof of Theorem 3.1.15, parts (7) and (8), using an arbitrary coordinate system $(x^i)$.

11. Prove (4).

12. Prove (5).

13. Let $\Lambda = T/I$ be the exterior algebra of $V$ as defined in 3.1.18. Show that the natural map $T \to \Lambda = T/I$ restricted to alternating $k$–tensors is given by (4) with an additional factor $k!$ on the right.

## 3.2 Differential calculus

On an arbitrary manifold it is not possible to define a derivative operation on arbitrary tensors in a general and natural way. But for differential forms (i.e. alternating covariant tensors) this is possible, as will now be discussed.

3.2.1 **Theorem**. *Let $\varpi = f_{ij\cdots}dx^i \wedge dx^j \cdots$ be a differential k-form. Then the differential (k+1)-form*

$$df_{ij\cdots} \wedge dx^i \wedge dx^j \cdots \ = \frac{\partial f_{ij\cdots}}{\partial x^k} dx^k \wedge dx^i \wedge dx^j \cdots$$

*is independent of the coordinate system $(x^i)$.*

**Proof**. Consider first the case of a differential 1-form. Thus assume $f_i dx^i = \tilde{f}_{\mathrm{a}} d\tilde{x}^{\mathrm{a}}$. This equation gives $f_i dx^i = \tilde{f}_{\mathrm{a}}(\partial\tilde{x}^{\mathrm{a}}/\partial x^i)dx^i$ hence $f_i = \tilde{f}_{\mathrm{a}}(\partial\tilde{x}^{\mathrm{a}}/\partial x^i)$, as we know. Now compute:

$$\begin{aligned}
\tfrac{\partial f_i}{\partial x^k} dx^k \wedge dx^i &= \tfrac{\partial}{\partial x^k}\Big(\tilde{f}_{\mathrm{a}} \tfrac{\partial \tilde{x}^{\mathrm{a}}}{\partial x^i}\Big) dx^k \wedge dx^i \\
&= \Big(\tfrac{\partial \tilde{f}_{\mathrm{a}}}{\partial x^k}\tfrac{\partial \tilde{x}^{\mathrm{a}}}{\partial x^i} + \tilde{f}_{\mathrm{a}} \tfrac{\partial^2 \tilde{x}^{\mathrm{a}}}{\partial x^k \partial x^i}\Big) dx^k \wedge dx^i \\
&= \tfrac{\partial \tilde{f}_{\mathrm{a}}}{\partial \tilde{x}^{\mathrm{b}}}\tfrac{\partial \tilde{x}^{\mathrm{b}}}{\partial x^k}\tfrac{\partial \tilde{x}^{\mathrm{a}}}{\partial x^i} dx^k \wedge dx^i + \tilde{f}_{\mathrm{a}} \tfrac{\partial^2 \tilde{x}^{\mathrm{a}}}{\partial x^k \partial x^i} dx^k \wedge dx^i \\
&= \tfrac{\partial \tilde{f}_{\mathrm{a}}}{\partial \tilde{x}^{\mathrm{b}}} d\tilde{x}^{\mathrm{b}} \wedge d\tilde{x}^{\mathrm{a}} + 0
\end{aligned}$$

because the terms $ki$ and $ik$ cancel in view of the symmetry of the second partials. The proof for a general differential $k$-form is obtained by adding some dots $\cdots$ to indicate the remaining indices and wedge–factors, which are not affected by the argument. □

3.2.2 **Remark**. If $T = (T_{i_1\cdots i_k})$ is the alternating $(0,k)$-tensor corresponding to $\varpi$ then the alternating (0,k+1)-tensor $dT$ corresponding to $d\varpi$ is given by the formula

$$(dT)_{\mathrm{j}_1\cdots\mathrm{j}_{\mathrm{k}+1}} = \sum_{q=1}^{\mathrm{k}+1} (-1)^{q-1} \frac{\partial T_{\mathrm{j}_1\cdots\mathrm{j}_{q-1}\mathrm{j}_{q+1}\cdots\mathrm{j}_{\mathrm{k}+1}}}{\partial x^{\mathrm{j}_q}}$$

3.2.3 **Definition**. Let $\varpi = f_{ij\cdots}dx^i \wedge dx^j \cdots$ be a differential k-form. Define a differential (k+1)-form $d\varpi$, called the exterior *derivative of* $\varpi$, by the formula

$$d\varpi := df_{ij\cdots} \wedge dx^i \wedge dx^j \cdots \ = \frac{\partial f_{ij\cdots}}{\partial x^k} dx^k \wedge dx^i \wedge dx^j \cdots .$$

Actually, the above recipe defines a form $d\varpi$ separately on each coordinate domain, even if $\varpi$ is defined on all of $M$. However, because of 3.2.1 the forms $d\varpi$ on any two coordinate domains agree on their intersection, so $d\varpi$ is really a single form defined on all of $M$ after all. In the future we shall take this kind of argument for granted.

3.2.4 **Example: exterior derivative in $\mathbb{R}^3$ and vector anaysis**.

0-forms (functions): $\varpi = f$, $\quad d\varpi = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$.

1-forms (covectors): $\varpi = Adx + Bdy + Cdz$,

$d\varpi = (\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z})dy \wedge dz - (\frac{\partial C}{\partial x} - \frac{\partial A}{\partial z})dz \wedge dx + (\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y})dx \wedge dy.$

2-forms: $\varpi = Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy,$

$d\varpi = (\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial dz})dxdydz.$

3-forms: $\varpi = Ddx \wedge dy \wedge dz, \quad d\varpi = 0.$

We see that the exterior derivative on $\mathbb{R}^3$ reproduces all three of the basic differentiation operation from vector calculus, i.e. gradient, the curl, and the divergence, depending on the kind of differential form it is applied to. Even in the special case of $\mathbb{R}^3$, the use of differential forms often simplies the calculation and clarifies the situation.

3.2.5 **Theorem (Product Rule).** *Let $\alpha, \beta$ be differential forms with $\alpha$ homogeneous of degree $|\alpha|$. Then*

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{|\alpha|}\alpha \wedge (d\beta).$$

**Proof**. By induction on $|\alpha|$ it suffices to prove this for $\alpha$ a 1-form, say $\alpha = a_k dx^k$. Put $\beta = b_{ij\cdots}dx^i \wedge dx^j \cdots$. Then

$$\begin{aligned} d(\alpha \wedge \beta) &= d(a_k b_{ij}\ dx^k \wedge dx^i \wedge dx^j \cdots) \\ &= \{(da_k)b_{ij} + a_i(db_{ij\cdots})\} \wedge dx^k \wedge dx^i \wedge dx^j \cdots \\ &= (da_k \wedge dx^k) \wedge (b_{ij}dx^i \wedge dx^j \cdots) - (a_i dx^k) \wedge (db_{ij\cdots} \wedge dx^i \wedge dx^j \cdots) \\ &= (d\alpha) \wedge \beta - \alpha \wedge (d\beta). \end{aligned}$$

□

**Remark**.. The exterior derivative operation $\varpi \mapsto d\varpi$ on differential forms $\varpi$ defined on open subsets of $M$ is uniquely characterized by the following properties.

a) If $f$ is a scalar function, then $df$ is the usual differential of $f$.
b) For any two forms $\alpha$, $\beta$, one has $d(\alpha + \beta) = d\alpha + d\beta$.
c) For any two forms $\alpha$, $\beta$ with $\alpha$ homogeneous of degree $p$ one has

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{|\alpha|}\alpha \wedge (d\beta).$$

This is clear since any differential form $f_{ij\cdots}\ dx^i \wedge dx^j \cdots$ on a coordinate domain can be built up from scalar functions and their differentials by sums and wedge products. Note that in this "axiomatic" characterization we postulate that $d$ operates also on forms defined only on open subsets. This postulate is natural, but actually not necessary.

3.2.6 **Theorem**. *For any differential form $\varpi$ of class $C^2$, $d(d\varpi) = 0$.*

**Proof**. First consider the case when $\varpi = f$ is a $C^2$ function. Then

$$dd\varpi = d(\frac{\partial f}{\partial x^i}dx^i) = \frac{\partial^2 f}{\partial x^j \partial x^i}dx^j \wedge dx^i$$

and this is zero, since the terms $ij$ and $ji$ cancel, because of the symmetry of the second partials. The general case is obtained by adding some dots, like

$\varpi = f_{\cdots}\cdots$ to indicate indices and wedge-factors, which are not affected by the argument. (Alternatively one can argue by induction.) □

3.2.7 **Corollary**. *If $\varpi$ is a differential form which can be written as an exterior derivative $\varpi = d\varphi$, then $d\varpi = 0$.*

This corollary has a local converse:

3.2.8 **Poincaré's Lemma.** *If $\varpi$ is a* smooth *differential from satisfying $d\varpi = 0$ then every point has a neighbourhood $U$ so that $\varpi = d\varphi$ on $U$ for some* smooth *form defined on $U$.*

However there may not be one single form $\varphi$ so that $\varpi = d\varphi$ on the whole manifold $M$. (This is true provided $M$ is *simply connected*, which means intuitively that every closed path in $M$ can be continuously deformed into a point.)

3.2.9 **Definition (pull-back of forms)**. Let $F : N \to M$ be a smooth mapping between manifolds. Let $(y^1, \cdots, y^m)$ be a coordinate system on $N$ and $(x^1, \cdots, x^n)$ a coordinate system on $M$. Then $F$ is given by equations $x^i = F^i(y^1, \cdots, y^m)$, $i = 1, \cdots, n$. Let $\varpi = g_{ij\cdots}(x^1, \cdots, x^n)dx^i \wedge dx^j \cdots$ be a differential form on $M$,. Then $F^*\varpi = h_{kl\cdots}(y^1, \cdots, y^m)dy^k \wedge dy^l \cdots$ is the differential form on $N$ obtained from $\varpi$ by the substitution

$$x^i = F^i(y^1, \cdots, y^m),\ dx^i = \frac{\partial F^i}{\partial y^j}dy^j. \tag{*}$$

$F^*\varpi$ is called the *pull-back* of $\varpi$ by $F$.

3.2.10 **Lemma**. *The differential form $F^*\varpi$ is independent of the coordinate systems $(x^i), (y^j)$.*

**Proof**. The equations (*) are those defining $dF$. So in terms of the alternating multilinear function $\varpi(v, w, \cdots)$ the definition of $F^*\varpi$ amounts to

$$(F^*\varpi)(v, w, \cdots) = \varpi(dF(v), dF(w), \cdots)$$

from which the independence of coordinates is evident. □

3.2.11 **Remark**. *Any* $(0, k)$-tensor $T$ on $M$ (not necessarily alternating) can be written as a linear combination of tensor products $dx^i \otimes dx^j \otimes \cdots$ of the coordinate differentials $dx^i$. By the same procedure just used one can define a $(0, k)$-tensor $F^*T$ on $N$.

3.2.12 **Theorem**. *The pull-back operation has the following properties*
*(a)* $F^*(\varpi_1 + \varpi_2) = F^*\varpi_1 + F^*\varpi_2$
*(b)* $F^*(\varpi_1 \wedge \varpi_2) = (F^*\varpi_1) \wedge (F^*\varpi_2)$
*(c)* $d(F^*\varpi) = F^*(d\varpi)$
*(d)* $(G \circ F)^*\varpi = G^*(F^*\varpi)$ *(if the composite $G \circ F$ makes sense.)*

**Proof**. Exercise.

3.2.13 **Examples**.

(a) Let $\varpi$ be the 1-form on $\mathbb{R}^2$ given by $\varpi = xdy$ in Cartesian coordinates $x, y$ on $\mathbb{R}^2$. Let $F : \mathbb{R} \to \mathbb{R}^2$, $t \to F(t) = (x, y)$ be the map given by $x = \sin t$, $y = \cos t$. Then $F^*\varpi$ is the 1-form on $\mathbb{R}$ given by $F^*\varpi = F^*(xdy) = \sin t \, d \cos t = -\sin^2 t dt$.
(b) Let $\varpi$ be the 1-form on $\mathbb{R}$ given by $\varpi = dt$ in the Cartesian coordinate $t$ on $\mathbb{R}$. Let $F : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \to F(x, y) =$t be the map given by $t = (x - y)^2$. Then $F^*\varpi = F^*(dt) = d(x - y)^2 = 2(x - y)(dx - dy)$.
(c) The symmetric (0,2)-tensor $g$ representing the Euclidean metric $dx^2 + dy^2 + dz^2$ on $\mathbb{R}^3$ has components $(\delta_i^j)$ in Cartesian coordinates. It can be written as $g = dx \otimes dx + dy \otimes dy + dz \otimes dz$ . Let $S^2 = \{(x, y, z) \in \mathbb{R}^2 \mid x^2 + y^2 + z^2 = R^2\}$ be the sphere (a submanifold of $\mathbb{R}^3$) and $F : S^2 \to \mathbb{R}^3$ the inclusion. Then $F^*g$ is the symmetric $(0, 2)$-tensor on $S^2$ which represents the Riemann metric on $S^2$ obtained from the Euclidean metric on $\mathbb{R}^3$ by restriction to $S^2$. ($F^*g$ is defined in accordance with remark 3.2.11)
(d) Let $T$ be any $(0, k)$-tensor on $M$. $T$ can be thought of as a multilinear function $T(u,\mathrm{v}, \cdots)$ on vectors on $M$. Let $S$ be a submanifold of $M$ and $i : S \to M$ the inclusion mapping $S$. The pull-back $i^*T$ of $T$ by $i$ is just the restriction of $T$ to tangent vectors to $S$. This means that $i^*T(u, v, \cdots) = T(u, v, \cdots)$ when $u, v, \cdots$ are tangent vectors to $S$. The tensor $i^*T$ is also denoted $T|_S$, called the *restriction* of $T$ to $S$. (This works only for tensors of type $(0, k)$.)

## Differential calculus on a Riemannian manifold

From now on we assume given a Riemann metric $ds^2 = g_{ij}dx^i dx^j$. Recall that the Riemann metric gives the star operator on differential forms.

3.2.14 **Definition**. Let $\varpi$ be a k-form. Then the $\delta\varpi$ is the $(k-l)$-form defined by

$$*\delta\varpi = d * \varpi.$$

One can also write

$$\delta\varpi = *^{-1} d * \varpi = (-1)^{(n-k+1)(k-1)} * \mathrm{d} * \varpi.$$

3.2.15 **Example: more vector analysis in** $\mathbb{R}^3$. Identify vectors and covectors on $\mathbb{R}^3$ using the Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$, so that the 1-form $F = F_x dx + F_y dy + F_z dz$ is identified with the vector field $F_x(\partial/\partial x) + F_y(\partial/\partial y) + F_z(\partial/\partial z)$. Then

$$*dF = \mathrm{curl} F \quad \text{and} \qquad d * F = *\mathrm{div} F.$$

The second equation says that $\delta F =$div$F$. Summary: if 1-forms and 2-forms on $\mathbb{R}^3$ are identified with vector fields and 3-forms with scalar functions (using the metric and the star operator) then $d$ and $\delta$ become curl and $d$iv, respectively. If $f$ is a scalar function, then

$$\delta df = \mathrm{div}\,\mathrm{grad} f = \Delta f,$$

the Laplacian of $f$.

**EXERCISES** 3.2

1. Let $f$ be a $C^2$ function (0-form). Prove that $d(df)=0$.

2. Let $\varpi=\sum_{i<j} f_{ij}dx^i\wedge dx^j$. Write $d\varpi$ in the form

$$d\varpi=\sum_{i<j<k} f_{ijk}dx^i\wedge dx^j\wedge dx^k$$

(Find $f_{ijk}$.)

3. Let $\varphi=f_i dx^i$ be a 1-form. Find a formula for $\delta\varphi$.

4. Prove the assertion of Remark 2, assuming *Theorem* 3.2.1.

5. Prove parts (a) and (b) of *Theorem* 3.2.11.

6. Prove part (c) of *Theorem* 3.2.11.

7. Prove part (d) of *Theorem* 3.2.11.

8. Prove from the definitions the formula

$$i^*T(u,v,\cdots)=T(u,v,\cdots)$$

of *Example* 3.2.15 (d).

9. Let $\varpi=\mathrm{P}dx+\mathrm{Q}dy+Rdz$ be a 1-form on $\mathbb{R}^3$.
(a) Find a formula for $\varpi$ in cylindrical coordinates $(r,\theta,z)$.
(b) Find a formula for $d\varpi$ in cylindrical coordinates $(r,\theta,z)$.

10. Let $x^1,x^2,x^3$ be an *orthogonal coordinate system* in $\mathbb{R}^3$, i.e. a coordinate system so that Euclidean metric $ds^2$ becomes "diagonal":

$$ds^2=(h_1dx^1)^2+(h_2dx^2)^2+(h_3dx^3)^2$$

for certain positive functions $h_1,h_2,h_3$. Let $\varpi=A_1dx^1+A_2dx^2+A_3dx^3$ be a 1-form on $\mathbb{R}^3$.
(a) Find $*d\varpi$. (b) Find $d*\varpi$.

11. Let $\varpi=\sum_{i<j} f_{ij}dx^i\wedge dx^j$. *Use the definition* 3 to prove that
$d\varpi=\sum_{i<j<k}\Big(\frac{\partial f_{ij}}{\partial x^k}+\frac{\partial f_{jk}}{\partial x^i}+\frac{\partial f_{ki}}{\partial x^j}\Big)dx^i\wedge dx^j\wedge dx^k$.

12. a) Let $\varpi=f_idx^i$ be a smooth 1-form. Prove that $\delta\varpi=-\frac{1}{\gamma}\frac{\partial(\gamma f^i)}{\partial x^i}$ where $\gamma=\sqrt{|\det g_{ij}|}$. Deduce that
$\delta df=-\frac{1}{\gamma}\frac{\partial}{\partial x^i}\Big(\gamma g^{ij}\frac{\partial f}{\partial x^j}\Big)$
[The operator $\Delta: f\to-\delta df$ is called the *Laplace-Beltrami* operator of the Riemann metric.]

13. Use problem 12 to write down $\Delta f=\delta df$

(a) in cylindrical coordinates $(r, \theta, z)$ on $\mathbb{R}^3$,
(b) in spherical coordinates $(\rho, \theta, \phi)$ on $\mathbb{R}^3$,
(b) in geographical coordinates $(\theta, \phi)$ on $S^2$.

14. Identify vectors and covectors on $\mathbb{R}^3$ using the Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$, so that the 1-form $F_x dx + F_y dy + F_z dz$ is identified with the vector field $F_x(\partial/\partial x) + F_y(\partial/\partial y) + F_z(\partial/\partial z)$. Show that
a) $*dF = $curl$F$   b) $d * F = *$div$F$
for any 1-form $F$, as stated in *Example* 3.2.15

15. Let $F = F_r(\partial/\partial r) + F_\theta(\partial/\partial\theta) + F_z(\partial/\partial z)$ be a vector field in cylindrical coordinates $(r, \theta, z)$ on $\mathbb{R}^3$. Use problem 14 to find a formula for curl$F$ and div$F$.

16. Let $F = F_\rho(\partial/\partial\rho) + F_\theta(\partial/\partial\theta) + F_\phi(\partial/\partial\phi)$ be a vector field in spherical coordinates $(\rho, \theta, \phi)$ on $\mathbb{R}^3$. Use problem 14 to find a formula for curl$F$ and div$F$.

17. Find all radial solutions $f$ to Laplace's equation $\Delta f = 0$ in $\mathbb{R}^2$ and in $\mathbb{R}^3$. ("Radial" means $f(p) = f(r)$. You may use problem 12 to find a formula of $\Delta f$.)

## 3.3 Integral calculus

It is not possible to define the integral of a scalar functions $f$ on a manifold in coordinates $(x^i)$ as

$$\int \cdots \int f(x^1, \cdots, x^n) dx^1 \cdots dx^n$$

if one wants the value of the integral to come out independent of the coordinates. The reason for this is the Change of Variables formula form calculus.

3.3.1 **Theorem (Change of Variables Formula)**. *Let $\tilde{x}^j = F^j(x^1, \cdots, x^n)$ be a* smooth *mapping $\mathbb{R}^n \to \mathbb{R}^n$ which maps an open set $U$ one-to-one onto an open set $\tilde{U}$. Then for any integrable function $f$ on $U$,*

$$\int \cdots \int f d\tilde{x}^1 \cdots d\tilde{x}^n = \int \cdots \int f |\det \frac{\partial \tilde{x}^i}{\partial x^j}| dx^1 \cdots dx^n.$$

The domain of integration $\tilde{U}$ on the left corresponds to the domain of integration $U$ on the right under the mapping $F$: $\tilde{U} = F(U)$. On the left, the function $f$ is expressed in terms of the $\tilde{x}^j$ by means of the mapping $\tilde{x}^i = F^i(x^1, \cdots, x^n)$. –You may consult your Calculus text for the proof (e.g. Marsden-Tromba), at least for the case $n = 1, 2, 3$.

From now on $(x^1, \cdots, x^n)$ denotes again a positively oriented coordinate system on an $n$–dimensional oriented manifold $M$.

3.3.2 **Theorem and definition**. *Let $M$ be an oriented $n$–dimensional manifold, $(x^i)$ a positively oriented coordinate system defined on an open set $U$. For any $n$–form $\varpi = f dx^1 \wedge \cdots \wedge dx^n$ on $U$ the integral*

$$\int_U \varpi := \int \cdots \int f\, dx^1 \cdots dx^n$$

*is independent of the coordinates $(x^i)$.*

**Proof**. Let $(\tilde{x}^j)$ be another coordinate system. Write $f dx^1 \wedge \cdots \wedge dx^n = \tilde{f} d\tilde{x}^1 \wedge \cdots \wedge d\tilde{x}^n$. Then $f = \tilde{f} \det(\partial \tilde{x}^j / \partial x^i)$. Therefore

$$\begin{aligned}
\int \cdots \int f\, dx^1 \wedge \ \cdots \wedge dx^n &= \int \cdots \int \tilde{f} \det \frac{\partial \tilde{x}^j}{\partial x^i}\, dx^1 \ \cdots dx^n \\
&= \pm \int \cdots \int \tilde{f} |\det \frac{\partial \tilde{x}^j}{\partial x^i}|\, dx^1 \ \cdots dx^n \\
&= \pm \int \cdots \int \tilde{f}\, d\tilde{x}^1 \cdots d\tilde{x}^n
\end{aligned}$$

□

This theorem–definition presupposes that the integral on the right exists. This is certainly the case if $f$ is continuous and the $(x^i(p))$, $p \in U$, lie in a bounded subset of $\mathbb{R}^n$. It also presupposes that the region of integration lies in the domain of the coordinate system $(x^i)$. If this is not the case the integral is defined by subdividing the region of integration into sufficiently small pieces each of which lies in the domain of a coordinate system, as one does in calculus. (See the appendix for an alternative procedure.)

3.3.3 **Example**. Suppose $M$ has a Riemann metric $g$. The volume element is the $n$–form

$$\mathrm{vol}_g = |\det g_{ij}|^{1/2} dx^1 \wedge \cdots \wedge dx^n.$$

It is independent of the coordinates up to sign. Therefore the positive number

$$\int_U \mathrm{vol}_g = \int \cdots \int |\det g_{ij}|^{1/2} dx^1 \cdots dx^n.$$

is independent of the coordinate system. It is called the volume of the region $U$. (When $\dim M = 1$ it is naturally called *length*, $\dim M = 2$ it is called area.) More generally, if $f$ is a scalar valued function, then $f\mathrm{vol}_g$ is an $n$–form, and

$$\int_U f\mathrm{vol}_g = \int \cdots \int f |\det g_{ij}|^{1/2} dx^1 \cdots dx^n.$$

is independent of the coordinates up to sign. It is called the integral of $f$ with respect to the volume element $\mathrm{vol}_g$.

3.3.4 **Definition**. Let $\omega$ be an m-form on $M$ ($m \le \dim M$), $S$ an $m$-dimensional submanifold of $M$. Then

$$\int_S \omega = \int_S \omega|_S$$

where $\omega|_S$ is the restriction of $\omega$ to $S$.

**Explanation**. Recall that the restriction $\omega|_S$ is the pull-back of $\omega$ by the inclusion map $S \to M$, an $m$-form on $S$. Since $m = \dim S$, the integral on the right is defined.

3.3.5 **Example: integration of forms on $\mathbb{R}^3$.**

0-*forms*. A 0-form is a function $f$. A zero dimensional manifold is a discrete set of points $\{p_i\}$. The "integral" of $f$ over $\{p_i\}$ is the sum $\sum_i f(p_i)$.The sum need not be finite. The integral exists provided the series converges.

1-*forms*. A 1-form on $\mathbb{R}^3$ may be written as $Pdx + Qdy + Rdz$. Let $C$ be a 1-dimensional submanifold of $\mathbb{R}^3$. Let $t$ be a coordinate on $C$. Let $p = p(t)$ be the point on $C$ with coordinate $t$ and write $x = x(t), y = y(t), z = z(t)$ for the Cartesian coordinates of $p(t)$. Thus $C$ can be considered a curve in $\mathbb{R}^3$ parametrized by $t$. The restriction of $Pdx + Qdy + Rdz$ to $C$ is

$$P\frac{dx}{dt} + Q\frac{dy}{dt} + R\frac{dz}{dt}$$

where $P, Q, R$ are evaluated at $p(t)$. Thus

$$\int_C Pdx + Qdy + Rdz = \int_C (P\frac{dx}{dt} + Q\frac{dy}{dt} + R\frac{dz}{dt})dt,$$

the usual line integral. If the coordinate $t$ does not cover all of $C$, then the integral must be defined by subdivision as remarked earlier.

2-*forms*. A 2-form on $\mathbb{R}^3$ may be written as $Ady \wedge dz - Bdx \wedge dz + Cdx \wedge dy$. Let $S$ be a 2-dimensional submanifold of $\mathbb{R}^3$. Let $(u, v)$ be coordinates on $S$. Let $p = p(u, v)$ be the point on $S$ with coordinates $(u, v)$ and write $x = x(u, v), y = y(u, v), z = z(u, v)$ for the Cartesian coordinates of $p(u, v)$. Thus $S$ can be considered a surface in $\mathbb{R}^3$ parametrized by $(u, v) \in D$. The restriction of $Ady \wedge dz - Bdx \wedge dz + Cdx \wedge dy$ to $S$ is

$$\Big\{A\Big(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial y}{\partial v}\Big) - B\Big(\frac{\partial x}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial x}{\partial v}\Big) + C\Big(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\Big)\Big\}du \wedge dv.$$

In calculus notation this expression is $V \cdot N$ where $V = Ai + Bj + Ck$ is the vector field corresponding to the given 2-form and

$$N = \frac{\partial p}{\partial u} \times \frac{\partial p}{\partial v} = \Big(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial y}{\partial v}\Big)i - \Big(\frac{\partial x}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial x}{\partial v}\Big)j + \Big(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\Big)k.$$

$N$ is the normal vector corresponding to the parametrization of the surface. Thus the integral above is the usual surface integral:

$$\int_S Ady \wedge dz - Bdx \wedge dz + Cdx \wedge dy = \iint\limits_D V \cdot N \, du \, dv.$$

If the coordinates $(u, v)$ do not cover all of $S$, then the integral must be defined by subdivision as remarked earlier.

3-*forms*. A 3-form on $\mathbb{R}^3$ may be written as $f dx \wedge dy \wedge dz$ and its integral is the usual triple integral

$$\int_U f\, dx \wedge dy \wedge dz = \iiint\limits_U f\, dx\, dy\, dz.$$

3.3.6 **Definitions**. Let $S$ be a $k$–dimensional manifold, $R$ subset of $S$ with the following property. $R$ can be covered by finitely many open *coordinate cubes* $Q$, consisting of all points $p \in S$ whose coordinates of $p$ in a given coordinate system $(x^1, \cdots, x^k)$ satisfy $-1 < x^i < 1$, in such a way that each $Q$ is either completely contained in $R$ or else intersects $R$ in a half-cube

$$R \cap Q = \{\text{all points in } Q \text{ satisfying } x^i \le 0\}. \tag{*}$$

The *boundary* of $R$ consists of all points in the in the sets (*) satisfying $x^1 = 0$. It is denoted $\partial R$. If $S$ is a submanifold of $M$, then $R$ is also called a *bounded k-dimensional submanifold* of $M$.

**Remarks**. a) It may happen that $\partial R$ is empty. In that case $R$ is itself a k-dimensional submanifold and we say that $R$ is *without boundary*.
b) Instead of cubes and half-cubes one can also use balls and half-balls to define "bounded submanifold".

3.3.7 **Lemma.** *The boundary $\partial R$ of a bounded k-dimensional submanifold is a bounded k-1-dimensional submanifold without boundary with coordinates the restrictions to $\partial R$ of the $n-1$ last coordinates $x^2, \cdots, x^n$ of the coordinates $x^1, \cdots, x^n$ on $S$ of the type entering into (*).*

**Proof**. Similar to the proof for submanifolds without boundary. □

3.3.8 **Conventions**. (a) On an oriented $n$–dimensional manifold only positively oriented coordinates are admitted in the definition of the integral of a $n$–form. This specifies the ambiguous sign $\pm$ in the definition of the integral.
(b) Any $n$-form $\varpi$ is called *positive* if on the domain of any positive coordinate system $(x^i)$

$$\varpi = D dx^1 \wedge \cdots \wedge dx^n \text{ with } D > 0.$$

Note that one can also specify an orientation by saying which $n$-forms are positive.

3.3.9 **Lemma and definition**. *Let $R \subset S$ be a bounded $k$–dimensional submanifold. Assume that $S$ is oriented. As $(x^1, \cdots, x^n)$ runs over the positively oriented coordinate systems on $S$ of the type referred to in Lemma* 3.3.7*, the corresponding coordinate systems $(x^2, \cdots, x^n)$ on $\partial R$ define an orientation on $\partial R$, called the induced orientation.*

*Proof.* Let $(x^i)$ and $(\tilde{x}^i)$ be two positively oriented coordinate systems of this type. Then on $\partial R$ we have $x^1 = 0$ and $\tilde{x}^1 = 0$, so if the $\tilde{x}^i$ are considered as functions of the $x^i$ by the coordinate transformation, then

$$0 \equiv \tilde{x}^1 = \tilde{x}^1(0, x^2, \cdots, x^k) \qquad \text{on } \partial R$$

Thus $\partial\tilde{x}^1/\partial x^2 = 0, \cdots, \partial\tilde{x}^1/\partial x^k = 0$ on $\partial R$. Expanding $\det(\partial\tilde{x}^i/\partial x^j)$ along the "row" $(\partial x^1/\partial x^j)$ one finds that

$$\det\Big(\frac{\partial\tilde{x}^i}{\partial x^j}\Big)_{1\leq ij\leq k} = \Big(\frac{\partial\tilde{x}^1}{\partial x^1}\Big)\det\Big(\frac{\partial\tilde{x}^i}{\partial x^j}\Big)_{2\leq ij\leq k} \qquad \text{on } \partial R.$$

The LHS is $> 0$, since $(x^i)$ and $(\tilde{x}^i)$ are positively oriented. The derivative $\partial x^1/\partial\tilde{x}^1$ cannot be $< 0$ on $\partial R$, since $\tilde{x}^1 = \tilde{x}^1(x^1, x^2_o, \cdots x^k_o)$ is $= 0$ for $x^1 = 0$ and is $< 0$ for $x^1 < 0$. It follows that in the above equation the first factor $(\partial x^1/\partial\tilde{x}^1)$ on the right is $> 0$ on $\partial R$, and hence so is the second factor, as required. □

3.3.10 **Example**. Let $S$ be a 2-dimensional submanifold of $\mathbb{R}^3$. For each $p \in S$, there are two unit vectors $\pm n(p)$ orthogonal to $T_pS$. Suppose we choose one of them, say $n(p)$, depending continuously on $p \in S$. Then we can specify an orientation on $S$ by stipulating that a coordinate system $(u, v)$ on $S$ is positive if

$$\frac{\partial p}{\partial u} \times \frac{\partial p}{\partial v} = Dn \text{ with } D > 0. \qquad (*)$$

Let $R$ be a bounded submanifold of $S$ with boundary $C = \partial R$. Choose a positive coordinate system $(u, v)$ on $S$ around a point of $C$ as above, so that $u < 0$ on $R$ and $u = 0$ on $C$. Then $p = p(0, v)$ defines the positive coordinate $v$ on $C$ and $\partial p/\partial v$ is the tangent vector along $C$. Along $C$, the equation(*) amounts to this. If we walk upright along $C$ (head in the direction of $n$) in the positive direction (direction of $\partial p/\partial v$), then $R$ is to our left ($\partial p/\partial u$ points outward from $R$, in the direction of increasing $u$, since the $u$-component of $\partial p/\partial u = \partial/\partial u$ is $+1$).

The next theorem is a kind of *change of variables formula* for integrals of forms.

3.3.11 **Theorem**. *Let $F : M \to N$ be a diffeomorphism of $n$–dimensional manifolds. For any $m$–dimensional oriented bounded submanifold $R$ of $N$ and any $m$–form $\varpi$ on $N$ one has*

$$\int_R F^*\varpi = \int_{F(R)} \varpi$$

*if $F(R)$ is given the orientation corresponding to the orientation on $R$ under the diffeomorphism $F$.*

**Proof**. Replacing $M$ by the $m$–dimensional submanifold $S$ containing $R$ and $N$ by $F(S)$ we may assume that $m = n$ in the first place. By subdivision we may assume that $R$ is contained in the domain of coordinates $(x^i)$ on $M$ and $F(R)$ in the domain of coordinates $(\tilde{x}^j)$ on $N$. Then we can write $\varpi =$

$fdx^1 \wedge \cdots \wedge dx^n$, $\tilde{x}^j = F^j(x^1, \cdots, x^n)$ and the formula reduces to the usual change of variables formula 3.3.1. $\square$

3.3.12 **Theorem (Stokes's Formula)**. *Let $\omega$ be a* smooth *$k$-form on $M$, $R$ a $k$-dimensional oriented bounded submanifold of $M$. Then*

$$\int_R d\omega = \int_{\partial R} \omega.$$

**Proof**. We first prove this for the special case of a $n$-cube $I^n$ in $\mathbb{R}^n$. Let

$$I^n = \{(x^1, \cdots, x^n) \in \mathbb{R}^n \mid 0 \leq x^j \leq 1\}.$$

We specify the orientation so that the Cartesian coordinates $(x^1, \cdots, x^n)$ form a positively oriented coordinate system. We have $\partial I^n = \bigcup_{\mathrm{j}=1}^n (I_j^0 \cup I_j^1)$ where

$I_j^0 = \{(x^1, \cdots, x^n) \in I^n \mid x^j = 0\}$, $\qquad I_j^1 = \{(x^1, \cdots, x^n) \in I^n \mid x^j = 1\}$.

$$\begin{tikzcd} * \arrow[d, "I_1^0"'] & * \arrow[l, "I_2^1"'] \arrow[d, "I_1^1"] \\ * \arrow[r, "I_2^0"'] & * \end{tikzcd}$$

Near $I_j^0$ the points of $I^n$ satisfy $x_j \geq 0$. If one uses $x^j$ as first coordinate, one gets a coordinate system $(x^j, x^1, \cdots [x^j] \cdots x^n)$ on $I^n$ whose orientation is positive or negative according to the sign of $(-1)^{j-1}$. The coordinate system $(-x^j, x^1, \cdots [x^j] \cdots x^n)$ on $I^n$ is of the type required in 3.3.11 for $R = I^n$ near a point of $S = I_j^0$. It follows that $(x^1, \cdots [x^j] \cdots x^n)$ is a coordinate system on $I_j^0$ whose orientation (specified according to 3.3.11) is positive or negative according to the sign of $(-1)^j$. Similarly, $(x^1, \cdots [x^j] \cdots x^n)$ is a coordinate system on $I_j^0$ whose orientation (specified according to 3.3.11) is positive or negative according to the sign of $(-1)^{\mathrm{j}-1}$. We summarize this by specifying the required sign as follows

Positive $n$–form on $I^n$: $dx^1 \wedge \cdots \wedge dx^n$

Positive (n$-$1)-form on $I_j^0$: $(-1)^j dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n$

Positive (n$-$1)-form on $I_j^1$: $(-1)^{\mathrm{j}-1} dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n$

A general (n$-$1)-form can be written as

$$\omega = \sum_{\mathrm{j}} f_j dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n.$$

To prove Stokes formula it therefore suffices to consider

$$\omega_j = f_j dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n.$$

Compute

$\int_{I^n} d\omega_j = \int_{I^n} \frac{\partial f_j}{\partial x^j} dx^j \wedge dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n$

[definition of $d\omega$]

$= \int_{I^n} \frac{\partial f_j}{\partial x^j} (-1)^{j-1} dx^1 \wedge \cdots \wedge dx^n$

[move $dx^j$ in place]

$$
\begin{aligned}
&= \int_0^1 \cdots \int_0^1 \tfrac{\partial f_j}{\partial x^j} (-1)^{j-1} dx^1 \cdots dx^n \\
&\hfill \text{[repeated integral: definition 3.3.2]} \\
&= \int_0^1 \cdots \int_0^1 \left\{ \int_0^1 \tfrac{\partial f_j}{\partial x^j} (-1)^{j-1} dx^j \right\} dx^1 \cdots [dx^j] \cdots dx^n \\
&\hfill \text{[integrate first over } dx^j\text{]} \\
&= \int_0^1 \cdots \int_0^1 \Big( f_j \Big)\Big|_{x_j=0}^1 (-1)^{j-1} dx^1 \cdots [dx^j] \cdots dx^n \\
&\hfill \text{[Fundamental Theorem of Calculus]} \\
&= \int_0^1 \cdots \int_0^1 f_j|_{x_j=1} (-1)^{j-1} dx^1 \cdots [dx^j] \cdots dx^n + \\
&\qquad + \int_0^1 \cdots \int_0^1 f_j|_{x_j=0} (-1)^{j} dx^1 \cdots [dx^j] \cdots dx^n \\
&= \int_{I_j^1} f_j dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n + \int_{I_j^0} f_j dx^1 \wedge \cdots [dx^j] \cdots \wedge dx^n \\
&\hfill \text{[definition 3.3.2]} \\
&= \int_{\partial I^n} \omega_j \\
&\hfill [\textstyle\int_{I_k^0} \omega_j = 0 \text{ if } k \neq j, \text{ because } dx^k = 0 \text{ on } I_k^0,\, I_k^1]
\end{aligned}
$$

This proves the formula for a cube. To prove it in general we need to appeal to a theorem in topology, which implies that any bounded submanifold can be subdivided into a finite number of coordinate-cubes, a procedure familiar from surface integrals and volume integrals in $\mathbb{R}^3$. (A coordinate cube is a subset of $M$ which becomes a cube in a suitable coordinate system).

□

**Remarks**. (a) Actually, the solid cube $I^n$ is not a bounded submanifold of $\mathbb{R}^n$, because of its edges and corners (intersections of two or more faces $I_j^{0,1}$). One way to remedy this sort of situation is to argue that $R$ can be approximated by bounded submanifolds (by rounding off edges and corners) in such a way that both sides of Stokes's formula approach the desired limit.
(b) There is another approach to integration theory in which integrals over a bounded $m$-dimensional submanifolds are replaced by integrals over formal linear combination (called "chains") of $m$-cubes $\gamma = \sum_k c_k \gamma_k$. Each $\gamma_k$ is a map $\gamma_k : Q \to M$ of the standard cube in $\mathbb{R}^m$ and the integral of an $m$-form $\varpi$ over $\gamma$ is simply defined as

$$\int_\gamma \varpi = \sum c_k \int_Q \gamma_k^* \omega.$$

The advantage of this procedure is that one does not have to worry about subdivisions into cubes, since this is already built in. This disadvantage is that it is often much more natural to integrate over bounded submanifolds rather than over chains. (Think of the surface and volume integrals from calculus, for example.)

## Appendix: Partition of Unity

We fix an $n$–dimensional manifold $M$. The definition of the integral by a subdivision of the domain of integration into pieces contained in coordinate domains is awkward for proofs. There is an alternative procedure, which cuts up the form rather than the domain, which is more suitable for theoretical considerations. It is based on the following lemma.

**Lemma.** *Let $D$ be a compact subset of $M$. There are continuous non–negative functions $g_1, \cdots, g_l$ on $M$ so that*

$$g_1 + \cdots + g_l \equiv 1 \text{ on } D, \; g_k \equiv 0 \text{ outside of some coordinate ball } B_k.$$

**Proof**. For each point $p \in M$ one can find a continuous function $h_p$ so that $h_p \equiv 0$ outside a coordinate ball $B_p$ around $p$ while $h_p > 0$ on a smaller coordinate ball $C_p \subset B_p$. Since $D$ is compact, it is covered by finitely many of the $C_p$, say $C_1, \cdots, C_l$. Let $h_1, \cdots, h_l$ be the corresponding functions $h_p$, $B_1, \cdots, B_l$ the corresponding coordinate balls $B_p$. The function $\sum h_k$ is $> 0$ on $D$, hence one can find a strictly positive continuous function $h$ on $M$ so that $h \equiv \sum h_k$ on $D$ (e.g. $h = \max\{\epsilon, \sum h_k\}$ for sufficiently small $\epsilon > 0$). The functions $g_k = h_k/h$ have all the required properties. □

Such a family of functions $\{g_k\}$ will be called a *(continuous) partition of unity on $D$* and we write $1 = \sum g_k$ (on $D$).

We now define integrals $\int_D \varpi$ without cutting up $D$. Let $\varpi$ be an $n$–form on $M$. Exceptionally, we do not require that it be smooth, only that locally $\varpi = f dx^1 \wedge \cdots \wedge dx^n$ with $f$ *integrable* (in the sense of Riemann or Lebesgue, it does not matter here). We only need to define integrals over all of $M$, since we can always make $\varpi \equiv 0$ outside of some domain $D$ and the integral $\int_M \varpi$ is already defined if $\varpi \equiv 0$ outside of some coordinate domain.

**Proposition and definition**. *Let $\varpi$ be an $n$–form on $M$ vanishing outside of a compact set $D$. Let $1 = \sum g_k$ (on $D$) be a continuous partition of unity on $D$ as in the lemma. Then the number*

$$\int_{B_1} g_1 \varpi + \cdots + \int_{B_l} g_l \varpi$$

*is independent of the partition of unity chosen and is called the* integral *of $\varpi$ over $M$, denoted $\int_M \varpi$.*

**Proof**. Let $\{g_i\}$ and $\{\tilde{g}_j\}$ be two partitions of unity for $D$. Then on $D$,

$$\varpi = \sum_i g_i \varpi = \sum_j \tilde{g}_j \varpi$$

Multiply through by $g_k$ for a fixed $k$ and integrate to get:

$$\int_M \sum_i g_k g_i \varpi = \int_M \sum_j g_k \tilde{g}_j \varpi$$

Since $g_k$ vanishes outside of a coordinate ball, the additivity of the integral on $\mathbb{R}^n$ gives

$$\sum_i \int_M g_k g_i \varpi = \sum_j \int_M g_k \tilde{g}_j \varpi.$$

Now add over $k$:

$$\sum_{i,k}\int_M g_k g_i \varpi = \sum_{j,k}\int_M g_k \tilde{g}_j \varpi.$$

Since each $g_i$ and $\tilde{g}_j$ vanishes outside a coordinate ball, the same additivity gives

$$\sum_i \int_M \sum_k g_k g_i \varpi = \sum_j \int_M \sum_k g_k \tilde{g}_j \varpi.$$

Since $\sum_k g_k = 1$ on $D$ this gives

$$\sum_i \int_M g_i \varpi = \sum_j \int_M \tilde{g}_j \varpi$$

as required. □

Thus the integral $\int_M \varpi$ is defined whenever $\varpi$ is locally integrable and vanishes outside of a compact set.

**Remark**. Partitions of unity $1 = \sum g_k$ as in the lemma can be found with the $g_k$ even smooth. But this is not required here.

**EXERCISES** 3.3

1. Verify in detail the assertion in the proof of Theorem 3.3.11 that "the formula reduces to the usual change of variables formula 3.3.1". Explain how the orientation on $F(R)$ is defined.

2. Let C be the parabola with equation $y = 2x^2$.
(a) Prove that C is a submanifold of $\mathbb{R}^2$.
(b) Let $\omega$ be the differential 2-form on $\mathbb{R}^2$ defined by

$$\omega = 3xydx + y^2dy.$$

Find $\int_U \omega$ where $U$ is the part of C between the points $(0, 0)$ and $(1, 2)$.

3. Let $S$ be the cylinder in $\mathbb{R}^3$ with equation $x^2 + y^2 = 16$.
(a) Prove that $S$ is a submanifold of $\mathbb{R}^3$.
(b) Let $\omega$ be the differential 1-form on $\mathbb{R}^3$ defined by

$$\omega = zdy \wedge dz - xdx \wedge dz - 3y^2zdx \wedge dy.$$

Find $\int_U \omega$ where $U$ is the part of $S$ between the planes $z = 2$ and $z = 5$.

4. Let $\omega$ be the 1-form in $\mathbb{R}^3$ defined by

$$\omega = (2x - y)dx - yz^2dy - y^2zdz.$$

Let $S$ be the sphere in $\mathbb{R}^3$ with equation $x^2 + y^2 + z^2 = 1$ and C the circle $x^2 + y^2 = 1, z = 0$ in the $xy$-plane. Calculate the following integral from the definitions given in this section. Explain the result in view of Stokes's Theorem.

(a) $\int_C \omega$.
(b) $\int_{S_+} d\omega$, where $S_+$ is half of $S$ in $z \geq 0$.

5. Let $S$ be the surface in $\mathbb{R}^3$ with equation $z = f(x, y)$ where $f$ is a smooth function.
(a) Show that $S$ is a submanifold of $\mathbb{R}^3$.
(b) Find a formula for the Riemann metric $ds^2$ on $S$ obtained by restriction of the Euclidean metric in $\mathbb{R}^3$.
(c) Show that the *area* (see *Example* 3.3.3) of an open subset $U$ of $S$ is given by the integral

$$\int_U \sqrt{\Big(\frac{\partial f}{\partial x}\Big)^2 + \Big(\frac{\partial f}{\partial y}\Big)^2 + 1}\; dx \wedge dy\ .$$

6. Let $S$ be the *hypersurface* in $\mathbb{R}^n$ with equation $F(x^1, \cdots, x^n) = 0$ where $F$ is a smooth function satisfying $(\partial F/\partial x^n)_p \neq 0$ at all points $p$ on $S$.
(a) Show that $S$ is an (n−1)-dimensional submanifold of $\mathbb{R}^n$ and that $x^1, \cdots, x^{\mathrm{n}-1}$ form a coordinate system in a neighbourhood of any point of $S$.
(b) Show that the Riemann metric $ds^2$ on $S$ obtained by restriction of the Euclidean metric in $\mathbb{R}^n$ in the coordinates $x^1, \cdots, x^{\mathrm{n}-1}$ on $S$ is given by $g_{ij} = \delta_{ij} + F_n^{-2} F_i F_j$ $(1 \leq i, j \leq n-1)$ where $F_i = \partial F / \partial x^i$.
(c) Show that the volume (see E*x*ample 3.3.3) of an open subset $U$ of $S$ is given by the integral

$$\int_U \frac{\sqrt{\Big(\frac{\partial F}{\partial x^1}\Big)^2 + \cdots + \Big(\frac{\partial F}{\partial x^n}\Big)^2}}{\frac{\partial F}{\partial x^n}} dx^1 \wedge \cdots \wedge dx^{\mathrm{n}-1}.$$

[Suggestion. To calculate $\det[g_{ij}]$ use an orthonormal basis $e_1, \cdots, e_{n-1}$ for $\mathbb{R}^{n-1}$ with first vector $e_1 = (F_1, \cdots, F_n)/\sqrt{F_1^2 + \cdots + F_n^2}$. ]

7. Let $B = \{p = (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 1\}$ be the solid unit ball in $\mathbb{R}^3$.
a) Show that $B$ is a bounded submanifold of $\mathbb{R}^3$ with boundary the unit sphere $S^2$.
b) Give $\mathbb{R}^3$ the usual orientation, so that the standard coordinate system $(x, y, z)$ is positive. Show that the induced orientation on $S^2$ (Definition 3.3.9) corresponds to the outward normal on $S^2$ .

8. Prove *Lemma* 3.3.7

## 3.4 Lie derivatives

Throughout this section we fix a manifold $M$. All curves, functions, vector fields, etc. are understood to be smooth. The following theorem is a consequence of the existence and uniqueness theorem for systems of ordinary differential equations, which we take for granted.

3.4.1 **Theorem**. *Let $X$ be a* smooth *vector field on $M$. For any $p_o \in M$ the differential equation*

$$p'(t) = X(p(t)) \tag{1}$$

*with the initial condition*

$$p(0) = p_o$$

*has a unique* smooth *solution $p(t)$ defined in some interval about $t = 0$.* □

Write momentarily $p(t) = p(X, t)$ to bring out its dependence on $X$. For any $a \in \mathbb{R}$ one has

$$p(aX, t) = p(X, at)$$

because as a function of $t$, both sides of this equation satisfy $p'(t) = aX(p(t))$, $p(0) = p_o$. Hence $p(X, t)$ depends only on $tX$, namely $p(X, t) = p(tX, 0)$, whenever defined. It also depends on $p_o$, of course; we shall now denote it by the symbol $\exp(tX)p_o$. Thus (with $p_o$ replaced by $p$) the defining property of $\exp(tX)p$ is

$$\frac{d}{dt}\ \exp(tX)p = X(\exp(tX)p), \qquad \exp(tX)p|_{t=0} = p. \tag{2}$$

We think of $\exp(tX) : M \cdots \to M$ as a partially defined transformation of $M$, whose domain consists of those $p \in M$ for which $\exp(tX)p$ exists in virtue of the above theorem. More precisely, the map $(t, p) \to M$, $\mathbb{R} \times M \cdots \to M$, is a smooth map defined in a neighbourhood of $[0] \times M$ in $\mathbb{R} \times M$. In the future we shall not belabor this point: expressions involving $\exp(tX)p$ are understood to be valid whenever defined.

3.4.2 **Theorem**. *(a) For all $s, t \in \mathbb{R}$ one has*

$$\exp((s + t)X)p = \exp(sX) \exp(tX)p, \tag{3}$$

$$\exp(-tX) \exp(tX)p = p, \tag{4}$$

*where defined.*
*(b) For every* smooth *function $f : M \cdots \to \mathbb{R}$ defined in a neighbourhood of $p$ on $M$,*

$$f(\exp(tX)p)\ \sim\ \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k f(p), \tag{5}$$

*as Taylor series at $t = 0$ in the variable $t$.*

**Comment**. In (b) the vector field $X$ is considered a differential operator, given by $X = \sum X^i\, \partial/\partial x^i$ in coordinates. The formula may be taken as justification for the notation $\exp(tX)p$.

**Proof**. a) Fix $p$ and $t$ and consider both sides of equation (3) as function of $s$ only. Assuming $t$ is within the interval of definition of $\exp(tX)p$, both sides of the equation are defined for $s$ in an interval about 0 and smooth there. We check that the right side satisfies the differential equation and initial condition defining the left side:

$$\frac{\mathrm{d}}{ds}\exp((s+t)X)p = \frac{d}{dr}\exp(rX)p\Big|_{r=s+t} = X\Big(\exp((s+t)X)p\Big),$$

$$\exp((s+t)X)p|_{s=0} = \exp(tX)p.$$

This proves (3), and (4) is a consequence thereof.
(b) It follows from (2) that

$$\frac{d}{dt}f(\exp(tX)p) = Xf(\exp(tX)p).$$

Using this equation repeatedly to calculate the higher-order derivatives one finds that the Taylor series of $f(p(t))$ at $t=0$ may be written as

$$f(\exp(tX)p) \sim \sum_k \frac{t^k}{k!}X^k f(p),$$

as desired. □

The family $\{\exp(tX) : M \cdots \to M \mid t \in \mathbb{R}\}$ of locally defined transformations on $M$ is called the one parameter group of (local) transformations, or the flow, generated by the vector field $X$.

3.4.3 **Example**. The family of transformations of $M = \mathbb{R}^2$ generated by the vector field

$$X = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$$

is the one–parameter family of rotations given by

$$\exp(tX)(x,y) = ((\cos t)x - (\sin t)y,\ (\sin t)x + (\cos t)y).$$

It is of course defined for all $(x,y)$.

3.4.4 **Pull–back of tensor fields**. Let $F : N \to M$ be a smooth map of manifolds. If $f$ is a function on $M$ we denote by $F^*f$ the function on $N$ defined by

$$F^*f(p) = f(F(p))$$

i.e. $F^*f = f \circ F$. More generally, if $\varphi$ is a covariant tensor field on $M$, say of type $(0,k)$, then we can consider $\varphi$ a multilinear function on vector fields on $M$ and define $F^*\varphi$ by the rule

$$F^*\varphi(X, Y, \cdots) = \varphi(dF(X), dF(Y), \cdots)$$

for all vector fields $X, Y, \cdots$ on $N$, i.e. $F^*\varphi = \varphi \circ dF$ as multilinear function on vector fields. For general tensor fields one cannot define such a pull–back operation in a natural way for all maps $F$. Suppose however that $F$ is a diffeomorphism from $N$ onto $M$. Then with every vector–field $X$ on $M$ we can associate a vector field $F^*X$ on $N$ by the rule

$$(F^*X)(p) = dF_p^{-1}(X(F(p))).$$

The pull–back operation $F^*$ can now be uniquely extended to arbitrary tensor fields so as to respect tensor addition and tensor multiplication, i.e. so that

$$F^*(S+T) = (F^*S) + (F^*T), \tag{6}$$

$$F^*(S \otimes T) = (F^*S) \otimes (F^*T) \tag{7}$$

for all tensor fields $S, T$ on $M$. In fact, if we use coordinates $(x^i)$ on $M$ and $(y^j)$ on $N$, write $p = F(q)$ as $x^j = x^j(y^1, \cdots, y^n)$, and express tensors as a linear combination of the tensor products of the $dx^i$, $\partial/\partial x^i$ and $dy^a$, $\partial/\partial y^b$ all of this amounts to the familiar rule

$$(F^*T)^{ab\cdots}_{cd\cdots} = T^{ij\cdots}_{kl\cdots} \frac{\partial y^a}{\partial x^i} \frac{\partial y^b}{\partial x^j} \cdots \frac{\partial x^k}{\partial y^c} \frac{\partial x^l}{\partial y^d} \cdots \tag{8}$$

as one may check. Still another way to characterize $F^*T$ is obtained by considering a tensor $T$ of type $(r, s)$ as a multilinear function of $r$ covectors and $s$ vectors. Then one has

$$(F^*T)(F^*\varphi, F^*\psi, \cdots, F^*X, F^*Y, \cdots) = T(\varphi, \psi, \cdots,\ X, Y, \cdots) \tag{9}$$

for any $r$ covector fields $\varphi, \psi, \cdots$ on $M$ and any $s$ vector fields $X, Y, \cdots$ on $M$. If $M$ is a scalar function on $M$ then the chain rule says that

$$d(F^*f) = F^*(df). \tag{10}$$

If $f$ is a scalar function and $X$ vector field, then the scalar functions $Xf = df(X)$ on $M$ and $(F^*X)(F^*f)$ satisfy

$$F^*(Xf) = (F^*X)(F^*f). \tag{11}$$

If we write $g = F^*f$ this equation says that

$$(F^*X)g = [F^* \circ X \circ (F^*)^{-1}]g \tag{12}$$

as operators on scalar functions $g$ on $N$, which is sometimes useful.
The pull–back operation for a composite of maps satisfies

$$(G \circ F)^* = F^* \circ G^* \tag{13}$$

whenever defined. Note the reversal of the order of composition! (The verifications of all of these rules are left as exercises.)
We shall apply these definitions with $F = \exp(tX) : M \cdots \to M$. Of course, this is only partially defined (indicated by the dots), so we need to take for its domain $N$ a suitably small open set in $M$ and then take $M$ for its image. As

usual, we shall not mention this explicitly. Thus if $T$ is a tensor field on $M$ then $\exp(tX)^*T$ will be a tensor field of the same type. They need not be defined on all of $M$, but if $T$ is defined at a given $p \in M$, so is $\exp(tX)^*T$ for all $t$ sufficiently close to $t = 0$. In particular we can define

$$L_X\ T(p) := \frac{d}{dt}\ \exp(tX)^*T(p)\Big|_{t=0} = \lim_{t\to 0}\ \frac{1}{t}\ [\exp(tX)^*T(p) - T(p)]. \qquad (14)$$

This is called the *Lie derivative of $T$ along $X$*. It is very important to note that for a general tensor field $T$ the Lie derivative $L_XT(p)$ at a point $p$ depends not only on the vector $X(p)$ at $p$, but on the whole vector field $X$ near $p$. Only in the case of scalar functions, when the Lie derivative reduces to the directional derivative $Xf(p) = df_p(X(p))$, does it depend only on the vector $X(p)$ at $p$. We shall momentarily see how to compute it in general. The following rules follow immediately from the definition.

3.4.5 **Lemma**. *Let $X$ be a vector field, $S$, $T$ tensor fields. Then*
(a) $L_X(S + T) = L_X(S) + L_X(T)$
(b) $L_X(S \cdot T) = L_X(S) \cdot T + S \cdot L_X(T)$
(c) $L_X(\varphi \wedge \psi) = L_X(\varphi) \wedge \psi + \varphi \wedge L_X(\psi)$
(d) $F^*(L_X(T)) = L_{F^*X}\ (F^*T)$

**Explanation**. In (a) we assume that $S$ and $T$ are of the same type, so that the sum is defined. In (b) we use the symbol $S \cdot T$ to denotes any (partial) contraction with respect to some components. In (c) $\varphi$ and $\psi$ are differential forms. (d) requires that $F$ be a diffeomorphism so that $F^*T$ and $F^*X$ are defined.

**Proof**. (a) is clear from the definition. (b) follows directly from the definition (14) of $L_XT(p)$ as a limit, just like the product rule for scalar functions, as follows. Let $f(t) = \exp(tX)^*S$ and $g(t) = \exp(tX)^*S$. We have

$$\frac{1}{t}[f(t) \cdot g(t) - f(0) \cdot g(0)] = \frac{1}{t}[f(t) - f(0)] \cdot g(t) + f(0) \cdot \frac{1}{t}[g(t) - g(0)], \quad (15)$$

which gives (b) as $t \to 0$. (c) is proved in the same way. To prove (d) it suffices to consider for $T$ scalar functions, vector fields, and covector fields, since any tensor is built from these using sums and tensor products, to which the rules (a) and (b) apply. The details are left as an exercise. $\square$

**Remark**. The only property of the "product" $S \cdot T$ needed to prove a product rule using (15) is its $\mathbb{R}$–bilinearity.

3.4.6 **Lie derivative of scalar functions**. Let $f$ be a scalar function. From (13) we get

$$L_X\ f(p) = \frac{d}{dt}\ \exp(tX)^*f(p)\Big|_{t=0} = \frac{d}{dt}\ f(\exp(tX)p)\Big|_{t=0} = df_p(X(p)) = Xf(p),$$

the usual directional derivative along $X$.

3.4.7 **Lie derivative of vector fields**. Let $Y$ be another vector field. From (12) we get

$$\exp(tX)^*(Y)f = \exp(tX)^* \circ Y \circ \exp(-tX)^* f$$

as operators on scalar functions $f$. Differentiating at $t = 0$, we get

$$(L_X Y)f = \frac{d}{dt}\,[\exp(tX)^* \circ Y \circ \exp(-tX)^*]f\bigg|_{s=t=0} \tag{16}$$

This is easily computed with the help of the formula (5), which says that

$$\exp(tX)^* f \sim \sum \frac{t^k}{k!}\, X^k f. \tag{17}$$

One finds that

$$\begin{aligned}[\exp(tX)^* \circ Y \circ \exp(-tX)^*]f &= [1 + tX + \cdots]Y[1 - tX + \cdots]f \\ &= [Y + t(XY - YX) + \cdots]f\end{aligned}$$

Thus (16) gives the basic formula

$$(L_X Y)f = (XY - YX)f \tag{18}$$

as operators on functions. At first sight this formula looks strange, since a vector field

$$Z = \sum Z^i(\partial/\partial x^i)$$

is a first–order differential operator, so the left side of (18) has order one as differential operator, while the right side appears to have order two. The explanation comes from the following lemma.

3.4.8 **Lemma**. *Let $X$, $Y$ be two vector fields on $M$. There is a unique vector field, denoted $[X,Y]$, so that $[X,Y] = XY - YX$ as operators on functions defined on open subsets of $M$.*

**Proof**. Write locally in coordinates $x^1, x^2, \cdots, x^n$ on $M$:

$$X = \sum_k X^k \frac{\partial}{\partial x^k},\ Y = \sum_k Y^k \frac{\partial}{\partial x^k}.$$

By a simple computation using the symmetry of second partials one sees that $Z = \sum_k Z^k \frac{\partial}{\partial x^k}$ satisfies $Zf = XYf - YXf$ for all analytic functions $f$ on the coordinate domain if and only if

$$Z^k = \sum_j X^j \frac{\partial Y^k}{\partial x^j} - Y^j \frac{\partial X^k}{\partial x^j}. \tag{19}$$

This formula defines $[X,Y]$ on the coordinate domain. Because of the uniqueness, the $Z$'s defined on the coordinate domains of two coordinate systems agree on the intersection, from which one sees that $Z$ is defined globally on the whole manifold (assuming $X$ and $Y$ are). □

The bracket operation (19) on vector fields is called the *Lie bracket*. Using this the formula (18) becomes

$$L_X Y = [X, Y]. \tag{20}$$

The Lie bracket $[X, Y]$ is evidently bilinear in $X$ and $Y$ and skew–symmetric:

$$[X, Y] = -[Y, X].$$

In addition it satisfies the *Jacobi identity*, whose proof is left as an exercise:

$$[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0. \tag{21}$$

**Remark**. A *Lie algebra* of vector fields is a family $L$ of vector fields which is closed under $\mathbb{R}$–linear combinations and Lie brackets, i.e.

$$X, Y \in L \ \Rightarrow \ aX + bY \in L \text{ (for all } a, b \in \mathbb{R})$$

$$X, Y \in L \ \Rightarrow \ [X, Y] \in L.$$

3.4.9 **Lie derivative of differential forms**. We first discuss another operation on forms.

3.4.10 **Lemma**. *Let $X$ be a vector field on $M$. There is a unique operation $\varphi \to i_X\varphi$ which associates to each $k$–form $\varphi$ a $(k-1)$–form $i_X\varphi$ so that*

*(a) $i_X\theta = \theta(X)$ if $\theta$ is a 1–form*

*(b) $i_X(\varphi + \psi) = i_X(\varphi) + i_X(\psi)$ for any forms $\varphi, \psi$*

*(c) $i_X(\alpha \wedge \beta) = (i_X\alpha) \wedge \beta + (-1)^{|\alpha|}\alpha \wedge (i_X\beta)$*

*if $\alpha$ is homogeneous of degree $|\alpha|$. This operator $i_X$ is given by*

*(d) $(i_X\varphi)(X_1, \cdots, X_{k-1}) = \varphi(X, X_1, \cdots, X_{k-1})$*

*for any vector fields $X_1, \cdots, X_{k-1}$. It satisfies the following rules*

*(e) $i_X \circ i_X = 0$*

*(f)$i_{X+Y} = i_X + i_Y$ for any two vector fields $X, Y$*

*(g)$i_{fX} = f i_X$ for any scalar function $f$*

*(h) $F^* \circ i_X = i_{F^*X} \circ F^*$ for any diffeomorphism $F$.*

**Remark**. If $f$ is a scalar function it is understood that $i_X f = 0$, since there are no $(-1)$–forms.

**Outline of proof**. It is clear that there is at most one operation satisfying (a)$-$(c), since any form can be built from 1–forms by wedge–products and sums. Property (d) does define some operation $\varphi \to i_X\varphi$ and one can check that it satisfies (a) $-$ (c). The rest consists of easy verifications. □

We now turn to the Lie derivative on forms.

3.4.11 **Lemma**. *For all vector fields $X, Y$ we have*

*a) $d \circ L_X = L_X \circ d$*

*b) $i_{[X,Y]} = L_X \circ i_Y - i_Y \circ L_X$*

*as operators on forms.*

**Proof**. a) Since generally $F^*(d\varphi) = d(F^*\varphi)$, we have $d(\exp(tX)^*\varphi) = \exp(tX)^*(d\varphi)$. Differentiation at $t = 0$ gives $d(L_X\varphi) = L_X(d\varphi)$, hence (a).
b) By 3.4.10, (h)

$$\exp(tX)^*(i_Y\varphi) = i_{\exp(tX)^*Y}\ (\exp(tX)^*\varphi).$$

The derivative at $t = 0$ of the right side can again be evaluated as a limit using (15) if we write it as a symbolic product $f(t) \cdot g(t)$ with $f(t) = \exp(tX)^*Y$ and $g(t) = \exp(tX)^*\varphi$. This gives

$$L_X(i_Y\varphi) = i_{[X,Y]}\varphi + i_Y L_X\varphi,$$

which is equivalent to (b). □

3.4.12 **Lemma**. *Let $\theta$ be a 1–form and $X, Y$ two vector fields. Then*

$$X\theta(Y) - Y\theta(X) = d\theta(X,Y) + \theta([X,Y]) \tag{22}$$

**Proof**. We use coordinates. The first term on the left is

$$X\theta(Y) = X^i\ \frac{\partial}{\partial x^i}(\theta^j Y^j) = X^i\ \frac{\partial \theta^j}{\partial x^i} Y^j + X^i\theta^j\ \frac{\partial Y^j}{\partial x^i}$$

Interchanging $X$ and $Y$ and subtracting gives

$$X\theta(Y) - Y\theta(X) = \frac{\partial \theta^j}{\partial x^i}(X^i\ Y^j - X^j\ Y^i) + \theta^j\Big(X^i\ \frac{\partial Y^j}{\partial x^i} - Y^i\ \frac{\partial X^j}{\partial x^i}\Big) = d\theta(X,Y) + \theta([X,Y])$$

as desired. □
The following formula is often useful.

3.4.13 **Lemma (Cartan's Homotopy Formula)**. *For any vector field $X$,*

$$L_X = d \circ i_X + i_X \circ d \tag{23}$$

*as operators on forms.*

**Proof**. We first verify that the operators $A$ on both sides of this relation satisfy

$$A(\varphi + \psi) = (A\varphi) + (A\psi),\ A(\varphi \wedge \psi) = (A\varphi) \wedge \psi + \varphi \wedge (A\psi). \tag{24}$$

For $A = L_X$ this follows from Lemma 3.4.5 (c). For $A = d \circ i_X - i_X \circ d$, compute:

$$i_X(\alpha \wedge \beta) = (i_X\alpha) \wedge \beta + (-1)^{|\alpha|}\alpha \wedge (i_X\beta)$$

$$d\circ i_X(\alpha\wedge\beta) = (di_X\alpha)\wedge\beta + (-1)^{|\alpha|-1}(i_X\alpha)\wedge(d\beta) + (-1)^{|\alpha|}(d\alpha)\wedge i_X\beta + \alpha\wedge(di_X\beta)$$

and similarly

$$i_X\circ d(\alpha\wedge\beta) = (i_Xd\alpha)\wedge\beta + (-1)^{|\alpha|-1}(d\alpha)\wedge(i_X\beta) + (-1)^{|\alpha|}(i_X\alpha)\wedge d\beta + \alpha\wedge(i_Xd\beta).$$

Adding the last two equations one gets the relation (24) for $A = d \circ i_X + i_X \circ d$. Since any form can be built form scalar functions $f$ and 1–forms $\theta$ using sums and wedge products one sees from (24) that it suffices to show that (23) is true for these. For a scalar function $f$ we have:

$$L_X f = df(X) = i_X df$$

For a 1–form $\theta$ we have

$$\begin{aligned}(L_X\theta)(Y) &= X\theta(Y) - \theta([X,Y]) \qquad &&[\text{by } 3.4.11\ (\text{b})]\\ &= Y\theta(X) + d\theta(X,Y) &&[\text{by}(22)]\\ &= (di_X\theta)(Y) + (i_X d\theta)(Y)\end{aligned}$$

as required. □

3.4.14 **Corollary**. For any vector field $X$, $L_X = (d + i_X)^2$.

3.4.15 **Example: Nelson's parking problem**. This example is intended to give some feeling for the Lie bracket $[X, Y]$. We start with the formula

$$\exp(tX)\exp(tY)\exp(-tX)\exp(-tY) = \exp(t^2[X,Y] + \mathrm{o}(t^2)) \tag{25}$$

which is proved like (16). In this formula, replace $t$ by $\sqrt{t/\mathrm{k}}$, raise both sides to the power k$= 0, 1, 2, \cdots$ and take the limit as k$\to \infty$. There results

$$\lim_{k\to\infty}\Big(\exp(\sqrt{\frac{1}{2}}X)\exp(\sqrt{\frac{1}{2}}Y)\exp(-\sqrt{\frac{1}{2}}X)\exp(-\sqrt{\frac{1}{2}}Y)\Big)^k = \exp(t[X,Y)), \tag{26}$$

which we interpret as an iterative formula for the one−parameter group generated by $[X, Y]$. We now join Nelson (*Tensor Analysis*, 1967, p.33).

*Consider a car. The configuration space of a car is the four dimensional manifold $M = \mathbb{R}^2\times S^1\times S^1$ parametrized by $(x, y, \phi, \theta)$, where $(x, y)$ are the Cartesian coordinates of the center of the front axle, the angle $\phi$ measures the direction in which the car is headed, an $\theta$ is the angle made by the front wheels with the car. (More realistically, the configuration space is the open submanifold $-\theta_{\max} < \theta < \theta_{\max}$.) See Figure 1.*

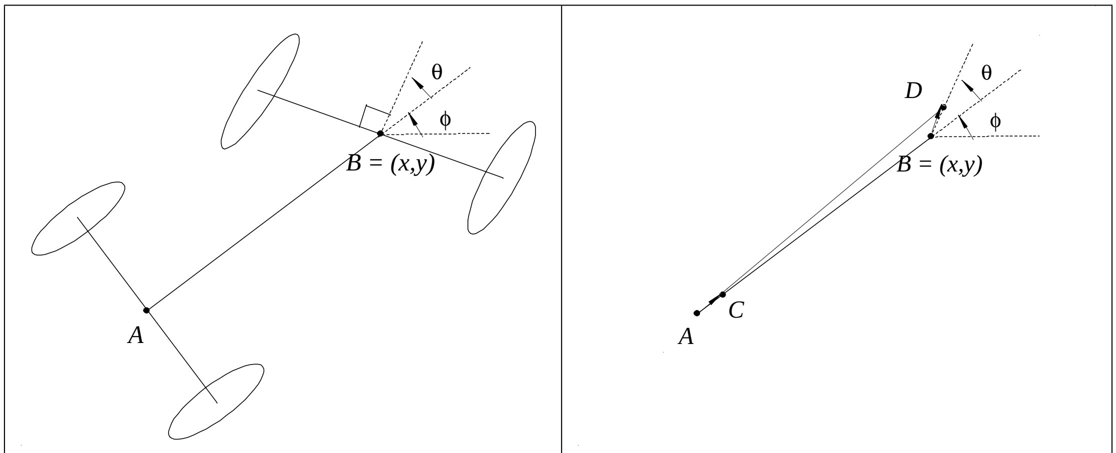
*There are two distinguished vector fields, called Steer and Drive, on $M$ corresponding to the two ways in which we can change the configuration of a car. Clearly*

$$\mathit{Steer} = \frac{\partial}{\partial\theta} \tag{27}$$

*since in the corresponding flow $\theta$ changes at a uniform rate while $x, y$ and $\phi$ remain the same. To compute Drive, suppose that the car, starting in the configuration $(x, y, \phi, \theta)$ moves an infinitesimal distance $h$ in the direction in which the front wheels are pointing.*

*In the notation of Figure 1,*

$$D = (x + h\cos(\phi + \theta) + o(h), y + h\sin(\phi + \theta) + o(h)).$$




*(a)A car* | *(b)A car in motion*

*Fig. 1. Nelson's car*

*Let* $1 = \overline{AB}$ *be the length of the tie rod (if that is the name of the thing connecting the front and rear axles). Then* $\overline{CD} = 1$ *too since the tie rod does not change length (in non−relativistic mechanics). It is readily seen that* $\overline{CE} = 1 + o(h)$*, and since* $\overline{DE} = h\sin\theta + o(h)$*, the angle* $BCD$ *(which is the increment in* $\phi$*) is* $(h\sin\theta)/l) + o(h))$ *while* $\theta$ *remains the same. Let us choose units so that* $l = 1$*. Then*

$$\text{Drive} = \cos(\phi+\theta)\frac{\partial}{\partial x} + \sin(\phi+\theta)\frac{\partial}{\partial y} + \sin\theta\frac{\partial}{\partial\phi}. \tag{28}$$

*By (27) and (28)*

$$[\text{Steer, Drive}] = -\sin(\phi+\theta)\frac{\partial}{\partial x} + \cos(\phi+\theta)\frac{\partial}{\partial y} + \cos(\theta)\frac{\partial}{\partial\phi}. \qquad (29))$$

*Let*

$$\text{Slide} = -\sin\phi\frac{\partial}{\partial x} + \cos\phi\frac{\partial}{\partial y}, \qquad \text{Rotate} = \frac{\partial}{\partial\phi}.$$

*Then the Lie bracket of* Steer *and* Drive *is equal to* Slide+ Rotate *on* $\theta = 0$*, and generates a flow which is the simultaneous action of sliding and rotating. This motion is just what is needed to get out of a tight parking spot. By formula (26) this motion may be approximated arbitrarily closely, even with the restrictions* $-\theta_{\max} < \theta < \theta_{\max}$ *with* $\theta_{\max}$ *arbitrarily small, in the following way: steer, drive reverse steer, reverse drive, steer, drive reverse steer,* $\cdots$*. What makes the process so laborious is the square roots in (26).*
*Let us denote the Lie bracket (29) of* Steer *and* Drive *by* Wriggle*. Then further simple computations show that we have the commutation relations*

$$\begin{aligned} &[\text{Steer, Drive}] = \text{Wriggle} \\ &[\text{Steer, Wriggle}] = -\text{Drive} \\ &[\text{Wriggle, Drive}] = \text{Slide} \end{aligned} \tag{30}$$

*and the commutator of* Slide *with* Steer, Drive *and* Wriggle *is zero. Thus the four vector fields span a four-dimensional Lie algebra over* $\mathbb{R}$*. To get out of an extremely tight parking spot,* Wriggle *is insufficient because it may produce too much rotation. The last commutation relation shows, however, that one may get out of an arbitrarily tight parking spot in the following way: wriggle, drive, reverse wriggle, (this requires a cool head), reverse drive, wriggle, drive,* $\cdots$.

### EXERCISES 3.4

1. Prove (8).

2. Prove (9).

3. Prove (11).

4. Prove (13).

5. Prove the formula for $\exp(tX)$ in example 3.4.3.

6. Fix a non–zero vector $v \in \mathbb{R}^3$ and let $X$ be the vector field on $\mathbb{R}^3$ defined by

$$X(p) = v \times p \text{ (cross-product)}.$$

(We identify $T_p\mathbb{R}^3$ with $\mathbb{R}^3$ itself.) Choose a right-handed orthonormal basis $(e_1, e_2, e_3)$ for $\mathbb{R}^3$ with $e_3$ a unit vector parallel to $v$. Show that $X$ generates the one–parameter group of rotations around $v$ given by the formula

$$\begin{aligned}
\exp(tX)e_1 &= (\cos ct)e_1 + (\sin ct)e_2 \\
\exp(tX)e_2 &= (-\sin ct)e_1 + (\cos ct)e_2 \\
\exp(tX)e_3 &= e_3
\end{aligned}$$

where $c = \|v\|$. [For the purpose of this problem, an ordered orthonormal basis $(e_1, e_2, e_3)$ is defined to be *right-handed* if it satisfies the usual cross–product relation given by the “right-hand rule”, i.e.

$$e_1 \times e_2 = e_3,\ e_2 \times e_3 = e_1,\ e_3 \times e_1 = e_2.$$

Any orthonormal basis can be ordered so that it becomes right–handed.]

7. Let $X$ be the coordinate vector $\partial/\partial\theta$ for the geographical coordinate system $(\theta, \phi)$ on $S^2$ (see example 3.2.8). Find a formula of $\exp(tX)p$ in terms of the Cartesian coordinates $(x, y, z)$ of $p \in S^2 \subset \mathbb{R}^3$.

8. Fix a vector $v \in \mathbb{R}^n$ and let $X$ be the constant vector field $X(p) = v \in T_p\mathbb{R}^n = \mathbb{R}^n$. Find a formula for the one–parameter group $\exp(tX)$ generated by $X$.

9. Let $X$ be a linear transformation of $\mathbb{R}^n$ ($n \times n$ matrix) considered as a vector field $p \to X(p) \in T_p\mathbb{R}^n = \mathbb{R}^n$. Show that the one–parameter group $\exp(tX)$ generated by $X$ is given by the *matrix exponential*

$$\exp(tX)p = \sum_{k=0}^{\infty} \frac{t^k}{k!}\ X^k p$$

where $X^k p$ is the $k$–th matrix power of $X$ operating on $p \in \mathbb{R}^n$.

10. Supply all details in the proof of Lemma 3.4.5.

11. Prove the Jacobi identity (21). [Suggestion. Use the operator formula $[X, Y] = XY - YX$.]

12. Prove that the operation $i_X$ defined by Lemma 3.4.10 (d) satisfies (a)–(c). [Suggestion. (a) and (b) are clear. For (c), argue that it suffices to take $X = \partial/\partial x_i$, $\alpha = dx_{j_1}$, $\beta = dx_{j_2} \wedge \cdots \wedge dx_{j_k}$. Argue further that it suffices to consider $i \in \{j_1, \cdots, j_k\}$. Relabel $(j_1, \cdots, j_k) = (1, \cdots, k)$ so that $i = 1$ or $i = 2$. Write $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k = (dx_1 \otimes dx_2 - dx_2 \otimes dx_2) \wedge dx_3 \wedge \cdots \wedge dx_k$ and complete the proof. ]

13. Prove Lemma 3.4.10, (e)–(h).

14. Supply all details in the proof of Lemma 3.4.11 (b).

15. Prove in detail that $A = d \circ i_X + i_X \circ d$ satisfies (24), as indicated.

16. Let $X$ be a vector field. Prove that for any scalar function $f$, 1–form $\theta$, and vector field $Y$,
a) $L_{fX}\theta = \theta(X)df + f(L_X\theta)$
b) $L_{fX}Y = f([X, Y]) - df(Y)X$

17. Let $X = P(\partial/\partial x) + Q(\partial/\partial x) + R(\partial/\partial x)$ be a vector field on $\mathbb{R}^3$, vol= $dx \wedge dy \wedge dz$ the usual volume form. Show that

$$L_X(\mathrm{vol}) = (\mathrm{div}X)\mathrm{vol}$$

where $\mathrm{div}X = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$ as usual.

18. Prove the following formula for the exterior derivative $d\varphi$ of a $(k-1)$–form $\varphi$. For any $k$ vector fields $X_1, \cdots, X_k$

$$\begin{aligned} d\varphi(X_1, \cdots, X_k) = \sum_{j=1}^{k} \ & (-1)^{j+1} X_j \varphi(X_1, \cdots, \widehat{X}_j \cdots, X_k) + \\ + \sum_{i<j} \ & (-1)^{i+j} \varphi([X_i, X_j], X_1, \cdots, \widehat{X}_i, \cdots, \widehat{X}_j, \cdots X_k) \end{aligned}$$

where the terms with a hat are to be omitted. The term $X_j\varphi(X_1, \cdots, \widehat{X}_j \cdots, X_k)$ is the differential operator $X_j$ applied to the scalar function $\varphi(X_1, \cdots, \widehat{X}_j \cdots, X_k)$. [Suggestion. Use induction on $k$. Calculate the Lie derivative $X\varphi(X_1, \cdots, X_k)$ of $\varphi(X_1, \cdots, X_k)$ by the product rule. Use $L_X = di_X + i_X d$.]

19. Let $M$ be an $n$–dimensional manifold, $\varpi$ an $n$–form, and $X$ a vector field. In coordinates $(x^i)$, write

$$\varpi = a dx^1 \wedge \cdots \wedge dx^n, \ X = X^i \frac{\partial}{\partial x^i}.$$

Show that

$$L_X\varpi = \Big(\sum_i \frac{\partial(aX^i)}{\partial x^i}\Big)dx^1\wedge\cdots\wedge dx^n.$$

20. Let $g$ be a Riemann metric on $M$, considered as a $(0,2)$ tensor $g = g_{ij}dx^i\otimes dx^j$, and $X = X^k(\partial/\partial x^k)$ a vector field. Show that $L_Xg$ is the $(0,2)$ tensors with components

$$(L_Xg)_{ij} = X^l\frac{\partial g_{ij}}{\partial x^l} + g_{kj}\frac{\partial X^k}{\partial x^i} + g_{ik}\frac{\partial X^k}{\partial x^j}.$$

[Suggestion. Consider the scalar function

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

as a contraction of the three tensors $g$, $\partial/\partial x^i$,and $\partial/\partial x^j$. Calculate its Lie derivative along $X$ using the product rule 3.4.5 (b), extended to triple products. Solve for $(L_Xg)_{ij}$.]

21. Let $g$ be a Riemann metric on $M$, considered as a $(0,2)$ tensor $g = g_{ij}dx^i\otimes dx^j$, and $X = X^k(\partial/\partial x^k)$ a vector field. Show that the 1–parameter group $\exp(tX)$ generated by $X$ consists of isometries of the metric $g$ if and only if $L_Xg = 0$. Find the explicit conditions on the components $X^k$ that this be the case. [Such a vector field is called a Killing vector field or an infinitesimal isometry of the metric $g$. See 5.22 for the definition of "isometry". You may assume that $\exp(tX) : M \to M$ is defined on all of $M$ and for all $t$, but this is not necessary if the result is understood to hold locally. For the last part, use the previous problem.]

22. Let $M$ be an $n$–dimensional manifold, $S$ a bounded, $m$–dimensional, oriented submanifold $S$ of $M$. Let $X$ be a vector field on $M$, which is tangential to $S$, i.e. $X(p)\in T_pS$ for all $p\in S$. Show:
a) $(i_X\varphi)\mid_S = 0$ for any smooth $(m+1)$–form $\varphi$ on $M$,
b) $\int_S L_X\varpi = \int_{\partial S} i_X\varpi$ for any smooth $m$–form $\varpi$ on $M$.
[The restriction $\varpi|_S$ of a form on $M$ to $S$ is the pull–back by the inclusion map $S\overset{\subseteq}{\to}M$.]

23. Let $M$ be an $n$–dimensional manifold, $S$ a bounded, $m$–dimensional, oriented submanifold of $M$. Let $X$ be a vector field on $M$. Show that for any smooth $m$–form $\varpi$ on $M$

$$\frac{d}{dt}\int_{\exp(tX)S}\varpi = \int_{\partial S} i_X(\exp(tX)^*\varpi)\ .$$

[Suggestion. Show first that for any diffeomorphism $F$ one has $\int_S F^*\varpi = \int_{F(S)}\varpi$.]

24. Prove (25).

25. Verify the bracket relation (29).

26. Verify the bracket relations (30).

27. In his Leçons of 1925-26 Élie Cartan introduces the exterior derivative on differential forms this way (after discussing the formulas of Green, Stokes, and Gauss in 3-space). *The operation which produces all of these formulas can be given in a very simple form. Take the case of a line integral* $\int \varpi$ *over a closed curve* $C$*. Let* $S$ *a piece of a* $2$*–surface (in a space of* $n$ *dimensions) bounded by* $C$*. Introduce on* $S$ *two interchangeable differentiation symbols* $d_1$ *and* $d_2$ *and partition* $S$ *into the corresponding family of infinitely small parallelograms. If* $p$ *is the vertex of on of these parallelograms (Fig. 2) and if* $p_1, p_2$ *are the vertices obtained from* $p$ *by the operations* $p_1$ *and* $p_2$*, then*

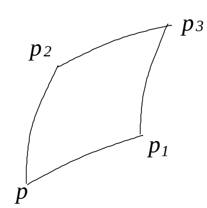



Fig. 2.

$\int_p^{p_1} \varpi = \varpi(d_1), \qquad \int_p^{p_2} \varpi = \varpi(d_2),$
$\int_{p_1}^{p_3} \varpi = \int_p^{p_1} \varpi + d_1 \int_p^{p_2} \varpi = \varpi(d_2) + d_1\varpi(d_2), \qquad \int_{p_2}^{p_3} \varpi = \varpi(d_1) + d_2\varpi(d_1).$

*Hence the integral* $\int \varpi$ *over the boundary of the parallelogram equals*
$\varpi(d_1) + [\varpi(d_2) + d_1\varpi(d_2)] - [\varpi(d_1) + d_2\varpi(d_1) - \varpi(d_2)$
$= d_2\varpi(d_1) - d_2\varpi(d_1).$
*The last expression is the exterior derivative* $d\varpi$*.*
(a)Rewrite Cartan's discussion in our language. [Suggestion. Take $d_1$ and $d_2$ to be the two vector fields $\partial/\partial u_1, \partial/\partial u_2$ on the 2-surface $S$ tangent to a parametrization $p = p(u_1, u_2)$.]
(b)Write a formula generalizing Cartan's $d\varpi = d_2\varpi(d_1) - d_2\varpi(d_1)$ in case $d_1, d_2$ are not necessarily interchangeable. [Explain what this means.]

# Chapter 4

# Special Topics

## 4.1 General Relativity

**The postulates of General Relativity**. The physical ideas underlying Einstein's relativity theory are deep, but its mathematical postulates are simple:

GR1. *Spacetime* is a four-dimensional manifold $M$.

GR2. $M$ has a Riemannian metric $g$ of signature $(-+++)$; the *world line* $p=p(s)$ of a freely-falling object is a geodesic:

$$\frac{\nabla}{ds}\frac{dp}{ds}=0. \tag{1}$$

GR3. In spacetime regions free of matter the metric $g$ satisfies the *field equations*

$$\mathrm{Ric}[g]=0. \tag{2}$$

**Discussion**. (1) The axiom GR1 is not peculiar to GR, but is at the basis of virtually all physical theories since time immemorial. This does not mean that it is carved in stone, but it is hard to see how mathematics as we know it could be applied to physics without it. Newton (and everybody else before Einstein) made further assumptions on how "space-time coordinates" are to be defined ("inertial coordinates for absolute space and absolute time"). What is especially striking in Einstein's theory is that the four-dimensional manifold spacetime is not further separated into a direct product of a three-dimensional manifold "space" and a one-dimensional manifold "time", and that there are no further restrictions on the space-time coordinates beyond the general manifold axioms.

(2) In the tangent space $T_pM$ at any point $p\in M$ one has a light cone, consisting of null-vectors ($ds^2=0$); it separates the vectors into timelike ($ds^2<0$) and spacelike ($ds^2>0$). The set of timelike vectors at $p$ consist of two connected components, one of which is assumed to be designated as forward in a manner

varying continuously with $p$. The world lines of all massive objects are assumed to have a forward, timelike directions, while the world lines of light rays have null directions.

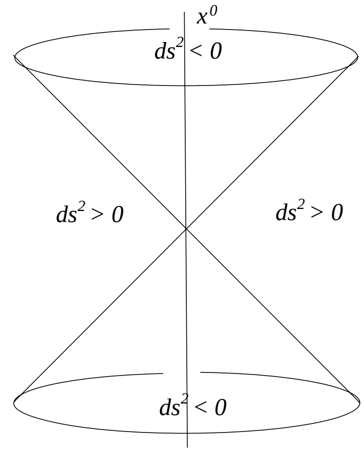



Fig.1

The coordinates on $M$ are usually labeled $(x^0, x^1, x^2, x^3)$. The postulate GR2 implies that around any given point $p_o$ one can choose the coordinates so that $ds^2$ becomes $-(dx^0)^2 + (dx^1)^2 + (dx^1)^2 + (dx^1)^2$ *at that point* and reduces to the usual positive definite $(dx^1)^2 + (dx^1)^2 + (dx^1)^2$ on the tangent space to the three dimensional spacelike manifold $x^0$ =const. By contrast, the tangent vector to a timelike curve has an imaginary "length" $ds$. In another convention the signature $(-+++)$ of $ds^2$ is replaced by $(+---)$, which would make the length real for timelike vectors but imaginary for spacelike ones.

(3 )In general, the parametrization $p = p(t)$ of a world line is immaterial. For a geodesic however the parameter $t$ is determined up to $t \to at + b$ with $a, b$ =constant. In GR2 the parameter $s$ is normalized so that $g(dp/ds, dp/ds) = -1$ and is then unique up to $s \to s + s_o$; it is called proper time along the world line. (It corresponds to parametrization by arclength for a positive definite metric.)

(4) Relative to a coordinate system $(x^0, x^1, x^2, x^3)$ on spacetime $M$ the equations (1) and (2) above read follows.

$$\frac{d^2x^k}{ds^2} + \Gamma^k_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds} = 0, \tag{1'}$$

$$-\frac{\partial\Gamma^k_{ql}}{\partial x^k} + \frac{\partial\Gamma^k_{qk}}{\partial x^l} - \Gamma^k_{\mathrm{pk}}\Gamma^p_{ql} + \Gamma^k_{\mathrm{pl}}\Gamma^p_{qk} = 0, \tag{2'}$$

where

$$\Gamma^l_{jk} = \frac{1}{2}g^{\mathrm{li}}(g_{ki,j} + g_{ij,k} - g_{jk,i}). \tag{3}$$

Equation (1), or equivalently (1′), takes the place of Newton's second law

$$\frac{d^2x^\alpha}{dt^2}+\frac{1}{m}\frac{\partial\varphi}{\partial x^\alpha}=0. \tag{4}$$

where $m$ is the mass of the object and $\varphi=\varphi(x^1,x^2,x^2)$ is a static gravitational potential. This is to be understood in the sense that there are special kinds of coordinates $(x^0,x^1,x^2,x^3)$, which we may call *Newtonian inertial frames*, for which the equations (4) hold for $\alpha=1,2,3$ provided the world line is parametrized so that $x^0=at+b$.
Einstein's field equations (2) represent a system of second-order, partial differential equations for the metric $g_{ij}$ analogous to Laplace's equations for the gravitational potential of a static field in Newton's theory, i.e.

$$\sum_\alpha \frac{\partial^2\varphi}{\partial x^\alpha\partial x^\alpha}=0 \tag{5}$$

(We keep the convention that the index $\alpha$ runs over $\alpha=1,2,3$ only. The equation is commonly abbreviated as $\Delta\varphi=0$.) Thus one can think of the metric $g$ as a sort of gravitational potential in spacetime.

*The field equations.* We wish to understand the relation of the field equations (2) in Einstein's theory to the field equation (4) of Newton's theory. For this we consider the mathematical description of the same physical situation in the two theories, namely the world lines of a collection of objects (say stars) in a given gravitational field. It will suffice to consider a one-parameter family of objects, say $p=p(r,s)$, where $r$ labels the object and $s$ is the parameter along its world-line.

(a) In *Einstein's theory* $p=p(r,s)$ is a family of geodesics depending on a parameter $r$:

$$\frac{\nabla}{\partial s}\frac{\partial p}{\partial r}=0. \tag{6}$$

For small $\Delta r$, the vector $(\partial p/\partial r)\Delta r$ at $p(r,s)$ can be thought of as the relative position at proper time $s$ of the object $r+\Delta r$ as seen from the object $\Delta r$.

To get a feeling for the situation, assume that at proper time $s=0$ the object $r$ sees the object $r+\Delta r$ as being contemporary. This means that the relative position vector $(\partial p/\partial r)\Delta r$ is orthogonal to the spacetime direction $\partial p/\partial s$ of the object $r$. By Gauss's lemma, this remains true for all $s$, so $r+\Delta r$ remains a contemporary of $r$. So it makes good sense to think of $\partial p/\partial r$ as a relative position vector.
The motion of the position vector $(\partial p/\partial r)\Delta r$ of $r+\Delta$s as seen from $r$ is governed by Jacobi's equation p.131

$$\frac{\nabla^2}{\partial s^2}\frac{\partial p}{\partial r}=R\Big(\frac{\partial p}{\partial r},\frac{\partial p}{\partial s}\Big)\frac{\partial p}{\partial s}. \tag{7}$$

This equation expresses its second covariant derivative (the "relative acceleration") in terms of the position vector $\partial p/\partial r$ of $r+\Delta r$ and the spacetime direction $\partial p/\partial s$ of $r$.
Introduce a local inertial frame $(x^0, x^1, x^2, x^3)$ for $\nabla$ at $p_o = p(r_o, s_o)$ so that $\partial/\partial x^0 = \partial p/\partial s$ at $p_o$. (This is a "local rest frame" for the object $r_o$ at $p_o$ in the sense that there $\partial x^\alpha/\partial s = 0$ for $\alpha = 1, 2, 3$ and $\partial x^0/\partial s = 1$.) At $p_o$ and in these coordinates (7) becomes

$$\frac{\partial^2}{\partial s^2}\frac{\partial x^i}{\partial r} = \sum_{j=0}^{3} F^i_j \frac{\partial x^j}{\partial r} \qquad \text{(at } p_0) \tag{8}$$

where $F^i_j$ is the matrix defined by the equation

$$R\Big(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^0}\Big)\frac{\partial}{\partial x^0} = \sum_{i=1}^{3} F^i_j \frac{\partial}{\partial x^i} \qquad \text{(at } p_0). \tag{9}$$

Note that the left-hand side vanishes for $j = 0$, hence $F^i_0 = 0$ for all $i$.

(b) In *Newton's theory* we choose a Newtonian inertial frame $(x^0, x^1, x^2, x^3)$ so that the equation of motion (4) reads

$$\frac{\partial^2 x^\alpha}{\partial t^2} = -\frac{1}{m}\frac{\partial \varphi}{\partial x^\alpha} \tag{4}$$

If we introduce a new parameter $s$ by $s = ct + t_o$ this becomes

$$c^2\frac{\partial^2 x^\alpha}{\partial s^2} = -\frac{1}{m}\frac{\partial \varphi}{\partial x^\alpha}\ .$$

By differentiation with respect to $r$ we find

$$\frac{\partial^2}{\partial s^2}\frac{\partial x^\alpha}{\partial r} = -\frac{1}{mc^2}\sum_{\beta=1}^{3}\frac{\partial^2 \varphi}{\partial x^\beta \partial x^\alpha}\frac{\partial x^\beta}{\partial r} \tag{10}$$

(c) Compare (8) and (10). The $x^i = x^i(r, s)$ in (8) refer the solutions of Einstein's equation of motion (7), the $x^\alpha = x^\alpha(r, s)$ in (10) to solutions of Newton's equations of motion (4). Assume now that the local inertial frame for $\nabla$ in (8) is the same as the Newtonian inertial frame in (10) and assume further that the derivatives on the left-hand side of (8) and (10) agree at $p_0$. Then we find that

$$\sum_{j=0}^{3} F^\alpha_j \frac{\partial x^j}{\partial r} = -\frac{1}{mc^2}\sum_{\beta=1}^{3}\frac{\partial^2 \varphi}{\partial x^\beta \partial x^\alpha}\frac{\partial x^\beta}{\partial r} \qquad \text{(at } p_0) \tag{11}$$

for all $\alpha = 1, 2, 3$. Since $\partial p/\partial r$ is an arbitrary spacelike vector at the point $p_o$ it follows from (11) that

$$F^\alpha_\beta = -\frac{1}{mc^2}\frac{\partial^2 \varphi}{\partial x^\beta \partial x^\alpha} \qquad \text{(at } p_0) \tag{12}$$

for $\alpha, \beta = 1, 2, 3$. Hence

$$\sum_{i=0}^{3} F_i^i = \sum_{\alpha=1}^{3} F_\alpha^\alpha = -\frac{1}{mc^2} \sum_{\alpha=1}^{3} \frac{\partial^2 \varphi}{\partial x^\alpha \partial x^\alpha} \qquad \text{(at } p_0\text{)} \tag{13}$$

By the definition (9) of $F_j^i$ this says that

$$\mathrm{Ric}(g)\Big(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^0}\Big) = -\frac{1}{mc^2} \sum_{\alpha=1}^{3} \frac{\partial^2 \varphi}{\partial x^\alpha \partial x^\alpha} \qquad \text{(at } p_0\text{)} \tag{14}$$

Newton's field equations (5) say that the right-hand side of (14) is zero. Since $\partial/\partial x^0$ is an arbitrary timelike vector at $p_o$ we find that $\mathrm{Ric}(g) = 0$ at $p_0$.
In summary, the situation is this. As a manifold (no metric or special coordinates), spacetime is the same in Einstein's and in Newton's theory. In Einstein's theory, write $x^i = x^i(g; r, s)$ for a one-parameter family of solutions of the equations of motion (1) corresponding to a metric $g$ written in a local inertial frame for $\nabla$ at $p_o$. In Newton's theory write $x^\alpha = x^\alpha(\varphi; r, t)$ for a one parameter family of solutions of the equations of motion (4) corresponding to a static potential $\varphi$ written in a Newtonian inertial frame. Assume the two inertial frames represent the same coordinate system on spacetime and require that the derivatives on the left-hand sides of (8) and (10) agree at $p_o$. Then the relation (14) between $g$ and $\varphi$ must hold. In particular, if $\varphi$ satisfies Laplace's equation (5) at $p_o$ then $g$ satisfies Einstein's field equation (2) at $p_o$.

One final comment. In Newton's theory, Laplace's equation $\Delta\varphi = 0$ gets replaced by Poisson's equation $\Delta\varphi = r$ in the presence of matter, where $r$ depends on the matter. In Einstein's theory the equation $\mathrm{Ric}(g) = 0$ gets analogously replaced by $\mathrm{Ric}(g) = T$, where $T$ is a 2–tensor, called the energy-momentum tensor. But in its present form this tensor is not an exact, fundamental representation of matter only a rough, macroscopic approximation,. Einstein once put it something like this. "*The field equations* $\mathrm{Ric}(g) = T$ *have the aspect of an edifice whose left wing is constructed from marble and whose right wing is constructed from inferior lumber.*" –We may be looking at the scaffolding; that right wing is still under construction.

## Appendix: some Special Relativity

*Special relativity* refers to the geometry in a four–dimensional vector space $W$ with an inner product of type $(- + ++)$, called Minkowski space. In terms of components with respect to an appropriate basis $(e_i)$ we can write $w \in W$ as $w = \xi_0 e_0 + \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3$ and

$$w^2 := (w, w) = -\xi_0^2 + \xi_1^2 + \xi_2^2 + \xi_3^2.$$

We shall think of this vector space as the tangent space $W = T_pM$ at some point (event) in the spacetime $M$ of general relativity, but this is not necessary and historically special relativity came first.

Let $e \in T_pM$ be a timelike spacetime vector. Think of $e$ as the tangent vector at $p$ to the world–line of some observer passing through $p$ (i.e. present at the event $p$). Let $w \in T_pM$ another spacetime vector at $p$, interpreted as the tangent vector to the world–line of some other object. Think of $w$ as an infinitesimal spacetime displacement from $p$. The observer $e$ splits $w$ as an orthogonal sum

$$w = \tau e + d$$

and interprets $\tau$ as the *relative time–duration* of $w$ and $d$ as the *relative space–displacement* of $w$, where relative refers to $e$. These components $\tau e$ and $d$ depend only on the spacetime direction of $e$, i.e. on the line $\mathbb{R}e$. So we assume that $e$ is normalized so that $(e, e) = -1$. Then evidently

$$\tau = -(w, e), \; d = w + (w, e)e.$$

Thus the observer $e$ would say the object $w$ moves through a space–displacement $d$ during a time–duration $\tau$ relative to himself. Hence $e$ would consider $d/\tau$ as the *relative space–velocity* of $w$. For a light ray one has $w^2 = 0$, so $-\tau^2 + d^2 = 0$ and $d^2/\tau^2 = 1$, i.e. the observer $e$ is using units so that velocity of light is 1.

Now suppose we have another observer $e' \in T_pM$ at $p$, again normalized to $(e', e') = -1$. Then $e'$ will split $w$ as

$$w = \tau' e' + d'.$$

The map $(\tau, d) \to (\tau', d')$ defined by the condition $\tau e + d = \tau' e' + d'$ is the *Lorentz transformation* which relates the (infinitesimal) spacetime displacements $w$ as observed by $e$ and $e'$. (The term *Lorentz transformation* is also applied to any linear transformation of $W$ which preserves the inner product.) It is easy to find a formula for it. Write

$$e' = ae + av$$

so that $v$ is the space–velocity of $e'$ relative to $e$. Taking inner products of this equation gives

$$-1 = a^2(-1 + v^2)$$

where $v^2 = (v, v)$. So

$$a^2 = (1 - v^2)^{-1}.$$

The equation $\tau e + d = \tau' e' + d'$ gives $\tau e + d = \tau' ae + (av + d')$ so the Lorentz transformation is

$$\tau = a\tau', \; d = d' + av.$$

In particular

$$\tau = \frac{\tau'}{\sqrt{1 - v^2}}.$$

Since $\sqrt{1 - v^2} < 1$ when real, $\tau \geq \tau'$; this phenomenon is called *time dilation.*

Consider now another situation, where $e$ and $e'$ observe some three–dimensional object. Such an object will not have a world–line but a world–tube, which we can picture as three–parameter family of world–lines representing the point–particles which make up the object. If the object is one–dimensional (a stick) we get a *world–band*, as we shall now assume. Locally around $p$ we approximate the world–band by the piece of a 2–plane between two parallel lines in $T_pM$. We shall assume that $e'$ is parallel to these lines, which means that the stick is a rest relative of $e'$. We assume further that the stick as seen by $e'$ points into the direction of $e$, which implies that $e, e'$ lie also in this 2–plane. The band $B$ intersects the 3–spaces orthogonal to $e, e'$ (the rest–spaces of $e, e'$) in line–segments, which we represent by vectors $d, d'$. The *length* of the stick as seen by $e, e'$ is then $l = |d|, l' = |d'|$ respectively. The relation between them can be found by simple vector algebra in the 2–plane in question. (The metric in this plane has type $(-,+)$, but the reasoning becomes more transparent if one draws upon one's intuition in a Euclidean space as a guide to the vector algebra.) As above we have

$$e^2 = -1,\ (e') = -1,\ e' = ae + av, a^2 = (1-v^2)^{-1}.$$

Since $d'$ is the component of $d$ orthogonal to $e'$, we have $d = d' + (d, e')e'$ and

$$d^2 = (d')^2 - (v, e')^2.$$

Since $d, v$ are both orthogonal to $e$ in the same 2–plane, $d/|d| = v/|v|$. Substituting $d$ in the previous equation we get

$$d^2 = (d')^2 - \frac{d^2}{v^2}(v, e')^2.$$

Since $e' = ae + av$ we find $d^2 = (d')^2 - d^2 a^2 v^2$, or

$$(d')^2 = d^2(1 + a^2v^2) = d^2(1 + (1-v^2)^{-1}v^2) = d^2(1-v^2)^{-1}.$$

This gives the desired relation between the relative lengths of the stick:

$$l = l'\sqrt{1-v^2}.$$

Thus $l \le l'$ and this is known as the *Lorentz contraction*. From a purely mathematical point of view all of this is elementary vector algebra, but its physical interpretation is startling.

## 4.2 The Schwarzschild metric

## 4.3

We now turn to *metrics with spherical symmetry*. This is not an entirely obvious concept in general relativity. To explain it, we start with two examples of symmetry in the familiar setting of $\mathbb{R}^3$ with the Euclidean metric $ds^2 = dx^2 + dy^2 + dz^2$. (Any unfamiliar terms in these examples should be self-explanatory, but will be defined precisely later.)

**Example 1: spherical symmetry in $\mathbb{R}^3$.** Consider the action of the group SO(3) of rotation about the origin on the Euclidean space $\mathbb{R}^3$, written $p \to a \cdot p$ with $a \in \mathrm{SO}(3)$ and $p \in \mathbb{R}^3$. We record the following facts.

(a) The group SO(3) acts on $\mathbb{R}^3$ by isometries, i.e. by transformations preserving the metric.

(b) The "orbit" $S(p_o) := \{p = a \cdot p_o \mid a \in \mathrm{SO}(3)\}$ of SO(3) through a point $p_o$ is the sphere about the origin through $p_o$.

(c) Fix $p_o$ and let $C = C(p_o) \approx \mathbb{R}$ be the line through $p_o$ orthogonal to $S(p_o)$. In a neighbourhood of $p_o$ the line $C$ intersects each sphere $S(p)$ orthogonally in a single point.

(d) Let $S = S^2$ be the unit sphere, $ds_S^2$ the standard metric on $S = S^2$, and $ds_C^2$ the induced metric on $C$. In a neighbourhood of $p_o$ there is a local diffeomorphism $C \times S \to \mathbb{R}^3$ so that

$$ds^2 = ds_C^2 + \rho^2 ds_S^2$$

where $\rho = \rho(p)$ is the radius of the sphere $S(p)$ though $p$.

The first three statements are again geometrically obvious. For the last one recall the expression for the Euclidean metric in spherical coordinates:

$$ds^2 = d\rho^2 + \rho^2(\sin\phi \, d\theta^2 + d\phi^2)$$

If we take $\rho$ as Cartesian coordinate on $C \approx \mathbb{R}$ and $(\theta, \phi)$ as coordinates on $S = S^2$, then the map $C \times S \to \mathbb{R}^3$ is just the spherical coordinate map $(\rho, \theta, \phi) \to (x, y, z)$.

**Example 2: cylindrical symmetry in $\mathbb{R}^3$.** Consider the group SO(2) of rotations about the $z$-axis on $\mathbb{R}^3$. We note that the statements (a)-(b) of the previous example have obvious analogs.

(a) The group SO(2) acts on $\mathbb{R}^3$ by isometries.

(b) The "orbit" $S(p_o) := \{p = a \cdot p_o \mid a \in \mathrm{SO}(2)\}$ of SO(2) through a point $p_o$ is the circle about the $z$-axis through $p_o$ .

(c) Fix $p_o$ and let $C = C(p_o) \approx \mathbb{R}$ be the plane through $p_o$ orthogonal to $S(p_o)$. In a neighbourhood of $p_o$ the plane $C$ intersects each circle $S(p)$ orthogonally in a single point.

(d) Let $S = S^1$ be the unit circle, $ds_S^2$ the standard metric on $S = S^2$, and $ds_C^2$ the induced metric on $C$. In a neighbourhood of $p_o$ there is a local diffeomorphism $C \times S \to \mathbb{R}^3$ so that

$$ds^2 = ds_C^2 + r^2 ds_S^2$$

where $r = s(p)$ is the radius of the circle $S(p)$ though $p$.

The first three statements are again geometrically obvious. For the last one recall the expression for the Euclidean metric in cylindrical coordinates:

$$ds^2 = (dr^2 + dz^2) + r^2 d\theta^2$$

If we take $(r, z)$ as Cartesian coordinates on $C \approx \mathbb{R}^2$ and $\theta$ as coordinate on $S = S^1$, then the map $C \times S \to \mathbb{R}^3$ is just the cylindrical coordinate map $(r, z; \theta) \to (x, y, z)$.

We now turn to some generalities. We take the view that *symmetry* (whatever that may mean) is characterized by a *group*. In general a *transformation group* of a space $M$ (any set, in our case a manifold) is a family $G$ of transformations of $M$ (mappings of $M$ into itself) satisfying

(1) If $a, b \in G$, then the composite $ab \in G$.
(2) If $a \in G$, then $a$ has an inverse $a^{-1}$ and $a^{-1} \in G$.

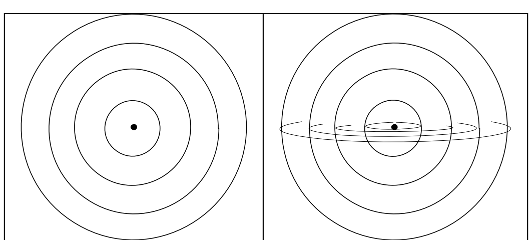
*The action of transformation $a \in G$ on a point $p \in M$ is denoted $p \to a \cdot p$.* The transformations in $G$ fixing a given point $p$ is $I(p) := \{a \in G \mid a \cdot p = p\}$, called the stabilizer of $p$ in $G$. It is itself a group, a subgroup of $G$.

4.2.1 **Example: rotation groups**. The *rotation group* in $\mathbb{R}^3$, denoted SO(3), consists of all orthogonal linear transformations of $\mathbb{R}^3$ of determinant $+1$:

$$\mathrm{SO}(3) = \{a \in \mathbb{R} \mid aa^* = 1, \ \det a = +1\}$$

SO stands for "special orthogonal". It can be realized geometrically as the group of transformations of Euclidean 3–space which can be obtained by a rotation by some angle about some axis through a given point (the origin): hence we can take $G = \mathrm{SO}(3)$, $M = \mathbb{R}^3$ in the above definition. Alternatively, SO(3) can be realized as the group of transformations of a 2–sphere $S^2$ which can be obtained by such rotations: hence we may also take $G = \mathrm{SO}(3)$, $M = \mathrm{S}^2$.

The rotation group SO(2) in $\mathbb{R}^2$ consists of linear transformations of $\mathbb{R}^2$ represented by matrices of the form $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. It acts naturally on $\mathbb{R}^2$ by rotations about the origin, but it also acts on $\mathbb{R}^3$ by rotations about the $z-$axis. In this way SO(2) can thought of as the subgroup of SO(3) which leaves the points on the $z$-axis fixed.



SO(2) orbits in $\mathbb{R}^2$ | SO(3) orbits in $\mathbb{R}^3$

Fig. 1

This example shows that the same group $G$ can be realized as a transformation group on different spaces $M$: we say $G$ acts on $M$ and we write $a \cdot p$ for the action of the transformation $a \in G$ on the point $p \in M$. The orbit a point $p \in M$ by $G$ is the set

$$G \cdot p = \{a \cdot p \mid a \in G\}$$

of all transforms of $p$ by elements of $G$. In the example $G = \mathrm{SO}(3)$, $M = \mathbb{R}^3$ the orbits are the spheres of radius $\rho > 0$ together with the origin $\{0\}$.

4.2.2 **Lemma**. *Let $G$ be a group acting on a space $M$. Then $M$ is the disjoint union of the orbits of $G$.*

**Proof** . Let $G \cdot p$ and $G \cdot q$ be two orbits. We have to show that they are either disjoint or identical. So suppose they have a point in common, say $b \cdot p = c \cdot q$ for some $b, c \in G$. Then $p = b^{-1}c \cdot q \in G \cdot q$, hence $a \cdot p = ab^{-1}c \cdot q \in G \cdot q$ for any $a \in G$, i.e. $G \cdot p \subset G \cdot q$. Similarly $G \cdot q \subset G \cdot p$, hence $G \cdot p = G \cdot q$.
From now on $M$ will be spacetime, a 4–dimensional manifold with a Riemann metric $g$ of signature $(-+++)$.

4.2.3 **Definition**. The metric $g$ is *spherically symmetric* if $M$ admits an action of SO(3) by isometries of $g$ so that every orbit of SO(3) in $M$ is isometric to a sphere in $\mathbb{R}^3$ of some radius $r > 0$ by an isometry preserving the action of SO(3).

**Remarks.** a) We exclude the possibility $r = 0$, i.e. the spheres cannot degenerate to points. Let us momentarily omit this restriction to explain where the definition comes from. The centre of symmetry in our spacetime should consist of a world line $L$ (e.g. the world line of the centre of the sun) where $r = 0$. Consider the group $G$ of isometries of the metric $g$ which leaves $L$ pointwise fixed. For a general metric $g$ this group will consist of the identity transformation only. In any case, if we fix a point $p$ on $L$, then $G$ acts on the 3–dimensional space of tangent vectors at $p$ orthogonal to $L$, so we can think of $L$ as a subgroup of SO(3). If it is all of SO(3) (for all points $p \in L$) then we have spherical symmetry in the region around $L$. However in the definition, we do not postulate the existence of such a world line (and in fact explicitly exclude it from consideration by the condition $r > 0$), since the metric (gravitational field) might not be defined at the centre, in analogy with Newton's theory.
b) It would suffice to require that the orbits of SO(3) in $M$ are 2–dimensional. This is because any 2–dimensional manifold on which SO(3) acts with a single orbit can be mapped bijectively onto a 2–sphere $S^2$ (possibly with antipodal points identified) so that the action becomes the rotation action. Furthermore, any Riemann metric on such a sphere which is invariant under SO(3) is isometric to the usual metric on a sphere of some radius $r > 0$, or the negative thereof. The latter is excluded here, because there is only one negative direction available in a spacetime with signature $(-+++)$.

We now assume given a spherically symmetric spacetime $M, g$ and fix an action of SO(3) on $M$ as in the definition. For any $p \in M$, we denote by $S(p)$ the SO(3) orbit through $p$; it is isometric to a 2–sphere of some radius $r(p) > 0$. Thus we have a *radius function* $r(p)$ on $M$. ($r(p)$ is intrinsically defined, for example by the fact that the area of $S(p)$ with respect to the metric $g$ is $4\pi r^2$.) For any $p \in M$, let $C(p) = \exp_p(T_pS(p)^\perp)$ be the union of the geodesics through $p$ orthogonal to $S(p)$. It is a 2–dimensional submanifold of $M$ which intersects $S(p)$ orthogonally in the single point $p$, at least as long as we restrict attention to a sufficiently small neighbourhood of $p$, as shall do throughout this discussion. Now fix $p_o \in M$ and let $C = C(p_o)$.

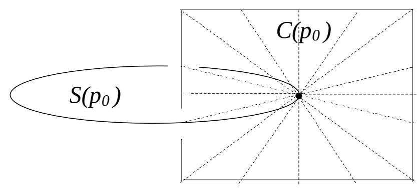



Fig.2

Let $S$ be a sphere about the origin in $\mathbb{R}^3$. The geographic coordinates $(\theta, \phi)$ on $S$ can be defined by the formula $p = a_z(\theta)a_y(\phi)e_z$ where $a_z(\theta)$ and $a_y(\phi)$ are rotations about the $z$ and $y$ axis and $e_z$ is the north-pole of $S$, i.e. its point of intersection with the positive $z$-axis. We transfer these coordinates to the orbit $S(p_o)$ as follows. By assumption, $S(p_o)$ is an isometric copy of a sphere $S$ in $\mathbb{R}^3$. We may choose this isometry so that it maps $p_o$ to $e_z$. It then maps $a \cdot p$ to $a \cdot e_z$ for any $a \in \mathrm{SO}(3)$, since the isometry preserves the action of $\mathrm{SO}(3)$. In particular, we can transfer $(\theta, \phi)$ to coordinates on $S(p_o)$ by setting $p = a_z(\theta)a_y(\phi)e_z$. We can extend these from $S(p_o)$ to all of $M$ as follows.

4.2.4 **Lemma**. *The map $f : C \times S \to M$, given by $p = a_z(\theta)a_y(\phi) \cdot q$, is local diffeomorphism near $(p_o, e_z)$ which maps $\{q\} \times S$ to $S(q)$.*

**Proof**. The second assertion is clear. To prove the first it suffices to show that the differential of $f$ at the point $(p_o, e_z)$ is invertible. Since $f$ maps $\{p_o\} \times S$ diffeomorphically onto $S(p_o)$ and $C \times \{e_z\}$ identically onto $C$, its differential restricts to a linear isomorphism of the corresponding tangent spaces and hence is a linear isomorphism onto $T_{p_o}M = T_{p_o}C \oplus T_{p_o}S(p_o)$. □

**Corollary**. *Near $p_o$, $C$ intersects each sphere $S(q)$ in a single point.*

**Proof**. This is clear, because the corresponding statement holds on $C \times S$. □

**Remark**. The group $I(p_o)$ acts also on the tangent space $T_{p_o}M$ by the differential of the map $p \to a \cdot p$ at $p = p_o$. This means that for curve $p(t)$ with $p(0) = p_o$ and $\dot{p}(0) = v$ we have

$$a \cdot v = \left.\frac{d}{dt}\right|_{t=0} a \cdot p(t).$$

Since $a \in I(p_o)$ operates on $M$ by isometries fixing $p_o$, it maps geodesics through $p_o$ into geodesics through $p_o$, i.e.

$$\exp_{p_o} : T_{p_o}M \to M \text{ satisfies } a \cdot (\exp_{p_o} v) = \exp_{p_o}(a \cdot v).$$

So if we identify vectors $v$ with points $p$ via $p = \exp_{p_o}(v)$, as we can near $p_o$, then the action of $I(p_o)$ on $T_{p_o}M$ turns into its action on $M$. We record this fact as follows

$$\textit{the action of } I(p_o) \textit{ on } M \textit{ looks locally like its linear action on } T_{p_o}M \tag{*}$$

In may help to keep this in mind for the proof of the following lemma.

4.2.5**Lemma**. *$C$ intersects each sphere $S(q)$, $q \in C$, orthogonally.*

**Proof**. Consider the action of the subgroup $I(p_o)$ of SO(3) fixing $p_o$. Since the action of SO(3) on the orbit $S(p_o)$ is equivalent to its action an ordinary sphere in Euclidean 3-space, the group $I(p_o)$ is just the rotation group SO(2) in the 2–dimensional plane $T_{p_o}S(p_o) \approx \mathbb{R}^2$. It maps $C(p_o)$ into itself as well as each orbit $S(q)$, hence also $C(p_o) \bigcap S(q) = \{q\}$, i.e. $I(p_o)$ fixes all $q \in C(p_o)$, i.e. $I(p_o) \subset I(q)$. The fixed-point set $C'(p_o)$ of $I(p_o)$ consists of the geodesics orthogonal to the 2-sphere $S(p_o)$ at $p_o$, as is evident for the corresponding sets in the Minkowski space $T_{p_o}M$. Thus $C'(p_o) = C(p_o)$. But $I(p_o) \subset I(q)$ implies $C'(p_o) \supset C'(q)$, hence $C(p_o) \supset C(q)$. Interchanging $p_o$ and $q$ in this argument we find also $C(q) \subset C(p_o)$. So $C(p_o) = C(q)$ and intersects $S(q)$ orthogonally.□

In summary, the situation is now the following. Under local diffeomorphism $C \times S \to M$ the metric $ds^2$ on $M$ decomposes as an orthogonal sum

$$ds^2 = ds_C^2 + r^2 ds_S^2 \tag{0}$$

where $ds_C^2$ is the induced metric on $C$, $ds_S^2$ the standard metric on $S = S^2$, and $r = r(q)$ is the radius function. To be specific, choose any orthogonal coordinates $\tau, \rho$ on $C$ with $\tau$ timelike and $r$ spacelike . (That this is possible follows from Gauss's Lemma, for example.) Let $\phi$, $\theta$ be usual coordinates $\phi, \theta$ on $S \approx S^2$. Then $\tau, \rho, \phi, \theta$ provide coordinates on $M$ and

$$ds^2 = -A^{-2}(\tau,\rho)d\tau^2 + B^2(\tau,\rho)d\rho^2 + r^2(\tau,\rho)(d\phi^2 + \sin^2\phi d\theta^2) \tag{1}$$

for some strictly positive functions $A(\tau,\rho)$, $B(\tau,\rho)$. We record the result as a theorem.

4.2.6 **Theorem**. *Any spherically symmetric metric is of the form (1) in suitable coordinates $\tau, \rho, \phi, \theta$.*

So far we have not used Einstein's field equations, just spherical symmetry. The field equations in empty space say that $\mathrm{Ric}[g] = 0$. For the metric (1) this amounts to the following equations (as one can check by some unpleasant but straightforward computations). Here $'$ denotes $\partial/\partial\rho$ and $^{\cdot}$ denotes $\partial/\partial\tau$.

$$\frac{2B}{A}\Big(\frac{r^{\cdot\prime}}{r} - \frac{B^{\cdot}r'}{Br} + \frac{r^{\cdot}A'}{rA}\Big) = 0 \tag{2}$$

$$\frac{1}{r^2} + \frac{2}{B}\Big(-\frac{r'}{Br}\Big)' - 3\Big(\frac{r'}{Br}\Big)^2 + 2A^2\frac{B^{\cdot}r^{\cdot}}{Br} + A^2\Big(\frac{r^{\cdot}}{r}\Big)^2 = 0 \tag{3}$$

$$\frac{1}{r^2}+2A\Big(A\frac{r^{\cdot}}{r}\Big)^{\cdot}+3\Big(\frac{r^{\cdot}}{r}\Big)^2A^2+\frac{2}{B^2}\frac{r'A'}{rA}-\Big(\frac{r'}{Br}\Big)^2=0 \tag{4}$$

$$\frac{1}{B}\Big(-\frac{A'}{AB}\Big)^2-A\Big(A\frac{B^{\cdot}}{B}\Big)^{\cdot}-2A\Big(A\frac{r^{\cdot}}{r}\Big)^{\cdot}-A^2\Big(\frac{B^{\cdot}}{B}\Big)^2- \tag{5}$$
$$-2A^2\Big(\frac{r^{\cdot}}{r}\Big)^2+\frac{1}{B^2}\Big(\frac{A'}{A}\Big)^2-\frac{2}{B^2}\frac{r'A'}{rA}=0$$

One has to distinguish three cases, according to the nature of the variable $r = r(p)$, the radius of the sphere $S(p)$.

(a) $r$ is a space-like variable, i.e. $g(\mathrm{grad}r,\mathrm{grad}r)>0$,
(b) $r$ is a time-like variable, i.e. $g(\mathrm{grad}r,\mathrm{grad}r)<0$,
(c) $r$ is a null variable, i.e. $g(\mathrm{grad}r,\mathrm{grad}r)\equiv 0$.

Here $\mathrm{grad}r = (g^{ij}\partial r/\partial x^j)(\partial/\partial x^i)$ is the vector-field which corresponds to the covector field $dr$ by the Riemann metric $g$. It is orthogonal to the 3–dimensional hypersurfaces $r =$constant. It is understood that we consider the cases where one of these conditions (a)–(c) holds identically on some open set.

We first dispose of the exceptional case (c). So assume $g(\mathrm{grad}r,\mathrm{grad}r)\equiv 0$. This means that $-(Ar^{\cdot})^2+(B^{-1}r')^2=0$ i.e.

$$\frac{r'}{B}=Ar^{\cdot} \tag{6}$$

up to sign, which may be adjusted by replacing the coordinate $\tau$ by $-\tau$, if necessary. But then one finds that $r^{\cdot\prime}$ determined by (2) is inconsistent with (3), so this case is excluded.

Now consider the case (a) when $g(\mathrm{grad}r,\mathrm{grad}r)>0$. In this case we can take $\rho = r$ as $\rho$-coordinate on $C$. Then $r^{\cdot}=0$ and $r'=1$. The equations (2)–(4) now simplify as follows.

$$B^{\cdot}=0 \tag{2$'$}$$

$$\frac{1}{r^2}+\frac{2}{B}\Big(-\frac{1}{Br}\Big)'-3\Big(\frac{1}{Br}\Big)^2=0 \tag{3$'$}$$

$$\frac{1}{r^2}+\frac{2}{B^2}\frac{1A'}{rA}-\Big(\frac{1}{Br}\Big)^2=0 \tag{4$'$}$$

$$\frac{1}{B}\Big(-\frac{A'}{AB}\Big)^2+\frac{1}{B^2}\Big(\frac{A'}{A}\Big)^2-\frac{2}{B^2}\frac{A'}{rA}=0 \tag{5$'$}$$

Equation (2$'$) shows that $B^{\cdot}=0$, i.e. $B=B(r)$. Equation (4$'$) differentiated with respect to $\tau$ shows that $(A^{\cdot}/A)'=0$, i.e. $(\log A)^{\cdot\prime}=0$. Hence $A = \tilde{A}(r)F(\tau)$. Now replace $\tau$ by $t=t(\tau)$ so that $dt = d\tau/F(\tau)$ and then drop the $\sim$ ; we get $A=A(r)$. Equation (3$'$) simplifies to

$$-2rB^{-3}B'+B^{-2}=1,\ \text{i.e}\ (rB^{-2})'=1.$$

By integration we get $rB^{-2}=r-2m$ where $-2m$ is a constant of integration, i.e. $B^2=(1-2m/r)^{-1}$ . Equation (4$'$) has the solution $A=B$ and this solution is unique up to a non-zero multiplicative constant $\gamma$, which one can take be 1 after replacing $t$ by $\gamma^{-1}t$. The metric (1) now becomes

$$ds^2 = -\Big(1 - \frac{2m}{r}\Big)dt^2 + \Big(1 - \frac{2m}{r}\Big)^{-1}dr^2 + r^2(d\phi^2 + \sin^2\phi d\theta^2). \tag{7}$$

It gives a solution for all $r \neq 0, 2m$. We shall however keep the restriction $r > 0$. Furthermore, only for $r > 2m$ is the radius function $r$ space-like, as we assumed. In the region where $r < 2m$ the coordinate $r$ becomes time-like, and to make the signs in (7) agree with those of (1) one can then set $\rho = t$, $\tau = r$ in (7). This gives at the same time the solution in the case (b), when $r$ is timelike, i.e. in the region $r < 2m$: it is still given by (7), only one has to set with $\tau = r, \rho = t$ if one compares (7) with (1). We note that $t$ is uniquely determined up to $t \to t + t_o$, because these are the only substitutions $t \to f(t)$ which leave the metric invariant. Thus all four *Schwarzschild coordinates* $t, r, \phi, \theta$ are essentially uniquely determined. We record the result meticulously as a theorem.

4.2.7 **Theorem (Schwarzschild)**. *Let $g$ be a spherically symmetric metric, which satisfies Einstein's field equations* $\mathrm{Ric}[g] = 0$. *In a neighbourhood of any point where the differential $dr$ of the radius function $r$ of $g$ is non–zero, the metric is the form (7) where $\phi, \theta$ are the usual coordinates on a sphere, $r$ is the radius function, and $t$ is uniquely determined up to $t \to t + t_o$. Such coordinates exist in a neighbourhood of any point where $r > 0$, $r \neq 2m$.*

The Schwarzschild coordinates cannot make sense at the *Schwarzschild radius* $r = 2m$, because the coefficients of the metric (7) in these coordinates become singular there. We momentarily suspend judgment as to whether the metric itself becomes singular there (if such points exist on the spacetime manifold) or if the singularity is an artifact of the coordinates (as is the case with polar coordinates in the plane at the origin).

The time translations $(t, r, \phi, \theta) \to (t + t_o, r, \phi, \theta)$ evidently leave the metric invariant, i.e. define isometries of the metric, in addition the rotations from SO(3). One says that the metric is static, which is known as *Birkhoff's Theorem*:

4.2.8 **Theorem (Birkhoff)**. *Any spherically symmetric metric, which satisfies Einstein's field equations* $\mathrm{Ric}[g] = 0$ *is static, i.e. one can choose the coordinates $\tau, \rho$ in (1) so that $A$ and $B$ are independent of $\tau$.*

**Remark**. The discussion above does not accurately reflect the historical development. Schwarzschild assumed from the outset that the metric is of the form (1) with $A$ and $B$ independent of $t$, so Birkhoff's theorem was not immediate from Schwarzschild's result. The definition of spherical symmetry used here came later.

The Schwarzschild metric (7) can be written in many other ways by introducing other coordinates $\tau, \rho$ instead of $t, r$ (there is no point in changing the coordinates $\phi, \theta$ on the sphere). For example, one can write the metric (7) in the form

$$ds^2 = -\Big(1 - \frac{2m}{r}\Big)dvdw + r^2(d\phi^2 + \sin^2\phi d\theta^2). \tag{8}$$

The coordinates $v, w$ are related to Schwarzschild's $t, r$ by the equations

$$v = t + r^*, \quad w = t - r^*$$

with

$$r^* = \int \frac{dr}{1 - 2m/r} = r + 2m\log(r - 2m).$$

In terms of these $v, w$, Schwarzschild's $r$ is determined by

$$\frac{1}{2}(v - w) = r + 2m\log(r - 2m).$$

Another possibility due to *Kruskal* (1960) is

$$ds^2 = \frac{16m^2 e^{-r/2m}}{r}\,(-d\tilde{t}^2 + d\tilde{x}^2) + r^2(d\phi^2 + \sin^2\phi d\theta^2). \tag{9}$$

The coordinates $\tilde{t}, \tilde{x}$ are defined by

$$\tilde{t} = \frac{1}{2}\left(e^{v/2} + e^{-w/2}\right), \quad \tilde{x} = \frac{1}{2}\left(e^{v/2} - e^{-w/2}\right)$$

and $r$ must satisfy

$$\tilde{t}^2 - \tilde{x}^2 = -(r - 2m)e^{r/2m}. \tag{10}$$

These coordinates $\tilde{t}, \tilde{x}$ must be restricted by

$$\tilde{t}^2 - \tilde{x}^2 < 2m \tag{11}$$

so that there is a solution of (10) with $r > 0$. What is remarkable is that the metric (9) is then defined for all $(\tilde{t}, \tilde{x}, \phi, \theta)$ satisfying (11) and the radius function $r$ can take all values $r > 0$. So the singularity at $r = 2m$ in the Schwarzschild metric has disappeared. This is explained by the fact that the coordinate transformation between the Scharzschild coordinates $t, r$ and the Kurskal coordinates $\tilde{t}, \tilde{x}$ is singular along $r = 2m$. One can now take the point of view that the whole region (11) belongs to $M$ with the metric given by (9). This means that the manifold $M$ is (by definition) the product $C \times S$ where $S$ is the sphere with coordinates $(\phi, \theta)$ and $C$ the region in the $(\tilde{t}, \tilde{x})$ plane described by (11). As in Schwarzschild's case, this region $C$ is composed of a region in which $r$ is spacelike $(r > 2m)$ and region in which $r$ is timelike $(0 < r < 2m)$, but now the metric stays regular at $r = 2m$. The Schwarzschild solution (7) now appears as the local expression of the metric in a subregion of $M$. One may now wonder whether the manifold $M$ can be still enlarged in a non-trivial way. But this is not the case: the Kruskal spacetime is the unique locally inextendible extension of the Schwarzschild metric, in a mathematically precise sense explained in the book by Hawking and Ellis (1973).

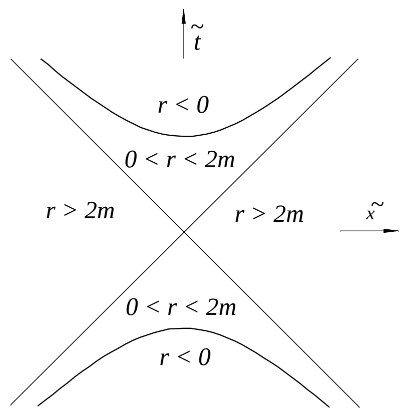



Fig.3

To get some feeling for this spacetime, consider again the situation in the $\tilde{t}\tilde{x}$–plane. The metric (9) has the agreeable property that the light cones in this plane are simply given by $-d\tilde{t}^2 + d\tilde{x}^2 = 0$. This means that along any timelike curve one must remain inside the cones $-d\tilde{t}^2 + d\tilde{x}^2 < 0$. This implies that anything that passes from the region $r > 2m$ into the region $r < 2m$ will never get back out, and this includes light. So we get the famous black hole effect. The hypersurface $r = 2m$ therefore still acts like a one–way barrier, even though the metric stays regular there. The region $0 < r < 2m$ is the fatal zone: once inside, you can count the time left to you by the "distance" $r$ from the singularity $r = 0$. In the region $r > 2m$ there is hope: some time-like curves go on forever, but others are headed for the fatal zone. (Whether there is hope for time-like *geodesics* is another matter.) The region $r < 0$ is out of this world. You will note that the singularity $r = 0$ consists of two pieces: a white hole in the past, a black hole in the future.

If one compares the equations §14- $(1'))$ and §14-(4) one comes to the conclusion that $c^2m$ is the mass of the centre, if $c$ is velocity of light, which was taken to be 1 in (7), but not in §14-(4). For a star like the sun the Schwarzschild radius $r = 2m = 2(\text{mass})/c^2$ is about 3 km, hence rather irrelevant, because it would lie deep inside, where one would not expect to apply the vacuum field equations anyway.

**Reference**. The discussion of spherical symmetry is taken from Hawking and Ellis *The large Scale Structure of Space–Time* (1973), who refer to original papers of Schmidt (1967) and Bergman, Cahen and Komar (1965).

**EXERCISES 4.3**

1. Prove the following statement from the proof of Lemma 4.2.5 in all detail. *The fixed-point set $C'(p_o)$ of $I(p_o)$ consists of the geodesics orthogonal to the 2-sphere $S(p_o)$ at $p_o$, as is evident for the corresponding sets in the Minkowski space.*

## 4.4 The rotation group SO(3)

What should we mean by a *rotation* in Euclidean 3–space $\mathbb{R}^3$? It should certainly be a transformation $a\colon \mathbb{R}^3 \to \mathbb{R}^3$ which fixes a point (say the origin) and preserves the Euclidean distance. You may recall from linear algebra that for a linear transformation $a \in M_3(\mathbb{R})$ of $\mathbb{R}^3$ the following properties are equivalent.

a) The transformation $p \to ap$ preserves Euclidean length $\|p\| = \sqrt{x^2 + y^2 + z^2}$:

$$\|ap\| = \|p\|$$

for all $p = (x, y, z) \in \mathbb{R}^3$.

b) The transformation $p \to ap$ preserves the scalar product $(p \cdot p') = xx' + yy' + zz'$:

$$(ap \cdot ap') = (p \cdot p')$$

for all $p, p' \in \mathbb{R}^3$.

c) The transformation $p \to ap$ is orthogonal: $aa^* = 1$ where $a^*$ is the adjoint (transpose) of $a$, i.e.

$$(ap \cdot p') = (p \cdot a^* p')$$

for all $p, p' \in \mathbb{R}^3$, and $1 \in M_3(\mathbb{R})$ is the identity transformation.

The set of all of these orthogonal transformations is denoted O(3). Let's now consider any transformation of $\mathbb{R}^3$ which preserves distance. We shall prove that it is a composite of a linear transformation $p \to ap$ and a translation $p \to p + b$, as follows.

4.3.1 **Theorem**. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be any map preserving Euclidean distance:*

$$\|T(q) - T(p)\| = \|q - p\|$$

*for all $q, p \in \mathbb{R}^3$. Then $T$ is of the form*

$$T(p) = ap + b$$

*where $a \in$O(3) and $b \in \mathbb{R}^3$.*

**Proof**. Let $p_o, p_1$ be two distinct points. The point $p_s$ on the straight line from $p_o$ to $p_1$ a distance $s$ from $p_o$ is uniquely characterized by the equations

$$\|p_1 - p_o\| = \|p_1 - p_s\| + \|p_s - p_o\|, \;\; \|p_s - p_o\| = s.$$

We have $p_s = p_o + t(p_1 - p_o)$ where $t = s/\|p_1 - p_o\|$. These data are preserved by $T$. Thus

$$T(p_o + t(p_1 - p_o)) = T(p_o) + t(T(p_1) - T(p_o)).$$

This means that for all $u, v \in \mathbb{R}^3$ and all $s, t \geq 0$ satisfying $s + t = 1$ we have

$$T(su + tv) = sT(u) + tT(v). \tag{1}$$

(Set $u = p_o, v = p_1, s = 1 - t$.) We may assume that $T(0) = 0$, after replacing $p \to T(p)$ by $p \to T(p) - T(0)$. The above equation with $u = 0$ then gives $T(tv) = tT(v)$. But then we can replace $s, t$ by $cs, ct$ in (1) for any $c \geq 0$, and (1) holds for all $s, t \geq 0$. Since $0 = T(v + (-v)) = T(v) + T(-v)$, we have $T(-v) = -T(v)$, hence (1) holds for all $s, t \in \mathbb{R}$. Thus $T(0) = 0$ implies that $T = a$ is linear, and generally $T$ is of the form $T(p) = a(p) + b$ where $b = T(0)$.
$\square$

The set O(3) of orthogonal transformations is a group, meaning the composite (matrix product) $ab$ of any two elements $a, b \in$O(3) belongs again to O(3), as does the inverse $a^{-1}$ of any element $a \in$O(3). These transformations $a \in$O(3) have determinant $\pm 1$ because $\det(a)^2 = \det(aa^*) = 1$. Those $a \in$O(3) with $\det(a) = +1$ form a subgroup, called the special orthogonal group or rotation group and is denoted SO(3). Geometrically, the condition $\det(a) = +1$ means that $a$ preserves orientation, i.e. maps a right–handed basis $v_1, v_2, v_3$ to a right–handed basis $av_1, av_2, av_3$. Right–handedness means that the triple scalar product $v_1 \cdot (v_2 \times v_3)$, which equals the determinant of the matrix $[v_1, v_2, v_3]$ of the matrix of components of the $v_i$ with respect to the *standard basis* $e_1, e_2, e_3$ of $\mathbb{R}^3$ should be positive. For an orthonormal basis this means that $v_1, v_2, v_3$ satisfies the same cross–product relations (given by the right hand rule) as $(e_1, e_2, e_3)$, i.e.

$$v_1 \times v_2 = v_3, \quad v_3 \times v_1 = v_2, \quad v_2 \times v_3 = v_1.$$

For reference we record the formula

$$v_1 \cdot (v_2 \times v_3) = \det[v_1, v_2, v_3]$$

which can in fact be used to *define* the cross–product. We need some facts about the *matrix exponential function*.

4.3.2 **Theorem**. *The series*

$$\exp X := \sum_{k=0}^{\infty} \frac{1}{k!} X^k$$

*converges for any* $X \in M_n(\mathbb{R})$ *and defines a smooth map* $\exp : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ *which maps an open neighbourhood* $U_0$ *of* $0$ *diffeomorphically onto an open neighbourhood* $U_1$ *of* $1$.

**Proof**. The Euclidean norm $\|X\|$ on $M_n(\mathbb{R})$, defined as the square-root of the sum of the squares of the matrix entries of $X$, satisfies $\|XY\| \leq \|X\|\|Y\|$. (We shall take this for granted, although it may be easily proved using the Schwarz inequality.) Hence we have *termwise*

$$\sum_{k=0}^{\infty} \frac{1}{k!} \|X^k\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} \|X\|^k. \tag{1}$$

It follows that the series for $\exp X$ converges in norm for all $X$ and defines a smooth function. We want to apply the Inverse Function Theorem to $X \to$

$\exp X$. From the series we see that

$$\exp X = 1 + X + o(X)$$

and this implies that the differential of exp at 0 is the identity map $X \to X$. $\square$

4.3.3 **Notation**. We write $a \to \log a$ for the local inverse of $X \to \exp X$. It is defined in $a \in U_1$. It can in fact be written as the log series

$$\log a = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(a-1)^k$$

but we don't need this.

We record some properties of the matrix exponential.

4.3.4 **Proposition**. *a) For all $X$,*

$$\frac{d}{dt}\exp tX = X\ (\exp tX) = \ (\exp tX)X$$

*b) If $XY = YX$, then*

$$\exp(X+Y) = \exp X \exp Y.$$

*c) For all $X$,*

$$\exp X \exp(-X) = 1.$$

*d) If $a(t) \in M_n(\mathbb{R})$ is a differentiable function of $t \in \mathbb{R}$ satisfying $\dot{a}(t) = Xa(t)$ then $a(t) = \exp(tX)a(0)$.*

**Proof.** a)Compute:

$$\begin{aligned}\frac{d}{dt}\exp tX &= \frac{d}{dt}\sum_{k=0}^{\infty}\frac{t^k}{k!}X^k = \sum_{k=1}^{\infty}\frac{kt^{k-1}}{k!}X^k = \sum_{k=0}^{\infty}\frac{t^k}{k!}X^{k+1} \\ &= X\sum_{k=0}^{\infty}\frac{t^k}{k!}X^k = X\ (\exp tX) = (\exp tX)X.\end{aligned}$$

b) If $XY = XY$, then $(X+Y)^n$ can be multiplied out and rearranged, which gives

$$\begin{aligned}e^{X+Y} &= \sum_k \frac{1}{k!}\ (X+Y)^k = \sum_k \frac{1}{k!} \sum_{i+j=k}\frac{k!}{i!j!}X^iY^j \\ &= \Big(\sum_i \frac{1}{i!}X^i\Big)\Big(\sum_j \frac{1}{j!}X^j\Big) = e^Xe^Y.\end{aligned}$$

These series calculations are permissible because of (1).

c) Follows form (b): $\exp X\ \exp(-X) = \exp(X-X) = \exp 0 = 1 + 0 + \cdots$.

d) Assume $\dot{a}(t) = Xa(t)$. Then

$$\begin{aligned}\frac{d}{dt}\Big(\exp(-tX)\,a(t)\Big) &= \frac{d\exp(-tX)}{dt}a(t) + \exp(-tX)\,\frac{da(t)}{dt}\\ &= (\exp(-tX))(-X)a(t) + (\exp(-tX))(Xa(t)) \equiv 0\end{aligned}$$

hence $(\exp(-tX))a(t)$ =constant= $a(0)$ and $a(t) = \exp(tX)a(0)$. $\square$

We now return to SO(3). We set

$$\mathbf{so}(3) = \{X \in M_3(\mathbb{R}) : X^* = -X\}.$$

Note that $\mathbf{so}(3)$ is a 3–dimensional vector space. For example, the following three matrices form a basis.

$$E_1 = \begin{bmatrix} 0 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{bmatrix},\ E_2 = \begin{bmatrix} 0 & 0 & 1\\ 0 & 0 & 0\\ -1 & 0 & 0\end{bmatrix},\ E_3 = \begin{bmatrix} 0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

4.3.5 **Theorem**. *For any* $a \in$SO(3) *there is a right–handed orthonormal basis* $u_1, u_2, u_3$ *of* $\mathbb{R}^3$ *so that*

$$au_1 = \cos\alpha u_1 + \sin\alpha u_2, \quad au_2 = -\sin\alpha u_1 + \cos\alpha u_2, \quad au_3 = u_3.$$

*Then* $a = \exp X$ *where* $X \in \mathbf{so}(3)$ *is defined by*

$$Xu_1 = \alpha u_2, \quad Xu_2 = -\alpha u_1 \quad Xu_3 = 0.$$

4.3.6 **Remark**. If we set $u = \alpha u_3$, then these equations say that

$$Xv = u \times v \tag{12}$$

for all $v \in \mathbb{R}^3$.

**Proof**. The eigenvalues $\lambda$ of a matrix $a$ satisfying $a^*a = 1$ satisfy $|\lambda| = 1$. For a real $3 \times 3$ matrix one eigenvalue will be real and the other two complex conjugates. If in addition $\det a = 1$, then the eigenvalues will have to be of the form $e^{\pm i\alpha}, 1$. The eigenvectors can be chosen to of the form $u_2 \pm iu_1, u_3$ where $u_1, u_2, u_3$ are real. If the eigenvalues are distinct these vectors are automatically orthogonal and otherwise may be so chosen. They may assumed to be normalized. The first relations then follow from

$$a(u_2 \pm iu_1) = e^{\pm i\alpha}(u_2 \pm iu_1),\ au_3 = u_3$$

since $e^{i\alpha} = \cos\alpha + i\sin\alpha$. The definition of $X$ implies

$$X(u_2 \pm iu_1) = \pm i\alpha(u_2 \pm iu_1),\ Xu_3 = 0.$$

Generally, if $Xv = \lambda v$, then

$$(\exp X)v = \sum_k \frac{1}{k!}X^k v = \sum_k \frac{1}{k!}\lambda^k v = e^\lambda v.$$

The relation $a = \exp X$ follows. $\square$

In the situation of the theorem, we call $u_3$ the *axis of rotation* of $a \in$SO(3), $\alpha$ the *angle of rotation*, and $a$ the right–handed rotation about $u_3$ with angle $\alpha$. We now set $V_0 = \mathbf{so}(3) \bigcap U_0$ and $V_1 = \mathrm{SO}(3) \bigcap U_1$ and assume $U_0$ chosen so that $X \in U_0 \Rightarrow X^* \in U_0$.

4.3.7 **Corollary**. *exp maps* **so**(3) *onto* SO(3) *and gives a bijection from* $V_0$ *onto* $V_1$.

**Proof**. The first assertion follows from the theorem. If $\exp X \in \mathrm{O}(3)$, then

$$\exp(X^*) = (\exp X)^* = (\exp X)^{-1} = \exp(-X).$$

If $\exp X \in U_1$, this implies that $X^* = -X$, i.e. $X \in \mathbf{so}(3)$, since $\exp : U_0 \to U_1$ is bijective. $\square$

We use the local inverse $\log : \mathrm{SO}(3) \cdots \to \mathbf{so}(3)$ of $\exp : \mathbf{so}(3) \to \mathrm{SO}(3)$ to define coordinates on SO(3) as follows. For $a \in V_1$ we write $a = \exp X$ with $X \in V_0$ and use this $X \in \mathbf{so}(3)$ as the coordinate point for $a \in$SO(3). If we want a coordinate point in $\mathbb{R}^3$ we use a basis $X_1, X_2, X_3$ for $\mathbf{so}(3)$, write $X = x^1 X_1 + x^2 X_2 + x^3 X_3$ and use $(x^1, x^2, x^3)$ as coordinate point. This gives coordinates around 1 in SO(3). To define coordinates around a general point $a_o \in$SO(3) we write

$$a = a_o \exp X, \ a \in a_o V_1, \ X \in V_0$$

and use this $X \in V_o$ as coordinate point for $a \in a_o V_1$. Thus $X = \log(a_o^{-1} a)$. We check the manifold axioms.

MAN 1. $a \to \log(a_o^{-1} a)$ maps $a_o V_1$ one–to–one onto $V_0$.

MAN 2. The coordinate transformation $X \to \tilde{X}$ between two such coordinate systems is defined by

$$a_o \exp X = \tilde{a}_o \exp \tilde{X}.$$

Its domain consists of the $X \in V_0$ for which $\tilde{a}_o^{-1} a_o \exp X \in U_1$ and then is given by

$$X = \log(\tilde{a}_o^{-1} a_o \exp X).$$

If follows that its domain is open and the map is smooth.

MAN 3. $a_o \in$SO(3) lies in the coordinate domain $a_o V_1$.

So the exponential coordinate systems make SO(3) into a manifold, directly from the axiom. But we could also use the following result.

4.3.8 **Theorem**. SO(3) *is a submanifold of* $M_3(\mathbb{R})$. *Its tangent space at* 1 *is* **so**(3) *and its tangent space at any* $a \in$SO(3) *is* $a\,\mathbf{so}(3) = \mathbf{so}(3)a$.

**Proof**. We first consider O(3)$= \{a \in M_3(\mathbb{R}) : aa^* = 1\}$. Consider the map $f$ from $M_3(\mathbb{R})$ into the vector space $\mathrm{Sym}_3(\mathbb{R})$ of symmetric $3 \times 3$ matrices given by $f(a) = aa^* - 1$. Its differential be calculated by writing $a = a(t)$ and taking the derivative:

$$df_a(\dot{a}) = \frac{d}{dt} f(a) = \frac{d}{dt}(aa^* - 1) = \dot{a}a^* + a\dot{a}^* = (\dot{a}a^*) + (\dot{a}a^*)^*.$$

As long as $a$ is invertible, any element of $\mathrm{Sym}_3(\mathbb{R})$ is of the form $Ya^* + (Ya^*)^*$ for some $Y$, so $df_a$ is surjective at such $a$. Since this is true for all $a \in$O(3) it follows that O(3) is a submanifold of $M_3(\mathbb{R})$. Furthermore, its tangent space at $a \in$O(3) consists of the $X \in M_3(\mathbb{R})$ satisfying $Ya^* + (a^*Y)^* = 0$. This means that $X = Ya^* = Ya^{-1}$ must belong to $\mathbf{so}(3)$ and so $Y \in \mathbf{so}(3)a$. The equality $a\,\mathbf{so}(3) = \mathbf{so}(3)a$ is equivalent to $a\,\mathbf{so}(3)a^{-1} = \mathbf{so}(3)$, which clear, since $a^{-1} = a^*$.

This proves the theorem with SO(3) replaced by O(3). But since SO(3) is the intersection of O(3) with the open set $\{a \in M_3(\mathbb{R}) : \det a > 0\}$ it holds for SO(3) as well. □

4.3.9 **Remark**. O(3) is the disjoint union of the two open subsets $\mathrm{O}^{\pm}(3) = \{a \in \mathrm{O}(3) \colon \pm \det a > 0\}$. $\mathrm{O}^+(3) = \mathrm{SO}(3)$ and is connected, because SO(3) is the image of the connected set $\mathbf{so}(3)$ under the continuous map exp. Since $\mathrm{O}^-(3) = c\mathrm{O}^+(3)$ for any $c \in$O(3) with $\det c = -1$ (e.g. $c = -1$), $\mathrm{O}^-(3)$ is connected as well. It follows O(3) has two connected components: the connected component of the identity element 1 is $\mathrm{O}^+(3) = \mathrm{SO}(3)$ and the other one is $\mathrm{O}^-(3) = c\mathrm{SO}(3)$. This means that SO(3) can be characterized as the set of elements O(3) which can be connected to the identity by a continuous curve $a(t)$ in O(3), i.e. transformations which can be obtained by a continuous motion starting from rest, preserving distance, and leaving the origin fixed, in perfect agreement with the notion of a "rotation".

We should verify that the submanifold structure on SO(3) is the same as the one defined by exponential coordinates. For this we have to verify that the exponential coordinates $a_oV_1 \to V_0$, $a \to X = \log(a_o^{-1}a)$, are also submanifold coordinates, which is clear, since the inverse map is $V_0 \to a_0V_1$, $X \to a_o \exp X$ .

There is a classical coordinate system on SO(3) that goes back to Euler (1775). (It was lost in Euler's voluminous writings until Jacobi (1827) called attention to it because of its use in mechanics.) It is based on the following lemma (in which $E_2$ and $E_3$ are two of the matrices defined before Theorem 4.3.5).

4.3.10 **Lemma**. *Every $a \in$SO(3) can be written in the form*

$$a = a_3(\theta)a_2(\phi)a_3(\psi)$$

*where*

$$a_3(\theta) = \exp(\theta E_3),\ \ a_2(\phi) = \exp(\phi E_2)$$

*and $0 \le \theta, \psi < 2\pi$, $0 \le \phi \le \pi$. Furthermore, $\theta, \phi, \psi$ are unique as long as $\phi \ne 0, \pi$.*

**Proof**. Consider the rotation-action of SO(3) on sphere $S^2$. The geographical coordinates $(\theta, \phi)$ satisfy

$$p = a_3(\theta)a_2(\phi)e_3.$$

Thus for any $a \in$O(3) one has an equation

$$ae_3 = a_3(\theta)a_2(\phi)e_3.$$

for some $(\theta, \phi)$ subject to the above inequalities and unique as long as $\phi \neq 0$. This equation implies that $a = a_3(\theta)a_2(\phi)b$ for some $b \in$SO(3) with $be_3 = e_3$. Such a $b \in$SO(3) is necessarily of the form $b = a_3(\psi)$ for a unique $\psi$, $0 \leq \psi \leq \pi$. $\square$

$\theta, \phi, \psi$ are the Euler angles of the element $a = a_3(\theta)a_2(\phi)a_3(\psi)$. They form a coordinate system on SO(3) with domain consisting of those $a$'s whose Euler angles satisfy $0 < \theta, \psi < 2\pi$, $0 < \phi < \pi$ (strict inequalities). To prove this it suffices to show that the map $\mathbb{R}^3 \to M_3(\mathbb{R})$, $(\theta, \phi, \psi) \to a_3(\theta)a_2(\phi)a_3(\psi)$, has an injective differential for these $(\theta, \phi, \psi)$. The partials of $a = a_3(\theta)a_2(\phi)a_3(\psi)$ are given by

$$\begin{aligned}
a^{-1}\frac{\partial a}{\partial \theta} &= -\sin\phi\cos\psi\ E_1 - \sin\phi\sin\psi\ E_2 + \cos\phi\ E_3 \\
a^{-1}\frac{\partial a}{\partial \phi} &= \sin\psi\ E_1 + \cos\psi\ E_2 \\
a^{-1}\frac{\partial a}{\partial \psi} &= E_3.
\end{aligned}$$

The matrix of coefficients of $E_1,\ E_2, E_3$ on the right has determinant$-\sin\phi$, hence the three elements of so(3) given by these equations are linearly independent as long as $\sin\phi \neq 0$. This proves the desired injectivity of the differential on the domain in question. $\square$

**EXERCISES 4.3**

1. Fix $p_o \in S^2$. Define $F$ :SO(3) $\to S^2$ by $F(a) = ap_o$. Prove that $F$ is a surjective submersion.

2. Identify $TS^2 = \{(p, v) \in S^2 \times \mathbb{R}^3 : p \in S^2, v \in T_pS^2 \subset \mathbb{R}^3\}$. The circle bundle over $S^2$ is the subset

$$S = \{(p, v) \in TS^2 \colon \|v\| = 1\}$$

of $TS^2$.
a) Show that $S$ is a submanifold of $TS^2$.
b) Fix $(p_o, v_o) \in S$. Show that the map $F$ :SO(3) $\to S$, $a \to (ap_o, av_o)$ is a diffeomorphism of SO(3) onto $S$.

3. a) For $u \in \mathbb{R}^3$, let $X_u$ be the linear transformation of $\mathbb{R}^3$ given by $X_u(v) := u \times v$ (cross product).
Show that $X_u \in \mathbf{so}(3)$ and that $u \to X_u$ is a linear isomorphism $\mathbb{R}^3 \to \mathbf{so}(3)$.
b) Show that $X_{au} = aX_ua^{-1}$ for any $u \in \mathbb{R}^3$ and $a \in$SO(3).
c) Show that $\exp X_u$ is the right–handed rotation about $u$ with angle $\|u\|$. [Suggestion. Use a right–handed o.n. basis $u_1, u_2, u_3$ with $u_3 = u/\|u\|$, assuming $u \neq 0$.]

d) Show that $u \to \exp X_u$ maps the closed ball $\{\|u\| \leq \pi\}$ onto SO(3), is one–to–one except that it maps antipodal points $\pm u$ on the boundary sphere $\{\|u\| = 1\}$ into the same point in SO(3). [Suggestion. Argue geometrically, using (c).]

4. Prove the formulas for the partials of $a = a_3(\theta)a_2(\phi)a_3(\psi)$. [Suggestion. Use the product rule on matrix products and the differentiation rule for $\exp(tX)$.]

5. Define an indefinite scalar product on $\mathbb{R}^3$ by the formula $(p \cdot p') = xx' + yy' - zz'$. Let $\mathrm{O}(2,1)$ be the group of linear transformations preserving this scalar product and $\mathrm{SO}(2,1)$ the subgroup of elements of determinant 1.
a) Make $\mathrm{O}(2,1)$ and $\mathrm{SO}(2,1)$ into manifolds using exponential coordinates.
b) Show that $\mathrm{O}(2,1)$ and $\mathrm{SO}(2,1)$ are submanifolds of $M_3(\mathbb{R})$ and determine the tangent spaces.
[Define now $a^*$ by $(ap \cdot p') = (p \cdot a^*p')$. Introduce $\mathrm{so}(2,1)$ as before. In this case the exponential map $\exp : \mathrm{so}(2,1) \to \mathrm{SO}(2,1)$ is not surjective and $\mathrm{SO}(2,1)$ is not connected, but you need not prove this and it does not matter for this problem.]

6. Continue with the setup of the preceding problem.
a) Show that for any two vectors $v_1, v_2 \in \mathbb{R}^3$ there is a unique vector, denoted $v_1 \times v_2$ satisfying

$$((v_1 \times v_2) \cdot v_3) = \det[v_1, v_2, v_3]$$

for all $v_1 \in \mathbb{R}^3$. [Suggestion. For fixed $v_1, v_2$ is linear in $v_3$. Any linear functional on $\mathbb{R}^3$ is of the form $v \to (u \cdot v)$ for a unique $u \in \mathbb{R}^3$. Why?]
b) For $u \in \mathbb{R}^3$, define $X_u \in M_3(\mathbb{R})$ by $X_u(v) = u \times v$. Show that $\exp X_u \in \mathrm{SO}(2,1)$ for all $u \in \mathbb{R}^3$ and that

$$X_{cu} = cX_uc^{-1}$$

for all $u \in \mathbb{R}^3$ and all $c \in \mathrm{SO}(2,1)$.
*c) Determine when $\exp X_u = \exp X_{u'}$.

7. Let $\mathrm{U}(2) = \{a \in M_2(\mathbb{C}) : aa^* = 1\}$ and $\mathrm{SU}(2) = \{a \in \mathrm{U}(2) : \det a = 1\}$.
[For $a \in M_2(\mathbb{C})$, $a^*$ denotes the adjoint of $a$ with respect to the usual Hermitian scalar product. $M_2(\mathbb{C})$ is considered as a real vector space in order to make it into a manifold. ]
a) Show that U(2) and SU(2) are submanifolds of $M_2(\mathbb{C})$ and determine their tangent spaces u(2) and su(2) at the identity 1.
b) Determine the dimension of U(2) and SU(2).
c) Let $V = \{Y \in M_2(\mathbb{C}) : Y^* = Y, \mathrm{tr}Y = 0\}$. Show that $V$ is a 3–dimensional real vector space with basis

$$F_1 = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix} \quad F_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad F_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and that the formula

$$(Y \cdot Y') = \frac{1}{2}\mathrm{tr}(YY')$$

defines a positive definite inner product on $V$.

d) For $a \in$U(2) define $T_a : V \to V$ by the formula $T_a(Y) = aYa^{-1}$. Show that $T_a$ preserves the inner product on $V$, i.e. $T_a \in$O$(V)$, the group of linear transformations of $V$ preserving the inner product. Show further that $a \to T_a$ is a group homomorphism, i.e. $T_{aa'} = T_a T_{a'}$ and $T_1 = 1$.
e) Let $Y \in V$. Show that $a(t) := \exp itY \in$SU(2) for all $t \in \mathbb{R}$ and that the transformation $T_{a(t)}$ of $V$ is the right–handed rotation around $Y$ with angle $2t\|Y\|$. [“Right–handed” refers to the orientation specified by the basis $(F_1, F_2, F_3)$ of $V$. Suggestion. Any $Y \in V$ is of the form $Y = \gamma c F_3 c^{-1}$ for some real $\gamma \geq 0$ and some $c \in$U(2). (Why?) This reduces the problem to $Y = F_3$. (How?)]
f) Show that $a \to T_a$ maps U(2) onto SO(3) and that $T_a = 1$ iff $a$ is a scalar matrix. Deduce that already SU(2) gets mapped onto SO(3) and that $a \in$SU(2) satisfies $T_a = 1$ iff $a = \pm 1$. Deduce further that $a, a' \in$SU(2) satisfy $T_a = T_{a'}$ iff $a' = \pm a$. [These facts are summarized by saying that SU(2)→SO(3) is a double covering of SO(3) by SU(2). Suggestion. If $a \in M_2(\mathbb{C})$ commutes with all matrices $Z$, then $a$ is a scalar matrix. Any matrix is of the form $Z = X + iY$ where $X, Y$ are Hermitian. (Why?)]

8. Let $\mathrm{SL}(2, \mathbb{R}) = \{a \in M_2(\mathbb{R}) : \det a = 1\}$.
a) Show that $\mathrm{SL}(2, \mathbb{R})$ is a submanifold of $M_2(\mathbb{R})$ and determine its tangent space $\mathrm{sl}(2, \mathbb{R}) = T_1\mathrm{SL}(2\mathbb{R})$ at the identity element 1. Show that the three matrices

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} .$$

form a basis for $\mathrm{sl}(2, \mathbb{R})$.
b) Show that

$$\exp(\tau E_1) = \begin{bmatrix} \mathrm{e}^{\tau} & 0 \\ 0 & \mathrm{e}^{-\tau} \end{bmatrix}, \quad \exp(\theta E_3) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

c) Show that the equation

$$a = \exp(\phi E_1) \exp(\tau E_2) \exp(\theta E_3)$$

defines a coordinate system $(\phi, \tau, \theta)$ in some neighbourhood of 1 in $\mathrm{SL}(2, \mathbb{R})$.

9. Let $G$ be one of the following groups of linear transformations of $\mathbb{R}^n$
   a) $\mathrm{GL}(n, \mathbb{R}) = \{a \in M_n(\mathbb{R}) : \det a \neq 0\}$
   b) $\mathrm{SL}(n, \mathbb{R}) = \{a \in \mathrm{GL}_n(\mathbb{R}) : \det a = 1\}$
   c) $\mathrm{O}(n, \mathbb{R}) = \{a \in M_n(\mathbb{R}) : aa^* = 1\}$
   d) $\mathrm{SO}(n) = \{a \in \mathrm{O}(n) = +1\}$
(1) Show that $G$ is a submanifold of $M_n(\mathbb{R})$ and determine its tangent space g at $1 \in G$.
(2) Determine the dimension of $G$.
(3) Show that exp maps g into $G$ and is a diffeomorphism from a neighbourhood of 0 in g onto a neighbourhood of 1 in $G$.
[For (b) you may want to use the differentiation formula for det:

$$\frac{d}{dt} \det a = (\det a) \mathrm{tr}(a^{-1} \frac{da}{dt})$$

if $a = a(t)$. Another possibility is to use the relation

$$\det \exp X = e^{\mathrm{tr} X}$$

which can be verified by choosing a basis so that $X \in M_n(\mathbb{C})$ becomes triangular, as is always possible.

## 4.5 Cartan's mobile frame

To begin with, let $S$ be any $m$-dimensional submanifold of $\mathbb{R}^n$. An *orthonormal frame* along $S$ associates to each point $p \in S$ a orthonormal $n$-tuple of vectors $(e_1, \cdots, e_n)$. The $e_i$ form a basis of $\mathbb{R}^n$ and $(e_i, e_j) = \delta_{ij}$ . We can think of the $e_i$ as functions on $S$ with values in $\mathbb{R}^n$ and these are required to be smooth. It will also be convenient to think of the variable point $p$ of $S$ as the function as the inclusion function $S \to \mathbb{R}^n$ which associates to each point of $S$ the same point as point of $\mathbb{R}^n$. So we have $n+1$ functions $p, e_1, \cdots, e_n$ on $S$ with values in $\mathbb{R}^n$; the for any function $f : S \to \mathbb{R}^n$, the differentials $dp, de_1, \cdots, de_n$ are linear functions $df_p : T_pS \to T_{f(p)}\mathbb{R}^n = \mathbb{R}^n$. Since $e_1,\ e_2, , e_n$ form a basis for $\mathbb{R}^n$, we can write the vector $df_p(v) \in \mathbb{R}^n$ as a linear combination $df(v) = \varpi^i(v)e_i$ . Everything depends also on $p \in S$, but we do not indicate this in the notation. The $\varpi^i$ are 1-forms on $S$ and we simply write $df = \varpi^i e_i$. In particular, when we take for $f$ the functions $p, e_1, \cdots, e_n$ we get 1-forms $\varpi^i$ and $\varpi^i_j$ on $S$ so that

$$dp = \varpi^i e_i, \quad de_i = \varpi^j_i e_j. \tag{1}$$

4.4.1 **Lemma.** *The 1-forms $\varpi^i$, $\varpi^i_j$ on $S$ satisfy Cartan's structural equations:*

*(CS1)* $d\varpi^i = \varpi^j \wedge \varpi^i_j$
*(CS2)* $d\varpi^j_i = \varpi^k_i \wedge \varpi^i_k$
*(CS3)* $\varpi^j_i = -\varpi^i_j$

**Proof**. Recall the differentiation rules $d(d\varphi) = 0$, $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{\deg \alpha} \alpha \wedge (d\beta)$.

(1)$0 = d(dp) = d(\varpi^i e_i) = (d\varpi^i)e_i - \varpi^i \wedge (de_i) = (d\varpi^i)e_i - \varpi^i \wedge \varpi^l_i e_l = (d\varpi^i - \varpi^j \wedge \varpi^i_j)e_i$.

(2)$0 = d(de_i) = d(\varpi^j_i e_j) = (d\varpi^j_i)e_j - \varpi^j_i \wedge (de_j) = (d\varpi^j_i)e_j - \varpi^j_i \wedge \varpi^l_j e_l = (d\varpi^j_i - \varpi^k_i \wedge \varpi^i_k)e_i$.

(3)$0 = d(e_i, e_j) = (de_i, e_j) + (e_i, de_j) = (\varpi^k_i e_k, e_j) + (e_i, \varpi^k_j e_k) = \varpi^j_i + \varpi^i_k$.
$\square$*From now on $S$ is a surface in $\mathbb{R}^3$.*

4.4.2 **Definition**. A *Darboux frame* along $S$ is an orthonormal frame $(e_1, e_2, e_3)$ along $S$ so that at any point $p \in S$,

(1) $e_1, e_2$ are tangential to $S$, $e_3$ is orthogonal to $S$
(2) $(e_1, e_2, e_3)$ is positively oriented for the standard orientation on $\mathbb{R}^3$.

The second condition means that the determinant of coefficient matrix $[e_1, e_2, e_3]$ of the $e_i$ with respect to the standard basis for $\mathbb{R}^3$ is positive. From now on $(e_1, e_2, e_3)$ will denote a Darboux frame along $S$ and we use the above notation $\varpi^i, \varpi_i^j$.

4.4.3 **Interpretation**. The forms $\varpi^i$ and $\varpi_i^j$ have the following interpretations.
a) For any $v \in T_pS$ we have $dp(v) = v$ considered as vector in $T_p\mathbb{R}^3$. Thus

$$v = \varpi^1(v)e_1 + \varpi^2(v)e_2. \tag{2}$$

This shows that $\varpi^1, \varpi^2$ are just the components of a general tangent vector $v \in T_pS$ to $S$ with respect to the basis $e_1, e_2$ of $T_pS$. We shall call the $\varpi^i$ the *component forms* of the frame. We note that $\varpi^1, \varpi^2$ depend only on $e_1, e_2$.
b) For any vector field $X = X^i e_i$ along $S$ in $\mathbb{R}^3$ and any tangent vector $v \in T_pS$ the (componentwise) directional derivative $D_vX$ (=covariant derivative in $\mathbb{R}^3$) is

$$D_vX = D_v(X^i e_i) = dX^i(v)e_i + X^i de_i(v) = dX^i(v)e_i + X^i \varpi_i^j(v)e_j\ .$$

Hence if $X = X^1e_1 + X^2e_2$ is a vector field on $S$, then the covariant derivative $\nabla_vX$ =tangential component of $D_vX$ is

$$\nabla_vX = \sum_{\alpha=1,2} (dX^\alpha(v) + \varpi_\beta^\alpha(v)X^\beta)e_\alpha$$

**Convention**. Greek indices $\alpha, \beta, \cdots$ run only over $\{1, 2\}$, Roman indices $i, j, k$ run over $\{1, 2, 3\}$.
(This is not important in the above equation, since a vector field $X$ on $S$ has normal component $X^3 = 0$ anyway.) In particular we have

$$\nabla_v e_\beta = \varpi_\beta^\alpha(v)e_\alpha \tag{3}$$

This shows that the $\varpi_\beta^\alpha$ determine the connection $\nabla$ on $S$. We shall call them the connection forms for the frame.

4.4.4 **Lemma**. *a) The Darboux frame $(e_1, e_2, e_3)$ satisfies the equations of motion*

$$dp = \varpi^1 e_1 + \varpi^2 e_2 + 0$$
$$de_1 = 0 + \varpi_1^2 e_2 + \varpi_1^3 e_3$$
$$de_2 = -\varpi_2^1 e_1 + 0 + \varpi_2^3 e_3$$
$$de_3 = -\varpi_1^3 e_1 - \varpi_2^3 e_2 + 0$$

*b) The forms $\varpi^i$, $\varpi_i^j$ satisfy Cartan's structural equations*

$$d\varpi^1 = -\varpi_2 \wedge \varpi_1^2, \quad d\varpi^2 = \varpi_1 \wedge \varpi_1^2$$
$$\varpi_1 \wedge \varpi_1^3 + \varpi_2 \varpi_2^3 = 0$$
$$d\varpi_1^2 = -\varpi_1^3 \wedge \varpi_2^3.$$

**Proof**. a) In the first equation, $\varpi^3 = 0$, because $dp/dt = dp(\dot{p})$ is tangential to $S$.

In the remaining three equation $\varpi_i^j = -\varpi_j^i$ by (CS3).
b) This follows from (CS1 –3) together with the relation $\varpi^3 = 0$ □

4.4.5 **Proposition**. *The Gauss curvature $K$ satisfies*

$$d\varpi_1^2 = -K\varpi^1 \wedge \varpi^2. \tag{4}$$

**Proof**. We use $N = e_3$ as unit normal for $S$. Recall that the Gauss map $N : S \to S^2$ satisfies

$$N^*(\text{area}_{S^2}) = K(\text{area}).$$

Since $e_1$, $e_2$ form an orthonormal basis for $T_pS = T_{N(p)}S^2$ the area elements at $p$ on $S$ and at $N(p)$ on $S^2$ are both equal to the 2–form $\varpi^1 \wedge \varpi^2$. Thus we get the relation

$$N^*(\varpi^1 \wedge \varpi^2) = K(\varpi^1 \wedge \varpi^2).$$

From the equations of motion,

$$dN = de_3 = -\varpi_1^3 e_1 - \varpi_2^3 e_2.$$

This gives

$$N^*(\varpi^1 \wedge \varpi^2) = (\varpi^1 \circ dN) \wedge (\varpi^2 \circ dN) \ = (-\varpi_1^3) \wedge (-\varpi_2^3) == \varpi_1^3 \wedge \varpi_2^3.$$

Hence

$$\varpi_1^3 \wedge \varpi_2^3 = K\varpi^1 \wedge \varpi^2.$$

From Cartan's structural equation $d\varpi_1^2 = -\varpi_1^3 \wedge \varpi_2^3$ we therefore get

$$d\varpi_1^2 = -K\varpi^1 \wedge \varpi^2. \qquad \square$$

**The Gauss–Bonnet Theorem**. Let $D$ be a bounded region on the surface $S$ and $C$ its boundary. We assume that $C$ is parametrized as a smooth curve $p(t)$, $t \in [t_o, t_1]$, with $p(t_o) = p(t_1)$. Fix a unit vector $v$ at $p(t_o)$ and let $X(t)$ be the vector at $p(t)$ obtained by parallel transport $t_o \to t$ along $C$. This is a smooth vector field along $C$. Assume further that there is a smooth orthonormal frame $e_1, e_2$ defined in some open set of $S$ containing $D$. At the point $p(t)$, write

$$X = (\cos\theta)e_1 + (\sin\theta)e_2$$

where $X, \theta, e_1, e_2$ are all considered functions of $t$. The angle $\theta(t)$ can be thought of the polar angle of $X(t)$ relative to the vector $e_1(p(t))$. It is defined only up to a constant multiple of $2\pi$. To say something about it, we consider the derivative of the scalar product $(X, e_i)$ at values of $t$ where $p(t)$ is differentiable. We have

$$\frac{d}{dt}(X, e_i) = (\frac{\nabla X}{dt}, e_i) + (X, \frac{\nabla e_i}{dt}).$$

This gives

$$\frac{d}{dt}\ (\cos\theta) = 0 + (\sin\theta)\ \varpi_1^2(\dot{p}), \qquad \frac{d}{dt}\ (\sin\theta) = 0 - (\cos\theta)\ \varpi_1^2(\dot{p})$$

hence

$$\dot{\theta} = -\varpi_1^2(\dot{p}).$$

Using this relation together with (4) and Stokes' Theorem we find that the total variation of $\theta$ around $C$, given by

$$\Delta_C\theta := \int_{t_0}^{t_1} \dot{\theta}(t)dt \tag{*}$$

is

$$\Delta_C\theta = \int_C -\varpi_1^2 = \int_D K\varpi^1 \wedge \varpi^2$$

This relation holds provided $D$ and $C$ are oriented compatibly. We now assume fixed an orientation on $S$ and write the area element $\varpi^1 \wedge \varpi^2$ as $dS$. With this notation,

$$\Delta_C\theta = \int_D K\ dS. \tag{**}$$

The relation (**) is actually valid in much greater generality. Up to now we have assume that the boundary $C$ of $D$ be a smooth curve. The equation (**) can be extended to any bounded region $D$ with a piecewise smooth boundary $C$ since we can approximate $D$ be "rounding off" the corners. The right side of (**) then approaches the integral of $KdS$ over $D$ under this approximation, the left side the sum of the integrals of $\dot{\theta}(t)dt = -\varpi_1^2|C$ over the smooth pieces of $C$. All of this still presupposes, however, that $D$ carries some smooth frame field $e_1, e_2$, although the value of (**) is independent of its particular choice (since the right side evidently is). One can paraphrase (**) by saying that the rotation angle of the parallel transport around a simple loop equals the surface integral of the Gauss–curvature over its interior. This is a local version of the classical Gauss–Bonnet theorem. ("Local" because it still assumes the existence of the frame field on $D$, which can be guaranteed only if $D$ is sufficiently small.)

**Calculation of the Gauss curvature**. The component forms $\varpi^1$, $\varpi^2$ for a basis of the 1–forms on $S$. In particular, one can write

$$\varpi_1^2 = \lambda_1\varpi^1 + \lambda_2\varpi^2$$

for certain scaler functions $\lambda_1, \lambda_2$. These can be calculated from Cartan's structural equations

$$d\varpi^1 = -\varpi_2 \wedge \varpi_1^2, \quad d\varpi^2 = \varpi_1 \wedge \varpi_1^2$$

which give

$$d\varpi^1 = \lambda_1\varpi_1 \wedge \varpi_2, \quad d\varpi^2 = \lambda_2\varpi_1 \wedge \varpi_2. \tag{5}$$

and find

$$-K\varpi^1 \wedge \varpi^2 = d\varpi_1^2 = d(\lambda_1 \wedge \varpi^1 + \lambda_2 \wedge \varpi^2). \tag{6}$$

We write the equations (5) symbolically as

$$\lambda_1 = \frac{d\varpi^1}{\varpi^1 \wedge \varpi^2} \quad \lambda_2 = \frac{d\varpi^2}{\varpi^1 \wedge \varpi^2}$$

(The "quotients" denote functions $\lambda_1, \lambda_2$ satisfying (5).) Then (6) becomes

$$-K\varpi^1 \wedge \varpi^2 = d\Big(\frac{d\varpi^1}{\varpi^1 \wedge \varpi^2}\varpi^1 + \frac{d\varpi^2}{\varpi^1 \wedge \varpi^2}\varpi^2\Big). \tag{7}$$

The main point is that $K$ is uniquely determined by $\varpi^1$ and $\varpi^2$, which in turn are determined by the Riemann metric on $S$: for any vector $v = \varpi^1(v)e_1 + \varpi^2(v)e_2$ we have $(v, v) = \varpi_1(v)^2 + \varpi_2(v)^2$, so

$$ds^2 = (\varpi_1)^2 + (\varpi_2)^2.$$

Hence $K$ depends only on the metric in $S$, not on the realization of $S$ as subset in the Euclidean space $\mathbb{R}^3$. One says that the Gauss curvature is an intrinsic property of the surface, which is what Gauss called the *theorema egregium*. It is a very surprising result if one thinks of the definition $K = \det dN$, which does use the realization of $S$ as a subset of $\mathbb{R}^3$; it is especially surprising when one considers that the space-curvature $\kappa$ of a curve $C$: $p = p(s)$ in $\mathbb{R}^3$ is not an intrinsic property of the curve, even though the definition $\kappa = \pm\|dT/ds\| =$ (scalar rate of change of tangential direction $T$ per unit length) looks superficially very similar to $K = \det dN =$(scalar rate of change of normal direction $N$ per unit area).

4.4.6 **Example: Gauss curvature in orthogonal coordinat**es. Assume that the surface $S$ is given parametrically as $p = p(s, t)$ so that the vector fields $\partial p/\partial s$ and $\partial p/\partial t$ are orthogonal. Then the metric is of the form

$$ds^2 = A^2 ds^2 + B^2 dt^2$$

for some scalar functions $A, B$. We may take

$$e_1 = A^{-1}\frac{\partial p}{\partial s}, \quad e_2 = B^{-1}\frac{\partial p}{\partial t}.$$

Then

$$\varpi^1 = Ads, \quad \varpi^2 = Bdt.$$

Hence

$$d\varpi^1 = -A_t ds \wedge dt, \quad d\varpi^2 = B_s ds \wedge dt, \quad \varpi^1 \wedge \varpi^2 = ABds \wedge dt$$

where the subscripts indicate partial derivatives. The formula (7) becomes

$$-KABds \wedge dt = d\Big(-\frac{A_t}{AB}Ads + \frac{B_s}{AB}Bdt\Big) = \Big(\Big(\frac{A_t}{B}\Big)_t + \Big(\frac{B_s}{A}\Big)_s\Big)ds \wedge dt$$

Thus

$$K = -\frac{1}{AB}\Big(\Big(\frac{A_t}{B}\Big)_t + \Big(\frac{B_s}{A}\Big)_s\Big) \tag{8}$$

**A footnote**. The theory of the mobile frame (*repère mobile*) on general Riemannian manifolds is an invention of Élie Cartan., expounded in his wonderful book *Leçons sur la géométrie des espaces de Riemann.* An adaptation to modern tastes can be found in th book by his son Henri Cartan entitled *Formes différentielles.*

**EXERCISES 4.4**

The following problems refer to a surface $S$ in $\mathbb{R}^3$ and use the notation explained above.

1. Show that the second fundamental form of $S$ is given by the formula

$$\Phi = \varpi^1 \varpi_1^3 + \varpi^2 \varpi_2^3.$$

2. a) Show that there are scalar functions $a, b, c$ so that

$$\varpi_1^3 = a\varpi^1 + b\varpi^2, \quad \varpi_2^3 = b\varpi^1 + c\varpi^2$$

[Suggestion. First write $\varpi_2^3 = d\varpi^1 + c\varpi^2$. To show that $d = b$, show first that $\varpi^1 \wedge \varpi^2 \wedge \varpi_2^3 = 0$.]

b) Show that the normal curvature in the direction of a unit vector $u \in T_pS$ which makes an angle $\theta$ with $e_1$ is

$$k(u) = a\cos^2\theta + 2b\sin\theta\cos\theta + c\sin^2\theta$$

[Suggestion. Use the previous problem.]

## 4.6 Weyl's gauge theory paper of 1929

The idea of gauge invariance is fundamental to present-day physics. It has a colourful history, starting with Weyl's 1918 attempt at a unification of Einstein's theory of gravitation and Maxwell's theory of electromagnetism –a false start as it turned out– then Weyl's second attempt in 1929 –another false start in a sense, because a quarter of a century ahead of Yang and Mills. The story is told thoroughly in O'Raifereartaigh's "*The Dawning of Gauge Theory*" and there is no point in retelling it here. Instead I suggest that Weyl's paper of 1929 may be read with profit and pleasure for what it has to say on issues very much of current interest. The paper should speak for itself and the purpose of the translation is to give it a chance to do so. This meant that I had to take a few liberties as translator, which I would not have taken with a document to be read for historical interest only, translating not only into another language but also into another time. Mainly I updated Weyl's notation to conform with what has become standard in the meantime. Everything of substance I have kept the way Weyl said it, including his delightful handling of infinitesimals, limiting the liberties (flagged in footnotes) to what I thought was the minimum compatible with the purpose of the translation.

WR

**ELECTRON AND GRAVITATION**
**Hermann Weyl**
Zeitschrift für Pysik 56, 330–352 (1929)

### Introduction

In this paper I develop in detailed form a theory comprising gravitation, electromagetism and matter, a short sketch of which appeared in the Proc. Nat. Acad., April 1929. Several authors have noticed a connection between Einstein's theory of parallelism at a distance and the spinor theory of the electron[1]. My approach is radically different, in spite of certain formal analogies, in that I reject parallelism at a distance and retain Einstein's classical relativity theory of gravitation.

The adaptation of the Pauli-Dirac theory of the electron spin promises to lead to physically fruitful results for two reasons. (1)The Dirac theory, in which the electron field is represented by a potential $\psi$ with four components, gives twice to many energy levels; one should therefore *be able to return to the two component field of Pauli's theory without sacrificing relativistic invariance. This is prevented by the term in Dirac's action which contains the mass $m$ of the electron as a factor. Mass, however, is a gravitational effect; so there is hope of finding a replacement for this term in the theory of gravitation which produces the desired correction. (2)The Dirac field equations for $\psi$ together with Maxwell's equations for the four potentials $A_\alpha$ for the electromagnetic field have an invariance property formally identical to the one I called gauge invariance in my theory of gravitation of 1918; the field equations remain invariant one replaces simultaneously

$$\psi \text{ by } e^{i\lambda}\psi \quad \text{and} \quad A_\alpha \text{ by } A_\alpha - \frac{\partial\lambda}{\partial x_\alpha}$$

where $\lambda$ denotes an arbitrary function of the place in four dimensional spacetime. A factor $e/c\hbar$ is incorporated into $A_\alpha$ ($-e$ is the charge of the electron, $c$ is the velocity of light, and $\hbar$ is Planck's constant divided by $2\pi$). The relation of "gauge invariance" to conservation of charge remains unchanged as well. But there is one essential difference, crucial for agreement with empirical data in that the exponent of the factor multiplying $\psi$ is not real but purely imaginary. $\psi$ now takes on the role played by the $ds$ of Einstein in that old gauge theory. It appears to me that this new principle of gauge invariance, which is derived from empirical data not speculation, indicates cogently that *the electric field is a necessary consequence of the electron field $\psi$ and not of the gravitational field.* Since gauge invariance involves an arbitrary function $\lambda$ it has the character of "general" relativity and can naturally be understood only in that framework.

There are several reasons why I cannot believe in parallelism at a distance. First of all, it a priori goes against my mathematical feeling to accept such an artificial geometry; I find it difficult to understand the force by which the local tetrads in at different spacetime points might have been rigidly frozen in their distorted positions. There are also two important physical reason, it seems to

[1] E. Wigner, ZS. f. Phys. 53, 592, 1929 and others.

me. It is exactly the removal of the relation between the local tetrads which changes the arbitrary gauge factor $e^{i\lambda}$ in $\psi$ from a constant to an arbitrary function; only this freedom explains the gauge invariance which in fact exists in nature. Secondly, the possibility of rotating the local tetrads independently of each other is equivalent with the *symmetry of the energy–momentum tensor* or with conservation of energy–momentum, as we shall see.

In any attempt to establish field equations one must keep in mind that these cannot be compared with experiment directly; only after quantization do they provide a basis for statistical predictions concerning the behaviour of matter particles and light quanta. The Dirac–Maxwell theory in its present form involves only the electromagnetic potentials $A_\alpha$ and the electron field $\psi$. No doubt, the proton field $\psi'$ will have to be added. $\psi, \psi'$ and $A_\alpha$ will enter as functions of the same spacetime coordinates into the field equations and one should not require that $\psi$ be a function of a spacetime point $(t, xyz)$ and $\psi'$ a function of an independent spacetime point $(t', x'y'z')$. It is natural to expect that one of the two pairs of components of Dirac's field represents the electron, the other the proton. Furthermore, there will have to be two charge conservation laws, which (after quantization) will imply that the number of electrons as well as the number of protons remains constant. To these will correspond a gauge invariance involving two arbitrary functions.

We first examine the situation in special relativity to see if and to what extent the increase in the number of components of $\psi$ from two to four is necessary because of the formal requirements of group theory, quite independently of dynamical differential equations linked to experiment. We shall see that two components suffice if the symmetry of left and right is given up.

### The two–component theory

**§1. The transformation law of $\psi$.** Homogeneous coordinates $j_0, \cdots, j_3$ in a 3 space with Cartesian coordinates $x, y, z$ are defined by the equations

$$x = \frac{j_1}{j_0}, y = \frac{j_2}{j_0}, z = \frac{j_3}{j_0}.$$

The equation of the unit sphere $x^2 + y^2 + z^2 = 1$ reads

$$(1) \qquad -j_0^2 + j_1^2 + j_2^2 + j_3^2 = 0.$$

If it is projected from the south pole onto the equator plane $z = 0$, equipped with the complex variable

$$x + iy = \zeta = \frac{\psi_2}{\psi_1}$$

then one has the equations

$$(2) \qquad \begin{array}{ll} j_0 = \bar{\psi}_1\psi_1 + \bar{\psi}_2\psi_2 & j_1 = \bar{\psi}_1\psi_2 + \bar{\psi}_2\psi_1 \\ j_2 = i(-\bar{\psi}_1\psi_2 + \bar{\psi}_2\psi_1) & j_3 = \bar{\psi}_1\psi_1 - \bar{\psi}_2\psi_2 \end{array}$$

The $j_\alpha$ are Hermitian forms of $\psi_1, \psi_2$. Only the ratios of variables $\psi_1, \psi_3$ and of the coordinates $j_\alpha$ are relevant. A linear transformation of the variables $\psi_1, \psi_2$ (with complex coefficients) produces a real linear transformation of the $j_\alpha$ preserving the unit sphere (1) and its orientation. It is easy to show, and well-known, that one obtains any such linear transformation of the $j_\alpha$ once and only once in this way.

Instead of considering the $j_\alpha$ as homogenous coordinates in 3 dimensional space, we now interpret them as coordinates in 4 dimensional spacetime and (1) as the equation of the light cone; we restrict the complex linear transformations $L$ of $\psi_1, \psi_2$ to those with determinant of absolute value 1. $L$ produces a Lorentz transformation $\Lambda = \Lambda(L)$ on the $j_\alpha$, i.e. a real linear transformation preserving the quadratic form

$-j_0^2 + j_1^2 + j_2^2 + j_3^2$.

The formulas for the $j_\alpha$ and the remarks on preservation of orientation imply that the Lorentz transformations thus obtained (1)do not interchange past and future and (2)have determinant+1, not $-1$. These transformations form a connected, closed continuum and are all so obtained, without exception. However, the linear transformation $L$ of the $\psi$ is not uniquely determined by $\Lambda$, but only up to an arbitrary factor $e^{i\lambda}$ of absolute value 1. $L$ can be normalized by the requirement that its determinant be equal to 1, but remains a double valued even then. The condition (1) is to be retained; it is one of the most promising aspects of the $\psi$–theory that is can account for the essential *difference between past and future*. The condition (2) removes the symmetry between left and right. It is only this symmetry, which actually exists in nature, that will force us (Part II[2]) to introduce a second pair of $\psi$–components.

Let $\sigma_\alpha$ , $\alpha = 0, \cdots, 3$, denote the coefficient matrix of the Hemitian form of the variables $\psi_1, \psi_2$ which represents $j_\alpha$ in (2):

(3) $\quad j_\alpha = \psi^* \sigma_\alpha \psi.$

$\psi$ is taken as a 2–column, $\psi^*$ is its conjugate transpose. $\sigma_0$ is the identity matrix; one has the equations

(4) $\quad \sigma_1^2 = 1, \quad \sigma_2\sigma_3 = \mathrm{i}\sigma_1$

and those obtained from these by cyclic permutation of the indices $1, 2, 3$.

It is formally more convenient to replace the real variable $j_0$ by the imaginary variable $\mathrm{i}j_0$. The Lorentz transformations $\Lambda$ then appear as orthogonal transformations of the four variables

$j(0) = \mathrm{i}j_0, j(\alpha) = j_\alpha$ for $\alpha = 1, 2, 3$.

Instead of (3) write

(5) $\quad j(\alpha) = \psi^* \sigma(\alpha) \psi.$

so that $\sigma(0) = \mathrm{i}\sigma_0$, $\sigma(\alpha) = \sigma_\alpha$ for $\alpha = 1, 2, 3$. The transformation law of the components of the $\psi$ field relative to a Lorentz frame in spacetime is characterized by the requirement that quantities $j(\alpha)$ in (5) undergo the Lorentz transformation $\Lambda$ under if the Lorentz frame does. A quantity of this type represents the field of a matter particle, as follows from the spin phenomenon. The $j(\alpha)$ are the components of vector relative to a the Lorentz frame $e(\alpha)$; $e(1), e(2), e(3)$ are real space-like vectors forming left-handed Cartesian coordinate system, $e(0)/\mathrm{i}$ is a real, time-like vector directed toward the future. The transformation $\Lambda$ describes the transition from one such Lorentz frame another, and will be referred to as a *rotation of the Lorentz frame*. We get the same coefficients $\Lambda(\alpha\beta)$ whether we make $\Lambda$ act on the basis vectors $e(\alpha)$ of the tetrad

[2]Never written, as far as I know. [WR]

or on the components $j(\alpha)$:

$j = \sum_\alpha j(\alpha)e(\alpha) = \sum_\alpha j'(\alpha)e'(\alpha),$

if

$e'(\alpha) = \sum_\beta \Lambda(\alpha\beta)e(\beta), \quad j'(\alpha) = \sum_\beta \Lambda(\alpha\beta)j(\beta);$

this follows from the orthoganality of $\Lambda$.

For what follows it is necessary to compute the infinitesimal transformation

(6) $\quad \delta\psi = \delta L.\psi$

which corresponds to an infinitesimal rotation $\delta j = \delta\Lambda.j$. [Note1.]

The transformation (6) is assumed to be normalized so that the trace of $\delta L$ is 0. The matrix $\delta L$ depends linearly on the $\delta\Lambda$; so we write

$\delta L = \frac{1}{2}\sum_{\alpha\beta}\delta\Lambda(\alpha\beta)\sigma(\alpha\beta) = \sum_{\alpha\beta}\delta\Lambda(\alpha\beta)\sigma(\alpha\beta)$

for certain complex $2\times 2$ matrices $\sigma(\alpha\beta)$ of trace 0 depending skew-symetrically on $(\alpha\beta)$ defined by this equation. The last sum runs only over the index pairs

$(\alpha\beta) = (01), (02), (03); \quad (23), (31), (12).$

One must not forget that the skew-symmetric coefficients $\delta\Lambda(\alpha\beta)$ are purely imaginary for the first three pairs, real for the last three, but arbitrary otherwise. One finds

(7) $\quad \sigma(23) = -\frac{1}{2\mathrm{i}}\sigma(1), \quad \sigma(01) = \frac{1}{2\mathrm{i}}\sigma(1)$

and two analogous pairs of equations resulting from cyclic permutation of the indices $1, 2, 3$. To verify this assertion one only has to check that the two infinitesimal transformations $\delta\psi = \delta L\psi$ given by

$\delta\psi = \frac{1}{2\mathrm{i}}\sigma(1)\psi, \quad \text{and} \quad \delta\psi = \frac{1}{2}\sigma(1)\psi,$

correspond to the infinitesimal rotations $\delta j = \delta\Lambda j$ given by

$\delta j(0) = 0, \quad \delta j(1) = 0, \quad \delta j(2) = -j(3), \qquad \delta j(3) = j(2),$

resp.

$\delta j(0) = \mathrm{i}j(1), \quad \delta j(1) = -\mathrm{i}j(0), \quad \delta j(2) = 0, \qquad \delta j(3) = 0.$

**§2. Metric and parallel transport**. We turn to general relativity. We characterize the metric at a point $x$ in spacetime by a *local* Lorentz frame or *tetrad* $e(\alpha)$. Only the class of tetrads related by rotations $\Lambda$ is determined by the metric; one tetrad is chosen arbitrarily from this class. Physical laws are invariant under arbitrary rotations of the tetrads and these can be chosen independently of each other at different point. Let $\psi_1(x)$, $\psi_2(x)$,be the components of the matter field at the point $x$ relative to the local tetrad $e(\alpha)$ chosen there. A vector $v$ at $x$ can be written in the form

$v = \sum_\alpha v(\alpha)e(\alpha);$

For the analytic representation of spacetime we shall a coordinate system $x^\mu, \mu = 0, \cdots, 3;$ the $x^\mu$ are four arbitrary, differentiable functions on spacetime distinguishing the spacetime points. The physical laws are therefore *invariant under arbitrary coordinate transformations*. Let $e^\mu(\alpha)$ denote the coordinates of the vector $e(\alpha)$ relative to the coordinate system $x^\mu$. These $4\times 4$ quantities $e^\mu(\alpha)$ characterize the gravitational field. The contravariant components $v^\mu$ of a vector $v$ with respect to the coordinate system are related to its components $v(\alpha)$ with respect to the tetrad by the equations

$v^\mu = \sum_\alpha v(\alpha)e^\mu(\alpha).$

On the other hand, the $v(\alpha)$ can be calculate from the contravariant components $v_\mu$ by means of

$$v(\alpha) = \sum_\mu v_\mu e^\mu(\alpha).$$

These equations govern the change of indices. The indices $\alpha$ referring to the tetrad I have written as arguments because for these upper and lower positions are indistinguishable. The change of indices in the reverse sense is accomplished by the matrix $(e_\mu(\alpha))$ inverse to $(e^\mu(\alpha))$ :

$$\sum_\alpha e_\mu(\alpha) e^\nu(\alpha) = \delta^\nu_\mu \quad \text{and} \quad \sum_\mu e_\mu(\alpha) e^\mu(\beta) = \delta(\alpha, \beta).$$

The symbol $\delta$ is 0 or 1 depending on whether its indices are the same or not. The convention of omitting the summation sign applies from now on to all repeated indices. Let $\epsilon$ denote the absolute value of the determinant $\det(e^\mu(\alpha))$. Division by of a quantity by $\epsilon$ will be indicated by a change to bold type, e.g.

$$\boldsymbol{e}^\mu(\alpha) = e^\mu(\alpha)/\epsilon.$$

A vector or a tensor can be represented by its components relative to the coordinate system as well as by its components relative to the tetrad; but the quantity $\psi$ can only be represented by its components relative to the tetrad. For its transformation law is governed by a representation of the Lorentz group which does not extend to the group of all linear transformations. In order to accommodate the matter field $\psi$ it is therefore necessary to represent the gravitational field in the way described[3] rather than in the form

$$\sum_{\mu\nu} g_{\mu\nu} dx^\mu dx^\nu.$$

The theory of gravitation must now be recast in this new analytical form. I start with the formulas for the *infinitesimal parallel transport* determined by the metric. Let the vector $e(\alpha)$ at the point $x$ go to the vector $e'(\alpha)$ at the infinitely close point $x'$ by parallel transport. The $e'(\alpha)$ form a tetrad at $x'$ arising from the tetrad $e(\alpha)$ at $x$ by an infinitesimal rotation $\omega$ :

(8) $$\nabla e(\beta) = \sum_\gamma \omega(\beta\gamma).e(\gamma), \qquad [\nabla e(\beta) = e'(\beta) - e(\beta; x')].$$

$\nabla e(\beta)$ depends linearly on the vector $v$ from $x$ to $x'$; if its components $dx^\mu$ equal $v^\mu = e^\mu(\alpha)v(\alpha)$ then $\omega(\beta\gamma) = \omega_\mu(\beta\gamma)dx^\mu$ equals

(9) $$\omega_\mu(\beta\gamma)v^\mu = \omega(\alpha; \beta\gamma)v(\alpha).$$

The infinitesimal parallel transport of a vector $w$ along $v$ is given by the well-known equations

$$\nabla w = -\Gamma(v).w, \quad \text{i.e.} \quad \nabla w^\mu = -\Gamma^\mu_\rho(v) w^\rho, \quad \Gamma^\mu_\rho(v) = \Gamma^\mu_{\rho\nu} v^\nu;$$

the quantities $\Gamma^\mu_{\rho\nu}$ are symmetric in $\rho\nu$ and independent of $w$ and $v$. We therefore have

$$e'(\beta) - e(\beta) = -\Gamma(v).e(\beta)$$

in addition to (8). Subtracting the two differences on the left-hand sides gives the differential $de(\beta) = e(\beta, x') - e(\beta, x)$ :

$$de^\mu(\beta) + \Gamma^\mu_\rho(v) e^\rho(\beta) = -\omega(\beta\gamma).e^\mu(\gamma),$$

or

$$\frac{\partial e^\mu(\beta)}{\partial x^\nu} e^\nu(\alpha) + \Gamma^\mu_{\rho\nu} e^\rho(\beta) e^\nu(\alpha) = -\omega(\alpha; \beta\gamma).e^\mu(\gamma).$$

Taking into account that the $\omega(\alpha; \beta\gamma)$ are skew symmetric in $\beta$ and $\gamma$, one can eliminate the $\omega(\alpha; \beta\gamma)$ and find the well-known equations for the $\Gamma^\mu_{\rho\nu}$. Taking

[3] In formal agreement with Einstein's recent papers on gravitation and electromagnetism, Sitzungsber. Preuß. Ak. Wissensch. 1928, p.217, 224; 1920 p.2. Einstein uses the letter $h$ instead of $e$.

into account that the $\Gamma^{\mu}_{\rho\nu}$ and the $\Gamma^{\mu}(\beta,\alpha) = \Gamma^{\mu}_{\rho\nu} e^{\rho}(\beta) e^{\nu}(\alpha)$ symmetric in $\rho$ and $\nu$ one can eliminate the $\Gamma^{\mu}_{\rho\nu}$ and finds

(10) $\quad \frac{\partial e^{\mu}(\alpha)}{\partial x^{\nu}} e^{\nu}(\alpha) - \frac{\partial e^{\mu}(\beta)}{\partial x^{\nu}} e^{\nu}(\alpha) = (\omega(\alpha;\beta\gamma) - \omega(\beta;\alpha\gamma))\, e^{\mu}(\gamma).$

The left-hand side is a component of the Lie bracket of the two vector fields $e(\alpha), e(\beta)$, which plays a fundamental role in Lie's theory of infinitesimal transformations, denoted $[e(\alpha), e(\beta)]$. Since $\omega(\beta;\alpha\gamma)$ is skew-symmetric in $\alpha$ and $\gamma$ one has

$$[e(\alpha), e(\beta)]^{\mu} = (\omega(\alpha;\beta\gamma) + \omega(\beta;\gamma\alpha))\, e^{\mu}(\gamma).$$

or

(11) $\quad \omega(\alpha;\beta\gamma) + \omega(\beta;\gamma\alpha) = [e(\alpha), e(\beta)](\gamma).$

If one takes the three cyclic permutations of $\alpha\beta\gamma$ in these equations and adds the resulting equations with the signs$+ - +$ then one obtains

$$2\omega(\alpha;\beta\gamma) = [e(\alpha), e(\beta)](\gamma) - [e(\beta), e(\gamma)](\alpha) + [e(\gamma), e(\alpha)](\beta).$$

$\omega(\alpha;\beta\gamma)$ is therefore indeed uniquely determined. The expression for it found satisfies all requirements, being skew-symmetric in $\beta$ and $\gamma$, as is easily seen.

In what follows we need in particular the contraction (sum over $\rho$)

$$\omega(\rho;\rho\alpha) = [e(\alpha), e(\rho)](\rho) = \frac{\partial e^{\mu}(\alpha)}{\partial x^{\mu}} e^{\nu}(\alpha) - \frac{\partial e^{\mu}(\rho)}{\partial x^{\nu}} e^{\nu}(\alpha) e_{\mu}(\rho),$$

Since $\epsilon = |\det e^{\mu}(\alpha)|$ satisfies

$$-\epsilon d(\tfrac{1}{\epsilon}) = \tfrac{d\epsilon}{\epsilon} = e_{\nu}(\rho) de^{\nu}(\rho)$$

one finds

(12) $\quad \omega(\rho;\rho\alpha) = \epsilon \frac{\partial \boldsymbol{e}^{\mu}(\alpha)}{\partial x\mu^{\nu}},$

where $\boldsymbol{e}(\alpha) = e(\alpha)/\epsilon$.

**§3. The matter action**. With the help of the parallel transport one can compute not only the covariant derivative of vector and tensor fields, but also that of the $\psi$ field. Let $\psi_a(x)$ and $\psi_a(x')$ $[a = 1, 2]$ denote the components relative to a local tetrad $e(\alpha)$ at the point $x$ and at an infinitely close point $x'$. The difference $\psi_a(x') - \psi_a(x) = d\psi_a$ is the usual differential. On the other hand, we parallel transport the tetrad $e(\alpha)$ at $x$ to a tetrad $e'(\alpha)$ at $x'$. Let $\psi'_a$ denote the components of $\psi$ at $x'$ relative to the tetrad $e'(\alpha)$. Both $\psi_a$ and $\psi\text{'}_a$ depend only on the choice of the tetrad $e(\alpha)$ at $x$; they have nothing to with the tetrad at $x'$. Under a rotation of the tetrad at $x$ the $\psi'_a$ transform like the $\psi_a$ and the same holds for the differences $\nabla\psi_a = \psi'_a - \psi_a$. These are the components of the *covariant differential* $\nabla\psi$ of $\psi$. The tetrad $e'(\alpha)$ arises from the local tetrad $e(\alpha) = e(\alpha, x')$ at $x'$ by the infinitesimal rotation $\omega$ of §2. The corresponding infinitesimal transformation $\theta$ is therefore of the form

$$\theta = \tfrac{1}{2}\omega(\beta\gamma)\sigma(\beta\gamma)$$

and transforms $\psi_a(x')$ into $\psi'_a$, i.e. $\psi' - \psi(x') = \theta.\psi$. Adding $d\psi = \psi(x') - \psi(x)$ to this one obtains

(13) $\quad \nabla\psi = d\psi + \theta.\psi.$

Everything depends linearly on the vector $v$ from $x$ to $x'$; if its components $dx^{\mu}$ equal $v^{\mu} = e^{\mu}(\alpha)v(\alpha)$, then $\nabla\psi = \nabla\psi_{\mu} dx^{\mu}$ equals

$$\nabla\psi_{\mu} v^{\mu} = \nabla\psi(\alpha) v(\alpha), \quad \theta = \theta_{\mu} v^{\mu} = \theta(\alpha) v(\alpha).$$

We find

$$\nabla\psi_{\mu} = (\tfrac{\partial}{\partial x^{\mu}} + \theta_{\mu})\psi \quad \text{or} \quad \nabla\psi(\alpha) = (e^{\mu}(\alpha)\tfrac{\partial}{\partial x^{\mu}} + \theta(\alpha))\psi$$

where

$$\theta(\alpha) = \tfrac{1}{2}\omega(\alpha;\beta\gamma)\sigma(\beta\gamma).$$

Generally, if $\psi'$ is a field the same type as $\psi$ then the quantities

$$\psi^*\sigma(\alpha)\psi'$$

are the components of a vector relative to the local tetrad. Hence

$$v'(\alpha) = \psi^*\sigma(\alpha)\nabla\psi(\beta)v(\beta)$$

defines a linear transformation $v \mapsto v'$ of the vector space at $x$ which is independent of the tetrad. Its trace

$$\psi^*\sigma(\alpha)\nabla\psi(\alpha)$$

therefore a scalar and the equation

$$\text{(14)} \qquad \mathrm{i}\epsilon\,\boldsymbol{m} = \psi^*\sigma(\alpha)\nabla\psi(\alpha)$$

defines a scalar density $\boldsymbol{m}$ whose integral

$$\int \boldsymbol{m}\,dx \quad [dx = dx^0dx^1dx^2dx^3]$$

can be used as the action of the matter field $\psi$ in the gravitational field represented by the metric defining the parallel transport. To find an explicit expression for $\boldsymbol{m}$ we need to compute

$$\text{(15)} \qquad \sigma(\alpha)\theta(\alpha) = \tfrac{1}{2}\sigma(\alpha)\theta(\beta\gamma).\omega(\alpha;\beta\gamma).$$

From (7) and (4) it follows that for $\alpha \neq \beta$

$$\sigma(\beta)\theta(\beta\alpha) = \tfrac{1}{2}\sigma(\alpha) \quad [\text{no sum over } \beta]$$

and for any odd permutation $\alpha\beta\gamma\delta$ of the indices $0\,1\,2\,3$,

$$\sigma(\beta)\theta(\gamma\delta) = \tfrac{1}{2}\sigma(\alpha).$$

These two kinds of terms to the sum (15) give the contributions

$$\tfrac{1}{2}\omega(\rho;\rho\gamma) = \frac{1}{2\epsilon}\,\frac{\partial \boldsymbol{e}^\mu(\alpha)}{\partial x^\mu}$$

resp.

$$\tfrac{\mathrm{i}}{2}\varphi(\alpha) := \omega(\beta;\gamma\delta) + \omega(\gamma;\delta\beta) + \omega(\delta;\beta\gamma).$$

If $\alpha\beta\gamma\delta$ is an odd permutation of the indices $0\,1\,2\,3$,then according to (11),

$$\tfrac{\mathrm{i}}{2}\varphi(\alpha) := [e(\beta), e(\gamma)] + +(\text{cycl.perm of } \beta\gamma\delta)$$

$$\text{(16)} \qquad\qquad = \sum \pm \frac{\partial e^\mu(\beta)}{\partial x^\nu}e^\nu(\gamma)e_\mu(\delta)$$

The sum run over the six permutations of $\beta\gamma\delta$ with the appropriate signs (and of course also over $\mu$ and $\nu$). With this notation,

$$\text{(17)} \qquad \boldsymbol{m} = \frac{1}{\mathrm{i}}\left(\psi^*\boldsymbol{e}^\mu(\alpha)\sigma(\alpha)\frac{\partial\psi}{\partial x^\mu} + \frac{1}{2}\frac{\partial \boldsymbol{e}^\mu(\alpha)}{\partial x^\mu}\,\psi^*\sigma(\alpha)\psi\right) + \frac{1}{4\epsilon}\varphi(\alpha)j(\alpha).$$

The second part is

$$\frac{1}{4\mathrm{i}\epsilon}\det\left[e_\mu(\alpha), e^\nu(\alpha), \frac{\partial e^\mu(\alpha)}{\partial x^\nu}, j(\alpha)\right]$$

(sum over $\mu$ and $\nu$). Each term of this sum is a $4\times 4$ determinant whose rows are obtained from the one written by setting $\alpha = 0, 1, 2, 3$. The quantity $j(\alpha)$ is

$$\text{(18)} \qquad j(\alpha) = \psi^*\sigma(\alpha)\psi.$$

Generally, it is not an action integral

$$\text{(19)} \qquad \int \boldsymbol{h}\,dx$$

itself which is of significance for the laws of nature, but only is variation $\delta\int \boldsymbol{h}\,dx$. Hence it is not necessary that $\boldsymbol{h}$ itself be real, but it is sufficient that $\bar{\boldsymbol{h}}-\boldsymbol{h}$ be a divergence. In that case we say that $\boldsymbol{h}$ is practically real.

We have to check this for $\boldsymbol{m}$. $e^\mu(\alpha)$ is real for $\alpha = 1, 2, 3$ and purely imaginary for $\alpha = 0$. So $e^\mu(\alpha)\sigma(\alpha)$ is a Hermitian matrix. $\varphi(\alpha)$ is also real fro $\alpha = 1, 2, 3$ and purely imaginary for $\alpha = 0$. Thus

$$\bar{\boldsymbol{m}} = -\tfrac{1}{\mathrm{i}}\left(\psi^{*\,\mu}\frac{\partial\psi}{\partial x^\mu} + \tfrac{1}{2}\frac{\partial \boldsymbol{e}^\mu(\alpha)}{\partial x^\mu}\,\psi^*\sigma(\alpha)\psi\right) + \tfrac{1}{4\epsilon}\varphi(\alpha)j(\alpha).$$

$$\mathrm{i}(\boldsymbol{m}-\bar{\boldsymbol{m}}) = \psi^*\boldsymbol{\sigma}^\mu\frac{\partial\psi}{\partial x^\mu} + \frac{\partial\psi^*}{\partial x^\mu}\boldsymbol{\sigma}^\mu\psi + \frac{\partial \boldsymbol{e}^\mu(\alpha)}{\partial x^\mu}\,\psi^*\sigma(\alpha)\psi$$
$$= \frac{\partial}{\partial x^\mu}(\psi^*\boldsymbol{\sigma}^\mu\psi) = \frac{\partial \boldsymbol{j}^\mu}{\partial x^\mu}.$$

Thus $\boldsymbol{m}$ is indeed partially real. We return to special relativity if we set

$$e^0(0) = -\mathrm{i}, \quad e^1(1) = e^2(2) = e^3(3) = 1,$$

and all other $e^\mu(\alpha) = 0$.

**§4. Energy**. Let (19) be the action of matter in an extended sense, represented by the $\psi$ field and by the electromagnetic potentials $A_\mu$. The laws of nature say that the variation

$$\delta\int \boldsymbol{h}\,dx = 0$$

when the $\psi$ and $A_\mu$ undergo arbitrary infinitesimal variations which vanish outside of a finite region in spacetime. [Note 2.]

The variation of the $\psi$ gives the equations of matter in the restricted sense, the variation of the $A_\mu$ the electromagnetic equations. If the $e^\mu(\alpha)$, which were kept fixed up to now, undergo an analogous infinitesimal variation $\delta e^\mu(\alpha)$, then there will be an equation of the form

$$\delta\int \boldsymbol{h}\,dx = \int \boldsymbol{T}_\mu(\alpha)\delta e^\mu(\alpha)\,dx, \tag{20}$$

the induced variations $\delta\psi_a$ and $\delta A_\mu$ being absent as a consequence of the preceding laws. The tensor density $\boldsymbol{T}_\mu(\alpha)$ defined in this way the energy-momentum. Because of the invariance of the action density $\boldsymbol{h}$, the variation of (20) must vanish when the variation $\delta \mathrm{e}^\mu(\alpha)$ is produced

(1)by infinitesimal rotations of the local tetrads $e(\alpha)$, the coordinates $x^\mu$ being kept fixed.

(2)by an infinitesimal transformation of the coordinates $x^\mu$, the tetrads $e(\alpha)$ being kept fixed. The first process is represented by the equations

$$\delta e^\mu\ (\alpha) = \delta\Lambda(\alpha\beta).e^\mu(\beta)$$

where $\delta\Lambda$ is a skew symmetric matrix, an infinitesimal rotation depending arbitrarily on the point $x$. The vanishing of (20) says that

$$\boldsymbol{T}(\beta,\alpha)= \boldsymbol{T}_\mu(\alpha)e^\mu(\beta)$$

is symmetric in $\alpha\beta$. *The symmetry of the energy-momentum tensor is therefore equivalent with the first invariance property.* This symmetry law is however not satisfied identically, but as a consequence of the law of matter and electromagnetism. For the components of a given $\psi$-field will change as a result of the rotation of the tetrads!

The computation of the variation $\delta e^\mu(\alpha)$ produced by the second process is somewhat more cumbersome. But the considerations are familiar from the theory of relativity in its previous analytical formulation[4]. [Consider an infinitesimal transformation of spacetime $\delta x = \xi(x)$. It induces an infinitesimal transformation of the vector field $e(\alpha)$ denoted $\delta e(\alpha)$. The transformed vector fields form a tetrad for the isometrically transformed metric and if the fields $\psi$

[4]At this point Weyl refers to his book Raum, Zeit,Materie, 5th ed., p233ff (quoted as RZM), Berlin 1923. According to a reference there, the argument is due to F. Klein. The sentence in brackets explaining it has been added. [WR]

and $A$ are taken along with unchanged components relative to the transformed tetrad, then the action integral remains unchanged as well, due to its invariant definition, which depends only on the metric. The infinitesimal transformation $\delta e$ of *any* vector field $e$ induced by an infinitesimal point transformation $\delta x = \xi$ is the difference between the vector $e(x)$ at $x$ and the image of $e(x - \delta x)$ under $\delta x = \xi$. It is given by the Lie bracket

$$\delta e^\mu = [\xi, e]^\mu = \frac{\partial \xi^\mu}{\partial x^\nu} e^\nu - \frac{\partial e^\mu}{\partial x^\nu} \xi^\nu,$$

as may be verified as follows.]

Thus (20) will vanish if one substitutes

$$\delta e^\mu(\alpha) = [\xi, e(\alpha)]^\mu = \frac{\partial \xi^\mu}{\partial x^\nu} e^\nu(\alpha) - \frac{\partial e^\mu(\alpha)}{\partial x^\nu} \xi^\nu(\alpha).$$

After an integration by parts one obtains

$$\int \{ \frac{\partial \boldsymbol{T}^\nu_\mu}{\partial x^\nu} + \boldsymbol{T}_\nu(\alpha) \frac{\partial e^\nu(\alpha)}{\partial x^\mu} \} \xi^\mu \, dx = 0$$

Since the $\xi^\mu$ are arbitrary functions of vanishing outside of a finite region of spacetime, this gives the quasi-conservation law of energy-momentum in the form

$$\frac{\partial \boldsymbol{T}^\nu_\mu}{\partial x^\nu} + \frac{\partial e^\nu(\alpha)}{\partial x^\mu} \boldsymbol{T}_\nu(\alpha) = 0. \tag{21}$$

Because of the second term it is a true conservation law only in special relativity. In general relativity it becomes one if the energy-momentum of the gravitational field is added[5].

In special relativity one obtains the components $P_\mu$ of the 4–momentum as the integral

$$P_\mu = \int_{x^0=t} \boldsymbol{T}^0_\mu \mathrm{dx} \quad [\mathrm{dx} = dx^1 dx^2 dx^3]$$

over a 3–space section

$$x^0 = t = \text{const.} \tag{22}$$

The integrals are independent of $t$[6]. Using the symmetry of $\boldsymbol{T}$ one finds further the divergence equations

$$\frac{\partial}{\partial x^\nu}(x^2 \boldsymbol{T}^\nu_3 - x^3 \boldsymbol{T}^\nu_2) = 0, \cdots,$$
$$\frac{\partial}{\partial x^\nu}(x^1 \boldsymbol{T}^\nu_1 - x^1 \boldsymbol{T}^\nu_0) = 0, \cdots.$$

The three equations of the first type show that the *angular momentum* $(J_1, J_2, J_3)$ is constant in time:

$$J_1 = \int_{x^0=t} (x^2 \boldsymbol{T}^0_3 - x^3 \boldsymbol{T}^0_2) \mathrm{dx}, \cdots.$$

The equations of the second type contain the law of *inertia of energy*[7].

We compute the energy–momentum density for the matter action $\boldsymbol{m}$ defined above; we treat separately the two parts of $\boldsymbol{m}$ appearing in (17). For the first part we obtain after an integration by parts

$$\int \delta \boldsymbol{m} . dx = \int u_\mu(\alpha) \delta \boldsymbol{e}^\mu(\alpha) . dx$$

where

$$\mathrm{i} u_\mu(\alpha) = \psi^* \sigma(\alpha) \frac{\partial \psi}{\partial x^\mu} - \frac{1}{2} \frac{\partial (\psi^* \sigma(\alpha) \psi)}{\partial x^\mu}$$
$$= \frac{1}{2} \left( \psi^* \sigma(\alpha) \frac{\partial \psi}{\partial x^\mu} - \frac{\partial \psi^*}{\partial x^\mu} \sigma(\alpha) \psi \right).$$

The part of the energy–momentum arising from the first part of $\boldsymbol{m}$ is therefore

[5] Cf. RZM §41. [WR]
[6] Cf. RZM p.206. [WR]
[7] Cf. RZM p.201.

$$\boldsymbol{T}_\mu(\alpha) = \boldsymbol{u}_\mu(\alpha) - e_\mu(\alpha)\boldsymbol{u}, \quad \boldsymbol{T}^\nu_\mu = \boldsymbol{u}^\nu_\mu - \delta^\nu_\mu \boldsymbol{u},$$

where $\boldsymbol{u}$ denotes the contraction $e_\mu(\alpha)\boldsymbol{u}_\mu(\alpha)$. These formulas also hold in general relativity, for non-constant $e^\mu(\alpha)$. To treat the second part of $\boldsymbol{m}$ we restrict ourselves to special relativity for the sake simplicity. For this second part in (17) one has

$$\int \delta\boldsymbol{m}.dx = \tfrac{1}{4\mathrm{i}} \int \det\left[e_\mu(\alpha), e^\nu(\alpha), \tfrac{\partial(\delta \boldsymbol{e}^\mu(\alpha))}{\partial x^\nu}, j(\alpha)\right] dx$$
$$= \tfrac{1}{4\mathrm{i}} \int \det\left[\delta e_\mu(\alpha), e_\mu(\alpha), e^\nu(\alpha), \tfrac{\partial j(\alpha)}{\partial x^\nu}, j(\alpha)\right] dx$$

The part of the energy–momentum arising from the second part of $\boldsymbol{m}$ is therefore

$$\boldsymbol{T}_\mu(0) = -\tfrac{1}{4\mathrm{i}} \det\left[e_\mu(i), e^\nu(i), \tfrac{\partial j(\alpha)}{\partial x^\nu}, j(i)\right]_{i=1,2,3.}$$

$\boldsymbol{T}^0_\mu$ arises from this by multiplication by $-\mathrm{i}$; hence $\boldsymbol{T}^0_0 = 0$ and

$$(23) \qquad \boldsymbol{T}^0_1 = \tfrac{1}{4}\left(\tfrac{\partial j(3)}{\partial x^2} - \tfrac{\partial j(2)}{\partial x^3}\right), \cdots.$$

Be combine both parts to determine the total energy, momentum, and angular momentum. The equation

$$\boldsymbol{T}^0_0 = -\tfrac{1}{2\mathrm{i}} \textstyle\sum_{i=1}^3 \left(\psi^* \sigma^i \tfrac{\partial \psi}{\partial x^i} - \tfrac{\partial \psi^*}{\partial x^i}\sigma^i \psi)\right)$$

gives after an integration by parts on the subtracted term,

$$P_0 = \int \boldsymbol{T}^0_0 \mathrm{dx} = -\tfrac{1}{\mathrm{i}} \int \psi^* \textstyle\sum_{i=1}^3 \sigma^i \tfrac{\partial \psi}{\partial x^i} \cdot d\mathrm{x}.$$

This leads to the operator

$$-P_0 := \tfrac{1}{\mathrm{i}} \textstyle\sum_{i=1}^3 \sigma^i \tfrac{\partial}{\partial x^i}$$

as quantum-mechanical representative for the energy $-P_0$ of a free particle. Further,

$$P_1 = \int \boldsymbol{T}^0_1 \mathrm{dx} = -\tfrac{1}{2\mathrm{i}} \int \left(\psi^* \tfrac{\partial \psi}{\partial x^1} - \tfrac{\partial \psi}{\partial x^1}\psi^*\right) \cdot \mathrm{dx}$$
$$= \tfrac{1}{\mathrm{i}} \int \left(\psi^* \tfrac{\partial \psi}{\partial x^1}\right) \cdot d\mathrm{x}.$$

The expression (23) does not contribute to the integral. The momentum $(P_1, P_2, P_3)$ will therefore be represented by the operators

$$P_1, P_2, P_3 := \tfrac{1}{\mathrm{i}}\tfrac{\partial}{\partial x^1}, \tfrac{1}{\mathrm{i}}\tfrac{\partial}{\partial x^2}, \tfrac{1}{\mathrm{i}}\tfrac{\partial}{\partial x^3}$$

as it must be, according to Schrödinger. From the complete expression for $x^2\boldsymbol{T}^0_3 - x^3\boldsymbol{T}^0_2$ one obtains by a suitable integration by parts, the angular momentum

$$J_1 = \int \left(\tfrac{1}{\mathrm{i}}\psi^*(x^2 \tfrac{\partial \psi}{\partial x^3} - x^3 \tfrac{\partial \psi}{\partial x^2}) + \tfrac{1}{2} j(1)\right) .\mathrm{dx}, \quad [j(1) = \psi^* \sigma(1)\psi].$$

The angular momentum $(J_1, J_2, J_3)$ will therefore be represented by the operators

$$J_1 := \tfrac{1}{\mathrm{i}}(x^2 \tfrac{\partial \psi}{\partial x^3} - x^3 \tfrac{\partial \psi}{\partial x^2}) + \tfrac{1}{2}\sigma(1), \cdots$$

in agreement with well-known formulas.

Spin –having been put into the theory from the beginning– must naturally reemerge here; but the manner in which this happens is rather surprising and instructive. From this point of view the fundamental assumptions of quantum theory seem less basic than one may have thought, since they are a consequence of the particular action density $\boldsymbol{m}$. On the other hand, these interrelations confirm the irreplaceable role of $\boldsymbol{m}$ for the matter part of the action. Only general relativity, which –because of the free mobility of the $e^\mu(\alpha)$– leads to a definition of the energy–momentum free of any arbitrariness, allows one to complete the circle of quantum theory in the manner described.

**§5. Gravitation**. We return to the transcription of Einstein's classical theory of gravitation and determine first of all the Riemann curvature tensor[8]. Consider a small parallelogram $x(s,t), 0 \le s \le s^*, 0 \le t \le t^*$and let $d_s, d_t$ be the differentials with respect to the parameters[9]. The tetrad $e(\alpha)$ at the vertex $x = x(0,0)$ is parallel-transported to the opposite vertex $x^* = x(s^*,t^*)$ along the edges, once via $x(s^*,0)$ and once via $x(0,t^*)$. The two tetrads at $x^*$ obtained in this way are related an in infinitesimal rotation $\Omega(d_sx, d_tx)$ depending on the line elements $d_sx, d_tx$ at $x$ in the form

$$\Omega(d_sx, d_tx) = \Omega_{\mu\nu}\, d_sx^\mu d_tx^\nu = \tfrac{1}{2}\Omega_{\mu\nu}(d_sx \wedge d_tx)^{\mu\nu}$$

where $\Omega_{\mu\nu}$ is skew-symmetric in $\mu$ and $\nu$ and $(d_sx \wedge d_tx)^{\mu\nu} = d_sx^\mu d_tx^\nu - d_tx^\mu d_sx^\nu$ are the components of the surface element spanned by $d_sx$ and $d_tx$. As infinitesimal rotation, $\Omega_{\mu\nu}$ is itself a skew-symmetric matrix $(\Omega_{\mu\nu}(\alpha\beta))$; this is the *Riemann curvature tensor.*

The tetrad at $x^*$ arising from the tetrad $e(\alpha)$ at $x$ by parallel transport via $x(0,t^*)$ is

$$(1 + d_s + \omega)(1 + d_t + \omega)e(\alpha)$$

in a notation easily understood. The difference of this expression and the one arising from it by interchange of $d_s$ and $d_t$ is

$$= \{d_s(d_t\omega) - d_t(d_s\omega)\} - \{d_s\omega.d_t\omega - d_t\omega.d_s\omega\}$$

One has

$$\omega = \omega_\mu dx^\mu, \quad d_t(d_s\omega) = \tfrac{\partial\omega_\mu}{\partial x^\nu} d_sx^\nu\, d_tx^\mu - \omega_\mu d_t\, d_sx^\mu,$$

and $d_td_sx^\mu = d_sd_tx^\mu$. Thus

$$\Omega_{\mu\nu} = \left(\tfrac{\partial\omega_\nu}{\partial x^\mu} - \tfrac{\partial\omega_\mu}{\partial x^\nu}\right) + (\omega_\mu\omega_\nu - \omega_\nu\omega_\mu).$$

The scalar curvature is

$$\rho = e^\mu(\alpha)e^\nu(\beta)\Omega_{\mu\nu}(\alpha\beta).$$

The first, differentiated term in $\Omega_{\mu\nu}$ gives the contribution

$$\left(e^\nu(\alpha)e^\mu(\beta) - e^\nu(\beta)e^\mu(\alpha)\right) \tfrac{\partial\omega_\mu(\alpha\beta)}{\partial x^\nu}.$$

Its contribution to $\boldsymbol{\rho} = \rho/\epsilon$ consists of the two terms,

$$-2\omega(\beta,\alpha\beta)\tfrac{\partial \mathbf{e}^\mu(\alpha\beta)}{\partial x^\nu} \quad \text{and} \quad \tfrac{\omega_\mu(\alpha,\beta)}{\epsilon}\left(\tfrac{\partial e^\mu(\alpha)}{x^\nu}e^\nu(\beta) - \tfrac{\partial e^\mu(\alpha)}{x^\nu}e^\nu(\alpha)\right)$$

after omission of a complete divergence. According to (12) and (10) these terms are

$$-2\omega(\beta;\rho\beta)\omega(\alpha;\alpha\rho) \quad \text{and} \quad 2\omega(\alpha;\beta\gamma)\omega(\gamma;\alpha\beta).$$

The result is the following expression for the action density $\boldsymbol{g}$ of gravitation[10]

$$\epsilon\boldsymbol{g} = \omega(\alpha;\beta\gamma)\omega(\gamma;\alpha\beta) + \omega(\alpha;\alpha\gamma)\omega(\beta;\beta\gamma). \tag{24}$$

The integral $\int \boldsymbol{g}\, dx$ is not truly invariant, but it is practically invariant: $\boldsymbol{g}$ differs from the true scalar density $\boldsymbol{\rho}$ by a divergence. Variation of the $e^\mu(\alpha)$ in the total action

$$\int(\boldsymbol{g} + \kappa\boldsymbol{h})\, dx$$

[8] Cf. RZM, p119f.

[9] Weyl's infinitesimal argument has been rewritten in terms of parameters.

[10] Cf RZM p.231. [WR]

gives the *gravitational equations.* ($\kappa$ is a numerical constant.) The gravitational energy–momentum tensor $V_\mu^\nu$ is obtained from $\boldsymbol{g}$ if one applies an infinitesimal translation $\delta x^\mu = \xi^\mu$ with constant coefficients $\xi^\mu$ in the coordinate space. The induce variation of the tetrad is

$$\delta e^\mu(\alpha) = -\frac{\partial e^\mu(\alpha)}{\partial x^\nu}\xi^\nu.$$

$\boldsymbol{g}$ is a function of $e^\mu(\alpha)$ and its derivative $e^\mu_\nu(\alpha)= \frac{\partial e^\mu(\alpha)}{\partial x^\nu}$; let $\delta\boldsymbol{g}$ be the total differential of $\boldsymbol{g}$ with respect to these variables:

$$\delta\boldsymbol{g} = \boldsymbol{g}_\mu(\alpha)\delta e^\mu(\alpha) + \boldsymbol{g}^\nu_\mu(\alpha)\delta e^\mu_\nu(\alpha).$$

The variation of the action caused by the infinitesimal translation $\delta x^\mu = \xi^\mu$ in the coordinate space must vanish:

$$(25) \qquad \int \delta\boldsymbol{g}.dx + \int \frac{\partial\boldsymbol{g}}{\partial x^\mu}\xi^\mu.dx = 0.$$

The integral is taken over an arbitrary piece of spacetime. One has

$$\int \delta\boldsymbol{g}.dx = \int(\boldsymbol{g}_\mu(\alpha) + \frac{\partial\boldsymbol{g}^\nu_\mu(\alpha)}{\partial x^\mu})\delta e^\mu(\alpha).dx + \int \frac{\partial(\boldsymbol{g}^\nu_\mu(\alpha)\delta e^\mu_\nu(\alpha))}{\partial x^\nu}.dx^\nu.$$

The expression in parenthesis in the first term equals $-\kappa\boldsymbol{T}_\mu(\alpha)$, in virtue of the gravitational equations, and the integral itself equals

$$-\kappa\int \boldsymbol{T}_\nu(\alpha)\frac{\partial e^\nu(\alpha)}{\partial x^\nu}\xi^\mu.dx.$$

Introduce the quantity

$$\boldsymbol{V}^\nu_\mu = \delta^\nu_\mu\boldsymbol{g} - \frac{\partial e^\rho(\alpha)}{\partial x^\rho}\boldsymbol{g}^\nu_\rho(\alpha).$$

The equation (25) says that

$$\int(\frac{\partial\boldsymbol{V}^\nu_\mu}{\partial x^\nu} - \kappa\boldsymbol{T}_\nu(\alpha)\frac{\partial e^\nu(\alpha)}{\partial x^{\nu\mu}})\xi^\mu.dx = 0$$

the integral being taken over any piece of spacetime. The integrand must therefore vanish. Since the $\xi^\mu$ are arbitrary constants, the factors in front of them must all be zero. Substituting from the resulting expression into (21) produces the pure divergence equation

$$\frac{\partial(\boldsymbol{V}^\nu_\mu+\kappa\boldsymbol{T}^\nu_\mu)}{\partial x^\nu} = 0,$$

the conservation law for the total energy momentum $\boldsymbol{V}^\nu_\mu/\kappa+\boldsymbol{T}^\nu_\mu$ of which $\boldsymbol{V}^\nu_\mu/\kappa$ proves to be the gravitational part. In order to formulate a true differential conservation law for the angular momentum in general relativity one has to specialize the coordinates so that a simultaneous rotation of all local tetrads appears as an orthogonal transformation of the coordinates. This is certainly possible, but I shall not discuss it here.

**§6. The electromagnetic field**. We now come to the critical part of the theory. In my opinion, the origin and necessity of the electromagnetic field is this. The components $\psi_1, \psi_2$ are in reality not uniquely specified by the tetrad, but only in so far as they can still be multiplied by an arbitrary "gauge factor" $e^{\mathrm{i}\lambda}$ of absolute value 1. Only up to such a factor is the transformation specified which $\psi$ suffers due to a rotation of the tetrad. In special relativity this gauge factor has to be considered as a constant, because here we have a single fame, not tied to a point. Not so in general relativity: every point has its own tetrad and hence its own arbitrary gauge factor; after the loosening of the ties of the tetrads to the points the gauge factor becomes necessarily an arbitrary function of the point. The infinitesimal linear transformation $\delta A$ of the $\psi$ corresponding to an infinitesimal rotation $\omega$ is then not completely specified either, but can be increased by an arbitrary purely imaginary multiple $\mathrm{i}\delta A$ of the identity matrix.

In order to specify the covariant differential $\nabla\psi$ completely one needs at each point $(x^\mu)$, in addition to the metric, one such a $\delta A$ for each vector $(dx^\mu)$. In order that $\nabla\psi$ remain a linear function of $(dx^\mu)$, $\delta A = A_\mu dx^\mu$ has to be a linear function of the components $dx^\mu$. If $\psi$ is replaced by $e^{\mathrm{i}\lambda}\psi$, then $\delta A$ must simultaneously be replaced by $\delta A - d\lambda$, as follows from the formula for the covariant differential. A a consequence, one has to add the term

(26) $\quad \frac{1}{\epsilon}A(\alpha)j(\alpha) = \frac{1}{\epsilon}A(\alpha)\psi^*\sigma(\alpha)\psi = A(\alpha)\psi^*\boldsymbol{\sigma}(\alpha)\psi$

to the action density $\boldsymbol{m}$. From now on $\boldsymbol{m}$ *shall denote the action density with this addition,* still given by (14), but now with

$$\nabla\psi(\alpha) = (e^\mu(\alpha)\tfrac{\partial}{\partial x^\mu} + \theta(\alpha) + \mathrm{i}A(\alpha))\psi.$$

It is necessarily gauge invariant, in the sense that this action density remains unchanged under the replacements

$$\psi \mapsto e^{\mathrm{i}\lambda}\psi, \quad A_\mu \mapsto A_\mu - \tfrac{\partial\lambda}{\partial x^\mu}$$

for an arbitrary function $\lambda$ of the point. Exactly in the way described by (26) does the electromagnetic potential act on matter empirically. We are therefore justified in identifying the quantities $A_\mu$ defined here with the components of the electromagnetic potential. The proof is complete if we show that the $A_\mu$ field is conversely acted on by matter in the way empirically known for the electromagnetic field.

$$F_{\mu\nu} = \tfrac{\partial A_\nu}{\partial x^\mu} - \tfrac{\partial A_\mu}{\partial x^\nu}$$

is a gauge invariant skew-symmetric tensor and

(27) $\quad \boldsymbol{l} = \frac{1}{4}F_{\mu\nu}\boldsymbol{F}^{\mu\nu}$

is the scalar action density characteristic of Maxwell's theory. The Ansatz

(28) $\quad \boldsymbol{h} = \boldsymbol{m} + a\boldsymbol{l}$

($a$ a numerical constant) gives Maxwell's equation by variation of the $A_\mu$, with

(28) $\quad -\boldsymbol{j}^\mu = -\psi^*\boldsymbol{\sigma}^\mu\psi$

as density of the electric four current.

Gauge invariance is closely related to the conservation law for the electric charge. Because of the gauge invariance of $\boldsymbol{h}$,the variation $\delta\int \boldsymbol{h}\,dx$ must vanish identical when the $\psi$ and $A_\mu$ are varied according to

$$\delta\psi = \mathrm{i}\lambda.\psi, \quad \delta A_\mu = -\tfrac{\partial\lambda}{\partial x^\mu}$$

while the $e^\mu(\alpha)$ are kept fixed; $\lambda$ is an arbitrary function of the point. This gives an identical relation between matter equations and the electromagnetic equations. If the matter equations (in the restricted sense) hold, then it follows that

$$\delta\int \boldsymbol{h}\,dx = 0,$$

provided only the $A_\mu$ are varied, in accordance with the equations $\delta A_\mu = -\partial\lambda/\partial x^\mu$. On the other hand, the electromagnetic equations imply the same, if only the $\psi$ is varied according to $\delta\psi = \mathrm{i}\lambda.\psi$. For $\boldsymbol{h} = \boldsymbol{m} + a\boldsymbol{l}$ one obtains in both cases

$$\int \delta\boldsymbol{h}.dx = \pm\int \psi^*\boldsymbol{\sigma}^\mu\psi.\tfrac{\partial\lambda}{\partial x^\mu}dx = \mp\int \lambda\tfrac{\partial \boldsymbol{j}^\mu}{\partial x^\mu}dx.$$

We had an analogous situation for the conservation laws for energy–momentum and for angular momentum. These relate the matter equations in the extended sense with the gravitational equations and the corresponding invariance under coordinate transformations resp. Invariance under arbitrary independent rotations of the local tetrads at different spacetime points.

It follows from

(30) $\frac{\partial \boldsymbol{j}^\mu}{\partial x^\mu} = 0$

that the flux

(31) $(\psi, \psi) = \int \boldsymbol{j}^0 d\mathrm{x} = \int e^0(\alpha)\psi^* \sigma(\alpha)\psi \, d\mathrm{x}$

of the vector density $\boldsymbol{j}^\mu$ through a three dimensional section of spacetime, in particular through the section $x^0 = t =$const. of (22), is independent of $t$. Not only does this integral have an invariant significance, but also the individual element of integration; however, the sign depends on which the direction of traversal of the three dimensional section is taken as positive. The Hermitian form

(32) $j^0 = e^0(\alpha)\psi^* \sigma(\alpha)\psi$

must be positive definite for $\boldsymbol{j}^0 d\mathrm{x}$ to be taken as probability density in three space. It is easy to see that this is the case if $x^0 =$const. is indeed a spacelike section though $x$, i.e. when its tangent vectors at $x$ are spacelike. The sections $x^0 =$const must ordered so that $x^0$ increases in the direction of the timelike vector $e(0)/\mathrm{i}$ for (30) to be positive. The sign of the flux is specified by these natural restrictions on the coordinate systems; the invariant quantity (31) shall be normalized in the usual way be the condition

(33) $(\psi, \psi) = \int \boldsymbol{j}^0 d\mathrm{x} = 1.$

The coupling constant $a$ which combines $\boldsymbol{m}$ and $\boldsymbol{l}$ is then a pure number $= c\hbar/e^2$, the reciprocal fine structure constant.

We treat $\psi_1, \psi_2; A_\mu; e^\mu(\alpha)$ as the independent fields for the variation of the action. To the energy–momentum density $\boldsymbol{T}^\nu_\mu$ arising from $\boldsymbol{m}$ one must add

$$A_\mu \boldsymbol{j}^\nu - \delta^\nu_\mu (A_\rho \boldsymbol{j}^\rho)$$

because of the additional term (26). In special relativity, this lead one to represent the energy by the operator

$$H = \sum_{i=1}^3 \sigma^i \left( \frac{1}{\mathrm{i}} \frac{\partial}{\partial x^\mu} + A_\mu \right),$$

because the energy value is

$$-P_0 = \int \psi^* H \psi \, d\mathrm{x} = (\psi, H\psi).$$

The matter equation then read

$$\left( \frac{1}{\mathrm{i}} \frac{\partial}{\partial x^\mu} + A_\mu \right) \psi + H\psi = 0$$

and not $\frac{1}{\mathrm{i}} \frac{\partial \psi}{\partial x^\mu} + H\psi = 0$, as had assumed in quantum mechanics up to now. Of course, to this matter energy one has to add the electromagnetic energy for which the classical expressions of Maxwell remain valid.

As far as *physical dimensions* are concerned, in general relativity it is natural to take the coordinates $x^\mu$ as pure numbers. The quantities under consideration are then invariant under change of scale, but under arbitrary transformations of the $x^\mu$. If all $e(\alpha)$ are transformed into $be(\alpha)$ by multiplication with an arbitrary factor $b$, then $\psi$ has to be replaced by $e^{3/2}\psi$ if the normalization (33) is to remain in tact. $\boldsymbol{m}$ and $\boldsymbol{l}$ remain thereby unchanged, hence are pure numbers. But $\boldsymbol{g}$ takes on the factor $1/b^2$, so that $\kappa$ becomes the square of a length $d$. $\kappa$ is not identical with Einstein's constant of gravitation, but arises from it by multiplication with $2\hbar/e$. $d$ lies far below the atomic scale, being $\sim 10^{-32}$ cm. So gravity will be relevant for astronomical problems only.

If we ignore the gravitational term, then the atomic constants in the field equations are dimensionless. In the two–component theory there is no place for a term involving mass as a factor as there is in Dirac's theory[11]. But one know how one can introduce mass using the conservation laws[12]. One assumes that the $\boldsymbol{T}^{\nu}_{\mu}$ vanish in "empty spacetime" surrounding the particle, i.e. outside of a channel in spacetime, whose sections $x^0$ =const. are of finite extent. There the $e^{\mu}(\alpha)$ may be taken to be constant as in special relativity. Then

$$P_\mu = \int (\boldsymbol{T}^0_\mu + \tfrac{1}{\kappa}\boldsymbol{V}^0_\mu)\, d\mathrm{x}$$

are the components of a four-vector in empty spacetime, which is independent of $t$ and independent of the arbitrary choice of the local tetrads (within the channel). The coordinate system in empty spacetime can be normalized further by the requirement that the 3–
momentum $(P_1, P_2, P_3)$ should vanish; $-P_0$ is then the invariant and time-independent mass of the particle. One then requires that the value $m$ of this mass be given once and for all.

The theory of the electromagnetic field discussed here I consider to be the correct one, because it arises so naturally from the arbitrary nature of the gauge factor in $\psi$ and hence explains the empirically observed gauge invariance on the basis of the conservation law for electric charge. But another theory relating electromagnetism and gravitation is presents itself. The term (26) has the same form as the section part of $\boldsymbol{m}$ in formula (17); $\varphi(\alpha)$ plays the same role for the for the latter as $A(\alpha)$ does for the former. One may therefore expect that matter and gravitation, i.e. $\psi$ and $e^{\mu}(\alpha)$, will by themselves be sufficient to explain the electromagnetic phenomena when one takes the $\varphi(\alpha)$ as electromagnetic potentials. These quantities $\varphi(\alpha)$ depend on the $e^{\mu}(\alpha)$ and on their first derivatives in such a manner that there is invariance under arbitrary transformations of the coordinates. Under rotations of the tetrads, however, the $\varphi(\alpha)$ transform as the components of a vector only if all tetrads undergo the same rotation. If one ignores the matter field and considers only the relation between electromagnetism and gravitation, then one arrives in this way at a theory of electromagnetism of exactly the kind Einstein recently tried to establish. Parallelism at a distance would only be simulated, however.

I convinced myself that this Ansatz, tempting as it may be at first sight, does not lead to Maxwell's equations. Besides, gauge invariance would remain completely mysterious; the electromagnetic potential itself would have physical significance, and not only the field strength. I believe, therefore, that this idea leads astray and that we should take the hint given by gauge invariance: *electromagnetism is a byproduct of the matter field, not of gravitation.*

***

[11]Proc. Roy. Soc. (A) 117, 610.
[12]Cf. RZM, p.278ff. [WR]

**Translator's Notes**

Note 1.[p.203] The discussion around equation (6) may be interpreted as follows. An "infinitesimal rotation" is a vector field generating a one-parameter group of orthogonal linear transformations of the $j$s. The vector attached to $j$ by such a vector field is given by multiplication with a skew-symmetric matrix. This situation is summarized by the equation $\delta j = \delta\Lambda.j$. In the symbol $\delta j$ the letter $\delta$ may be taken to stand for the vector field itself. The symbol $\delta\Lambda$ is best not taken apart into $\delta$ and $\Lambda$; $\delta\Lambda$ denotes a skew symmetric matrix, a tangent vector at the identity on the manifold of the rotation group itslef. The infinitesimal transformation $\delta\psi = \delta L.\psi$ of the $\psi$ "corresponding" to $\delta j = \delta\Lambda.j$ is defined by the condition that $\delta\psi = \delta L.\psi$ if $j = j(\psi)$ and $\delta j = \delta\Lambda.j$, i.e. $\delta\Lambda.j$ is the image of $\delta L.\psi$ under the differential of the map $\psi \mapsto j = j(\psi)$. In particular, $\delta\Lambda.j$ depends linearly on $\delta L.\psi$, and this implies that $\delta\Lambda$ depends linearly on $\delta L$.

**Note 2**.[p.203] The equation $\delta\int \boldsymbol{h}\,dx = 0$ may be interpreted in the way standard in the calculus of variation, as follows. Relative to a coordinate system $(x^\mu)$, the integrand $\boldsymbol{h}$ is a function of the values of a number of variable "fields" $\varphi$ and their derivatives. In the present case, the tetrad $e^\mu(\alpha)$ must be included among the variable fields, in addition to $\psi$ and $A_\mu$. If the fields $\varphi$ are made depend on an auxiliary small parameter $\epsilon$ ("given an infinitesimal variation $\delta\varphi$"), then the integral $\int \boldsymbol{h}\,dx$ will depend on $\epsilon$ as well and $\delta\int \boldsymbol{h}\,dx$ denotes its derivative at $\epsilon = 0$. Naturally, the integrand $\boldsymbol{h}$ should depend on the fields in such a manner that the integral is independent of the coordinate system, or at least the equation $\delta\int \boldsymbol{h}\,dx = 0$ is so.

**Translator's comments**.

No doubt, the protagonist of Weyl's paper is the vector–valued Hermitian form[13]

$$j(\alpha) = \psi^*\sigma(\alpha)\psi.$$

with which it opens. Weyl's whole theory is based on the fact that the most general connection on $\psi$ whose parallel transport is compatible with the metric parallel transport of $j(\psi)$ differs by a differential 1-form i$A$ from the connection on $\psi$ induced by the metric itself. Attached to the action through the cubic $A\psi$ interaction $A(\alpha)\psi^*\sigma(\alpha)\psi$, the vector field $j(\psi)$ then reemerges as the Nöther current corresponding to gauge transformations of the spinor field $\psi$.

The current $j(\psi)$ is a very special object indeed, in peculiar to four dimensions in this form. (Related objects do exist in a few other dimensions, e.g. in eight, depending somewhat on what one is willing to accept as "analog".) No such canonical current present in the setting of general gauge theory. The cutting of this link between $\psi$ and $j$ by Yang and Mills made room for other fields, but at the same time destroyed the beautifully rigid construction of Weyl's theory. This may be one reason why Weyl never wrote the promised second part of the paper. Anderson's discovery of the positron in 1932 must have dashed

[13]This form had shown up long before Dirac and Weyl, in the context of projective geometry, in Klein's work on automorphic functions, cf. Fricke and Klein book of 1897, vol. 1, Einleitung §12, and probably even earlier e.g. in the work of Möbius, Hermite, and Cayley.

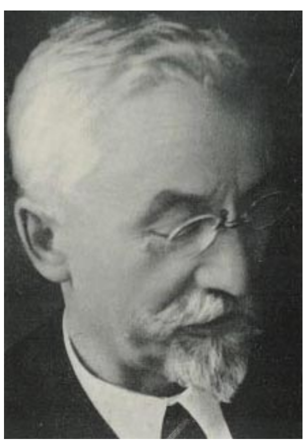
Weyl's hope for a theory of everything. (One is strangely reminded of Kepler's planetary model based on the five perfect solids; happily Kepler was spared the discovery of Uranus, Neptune, and Pluto.)

WR

# Time chart

Bernhard Riemann (1826–1866)

C. F. Gauss (1777–1855)



Henri Cartan (1869–1951)

Hermann Weyl (1885–1955)

Pictures selected from MacTutor History of Mathematics.

# Annotated bibliography

*Modern Geometry–Methods and Applications,* vol.1-2, by B.A. Dubrovin, A.T. Fomenko, S.P. Novikov. Springer Verlag, 1984. [A textbook which could supplement –or replace– the presentation here. Volume 1 is at the undergraduate level and has a similar point of view, but does not introduce manifolds explicitly. Volume 2 is much more advanced.]

*A comprehensive Introduction to Differential Geometry*, vol. 1-5, by M. Spivak. Publish or Perish, Inc., 1979. [Comprehensive indeed: a resource for its wealth of information. Volume I goes at manifolds from the abstract side, starting from topological spaces. At the graduate level.]

*Geometry and the Imagination* by D. Hilbert and S. Cohn-Vossen. German original published in 1932. Republished by the AMS in 1999. [A visual tour for the mind's eye, revealing an unexpected side of Hilbert himself. The original title "Anschauliche Geometrie" means "Visual Geometry".]

Riemann's *On the hypotheses which lie at the bases of geometry*, his Habilitationschrift of 1854, translated by William Kingdon Clifford. Available on the web thanks to D.R.Wilkins. [Riemann's only attempt to explain his revolutionary ideas on geometry, and that in non-technical terms. Explanations of his explanations are available; one may also just listen to him.]

Weyl's *Raum–Zeit–Materie* of 1918, republished by Springer Verlag in 1984. [Weyl's youthful effort to communicate his insights into geometry and physics. Not meant for the practical man. Its translation "Space–Time-Matter" (Dover,1984) casts Weyl's poetry in leaden prose.]

Élie Cartan's *Leçons sur la géométrie des espaces de Riemann* dating from 1925–1926, published in 1951 by Gauthier-Villars. [An exposition of the subject by the greatest differential geometer of the 20th century, based on his own methods and on some of his own creations.]

# Index