Winter 1-26-2016

An Introduction to Differential Geometry through Computation

Mark E. Fels
Utah State University

An Introduction to Differential Geometry through Computation

Mark E. Fels

Contents

Preface

This book was conceived after numerous discussions with my colleague Ian Anderson about what to teach in an introductory one semester course in differential geometry. We found that after covering the classical differential geometry of curves and surfaces that it was difficult to make the transition to more advanced texts in differential geometry such as [4], or to texts which use differential geometry such as in differential equations [9] or general relativity [11], [13]. This book aims to make this transition more rapid, and to prepare upper level undergraduates and beginning level graduate students to be able to do some basic computational research on such topics as the isometries of metrics in general relativity or the symmetries of differential equations.

This is not a book on classical differential geometry or tensor analysis, but rather a modern treatment of vector fields, push-forward by mappings, one-forms, metric tensor fields, isometries, and the infinitesimal generators of group actions, and some Lie group theory using only open sets in $\mathbb{R}^n$. The definitions, notation and approach are taken from the corresponding concept on manifolds and developed in $\mathbb{R}^n$. For example, tangent vectors are defined as derivations (on functions in $\mathbb{R}^n$) and metric tensors are a field of positive definite symmetric bilinear functions on the tangent vectors. This approach introduces the student to these concepts in a familiar setting so that in the more abstract setting of manifolds the role of the manifold can be emphasized.

The book emphasizes liner algebra. The approach that I have taken is to provide a detailed review of a linear algebra concept and then translate the concept over to the field theoretic version in differential geometry. The level of preparation in linear algebra effects how many chapters can be covered in one semester. For example, there is quite a bit of detail on linear transformations and dual spaces which can be quickly reviewed for students with advanced training in linear algebra.

The outline of the book is as follows. Chapter 1 reviews some basic facts about smooth functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, as well as the basic facts about vector spaces, basis, and algebras. Chapter 2 introduces tangent vectors and vector fields in $\mathbb{R}^n$ using the standard two approaches with curves and derivations. Chapter 3 reviews linear transformations and their matrix representation so that in Chapter 4 the push-forward as an abstract linear transformation can be defined and its matrix representation as the Jacobian can be derived. As an application, the change of variable formula for vector fields is derived in Chapter 4. Chapter 5 develops the linear algebra of the dual space and the space of bi-linear functions and demonstrates how these concepts are used in defining differential one-forms and metric tensor fields. Chapter 6 introduces the pullback map on one-forms and metric tensors from which the important concept of isometries is then defined. Chapter 7 investigates hyper-surfaces in $\mathbb{R}^n$, using patches and defines the induced metric tensor from Euclidean space. The change of coordinate formula on overlaps is then derived. Chapter 8 returns to $\mathbb{R}^n$ to define a flow and investigates the relationship between a flow and its infinitesimal generator. The theory of flow invariants is then investigated both infinitesimally and from the flow point of view with the goal of proving the rectification theorem for vector fields. Chapter 9 investigates the Lie bracket of vector-fields and Killing vectors for a metric. Chapter 10 generalizes chapter 8 and introduces the general notion of a group action with the goal of providing examples of metric tensors with a large number of Killing vectors. It also introduces a special family of Lie groups which I've called multi-parameter groups. These are Lie groups whose domain is an open set in $\mathbb{R}^n$. The infinitesimal generators for these groups are used to construct the left and right invariant vector-fields on the group, as well as the Killing vectors for some special invariant metric tensors on the groups.

Chapter 1

Preliminaries

1.1 Open sets

The components (or Cartesian coordinates) of a point $\mathbf{x} \in \mathbb{R}^n$ will be denoted by

$$\mathbf{x} = (x^1, x^2, \ldots, x^n).$$

Note that the labels are in the up position. That is x^2 is not the square of x unless we are working in $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$ where we will use the standard notation of x, y, z. The position of indices is important, and make many formulas easier to remember or derive. The Euclidean distance between the points $\mathbf{x} = (x^1, \ldots, x^n)$ and $\mathbf{y} = (y^1, \ldots, y^n)$ is

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x^1 - y^1)^2 + \ldots + (x^n - y^n)^2}.$$

The *open ball* of radius $r \in \mathbb{R}^+$ at the point $p \in \mathbb{R}^n$ is the set $B_r(p) \subset \mathbb{R}^n$, defined by

$$B_r(p) = \{ \ \mathbf{x} \in \mathbb{R}^n \mid d(\mathbf{x}, p) < r\}.$$

A subset $U \subset \mathbb{R}^n$ is an *open set* if given any point $p \in U$ there exists an $r \in \mathbb{R}^+$ (which depends on p) such that the open ball $B_r(p)$ satisfies $B_r(p) \subset U$. The empty set is also taken to be open.

Example 1.1.1. The set $\mathbb{R}^n$ is an open set.

Example 1.1.2. Let $p \in \mathbb{R}^n$ and $r \in \mathbb{R}^+$. Any open ball $B_r(p)$ is an open set.

Example 1.1.3. The upper half plane is the set

$$U = \{ \ (x, y) \in \mathbb{R}^2 \mid y > 0 \ \}$$

and is open.

Example 1.1.4. The set

$$V = \{ \ (x, y) \in \mathbb{R}^2 \mid y \geq 0 \ \}$$

is not open. Any point $(x, 0) \in V$ can not satisfy the open ball condition.

Example 1.1.5. The unit n-sphere $S^n \subset \mathbb{R}^{n+1}$ is the subset

$$S^n = \{ \ \mathbf{x} \in \mathbb{R}^{n+1} \mid d(\mathbf{x}, 0) = 1 \ \}$$

and S^n is not open. No point $\mathbf{x} \in S^n$ satisfies the open ball condition. The set S^n is the boundary of the open ball $B_1(0) \subset \mathbb{R}^{n+1}$.

Roughly speaking, open sets contain no boundary point. This can be made precise using some elementary topology.

1.2 Smooth functions

In this section we recall some facts from multi-variable calculus. A real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ has the form,

$$f(\mathbf{x}) = f(x^1, x^2, \ldots, x^n).$$

We will only be interested in functions whose domain $\mathrm{Dom}(f)$, is either all of $\mathbb{R}^n$ or an open subset $U \subset \mathbb{R}^n$. For example $f(x, y) = \log xy$ is defined only on the set $U = \{(x, y) \in \mathbb{R}^2 \mid xy > 0\}$, which is an open set in $\mathbb{R}^2$, and $\mathrm{Dom}(f) = U$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at $p \in \mathbb{R}^n$ if

$$\lim_{\mathbf{x} \to p} f(\mathbf{x}) = f(p).$$

If $U \subset \mathbb{R}^n$ is an open set, then $C^0(U)$ denotes the functions defined on U which are continuous at every point of U.

Example 1.2.1. Let $U = \{\ (x, y) \mid (x, y) \neq (0, 0)\ \}$, the function

$$f(x, y) = \frac{1}{x^2 + y^2}$$

is continuous on the open set $U \subset \mathbb{R}^2$

Note that if $f \in C^0(\mathbb{R}^n)$ then $f \in C^0(U)$ for any open subset $U \subset \mathbb{R}^n$.

The partial derivatives of f at the point $p = (x_0^1, \ldots, x_0^n)$ in the x^i direction is

$$\left.\frac{\partial f}{\partial x^i}\right|_p = \lim_{h \to 0} \frac{f(x_0^1, x_0^2, \ldots, x_0^i + h, x_0^{i+1}, \ldots, x_0^n) - f(x_0^1, x_0^2, \ldots, x_0^i, x_0^{i+1}, \ldots, x_0^n)}{h}$$

which is also written $(\partial_{x^i} f)|_p$. Let $U \subset \mathbb{R}^n$ be an open set. A function $f : U \to \mathbb{R}$ is said to be $C^1(U)$ if all the partial derivatives $\partial_{x^i} f, 1 \leq i \leq n$ exists at every point in U and these n-functions are continuous at every point in U.

The partial derivatives of order k are denoted by

$$\frac{\partial^k f}{\partial x^{i_1} \partial x^{i_2} \ldots \partial x^{i_k}}$$

where $1 \leq i_1, i_2, \ldots, i_k \leq n$. We say for a function $f : U \to \mathbb{R}$, U an open set in $\mathbb{R}^n$, that $f \in C^k(U)$ if all the partial derivatives up to order k exist at every point in the open set U and they are also continuous at every point in U. A function $f : U \to \mathbb{R}$ is said to be *smooth* or $f \in C^\infty(U)$ if $f \in C^k(U)$ for all $k \geq 0$. In other words a function is smooth if its partial derivatives exist to all orders at every point in U, and the resulting functions are continuous.

Example 1.2.2. Let $i \in \{1, \ldots, n\}$. The *coordinate functions* $f^i : \mathbb{R}^n \to \mathbb{R}$, where $f^i(\mathbf{x}) = x^i$ (so the i^{th} coordinate) satisfy $f^i \in C^\infty(\mathbb{R}^n)$, and are smooth functions. The coordinate functions f^i will just be written as x^i. Any polynomial in the coordinate functions

$$P(\mathbf{x}) = a_0 + \sum_{1 \leq i \leq n} a_i x^i + \sum_{1 \leq i_1, i_2 \leq n} a_{i_1 i_2} x^{i_1} x^{i_2} + \ldots \text{up to finite order}$$

satisfies $P(\mathbf{x}) \in C^\infty(\mathbb{R}^n)$, and are smooth functions.

Example 1.2.3. Let $U \subset I\!R^n$ be an open set and define the functions $1_U, 0_U : U \to I\!R$ by

$$
\begin{aligned}
1_U &= \{\ 1,\ \textit{for all} \quad \mathbf{x} \in U\ \},\\
0_U &= \{\ 0,\ \textit{for all} \quad \mathbf{x} \in U\ \}.
\end{aligned}
\tag{1.1}
$$

The function $1_U, 0_U \in C^\infty(U)$. The function 1_U is the unit function on U, and 0_U is the 0 function on U. All the partial derivatives are 0 for these functions.

One reason we work almost exclusively with smooth functions is that if $f \in C^\infty(U)$ then $\partial_{x^i} f \in C^\infty(U), 1 \leq i \leq n$, and so all the partial derivatives are again smooth functions. While working with this restricted class of functions is not always necessary, by doing so the exposition is often simpler.

The set of functions $C^k(U)$ have the following algebraic properties [12].

Proposition 1.2.4. *Let $f, g \in C^k(U)$ ($k \geq 0$ including $k = \infty$), and let $\alpha \in I\!R$. Then*

1. $(\alpha f)(\mathbf{x}) = \alpha f(\mathbf{x}) \in C^k(U)$,
2. $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) \in C^k(U)$,
3. $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x}) \quad \in C^k(U)$,
4. $(\frac{f}{g})(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x})} \in C^k(V)$, *where* $V = \{\ \mathbf{x} \in U \mid g(\mathbf{x}) \neq 0\ \}$.

Example 1.2.5. Let $P(\mathbf{x})$ and $Q(\mathbf{x})$ be polynomials on $I\!R^n$. Then by 4 in Lemma 1.2.4

$$
f(\mathbf{x}) = \frac{P(\mathbf{x})}{Q(\mathbf{x})}
$$

is a smooth function on the open set $V = \{\ \mathbf{x} \in I\!R^n \mid Q(\mathbf{x}) \neq 0\ \}$.

A function $\Phi : I\!R^n \to I\!R^m$ is written in components as

$$
\Phi(\mathbf{x}) = (\Phi^1(\mathbf{x}), \Phi^2(\mathbf{x}), \ldots, \Phi^m(\mathbf{x})), \quad \mathbf{x} \in I\!R^n.
$$

The function Φ is smooth or $\Phi \in C^\infty(I\!R^n, I\!R^m)$ if each component $\Phi^1, \Phi^2, \ldots \Phi^n \in C^\infty(I\!R^n)$. If $U \subset I\!R^n$ is open, then $C^\infty(U, I\!R^m)$ denotes the C^∞ functions $f : U \to I\!R^n$.

Example 1.2.6. The function $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\Phi(x,y) = (x+y, x-y, x^2+y^2)$$

has components

$$\Phi^1(x,y) = x+y,\ \Phi^2(x,y) = x-y,\ \Phi^3(x,y) = x^2+y^2.$$

Therefore $\Phi \in C^\infty(\mathbb{R}^2, \mathbb{R}^3)$.

1.3 Smooth Curves

parametrized

Let $a, b \in \mathbb{R}, a < b$ then $I = (a,b)$ is the open interval

$$I = \{x \in \mathbb{R} \mid a < x < b\}.$$

A function $\sigma \in C^\infty(I, \mathbb{R}^n)$ is a mapping $\sigma : I \to \mathbb{R}^n$, and is called a smooth (parameterized) or C^∞ curve. If t denotes the coordinate on I the curve σ has components

$$\sigma(t) = (\sigma^1(t), \sigma^2(t), \ldots, \sigma^n(t)).$$

The derivative $\dot\sigma(t)$ of the curve $\sigma(t)$ is

$$\dot\sigma(t) = \frac{d\sigma}{dt} = \left(\frac{d\sigma^1}{dt}, \frac{d\sigma^1}{dt}, \ldots, \frac{d\sigma^n}{dt}\right).$$

If $t_0 \in I$ then $\dot\sigma(t_0)$ is the *tangent vector* to σ at the point $\sigma(t_0)$. The *Euclidean arc-length* of a curve σ (when it exists) is

$$L(\sigma) = \int_a^b \sqrt{\sum_{i=1}^n \left(\frac{d\sigma^i}{dt}\right)^2} dt = \int_a^b ||\dot\sigma|| dt$$

where $||\dot\sigma|| = \sqrt{\sum_{i=1}^n (\dot\sigma^i)^2}$.

Example 1.3.1. Let $\sigma : \mathbb{R} \to \mathbb{R}^3$ be the smooth curve

$$\sigma(t) = (\cos t, \sin t, t)\ , \quad t \in \mathbb{R}$$

which is known as the helix. The tangent vector at an arbitrary t value is

$$\dot{\sigma}(t) = \frac{d\sigma}{dt} = (-\sin t, \cos t, 1).$$

When $t = \frac{\pi}{4}$ we have the tangent vector $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1)$, which looks like,

diagram

The arc-length of σ doesn't exists on $I\!R$. If we restrict the domain of σ to $I = (0, 2\pi)$ we get

$$L(\sigma) = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t + 1}dt = 2\sqrt{2}\pi$$

1.4 Composition and the Chain-rule

An easy way to construct smooth functions is through function composition. Let $m, n, k \in Z^+$ and let $\Phi : I\!R^n \to I\!R^m$, $\Psi : I\!R^m \to I\!R^l$. The composition of the two functions Ψ and Φ is the function $\Psi \circ \Phi : I\!R^n \to I\!R^l$ defined by

$$(\Psi \circ \Phi)(\mathbf{x}) = \Psi(\Phi(\mathbf{x})) \quad for\ all\ \mathbf{x} \in I\!R^n.$$

Note that unless $l = n$ the composition of $\Phi \circ \Psi$ cannot be defined.

Let $(x^i)_{1 \leq i \leq n}$ be coordinates on $I\!R^n$, $(y^a)_{1 \leq a \leq m}$ be coordinates on $I\!R^m$ and $(u^\alpha)_{1 \leq \alpha \leq l}$ be coordinates on $I\!R^l$. In terms of these coordinates the components of the functions Φ and Ψ can be written

$$y^a = \Phi^a(x^1, \ldots, x^n) \quad 1 \leq a \leq m,$$
$$u^\alpha = \Psi^\alpha(y^1, \ldots, y^m) \quad 1 \leq \alpha \leq l.$$

The components of the composition $\Psi \circ \Phi$ are then

$$u^\alpha = \Psi^\alpha(\Phi(x^1, \ldots, x^n)) \quad 1 \leq \alpha \leq l.$$

Example 1.4.1. Let $\sigma : I\!R \to I\!R^3$ be the helix from example 1.3.1, and let $\Phi : I\!R^3 \to I\!R^2$ be

$$\Phi(x, y, z) = (xy + 2yz, x + y). \tag{1.2}$$

The composition $\Phi \circ \sigma : I\!R \to I\!R^2$ is the curve

$$\Phi \circ \sigma(t) = (\sin t \cos t + 2t \sin t, \cos t + \sin t).$$

Now let $\Psi : I\!R^2 \to I\!R^2$ be given by

$$\Psi(u, v) = (u - v, uv). \tag{1.3}$$

The composition $\Psi \circ \Phi : I\!R^3 \to I\!R^2$ is then

$$\Psi \circ \Phi(x, y, z) = (xy + 2yz - x - y, x^2 y + 2xyz + xy^2 + 2y^2 z). \tag{1.4}$$

The formula for first partial derivatives of a composition of two functions is known as the chain-rule.

Theorem 1.4.2. *(The chain-rule). Let $\Phi \in C^1(I\!R^n, I\!R^m)$, and $\Psi \in C^1(I\!R^m, I\!R^l)$. Then $\Psi \circ \Phi \in C^1(I\!R^n, I\!R^l)$, and*

$$\frac{\partial (\Psi \circ \Phi)^\alpha}{\partial x^i} = \sum_{a=1}^{m} \left. \frac{\partial \Psi^\alpha}{\partial y^a} \right|_{y^a = \Phi^a(\mathbf{x})} \frac{\partial \Phi^a}{\partial x^i}, \quad 1 \le i \le n, \ 1 \le \alpha \le l. \tag{1.5}$$

Example 1.4.3. We verify the chain-rule for the functions Ψ and Φ in example 1.4.1. For the left side of equation 1.5, we have using equation 1.4,

$$\partial_x (\Psi \circ \Phi) = (y - 1, 2xy + 2yz + y^2). \tag{1.6}$$

While for the right side we need

$$\begin{aligned} \left. \frac{\partial \Psi}{\partial u} \right|_{(u,v)=\Phi(x,y,z)} &= (\partial_u \Phi^1, \partial_u \Phi^2)|_{(u,v)=\Phi(x,y,z)} &&= (1, v)|_{(u,v)=\Phi(x,y,z)} = (1, x + y) \\ \left. \frac{\partial \Psi}{\partial v} \right|_{(u,v)=\Phi(x,y,z)} &= (\partial_u \Phi^1, \partial_u \Phi^2)|_{(u,v)=\Phi(x,y,z)} &&= (-1, u)|_{(u,v)=\Phi(x,y,z)} = (-1, xy + 2yz) \end{aligned} \tag{1.7}$$

and

$$\partial_x \Phi = (y, 1). \tag{1.8}$$

Therefore the two terms on the right side of 1.5 for $\alpha = 1, 2$ can be computed from equations 1.7 and 1.8 to be

$$\begin{aligned} \frac{\partial \Psi^1}{\partial u} \frac{\partial \Phi^1}{\partial x} + \frac{\partial \Psi^1}{\partial v} \frac{\partial \Phi^2}{\partial x} &= y - 1 \\ \frac{\partial \Psi^2}{\partial u} \frac{\partial \Phi^1}{\partial x} + \frac{\partial \Psi^2}{\partial v} \frac{\partial \Phi^2}{\partial x} &= (x + y)y + (xy + 2yz) \end{aligned}$$

which agrees with 1.6.

Theorem 1.4.2 generalizes to the composition of C^k functions.

Theorem 1.4.4. *Let $k \geq 0$ (including $k = \infty$), and let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ be open sets. If $\Phi \in C^k(U, V)$ and $\Psi \in C^k(V, \mathbb{R}^l)$, then $\Psi \circ \Phi \in C^k(U, \mathbb{R}^l)$.*

Therefore the composition of two smooth functions is again a smooth function.

Example 1.4.5. Let $g(x, y, z) = x^2 + y^2 + z^2$. Clearly $g \in C^\infty(\mathbb{R}^3)$ because any polynomial is C^∞. Let $h(u) = e^u$, and $h \in C^\infty(\mathbb{R})$. Therefore by Theorem 1.4.4 above $e^{x^2+y^2+z^2} \in C^\infty(\mathbb{R}^3)$. Likewise all the compositions in example 1.4.1 are C^∞.

Example 1.4.6. The function

$$f(x, y, z) = \log(x + y + z)$$

is smooth on $U = \{(x, y, z) \mid x + y + z > 0\}$.

Example 1.4.7. Let $\sigma : I \to \mathbb{R}^n$ be a smooth curve in $\mathbb{R}^n$ defined on an open interval $I \subset \mathbb{R}$. Let $\Phi \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ then $\Psi \circ \sigma \in C^\infty(I, \mathbb{R}^m)$ and is a smooth curve in $\mathbb{R}^m$. This composition produces a smooth curve in the range space of Φ. The chain-rule produces

$$\frac{d}{dt}\Phi^a(\sigma(t)) = \sum_{i=1}^{n} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_{\sigma(t)} \frac{d\sigma^i}{dt} \tag{1.9}$$

The next theorem is technical but will be needed in Chapter 3.

Theorem 1.4.8. *Let $f \in C^\infty(U)$ where $U \subset \mathbb{R}^n$ is an open set, and let $p = (x_0^1, \ldots, x_0^n) \in U$. There exists an open ball $B_r(p) \subset U$ and n functions $g_i \in C^\infty(B_r(p)), 1 \leq i \leq n$ such that*

$$f(\mathbf{x}) = f(p) + \sum_{i=1}^{n} (x^i - x_0^i) g_i(\mathbf{x}) \quad \textit{for all } \mathbf{x} \in B_r(p) \tag{1.10}$$

and where

$$g_i(p) = \left.\frac{\partial f}{\partial x^i}\right|_p . \tag{1.11}$$

Proof. Let $B_r(p)$ be an open ball about p contained in U. Let $\mathbf{x} \in B_r(p)$ then the line $l : [0,1] \to I\!R^n$ from p to $\mathbf{x}$ given by

$$l(t) = p + t(\mathbf{x} - p),$$

has the properties $l(t) \subset B_r(p), 0 \leq t \leq 1$ and $l(0) = p$, $l(1) = \mathbf{x}$.

XXXXXXXXXXXXXXXX Picture XXXXXXXXXXXXXXXXXXXXXX

Therefore we can evaluate $f(l(t))$, and use the fundamental theorem of calculus to write,

$$\begin{aligned} f(\mathbf{x}) &= f(p) + \int_0^1 \frac{d}{dt} f(p + t(\mathbf{x} - p))dt \\ &= f(p) + f(l(1)) - f(l(0)) \\ &= f(\mathbf{x}). \end{aligned} \tag{1.12}$$

We expand out the derivative on the first line in equation 1.12 using the chain-rule 1.4.2 to get

$$\frac{d}{dt} f(l(t)) = \frac{d}{dt} f(p + t(\mathbf{x} - p)) = \sum_{i=1}^n (x^i - x_0^i) \left. \frac{\partial f}{\partial \xi^i} \right|_{\xi = p + t(\mathbf{x} - p)}, \tag{1.13}$$

where $p = (x_0^1, \ldots, x_0^n)$. Substituting from equation 1.13 into the first line in equation 1.12 gives

$$\begin{aligned} f(\mathbf{x}) &= f(p) + \int_0^1 \frac{d}{dt} f(p + t(\mathbf{x} - p))dt, \\ &= f(p) + \sum_{i=1}^n (x^i - x_0^i) \int_0^1 \left. \frac{\partial f}{\partial \xi^i} \right|_{\xi = p + t(\mathbf{x} - p)} dt. \end{aligned} \tag{1.14}$$

Therefore let

$$g_i(\mathbf{x}) = \int_0^1 \left. \frac{\partial f}{\partial \xi^i} \right|_{\xi = p + t(\mathbf{x} - p)} dt, \tag{1.15}$$

which satisfy 1.10 on account of equation 1.14. The smoothness property of the functions $g_i(\mathbf{x})$ follows by differentiation under the integral sign (see [12] where this is justified).

Finally substituting $\mathbf{x} = p$ into equation 1.15 gives

$$g_i(p) = \int_0^1 \left. \frac{\partial f}{\partial \xi^i} \right|_{\xi = p} dt = \left. \frac{\partial f}{\partial x^i} \right|_{\mathbf{x} = p}$$

which verifies equation 1.11. □

1.5 Vector Spaces, Basis, and Subspaces

We begin by reviewing the algebraic properties of matrices. Let $M_{m\times n}(I\!R)$ denotes the set of $m\times n$ matrices with real entries where $m,n\in Z^+$. A matrix $A\in M_{m\times n}(I\!R)$ has m rows and n columns. The components or entries of A are given by $A^a_j, 1\leq a\leq m, 1\leq j\leq n$. If $A,B\in M_{m\times n}(I\!R)$ and $c\in I\!R$ then $A+B, cA\in M_{m\times n}(I\!R)$, where in components

$$
\begin{aligned}
(A+B)^a_j &= A^a_j + B^a_j,\\
(cA)^a_j &= cA^a_j, \quad 1\leq a\leq m, 1\leq j\leq n.
\end{aligned}
\tag{1.16}
$$

If $A\in M_{m\times n}(I\!R), B\in M_{n\times p}(I\!R)$ then the product of A and B is the matrix $AB\in M_{m\times p}(I\!R)$ defined by

$$(AB)^a_s = \sum_{j=1}^n A^a_j\, B^j_s\ , \quad 1\leq a\leq m, 1\leq s\leq p.$$

Example 1.5.1. Let $\mathbf{x}\in I\!R^n$ and let $A\in M_{m\times n}(I\!R)$. If we view $\mathbf{x}$ as $\mathbf{x}\in M_{n\times 1}(I\!R)$ (so having n rows and one column) then $A\mathbf{x}\in M_{m\times 1}(I\!R)$ is a vector having m-components. The vector $A\mathbf{x}$ is given by just standard matrix vector multiplication.

The transpose of $A\in M_{m\times n}(I\!R)$ is the matrix $A^T\in M_{n\times m}$ with the property

$$(A^T)^i_a = A^a_i, \quad 1\leq i\leq n, 1\leq a\leq m.$$

If $A\in M_{n\times n}(I\!R)$ and $A^T=A$ then A is a symmetric matrix , if $A^T=-A$ then A is a skew-symmetric matrix . Finally if $A\in M_{n\times n}(I\!R)$ then the trace of A is

$$\mathrm{trace}(A) = \sum_{i=1}^n A^i_i$$

which is the sum of the diagonal elements of A.

Definition 1.5.2. A **vector space** V over $I\!R$ is a non-empty set with a binary operation $+ : V\times V\to V$, and a scalar multiplication $\cdot : I\!R\times V\to V$ which satisfy

V1) $(u+v)+w=(u+v)+w$,

V2) $u+v=v+u$,

V3) there exists $\mathbf{0} \in V$ such that $u + \mathbf{0} = u$,
V4) for all u there exists v such that $u + v = \mathbf{0}$,
V5) $1 \cdot u = u$,
V6) $(ab) \cdot u = a \cdot (b \cdot u)$,
V7) $a \cdot (u + v) = a \cdot u + a \cdot v$
V8) $(a + b) \cdot u = a \cdot u + b \cdot v$.
for all $u, v, w \in V$, $a, b \in I\!R$.

For vector-spaces we will drop the symbol $\cdot$ in $a \cdot u$ and write au instead. For example rule V8 is then $(a + b)u = au + bu$.

Example 1.5.3. Let $V = I\!R^n$, and let $+$ be ordinary component wise addition and let $\cdot$ be ordinary scalar multiplication.

Example 1.5.4. Let $V = M_{m\times n}(I\!R)$, and let $+$ be ordinary matrix addition, and let $\cdot$ be ordinary scalar multiplication as defined in equation 1.16. With these operations $M_{m\times n}(I\!R)$ is a vector-space.

Example 1.5.5. Let $f, g \in C^k(U)$ and $c \in I\!R$, and let

$$(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \quad (c \cdot f)(\mathbf{x}) = cf(\mathbf{x}) \quad \text{for all } \mathbf{x} \in U,$$

Properties 1 and 2 in Lemma 1.2.4 show that $f+g, c\cdot f \in C^k(U)$. Let $\mathbf{0} = 0_U$, the zero function on U defined in equation 1.1. With these definitions $C^k(U)$ is a vector-space over $I\!R$ (for any k including $k = \infty$).

Let $\mathcal{S}$ be a non-empty subset of V. A vector $v \in V$ is a *linear combination* of elements of S is there exists $c^i \in I\!R$ (recall an up index never means to the power), and $v_i \in S$ such that

$$v = \sum_{i=1}^{k} c^i v_i.$$

Note that the zero-vector $\mathbf{0}$ will always satisfy this condition with $c^1 = 0, v_1 \in \mathcal{S}$. The set of all vector which are a linear combination of S is called the span of S and denoted by span(S) .

A subset $\mathcal{S} \subset V$ is *linearly independent* if for every choice $\{v_i\}_{1\leq i\leq k} \subset S$, the only combination

$$\sum_{i=1}^{k} c^i v_i = \mathbf{0} \tag{1.17}$$

is $c^i = 0, 1 \leq i \leq k$. The empty set is taken to be linear independent.

Example 1.5.6. Let $V = \mathbb{R}^3$, and let

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \text{ and } \quad v = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \tag{1.18}$$

1. Is v a linear combination of elements in $\mathcal{S}$?
2. Is $\mathcal{S}$ a linearly independent set?

To answer the first question, we try to solve the system of equations

$$c^1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + c^2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c^3 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

for $c^1, c^2, c^3 \in \mathbb{R}$. This is the matrix equation

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \tag{1.19}$$

The augmented matrix and row reduced form are,

$$\begin{bmatrix} 1 & 1 & 0 & | & 2 \\ 2 & 0 & 1 & | & 3 \\ 1 & -1 & 1 & | & 1 \end{bmatrix} \text{rref} \rightarrow \begin{bmatrix} 1 & 0 & \frac{1}{2} & | & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} & | & \frac{1}{2} \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.$$

The system of equations in consistent. Therefore v is a linear combination of vectors in $\mathcal{S}$. We can determine the values of c_1, c_2, c_3 from the row reduced form of the coefficient matrix. This corresponding reduced form gives the equations

$$\begin{aligned} c_1 + \frac{1}{2}c_3 &= \frac{3}{2} \\ c_2 - \frac{1}{2}c_3 &= \frac{1}{2} \end{aligned}$$

There are an infinite number of solutions, given in parametric form by

$$\begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} + t \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \qquad t \in \mathbb{R}.$$

In this solution we let c^3 be the parameter t. If we choose for example $t = 0$, then $c^1 = \frac{3}{2}$ and $c^2 = \frac{1}{2}$ and we note that

$$\frac{3}{2}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix} = \begin{bmatrix}2\\3\\1\end{bmatrix}.$$

To answer the second questions on whether $\mathcal{S}$ is linearly independent we check for solution to equation 1.17, by looking for solutions to the *homogeneous* form of the systems of equations in (1.19),

$$\begin{bmatrix}1 & 1 & 0\\2 & 0 & 1\\1 & -1 & 1\end{bmatrix}\begin{bmatrix}c^1\\c^2\\c^3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}, \tag{1.20}$$

If the only solution is $c^1 = c^2 = c^3 = 0$, then the set $\mathcal{S}$ is a linearly independent set. The row reduced echelon form of the coefficient matrix for this system of equations is

$$\begin{bmatrix}1 & 0 & \frac{1}{2}\\0 & 1 & -\frac{1}{2}\\0 & 0 & 0\end{bmatrix}.$$

Therefore, there are an infinite number of solutions to the system (1.20), given by

$$\begin{bmatrix}c^1\\c^2\\c^3\end{bmatrix} = t\begin{bmatrix}-\frac{1}{2}\\\frac{1}{2}\\1\end{bmatrix} \quad t \in \mathbb{R}.$$

For example choosing $t = 1$ gives,

$$-\frac{1}{2}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Therefore $\mathcal{S}$ is not a linearly independent set.

A subset $\mathcal{S} \subset V$ is a *spanning set* if every $v \in V$ is a linear combination of elements of S, or $\text{span}(S) = V$.

Example 1.5.7. Continuing with example 1.5.6 we determine if $\mathcal{S}$ in example 1.18 is a spanning set. In order to do so, we try to solve the system of equations

$$c^1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + c^2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c^3 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

for $c^1, c^2, c^3 \in I\!R$ where the right hand side is **any** vector in $I\!R^3$. This is the matrix equation

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \tag{1.21}$$

The augmented matrix and a reduced form are,

$$\begin{bmatrix} 1 & 1 & 0 & | & a \\ 2 & 0 & 1 & | & b \\ 1 & -1 & 1 & | & c \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 & | & a \\ 0 & -2 & 1 & | & b-2a \\ 0 & 0 & 0 & | & c-b+a \end{bmatrix}.$$

Therefore if $c - b + a \neq 0$, the system has no solution and $\mathcal{S}$ is not a spanning set.

Lastly, a subset $\beta \subset V$ is a basis if β is linearly independent and a spanning set for V. We will always think of a basis β as an ordered set.

Example 1.5.8. The set $\mathcal{S}$ in equation 1.18 of example 1.5.6 is not a basis for $I\!R^3$. It is not linearly independent a linearly independent set, nor is it a spanning set.

Example 1.5.9. Let $V = I\!R^n$, and let

$$\beta = \{e_1, e_2, \dots, e_n\},$$

where

$$e_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad 1 \text{ in the } i^{th} \text{ row}, 0 \text{ otherwise.} \tag{1.22}$$

The set β is the standard basis for $I\!R^n$, and the dimension of $I\!R^n$ is n.

Example 1.5.10. Let $V = \mathbb{R}^3$, and let

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Is $\mathcal{S}$ a basis for $\mathbb{R}^3$? We first check if $\mathcal{S}$ is a linearly independent set. As in the example above we need to find the solutions to the homogeneous system $c^1 v_1 + c^2 v_2 + c^3 v_3 = \mathbf{0}$ (where v_1, v_2, v_3 are the three vectors in $\mathcal{S}$ given above). We get

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \text{rref} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Therefore the only solution to $c^1 v_1 + c^2 v_2 + c^3 v_3 = \mathbf{0}$ (the homogeneous system) is $c^1 = 0, c^2 = 0, c^3 = 0$. The row reduced form of the coefficient matrix also shows that $\mathcal{S}$ is a spanning set. Also see the theorem below.

Example 1.5.11. Let $V = M_{m \times n}(\mathbb{R})$, the vector-space of $m \times n$ matrices. Let $E^i_j \in V, 1 \le i \le m, 1 \le j \le n$ be the matrices

$$E^i_j = \begin{cases} 1 \text{ in the i}^{\text{th}} \text{ row j}^{\text{th}} \text{ column} \\ 0 \text{ everywhere else} \end{cases} \qquad 1 \le i \le m, 1 \le j \le n.$$

The collection $\{E^i_j\}_{1 \le i \le m, 1 \le j \le n}$ forms a basis for $M_{m \times n}(\mathbb{R})$ called the standard basis. We order them as $\beta = \{E^1_1, E^1_2, \dots, E^2_1, \dots, E^m_n\}$

A vector space is finite dimensional if there exists a basis β containing only a finite number of elements. It is a fundamental theorem that any two basis for a finite dimensional vector space have the same number of elements (cardinality) and so we define the dimension of V to be the cardinality of a basis β. A second fundamental theorem about finite dimensional vector spaces is the following

Theorem 1.5.12. *Let V be an n-dimensional vector space.*

1. *A spanning set $\mathcal{S}$ has at least n elements.*
2. *If $\mathcal{S}$ is a spanning set having n elements, then $\mathcal{S}$ is a basis.*
3. *A linearly independent set $\mathcal{S}$ has at most n elements.*

4. *If $\mathcal{S}$ is a linearly independent set with n elements, then $\mathcal{S}$ is a basis.*

Using part 4 of Theorem 1.5.12, we can concluded that $\mathcal{S}$ in example 1.5.10 is a basis, since we determined it was a linearly independent set. The set $\mathcal{S}$ in example 1.5.6 is not a basis for $\mathbb{R}^3$ by 4 of Theorem 1.5.12.

A useful property of a basis is the following.

Theorem 1.5.13. *Let V be an n-dimensional vector-space with basis $\beta = \{v_i\}_{1\leq i\leq n}$. Then every vector $v \in V$ can be written as a unique linear combination of elements of β.*

Proof. Since β is a spanning set, suppose that $v \in V$ can be written as

$$v = \sum_{i=1}^{n} c^i v_i \quad \text{and} \quad v = \sum_{i}^{n} d^i v_i. \tag{1.23}$$

Taking the difference of these two expressions gives,

$$\mathbf{0} = \sum_{i=1}^{n} (c^i - d^i) v_i.$$

Since the set β is a linearly independent set, we conclude $c^i - d^i = 0$, and the two expressions for v in (1.23) agree. □

The (unique) real numbers $c^1, \ldots, c^n$ are the *coefficients of the vector v in the basis β*. Also note that this theorem is true (and not hard to prove) for vector-spaces which are not necessarily finite dimensional.

A subset $W \subset V$ is a *subspace* if the set W is a vector-space using the vector addition and scalar-multiplication from V. The notion of a subspace is often more useful than that of a vector-space on its own.

Lemma 1.5.14. *A subset $W \subset V$ is a subspace if and only if*

1. $\mathbf{0} \in W$,
2. $u + v \in W, \ for\ all u, v \in W$,
3. $cu \in W, \ for\ all u \in W, c \in \mathbb{R}$.

Another way to restate these conditions is that a non-empty subset $W \subset V$ is a subspace if and only if it is closed under + and scalar multiplication. In order to prove this lemma we would need to show that W satisfies the axioms $V1$) through $V8$). This is not difficult because the set W inherits these properties from V.

Example 1.5.15. Let $S \subset V$ non-empty, and let $W = \text{span}(S)$. Then W is a subspace of V. We show this when S is finite, the infinite case is similar.

Let $v_1 \in S$, then $0v_1 = \mathbf{0}$ so 1) in Lemma 1.5.14 is true. Let $v, w \in \text{span}(S)$ then

$$v = \sum_{i=1}^{k} c^i v_i \quad w = \sum_{i=1}^{k} d^i v_i$$

where $S = \{v_1, \ldots, v_k\}$. Then

$$v + w = \sum_{i=1}^{k} (c^i + d^i) v_i$$

and so $v+w \in \text{span}(S)$, and 2) in Lemma 1.5.14 hold. Property 3) in Lemma 1.5.14 is done similarly.

1.6 Algebras

Definition 1.6.1. An algebra (over $\mathbb{R}$) denoted by $(V, *)$, is a vector space V (over $\mathbb{R}$) together with an operation $* : V \times V \to V$ satisfying

1. $(av_1 + bv_2) * w = a(v_1 * w) + b(v_2 * w)$,
2. $v * (aw_1 + bw_2) = av * w_1 + bv * w_2$.

The operation $*$ in this definition is called vector-multiplication. Properties 1 and 2 are referred to as the bi-linearity of $*$.

An algebra is *associative* if

$$v_1 * (v_2 * v_3) = (v_1 * v_2) * v_3, \tag{1.24}$$

commutative if

$$v_1 * v_2 = v_2 * v_1, \tag{1.25}$$

and *anti-commutative* if

$$v_1 * v_2 = -v_2 * v_1$$

for all $v_1, v_2, v_3 \in V$.

Example 1.6.2. Let $V = \mathbb{R}^3$ with its usual vector-space structure. Let the multiplication on V be the cross-product. Then $(V, \times)$ is an (anti-commutative) algebra which is not associative.

Example 1.6.3. Let $n \in Z^+$ and let $V = M_{n\times n}(\mathbb{R})$ be the vector-space of $n \times n$ matrices with ordinary matrix addition and scalar multiplication defined in equation 1.16. Let $*$ be matrix multiplication. This is an algebra because of the following algebraic properties of matrix multiplication:

$$(cA + B) * C = cA * C + B * C$$
$$A * (cB + C) = cA * B + A * C$$

for all $c \in \mathbb{R},\ A, B, C \in M_{n\times n}(\mathbb{R})$. These are properties 1 and 2 in Definition 1.6.1. This algebra is associative because matrix multiplication is associative.

Example 1.6.4. Let $V = C^k(U)$, where U is an open set in $\mathbb{R}^n$. This is vector-space (see example 1.5.5). Define multiplication of vectors by $f * g = f \cdot g$ by the usual multiplication of functions. Part 3) in Lemma 1.2.4 implies $f * g \in C^k(U)$. Therefore $C^k(U)$ is an algebra for any k (including $k = \infty$). This algebra is associative and commutative.

Example 1.6.5. Let $V = \mathbb{R}^2$ with the standard operations of vector addition and scalar multiplication. We define vector multiplication by considering V as the complex plane. The multiplication is determined by multiplying complex numbers,

$$(x + \mathbf{i}y)(u + \mathbf{i}v) = xu - yv + \mathbf{i}(xv + yu).$$

Therefore on V we define $(x, y) * (u, v) = (xu - yv, xv + yu)$ which makes $V = \mathbb{R}^2$ into a commutative and associative algebra.

Now let $V = \mathbb{R}^4$ and consider points in V as pairs of complex numbers (z_1, z_2). We can define a multiplication on $V = \mathbb{R}^4$ in the following way,

$$(z_1, z_2) * (w_1, w_2) = (z_1 w_1 - z_2 \bar{w}_2, z_1 w_2 + z_2 \bar{w}_1) \tag{1.26}$$

where $\bar{w}_1$ and $\bar{w}_2$ are the complex conjugates. This make $V = \mathbb{R}^4$ into an algebra called the quaternions.

If $(V, *)$ is an algebra and $W \subset V$ is a subset, then W is called a subalgebra if W is itself an algebra using the operation $*$ from V.

Lemma 1.6.6. *A subset $W \subset V$ of the algebra $(V, *)$ is a subalgebra if and only if*

1. *W is a subspace of V and*
2. *for all $w_1, w_2 \in W$, $w_1 * w_2 \in W$.*

Proof. If $W \subset V$ is a subalgebra, then it is necessarily a vector-space, and hence a subspace of V. In order that $*$ be well-defined on W it is necessary that for all $w_1, w_2 \in W$, that $w_1 * w_2 \in W$. Therefore conditions 1 and 2 are clearly necessary.

Suppose now that $W \subset V$ and conditions 1 and 2 are satisfied. By Lemma 1.5.14 condition 1 implies that W is a vector-space. Condition 2 implies that $*$ is well-defined on W, while the bi-linearity of $*$ on W follows from that on V. Therefore conditions 1 and 2 are sufficient. □

Example 1.6.7. Let $W \subset M_{2\times 2}(\mathbb{R})$ be the subset of upper-triangular 2×2 matrices. The set W is a subalgebra of $M_{2\times 2}(\mathbb{R})$ with ordinary matrix multiplication (see example 1.6.3). Properties 1 and 2 in Lemma 1.6.6 are easily verified.

Lemma 1.6.8. *Let $W \subset V$ be a subspace and $\beta = \{w_i\}_{1\leq i\leq m}$ a basis for W. Then W is a subalgebra if and only if $w_i * w_j \in W, 1 \leq i, j \leq m$.*

Proof. If W is a subalgebra then clearly $w_i * w_j \in W, 1 \leq i, j \leq m$ holds, and the condition is clearly necessary. We now prove that it is sufficient.

By Lemma 1.6.6 we need to show that if $u, v \in W$, then $u * v \in W$. Since β is a basis, there exist $a^i, b^i \in \mathbb{R}, 1 \leq i \leq m$ such that

$$u = \sum_{i=1}^{m} a^i \, w_i, \quad v = \sum_{i=1}^{m} b^i \, w_i.$$

Then using bi-linearity of $*$

$$u * v = \sum_{i,j=1}^{n} a^i b^i \, w_i * w_j. \tag{1.27}$$

By hypothesis $w_i * w_j \in W$, and since W is a subspace the right side of equation 1.27 is in W. Therefore by Lemma 1.6.6 W is a subalgebra. □

Let $(V,*)$ be a finite-dimensional algebra, and let $\beta=\{e_i\}_{1\leq i\leq n}$ be a basis of V. Since $e_i * e_j \in V$, there exists $c^k \in I\!R$ such that

$$e_i * e_j = \sum_{k=1}^{n} c^k e_k. \tag{1.28}$$

Now equation (1.28) holds for each choice $1 \leq i,j \leq n$ we can then write, for each $1 \leq i,j \leq n$ there exists $c^k_{ij} \in I\!R$, such that

$$e_i * e_j = \sum_{k=1}^{n} c^k_{ij} e_k. \tag{1.29}$$

The real numbers c^k_{ij} are called the structure constants of the algebra in the basis β.

Example 1.6.9. Let $\beta=\{E^1_1, E^1_2, E^2_1, E^2_2\}$ be the standard basis for $M_{2\times 2}(I\!R)$. Then

$$\begin{array}{llll} E^1_1 * E^1_1 = E^1_1, & E^1_1 * E^1_2 = E^1_2, & E^1_1 * E^2_1 = 0, & E^1_1 * E^2_2 = 0, \\ E^1_2 * E^1_1 = 0, & E^1_2 * E^1_2 = 0, & E^1_2 * E^2_1 = E^1_1, & E^1_2 * E^2_2 = E^1_2, \\ E^2_1 * E^1_1 = E^2_1, & E^2_1 * E^1_2 = E^2_2, & E^2_1 * E^2_1 = 0, & E^2_1 * E^2_2 = 0, \\ E^2_2 * E^1_1 = 0, & E^2_2 * E^1_2 = 0, & E^2_2 * E^2_1 = E^2_1, & E^2_2 * E^2_2 = E^2_2, \end{array} \tag{1.30}$$

which can also be written in table form

$*$	E^1_1	E^1_2	E^2_1	E^2_2
E^1_1	E^1_1	E^1_2	0	0
E^1_2	0	0	E^1_1	E^2_2
E^2_1	E^2_1	E^2_2	0	0
E^2_2	0	0	E^2_1	E^2_2

Therefore the non-zero structure constants (equation 1.29) are found from equation 1.30 to be

$$c^1_{11}=1, c^2_{12}=1, c^1_{22}=1, c^1_{23}=1, c^1_{24}=1, c^2_{31}=1, c^4_{32}=1, c^3_{43}=1, c^4_{44}=1.$$

1.7 Exercises

1. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$, and $\Psi : U \to \mathbb{R}^2$ be

$$\Phi(x,y) = (x+y, x^2+y^2), \Psi(x,y) = (\frac{x}{y}, x-y)$$

where $U = \{(x,y) \mid y \neq 0\}$.

   (a) Why are Φ and Ψ smooth functions?

   (b) Compute $\Psi \circ \Phi$ and find its domain, and state why it is C^∞.

   (c) Compute $\Phi \circ \Psi$ and find its domain.

   (d) Verify the chain-rule (1.5) for $\Psi \circ \Phi$.

2. Find the functions $g_1(x,y)$, $g_2(x,y)$ at $p = (1,2)$ for $f(x,y) = e^{x-y}$ in Theorem 1.4.8 and verify equation 1.10.

3. Let $\sigma(\tau) : I \to \mathbb{R}^n$ be a smooth curve where $I = (a,b), a, b \in \mathbb{R}$. Define the function $s : (a,b) \to \mathbb{R}$ by

$$s(t) = \int_a^t ||\frac{d\sigma}{d\tau}||d\tau \quad t \in I. \tag{1.31}$$

Note that the image of I is the open interval $(0, L(\sigma))$, that

$$\frac{ds}{dt} = ||\dot{\sigma}||,$$

and if $||\dot{\sigma}|| \neq 0$ then $s(t)$ is an invertible function.

Suppose that the inverse function of the function $s(t)$ in 1.31 exists and is C^∞. Call this function $t(s)$, and let $\gamma : (0, L(\sigma)) \to \mathbb{R}^n$ be the curve $\gamma(s) = \sigma \circ t(s)$. This parameterization of the curve σ is called the arc-length parameterization . The function $\kappa : (0, L(\sigma)) \to \mathbb{R}$

$$\kappa = ||\frac{d^2\gamma}{ds^2}||$$

is called the curvature of γ.

   (a) Compute the arc-length parameterization for the helix from example 1.3.1 on $I = (0, 2\pi)$.

(b) Compute $||\gamma'||$ for the helix ($'$ is the derivative with respect to s).

(c) Prove $||\gamma'|| = 1$ for any curve σ.

(d) Compute κ (as a function of s) for the helix.

(e) Show that $\kappa \circ s(t)$ for a curve σ can be computed by

$$\kappa(t) = \left(\frac{ds}{dt}\right)^{-1} ||\frac{dT}{dt}|| \tag{1.32}$$

where $T = ||\dot{\sigma}||^{-1}\dot{\sigma}$ is the unit tangent vector of $\sigma(t)$. (Hint: Apply the chain-rule to $\gamma(s) = \sigma(t(s))$.)

(f) Compute $\kappa(t)$ for the helix using the formula in 1.32.

(g) Compute the curvature for the curve

$$\sigma(t) = (e^t \cos t, e^t \sin t, t), \quad t \in (0, 2\pi).$$

4. Compute the matric products AB and BA if they are defined.

(a)

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 1 & 1 \end{bmatrix}$$

(b)

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & -2 & 3 \\ -1 & 1 & 1 \end{bmatrix}$$

5. Which of the following sets β of vector define a basis for $I\!R^3$. In cases when β is not a basis, state which property fails and prove that it fails.

(a) $$\beta = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \right\}.$$

(b) $$\beta = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix} \right\}.$$

(c) $\beta = \left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$

(d) $\beta = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} \right\}.$

6. Show that $\mathbb{R}^3$ with the cross-product is an algebra. Show that it is not associative.

7. In example 1.6.5 a multiplication on $\mathbb{R}^4$ is defined by equation 1.26. Compute the multiplication table for the standard basis for $\mathbb{R}^4$. Show that this multiplication is not commutative.

8. Show that algebra $M_{n\times n}(\mathbb{R})$ with matrix multiplication is an algebra and that it is also associative (see 1.24). Is it a commutative algebra (see 1.25)?

9. Let $M^0_{n\times n}(\mathbb{R}) \subset M_{n\times n}(\mathbb{R})$ be the subset of trace-free matrices,

   (1.33) $$M^0_{n\times n}(n, \mathbb{R}) = \{ \ A \in M_{n\times n}(\mathbb{R}) \mid \mathrm{trace}(A) = 0 \ \}.$$

   Show that,

   (a) $\mathrm{trace}(cA+B) = c\,\mathrm{trace}(A) + \mathrm{trace}(B)$, $c \in \mathbb{R}$, $A, B \in M_{n\times n}(\mathbb{R})$,

   (b) $\mathrm{trace}(AB) = \mathrm{trace}(BA)$, $A, B \in M_{n\times n}(\mathbb{R})$,

   (c) $M^0_{n\times n}(n, \mathbb{R}) \subset M_{n\times n}(\mathbb{R})$ is a subspace, and

   (d) that for $n > 1$, $M^0_{n\times n}(n, \mathbb{R}) \subset M_{n\times n}(\mathbb{R})$, is not a subalgebra of the algebra in example 1.6.3.

10. Consider the vector-space $V = M_{n\times n}(\mathbb{R})$ and define the function $[\ ,\ ] : V \times V \to V$ by

    $$[A, B] = AB - BA \quad A, B \in M_{n\times n}(\mathbb{R}).$$

    (a) Show that $(M_{n\times n}(\mathbb{R}), [\ ,\ ])$ is an algebra. This algebra is called $\mathbf{gl}(n, \mathbb{R})$, where $\mathbf{gl}$ stands for general linear.

    (b) Is $\mathbf{gl}(n, \mathbb{R})$ commutative or anti-commutative? (Consider $n = 1, n > 1$ separately.)

(c) Is $\mathbf{gl}(n, \mathbb{R})$ associative for $n > 1$?

(d) Show $[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0, \ \ for\ all A, B, C \in M_{n\times n}(\mathbb{R})$. Compare this with problem (b) above. (This is called the Jacobi identity for $\mathbf{gl}(n, \mathbb{R})$.)

11. Compute the structure constants for $\mathbf{gl}(2, \mathbb{R})$ using the standard basis for $M_{2\times 2}(\mathbb{R})$.

12. Let $\mathbf{sl}(n, \mathbb{R}) \subset \mathbf{gl}(n, \mathbb{R})$ be the subspace of trace-free matrices (see equation 1.33, and problem 7c),

    (a) Show that $\mathbf{sl}(n, \mathbb{R})$ is a subalgebra of $\mathbf{gl}(n, \mathbb{R})$. Compare with problem 7. (Hint Use part a and b of problem 7)

    (b) Find a basis for the subspace $\mathbf{sl}(2, \mathbb{R}) \subset \mathbf{gl}(2, \mathbb{R})$, and compute the corresponding structure constants. (Hint: $\mathbf{sl}(2, \mathbb{R})$ is 3 dimensional)

Chapter 2

Linear Transformations

2.1 Matrix Representation

Let V and W be two vector spaces. A function $T : V \to W$ is a *linear transformation* if

$$T(au + bv) = aT(u) + bT(v) \quad for\ all\ u, v \in V,\ a, b \in I\!R\,.$$

The abstract algebra term for a linear transformation is a *homomorphism* (of vector-spaces).

Example 2.1.1. The function $T : I\!R^2 \to I\!R^3$,

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x + 3y \\ x + y \\ x - y \end{bmatrix} \tag{2.1}$$

is a linear transformation. This is easily check by computing,

$$T\left(a\begin{bmatrix} x \\ y \end{bmatrix} + b\begin{bmatrix} x' \\ y' \end{bmatrix}\right) = \begin{bmatrix} 2(ax + bx') + 3(ay + by') \\ ax + bx' + ay + by' \\ ax + bx' - ay - by' \end{bmatrix} = aT\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + bT\left(\begin{bmatrix} x' \\ y' \end{bmatrix}\right).$$

Example 2.1.2. Let $A \in M_{m\times n}(I\!R$ be a $m \times n$ matrix and define $L_A : I\!R^n \to I\!R^m$ by

$$L_A(\mathbf{x}) = A\mathbf{x}$$

where $A\mathbf{x}$ is matrix vector multiplication (see example 1.5.1). It follows immediately from properties 1.16 of matrix multiplication that the function L_A is a linear transformation.

Note that in example 2.1.1

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Let V be an n dimensional vector-space with basis $\beta = \{v_1, \ldots, v_n\}$, and let W be an m dimensional vector space with basis $\gamma = \{w_1, \ldots, w_m\}$. Given a linear transformation $T : V \to W$ a linear transformation, we begin by applying T to the vector v_1, so that $T(v_1) \in W$. Since γ is a basis for W, there exists real numbers $A_1^1, A_1^2, \ldots, A_1^m$ such that

$$T(v_1) = \sum_{a=1}^{m} A_1^a w_a.$$

The role of the "extra index" 1 the coefficients A will be clear in a moment. Repeating this argument with all of the basis vector $v_j \in \beta$ we find that for each $j, 1 \leq j \leq n$ there $A_j^1, A_j^2, \ldots, A_j^m \in I\!R$ such that

$$T(v_j) = \sum_{a=1}^{m} A_j^a\, w_a. \tag{2.2}$$

The set of numbers $A_j^a, 1 \leq j \leq n, 1 \leq a \leq m$ form a matrix (A_i^a) with m rows, and n columns (so an $m \times n$ matrix). This is the *matrix representation of* T *in the basis'* β *and* γ and is denoted by

$$[T]_\beta^\gamma = (A_j^a). \tag{2.3}$$

Example 2.1.3. Continuing with $T : I\!R^2 \to I\!R^3$ in example 2.1.1 above, we compute $[T]_\beta^\gamma$ where

$$\beta = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad \gamma = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

So β and γ are the standard basis for $I\!R^2$ and $I\!R^3$ respectively.

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)=\begin{bmatrix}3\\1\\-1\end{bmatrix}=3\begin{bmatrix}1\\0\\0\end{bmatrix}+1\begin{bmatrix}0\\1\\0\end{bmatrix}-1\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Therefore

$$[T]_\beta^\gamma=\begin{bmatrix}2&3\\1&1\\1&-1\end{bmatrix}.$$

Note that the coefficients of $T(e_1)$ in the basis γ are in the first column, and those of $T(e_2)$ are in the second column. We now compute $[T]_\beta^\gamma$ in the following basis'

$$\beta=\left\{\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix}\right\},\quad \gamma=\left\{\begin{bmatrix}1\\2\\1\end{bmatrix},\begin{bmatrix}1\\0\\-1\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right\} \tag{2.4}$$

We get,

$$T\left(\begin{bmatrix}1\\-1\end{bmatrix}\right)=\begin{bmatrix}-1\\0\\2\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\2\\1\end{bmatrix}-\frac{3}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}-\begin{bmatrix}0\\1\\0\end{bmatrix}$$

and

$$T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=\begin{bmatrix}2\\1\\1\end{bmatrix}=\frac{3}{2}\begin{bmatrix}1\\2\\1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}-2\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Therefore

$$[T]_\beta^\gamma=\begin{bmatrix}\frac{1}{2}&\frac{3}{2}\\-\frac{3}{2}&\frac{1}{2}\\-1&-2\end{bmatrix} \tag{2.5}$$

Again note that the coefficients of $T(v_1)$ in the basis γ are in the first column, and those of $T(v_2)$ are in the second column.

Expanding on the remark at the end of the example, the **columns** of $[T]_\beta^\gamma$ are the coefficients of $T(v_1), T(v_2), \ldots, T(v_n)$ in the basis γ!

Example 2.1.4. Let $A \in M_{m\times n}$ and let $L_A : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformation in example 2.1.2. Let $\beta = \{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$

and let $\gamma = \{f_1, \ldots, f_m\}$ be the standard basis for $I\!R^m$. Then

$$L_A(e_j) = \begin{bmatrix} A^1_j \\ A^2_j \\ \vdots \\ A^m_j \end{bmatrix} \sum_{a=1}^{m} A^a_j f_a, \tag{2.6}$$

and therefore

$$[L_A]^\gamma_\beta = A.$$

The following lemma shows that the above example essentially describes all linear transformations from $I\!R^n$ to $I\!R^m$.

Lemma 2.1.5. *Let $T : I\!R^n \to I\!R^m$ be a linear transformation. There exists a matrix $A \in M_{m\times n}(I\!R)$ such that*

$$T(\mathbf{x}) = L_A(\mathbf{x}) \quad for\ all\ \mathbf{x} \in I\!R^n.$$

Suppose now that $T : V \to W$ is a linear transformation between the finite dimensional vector spaces V and W. Let $\beta = \{v_i\}_{1\leq i\leq n}$ be a basis for V and $\gamma = \{w_a\}_{1\leq a\leq n}$ for W. Let $v \in V$ which in the basis β is

$$v = \sum_{i=1}^{n} \xi^i\, v_i, \tag{2.7}$$

where $\xi^i \in I\!R, 1 \leq i \leq n$ are the coefficients of v in the basis β. Now let $w = T(v)$. Then $w \in W$ and so can be written in terms of the basis γ as

$$w = \sum_{a=1}^{m} \eta^a\, f_a$$

where $\eta_a \in I\!R, 1 \leq a \leq m$. We then find

Lemma 2.1.6. *The coefficients of the vector $w = T(v)$ are given by*

$$\eta^a = \sum_{i=1}^{n} A^a_i \xi^i$$

where A is the $m \times n$ matrix $A = [T]^\gamma_\beta$.

Proof. We simply expand out $T(v)$ using equation (2.7), and using the linearity of T to get

$$T(v) = T(\sum_{i=1}^{n} \xi^i \, v_i) = \sum_{i=1}^{n} \xi^i \, T(v_i).$$

Now substituting for $T(v_i) = \sum_{a=1}^{m} A_i^a w_a$, we get

$$T(v) = \sum_{i=1}^{n} \xi^i \left(\sum_{a=1}^{m} A_i^a \, w_a \right) = \sum_{a=1}^{m} \left(\sum_{i=1}^{n} A_i^a \xi^i \right) w_a. \tag{2.8}$$

Since $\{w_a\}_{1 \le a \le m}$ is a basis the coefficients of w_a in equation 2.8 must be the same, and so

$$\eta^a = \sum_{i=1}^{n} A_i^a \xi^i. \tag{2.9}$$

□

The m-coefficients η^a of w can be thought of as an column vector, and the n coefficients ξ^i of v can be thought of as a column vector. Equation (2.9) then reads

$$\begin{bmatrix} \eta^1 \\ \vdots \\ \eta^m \end{bmatrix} = A \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^n \end{bmatrix}$$

where the right side is standard matrix vector multiplication. This can also be written,

$$[w]_\gamma = A[v]_\beta \tag{2.10}$$

where $[w]_\gamma$ is the column m-vector of coefficients of w in the basis γ, and $[v]_\beta$ is the column n-vector of the coefficients of v in the basis β.

What we have just seen by Lemma 2.1.6 is that every linear transformation $T \in L(V, W)$ where V is dimension n and W is dimension m is completely determined by its value on a basis, or by its matrix representation. That is if $\beta = \{v_i\}_{1 \le i \le n}$ is a basis then given $T(v_i)$ we can compute $T(v)$ where $v = \sum_{i=1}^{n} \xi^i v_i$, $\xi^i \in I\!R$ to be

$$T(\sum_{i=1}^{n} \xi^i v_i) = \sum_{i=1}^{n} \xi_i T(v^i).$$

Conversely any function $\hat{T} : \beta \to W$ extends to a unique linear transformation $T : V \to W$ defined by

$$T(\sum_{i=1}^{n} \xi^i v_i) = \sum_{i=1}^{n} \xi_i \hat{T}(v^i)$$

which agrees with $\hat{T}$ on the basis β. We have therefore proved the following lemma.

Lemma 2.1.7. *Let β be a basis for the vector space V. Every linear transformation $T \in L(V, W)$ uniquely determines a function $\hat{T} : \beta \to W$. Conversely every function $\hat{T} : \beta \to W$ determines a unique linear transformation $T \in L(V, W)$ which agrees with $\hat{T}$ on the basis β.*

A simple corollary is then

Corollary 2.1.8. *Let $T, U \in L(V, W)$, with the dimension of V being n and the dimension of W being m. Then $T = U$ if and only if $[T]^\gamma_\beta = [U]^\gamma_\beta$ for any (and hence all) basis β for V and γ for W.*

2.2 Kernel, Rank, and the Rank-Nullity Theorem

Let $T : V \to W$ be a linear transformation the *kernel* of T (denoted $\ker(T)$) or the *null space* of T is the set

$$\ker(T) = \{\ v \in V \mid T(v) = \mathbf{0}_W\ \}$$

where $\mathbf{0}_W$ is the zero-vector in W. The *image* or *range* of T (denoted $R(T)$) is

$$R(T) = \{\ w \in W \mid w = T(v) \text{ for some } v \in V\ \}.$$

Lemma 2.2.1. *Let $T : V \to W$ be a linear transformation, then $\ker(T)$ is a subspace of V, and $R(T)$ is a subspace of W.*

Proof. We need to show that $\ker(T)$ satisfies conditions 1),2),3) from the subspace Lemma 1.5.14. We begin by computing

$$T(\mathbf{0}) = T(a\mathbf{0}) = aT(\mathbf{0}) \quad \text{for all } a \in \mathbb{R}.$$

Therefore $T(\mathbf{0}) = \mathbf{0}_W$, and $\mathbf{0} \in \ker(T)$. Now suppose $u, v \in \ker(T)$, then

$$T(au + v) = aT(u) + T(v) = \mathbf{0}_W + \mathbf{0}_W = 0.$$

This shows property 2) and 3) hold from Lemma 1.5.14, and so $\ker(T)$ is a subspace of V.

The proof that $R(T)$ is a subspace is left as an exercise. □

The *rank* of T denoted by $\text{rank}(T)$, is defined to be the dimension of $R(T)$,

$$\text{rank}(T) = \dim R(T).$$

The *nullity* of T denoted by $\text{nullity}(T)$, is defined to be

$$\text{nullity}(T) = \dim \ker(T).$$

Example 2.2.2. Let $L_A : I\!R^n \to I\!R^m$ be the linear transformation $L_A(\mathbf{x}) = A\mathbf{x}$ from example 2.1.2. Let $\beta = \{e_i\}_{1 \le i \le n}$ is the standard basis for $I\!R^n$ defined in 1.22. Then $\mathbf{x} = x^1 e_1 + x^2 e_2 + \ldots x^n e_n$ and

$$L_A(\mathbf{x}) = A(\sum_{i=1}^{n} x^i\, e_i) = \sum_{i=1}^{n} x^i\, A(e_i). \tag{2.11}$$

The kernel of L_A is also called the the kernel of A, that is

$$\ker(A) = \{\ \mathbf{x} \ \in\ I\!R^n \mid A\mathbf{x} = \mathbf{0}\}. \tag{2.12}$$

where $\mathbf{0}$ is the zero vector in $I\!R^m$.

The range space $R(L_A)$ of L_A is then found from equation 2.11 to be

$$R(L_A) = \text{span}\{A(e_1), \ldots, A(e_n)\}.$$

By equation 2.6 we have

$$R(L_A) = \text{span}\left\{ \begin{bmatrix} A^1_1 \\ A^2_1 \\ \vdots \\ A^m_1 \end{bmatrix}, \ldots, \begin{bmatrix} A^1_n \\ A^2_n \\ \vdots \\ A^m_n \end{bmatrix} \right\} \tag{2.13}$$

or $R(L_A)$ is the span of the columns of A.

Example 2.2.3. Let $Z : I\!R^5 \to I\!R^4$ be the linear transformation

$$Z\left(\begin{bmatrix} v \\ w \\ x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} v + 2w + 3y - 2z \\ w + 2x + 3y + 3z \\ v + w - x + y - 3z \\ w + 3x + 4y + 5z \end{bmatrix}. \tag{2.14}$$

To find $\ker(Z)$ we therefore need to solve

$$\begin{bmatrix} v + 2w + 3y - 2z \\ w + 2x + 3y + 3z \\ v + w - x + y - 3z \\ w + 3x + 4y + 5z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The parametric solution to these equation is

$$\begin{bmatrix} v \\ w \\ x \\ y \\ z \end{bmatrix} = t_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} 0 \\ 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

Note that $\ker(Z)$ is a two-dimensional subspace with basis

$$\beta_{\ker(T)} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Writing the linear transformation as

$$Z\left(\begin{bmatrix} v \\ w \\ x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 0 & 3 & -2 \\ 0 & 1 & 2 & 3 & 3 \\ 1 & 1 & -1 & 1 & -3 \\ 0 & 1 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} v \\ w \\ x \\ y \\ z \end{bmatrix}$$

then from equation 2.13

$$R(Z)\text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \\ -3 \\ 5 \end{bmatrix} \right\}.$$

We leave as an exercise the following lemma.

Lemma 2.2.4. *Let $T : V \to W$ be a linear transformation with $\beta = \{v_i\}_{1 \leq i \leq n}$ a basis for V and $\gamma = \{w_a\}_{1 \leq a \leq m}$ a basis for W. The function T is injective if and only if (see equation 2.12),*

$$\ker(A) = \mathbf{0}$$

where $\mathbf{0}$ is the zero vector in $\mathbb{R}^n$.

The next proposition provides a good way to find a basis for $R(T)$.

Proposition 2.2.5. *Let $T : V \to W$ be a linear transformation, where $\dim V = n$ and $\dim W = m$. Suppose $\beta_1 = \{v_1, \ldots, v_k\}$ form a basis for $\ker T$, and that $\beta = \{v_1, \ldots, v_w, u_{k+1}, \ldots, u_n\}$ form a basis for V, then $\{T(u_{k+1}), \ldots, T(u_n)\}$ form a basis for $R(T)$.*

In practice this means we first find a basis for $\ker(T)$ and then extend it to a basis for V. Apply T to the vectors not in the kernel, and these are a basis for $R(T)$.

Example 2.2.6. Let $Z : \mathbb{R}^5 \to \mathbb{R}^4$ be the linear transformation in equation 2.14 in example 2.2.3. We extend $\ker Z$ to a basis as in Lemma 2.2.5

$$\beta = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

(Exercise: Check this is a basis!). We then find

$$Z\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad Z\left(\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad Z\left(\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 2 \\ -1 \\ 3 \end{bmatrix},$$

which by Lemma 2.2.5 form a basis for $R(Z)$. Note that $\operatorname{rank}(Z) = 3$.

By counting the basis elements for V in the Lemma 2.2.5 leads to a Theorem known as the rank-nullity theorem, or the dimension theorem.

Theorem 2.2.7. *If V is a finite dimensional vector-space and $T : V \to W$ is a linear transformation, then*

$$\dim V = \dim \ker(T) + \dim R(T) = \operatorname{nullity}(T) + \operatorname{rank}(T).$$

2.3 Composition, Inverse and Isomorphism

The composition of two linear transformations plays a significant role in linear algebra. We begin with a straight-forward proposition.

Proposition 2.3.1. *Let U, V, W be vector-spaces, and let $T : U \to V$, and $S : V \to W$ be linear transformations. The composition $S \circ T : U \to W$ is a linear transformation.*

Proof. Exercise 3. □

Example 2.3.2. Let $S : \mathbb{R}^4 \to \mathbb{R}^2$ be the linear transformation

$$S\left(\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2w + 4y + 4z \\ x - 2z \end{bmatrix}. \tag{2.15}$$

Let T be the linear transformation given in equation 2.1 in example 2.1.1. Compute $T \circ S$ or $S \circ T$ if they are defined. Only $T \circ S$ is defined and we have,

(2.16)

$$T \circ S\left(\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}\right) = T\left(\begin{bmatrix} 2w + 4y + 4z \\ x - 2z \end{bmatrix}\right) = \begin{bmatrix} 2(2w + 4y + 4z) + 3(x - 2z) \\ 2w + 4y + 4z + x - 2z \\ 2w + 4y + 4z - (x - 2z) \end{bmatrix}$$

$$= \begin{bmatrix} 4w + 3x + 8y + 2z \\ 2w + x + 4y + 2z \\ 2w - x + 4y + 6z \end{bmatrix}$$

Note that it is not possible to compute $S \circ T$.

Finally, suppose that $T : U \to V$, $S : V \to W$ are linear transformations and that

$$\beta = \{u_1, \ldots, u_n\},\ \gamma = \{v_1, \ldots, v_m\},\ \delta = \{w_1, \ldots, w_p\}$$

are basis for U, V and W respectively. Let

$$[T]^\gamma_\beta = (B^a_j) \qquad 1 \leq a \leq m, 1 \leq j \leq n$$

and

$$[S]^\delta_\gamma = (A^\alpha_a) \qquad 1 \leq \alpha \leq p,\ 1 \leq a \leq m,$$

and

$$[S \circ T]^\delta_\beta = (C^\alpha_j) \qquad 1 \leq \alpha \leq p,\ 1 \leq j \leq n$$

be the matrix representations of T $(m \times n)$,S $(p \times n)$ and $S \circ T$ $(p \times n)$.

Theorem 2.3.3. *The coefficients of the $p \times n$ matrix C are*

$$C^\alpha_j = \sum_{b=1}^{m} A^\alpha_a B^a_j. \tag{2.17}$$

Proof. Let's check this formula. We compute

$$\begin{aligned} S \circ T(u_j) &= S\left(\sum_{a=1}^{m} B^a_j v_a\right) \\ &= \sum_{a=1}^{m} B^a_j S(v_a) \quad \text{by linearity of } S \\ &= \sum_{a=1}^{m} B^a_j \left(\sum_{\alpha=1}^{p} A^\alpha_a w_\alpha\right) \\ &= \sum_{\alpha=1}^{p} \left(\sum_{a=1}^{m} A^\alpha_a B^a_j\right) w_\alpha \quad \text{rearrange the summation.} \end{aligned}$$

This is formula (2.17). □

This theorem is the motivation on how to define matrix multiplication. If $A \in M_{p\times m}(\mathbb{R})$ and $B \in M_{m\times n}(\mathbb{R})$ then the product $AB \in M_{p\times n}$ is the $p \times n$ matrix C whose entries are given by (2.17).

Example 2.3.4. Let T and S be from equations (2.1) and (2.15), we then check (2.17) using the standard basis for each space. That is we check $[T \circ S] = [T][S]$ where these are the matrix representations in the standard basis. We have

$$[T] = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad [S] = \begin{bmatrix} 2 & 0 & 4 & 4 \\ 0 & 1 & 4 & 6 \end{bmatrix}$$

while from equation (2.16),

$$[T \circ S] = \begin{bmatrix} 4 & 3 & 8 & 2 \\ 2 & 1 & 4 & 2 \\ 2 & -1 & 4 & 6 \end{bmatrix}.$$

Multiplying,

$$[T][S] = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 4 & 4 \\ 0 & 1 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 8 & 2 \\ 2 & 1 & 4 & 2 \\ 2 & -1 & 4 & 6 \end{bmatrix}.$$

so we have checked (2.17) in the standard basis.

We check formula (2.17) again but this time we will use the basis (2.4) for $I\!R^2, I\!R^3$, and $\delta = \{e_1, e_2, e_3, e_4\}$ is again the standard basis for $I\!R^4$. We've got $[T]^\gamma_\beta$ is equation (2.5), so we need to compute $[S]^\beta_\delta$. This is

$$[S]^\beta_\delta = \begin{bmatrix} 0 & -1 & 0 & 2 \\ 2 & 1 & 4 & 2 \end{bmatrix}$$

while using (2.16), we have

$$[T \circ S]^\gamma_\delta = \begin{bmatrix} 3 & 1 & 6 & 4 \\ 1 & 2 & 2 & -2 \\ -4 & -1 & -8 & -6 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ -\frac{3}{2} & \frac{1}{2} \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 & 2 \\ 2 & 1 & 4 & 2 \end{bmatrix}$$

This verifies equation (2.16).

A linear transformation $T : V \to W$ is an *isomorphism* if T is an invertible function. That is T is a bijective function, and so one-to-one and onto.

Proposition 2.3.5. *A linear transformation $T : V \to W$ between two n-dimensional vector spaces is an isomorphism if and only if* $\ker(T) = \{\mathbf{0}\}$.

Proof. By exercise 2 in this chapter T is injective if and only if $\ker(T) = \{\mathbf{0}\}$. By the dimension Theorem 2.2.7 *dim* $\ker(T) = 0$ if and only if $R(T) = W$. In other words T is injective if and only if T is surjective. □

An $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix B such that

$$AB = BA = I$$

where I is the $n \times n$ identity matrix, in which case we write $B = A^{-1}$.

The standard test for invertibility is the following.

Proposition 2.3.6. *A matrix $A \in M_{n\times n}$ is invertible if and only if* $\det(A) \neq 0$*. Furthermore if A is invertible then A^{-1} is obtained by row reduction of the augmented system*

$$(\; A \mid I \;) \to (\; I \mid A^{-1} \;)$$

The invertibility of a matrix and an isomorphism are related by the next lemma.

Proposition 2.3.7. *Let $T : V \to W$ be an isomorphism from the n-dimensional vector-space V to the m dimensional vector-space W, with β a basis for V and γ a basis for W. Then*

1. *W is n-dimensional, and*
2. *$T^{-1} : W \to V$ is linear.*
3. *$[T^{-1}]^{\beta}_{\gamma} = \left([T]^{\gamma}_{\beta}\right)^{-1}$, where $\left([T]^{\gamma}_{\beta}\right)^{-1}$ is the* **inverse matrix** *of $[T]^{\gamma}_{\beta}$.*

Example 2.3.8. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation,

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x+y \\ x+y \end{bmatrix}.$$

We have according to Lemma 2.3.7,

$$[T] = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{and } [T]^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

Therefore,

$$T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = [T]^{-1}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x-y \\ 2y-x \end{bmatrix}.$$

Double checking this answer we compute

$$T\circ T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = T\left(\begin{bmatrix} x-y \\ 2y-x \end{bmatrix}\right) = \left(\begin{bmatrix} 2(x-y)+2y-x \\ x-y+2y-x \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$$

By applying the dimension theorem we have a simple corollary.

Corollary 2.3.9. *A function $T \in L(V,W)$ with* $\dim V = \dim W$ *is an isomorphism if and only if* $\ker T = \{\mathbf{0}\}$.

Let $L(V,W)$ be the set of all linear transformations from V to W.

Lemma 2.3.10. *Let $S,T \in L(V,W)$ and $a \in I\!R$ then $aS+T \in L(V,W)$ where*

$$(aS+T)(v) = aS(v) + T(v).$$

In particular this makes $L(V,W)$ is a vector-space.

Proof. We need to show $aS+T$ is indeed linear. Let $v_1, v_2 \in V$, $c \in I\!R$, then

$$\begin{aligned}(aS+T)(cv_1+v_2) &= aS(cv_1+v_2) + T(cv_1+v_2)\\ &= acS(v_1) + aS(v_2) + cT(v_1) + T(v_2)\\ &= c(aS+T)(v_1) + (aS+T)(v_2)\end{aligned}$$

Therefore $aS+T \in L(V,W)$. The fact that $L(V,W)$ is a vector-space can be shown in a number of ways. □

We now find

Lemma 2.3.11. *Let V be an n dimensional vector-space and W an m dimensional vector space. Then*

$$\dim L(V,W) = nm$$

Proof. Let $\beta = \{v_i\}\{1 \leq i \leq n\}$ and $\gamma = \{w_a\}_{1 \leq a \leq m}$ be basis for V and W respectively. Define $\Phi : L(V,W) \to M_{m\times n}(I\!R)$ by

$$\Phi(T) = [T]^\gamma_\beta. \tag{2.18}$$

We claim Φ is an isomorphism. First Φ is a linear transformation, which is an exercise. It follows from Corollary 2.1.8, that $\Phi(T_0) = 0_{m\times n}$ is the unique linear transformation with the 0 matrix for its representation, and so Φ is injective. If $A \in M_{m\times n}$ then

$$\hat{T}(v_i) = \sum_{a=1}^{m} A^a_i w_a\,,$$

by Lemma 2.1.7 extends to a linear transformation $T : V \to W$ with $[T]^\gamma_\beta = A$. Therefore Φ is onto, and so an isomorphism. □

2.4 Exercises

1. A transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by

$$T(\begin{bmatrix} x \\ y \\ z \end{bmatrix}) = \begin{bmatrix} x+y-z \\ 2y+z \end{bmatrix}. \tag{2.19}$$

   (a) Show that T is linear.

   (b) Find the matrix representing T for each of the following basis.

   i. $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.

   ii. $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.

   iii. $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$.

   iv. $\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$.

   (c) Compute the kernel of T using the matrix representations in (i), (ii) and (iii). Show that you get the same answer in each case.

2. Prove that if $T : V \to W$ is a linear transformation, then T is injective if and only if kernel$(T) = \mathbf{0}_V$.

3. Prove Lemma 2.2.4. (Hint use exercise 2.)

4. Let $T : V \to W$ be a linear transformation. Prove that $R(T) \subset W$ is a subspace.

5. For the linear transformation S in equation 2.15 from example 2.3.2, find a basis for $\ker(S)$ and $R(S)$ using lemma 2.2.5

6. (a) Find a linear transformation $T : I\!R^4 \to I\!R^3$ with

$$\ker(T) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \right\} \quad \text{and} \quad \text{image}(T) = \text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Express your answer as $T(xe_1 + ye_2 + ze_3 + we_4)$.

(b) Is your answer unique?

(c) Can you repeat a) with

$$\ker(T) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 3 \\ 1 \end{bmatrix} \right\} \quad \text{and } R(T) = \text{span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \right\}?$$

7. Prove Lemma 2.1.5.

8. Prove Proposition 2.2.5.

9. If $T : V \to W$ and $S : W \to U$ are linear transformations, show that $S \circ T : V \to U$ is a linear transformation.

10. Let $S : I\!R^2 \to I\!R^2$ be given by $S(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x - y \\ 2y - x \end{bmatrix}$.

(a) Show that S is a linear transformation.

(b) Find $S \circ T \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right)$, where T is the linear transformation given in equation 2.19 of problem 1.

(c) Find $[T]$, $[S]$, and $[S \circ T]$ the matrices of these linear transformations with respect to the standard basis for $I\!R^3$ and $I\!R^2$

(d) Check that $[S \circ T] = [S][T]$.

(e) Find the matrix for the linear transformation $S^2 = S \circ S$.

(f) Find the inverse function $S^{-1} : I\!R^2 \to I\!R^2$. Show that S^{-1} is linear and that $[S^{-1}] = [S]^{-1}$.

11. Prove that $\Phi : L(V, W) \to M_{m\times n}(\mathbb{R})$ defined in equation 2.18 is a linear transformation, by showing

$$\Phi(aT + S) = a\Phi(T) + \Phi(S) \quad , a \in \mathbb{R}, \; T, S \in L(V, W).$$

Chapter 3

Tangent Vectors

In multi-variable calculus a vector field on $\mathbb{R}^3$ is written

$$\mathbf{v} = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

where $P, Q, R \in C^\infty(\mathbb{R}^3)$. While in differential geometry the same vector field would be written as a differential operator

$$\mathbf{v} = P(x, y, z)\partial_x + Q(x, y, z)\partial_y + R(x, y, z)\partial_z.$$

This chapter will show why vector fields are written as differential operators and then examine their behavior under a changes of variables.

3.1 Tangent Vectors and Curves

Let $p \in \mathbb{R}^n$ be a point. We begin with a preliminary definition that a tangent vector at p is an ordered pair $(p\,, \mathbf{a})$ where $\mathbf{a} \in \mathbb{R}^n$ which we will write as $(\mathbf{a})_p$. The point p is always included - we can not think of tangent vectors as being at the origin of $\mathbb{R}^n$ unless p is the origin.

Let V_p be the set of tangent vectors at the point p. This set has the structure of a vector space over $\mathbb{R}$ where we define addition and scalar multiplication by

$$c(\mathbf{a})_p + (\mathbf{b})_p = (c\mathbf{a} + \mathbf{b})_p, \ \ c \in \mathbb{R}\,, \mathbf{a}, \mathbf{b} \in \mathbb{R}^n.$$

This purely set theoretical discussion of tangent vectors does not reflect the geometric meaning of tangency.

Recall that a smooth curve is function $\sigma : I \to I\!\!R^n$, where $I \subset I\!\!R$ is an open interval. Let $t_0 \in I$, and let $p = \sigma(t_0)$. The tangent vector to σ at $p = \sigma(t_0) \in I\!\!R^n$ is

$$(\mathbf{a})_p = \left(\left. \frac{d\sigma}{dt} \right|_{t=t_0} \right)_{\sigma(t_0)} = (\dot\sigma(t_0))_{\sigma(t_0)} .$$

Example 3.1.1. For the helix $\sigma(t) = (\cos t, \sin t, t)$ the tangent vector at $t = \frac{\pi}{4}$,is

$$\left(\left. \frac{d\sigma}{dt} \right|_{t=\frac{\pi}{4}} \right)_{(\frac{1}{\sqrt2}, \frac{1}{\sqrt2}, \frac{\pi}{4})} = \left(-\frac{1}{\sqrt2}, \frac{1}{\sqrt2}, 1 \right)_{(\frac{1}{\sqrt2}, \frac{1}{\sqrt2}, \frac{\pi}{4})} . \tag{3.1}$$

Consider the curve

$$\alpha(t) = \left(-\frac{1}{\sqrt2} t + \frac{1}{\sqrt2}, \frac{1}{\sqrt2} t + \frac{1}{\sqrt2}, t + \frac{\pi}{4} \right)$$

The tangent vector to $\alpha(t)$ at $t = 0$ is exactly the same as that in (3.1). Two completely different curves can have the same tangent vector at a point.

A *representative curve* of a given vector $(\mathbf{a})_p \in V_p$ is a smooth curve $\sigma : I \to I\!\!R^n$ defined on a non-empty open interval I that satisfies $\sigma(t_0) = p$ and $\dot\sigma(t_0) = \mathbf{a}$ for some $t_0 \in I$.

Example 3.1.2. Let $(\mathbf{a})_p \in V_p$. The curve $\sigma : I\!\!R \to I\!\!R^n$ given by

$$\sigma(t) = p + t\mathbf{a} \tag{3.2}$$

satisfies $\sigma(0) = p$ and $\dot\sigma(0) = \mathbf{a}$, and is a representative of the tangent vector $(\mathbf{a})_p$. The curve with components

$$\sigma^i(t) = x^i_0 + e^{a^i(t-1)} \tag{3.3}$$

where $p = (x^i_0)_{1 \le i \le n}$ is also a representative of $(\mathbf{a})_p$ with $t_0 = 1$

3.2 Derivations

The second geometric way to think about tangent vectors is by using the directional derivative. Let $(\mathbf{a})_p$ be a tangent vector at the point p and let $\sigma : I \to \mathbb{R}^n$ be representative curve (so $\sigma(t_0) = p, \dot\sigma(t_0) = \mathbf{a}$). Recall that the directional derivative at p of a smooth function $f \in C^\infty(\mathbb{R}^n)$ along $(\mathbf{a})_p$ and denoted $D_{\mathbf{a}}f(p)$ is

$$
\begin{aligned}
D_{\mathbf{a}}f(p) &= \left(\frac{d}{dt} f\circ\sigma\right)\bigg|_{t=t_0} \\
&= \sum_{i=1}^n \frac{\partial f}{\partial x^i}\bigg|_{\sigma(t_0)} \frac{d\sigma^i}{dt}\bigg|_{t=t_0} \\
&= \sum_{i=1}^n a^i \frac{\partial f}{\partial x^i}\bigg|_p ,
\end{aligned} \tag{3.4}
$$

where the chain-rule 1.4.2 was used in this computation. Using this formula let's make a few observations about the mathematical properties of the directional derivative.

Lemma 3.2.1. *Let $(\mathbf{a})_p \in V_p$, then the directional derivative given in equation 3.4 has the following properties:*

1. $D_{\mathbf{a}}f(p) \in \mathbb{R}$,
2. *The directional derivative of f only depends only the tangent vector $(\mathbf{a})_p$ and not on the curve σ used to represent it.*
3. *The function f in equation 3.4 need not be C^∞ everywhere but only C^∞ on some open set U in $\mathbb{R}^n$ with $p \in U$.*
4. *If g is a smooth function defined on some open set V with $p \in V$ we can compute the directional derivatives*

$$
\begin{aligned}
D_{\mathbf{a}}(cf+g)(p) &= cD_{\mathbf{a}}f(p) + D_{\mathbf{a}}g(p) \\
D_{\mathbf{a}}(fg)(p) &= D_{\mathbf{a}}(f)(p)g(p) + f(p)D_{\mathbf{a}}g(p)
\end{aligned} \tag{3.5}
$$

where $c \in \mathbb{R}$.

Proof. These claims are easily verified. For example, to verify the second property in equation 3.5,

$$\begin{aligned} D_{\mathbf{a}}(fg)(p) &= \frac{d}{dt}\left(f\circ\sigma\, g\circ\sigma\right)|_{t=t_0} \\ &= \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x^i}\right|_{\sigma(t_0)} \left.\frac{d\sigma^i}{dt}\right|_{t=t_0} g(\sigma(t_0)) + f(\sigma(t_0)) \left.\frac{\partial g}{\partial x^i}\right|_{\sigma(t_0)} \left.\frac{d\sigma^i}{dt}\right|_{t=t_0} \\ &= D_{\mathbf{a}}f(p)g(p) + f(p)D_{\mathbf{a}}g(p) \quad \text{by equation 3.4.} \end{aligned}$$

□

The directional derivative leads to the idea that a tangent vector at a point p is something that differentiates smooth function defined in an open set about that point, and satisfies properties 1-4 above. This is exactly what is done in modern differential geometry and we now pursue this approach.

The next definition, which is a bit abstract, is motivated by point 3 in Lemma 3.2.1 above. Let $p \in \mathbb{R}^n$, and define

$$C^\infty(p) = \bigcup_{U \subset \mathbb{R}^n} C^\infty(U), \quad \text{where } p \in U \text{ and } U \text{ is open.}$$

If $f \in C^\infty(p)$ then there exists an open set U containing p and $f \in C^\infty(U)$. Therefore $C^\infty(p)$ consists of all functions which are C^∞ on some open set containing p.

The set $C^\infty(p)$ has the similar algebraic properties as $C^\infty(U)$. For example if $f, g \in C^\infty(p)$ with $\mathrm{Dom}(f) = U$ and $\mathrm{Dom}(g) = V$, then define

$$(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}),\ (fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x}), \quad \mathbf{x} \in U \cap V,$$

Note that $p \in U \cap V$, $U \cap V$ is an open set, and therefore $f+g$ and $fg \in C^\infty(p)$. We now come to the fundamental definition.

Definition 3.2.2. A *derivation* of $C^\infty(p)$ is a function $X_p : C^\infty(p) \to \mathbb{R}$ which satisfies for all $f, g \in C^\infty(p)$ and $a, b \in \mathbb{R}$,

$$\begin{array}{lll} & \text{linearity} & X_p(af+bg) = aX_p(f(\mathbf{x})) + X_p(g), \\ (3.6) & \text{Leibniz Rule} & X_p(fg) = X_p(f)g(p) + f(p)X_p(g)\ . \end{array}$$

Let $Der(C^\infty(p))$ denote the set of all derivations of $C^\infty(p)$.

If $X_p, Y_p \in Der(C^\infty(p))$ and $a \in I\!R$ we define $aX_p + Y_p \in Der(C^\infty(p))$ by

$$(aX_p + Y_p)(f) = aX_p(f) + Y_p(f), \quad for\ all f \in C^\infty(p). \tag{3.7}$$

Lemma 3.2.3. *With the operations 3.7 the set $Der(C^\infty(p))$ is a vector-space.*

Proof. The zero vector $\mathbf{0}_p$ is the derivation

$$\mathbf{0}_p(f) = 0.$$

The vector-space properties are now easy to verify. □

Example 3.2.4. Let $f \in C^\infty(p)$ and let $X_p(f) = \partial_{x^i} f|_p$, where $i \in \{1, \ldots, n\}$. Then $X_p \in Der(C^\infty(I\!R^n))$. More generally, if $(\xi^i)_{1 \leq i \leq n} \in I\!R^n$ then

$$X_p = (\xi^1 \partial_1 + \xi^2 \partial_2 + \ldots + \xi^n \partial_n)|_p$$

satisfies $X_p \in Der(C^\infty(p))$.

Example 3.2.5. Let $p \in I\!R^n$ and let $(\mathbf{a})_p \in V_p$. Define the function $T : V_p \to Der(C^\infty(p))$ by

$$T((\mathbf{a})_p)(f) = D_{\mathbf{a}} f(p). \tag{3.8}$$

The function T takes the vector $(\mathbf{a})_p$ to the corresponding directional derivative. We need to check that $T((\mathbf{a})_p)$ is in fact a derivation. This is clear though from property 4 in Lemma 3.2.1.

Lemma 3.2.6. *The function $T : V_p \to Der(C^\infty(p))$ is an injective linear transformation.*

Proof. The fact that T is linear is left as an exercise. To check that it is injective we use exercise 2 in Chapter 2. Suppose $T((\mathbf{a})_p) = \mathbf{0}_p$ is the zero derivation. Then

$$T((\mathbf{a})_p)(x^i) = D_{\mathbf{a}} x^i(p) = a^i = 0.$$

Therefore $(\mathbf{a})_p = (\mathbf{0})_p$ and T is injective. □

We now turn towards proving the important property that $Der(C^\infty(p))$ is an n-dimensional vector-space. In order to prove this, some basic properties of derivations are needed.

Lemma 3.2.7. *Let $X_p \in Der(C^\infty(p))$ then*

1. *for any open set U containing p, $X_p(1_U) = 0$ where 1_U is defined in equation 1.1, and*

2. *$X_p(c) = 0$, for all $c \in \mathbb{R}$*

Proof. We prove 1 by using the Leibniz property from 3.6,

$$X_p(1_U) = X_p(1_U 1_U) = X_p(1_U)1 + X_p(1_U)1 = 2X_p(1_U).$$

Therefore $X_p(1_U) = 0$. To prove 2, use linearity from equation 3.6 and part 1,

$$X_p(c) = X_p(c \cdot 1) = cX_p(1) = 0 \quad for\ all\ , c \in \mathbb{R}.$$

□

Corollary 3.2.8. *If $f \in C^\infty(p)$ and $U \subset \text{Dom}(f)$ then $X_p(f) = X_p(1_U \cdot f)$.*

Corollary 3.2.9. *If $f, g \in C^\infty(p)$ and there exists an open set $V \subset \mathbb{R}^n$ with $p \in V$, and $f(\mathbf{x}) = g(\mathbf{x}),\ for\ all\ \mathbf{x} \in V$, then $X_p(f) = X_p(g)$.*

Proof. Note that $1_V \cdot f = 1_V \cdot g$. The result is then immediate from the previous corollary. □

The main theorem is the following.

Theorem 3.2.10. *Let $X_p \in Der(C^\infty(p))$, then there exists $\xi^i \in \mathbb{R}, 1 \leq i \leq n$ such that*

$$X_p(f) = \sum_{1 \leq i \leq n} \xi^i \left. \frac{\partial f}{\partial x^i} \right|_p . \tag{3.9}$$

The real numbers ξ^i are determined by evaluating the derivation X_p on the coordinate functions x^i,

$$\xi^i = X_p(x^i),$$

and

$$X_p = \sum_{i=1}^n \xi^i \partial_{x^i}|_p. \tag{3.10}$$

Proof. Let $f \in C^\infty(p)$, and let $U = \text{Dom}(f)$ be open, and $p = (x_0^1, \ldots, x_0^n)$. By Lemma 1.4.8 there exists functions $g_i \in C^\infty(p), 1 \leq i \leq n$, defined on an open ball $B_r(p) \subset U$ such that the function $F : B_r(p) \to I\!R$ given by

$$F(\mathbf{x}) = f(p) + \sum_{i=1}^{n} (x^i - x_0^i) g_i(\mathbf{x}), \quad \mathbf{x} \in B_r(p), \tag{3.11}$$

where

$$g_i(p) = \left. \frac{\partial f}{\partial x^i} \right|_p ,$$

agrees with $f(\mathbf{x})$ on $B_r(p)$. Since f and F agree on $B_r(p)$, Corollary 3.2.9 implies

$$X_p(f) = X_p(F).$$

Finally by using the properties for X_p of linearity, Leibniz rule and $X_p(f(p)) = 0$ (Lemma 3.2.7) in equation 3.11 we have

$$\begin{aligned} X_p(f) = X_p(F) &= X_p\left(\sum_{i=1}^{n} (x^i - x_0^i) g_i(\mathbf{x}) \right) \\ &= \sum_{i=1}^{n} X_p(x^i - x_0^i) g_i(p) + (x^i - x_0^i)|_{\mathbf{x}=p} X_p(g^i) \qquad (3.12) \\ &= \sum_{i=1}^{n} X_p(x^i) g_i(p). \end{aligned}$$

Property 2 in Lemma 1.4.8 gives $g_i(p) = (\partial_{x^i} f)|_p$ which in equation 3.12 gives equation 3.9 and the theorem is proved. □

Corollary 3.2.11. *The set* $\beta = \{\partial_{x^i}|_p\}_{1 \leq i \leq n}$ *forms a basis for* $Der(C^\infty(p))$.

Corollary 3.2.11 leads to the modern definition of the tangent space.

Definition 3.2.12. Let $p \in I\!R^n$. The tangent space at p denoted by $T_p I\!R^n$ is the n-dimensional vector-space $Der(C^\infty(p)$.

The linear transformation $T : V_p \to T_p I\!R^n$ given in equation 3.8 is then an isomorphism on account of Lemma 3.2.6 and the dimension theorem (2.2.7). If $\sigma : I \to I\!R^n$ is a curve with tangent vector $(\dot{\sigma}(t_0))_{\sigma(t_0)}$ then the corresponding derivation $X_p = T((\dot{\sigma}(t_0))_{\sigma(t_0)}$ is

$$X_p = \sum_{i=1}^{n} \left. \frac{d\sigma^i}{dt} \right|_{t=t_0} \left. \frac{\partial}{\partial x^i} \right|_{p=\sigma(t_0)} . \tag{3.13}$$

Example 3.2.13. Let $p = (1, 1, 1) \in I\!R^3$ and let $(1, -2, 3)_p \in V_p$. The curve

$$\sigma(t) = (1 + t, e^{-2t}, 1 + 3t)$$

is a representative curve for $(1, -2, 3)_p$. The corresponding tangent vector $X_p \in T_p I\!R^3$ is

$$X_p = (\partial_x - 2\partial_y + 3\partial_z)_p .$$

Example 3.2.14. Let $p = (1, 2) \in I\!R^2$. Find $X_p \in Der(C^\infty(p)) = T_p I\!R^2$ in the coordinate basis where $X_p(x^2 + y^2) = 3$ and $X_p(xy) = 1$. We begin by writing $X_p = a\partial_x + b\partial_y,\ a, b \in I\!R$, then note by Theorem 3.2.10 above, that $a = X_p(x)$ and $b = X_p(y)$. Applying the two rules of derivations in definition 3.2.2 we get

$$\begin{aligned} X_p(x^2 + y^2) &= (2x)|_{(1,2)} X_p(x) + (2y)|_{(1,2)} X_p(y) \\ &= 2X_p(x) + 4X_p(y) \end{aligned}$$

and

$$\begin{aligned} X_p(xy) &= = (y)|_{(1,2)} X_p(y) + (x)|_{(1,2)} X_p(y) \\ &= 2X_p(x) + X_p(y) \end{aligned}$$

Using $X_p(x^2 + y^2) = 3, X_p(xy) = 1$ this gives the system of equations for $a = X_p(x)$ and $b = X_p(y)$ to be

$$\begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

The solution to which is

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{6} \\ \frac{2}{3} \end{pmatrix} .$$

Therefore

$$X_p = \left(\frac{1}{6}\partial_x + \frac{2}{3}\partial_y \right)_p$$

Example 3.2.15. Generalizing the previous example, suppose we want to find $X_p \in T_p I\!R^n = Der(C^\infty(I\!R^n))$, where $f^1, f^2, \ldots, f^n \in C^\infty(p)$ are given together with

$$X_p(f^1) = c^1,\ X_p(f^2) = c^2, \ldots,\ X_p(f^n) = c^n. \tag{3.14}$$

Under what conditions on $f^i, 1 \leq i \leq n$ do these equations completely determine X_p? By Theorem 3.2.10 or Corollary 3.2.11 above we know that $X_p = \sum_{i=1}^n \xi^i \partial_{x^i}|_p$. Applying this to (3.14) we find

$$\begin{aligned} \sum_{i=1}^n \xi^i \left.\frac{\partial f^1}{\partial x^i}\right|_p &= c^1 \\ \sum_{i=1}^n \xi^i \left.\frac{\partial f^2}{\partial x^i}\right|_p &= c^2 \\ &\vdots \\ \sum_{i=1}^n \xi^i \left.\frac{\partial f^n}{\partial x^i}\right|_p &= c^n \end{aligned} \tag{3.15}$$

This system of equations can be written as a matrix/vector equation

$$J\boldsymbol{\xi} = \mathbf{c}$$

where $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^n)$, $\mathbf{c} = (c^1, \ldots, c^n)$ and J is the $n \times n$ Jacobian matrix

$$J^j_i = \left.\frac{\partial f^j}{\partial x^i}\right|_p,$$

and $J\boldsymbol{\xi}$ is standard matrix vector multiplication. Equation (3.2.15) has a unique solution if and only if J is invertible, in other words if and only if $\det J \neq 0$.

A set of n-functions $f^1, f^2, \ldots, f^n \in C^\infty(p)$ which satisfy

$$\det\left(\left.\frac{\partial f^j}{\partial x^i}\right|_p\right) \neq 0,$$

are said to be functionally independent at p. The term functionally independent will be discussed in more detail in Section 8.3.

3.3 Vector fields

A vector field X on $I\!R^n$ is a function that assigns to each point $p \in I\!R^n$ a tangent vector at that point. That is $X(p) \in T_p I\!R^n$. Therefore there exists functions $\xi^i(\mathbf{x})$, $1 \leq i \leq n$ on $I\!R^n$ such that,

$$X = \sum_{i=1}^{n} \xi^i(\mathbf{x}) \partial_{x^i}|_{\mathbf{x}}, \quad \mathbf{x} \in I\!R^n.$$

We will write this as $X = \sum_{i=1}^{n} \xi^i(\mathbf{x}) \partial_{x^i}$ (dropping the $|_{\mathbf{x}}$) since the point at which ∂_{x^i} is evaluated can be inferred from the coefficients. The vector field X is smooth or a C^∞ vector field if the coefficients $\xi^i(\mathbf{x})$ are smooth functions. Vector fields will play a prominent role in the rest of the book.

Example 3.3.1. The vector field X on $I\!R^3$ given by

$$X = xy^2 \partial_x + xz \partial_y + e^{y+z} \partial_z$$

is smooth.

Vector fields have algebraic properties that are similar to tangent vectors. Let $U \subset I\!R^n$ be an open set. A function $X : C^\infty(U) \to C^\infty(U)$ is called a derivation of $C^\infty(U)$ if for all $f, g \in C^\infty(U)$ and $a, b \in I\!R$,

$$\text{(3.16)} \qquad \begin{array}{ll} \text{linearity} & X(af + bg) = aX(f) + X(g), \\ \text{Leibniz rule} & X(fg) = X(f)g + fX(g). \end{array}$$

We let $Der(C^\infty(U))$ be the set of all derivations of $C^\infty(U)$.

Example 3.3.2. Consider ∂_{x^i} where $i \in \{1, \ldots, n\}$, and $C^\infty(I\!R^n)$. The partial derivatives ∂_{x^i} satisfy properties one and two in equation 3.16, and so $\partial_{x^i} \in Der(C^\infty(I\!R^n))$. More generally, let

$$X = \xi^1(\mathbf{x}) \partial_{x^1} + \xi^2(\mathbf{x}) \partial_{x^2} + \ldots + \xi^n(\mathbf{x}) \partial_{x^n}$$

where $\xi^1(\mathbf{x}), \ldots, \xi^n(\mathbf{x}) \in C^\infty(U)$. The first order differential operator X is a derivation of $C^\infty(U)$. In $I\!R^3$ the differential operator

$$X = yz \partial_x + x(y+z) \partial_z$$

is a derivation of $C^\infty(I\!R^3)$, and if $f = xe^{yz}$, then

$$X(f) = (yz \partial_x + x(y+z) \partial_z)(xe^{yz}) = yze^{yz} + xy(y+z)e^{yz}.$$

Example 3.3.3. Let $X \in Der(C^\infty(\mathbb{R}))$, and $n \in Z^+$ then

$$X(x^n) = nx^{n-1}X(x)$$

where $x^n = x \cdot x \cdot x \ldots \cdot x$. This is immediately true for $n = 1$. While by the Leibniz rule,

$$X(xx^{n-1}) = X(x)x^{n-1} + xX(x^{n-1}).$$

It then follows by induction that

$$X(x^n) = x^{n-1}X(x) + x(n-1)x^{n-2}X(x) = nx^{n-1}X(x).$$

3.4 Exercises

1. Let $(\dot\sigma(t_0))_{\sigma(t_0)}$ be the tangent vector to the curve at the indicated point as computed in section 3.1. Give two other representative curves for each resulting tangent vector (different than in equations 3.2 and 3.3), and give the corresponding derivation in the coordinate basis (see example 3.2.13).

   (a) $x = t^3 + 2$, $y = t^2 - 2$ at $t = 1$,

   (b) $r = e^t$, $\theta = t$ at $t = 0$,

   (c) $x = \cos(t)$, $y = \sin(t)$, $z = 2t$, at $t = \pi/2$.

2. Let X_p be a tangent vector at a point p in $\mathbb{R}^n$. Use the properties of a derivation to prove that $X(\frac{1}{f}) = -\frac{X(f)}{f(p)^2}$, $f \in C^\infty(p), f(p) \neq 0$. Hint: Write $1 = f \cdot (\frac{1}{f})$.

3. Let $p = (3, 2, 2) \in \mathbb{R}^3$ and suppose $X_p \in T_p\mathbb{R}^3$ is a tangent vector satisfying $X_p(x) = 1$, $X_p(y) = 1$, $X_p(z) = 2$.

   (a) Calculate $X_p(x^2 + y^2)$ and $X_p(z/y)$ using just the properties of a derivation.

   (b) Calculate $X_p(f)$ where $f(x, y, z) = \sqrt{x^2 - zy - 1}$ using just the properties of a derivation. Hint: Find a formula for $X_p(f^2)$.

   (c) Find the formula for X_p as a derivation in the coordinate basis.

4. With $p = (3, -4) \in \mathbb{R}^2$ find $X_p \in T_p\mathbb{R}^2$ (with (x, y) coordinates) such that $X_p(x + xy) = 4$ and $X_p(\sqrt{x^2 + y^2}) = \frac{1}{5}$.

5. With $p = (1, -1) \in \mathbb{R}^2$ find all $X_p \in T_p\mathbb{R}^2$ (with (x, y) coordinates) such that $X_p(x^2 + y^2) = 2$.

6. Prove Corollary 3.2.11.

7. Let $f_1 = x + y + z$, $f_2 = x^2 + y^2 + z^2$, $f_3 = x^3 + y^3 + z^3$ be functions on $\mathbb{R}^3$.

(a) Show f_1, f_2, f_3 are functionally independent at all point (a, b, c) such that $(a-b)(b-c)(c-a) \neq 0$.

(b) Find the derivation X_p at the point $p = (0, -1, 1)$ such that $X_p(f_1) = 2$, $X_p(f_2) = 3$, $X_p(f_3) = -4$.

8. Given the vector fields X, Y on $\mathbb{R}^3$,

$$X = x\partial_x - 3y\partial_y + z\partial_z, \quad Y = \partial_x + x\partial_y - \partial_z$$

and the functions

$$f = x^2yz, \quad g = y^2 - x + z^2,$$

compute

(a) $X(f)$,$X(g)$, $Y(f)$, and $Y(g)$,

(b) $Y(X(f))$, $X(Y(f)$, and

(c) $X(Y(g)) - Y(X(g))$.

9. Let X, Y be any two vector fields on $\mathbb{R}^n$ and define $X \circ Y : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ by

$$X \circ Y(f) = X(Y(f)).$$

Does $X \circ Y$ define a derivation?

10. Another common way to define tangent vectors uses the notion of *germs*. On the set $C^\infty(p)$ define two functions $f, g \in C^\infty(p)$ to be equivalent if there exists an open set V with $p \in V$, and $V \subset \mathrm{Dom}(f), V \subset \mathrm{Dom}(g)$ such that $f(\mathbf{x}) = g(\mathbf{x}),\ for\ all \mathbf{x} \in V$. Show that this is an equivalence relation on $C^\infty(p)$. The set of equivalence classes are called germs of C^∞ functions at p. The tangent space $T_p\mathbb{R}^n$ is then defined to be the vector-space of derivations of the germs of C^∞ functions at p.

Chapter 4

The push-forward and the Jacobian

Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ be a C^∞ function and let $p \in \mathbb{R}^n$. The goal of this section is to use the function Φ to define a function

$$\Phi_{*,p} : T_p\mathbb{R}^n \to T_{\Phi(p)}\mathbb{R}^m.$$

The map $\Phi_{*,p}$ is called the *push-forward*, *the differential*, or the *Jacobian* of Φ at the point p. In section 4.2 we give a second definition of $\Phi_{*,p}$ and show it agrees with the first. The first definition has a simple geometric interpretation, while the second definition is more convenient for studying the properties of $\Phi_{*,p}$.

4.1 The push-forward using curves

Before giving the definition we consider an example.

Example 4.1.1. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ be given as in 4.17 by

$$\Phi(x, y) = (u = x^2 + y^2, v = x^2 - y^2, w = xy). \tag{4.1}$$

Let $\sigma : \mathbb{R} \to \mathbb{R}^2$ be the curve

$$\sigma(t) = (1 + 3t, 2 - 2t),$$

which has the tangent vector at $t = 0$ given by

$$(\dot{\sigma}(0))_{\sigma(0)} = (3, -2)_{(1,2)}. \tag{4.2}$$

The image curve $\Phi \circ \sigma : I\!R \to I\!R^3$ is given by

$$\Phi(\sigma(t)) = \big(((1+3t)^2 + (2-2t)^2), ((1+3t)^2 - (2-2t)^2), (1+3t)(2-2t)\big).$$

The tangent vector to the image curve $\Phi \circ \sigma$ when $t = 0$ is then

$$\left(\frac{d}{dt}\Phi(\sigma(t)\Big|_{t=0} \right)_{\Phi(\sigma(0))} = (-2, 14, 4)_{(5,-3,2)}. \tag{4.3}$$

The map $\Phi_{*,p}$ we define below has the property that if $(\dot\sigma(t_0))_{\sigma(t_0)}$ is the tangent vector to the curve σ at $p = \sigma(t_0)$ with corresponding derivation X_p as in equation 3.13, then $Y_q = \Phi_{*,p}X_p$ is the derivation Y_q corresponding to the tangent vector $\Phi \circ \sigma$ at the image point $\Phi(p) = \Phi(\sigma(t_0))$! In example 4.1.1 this means with $p = (1,2)$ and $X_p = (3\partial_x - 2\partial_y)_{(1,2)}$ from 4.2, and $Y_q = (-2\partial_u + 14\partial_v + 4\partial_w)_{(5,-3,2)}$ from 4.3, that $Y_q = \Phi_{*,p}X_p$.

Let $X_p \in T_pI\!R^n$ and let $\sigma : I \to I\!R^n$ be a smooth curve which represents X_p by $\dot\sigma(t_0) = X_p$. By composing $\Phi : I\!R^n \to I\!R^m$ with σ we get the image curve $\Phi \circ \sigma : I \to I\!R^m$ from which we prove

Definition 4.1.2. The pushforward of the tangent vector to σ at $p = \sigma(t_0)$ is the tangent vector of the image curve $\Phi(\sigma)$ at $q = \Phi(\sigma(t_0))$. That is,

$$\Psi_{*,p}\dot\sigma(t_0) = \left(\frac{d}{dt}\Phi \circ \sigma \right)\Big|_{t=0} \tag{4.4}$$

XXXXXXXXXXXXXXXXX Diagram XXXXXXXXXXXXXXXXX

We now derive a formula for $\Phi_{*,p}X_p$ for the tangent vector $X_p = \dot\sigma(t_0)$ where $p = \sigma(t_0)$, and in coordinates

$$X_p = \sum_{i=1}^{n} \frac{d\sigma^i}{dt}\Big|_{t=t_0} \frac{\partial}{\partial x^i}\Big|_p . \tag{4.5}$$

For the curve $\Phi(\sigma(t))$, let Y_q be the tangent vector at $q = \Phi(\sigma(t_0))$ which is

$$Y_q = \sum_{i=1}^{n} \frac{d\Phi^i(\sigma(t))}{dt}\Big|_{t=t_0} \frac{\partial}{\partial y^i}\Big|_q = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial \Phi^i}{\partial x^j}\Big|_p \frac{d\sigma^j)}{dt}\Big|_{t=t_0} \right) \frac{\partial}{\partial y^i}\Big|. \tag{4.6}$$

where $(y^i)_{1 \le i \le n}$ denotes the coordinates on the image space of Φ.

4.2 The push-forward using derivations

Let $X_p \in T_p I\!R^n$ be a tangent vector which by definition is completely determined by its action on $C^\infty(p)$. In order that $\Phi_{*,p}(X_p)$ to be a tangent vector at $q = \Phi(p) \in I\!R^m$ we need to define

$$\Phi_{*,p}(X_p)(g) \quad for\ all\ g \in C^\infty(q),$$

and show that the result is a derivation (see 3.2.2). Before giving the definition we make the following simple observation. Let $g \in C^\infty(q)$ which is a function on the image space of Φ. The composition $g \circ \Phi : I\!R^n \to I\!R$, is a smooth function on an open subset of $I\!R^n$ **which contains** p, and so $X_p(g \circ \Phi)$ is well defined! Using this calculation, we are now ready to define $\Phi_{*,p}(X_p)$.

Theorem 4.2.1. *Let $\Phi : I\!R^n \to I\!R^m$ be a smooth function, let $p \in I\!R^n$ and $q = \Phi(p)$. Given $X_p \in T_p I\!R^n$ define $\Phi_{*,p}(X_p) : C^\infty(q) \to I\!R$ by*

$$\Phi_{*,p}(X_p)(g) = X_p(g \circ \Phi) \quad for\ all\ g \in C^\infty(q). \tag{4.7}$$

Then $\Phi_{*,p}(X_p) \in T_q I\!R^n$.

Proof. The function $g \in C^\infty(q)$ in 4.7 is arbitrary and so $\Phi_{*,p}(X_p) : C^\infty(q) \to I\!R$. Denote this function by $Y_q = \Phi_{*,p}(X_p)$. If we check that Y_q is a derivation of $C^\infty(q)$, then by definition 3.2.2 $Y_q \in T_q I\!R^m$. This is easy and demonstrates the power of using derivations.

Let $g, h \in C^\infty(q)$ $a, b \in I\!R$, then we compute using the fact that X_q is a derivation,

$$\begin{aligned} Y_q(ag + bh) &= X_q(ag \circ \Phi + bh \circ \Phi) \\ &= aX_p(g \circ \Phi) + bX_p(h \circ \Phi) \\ &= aY_q(g) + bY_q(h) \end{aligned}$$

and using $\cdot$ for multiplication of functions we have,

$$\begin{aligned} Y_q(g \cdot h) &= X_p(g \circ \Phi \cdot h \circ \Phi) \\ &= X_p(g \circ \Phi) \cdot h \circ \Phi(p) + g \circ \Phi(p) \cdot X_p(h \circ \Phi) \\ &= Y_q(g) \cdot h(q) + g(q) \cdot Y_q(h). \end{aligned}$$

Therefore Y_q in (4.7) is derivation of $C^\infty(q)$ and so is an element of $T_q I\!R^m$. □

We now check that $\Phi_{*,p}$ is a linear transformation.

Theorem 4.2.2. *The function $\Phi_{*,p} : T_p\mathbb{R}^n \to T_{\Phi(p)}\mathbb{R}^m$ is a linear transformation.*

Proof. Let $X_p, Y_p \in T_p\mathbb{R}^n$ and $a \in \mathbb{R}$. We use equation 4.7 to compute

$$\begin{aligned}\Phi_{*,p}(aX_p + Y_p)(g) &= (aX_p + Y_p)(g \circ \Phi)\\ &= aX_p(g \circ \Phi) + Y_p(g \circ \Phi)\\ &= a\Phi_{*,p}X_p(g) + \Phi_{*,p}Y_p(g).\end{aligned}$$

Therefore $\Phi_{*,p}(aX_p + Y_p) = a\Phi_{*,p}X_p + \Phi_{*,p}Y_p$. □

To gain a better understanding of definition (4.7) we now write out this equation in coordinates using the coordinate basis.

Proposition 4.2.3. *Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function and let $q = \Phi(p)$. Let $\beta = \{\partial_{x^i}|_p\}_{1\leq i\leq n}$ be the coordinate basis for $T_p\mathbb{R}^n$ and let $\gamma = \{\partial_{y^a}|_q\}_{1\leq a\leq m}$ be the coordinate basis for $T_q\mathbb{R}^m$. If $X_p \in T_p\mathbb{R}^n$ is given by*

$$X_p = \sum_{i=1}^n \xi^i\, \partial_{x^i}|_p, \quad \xi^i \in \mathbb{R}, \tag{4.8}$$

then

$$\Phi_{*,p}X_p = \sum_{a=1}^m \left(\sum_{i=1}^n \xi^i \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \right) \left.\frac{\partial}{\partial y^a}\right|_q \tag{4.9}$$

Proof. By definition we have

$$(\Phi_{*,p}X_p)(g) = \left(\sum_{i=1}^n \xi^i \partial x^i|_p \right)(g \circ \Phi) \tag{4.10}$$

If we expand out the derivative term in 4.10 using the chain rule we get,

$$\left.\frac{\partial g \circ \Phi}{\partial x^i}\right|_p = \sum_{a=1}^m \left.\frac{\partial g}{\partial y^a}\right|_q \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p .$$

Using this in (4.10), we get

$$\begin{aligned}X_p(g \circ \Phi) &= \sum_{i=1}^n \xi^i \left(\sum_{a=1}^m \left.\frac{\partial g}{\partial y^a}\right|_q \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \right)\\ &= \sum_{a=1}^m \left(\sum_{i=1}^n \xi^i \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \right) \left.\frac{\partial g}{\partial y^a}\right|_q\end{aligned} \tag{4.11}$$

where in the second line we have switched the order of the summations. By taking the coefficients of $\partial_{y^a}|_q$ equation 4.9 now follows directly from 4.11. □

The following corollary gives us the important interpretation of the function $\Phi_{*,p}$.

Corollary 4.2.4. *The matrix representation of the linear transformation* $\Phi_{*,p} : T_p\mathbb{R}^n \to T_q\mathbb{R}^m$ *in the basis* $\beta = \{\partial_{x^i}|_p\}_{1\leq i\leq n}$ *for* $T_p\mathbb{R}^n$ *and the basis* $\gamma = \{\partial_{y^a}|_q\}_{1\leq a\leq m}$ *for* $T_{\Phi(p)}\mathbb{R}^m$ *is given by the Jacobian*

$$[\Phi_{*,p}] = \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \tag{4.12}$$

Proof. Suppose the matrix representation of $\Phi_{*,p}$ is given by the matrix $(J^a_i) \in M_{m\times n}(\mathbb{R})$ (see equation 2.3)

$$\Phi_{*,p}(\partial_{x^i}|_p) = \sum_{a=1}^{m} J^a_i \partial_{y^a}|_q. \tag{4.13}$$

This entries in the matrix J^a_i are easily determined using equation 4.9 which gives

$$\Phi_{*,p}(\partial_{x^i}|_p) = \sum_{a=1}^{m} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \partial_{y^a}|_q, \tag{4.14}$$

therefore

$$[\Phi_{*,p}] = \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p . \tag{4.15}$$

□

A particulary useful way to compute $\Phi_{*,p}$ is the next corollary.

Corollary 4.2.5. *The coefficients of the image vector* $Y_q = \Phi_{*,p}X_p$ *in the coordinate basis* $\gamma = \{\partial_{y^a}|_q\}_{1\leq a\leq m}$ *are given by*

$$\eta^a = \Phi_{*,p}X_p(y^a) = \sum_{i=1}^{n} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \xi^i. \tag{4.16}$$

Example 4.2.6. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$\Phi(x, y) = (u = x^2 + y^2, v = x^2 - y^2, w = xy). \tag{4.17}$$

Let$p = (1, 2)$ and $X_p = (3\partial_x - 2\partial_y)|_p$. Compute $\Phi_{*,p}X_p$ first using derivation approach 4.7, and then the Jacobian (4.15.

With $q = \Phi(p) = (5, -3, 2)$ we now find the coefficients of $Y_q = (a\partial_u + b\partial_v + c\partial_w)|_q = \Phi_{*,p}X_p$ by using the definition 4.7 in the form of equation 4.16,

$$\begin{aligned}
a = \quad & \Phi_{*,p}(X_p)(u) = & (3\partial_x - 2\partial_y)_{(1,2)}(x^2 + y^2) = \quad & 6 - 8 = -2 \\
b = \quad & \Phi_{*,p}(X_p)(v) = & X_{(1,2)}(x^2 - y^2) = \quad & 6 + 8 = 14 \\
c = \quad & \Phi_{*,p}(X_p)(w) = & X_{(1,2)}(xy) = \quad & 6 - 2 = 4.
\end{aligned}$$

This gives

$$Y_q = (-2\partial_u + 14\partial_v + 4\partial_w)|_q. \tag{4.18}$$

We now compute $\Phi_{*,p}X_p$ using the Jacobian matrix at $(1, 2)$, which is

$$\left[\frac{\partial \Phi^a}{\partial x^i}\right]_{(1,2)} = \begin{bmatrix} 2x & 2y \\ 2x & -2y \\ y & x \end{bmatrix}_{(1,2)} = \begin{bmatrix} 2 & 4 \\ 2 & -4 \\ 2 & 1 \end{bmatrix},$$

and the coefficients of $\Phi_{*,p}(X_p)$ are

$$\begin{bmatrix} 2 & 4 \\ 2 & -4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 14 \\ 4 \end{bmatrix}.$$

This gives the same coefficients for $\Phi_{*,p}X_p$ in the coordinate basis $\{\partial_u|_q, \partial_v|_q, \partial_w|_q\}$ as in equation (4.18).

Remark 4.2.7. Note that the definition of $\Phi_{*,p}$ in 4.7 was given independent of coordinates! Then the coordinate dependent formulas 4.16 and 4.12 for $\Phi_{*,p}$ were derived from its definition. This is what we strive for when giving definitions in differential geometry.

4.3 The Chain rule, Immersions, Submersions, and Diffeomorphisms

Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ and let

$$T\mathbb{R}^n = \bigcup_{p\in\mathbb{R}^n} T_p\mathbb{R}^n. \tag{4.19}$$

The set $T\mathbb{R}^n$ consists of all possible tangent vectors at every possible point and an element of $T\mathbb{R}^n$ is just a tangent vector X_p at a particular point $p \in \mathbb{R}^n$. We now define the map $\Phi_* : T\mathbb{R}^n \to T\mathbb{R}^m$ by the point-wise formula

$$\Phi_*(X_p) = \Phi_{*,p}X_p. \tag{4.20}$$

At a generic point $p = (x^1, \ldots, x^n) \in \mathbb{R}^n$ with standard basis $\{\partial_{x^i}|_p\}_{1\le i\le n}$ for $T_p\mathbb{R}^n$ and basis $\{\partial_{y^a}|_{\Phi(p)}\}_{1\le a\le m}$ for $T_{\Phi(p)}\mathbb{R}^m$ the matrix representation of Φ_* is again computed by using 4.2.4 to give the $m \times n$ functions on $\mathbb{R}^n$

$$[\Phi_*] = \left[\frac{\partial \Phi^a}{\partial x^i}\right]. \tag{4.21}$$

Example 4.3.1. Let $\alpha : \mathbb{R} \to \mathbb{R}^n$ be a smooth curve. Then

$$\alpha_*\partial_t = \sum_{i=1}^n \frac{d\alpha^i}{dt}\, \frac{\partial}{\partial x^i}\bigg|_{\alpha(t)} \quad for\ all\ t \in \mathbb{R}.$$

This formula agrees with the point-wise given formula in equation 3.13.

We now prove the chain-rule using the derivation definition of Φ_*.

Theorem 4.3.2. *(The chain rule) Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ and let $\Psi : \mathbb{R}^m \to \mathbb{R}^p$. Then*

$$(\Psi \circ \Phi)_* = (\Psi_*) \circ (\Phi_*) \tag{4.22}$$

Proof. It suffices to check this point-wise on account of 4.20. Let $p \in \mathbb{R}^n$, $r = \Psi \circ \Phi(p)$, and let $g \in C^\infty(r)$. Let $X_p \in T_p\mathbb{R}^n$ then by definition 4.7

$$(\Psi \circ \Phi)_{*,p}X_p(g) = X_p\left(g \circ \Psi \circ \Phi\right).$$

On the right side we get by definition 4.7

$$\begin{aligned}((\Psi_*)\circ(\Phi_*)X_p)\,(g) &= (\Phi_* X_p)\,(g\circ\Phi)\\ &= X_p(g\circ\Phi\circ\Psi).\end{aligned}$$

We have shown $(\Psi\circ\Phi)_* X_p(g) = ((\Psi_*)\circ(\Phi_*)X_p)\,(g)$ for all $g\in C^\infty(r)$. Therefore $(\Psi\circ\Phi)_* X_p = ((\Psi_*)\circ(\Phi_*))\,X_p$ for all $X_p\in T\mathbb{R}^n$ and the theorem is proved. □

Recall in section 3.1 that the rank of a linear transformation T is the dimension of $R(T)$, the range space of T.

Let $\Phi : U\to V$, where $U\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^m$ be a C^∞ function.

Definition 4.3.3. The function Φ is an *immersion* at $p\in U$ if the rank of $\Phi_{*,p} : T_pU\to T_{\Phi(p)}V$ is n. The function Φ is an immersion if it is an immersion at each point in $p\in U$.

By fact that $\dim T_pU$ is n-dimensional, the dimension Theorem 2.2.7 shows that the map Φ_* is an immersion if it is injective at each point $p\in U$. This also implies by exercise 2 in the section 3 that the kernel $\Phi_{*,p}$ at each point $p\in U$ consists of only the zero tangent vector at p.

Example 4.3.4. Let $\Phi:\mathbb{R}^2\mathbb{R}^3$ be given by

$$\Phi(x,y) = (u = x^2+y^2, v = x^2-y^2, w = xy). \tag{4.23}$$

At a generic point $p=(x,y)\in\mathbb{R}^2$ with standard basis $\{\partial_x|_p,\partial_y|_p\}$ for $T_p\mathbb{R}^2$ and basis $\{\partial_u|_{\Phi(p)},\partial_v|_{\Phi(p)},\partial_w|_{\Phi(p)}\}$ for $T_{\Phi(p)}\mathbb{R}^3$ the matrix representation of Φ_* is (see equation 4.21)

$$[\Phi_*] = \begin{bmatrix} 2x & 2y\\ 2x & -2y\\ y & x\end{bmatrix}.$$

According to Lemma 2.2.4 (see also exercise 2), Φ_* is injective if and only if $\ker[\Phi_*]$ is the zero vector in $\mathbb{R}^2$. This fails to happen only when $x=y=0$. Therefore Φ is an immersion at every point except the origin in $\mathbb{R}^2$.

Definition 4.3.5. The function Φ is a *submersion* at $p\in U$ if the rank of $\Phi_{*,p} : T_pU\to T_{\Phi(p)}V$ is m. The function Φ is an submersion if it is a submersion at each point in $p\in U$.

In other words, Φ is a submersion if $\Phi_{*,p}$ is surjective for each point $p \in I\!R^n$.

For the next definitions we consider the case $m = n$.

Definition 4.3.6. A smooth function $\Phi : U \to V$, $U, V \subset I\!R^n$ which is invertible, with inverse function $\Phi^{-1} : V \to U$, which is also smooth, is a *diffeomorphism*

A less restrictive notion is a *local diffeomorphism.*

Definition 4.3.7. A smooth function $\Phi : U \to V$ is a local diffeomorphism at $p \in U$ if there exists an open set W with $W \subset U$ and $p \in W$, such that $\Phi : W \to \Phi(W)$ is a diffeomorphism.

Theorem 4.3.8. *Inverse Function Theorem. Let $\Phi : U \to V$ be C^∞. The function Φ is a local diffeomorphism at $p \in U$ if and only if $\Phi_{*,p} : T_pU \to T_{\Phi(p)}V$ is an isomorphism.*

Proof. The proof that this condition is necessary is simple. If $W \subset U$ is an open set on which Φ restricts to a diffeomorphism, let $Z = \Phi(W)$, and $\Psi : Z \to W$ be the inverse of Φ. By the chain-rule

$$(\Psi \circ \Phi)_{*,p} = \Psi_{*,\Phi(p)} \Phi_{*,p} = I$$

and so $\Phi_{*,p}$ is an isomorphism at $p \in W$.

The proof of converse is given by the inverse function theorem and can be found in many texts [12]. □

Example 4.3.9. Let $f : I\!R \to I\!R$ be

$$f(t) = t^3$$

then

$$f^{-1}(t) = t^{\frac{1}{3}}.$$

We have for $W = I\!R - 0$, the function $f : W \to I\!R$ is a local diffeomorphism.

4.4 Change of Variables

Let $U, V \subset \mathbb{R}^n$ be open sets. A change of coordinates is a diffeomorphism $\Phi : U \to V$. If $x^1, \ldots, x^n$ are coordinates on U and $y^1, \ldots, y^n$ are coordinates on V, then the map Φ is given by

$$y^i = \Phi^i(x^1, \ldots, x^n), \quad 1 \leq i \leq n.$$

If $p \in U$ and it has x-coordinates $(x_0^1, \ldots, x_0^n)$, then $y_0^i = \Phi^i(x_0^1, \ldots, x_0^n)$ are the y-coordinates of $\Phi(p)$.

Since Φ is a diffeomorphism, then $\Phi^{-1} : V \to U$ is a diffeomorphism.

Example 4.4.1. Let $V = \mathbb{R}^2 - \{(x, 0), \quad x \geq 0\}$ and let $U = \mathbb{R}^+ \times (0, 2\pi)$ then

$$\Phi(r, \theta) = (x = r\cos\theta, y = r\sin\theta) \tag{4.24}$$

is a change of coordinates.

Example 4.4.2. Let $V = \mathbb{R}^3 - \{(x, 0, z), \quad x \geq 0\}$ and let $U = \mathbb{R}^+ \times (0, 2\pi) \times (0, \pi)$ then

$$\Phi(\rho, \theta, \phi) = (x = \rho\cos\theta\sin\phi, y = \rho\sin\theta\sin\phi, \rho = \cos\phi)$$

is a change of coordinates.

We now look at how vector-fields behave under a change of coordinates. Let $U, V \subset \mathbb{R}^n$ and $\Phi : U \to V$ be a change of coordinates, and let X be a vector-field on U. Since for each point $p \in U$, X_p is a tangent vector, we can map this tangent vector to the image using $\Phi_{*,p}$ as in sections 4.1, 4.2

$$\Phi_{*,p}X(p) \in T_{\Phi(p)}V.$$

If we do this for every point $p \in U$, and use the fact that Φ is a diffeomorphism (so one-to-one and onto) we find that $\Phi_{*,p}$ applied to X defines a tangent vector at each point $q \in V$, and therefore a vector-field. To see how this is defined, let $q \in V$, and let $p \in U$ be the unique point in U such that $p = \Phi^{-1}(q)$. We then define the vector-field Y point-wise on V by

$$Y_q = \Phi_{*,p}(X_p).$$

Let's give a coordinate version of this formula. Suppose $(x^i)_{1\leq i\leq n}$ are coordinates on U, and $(y^i)_{1\leq i\leq n}$ are coordinates on V. We have a vector-field

$$X = \sum_{i=1}^{n} \xi^i(\mathbf{x}) \frac{\partial}{\partial x^i},$$

and want to find the vector-field

$$Y = \sum_{i=1}^{n} \eta^i(\mathbf{y}) \frac{\partial}{\partial y^i}.$$

Let $p \in U$, and $q = \Phi(p)$ the formula we have is $Y_q = \left(\Phi_{*,p} X_p\right)_{p=\Phi^{-1}(q)}$ and by equation 4.9 is

$$Y_q = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \xi^i(p) \left.\frac{\partial \Phi^j}{\partial x^i}\right|_p \right)\Bigg|_{p=\Phi^{-1}(q)} \left.\frac{\partial}{\partial y^j}\right|_q . \tag{4.25}$$

The coefficients of the vector-field Y are then

$$\eta^j(q) = \left(\sum_{i=1}^{n} \xi^i(p) \left.\frac{\partial \Phi^j}{\partial x^i}\right|_p \right)\Bigg|_{p=\Phi^{-1}(q)} , \tag{4.26}$$

or by equation 4.16

$$\eta^j(q) = Y(y^j)||_{p=\Phi^{-1}(q)} = X(\Phi^j)||_{p=\Phi^{-1}(q)}, \tag{4.27}$$

The formulas 4.26 and 4.27 for Y is called the change of variables formula for a vector-field. The vector-field Y is also called the push-forward by Φ of the vector-field X and is written $Y = \Phi_* X$.

Example 4.4.3. The change to polar coordinates is given in 4.24 in example 4.4.1 above. Let

$$X = r\partial_r.$$

We find this vector-field in rectangular coordinates, by computing $\Phi_* X$ using equation 4.27.This means

$$\Phi_* X = a\partial_x + b\partial_y$$

where $a = \Phi_* X(x)|_{\Phi^{-1}(x,y)}$ and $b = \Phi_* X(y)|_{\Phi^{-1}(x,y)}$. We compute these to be

$$
\begin{aligned}
a &= \Phi_* X(x) = (r\partial_r(r\cos\theta))\,|_{\Phi^{-1}(x,y)} = x \\
b &= \Phi_* X(y) = (r\partial_r(r\sin\theta))\,|_{\Phi^{-1}(x,y)} = y
\end{aligned}
$$

This gives the vector-field,

$$\Phi_*(r\partial_r) = x\partial_x + y\partial_y. \tag{4.28}$$

Similarly we find

$$\Phi_*(\partial_\theta) = -y\partial_x + x\partial_y. \tag{4.29}$$

We now compute $\Phi_*^{-1}\partial_x$ and $\Phi_*^{-1}\partial_y$. Using the formula $\Phi^{-1}(x,y) = (r = \sqrt{x^2+y^2}, \theta = \arctan(yx^{-1})$, gives

$$
\begin{aligned}
(\Phi_*^{-1}\partial_x)(r) &= \frac{x}{r} \\
(\Phi_*^{-1}\partial_x)(\theta) &= -\frac{y}{r^2}.
\end{aligned}
$$

Therefore

$$\Phi_*^{-1}\partial_x = \cos\theta\partial_r - \frac{\sin\theta}{r}\partial_\theta. \tag{4.30}$$

Similarly we have

$$\Phi_*^{-1}\partial_y = \sin\theta\partial_r + \frac{\cos\theta}{r}\partial_\theta. \tag{4.31}$$

There is a simple way to compute Φ_*^{-1} without having to compute Φ^{-1} directly which we now describe. Suppose we are given $\Phi : U \to V$ a diffeomorphism, we have for the push forward of the vector-fields ∂_{x^j} from equation 4.25,

$$\Phi_*(\partial_{x^j}) = \sum_{i=1}^n \left.\frac{\partial \Phi^i}{\partial x^j}\right|_{\Phi^{-1}(y)} \frac{\partial}{\partial x^i}.$$

Similarly we have

$$\Phi_*^{-1}(\partial_{y^j}) = \sum_{i=1}^n \left.\frac{\partial (\Phi^{-1})^i}{\partial y^j}\right|_{\Phi(x)} \frac{\partial}{\partial x^i}. \tag{4.32}$$

However by the chain-rule $\Phi_* \circ \Phi_*^{-1} = I_*$ and therefore in a basis this yields $[\Phi_*][\Phi_*^{-1}] = I_n$ which in the coordinate basis yields

$$\left[\frac{\partial(\Phi^{-1})^i}{\partial y^j}\right]_{\Phi(x)} = \left[\frac{\partial \Phi^i}{\partial x^j}\right]^{-1}. \tag{4.33}$$

Therefore using equation 4.33 in equation 4.32 we have

$$\Phi_*^{-1}\partial_{y^j} = \sum_{i=1}^{n} \left[\frac{\partial \Phi^i}{\partial x^j}\right]^{-1} \frac{\partial}{\partial x^i}. \tag{4.34}$$

Example 4.4.4. Continuing from example 4.4.3 we have

$$\left[\frac{\partial \Phi^a}{\partial x^i}\right] = \begin{bmatrix}\cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{bmatrix}$$

and

$$\left[\frac{\partial \Phi^a}{\partial x^i}\right]^{-1} = \begin{bmatrix}\cos\theta & \sin\theta \\ -\frac{\sin\theta}{r} & \frac{\cos\theta}{r}\end{bmatrix}.$$

Therefore equation 4.34 yields

$$\Phi_*^{-1}\partial_x = [\partial_r, \partial_\theta]\begin{bmatrix}\cos\theta & \sin\theta \\ -\frac{\sin\theta}{r} & \frac{\cos\theta}{r}\end{bmatrix} = \cos\theta\partial_r - \frac{\sin\theta}{r}\partial_\theta,$$

which agrees with 4.30. A similar computation reproduces $\Phi_*^{-1}\partial_y$ (compare with in equation 4.31)

Example 4.4.5. Let Φ be the change to spherical coordinates

$$\Phi(\rho,\theta,\phi) = (x = \rho\cos\theta\sin\phi, y = \rho\sin\theta\sin\phi, z = \rho\cos\phi),$$

from example 4.4.2. We compute Φ_* by using equation 4.26,

$$\begin{aligned}(\Phi_*\partial_\rho)(x) = &\quad \partial_\rho(\rho\cos\theta\sin\phi) = \cos\theta\sin\phi \\ (\Phi_*\partial_\rho)(y) = &\quad \partial_\rho(\rho\sin\theta\sin\phi)\sin\theta\sin\phi \\ (\Phi_*\partial_\rho)(z) = &\quad \partial_\rho(\rho\cos\phi)\cos\phi.\end{aligned}$$

Therefore substituting $\Phi^{-1}(x,y,z)$ gives

$$\begin{aligned}\Phi_*\partial_\rho &= (\cos\theta\sin\phi\partial_x + \sin\theta\sin\phi\partial_y + \cos\phi\partial_z)_{\Phi^{-1}(\rho,\theta,\phi)} \\ &= \frac{1}{\sqrt{x^2+y^2+z^2}}\left(x\partial_x + y\partial_y + z\partial_z\right).\end{aligned}$$

Similarly we get

$$\Phi_* \partial_\theta = -y\partial_x + x\partial_y,$$

and using $\rho \sin\phi = \sqrt{x^2+y^2}$ we get

$$\Phi_* \partial_\phi = \frac{z}{\sqrt{x^2+y^2}}(x\partial_x + y\partial_y) - \sqrt{x^2+y^2}\partial_z.$$

Finally suppose we have a vector-field

$$X = A^\rho \partial_\rho + A^\theta \partial_\theta + A^\phi \partial_\phi$$

where A^ρ, A^θ, A^ϕ are functions of ρ, θ, ϕ. We then compute $\Phi_* X = A^x \partial_x + A^y \partial_y + A^z \partial_z$ which is a vector-field in rectangular coordinates by

$$\begin{aligned}
&\Phi_*(A^\rho \partial_\rho + A^\theta \partial_\theta + A^\phi \partial_\phi) = \frac{A_\rho}{\sqrt{x^2+y^2+z^2}}\left(x\partial_x + y\partial_y + z\partial_z\right) + \\
&+ A_\theta(-y\partial_x + x\partial_y) + A_\phi\left(\frac{z}{\sqrt{x^2+y^2}}(x\partial_x + y\partial_y) - \sqrt{x^2+y^2}\partial_z\right) = \\
&\left(\frac{xA_\rho}{\sqrt{x^2+y^2+z^2}} - yA_\theta + \frac{xzA_\phi}{\sqrt{x^2+y^2}}\right)\partial_x + \\
&\left(\frac{yA_\rho}{\sqrt{x^2+y^2+z^2}} + xA_\theta + \frac{yzA_\phi}{\sqrt{x^2+y^2}}\right)\partial_y + \\
&\left(\frac{zA_\rho}{\sqrt{x^2+y^2+z^2}} - \sqrt{x^2+y^2}A_\phi\right)\partial_z.
\end{aligned}$$

Therefore (see Appendix VII, [3]),

$$\begin{aligned}
A_x &= \frac{xA_\rho}{\sqrt{x^2+y^2+z^2}} - yA_\theta + \frac{xzA_\phi}{\sqrt{x^2+y^2}} \\
A_y &= \frac{yA_\rho}{\sqrt{x^2+y^2+z^2}} + xA_\theta + \frac{yzA_\phi}{\sqrt{x^2+y^2}} \\
A_z &= \frac{zA_\rho}{\sqrt{x^2+y^2+z^2}} - \sqrt{x^2+y^2}A_\phi,
\end{aligned}$$

where A_ρ, A_θ, A_ϕ are expressed in terms or x, y, z.

Example 4.4.6. Let $U = \{\ (x,y,z) \in I\!R^3 \mid z \neq 0\ \}$ and let $\Phi : U \to V$ be

$$\Phi(x,y,z) = (x - \frac{y^2}{2z}, z, \frac{y}{z}).$$

where $V = \{\ (u,v,w) \in I\!R^3 \mid v \neq 0\ \}$. We write the vector-field

$$X = y\partial_x + z\partial_y.$$

in (u,v,w) coordinates by computing

$$\begin{aligned}
\Phi_*(X)(u) &= (y\partial_x + z\partial_y)(x - \frac{y^2}{2z}) = 0 \\
\Phi_*(X)(v) &= (y\partial_x + z\partial_y)(z) = 0 \\
\Phi_*(X)(u) &= (y\partial_x + z\partial_y)(\frac{y}{z}) = 1.
\end{aligned}$$

Therefore

$$\Phi_* X = \partial_w.$$

Remark 4.4.7. Given a smooth function $\Phi : I\!R^n \to I\!R^m$, it is not possible to push-forward a generic vector-field X on $I\!R^n$ and get another vector-field on $I\!R^m$ unless $m = n$.

4.5 Exercises

1. For each of the following maps and derivations compute $\Phi_{*,p}(X)$ using the definition of $\Phi_{*,p}$ in terms of derivations defined in Theorem 4.2.1 (see Example 4.2.6).

   (a) $\Phi(x,y) = (u = xy, v = x^2y, w = y^2),\ p = (1,-1),\ X_p = 3\partial_x - 2\partial_y$.

   (b) $\Phi(x,y,z) = (r = x/y, s = y/z),\ p = (1,2,-2),\ X_p = \partial_x - \partial_y + 2\partial_z$.

   (c) $\Phi(t) = (x = \exp(3t)\sin(\pi t), y = \exp(t)\cos(2\pi t), z = t),\ p = (0),\ X = \partial_t$.

2. Repeat question 1 by choosing a curve σ with $\dot\sigma(t_0) = X_p$ and using Definition 4.1.2.

3. Let
$$\Phi(x,y) = (u = x^3 + 3xy^2, v = 3x^2y + y^3),$$
and let $p = (1,-1)$. Compute the matrix representation of $\Phi_{*,p} : T_p\mathbb{R}^2 \to T_{\Phi(p)}\mathbb{R}^2$ with respect to the basis $\{X_1, X_2\}$ for the tangent space for T_pR^2 and $\{Y_1, Y_2\}$ for $T_{\Phi(p)}R^2$.

   (a) $X_1 = \partial_x|_p, X_2 = \partial_y|_p, Y_1 = \partial_u|_q, Y_2 = \partial_v|_q$.

   (b) $X_1 = (\partial_x - 2\partial_y)|_p, X_2 = (\partial_x + \partial_y)|_p, Y_1 = \partial_u|_q, Y_2 = \partial_v|_q$.

   (c) $X_1 = (\partial_x - 2\partial_y)|_p, X_2 = (\partial_x + \partial_y)|_p, Y_1 = (2\partial_u + 3\partial_v)|_q, Y_2 = (\partial_u + 2\partial_v)|_q$.

4. (a) Write the tangent vector $X_p = \left.\frac{\partial}{\partial x}\right|_p - \left.\frac{\partial}{\partial y}\right|_p$ $p = (1,1)$, in polar coordinates using the coordinate basis.

   (b) Write the tangent vector $X_p = -2\left.\frac{\partial}{\partial r}\right|_p + 3\left.\frac{\partial}{\partial \theta}\right|_p$, $p = (r = 1, \theta = \pi/3)$ given in polar coordinates, in Cartesian coordinates using the coordinate basis.

   (c) Let
$$\Phi(u,v) = \left(uv, \frac{1}{2}(v^2 - u^2)\right)$$

be the change to parabolic coordinates. Write the tangent vector $X_p = \left.\frac{\partial}{\partial u}\right|_p - \left.\frac{\partial}{\partial v}\right|_p$, $p = (u = 1, v = 1)$ given in parabolic coordinates, in Cartesian coordinates using the coordinate basis.

(d) Write the tangent vector X_p in part c) in polar coordinates using the coordinate basis.

5. Check the chain rule $(\Psi \circ \Phi)_{*,p} = \Psi_{*,p} \circ \Phi_{*,p}$ for each of the following maps at the point indicated. You may calculate in the coordinate basis.

   (a) $\Phi(x,y) = (x^2 - y^2, xy, x+y+2)$, $\Psi = (u,v,w) = (1/u, 1/v, 1/w)$, $p = (2,1)$

   (b) $\Phi(t) = (t, t^2, t^3)$, $\Psi = (u,v,w) = (u/v, v/w, uw)$, $p = (1)$

   (c) $\Phi(t) = (t, t^2, t^3)$, $\Psi = (u,v,w) = (u^2 + v^2 + w^2)$, $p = (1)$

6. Find the inverses of the following maps Φ and check that $(\Phi_{*,p})^{(-1)} = (\Phi^{(-1)})_{*,p}$ at the indicated point.

   (a) $\Phi(x,y,z) = (x+1, y+x^2, z-xy)$, $p = (1,1,2)$.

   (b) $\Phi(x,y) = (\frac{1}{2}(x^2 - y^2), xy)$ $p = (2,1)$, on $U = \{(x,y) \in I\!R^2 \mid x > y > 0\}$.

7. Find the points, if any, in the domain of the following maps about which the map fails to be a local diffeomorphism.

   (a) $\Phi(x,y) = (x^3 - 3xy^2, -y^3 + 3x^2 y)$.

   (b) $\Phi(x,y,z) = (x^2 + y + z, y^2 - z + y, y - z)$.

8. Show that the following maps are immersions.

   (a) $\Phi(u,v) = (u+v, u-v, u^2 + v^2)$

   (b) $\Phi(u,v,w) = (u+v+w, u^2+v^2+w^2, u^3+v^3+w^3, u^4+v^4+w^4)$, $w > v > u$.

9. Show that the following maps are submersions.

   (a) $\Phi(x,y,z) = (x+y-z, x-y)$

(b) $\Phi(x, y, z) = (x^2 + y^2 - z^2, x^2 - y^2) \quad y > 0$.

10. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$F(x, y, z) = (x^2 + y^2, x^2 + 2z, x - y + z)$$

(a) Find the kernel of $F_{*,p} : T_{(0,0,0)}\mathbb{R}^3 \to T_{(0,0,0)}\mathbb{R}^3$.

(b) Does there exists a vector $X_p \in T_{(1,0,1)}\mathbb{R}^3$ such that $F_{*,p}X_p = Y_q$ where $Y_q = (2\partial_x - 3\partial_y + \partial_z)_{(1,3,2)}$

(c) Compute $F_{*,p}(\partial_x - \partial_z)|_{(1,1,1)}$ in two different ways.

(d) Determine the set of points where F is an immersion, submersion, and local diffeomorphism. Are these three answers different? Why or why not.

11. Let $\Phi(u, v) = \big(x = uv, y = \frac{1}{2}(v^2 - u^2)\big)$ be the change to parabolic coordinates.

(a) Find $u\partial_u + v\partial_v$ in rectangular coordinates using the coordinate basis ∂_x, ∂_y.

(b) Find $u\partial_v - v\partial_u$ in rectangular coordinates.

(c) Find $y\partial_x - x\partial_y$ in parabolic coordinates.

12. Let $X = A^x\partial_x + A^y\partial_y + A^z\partial_z$ be a vector-field in rectangular.

(a) Compute ∂_x, ∂_y and ∂_z in spherical coordinates using the technique from example 4.4.4.

(b) Suppose $Y = A^\rho\partial_\rho + A^\theta\partial_\theta + A^\phi\partial_\phi$ is the formula for X in spherical coordinates. Write A^ρ, A^θ, A^ϕ in terms of A^x, A^y, A^z. (See appendix VII, [3])

Chapter 5

Differential One-forms and Metric Tensors

5.1 Differential One-Forms

Recall that if V is an n dimensional vector space W is an m dimensional vector-space that $L(V, W)$ the set of linear transformations from V to W was shown in Lemma 2.3.11 to be an mn dimensional vector-space. If $W = \mathbb{R}$ then $L(V, \mathbb{R})$ is n dimensional and is called the dual space to V, denoted V^*. The set V^* is also called the space of linear functionals on V, the co-vectors on V or the one-forms on V.

Example 5.1.1. Let $V = \mathbb{R}^3$, then

$$\alpha\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = x + 2y + z \tag{5.1}$$

satisfies $T \in V^*$. That is α is a linear transformation from V to $\mathbb{R}$. In fact it is not difficult to show (see Exercise 1 that if $\alpha \in V^*$, there exists $a, b, c \in \mathbb{R}$ such that

$$\alpha\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = ax + by + cz. \tag{5.2}$$

Suppose that $\beta = \{v_i\}_{1 \leq i \leq n}$ is a basis for the n dimensional vector-space V is an $n-$dimensional vector-space. By Lemma 2.1.7 every linear transformation is uniquely determined by its values on a basis. In particular if

$\alpha \in V^*$, then α is determined by the real numbers $\alpha(v_i)$, and again by Lemma 2.1.7 every function $\hat{\alpha} : \beta \to I\!R$ extends to a unique element $\alpha \in V^*$ (which agrees with $\hat{\alpha}$ on β). Define the function $\hat{\alpha}^1 : \beta \to I\!R$ by

$$\begin{aligned} &\hat{\alpha}^1(v_1) = 1 \\ &\hat{\alpha}^1(v_i) = 0 \quad 2 \leq i \leq n, \end{aligned}$$

which extends to a unique linear transformation $\alpha^1 : V \to I\!R$, where

$$\alpha^1(\sum_i^n c_i v^i) = c_1.$$

We can then define a sequence of functions $\alpha^i, 1 \leq i \leq n$ in a similar way. For each fixed $i \in \{1, \ldots, n\}$, let α^i be the unique element of V^* satisfying

$$\alpha^i(v_j) = \delta^i_j = \begin{cases} 1 & if\ i = j, \\ 0 & otherwise. \end{cases} \tag{5.3}$$

Example 5.1.2. Let

$$\beta = \left\{ v_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}. \tag{5.4}$$

be a basis for $I\!R^3$. We compute $\alpha^1, \alpha^2, \alpha^3$ and express them in the form of equation 5.2. One way to do this is to note

$$e_1 = v_1 - v_2 + v_3,\ e_2 = 2v_2 - v_1 - v_3,\ e_3 = v_3 - v_2.$$

Then using the properties in equation 5.3 we get,

$$\alpha^1(e_1) = \alpha^1(v_1 - v_2 + v_3) = 1,\ \alpha^1(e_2) = -1,\ \alpha^1(e_3) = 0.$$

Therefore

$$\alpha^1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \alpha^1(xe_1 + ye_2 + ze_3) = x - y. \tag{5.5}$$

Similarly

$$\alpha^2\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = -x + 2y - z, \quad \alpha^3\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = x - y + z. \tag{5.6}$$

Theorem 5.1.3. *The set of linear transformations $\alpha^1, \dots, \alpha^n \in V^*$ defined by equation 5.3 form a basis for V^*. If $\alpha \in V^*$ then*

$$\alpha = \sum_{i=1}^{n} \alpha(v_i)\alpha^i. \tag{5.7}$$

Proof. To show that these are linearly independent, suppose

$$Z = \sum_{i=1}^{n} c_i \alpha^i$$

is the zero element of V^*. The zero element is the linear transformation which maps every vector in V to $0 \in I\!R$. Therefore

$$0 = Z(v_1) = (\sum_{i=1}^{n} c_i \alpha^i)(v_1) = \sum_{i=1}^{n} c_i \alpha^i(v_1) = c_1 \alpha^1(v_1) = c_1$$

and so $c_1 = 0$. Likewise applying Z to the rest of the basis elements v_j we get zero. That is,

$$0 = Z(v_j) = c_j$$

and so $c_j = 0$, and $\{\alpha^i\}_{1 \leq i \leq n}$ is a linearly independent set.

To prove that $\{\alpha^i\}_{1 \leq i \leq n}$ are a spanning set, let $\alpha \in V^*$ and consider $\tau \in V^*$,

$$\tau = \sum_{i=1}^{n} \alpha(v_i)\alpha^i.$$

Then

$$\tau(v_j) = \sum_{i=1}^{n} \alpha(v_i)\alpha^i(v_j) = \sum_{i=1}^{n} \alpha(v_i)\delta^i_j = \alpha(v_j).$$

Therefore the two linear transformation $\tau, \alpha \in V^*$ agree on a basis and so by Lemma 2.1.7 are equal. This proves equation 5.7 holds and that $\{\alpha^i\}_{1 \leq i \leq n}$ is a spanning set. □

Given the basis β the basis $\{\alpha^i\}_{1 \leq i \leq n}$ for V^* in 5.3 is called the dual basis.

Example 5.1.4. Let $\alpha : \mathbb{R}^3 \to \mathbb{R}$ be the linear functional in equation 5.1 and let $\beta^* = \{\alpha^1, \alpha^2, \alpha^3\}$ be the dual basis given in equations 5.5, 5.6 to β in equation 5.4. Then by equation 5.7 we have

$$\begin{aligned}\alpha &= \alpha\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}\right)\alpha^1 + \alpha\left(\begin{bmatrix}1\\1\\0\end{bmatrix}\right)\alpha^2 + \left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right)\alpha^3\\ &= 3\alpha^2 + 4\alpha^3.\end{aligned}$$

We now consider the case $V = T_p\mathbb{R}^n$ and the corresponding dual space $T_p^*\mathbb{R}^n$ which is called the space of one-forms, co-vectors, or dual vectors at the point $p \in \mathbb{R}^n$. In this case it turns out that elements in $T_p^*\mathbb{R}^n$ appear in a natural way. Recall that if $X_p \in T_p\mathbb{R}^n$, then X_p is a derivation of $C^\infty(p)$. Let $f \in C^\infty(p)$ and so there exists an open set $U \subset \mathbb{R}^n$ with $p \in U$ and $f \in C^\infty(U), p \in U$. Then by definition

$$X_p(f) \in \mathbb{R}. \tag{5.8}$$

If $Y_p \in T_p\mathbb{R}^n$ then

$$(aX_p + Y_p)(f) = aX_p(f) + Y_p(f). \tag{5.9}$$

Now in the formula $X_p(f) \in \mathbb{R}$ from equation 5.8, we **change the way we think of this formula** - instead of this formula saying $X_p : f \to \mathbb{R}$ so that f is the argument of X_p, we instead think of X_p as being the argument of the function f! That is "f'' : $T_p\mathbb{R}^n \to \mathbb{R}$. We use a new notation to distinguish this way of thinking of f. Define the function $df_p : T_p\mathbb{R}^n \to \mathbb{R}$ by

$$df_p(X_p) = X_p(f). \tag{5.10}$$

Proposition 5.1.5. *The function df_p satisfies $df_p \in T_p^*\mathbb{R}^n$.*

Proof. We need to prove that df_p is a linear function. That is we need to show

$$df_p(aX_p + Y_p) = a\, df_p(X_p) + df_p(Y_p). \tag{5.11}$$

The left side of this on account of equation 5.10 is

$$df_p(aX_p + Y_p) = (aX_p + Y_p)(f) \tag{5.12}$$

while again on account of 5.10 is

$$a\, df_p(X_p) + df_p(Y_p) = a\, X_p(f) + Y_p(f). \tag{5.13}$$

Equation 5.12 and 5.13 agree by equation 5.9. □

Example 5.1.6. Let $f \in C^\infty(I\!R)$ be given by $f(x,y) = xy^2$, and let $p = (1,2) \in I\!R^2$. We compute and $df_p(X_p)$ where $X_p = (3\partial_x - 2\partial_y)_p$ by equation 5.10,

$$df_p(X_p) = (3\partial_x - 2\partial_y)(xy^2)|_{(1,2)} = 4.$$

We know that for each point $p \in I\!R^n$ the set $\beta = \{\partial_{x^i}\}_{1\leq i\leq n}$ is a basis for $T_pI\!R^n$. Let's calculate the dual basis. We begin by considering the function

$$f^1(x^1, \ldots, x^n) = x^1$$

which we just call x^1. Then by equation 5.10

$$dx^1_p(\partial_{x^1}|_p) = \partial_{x^1}(x^1) = 1,\ dx^1_p(\partial_{x^2}|_p) = \partial_{x^2}(x^1)0, \ldots, dx^1_p(\partial_{x^n}|_p) = 0.$$

This leads to the general case

$$dx^i_p(\partial_{x^j}|_p) = \delta^i_j, \tag{5.14}$$

where δ^i_j is the Kronecker delta. From equation 5.14 we conclude that for each $p \in I\!R^n$ the set of one-forms

$$\beta^* = \{dx^i_p\}_{1\leq i\leq n} \tag{5.15}$$

form a basis for $T^*_pI\!R^n$ which satisfies equation 5.3. Equation 5.15 is the dual basis to the coordinate basis $\beta = \{\partial_{x^i}\}_{1\leq i\leq n}$ to the tangent space $T_pI\!R^n$. We will call the basis $\{dx^i_p\}$ the coordinate basis for $T^*_pI\!R^n$.

We now express $df_p \in T^*_pI\!R^n$ in terms of our basis β^* in equation 5.15. Let $p \in U$ then by Theorem 5.1.3 we have

$$df_p = \sum_{i=1}^n c^i \in I\!R,$$

where $c_i = df_p(\partial_{x^i}|_p) = \partial_{x^i}|_p(f)$. Therefore

$$df_p = \sum_{i=1}^n \left.\frac{\partial f}{\partial x^i}\right|_p dx^i_p \tag{5.16}$$

holds at each point $p \in U$.

Example 5.1.7. Let $f : I\!R^3 \to I\!R$ be

$$f(x, y, z) = x^2 e^{y-z},$$

and $p = (1, 1, 1)$.

$$\begin{aligned} df_p &= \left(2xe^{y-z}dx + x^2e^{y-z}dy - x^2e^{y-z}dz\right)_p, \\ &= 2dx_p + dy_p - dz_p \end{aligned}$$

Let $X_p = (2\partial_x - \partial_y - 3\partial_z)|_p$, and we compute $df_p(X_p)$ in two ways. First by definition 5.10 we have

$$df_p(X_p) = X_p(x^2e^{y-z}) = 4 - 1 + 3 = 6.$$

The second way we use the properties of the vector space $T_p^* I\!R^3$ we have

$$\begin{aligned} df_p(X_p) &= (2dx_p + dy_p - dz_p)(2\partial_x|_p - \partial_y|_p - 3\partial_z|_p) \\ &= 2dx_p(2\partial_x|_p - \partial_y|_p - 3\partial_z|_p) + dy_p(2\partial_x|_p - \partial_y|_p - 3\partial_z|_p) \\ &\quad - dz_p(2\partial_x|_p - \partial_y|_p - 3\partial_z|_p) \\ &= 4 - 1 + 3 = 6. \end{aligned}$$

In this computation the fact that $dx_p(\partial_x|_p) = 1, dx_p(\partial_y|_p) = 0, dx_p(\partial_z|_p) = 0, \ldots, dz_p(\partial_z|_p) = 1$ has been used.

In direct analogy with a vector-field on $I\!R^n$ being a smoothly varying choice of a tangent vector at each point we define a differential one-form (or a one-form field) on $I\!R^n$ as a smoothly varying choice of one-form at each point. Since $\{dx^i|_p\}_{1\leq i\leq n}$ form a basis for $T_p^* I\!R^n$ at each point then every differential one-form α can be written

$$\alpha = \sum_{i=1}^{n} \alpha_i(\mathbf{x}) dx^i|_{\mathbf{x}}$$

where $\alpha_i(\mathbf{x})$ are n smooth functions on $I\!R^n$. As with vector-fields we will drop the subscript on dx^i, and write $\alpha = \sum_{i=1}^n \alpha_i(\mathbf{x})dx^i$.

Example 5.1.8. On $I\!R^3$,

$$\alpha = ydx - xdy + zydz$$

is a one-form field. At the point $p = (1, 2, 3)$,

$$\alpha_p = 2dx_p - dy_p + 6dz_p$$

If $f \in C^\infty(I\!R^n)$ then for each $p \in I\!R^n$, by equation 5.10, $df_p \in T^*_p I\!R^n$. This holds for each $p \in I\!R^n$ and so df is then a one-form field called the differential of f. The expansion of df in terms of the differential one-forms dx^i which form the dual basis at each point is obtained from equation 5.16 to be

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i. \tag{5.17}$$

Example 5.1.9. Find the differential of $f = x^2 y e^{xz} \in C^\infty(I\!R^3)$. By equation 5.17

$$df = (2xy + x^2 yz)e^{xz} dx + x^2 e^{xz} dy + x^3 y e^{xz} dz.$$

5.2 Bilinear forms and Inner Products

In the previous section we began by with the linear algebra of linear functions $T : V \to I\!R$. A generalization of this is to consider a function $B : V \times V \to I\!R$ which satisfies the properties

$$\begin{aligned} B(a_1 v_1 + a_2 v_2, w) &= a_1 B(v_1, w) + a_2 B(v_2, w), \\ B(v, a_1 w_1 + a_2 w_2) &= a_1 B(v, w_1) + a_2 B(v, w_2). \end{aligned} \tag{5.18}$$

These equations imply that B is linear as a function of each of its two arguments. A function B which satisfies these conditions is called a *bilinear form.*

Example 5.2.1. Let $A \in M_{n\times n}(I\!R)$. Define $B : I\!R^n \times I\!R^n \to I\!R$ by

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} x^i A^i_j y^j = \mathbf{x}^T A \mathbf{y},$$

where $\mathbf{x}^T$ is the transpose of the column vector $\mathbf{x}$. The function B is easily checked to be bilinear form.

Let $B, C : V \times V \to I\!R$ be bilinear forms. Then $aB + C$ is a bilinear form where

$$(aB + C)(v, w) = aB(v, w) + C(v, w). \tag{5.19}$$

Denote by $\mathbf{B}(V)$ the space of bilinear forms, which is a vector-space (see exercise 5 in this chapter).

For the rest of this section, let V be a finite dimensional vector-space with basis $\beta = \{v_i\}_{1\leq i\leq n}$. Let $B \in \mathbf{B}(V)$ be a bilinear form. The $n \times n$ matrix (B_{ij}) whose entries are

$$B_{ij} = B(v_i, v_j)\ , \quad 1 \leq i, j \leq n, \tag{5.20}$$

is the matrix representation of the bilinear form B in the basis β.

Example 5.2.2. Let $B : I\!R^3 \times I\!R^3 \to I\!R$ be the bilinear function

$$B\left(\begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}, \begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix}\right) = 2x^1y^1 - x^1y^2 + x^1y^3 - x^2y^1 - x^2y^2 + x^3y^1 + x^3y^3.$$

Using the basis

$$\beta = \left\{ v_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}. \tag{5.21}$$

we compute the entries of the matrix representation $B_{ij} = B(v_i, v_j)$. We find

$$\begin{array}{lll} B(v_1, v_1) = 1, & B(v_1, v_2) = 0, & B(v_1, v_3) = 0, \\ B(v_1, v_2) = 0, & B(v_2, v_2) = -1, & B(v_2, v_3) = 0, \\ B(v_3, v_1) = 0, & B(v_3, v_2) = 0, & B(v_3, v_3) = 2 \end{array} \tag{5.22}$$

and so the matrix representation of B in the basis $\beta = \{v_1, v_2, v_3\}$ is

$$[B_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{5.23}$$

Example 5.2.3. As in example 5.2.1 let $B : I\!R^n \times I\!R^n \to I\!R$ be given by $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A\mathbf{y}$ where $A \in M_{n\times n}(I\!R)$. In the standard basis $\beta = \{e_i\}_{1\leq i\leq n}$,

$$B(e_i, e_j) = A_{ij}$$

the entry in the i^{th} row and j^{th} column of A.

Theorem 5.2.4. *Let $v, w \in V$ with*

$$v = \sum_{i=1}^{n} a^i \, v_i, \quad w = \sum_{i=1}^{n} b^i \, v_i. \tag{5.24}$$

Then

$$B(v, w) = \sum_{i,j=1}^{n} B_{ij} a^i b^j = [v]^T (B_{ij})[w], \tag{5.25}$$

where $[v] = [a^i], [w] = [b^i]$ are the column vectors of the components of v and w in (5.24).

Proof. We simply expand out using (5.24), and the bi-linearity condition,

$$\begin{aligned} B(v, w) &= B\left(\sum_{i=1}^{n} a^i \, v_i, \sum_{j=1}^{n} b^j \, v_j\right) \\ &= \sum_{i=1}^{n} a^i B\left(v_i, \sum_{j=1}^{n} b^j \, v_j\right) \quad = \sum_{i=1}^{n} \sum_{j=1}^{n} a^i b^j B\left(v_i, w_j\right). \end{aligned}$$

This is the formula in equation 5.25 in the theorem. □

An easy consequence of this theorem is the following.

Corollary 5.2.5. *Let $B, C \in \mathbf{B}(V)$. The bilinear forms satisfy $B = C$ if and only if for any basis β, $B_{ij} = C_{ij}$.*

A bilinear form $B \in \mathbf{B}(V)$ is symmetric if $B(w, v) = B(v, w)$ for all $v, w \in V$. The bilinear for B is skew-symmetric or alternating if $B(w, v) = -B(v, w)$ for all $v, w \in V$.

Example 5.2.6. Let $V = I\!R^n$ and let Q_{ij} be any symmetric matrix. Define

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} Q_{ij} x^i y^j = \mathbf{x}^T Q \mathbf{y}.$$

Then the bilinear form B is symmetric.

We have the following simple proposition.

Proposition 5.2.7. *A bilinear form $B : V \times V \to \mathbb{R}$ is symmetric if and only if its matrix representation (in any basis) is a symmetric matrix. The bilinear form B is skew-symmetric if and only if its matrix representation in a basis is a skew-symmetric matrix.*

It is easy to show that if the matrix representation of B is symmetric or skew-symmetric in a basis, then it will have this property in every basis.

The matrix representation of the bilinear form B in example 5.2.2 is 5.23 and is symmetric. Therefore Proposition implies B is a symmetric bilinear form.

Definition 5.2.8. An inner product is a bilinear form $\gamma : V \times V \to \mathbb{R}$ that satisfies

1. $\gamma(v, w) = \gamma(w, v)$, and
2. $\gamma(v, v) \geq 0$,
3. $\gamma(v, v) = 0$ if and only if $v = \mathbf{0}_V$.

If γ is an inner product on V, then the length of a vector $v \in V$ is

$$||v|| = \sqrt{\gamma(v, v)}.$$

Theorem 5.2.9. *(Cauchy-Schwarz) Let $\gamma : V \times V \to \mathbb{R}$ be an inner product. Then*

$$|\gamma(v, w)| \leq ||v||||w||. \tag{5.26}$$

A proof of the Cauchy-Schwarz inequality 5.26 can be found in most books on linear algebra, see for example [6].

Given an inner product γ, the Cauchy-Schwartz inequality 5.26 allows us to define the angle θ between v, w (in the plane $W = \text{span}\{v, w\}$) by the usual formula

$$\gamma(v, w) = \cos\theta \, ||v|| \, ||w||. \tag{5.27}$$

Example 5.2.10. Let $V = \mathbb{R}^n$. The standard inner product is

$$\gamma(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x^i y^i = \mathbf{x}^T \mathbf{y}$$

where $\mathbf{x} = \sum_{i=1}^{n} x^i\, e_i$, $\mathbf{y} = \sum_{i=1}^{n} y^i\, e_i$. which is often called the dot product. It is easy to check that γ is an inner product. In the standard basis $\beta = \{e_i\}_{1\leq i\leq n}$,

$$g_{ij} = \gamma(e_i, e_j) = \delta_{ij}, \quad 1 \leq i, j \leq n,$$

and is the identity matrix. If $v = \sum_{i=1}^{n} a^i e_i$ then

$$||v|| = \sum_{i=1}^{n} (a^i)^2.$$

Let $\beta = \{v_i\}_{1\leq i\leq n}$ be a basis for V and let γ is an inner product. The $n \times n$ matrix $[g_{ij}]$ with entries

$$g_{ij} = \gamma(v_i, v_j)\ , \quad 1 \leq i, j \leq n, \tag{5.28}$$

is the matrix representation 5.20 of the bilinear form γ in the basis β. Properties 1 in 5.2.8 implies the matrix $[g_{ij}]$ is symmetric.

Property 2 and 3 for γ in definition 5.2.8 can be related to properties of its matrix representation $[g_{ij}]$. First recall that A real symmetric matrix is always diagonalizable over the $\mathbb{R}$ (Theorem 6.20 of [6]). This fact gives a test for when a symmetric bilinear form γ is positive definite in terms of the eigenvalues of a matrix representation of γ.

Theorem 5.2.11. *Let γ be a bilinear form on V and let $g_{ij} = \gamma(v_i, v_j)$ be the coefficients of the matrix representation of γ in the basis $\beta = \{v_i\}_{1\leq i\leq n}$. The bilinear form γ is an inner product if and only if its matrix representation $[g_{ij}]$ is a symmetric matrix with strictly positive eigenvalues.*

The property that $[g_{ij}]$ is symmetric with positive eigenvalues does not depend on the choice of basis.

Example 5.2.12. Consider again the bilinear form $\beta : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ from example 5.2.2,

$$B\left(\begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}, \begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix} \right) = 2x^1y^1 - x^1y^2 + x^1y^3 - x^2y^1 - x^2y^2 + x^3y^1 + x^3y^3.$$

In the basis in equation 5.21 the matrix representation is

$$[B_{ij}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Therefore by proposition 5.2 B is a symmetric bilinear form. However by Theorem 5.2.11 B is not positive definite, and hence B is not an inner product.

Example 5.2.13. Let $V = I\!R^n$ and let Q_{ij} be any symmetric matrix with positive eigenvalues. Define

$$\gamma(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} Q_{ij} x^i y^j = \mathbf{x}^T Q \mathbf{y}.$$

Then by Theorem 5.2.11 the bilinear form γ is an inner product.

5.3 Tensor product

There is a very important way to construct bilinear forms using linear ones called the tensor product. Let V be a vector-space and let $\alpha^1, \alpha^2 \in V^*$. Now define $\alpha^1 \otimes \alpha^2 : V \times V \to I\!R$ by

$$\alpha^1 \otimes \alpha^2 (v, w) = \alpha^1(v)\alpha^2(w), \quad for\ all\ v, w \in V.$$

Theorem 5.3.1. *The function $\alpha^1 \otimes \alpha^2 : V \times V \to I\!R$ is bilinear.*

Proof. This is simple to check. Using the linearity of α^1 we find,

$$\begin{aligned} \alpha^1 \otimes \alpha^2(av_1 + v_2, w) &= \alpha^1(av_1 + v_2)\alpha^2(w) \\ &= a\alpha^1(v_1)\alpha^2(w) + \alpha^1(v_2)\alpha^2(w) \\ &= a\alpha^1 \otimes \alpha^2(v_1, w) + \alpha^1 \otimes \alpha^2(v_2, w) \end{aligned}$$

and so $\alpha^1 \otimes \alpha^2$ is linear in the first argument. The linearity of $\alpha^1 \otimes \alpha^2$ in the second argument is shown using the linearity of α^2. □

Given $\alpha^1, \alpha^2 \in V^*$, the bilinear form $\alpha^1 \otimes \alpha^2$ and is called the *tensor product* of α^1 and α^2. This construction is the beginning of the subject multi-linear algebra, and the theory of tensors.

Let $\beta = \{v_i\}_{1 \leq i \leq n}$ be a basis for V and let $\beta^* = \{\alpha^j\}_{1 \leq j \leq n}$ be the dual basis for V^*, and let

$$\Delta = \{\alpha^i \otimes \alpha^j\}_{1 \leq i,j \leq n}. \tag{5.29}$$

The next theorem is similar to Theorem 5.1.3.

Theorem 5.3.2. *The n^2 elements of Δ in equation 5.29 form a basis for $\mathbf{B}(V)$. Moreover*

$$B = \sum_{1\leq i,j\leq n} B_{ij}\alpha^i \otimes \alpha^j \tag{5.30}$$

where $B_{ij} = B(v_i, v_j)$.

Proof. Let $B \in \mathbf{B}(V)$, and let

$$B_{ij} = B(v_i, v_j), \quad v_i, v_j \in \beta$$

be the $n \times n$ matrix representation of B in the basis β. Now construct $C \in \mathbf{B}(V)$,

$$C = \sum_{1\leq i,j\leq n} B_{ij}\alpha^i \otimes \alpha^j.$$

We compute the matrix representation of C in the basis β by computing

$$\begin{aligned} C(v_k, v_l) &= \sum_{1\leq i,j\leq n} B_{ij}\alpha^i(v_k)\alpha^j(v_l) \\ &= \sum_{1\leq i,j\leq n} B_{ij}\delta^i_k\delta^j_l \\ &= B_{kl}. \end{aligned}$$

Here we have used $\alpha^i(v_k) = \delta^i_k, \alpha^j(v_l) = \delta^j_l$. Therefore by corollary 5.2.5 $B = C$, and so Δ is a spanning set. The proof that Δ is a linearly independent set is left for the exercises. □

Note that formula 5.30 for bilinear forms is the analogue to formula 5.7 for linear function.

Example 5.3.3. Let $B : I\!R^3 \times I\!R^3 \to I\!R$ be the bilinear form from example 5.2.2. Let $\{\alpha^i\}_{1\leq i\leq 3}$ be the dual basis of the basis in equation 5.21. Then by Theorem 5.3.2 and equation 5.22 we have

$$B = \alpha^1 \otimes \alpha^1 - \alpha^2 \otimes \alpha^2 + 2\alpha^3 \otimes \alpha^3. \tag{5.31}$$

Let $\beta = \{v_i\}_{1\leq i\leq n}$ be a basis for V and $\Delta = \{\alpha^i\}_{1\leq i\leq n}$ be a basis for the dual space V^* and let B be symmetric bilinear form on V. Theorem 5.3.2 allows us to write the inner product B as in equation 5.30 we have

$$B = \sum_{1\leq i,j\leq n} B_{ij}\alpha^i \otimes \alpha^j, \tag{5.32}$$

where $B_{ij} = B(v_i, v_j)$. In equation 5.32 we note that by using the symmetry $B_{ij} = B_{ji}$ we can write

$$B_{ij}\alpha^i \otimes \alpha^j + B_{ji}\alpha^j \otimes \alpha^i = B_{ij}(\alpha^i \otimes \alpha^j + \alpha^j \otimes \alpha^i).$$

Therefore we define

$$\alpha^i\alpha^j = \frac{1}{2}(\alpha^i \otimes \alpha^j + \alpha^j \otimes \alpha^i),$$

so that equation 5.32 can be written

$$B = \sum_{1 \leq i,j \leq n} B_{ij}\, \alpha^i\, \alpha^j. \tag{5.33}$$

This notation will be used frequently in the next section. Equation 5.31 of example we can be rewritten as

$$B = \alpha^1 \otimes \alpha^1 - \alpha^2 \otimes \alpha^2 + 2\alpha^3 \otimes \alpha^3 = (\alpha^1)^2 - (\alpha^2)^2 + 2(\alpha^3)^3.$$

Example 5.3.4. Let V be a vector-space and $\beta = \{v_i\}_{1 \leq i \leq n}$ a basis and $\beta^* = \{\alpha^j\}_{1 \leq j \leq n}$ the dual basis. Let

$$\gamma = \sum_{i=1}^{n} c_i\alpha^i \otimes \alpha^i = \sum_{i=1}^{n} c_i(\alpha^i)^2 \quad c_i \in I\!R\,, c_i > 0.$$

We claim γ is an inner-product on V. First γ is a sum of bilinear forms and therefore bilinear. Let's compute the matrix representation of γ in the basis β. First note

$$\begin{aligned}
\gamma(v_1, v_1) &= (\sum_{i=1}^{n} c_i\alpha^i \otimes \alpha^i)(v_1, v_1), \\
&= \sum_{i=1}^{n} c_i\alpha^i(v_1) \otimes \alpha^i(v_1) \\
&= c_1
\end{aligned}$$

where we've used $\alpha^i(v_1) = 0, i \neq 1$. Similarly then $\gamma(v_1, v_2) = 0$, and in general

$$\gamma(v_i, v_i) = c_i, \text{ and } \quad \gamma(v_i, v_j) = 0,\ i \neq j.$$

Therefore the matrix representation is the diagonal matrix,

$$[g_{ij}] = \text{diag}(c_1, c_2, \ldots, c_n).$$

By Theorem 5.2.11 and that $c_i > 0$, γ is positive definite, and so an inner product.

5.4 Metric Tensors

Let $p \in \mathbb{R}^n$ and suppose we have an inner product $\gamma_p : T_p\mathbb{R}^n \times T_p\mathbb{R}^n \to \mathbb{R}$ on the tangent space at p. The matrix representation 5.28 in the coordinate basis $\{\partial_{x^i}\}_{1\le i\le n}$, is the $n \times n$ matrix $[g_{ij}]_{1\le i,j\le n}$ with entries

$$[\gamma_p]_{ij} = g_{ij} = \gamma_p\left(\partial_{x^i}|_p, \partial_{x^j}|_p\right), \quad g_{ij} \in \mathbb{R}.$$

If we use the dual basis $\{dx^i|_p\}_{1\le i\le n}$, then equation 5.30 in Theorem 5.3.2 says

$$\begin{aligned} \gamma(p) &= \sum_{1\le i,j\le n} g_{ij}\, dx^i|_p \otimes dx^j|_p \\ &= \sum_{1\le i,j\le n} g_{ij}\, dx^i|_p dx^j|_p \qquad \text{by equation 5.33.} \end{aligned} \tag{5.34}$$

Definition 5.4.1. A metric tensor $\boldsymbol{\gamma}$ on $\mathbb{R}^n$ is a choice of inner product

$$\boldsymbol{\gamma}_p : T_p\mathbb{R}^n \times T_p\mathbb{R}^n \to \mathbb{R}$$

for **each** point $p \in \mathbb{R}^n$, which varies smoothly with p.

A metric tensor $\boldsymbol{\gamma}$ is also called a *Riemannian metric*. It is a field of bilinear forms on the tangent space satisfying the conditions of an inner product at each point.

We now say precisely what it means for $\boldsymbol{\gamma}$ to vary smoothly. The derivations ∂_{x^i} form a basis at every point, therefore given a metric tensor $\boldsymbol{\gamma}$, we define the n^2 functions $g_{ij} : \mathbb{R}^n \to \mathbb{R}$ by

$$g_{ij}(\mathbf{x}) = \boldsymbol{\gamma}(\partial_{x^i}|_{\mathbf{x}}, \partial_{x^j}|_{\mathbf{x}}) \quad \textit{for all } \mathbf{x} \in \mathbb{R}^n. \tag{5.35}$$

A smooth metric tensor is one where the functions $g_{ij}(\mathbf{x})$ are C^∞ functions on $\mathbb{R}^n$. Using $\{dx^i|_p\}_{1\le i\le n}$ as the dual basis for $\partial_{x^i}|_p$, applying the formula 5.34 pointwise for $\boldsymbol{\gamma}$ using equation 5.35 yields

$$\boldsymbol{\gamma} = \sum_{1\le i,j\le n} g_{ij}(\mathbf{x})\, dx^i dx^j. \tag{5.36}$$

Note that at each fixed point $p \in \mathbb{R}^n$, the matrix $[g_{ij}(p)]$ will be symmetric and positive definite (by Theorem 5.2.11) because $\boldsymbol{\gamma}_p$ is an inner product.

Conversely we have the following.

Theorem 5.4.2. *Let $[g_{ij}(\mathbf{x})]$ be a matrix of smooth functions on $I\!R^n$ with the property that at each point $p \in I\!R^n$ that $[g_{ij}(p)]$ is a positive definite symmetric matrix, then*

$$\gamma = \sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) dx^i dx^j \tag{5.37}$$

is a metric tensor on $I\!R^n$.

This theorem then states that every metric tensor is of the form 5.37. The functions $g_{ij}(\mathbf{x})$ are be called the components of γ in the coordinate basis, or the components of γ.

Given a metric tensor $\gamma = \sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) dx^i dx^j$, a point $p \in I\!R^n$ and vectors $X_p = \sum_{i=1}^{n} \xi^i \partial_{x^i}|_p, Y_p = \sum_{i=1}^{n} \eta^i \partial_{x^i}|_p$, Theorem 5.2.4 gives,

$$\gamma_p(X_p, Y_p) = [X_p]^T [g_{ij}(p)][Y_p] = \sum_{i,j=1}^{n} g_{ij}(p) \xi^i \eta^j \tag{5.38}$$

where $[X_p] = [\xi^i], [Y_p] = [\eta^i]$ are the column vectors of the coefficients of X_p and Y_p in the coordinate basis.

Given two vector-fields $X = \sum_{i=1}^{n} \xi^i(\mathbf{x}) \partial_{x^i}, Y = \sum_{i=1}^{n} \eta^i(\mathbf{x}) \partial_{x^i}$ on $I\!R^n$ we can also evaluate

$$\gamma(X, Y) = [X]^T [g_{ij}(\mathbf{x})][Y] = \sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) \xi^i(\mathbf{x}) \eta^j(\mathbf{x}) \ \in C^\infty(I\!R^n). \tag{5.39}$$

A metric tensor γ on an open set $U \subset I\!R^n$ is defined exactly as for $I\!R^n$.

Example 5.4.3. Let $U \subset I\!R^3$ be the open set $U = \{(x, y, z) \in I\!R^3 \mid xz \neq 0\}$. Then

$$\gamma = \frac{1}{x^2} dx^2 + \frac{1}{x^2} dy^2 - 2\frac{y}{zx^2} dy dz + \left(\frac{y^2}{x^2 z^2} + \frac{1}{z^2} \right) dz^2, \tag{5.40}$$

is a metric tensor on U. The components of γ are

$$[g_{ij}(\mathbf{x})] = \begin{bmatrix} \frac{1}{x^2} & 0 & 0 \\ 0 & \frac{1}{x^2} & -\frac{y}{zx^2} \\ 0 & -\frac{y}{zx^2} & \frac{y^2}{x^2 z^2} + \frac{1}{z^2} \end{bmatrix}$$

Let $p = (1, 2, -1)$ and $X_p = (\partial_x + \partial_y - \partial_z)|_p, Y_p = (\partial_x + \partial_z)|_p$. Using equation 5.38 we have

$$\gamma_p(X_p, Y_p) = [1, 1, -1]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = -2.$$

If $X = z\partial_x + y\partial_y + x\partial_z$ and $Y = y\partial_x - x\partial_y$ then equation 5.39 gives

$$\boldsymbol{\gamma}(X, Y) = \frac{yz}{x^2} - \frac{y}{x} + \frac{y}{z}.$$

Example 5.4.4. Let $\boldsymbol{\gamma}$ be a metric tensor on $\mathbb{R}^2$. By equation 5.36 there exists functions $E, F, G \in C^\infty(\mathbb{R}^2)$ such that

$$\begin{aligned} \boldsymbol{\gamma} &= E\, dx \otimes dx + F(dx \otimes dy + dy \otimes dx) + G\, dy \otimes dy \\ &= E(dx)^2 + 2F dxdy + G(dy)^2 = Edx^2 + 2Fdxdy + Gdy^2. \end{aligned} \tag{5.41}$$

The matrix $[g_{ij}(\mathbf{x})$ of components of $\boldsymbol{\gamma}$ are then

$$[g_{ij}(\mathbf{x})] = \begin{bmatrix} E & F \\ F & G \end{bmatrix} \tag{5.42}$$

which is positive definite at each point in $\mathbb{R}^2$, since $\boldsymbol{\gamma}$ was assumed to be a metric tensor field.

Example 5.4.5. On $\mathbb{R}^n$ let

$$\boldsymbol{\gamma}^E = \sum (dx^i)^2. \tag{5.43}$$

The components of $\boldsymbol{\gamma}^E$ at a point p in the coordinate basis $\{\partial_{x^i}|_p\}_{1 \le i \le n}$ are

$$\boldsymbol{\gamma}^E(\partial_{x^i}|_p, \partial_{x^j}|_p) = \delta_{ij}.$$

The metric $\boldsymbol{\gamma}^E$ is called the *Euclidean metric tensor*. If $X_p, Y_p \in T_p\mathbb{R}^n$ then equation 5.38 gives

$$\boldsymbol{\gamma}^E_p(X_p, Y_p) = [X_p]^T[Y_p].$$

Example 5.4.6. Let $U = \{\ (x, y) \in \mathbb{R}^2 \mid y > 0\ \}$, and let

$$[g_{ij}] = \begin{bmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{bmatrix}. \tag{5.44}$$

This 2×2 matrix defines a metric tensor on U given by formula 5.37 by

$$\boldsymbol{\gamma} = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy) = \frac{dx^2 + dy^2}{y^2}.$$

Let $p = (1, 2) \in U$ and let

$$X_p = (2\partial_x - 3\partial_y)|_p \ , Y_p = (-\partial_x + \partial_y)|_p. \tag{5.45}$$

The real number $\boldsymbol{\gamma}_p(X_p, Y_p)$ can be computed using formula 5.38 or by expanding,

$$\boldsymbol{\gamma}(X_p, Y_p) = \frac{1}{4}\left(dx(X_p)dx(X_p) + dy(X_p)dy(X_p)\right) = \frac{1}{4}(2)(-1)+\frac{1}{4}(-3)(1) = -\frac{5}{4}$$

If the point were $p = (2, 5)$ with X_p, Y_p from equation 5.45 we would have

$$\boldsymbol{\gamma}(X_p, Y_p) = \frac{1}{25}(2)(-1) + \frac{1}{25}(-3)(1) = -\frac{1}{5}.$$

Notice how the computation depends on the point $p \in U$.

Example 5.4.7. Let $U \subset \mathbb{R}^2$ an open set, and let

$$[g_{ij}] = \begin{bmatrix} E & F \\ F & G \end{bmatrix}$$

where $E, F, G \in C^\infty(U)$ and this 2×2 matrix is positive definite at each point in U. This 2×2 matrix defines the metric tensor using equation 5.37 producing the metric in equation 5.41 on the open set $U \subset \mathbb{R}^2$.

Remark 5.4.8. In general relativity one is interested in symmetric bilinear forms $\gamma : V \times V \to \mathbb{R}$ which are non-degenerate (see exercise 9) but are not positive definite. A famous example of a metric tensor in general relativity is the Schwartzschild metric. This is the metric tensor whose coefficients in coordinates (t, r, θ, ϕ) are given by

$$[g_{ij}(t, r, \theta, \phi)] = \begin{bmatrix} -1 + \frac{2M}{r} & 0 & 0 & 0 \\ 0 & \frac{1}{1 - \frac{2M}{r}} & 0 & 0 \\ 0 & 0 & r^2 \sin^2 \phi & 0 \\ 0 & 0 & 0 & r^2 \end{bmatrix},$$

which written in terms of differentials is

$$ds^2 = \left(-1 + \frac{2M}{r}\right) dt^2 + \frac{1}{1 - \frac{2M}{r}} dr^2 + r^2 \sin^2 \phi d\theta^2 + r^2 d\phi^2.$$

This metric tensor is not a Riemannian metric tensor. It does not satisfy the positive definite criteria as a symmetric bilinear form on the tangent space at each point. It is however non-degenerate.

The Einstein equations in general relativity are second order differential equations for the coefficients of a metric $[g_{ij}]$, $1 \leq i, j, \leq 4$. The differential equations couple the matter and energy in space and time together with the second derivatives of the coefficients of the metric. The idea is that the distribution of energy determines the metric tensor $[g_{ij}]$. This then determines how things are measured in the universe. The Scwartzschild metric represents the geometry outside of a fixed spherically symmetric body.

Remark 5.4.9. You need to be careful about the term *metric* as it is used here (as in metric tensor). It is not the same notion as the term *metric* in topology! Often in differential geometry the term metric is used, instead of the full name metric tensor, which further confuses the issue. There is a relationship between these two concepts - see remark 5.4.13 below.

5.4.1 Arc-length

Let $\boldsymbol{\gamma} = \sum_{i,j=1}^n g_{ij}(\mathbf{x}) dx^i dx^j$ be a metric tensor on $I\!R^n$, and $\sigma : [a, b] \to I\!R^n$ a continuous curve on the interval $[a, b]$, and smooth on (a, b). Denote by $\sigma(t) = (x^1(t), x^2(t), \ldots, x^n(t))$ the components of the curve. At a fixed value of t, $\dot{\sigma}$ is the tangent vector (see equation 3.13), and it's length with respect to $\boldsymbol{\gamma}$ is compute using 5.38 to be

$$\sqrt{\boldsymbol{\gamma}(\dot{\sigma}, \dot{\sigma})} = \sqrt{\sum_{i,j=1}^n g_{ij}(\sigma(t)) \frac{dx^i}{dt} \frac{dx^j}{dt}}.$$

Integrating this function with respect to t gives

$$L(\sigma) = \int_a^b \sqrt{\boldsymbol{\gamma}(\dot{\sigma}, \dot{\sigma})} dt = \int_a^b \sqrt{\sum_{i,j=1}^n g_{ij}(\sigma(t)) \frac{dx^i}{dt} \frac{dx^j}{dt}}\, dt \tag{5.46}$$

which is the *arc-length* of σ with respect to the metric $\boldsymbol{\gamma}$. Note that the components $g_{ij}(\mathbf{x})$ of the metric tensor are evaluated along the curve σ.

Example 5.4.10. For the Euclidean metric tensor γ_E

$$L(\sigma) = \int_a^b \sqrt{\sum_i^n \left(\frac{dx^i}{dt}\right)^2}\, dt\,,$$

as expected.

Example 5.4.11. Compute the arc-length of the line

$$\sigma(t) = (t+1, 2t+1) \quad 0 \leq t \leq 1$$

with respect to the metric in equation (5.44). We get,

$$\int_0^1 \sqrt{\frac{1}{(2t+1)^2}(1+4))}\, dt = \frac{\sqrt{5}}{2}\log 3.$$

This is not the same as the arc-length using the Euclidean metric tensor which is found to be

$$\int_0^1 \sqrt{1+4}dt = \sqrt{5}.$$

Remark 5.4.12. Another name for a metric tensor is a "line-element field" (written ds^2), because the metric can be used to measure the length of a line.

Remark 5.4.13. We now explain a relationship between metric tensors and the term metric in topology. Recall a metric on a set U is a function $d : U \times U \to I\!R$ satisfying for all $x, y, z \in U$,

1. $d(x,y) = d(y,x)$,
2. $d(x,y) \geq 0$,
3. $d(x,y) = 0$ if and only if $y = x$,
4. $d(x,y) \leq d(x,y) + d(y,z)$ (the triangle inequality).

Let $\boldsymbol{\gamma}$ be a metric tensor on $I\!R^n$ and let $p, q \in I\!R^n$. Define $d(p,q)$ by

$$d(p,q) = \inf_\sigma \int_a^b \sqrt{\gamma(\dot\sigma, \dot\sigma)}dt \tag{5.47}$$

where $\sigma : [a,b] \to I\!R^n$ is a curve satisfying $\sigma(a) = p, \sigma(b) = q$. The function $d(p,q)$ in equation 5.47 defines a metric on $I\!R^n$.

5.4.2 Orthonormal Frames

Recall that a basis $\beta = \{u_i\}_{1\leq i \leq n}$ for an n-dimensional vector space V with inner product γ, is an orthonormal basis if

$$\gamma(u_i, u_j) = \delta_{ij}, \quad 1 \leq i, j \leq n.$$

The theory of orthonormal frames begins with an extension of the Gram-Schmidt process for constructing orthonormal basis on inner product spaces which we now recall.

Starting with an basis $\beta = \{v_i\}_{1\leq i\leq n}$ for the n-dimensional vector space V with inner product γ, the Gram-Schmidt process constructs by induction the set of vector $\tilde{\beta} = \{w_i\}_{1\leq i\leq n}$ by

$$\begin{aligned} w_1 &= v_1, \\ w_j &= v_j - \sum_{k=1}^{j-1} \frac{\gamma(v_j, w_k)}{\gamma(w_k, v_w)} w_k, \quad 2 \leq j \leq n. \end{aligned}$$

Theorem 5.4.14. *The set of vector $\tilde{\beta}$ form a basis for V and are orthogonal to each other*

$$\gamma(w_i, w_j) = 0 \quad i \neq j.$$

The proof of this is standard see [6]. From the set β' the final step in the Gram-Schmidt process is to let

$$u_i = \frac{1}{\sqrt{\gamma(w_i, w_i)}} w_i \quad 1 \leq i \leq n$$

so that $\beta = \{u_i\}_{1\leq i\leq n}$ is an orthonormal basis for V.

We now apply the Gram-Schmidt algorithm in the setting of a metric tensor. Let $\beta' = \{X_i\}_{1\leq i\leq n}$ be n vector-fields on $\mathbb{R}^n$ which are linearly independent at each point. A basis of $T_p\mathbb{R}^n$ is also called a *frame*, and the collection of vector-fields β' a *frame field* or moving frame. An orthonormal basis of $T^p\mathbb{R}^n$ is called an orthonormal frame and if the vector-fields $\{X_i\}_{1\leq i\leq n}$ satisfy

$$\boldsymbol{\gamma}(X_i, X_j) = \delta_{ij}$$

and so are are orthonormal at each point in $\mathbb{R}^n$, then the collection β' is called an *orthonormal frame field*.

We now show how to construct orthonormal frame fields using the Gram-Schmidt process. Let γ be a Riemannian metric tensor on $I\!R^n$. Construct the following vector fields from the frame field $\beta' = \{X_i\}_{1\leq i\leq n}$,

$$
\begin{aligned}
&Y_1 = X_1, \\
&Y_j = X_j - \sum_{k=1}^{j-1} \frac{\gamma(X_j, Y_k)}{\gamma(Y_k, Y_k)} Y_k, \quad 2 \leq j \leq n.
\end{aligned}
\tag{5.48}
$$

As in Theorem 5.4.14

Theorem 5.4.15. *The vector-fields $\{Y_i\}_{1\leq i\leq n}$ are smooth and linearly independent at each point in $I\!R^n$, and are mutually orthogonal at each point*

$$\gamma(Y_i, Y_j) = 0, \quad i \neq j.$$

Again as in the final step of the Gram-Schmidt process let

$$Z_i = \frac{1}{\sqrt{\gamma(Y_i, Y_i)}} Y_i \quad 1 \leq i \leq n,$$

and the set $\beta = \{Z_i\}_{1\leq i\leq n}$ is a set of vector-fields that form an orthonormal basis with respect to γ_p for $T_p I\!R^n$ for each point $p \in I\!R^n$, and so form an orthonormal frame field.

Example 5.4.16. Let γ be the metric tensor on $U = \{(x, y, z) \in I\!R^3 \mid xz \neq 0\}$ given by

$$\gamma = \frac{1}{x^2} dx^2 + \frac{1}{x^2} dy^2 - 2\frac{y}{zx^2} dydz + \left(\frac{y^2}{x^2z^2} + \frac{1}{z^2} \right) dz^2.$$

We find an orthornomal frame field starting with the coordinate frame $\partial_x, \partial_y, \partial_z$ and then using equation 5.48. The first vector-field is $Y_1 = \partial_x$, the second is

$$Y_2 = \partial_y - \frac{\gamma(\partial_y, Y_1)}{\gamma(Y_1, Y_1)} Y_1 = \partial_y,$$

and the third is

$$
\begin{aligned}
Y_3 &= \partial_z - \frac{\gamma(\partial_z, Y_1)}{\gamma(Y_1, Y_1)} Y_1 - \frac{\gamma(\partial_z, Y_2)}{\gamma(Y_2, Y_2)} Y_2 \\
&= \partial_z - 0\partial_x + \frac{y}{z}\partial_y.
\end{aligned}
$$

Finally the resulting orthonormal frame field is

$$Z_1 = x\partial_x, \ Z_2 = x\partial_y, \ Z_3 = z\partial_z + y\partial_y. \tag{5.49}$$

5.5 Raising and Lowering Indices and the Gradient

Given a function $f \in C^\infty(\mathbb{R}^n)$ its differential df (see equation 5.10) defines at each point $p \in \mathbb{R}^n$ an element of $T_p^*\mathbb{R}^n$. This is **not** the gradient of f. We will show that when given a metric tensor-field $\boldsymbol{\gamma}$ on $\mathbb{R}^n$, it can be used to convert the differential one-form df into a vector-field X, which is called the gradient of f (with respect to $\boldsymbol{\gamma}$). This highlights the fact that in order to define the gradient of a function a metric tensor is needed.

Again we need some linear algebra. Suppose that B is a bilinear form on V (see exercise 9). In particular B could be an inner product. Now as a function B requires two vectors as input. Suppose we fix one vector in the input like $B(v, -)$ and view this as a function of one vector. That is, let $v \in V$ be a fixed vector and define the function $\alpha_v : V \to \mathbb{R}$ by

$$\alpha_v(w) = B(v, w) \quad for\ all\ w \in V. \tag{5.50}$$

Alternatively we could have defined a function $\tilde{\alpha}_v : V \to \mathbb{R}$ by $\tilde{\alpha}_v(w) = B(w, v)$ *for all* $w \in V$, which if B is symmetric would be the same function α_v defined in 5.50.

The notation α_v is used to emphasize that the form α_v depends on the initial choice of $v \in V$. Let's check $\alpha_v \in V^*$,

$$\begin{aligned}\alpha_v(aw_1 + w_2) &= B(v, aw_1 + w_2)\\ &= a\,B(v, w_1) + B(v, w_2)\\ &= a\alpha_v(w_1) + \alpha_v(w_2) \quad a \in \mathbb{R}, w_1, w_2 \in V.\end{aligned}$$

Therefore $\alpha \in V^*$. From this we see that given $v \in V$ we can construct from v an element $\alpha_v \in V^*$ using the bilinear form B. That is we can use B to convert a vector to an element of the dual space α_v called *the dual of* v *with respect to* B.

From this point of view the bilinear form B allows us to define a function $T_B : V \to V^*$ by

$$T_B(v)(w) = B(v, w), \quad for\ all\ w \in V. \tag{5.51}$$

How does T_B depend on V? Let's compute

$$\begin{aligned}T_B(av_1 + v_2)(w) &= B(av_1 + v_2, w) \quad \text{by } 5.51\\ &= a\,B(v_1, w) + B(v_2, w) = \left(aT_B(v_1) + T_B(v_2)\right)(w).\end{aligned}$$

Therefore $T_B(av_1+v_2) = aT_B(v_2)+T_B(v_2)$ and T_B is a linear transformation.

We now work out T_B in a basis!

Proposition 5.5.1. *Let $\beta = \{v_i\}_{1\leq i\leq n}$ for V and let $\beta^* = \{\alpha^i\}_{1\leq i\leq n}$ be the dual basis. The matrix representation of $T_B : V \to V^*$ is*

$$[T_B]_\beta^{\beta^*} = [g_{jk}]$$

where $g_{ij} = B(v_i, v_j)$.

Proof. We begin by writing

$$T_B(v_i) = \sum_{k=1}^{n} g_{ik}\alpha^k \tag{5.52}$$

and determine g_{ik}. By equation 5.51,

$$T_B(v_i)(v_j) = B(v_i, v_j)$$

and therefore evaluating equation 5.52 on v_j gives

$$B(v_i, v_j) = \left(\sum_{k=1}^{n} g_{ik}\alpha^k\right)(v_j) = g_{ij}.$$

This proves the theorem. □

Now let $v \in V$ and $\alpha_v = T_B(v)$ which we write in the basis β and dual basis β^* as

$$\begin{aligned} v &= \sum_{i=1}^{n} a^i v_i, \text{ and} \\ \alpha_v &= T_B(v) = \sum_{i=1}^{n} b_i\, \alpha^i, \quad a^i, b_i \in I\!R, \end{aligned} \tag{5.53}$$

where we assume a^i are known since v is given, and we want to find b_j in terms of a^i. It follows immediately from Lemma 2.1.6 and Proposition that the coefficients b_j of the image form α_v are

$$b_j = \sum_{i=1}^{n} g_{ij}a^i. \tag{5.54}$$

We can then write

$$T_B(v) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} g_{ij} a^i \right) \alpha^j. \tag{5.55}$$

The form $\alpha_v = T_B(v)$ in equation 5.55 is called the dual of v with respect to B, and this process is also sometimes called "lowering the index" of v with B.

Another way to compute $T_B(v)$ is given by the following corollary (see Theorem 5.3.2).

Corollary 5.5.2. *Let*

$$B = \sum_{i,j=1}^{n} g_{ij} \alpha^i \otimes \alpha^j$$

where $g_{ij} = B(v_i, v_j)$. *Then*

$$T_B(v) = \sum_{i,j=1}^{n} g_{ij} \alpha^i(v) \alpha^j. \tag{5.56}$$

Example 5.5.3. Let B be the symmetric non-degenerate bilinear form from example 5.2.2

$$B\left(\begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}, \begin{bmatrix} y^1 \\ y^2 \\ y^3 \end{bmatrix} \right) = 2x^1y^1 - x^1y^2 + x^1y^3 - x^2y^1 - x^2y^2 + x^3y^1 + x^3y^3. \tag{5.57}$$

Let $v = e_1 - e_2 + 2e_3$, then we compute $\alpha_v = T_B(v)$ by noting

$$\alpha_v(e_1) = B(v, e_1) = 5, \ \alpha_v(e_2) = 0 \ \alpha_v(e_3) = B(v, e_3) = 3.$$

Therefore $\alpha_v(xe_1 + ye_2 + ze_3) = 5x + 3z$.

Theorem 5.5.4. *If the bi-linear form B is non-degenerate then the linear transformation T_B in equation 5.51 is an isomorphism.*

Proof. Suppose $T_B(v) = 0$, where $0 \in V^*$, then

$$0 = T_B(v)(w) = B(v, w) \quad for \ all \ w \in V.$$

Since B is non-degenerate (see exercise 8), this implies $v = 0$. Which proves the lemma. □

Suppose from now on that B is non-degenerate. Let $T_B^{-1} : V^* \to V$ be the inverse of T_B, and let $\alpha \in V^*$ then by equation 5.51

$$T_B \circ T_B^{-1}(\alpha)(w) = B(T_B^{-1}(\alpha), w), \quad for\ all\ w \in V.$$

However $T_B \circ T_B^{-1} = I$ the identity, and therefore

$$\alpha(w) = B(T_B^{-1}(\alpha), w), \quad for\ all\ w \in V. \tag{5.58}$$

Let $\beta = \{v_i\}_{1 \leq i \leq n}$ be a basis for V, $\beta^* = \{\alpha^i\}_{1 \leq i \leq n}$ the dual basis, and let $\alpha = \sum_{i=1}^n b_i \alpha^i$. We now find $v = T_B^{-1}(\alpha)$ in the basis β. By Proposition 5.5 $[T_B]_\beta^{\beta^*} = [g_{ij}]$. Using Proposition 2.3.7, $[T_B^{-1}]_{\beta^*}^\beta = ([T_B]_\beta^{\beta^*})^{-1}$, and so let $[g^{ij}]$ denote the inverse matrix of $[g_{ij}] = [T_B]_\beta^{\beta^*}$. Utilizing Lemma 2.1.6 the coefficients a^j of the image vector $T_B(\alpha)$ in terms of the coefficients b_i of α are given by

$$a^j = \sum_{i=1}^n g^{ij} b_i, \tag{5.59}$$

and

$$v = T_B^{-1}(\alpha) = \sum_{j=1}^n \left(\sum_{i=1}^n g^{ij} b_i \right) v_j. \tag{5.60}$$

The vector v in formula 5.59 is called the dual of α with respect to B. The process is also sometimes called "raising the index" with B.

Example 5.5.5. We continue with the example 5.5.3 and we compute the dual of $\alpha(xe_1 + ye_2 + ze_3) = 3x - y + 2z$ with respect to B (raise the index). From equation 5.57, we find the matrix $[g^{ij}] = [g_{ij}]^{-1}$ is

$$[g^{ij}] = \frac{1}{4} \begin{bmatrix} 1 & -1 & -1 \\ -1 & -3 & 1 \\ 1 & -1 & 3 \end{bmatrix}.$$

Therefore

$$[T_B^{-1}(\alpha)] = \frac{1}{2}[1, 1, 5]$$

and

$$T_B^{-1}(\alpha) = \frac{1}{2}e_1 + \frac{1}{2}e_2 + \frac{5}{2}e_3.$$

In summary using $T_B : V \to V^*$ we can map vectors to dual vectors, and using the inverse $T_\gamma^{-1} : V^* \to V$ we can map dual vectors to vectors. This will be the essential part of the gradient.

Let $\boldsymbol{\gamma}$ be a Riemannian metric on $I\!R^n$, and suppose X is a vector-field. We can convert the vector-field X to a differential one-form α_X by using the formula 5.51 at each point

$$\alpha_X(Y_p) = \boldsymbol{\gamma}_p(X_p, Y_p), \quad for\ all\ T_p \in T_p I\!R^n. \tag{5.61}$$

We then define $T_{\boldsymbol{\gamma}}$ from vector-fields on $I\!R^n$ to differential one-forms on $I\!R^n$ by

$$T_{\boldsymbol{\gamma}}(X)(Y_p) = \boldsymbol{\gamma}(X(p), Y_p) \quad for\ all Y_p \in T_p I\!R^n.$$

If $X = \sum_{i=1}^n \xi^i(x)\partial_{x^i}$, and the metric components are

$$g_{ij}(\mathbf{x}) = \boldsymbol{\gamma}(\partial_{x^i}, \partial_{x^j})$$

then formula 5.55 or equation 5.56 applied point-wise for lowering the index gives

$$\alpha_X = T_{\boldsymbol{\gamma}}(X) = \sum_{j=1}^n \left(\sum_{i=1}^n g_{ij}(\mathbf{x}) dx^i(X) \right) dx^j = \sum_{j=1}^n \left(\sum_{i=1}^n g_{ij}(\mathbf{x}) \xi^i(\mathbf{x}) \right) dx^j. \tag{5.62}$$

The differential one-form α_X is the called dual of the vector-field X with respect to the metric $\boldsymbol{\gamma}$.

The function $T_{\boldsymbol{\gamma}}$ is invertible (because $\boldsymbol{\gamma}$ is non-degenerate at each point), and given a differential form $\alpha = \sum_{i=1}^n \alpha_i(\mathbf{x}) dx^i$, then its dual with respect to the metric $\boldsymbol{\gamma}$ is given by the raising index formula 5.60

$$X = T_{\boldsymbol{\gamma}}^{-1}(\alpha) = \sum_{j=1}^n \left(\sum_{i=1}^n g^{ij}(\mathbf{x}) \alpha_i(\mathbf{x}) \right) \partial_{x^j} \tag{5.63}$$

where again $[g^{ij}(\mathbf{x})]$ is the inverse of the matrix $[g_{ij}(\mathbf{x})]$.

Example 5.5.6. Let

$$\boldsymbol{\gamma} = \frac{1}{1+e^x}(dx^2 + dy^2)$$

be a Riemannian metric on $I\!R^2$, and let $X = x\partial_x + y\partial_y$. The dual of X with respect to $\boldsymbol{\gamma}$ is computed using 5.62

$$\alpha = \frac{1}{1+e^x} dx(X)dx + \frac{1}{1+e^x} dy(Y)dy = \frac{1}{1+e^x} xdx + \frac{1}{1+e^x} ydy.$$

If $\alpha = ydx - xdy$ then by equation 5.63 the dual of α with respect to $\boldsymbol{\gamma}$ is

$$X = y(1+e^x)\partial_x - x(1+e^x)\partial_y.$$

More generally if

$$\boldsymbol{\gamma} = Edx^2 + 2Fdxdy + Gdy^2 ,$$

is a metric tensor on $I\!R^2$ and and $X = a\partial_x + b\partial_y$ a vector-field on U then the dual of X is computed using 5.62 to be

$$T_{\boldsymbol{\gamma}}(X) = (aE + bF)dx + (aF + bG)dy.$$

If $\alpha = adx + bdy$ is a differential one-form then by equation 5.63 its dual is

$$T_{\boldsymbol{\gamma}}^{-1}(\alpha) = \frac{1}{EG - F^2}\left((aG - bF)\partial_x + (bE - aF)\partial_y\right)$$

where we have used,

$$[g^{ij}] = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

Finally we can define the gradient of a function. Let $f \in C^\infty(U)$, and $\boldsymbol{\gamma}$ a Riemannian metric tensor on U (or at least a non-degenerate bilinear form-field). Let

$$\text{grad}(f) = \{\ \textit{the dual of } df \textit{ with respect to } \boldsymbol{\gamma}\ \} = T_{\boldsymbol{\gamma}}^{-1}(df). \tag{5.64}$$

By equation 5.63 the formula for $\text{grad}(f)$ is

$$\text{grad}(f) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} g^{ij}(\mathbf{x})\frac{\partial f}{\partial x^i}\right)\partial_{x^j} \tag{5.65}$$

Example 5.5.7. Continuing from Example 5.5.6 with the metric tensor $\boldsymbol{\gamma} = (1+e^x)^{-1}(dx^2 + dy^2)$ on $I\!R^2$, if $f \in C^\infty(I\!R^2)$ then $df = f_x dx + f_y dy$ and its dual with respect to $\boldsymbol{\gamma}$ is

$$\text{grad}(f) = (1+e^x)f_x\partial_x + (1+e^x)f_y\partial_y.$$

Example 5.5.8. Let (x, y, z) be coordinates on $\mathbb{R}^3$, and

$$\boldsymbol{\gamma} = dx^2 + dy^2 - 2xdydz + (1 + x^2)dz^2.$$

The components of $\boldsymbol{\gamma}$ in the coordinate basis are

$$[g_{ij}(\mathbf{x})] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -x \\ 0 & -x & 1 + x^2 \end{bmatrix},$$

while

$$[g^{ij}(\mathbf{x})] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 + x^2 & x \\ 0 & x & 1 \end{bmatrix}.$$

Given $f(x, y, z)$ then its gradient with respect to this metric is

$$f_x \partial_x + \left((1 + x^2)f_y + xf_z\right)\partial_y + (xf_y + f_z)\partial_z.$$

Example 5.5.9. If

$$\boldsymbol{\gamma}_E = \sum_{i=1}^{n} (dx^i)^2$$

is the Euclidean metric tensor then

$$\mathrm{grad}(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \partial_{x^i}.$$

In this case the coefficients of the gradient of a function are just the partial derivatives of the function. This is what occurs in a standard multi-variable calculus course.

Let $U_p \in T_p\mathbb{R}^n$ be a unit vector with respect to a metric tensor $\boldsymbol{\gamma}$. The rate of change of $f \in C^\infty(\mathbb{R}^n)$ at the point p in the direction U_p is

$$U_p(f).$$

As in ordinary calculus, the following is true.

Theorem 5.5.10. *Let $f \in C^\infty(\mathbb{R}^n)$ with $\mathrm{grad}_p(f) \neq \mathbf{0}$. Then $\mathrm{grad}_p(f)$ is the direction at p in which f increases the most rapidly, and the rate of change is*

$$||\mathrm{grad}_p(f)||_{\boldsymbol{\gamma}}.$$

Proof. We begin by using equation 5.58 with $\boldsymbol{\gamma}_p : T_p\mathbb{R}^n \times T_p\mathbb{R}^n \to \mathbb{R}$ and definition 5.64 to get

$$U_p(f) = df_p(U_p) = \boldsymbol{\gamma}_p(T_{\boldsymbol{\gamma}}^{-1}(df_p), U_p) = \boldsymbol{\gamma}_p(\mathrm{grad}_p(f), U_p).$$

The Cauchy-Schwartz inequality 5.26 applied to this formula then gives,

$$|U_p(f)| = |\boldsymbol{\gamma}(\mathrm{grad}_p(f), U_p)| \leq ||\mathrm{grad}(f)||_{\boldsymbol{\gamma}}.$$

The result follows by noting the maximum rate $||\mathrm{grad}_p(f)||_{\boldsymbol{\gamma}}$ of $|U_p(f)|$ is obtained when $U_p = ||\mathrm{grad}_p(f)||_{\boldsymbol{\gamma}}^{-1}\mathrm{grad}_p(f)$. □

This next bit is not necessary for doing the assignment but is another way to define the raising the index procedure. Since T_γ is an invertible linear transformation it can be used to define a non-degenerate bilinear form $\gamma^* : V^* \times V^* \to \mathbb{R}$, defined by

$$\gamma^*(\alpha, \beta) = \gamma(T_\gamma^{-1}(\alpha), T_\gamma^{-1}(\beta)).$$

Theorem 5.5.11. *The function γ^* is bilinear, non-degenerate. If γ is symmetric, then so is γ^*. If γ is positive definite then so is γ^*. Furthermore if $\{v_i\}_{1\leq i\leq n}$ is a basis for V, and $\{\alpha^i\}_{1\leq i\leq n}$ is the dual basis then,*

$$\gamma^*(\alpha^i, \alpha^j) = g^{ij}$$

where $[g^{ij}]$ is the inverse matrix of $[g_{ij}] = [\gamma(v_i, v_j)]$.

If $\boldsymbol{\gamma}$ is a metric tensor on $\mathbb{R}^n$ then γ^* is called the contravariant form of the metric $\boldsymbol{\gamma}$. The raising the index procedure (and the gradient) can then be defined in terms of γ^* using the isomorphism $T_{\gamma^*} : V^* \to V$

$$T_{\gamma^*}(\alpha)(\tau) = \gamma^*(\alpha, \tau).$$

In this formula we have identified $(V^*)^* = V$.

Remark 5.5.12. Applications of the gradient in signal processing can be found in [5], [1], [2].

5.6 A tale of two duals

Given a vector $v \in V$, where V is a finite dimensional vector-space, there is no notion of the dual of v unless there is an inner product γ on V. In this case the inner product γ can be used to define the function T_γ as in equation 5.51 giving $T_\gamma(v) \in V^*$ which is the dual with respect to γ.

The matter is quite different if we are given a basis $\beta = \{v_i\}_{1\leq i\leq n}$ for V. We then have the dual basis $\beta^* = \{\alpha^i\}_{1\leq i\leq n}$ defined by (see equation 5.3),

$$\alpha^i(v_j) = \delta^i_j, \quad 1 \leq i, j \leq n.$$

Suppose we also have γ as an inner product on V, and let $\tilde{\beta}^* = \{\sigma^i\}_{1\leq i\leq n}$ be the forms dual with respect to γ,

$$\sigma^i = T_\gamma(v_i), \quad 1 \leq i \leq n.$$

These two duals are related by the following theorem.

Theorem 5.6.1. *The set $\tilde{\beta}^*$ is a basis for V^* and $\alpha^i = \sigma^i$ if and only if v_i is an orthonormal basis.*

Proof. The fact that $\tilde{\beta}^*$ is a basis will be left as an exercise. Suppose that β is an orthornormal basis, then

$$\sigma^i(v_j) = T_\gamma(v_i)(v_j) = \gamma(v_i, v_j) = \delta^i_j.$$

However by definition $\alpha^i(v_j) = \delta^i_j$, and so for each $i = 1, \ldots, n$, α^i and σ^i agree on a basis, and hence are the same elements of V^*. This proves the sufficiency part of the theorem.

Finally assume that $\alpha^i = \sigma^i$, $1 \leq i \leq n$. Then

$$\delta^i_j = \alpha^i(v_j) = \sigma^i(v_j) = \gamma(v_i, v_j),$$

and β is an orthonormal basis. □

Theorem 5.6.1 has an analogue for frames fields. Let $\{X_i\}_{1\leq i\leq n}$ be a frame field on $\mathbb{R}^n$. The algebraic dual equations

$$\alpha^i(X_j) = \delta^i_j \quad \textit{for all } p \in \mathbb{R}^n, \tag{5.66}$$

define α^i, $1 \leq i \leq n$ as a field of differential one-forms, which form a basis for $T_p^* I\!R^n$ for each $p \in I\!R^n$. Given a Riemmanian metric tensor $\boldsymbol{\gamma}$ define the differential one-forms as in equation 5.61,

$$\sigma^j = T_{\boldsymbol{\gamma}}(X_j) = \boldsymbol{\gamma}(X_j, -), \quad 1 \leq j \leq n. \tag{5.67}$$

We then have the field version of Theorem 5.6.1.

Corollary 5.6.2. *The one-form fields $\{\sigma^i\}_{1 \leq i \leq n}$ from equation 5.67 define a basis for $T_p^* I\!R^n$ for each point $p \in I\!R^n$. The fields satisfy $\alpha^i = \sigma^i, 1 \leq i \leq n$, if and only if X_i is an orthonormal frame field.*

Example 5.6.3. In equation 5.68 of example 5.4.16 we found the orthonormal frame field

$$Z_1 = x\partial_x, \ Z_2 = x\partial_y, \ Z_3 = z\partial_z + y\partial_y. \tag{5.68}$$

for the metric tensor

$$\boldsymbol{\gamma} = \frac{1}{x^2}dx^2 + \frac{1}{x^2}dy^2 - 2\frac{y}{zx^2}dydz + \left(\frac{y^2}{x^2z^2} + \frac{1}{z^2}\right)dz^2$$

on $U = \{(x, y, z) \in I\!R^3 \mid xz \neq 0\}$. The algebraic dual defined in equation 5.66 of Z_1, Z_2, Z_3 is easily computed by using Corollary 5.6.2 by taking the dual with respect to $\boldsymbol{\gamma}$. We find

$$\alpha^1 = T_{\boldsymbol{\gamma}}(Z_1) = \frac{1}{x}dx, \quad \alpha^2 = T_{\boldsymbol{\gamma}}(Z_2) = \frac{1}{x}dy, \quad \alpha^3 = T_{\boldsymbol{\gamma}}(Z_3) = \frac{1}{z}dz.$$

5.7 Exercises

1. Let $V = \mathbb{R}^3$ and $T \in V^*$. Prove there exists $a, b, c \in \mathbb{R}$ such that

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = ax + by + cz.$$

2. Let $X_p, Y_p \in T_p\mathbb{R}^2$, with $p = (1 - 2)$ be

$$X_p = (2\partial_x + 3\partial_y)|_p, \ Y_p = (3\partial_x + 4\partial_y)|_p.$$

   Compute the dual basis to $\beta = \{X_p, Y_p\}$

3. Let $f = 2xyz$, let $p = (1, 1, 1)$, and $X_p = (-3\partial_x + \partial_y + \partial_z)|_p$.

   (a) Compute $df_p(X_p)$.

   (b) Find df in the coordinate basis.

   (c) Let $g = x^2 + y^2 + z^2$. Are dg and df linear dependent at any point?

4. Show that $\mathbf{B}(V)$ the space of bilinear functions on V with addition and scalar multiplication defined by equation 5.19 is a vector-space. (Do not assume that V is finite dimensional.)

5. Show that $\mathbf{S}(V) \subset \mathbf{B}(V)$ the symmetric bilinear functions, form a subspace. (Do not assume that V is finite dimensional.)

6. Prove corollary 5.2.5.

7. Finish the proof of Theorem 5.3.2 by showing $\{\alpha^i \otimes \alpha^j\}_{1 \leq i,j \leq n}$ is a linearly independent set.

8. A bilinear function $B : V \times V \to \mathbb{R}$ is *non-degenerate* if

$$B(v, w) = 0 \quad for\ all\ w \in V$$

   then $v = 0$.

   (a) Prove that an inner product on V is non-degenerate.

(b) Given a basis $\beta = \{v_i\}_{1\leq i\leq n}$ for V, prove that B is non-degenerate if and only if the matrix $[g_{ij}] = [B(v_i, v_j)]$ is invertible.

9. Let $B : I\!R^3 \times I\!R^3 \to I\!R$ be the bilinear function from example 5.2.2,

   (a) Show B is a symmetric non-degenerate bilinear function on $I\!R^3$.

   (b) Is B positive definite?

   (c) Compute α_v the dual of $v = -2e_1 + e_2 + e_3$ with respect to B as defined in equation 5.50.

   (d) Compute the dual of the form $\alpha(xe_1 + ye_2 + ze_3) = 4x - 3y + z$ with respect to B (raise the index). (Answer = $(3, 0, -2)$).

10. Let $\eta : I\!R^2 \times I\!R^2 \to I\!R$ be the function

$$\eta(\mathbf{x}, \mathbf{y}) = x^1y^2 + x^2y^1,$$

where $\mathbf{x} = (x^1, x^2), \mathbf{y} = (y^1, y^2)$.

   (a) Is η a symmetric non-degenerate bilinear function on $I\!R^2$? If so, is η positive definite ?

   (b) Let $\beta = \{v_1 = (1, 1), v_2 = (1, -1)\}$ write η as a linear combination of tensor products of the dual basis $\beta^* = \{\alpha^1, \alpha^2\}$.

   (c) Compute α_{v_1} and α_{v_2} as defined in equation 5.50 where v_1, v_2 are from part (b) (lower the index of v_1 and v_2). Compare to part (b).

   (d) Compute the dual of the $\alpha(xe_1 + ye_2) = 4x - 3y$ with respect to η (raise the index).

11. Let γ be a symmetric bilinear forms and $\beta = \{v_i\}_{1\leq i\leq n}$ a basis for V, and $\beta^* = \{\alpha^j\}_{1\leq j\leq n}$ a basis for V^*.

   (a) Show that

$$\gamma = \sum_{1\leq i,j\leq n} \gamma(v_i, v_j)\alpha^i\alpha^j,$$

   where $\alpha^i\alpha^j$ is given in equation 5.33.

   (b) Show that $\Delta_s = \{\alpha^i\alpha^j\}_{1\leq i\leq j\leq n}$ forms a basis for $\mathbf{S}(V)$, the symmetric bilinear forms on V.

12. For the metric tensor on $\mathbb{R}^3$ given by

$$\boldsymbol{\gamma} = dx^2 + dy^2 - 2x dy dz + (1 + x^2) dz^2,$$

   (a) compute the dual of the vector-field $z\partial_x + y\partial_y + x\partial_z$ with respect to $\boldsymbol{\gamma}$ (lower the index).
   (b) Compute the dual of the differential form $ydx + zdy - (1 + x^2)dz$ with respect to the metric $\boldsymbol{\gamma}$ (raise the index).
   (c) Find an orthonormal frame field and its dual. (Hint: See Corollary 5.6.2)

13. Let $U = \{(x, y) \mid y > 0\}$ with metric tensor

$$\boldsymbol{\gamma} = \frac{1}{y^2}(dx^2 + dy^2)$$

   (a) Compute the arc-length of a "straight line" between the points $(0, \sqrt{2})$ and $(1, 1)$.
   (b) Compute the arc-length of a circle passing through the points $(0, \sqrt{2})$ and $(1, 1)$ which has its center on the x-axis. Compare to part (a). Hint: You will need to find the circle.
   (c) Find an orthonormal frame field and its dual.
   (d) Find the gradient of $f \in C^\infty(U)$.

14. For the metric

$$\boldsymbol{\gamma} = d\phi^2 + \sin^2\phi \, d\theta^2$$

   on the open set $0 < \phi < \pi, \ 0 < \theta < 2\pi$, find

   (a) an orthonormal frame field and its dual.
   (b) Compute the gradient of $f(\theta, \phi)$.

Chapter 6

The Pullback and Isometries

6.1 The Pullback of a Differential One-form

Recall that in Chapter 4 that a function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ induces a linear transformation at each point $p \in \mathbb{R}^n$, $\Phi_{*,p} : T_p\mathbb{R}^n \to T_q\mathbb{R}^m$, where $q = \Phi(p)$, defined on derivations by

$$(\Phi_{*,p}X_p)(g) = X_p(g \circ \Phi) \quad g \in C^\infty(q).$$

If $X_p = \sum_{i=1}^n \xi^i \partial_{x^i}|_p, \ \xi^i \in \mathbb{R}$ then

$$\Phi_{*,p}X_p = \sum_{j=1}^n \sum_{i=1}^n \xi^i \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \left.\frac{\partial}{\partial y^a}\right|_q .$$

The map $\Phi_{*,p} : T_p\mathbb{R}^n \to T_q\mathbb{R}^m$ induces a map Φ_q^* going in the *other direction*,

$$\Phi_q^* : T_q^*\mathbb{R}^m \to T_p^*\mathbb{R}^n$$

on the dual space. The definition of the function Φ_q^* is easy once we examine the linear algebra.

Let V, W be real vector-spaces and $T : V \to W$ a linear transformation. There exists a map $T^t : W^* \to V^*$ defined as follows. If $\tau \in W^*$ then $T^*(\tau) \in V^*$ is defined by its value on $v \in V$ through

$$T^t(\tau)(v) = \tau(T(v)) \quad for\ all v \in V. \tag{6.1}$$

Lemma 6.1.1. *If $\tau \in W^*$ then $T^t(\tau) \in V^*$. Furthermore $T^t : W^* \to V^*$ is a linear transformation.*

Proof. The first part of this lemma is proved by showing that $T^t(\tau)$ is a linear function of v in equation 6.1. Suppose $v_1, v_2 \in V$, and $c \in I\!R$, then

$$
\begin{aligned}
T^t(\tau)(cv_1 + v_2) &= \tau(T(cv_1 + v_2)) \\
&= \tau(cT(v_1) + T(v_2)) \\
&= c\tau(T(v_1)) + \tau(T(v_2)) \\
&= cT^t(\tau)(v_1) + T^t(\tau)(v_2).
\end{aligned} \tag{6.2}
$$

Therefore $T^t(\tau) \in V^*$. The proof that T^t is linear is an exercise. □

To be more concrete about what T^t it is useful to write it in a basis. Suppose that V and W are finite dimensional vector-spaces of dimension n and m respectively, and that $\beta = \{v_i\}_{1 \le i \le n}$ is a basis for V, and $\gamma = \{w_a\}_{1 \le a \le n}$ is a basis for W. Let $A = [T]^\gamma_\beta$ be the matrix representation of T then A is the $m \times n$ matrix determined by the equations (see equation 2.2)

$$
T(v_i) = \sum_{a=1}^{m} A^a_i w_a. \tag{6.3}
$$

Furthermore if $v = \sum_{i=1}^n c^i v_i$ then the coefficients of the image vector $T(v)$ are by Lemma 2.1.6 or equation 2.10,

$$
[T(v)]_\gamma = [\sum_{i=1}^{n} A^a_i c^i] = A[v]_\beta.
$$

Now let $\beta^* = \{\alpha^i\}_{1 \le i \le n}$ be the basis of V^* which is the dual basis of β for V, and $\gamma = \{\tau^a\}_{1 \le a \le n}$ the basis of W^* which is the dual basis of γ for W^* as defined in equation 5.3. The matrix representation of function the $T^t : W^* \to V^*$ will now be computed in the basis γ^*, β^*. Let $B = [T^t]^{\beta^*}_{\gamma^*}$ which is an $n \times m$ matrix, which is determined by

$$
T^t(\tau^a) = \sum_{i=1}^{n} B^a_i \alpha^i. \tag{6.4}
$$

By evaluate the right side of equation 6.4 on $v_k \in \beta$ we get

$$
(\sum_{i=1}^{n} B^a_i \alpha^i)(v_k) = \sum_{i=1}^{n} B^a_i \alpha^i(v_k) = \sum_{i=1}^{n} B^a_i \delta^i_k = B^a_k.
$$

Therefore equation 6.4 gives

$$
\begin{aligned}
B^a_k &= T^t(\tau_a)(v_k) \\
&= \tau^a(T(v_k)) && \text{by equation 6.1} \\
&= \tau^a(\sum_{i=1}^n A^b_k w_b) && \text{by equation 6.3} \\
&= \sum_{i=1}^n A^b_k \tau^a(w_b) \\
&= \sum_{i=1}^n A^b_k \delta^a_b \\
&= A^a_k \,.
\end{aligned}
\tag{6.5}
$$

Therefore equation 6.5 gives

$$
\begin{aligned}
T(v_i) &= \sum_{a=1}^m A^a_i w_a, \\
T^t(\tau^a) &= \sum_{i=1}^n A^a_i \alpha^i \,.
\end{aligned}
\tag{6.6}
$$

Suppose $\tau \in W^*$ and $\tau = \sum_{a=1}^m c_a \tau^a$. We then write out what are the coefficients of $T^t(\tau)$ in the basis β^* by computing,

$$
\begin{aligned}
T^t(\sum_{a=1}^m c_a \tau^a) &= \sum_{a=1}^m c_a T^t(\tau^a) \\
&= \sum_{a=1}^m c_a (\sum_{i=1}^n A^a_i \alpha^i) && \text{by equation 6.6} \\
&= \sum_{i=1}^n (\sum_{a=1}^m c_a A^a_i) \alpha^i \,.
\end{aligned}
\tag{6.7}
$$

In other words the coefficients of the image $[T^t(\tau)]_{\beta^*}$ are the row vector we get by multiplying A on the left by the row vector $[\tau]_{\gamma^*} = [c^1, \ldots, c^m]$,

$$[T^t(\tau)]_{\beta^*} = [\tau]_{\gamma^*} A.$$

Now let's put formula 6.6 to use in the case $\Phi_{*,p} : T_p I\!R^n \to T_q I\!R^m$, $q = \Phi(p)$. Denote by $\Phi^*_q : T^*_q I\!R^m \to T^*_p I\!R^n$ the map on the dual, and let

$\tau_q \in T^*_q I\!R^m$ (note a dual vector at a point in $I\!R^m$ in the image of Φ). Then the corresponding definition from 6.1 of the map $\Phi^*_q(\tau) \in T^*_p I\!R^n$ is

$$(\Phi^*_q \tau_q)(X_p) = \tau_q(\Phi_{*,p} X_p). \tag{6.8}$$

Let $\beta = \{\partial_{x^i}|_p\}_{1\leq i\leq n}$ be the coordinate basis for $T_p I\!R^n$, and $\gamma = \{\partial_{y^a}|_q\}_{1\leq a\leq m}$ the coordinate basis for $T_q I\!R^n$, and the corresponding dual basis are $\{dx^i|_p\}_{1\leq i\leq n}$, and $\{dy^a|_q\}_{1\leq a\leq m}$. The matrix representation of $\Phi_{*,p}$ in the coordinate basis is

$$\Phi_{*,p}(\partial_{x^i}|_p) = \sum_{a=1}^{m} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \partial_{y^a}|_q,$$

and so equation 6.6 gives

$$\Phi^*_q(dy^a|_q) = \sum_{i=1}^{n} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p dx^i|_p. \tag{6.9}$$

Note the difference in the summation index in these last two equations.

An important observation from equation 6.9 needs to be made. Equation 5.16 is

$$(d\Phi^a)_p = \sum_{i=1}^{n} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p dx^i|_p,$$

and so equation 6.9 can then be written,

$$\Phi^*_q(dy^a|_q) = (d\Phi^a)_p. \tag{6.10}$$

This motivates the definition.

Definition 6.1.2. Let $\Phi : I\!R^n \to I\!R^m$, and $g \in C^\infty(I\!R^m)$. The pullback of g to $I\!R^n$ denoted by $\Phi^* g$ is the function

$$\Phi^* g = g \circ \Phi, \tag{6.11}$$

and $\Phi^* g \in C^\infty(I\!R^n)$.

Using definition 6.1.2 we have for the coordinate functions $\Phi^* y^a = \Phi^a$, and equation 6.10 can be written

$$\Phi^*_q(dy^a)_p = d(\Phi^* y^a)_p = (d\Phi^a)_p. \tag{6.12}$$

Finally for a general element of $\tau_q \in T_q^*\mathbb{R}^m$, $\tau_q = \sum_{a=1}^m c_a dy^a|_q$, $c_a \in \mathbb{R}$, we then find as in equation 6.7, by using equations 6.10, and 6.9 that

$$\Phi_q^*(\sum_{a=1}^m c_a dy^a|_q) = \sum_{a=1}^m c_a (d\Phi^a)_p , \tag{6.13}$$
$$= \sum_{i=1}^n \left(\sum_{a=1}^m c_a \left. \frac{\partial \Phi^a}{\partial x^i} \right|_p \right) dx^i|_p .$$

Example 6.1.3. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$\Phi(x, y, z) = (u = x + y, v = x^2 - y^2, w = xy), \tag{6.14}$$

and let $\tau_q = (2du - 3dv + dw)_q$ where $q = \Phi(1, 2) = (3, -3, 2)$. We compute $\Phi_q^*\tau_q$ by first using 6.10,

$$\begin{aligned} \Phi_q^* du_q &= d(x + y)_{(1,2)} = (dx + dy)_{(1,2)} \\ \Phi_q^* dv_q &= d(x^2 - y^2)_{(1,2)} = (2xdx - 2ydy)_{(1,2)} = (2dx - 4dy)_{(1,2)} \\ \Phi_q^* dw_q &= d(xy)_{(1,2)} = (2dx + dy)_{(1,2)}. \end{aligned} \tag{6.15}$$

Therefore by equation 6.13 and the equation 6.15,

$$\begin{aligned} \Phi_q^*(2du - 3dv + dw)|_q &= 2\Phi_q^*(du|_q) - 3\Phi_q^*(dv|_q) + \Phi_q^*(dw|_q) \\ &= 2(dx + dy) - 3(2dx - 4dy) + 2dx + dy \\ &= (-2dx + 15dy)_{(1,2)} . \end{aligned}$$

We now come to a fundamental observation. Recall if X is a vector-field on $\mathbb{R}^n$ and $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, that we cannot use Φ_* to define a vector-field on $\mathbb{R}^m$. In the case $m = n$ and Φ is a diffeomorphism then it is possible to push-forward a vector-field as in section 4.5, but in general this is not the case. Let's compare this with what happens for one-form fields.

Suppose that τ is now a one-form field on $\mathbb{R}^m$. (τ is specified on the image space of Φ). For **any** point $p \in \mathbb{R}^n$ (the domain of Φ) we can define $\alpha_p = \Phi_q^*(\tau_{\Phi(p)})$. We call the differential one-form α the pullback of τ and we write $\alpha = \Phi^*\tau$ for the map on the differential one-form τ. Therefore we can **always** pullback a differential one-form on the image to a differential one-form on the domain, something we can not do with the push-forward for vector-fields!

We now give a formula for $\Phi^*\tau$ in coordinates. If $\tau = \sum_{a=1}^m f_a(\mathbf{y})dy^a$, then with $\mathbf{y} = \Phi(\mathbf{x})$ we have

$$\tau_{\mathbf{y}=\Phi(\mathbf{x})} = \sum_{a=1}^m f_a(\Phi(\mathbf{x}))dy^a|_{\Phi(\mathbf{x})}.$$

From equation 6.13

$$\begin{aligned}(\Phi^*\tau)_{\mathbf{x}} &= \Phi_q^*(\tau_{\Phi(\mathbf{x})}) = \sum_{a=1}^m f_a(\Phi(\mathbf{x}))d\Phi^a \\ &= \sum_{i=1}^n \left(\sum_{a=1}^m f_a(\Phi(\mathbf{x})) \left.\frac{\partial \Phi^a}{\partial x^i}\right|_{\mathbf{x}} \right) dx^i|_{\mathbf{x}}\end{aligned} \tag{6.16}$$

which holds at every point $\mathbf{x}$ in the domain of Φ. In particular note that equation 6.16 implies that using definition 6.1.2 the pullback of the coordinate differential one-forms are ,

$$\Phi^*dy^a = d\left(\Phi^*y^a\right) = d\Phi^a = \sum_{i=1}^n \frac{\partial \Phi^a}{\partial x^i}dx^i. \tag{6.17}$$

We then rewrite equation 6.16 as (dropping the subscript $\mathbf{x}$)

$$\begin{aligned}\Phi^*\tau &= \sum_{a=1}^m f_a(\Phi(\mathbf{x}))\Phi^*dy^a \\ &= \sum_{i=1}^n \left(\sum_{a=1}^m f_a(\Phi(\mathbf{x}))\frac{\partial \Phi^a}{\partial x^i}\right) dx^i\end{aligned} \tag{6.18}$$

Example 6.1.4. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ be as in equation 6.14 from example 6.1.3. We compute $\Phi^*du, \Phi^*dv, \Phi^*dw$ using 6.17,

$$\begin{aligned}\Phi^*du = d(\Phi^*u) &= d(x+y) = dx + dy, \\ \Phi^*dv = d(\Phi^*v) &= d(x^2-y^2) = 2xdx - 2ydy, \\ \Phi^*dw = d(\Phi^*w) &= d(xy) = ydx + xdy\,.\end{aligned} \tag{6.19}$$

Let's calculate $\Phi^*(vdu - udv + wdw)$. In this case we have by formula 6.18 and 6.19

$$\begin{aligned}\Phi^*(vdu - udv + wdw) &= (x^2-y^2)\Phi^*du - (x+y)\Phi^*dv + xy\Phi^*dw\,, \\ &= (x^2-y^2)d(x+y) - (x+y)d(x^2-y^2) + xy\,d(xy)\,, \\ &= (x^2-y^2)(dx+dy) - (x+y)(2xdx - 2ydy) + xy(ydx + xdy)\,, \\ &= (xy^2 - 3x^2 - 2xy)dx + (x^2 + y^2 + 2xy + x^2y)dy.\end{aligned}$$

Finally we give a fundamental theorem relating the pull back and the exterior derivative.

Theorem 6.1.5. *Let $\Phi : I\!R^n \to I\!R^m$ be a smooth function. Then*

$$d(\Phi^* g) = \Phi^*(dg) \tag{6.20}$$

where $g \in C^\infty(I\!R^m)$.

This is states that Φ^* and d commute. A proof of this is easily given by writing both sides out in coordinates, but here is an alternative one which is the standard proof in differential geometry.

Proof. To prove equation 6.37 holds, let $p \in I\!R^n$ and X_p be in $T_pI\!R^n$. We then show

$$\left(d(\Phi^* g)\right)(X_p) = (\Phi^* dg)(X_p). \tag{6.21}$$

Since p and X_p are arbitrary, equation 6.21 implies 6.37.

We first expand out the left side of equation 6.21 using the definition of d in equation 5.10 applied to the function $\Phi^* g$ on $C^\infty(I\!R^n)$ and then equation 6.11 to get,

$$\left(d(\Phi^* g)\right)(X_p) = X_p(\Phi^* g) = X_p(g \circ \Phi). \tag{6.22}$$

Next we expand the right hand side of equation 6.21 using 6.8, 5.10 , and 4.7 to get

$$(\Phi^* dg)(X_p) = dg(\Phi_* X_p) = (\Phi_* X_p)(g) = X_p(g \circ \Phi). \tag{6.23}$$

The equality of equations 6.22 and 6.23 proves equation 6.37 holds. □

6.2 The Pullback of a Metric Tensor

Generalizing what we did in the previous section, suppose that $T : V \to W$ is a linear transformation we define the function $T^t : \mathbf{B}(W) \to \mathbf{B}(V)$ from the bilinear functions on W to those on V as

$$T^t(B)(v_1, v_2) = B(T(v_1), T(v_2)), \quad B \in \mathbf{B}(W). \tag{6.24}$$

We have a lemma analogous to Lemma 6.1.1 in the previous section checking that $T^t(B)$ is really bilinear.

Lemma 6.2.1. *Let $T : V \to W$ be a linear transformation, and $B \in \mathbf{B}(W)$, then $T^t(B) \in \mathbf{B}(V)$. Furthermore if B is symmetric, then $T^t(B)$ is symmetric. If T is injective and B is positive definite (or non-degenerate) then T^*B is positive definite (or non-degenerate).*

Proof. The fact that $T^t(B)$ is bilinear is similar to Lemma 6.1.1 above and won't be repeated.

Suppose that B is symmetric then for all $v_1, v_2 \in V$,

$$T^t(B)(v_1, v_2) = B(T(v_1), T(v_2)) = B(T(v_2), T(v_1)) = T^t(B)(v_2, v_1).$$

Therefore $T^t(B)$ is symmetric.

Suppose that T is injective, and that B is positive definite. Then

$$T^t(B)(v, v) = B(T(v), T(v)) \geq 0$$

because B is positive definite. If $B(T(v), T(v)) = 0$ then $T(v) = 0$, which by the injectivity of T implies $v = 0$. Therefore $T^t(B)$ is positive definite. □

Suppose now that V and W are finite dimensional vector-spaces of dimension n and m respectively, and that $\beta = \{v_i\}_{1 \leq i \leq n}$ is a basis for V, and $\gamma = \{w_a\}_{1 \leq a \leq n}$ is a basis for W. Denoting as usual $A = [T]^\gamma_\beta$, then A is the $m \times n$ matrix determined by

$$T(v_i) = \sum_{a=1}^{m} A^a_i w_a.$$

For $B \in \mathbf{B}(V)$ the matrix representation of B is

$$B_{ab} = B(w_a, w_b). \tag{6.25}$$

We now compute the matrix representation of the bilinear function $T^t(B)$,

$$\begin{aligned} T^t(B)_{ij} &= T^t(B)(v_i, v_j) = B(T(v_i), T(v_j)) \\ &= B(\sum_{a=1}^{m} A^a_i w_a, \sum_{b=1}^{m} A^b_i w_b) \\ &= \sum_{1 \leq a,b \leq m} A^a_i A^b_j B(w_a, w_b) \\ &= \sum_{1 \leq a,b \leq m} A^a_i A^b_j B_{ab} \end{aligned} \tag{6.26}$$

In terms of matrix multiplication one can write equation 6.26 as

$$(T^t(B)) = A^T(B)A.$$

Now let $\beta^* = \{\alpha^i\}_{1\le i\le n}$ be the basis of V^* which is the dual basis of β for V, and $\gamma = \{\tau^a\}_{1\le a\le n}$ the basis of W^* which is the dual basis of γ for W^*. Using equation 6.25 and the tensor product basis as in equation 5.30 we have

$$B = \sum_{1\le a,b\le m} B_{ab}\tau^a\otimes\tau^b. \tag{6.27}$$

While using the coefficients in equation 6.26 and the tensor product basis as in equation 5.30 we have

$$T^t(B) = \sum_{1\le i,j\le n}\left(\sum_{1\le a,b\le m} A^a_i A^b_j B_{ab}\right)\alpha^i\otimes\alpha^j. \tag{6.28}$$

By using the formula for $T^t(\tau^a), T^t(\tau^b)$ from equation 6.6, this last formula can also be written as

$$\begin{aligned} T^t(B) &= T^t\left(\sum_{1\le a,b\le m} B_{ab}\,\tau^a\otimes\tau^b\right) \\ &= \sum_{1\le a,b\le m} B_{ab}\,T^t(\tau^a)\otimes T^t(\tau^b)\,. \end{aligned} \tag{6.29}$$

Let $\Phi : I\!R^n \to I\!R^m$ be a smooth function and let $p\in I\!R^n$, and let $B : T_qI\!R^m\times T_qI\!R^m \to I\!R$, with $q=\Phi(p)$, be a bilinear function. Then $\Phi^*_q(B) : T_pI\!R^n\times T_pI\!R^n\to I\!R$ is a bilinear function defined exactly as in 6.24 by

$$\Phi^*_q(B)(X_p,Y_p) = B(\Phi_{*,p}X_p, \Phi_{*,p}Y_p).$$

Suppose that we have for $T_pI\!R^n$ the standard coordinate basis $\{\partial_{x^i}|_p\}_{1\le i\le n}$, and for $T_qI\!R^m$ the basis $\{\partial_{y^a}|_q\}$, with the corresponding dual basis $\{dx^i|_p\}_{1\le i\le n}$ and $\{dy^a|_q\}_{1\le a\le m}$. Recall equation 6.9,

$$\Phi^*_q(dy^a|_q) = \sum_{i=1}^n \left.\frac{\partial\Phi^a}{\partial x^i}\right|_p dx^i|_p.$$

Therefore writing

$$B = \sum_{1\leq a,b\leq m} B_{ab} dy^a|_q \otimes dy^b|_q, \quad B_{ab} \in I\!R\,,$$

formula 6.28 gives

$$(\Phi_q^* B)_p = \sum_{1\leq i,j\leq n} \left(\sum_{1\leq a,b\leq m} B_{ab} \left.\frac{\partial \Phi^a}{\partial x^i}\right|_p \left.\frac{\partial \Phi^b}{\partial x^j}\right|_p \right) dx^i|_p \otimes dx^j|_p. \tag{6.30}$$

Equation 6.30 can also be written using equation 6.29 as

$$(\Phi_q^* B)_p = \sum_{1\leq a,b\leq m} B_{ab} (\Phi_q^* dy^a|_{\Phi(p)}) \otimes (\Phi_q^* dy^b|_{\Phi(p)}) \tag{6.31}$$

Example 6.2.2. Let $\Phi : U \to I\!R^3$, $U = \{(\theta,\phi) \in I\!R^2 \mid 0 < \theta < 2\pi, 0 < \phi < \pi\}$, be the function

$$\Phi(\theta,\phi) = (x = \cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi). \tag{6.32}$$

Let $p = \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$, $q = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ and let $B = (dx^2 + dy^2 + dz^2)|_q$. From equation 6.31 we find

$$\Phi_q^*(B)_p = (\Phi_q^* dx|_q)^2 + (\Phi_q^* dy|_q)^2 + (\Phi_q^* dz|_q)^2.$$

Then by equation 6.9,

$$\Phi^* dx|_q = (-\sin\theta\sin\phi d\theta + \cos\theta\cos\phi d\phi)_p = -\frac{1}{2} d\theta|_p + \frac{1}{2} d\phi|_p.$$

Similarly

$$\Phi_q^* dy_q = \frac{1}{2} d\theta|_p + \frac{1}{2} d\phi|_p, \quad \Phi_q^* dz|_q = -\frac{1}{\sqrt{2}} d\phi|_p,$$

and so

$$\Phi_q^*(dx^2 + dy^2 + dz^2)_q = (d\phi^2 + \frac{1}{2} d\theta^2)|_p\,.$$

Recall from the previous section the important property that a differential one-form (or one-form field) that is defined on the image of a smooth function $\Phi : I\!R^n \to I\!R^m$ pulls-back to a differential one-form on the domain of Φ. A similar property holds for fields of bilinear functions such as Riemannian metric tensors. In particular, if $\boldsymbol{\gamma}$ is a field of bilinear functions (such as a metric tensor field), then the pullback of $\Phi^*\boldsymbol{\gamma}$ is a field of bilinear forms defined by,

(6.33)
$$(\Phi^*\boldsymbol{\gamma})_p(X_p, Y_p) = \boldsymbol{\gamma}_p(\Phi_{*,p}X_p, \Phi_{*,p}Y_p), \quad for\ all\ p \in I\!R^n,\ X_p, Y_p \in T_pI\!R^n.$$

For each $p \in I\!R^n$(the domain of Φ), then $\Phi^*\boldsymbol{\gamma}$ is a bi-linear function on $T_pI\!R^n$, and so a field of bi-linear function.

Suppose that
$$\boldsymbol{\gamma} = \sum_{1 \leq a,b \leq m} g_{ab}(y) dy^a \otimes dy^b$$

is a a field of bilinear functions on $I\!R^m$ then $\Phi^*\boldsymbol{\gamma}$ is the field of bilinear function defined by equation 6.28 at every point in $I\!R^n$. If $\mathbf{y} = \Phi(\mathbf{x})$ then

$$\boldsymbol{\gamma}_{\Phi(\mathbf{x})} = \sum_{1 \leq a,b \leq m} g_{ab}(\Phi(x)) dy^a|_{\Phi(\mathbf{x})} \otimes dy^b|_{\Phi(\mathbf{x})}$$

and by equation 6.30

$$\Phi^*\boldsymbol{\gamma} = \sum_{1 \leq a,b \leq m} g_{ab}(\Phi(x))(\Phi^* dy^a) \otimes (\Phi^* dy^b). \tag{6.34}$$

This can be further expanded to using equation 6.17 to

$$\Phi^*\boldsymbol{\gamma} = \sum_{i,j=1}^{n} \sum_{a,b=1}^{m} g_{ab}(\Phi(x)) \frac{\partial \Phi^a}{\partial x^i} \frac{\partial \Phi^b}{\partial x^j} dx^i \otimes dx^j \tag{6.35}$$

Example 6.2.3. Continue with Φ in equation 6.32 in example 6.2.2 we compute $\Phi^*\boldsymbol{\gamma}^E$ by using equation 6.34

$$\begin{aligned}\Phi^*(dx^2 + dy^2 + dz^2) &= (\Phi^* dx)^2 + (\Phi^* dy)^2 + (\Phi^* dz)^2 \\ &= (-\sin\theta \sin\phi\, d\theta + \cos\theta \cos\phi\, d\phi)^2 + \\ &(\cos\theta \sin\phi\, d\theta + \sin\theta \cos\phi\, d\phi)^2 + \sin^2\phi\, d\phi^2 \\ &= d\phi^2 + \sin^2\phi d\theta^2\end{aligned}$$

Example 6.2.4. With $\Phi : I\!R^2 \to I\!R^3$ given by

$$\Phi(x, y) = (x, y, z = f(x, y))$$

we compute $\Phi^* \boldsymbol{\gamma}^E$ using equation 6.34 by first computing

$$\Phi^* dx = dx,\ \Phi^* dy = dy,\ \Phi^* dz = f_x dx + f_y dy.$$

Therefore

$$\begin{aligned}\Phi^*(dx^2 + dy^2 + dz^2) &= dx^2 + dy^2 + (f_x dx + f_y dy)^2 \\ &= (1 + f_x^2) dx^2 + 2 f_x f_y dx dy + (1 + f_y^2) dy^2.\end{aligned}$$

Example 6.2.5. Let $U \subset I\!R^3$ and $\boldsymbol{\gamma}$ be the metric tensor in 5.4.3, and let $\Phi : I\!R^3 \to U$ be

$$\Phi(u, v, w) = (x = e^u, y = ve^u, z = e^w).$$

By equation 6.34 the pullback $\Phi^*\boldsymbol{\gamma}$ is

(6.36)

$$\begin{aligned}\Phi^\gamma &= \frac{1}{e^{2u}}(d(e^u))^2 + \frac{1}{e^{2u}}dv^2 - 2\frac{v}{e^w e^{2u}}dvd(e^w) + \left(\frac{v^2}{e^{2(u+w)}} + \frac{1}{e^{2w}}\right)(d(e^w))^2 \\ &= du^2 + e^{-2u}dv^2 - 2ve^{-2u}dvdw + (1 + v^2 e^{-2u})dw^2.\end{aligned}$$

Theorem 6.2.6. *Let $\Phi : I\!R^m \to I\!R^n$ be an immersion, and $\boldsymbol{\gamma}$ a Riemannian metric tensor on $I\!R^n$. Then the pullback $\Phi^*\boldsymbol{\gamma}$ is a Riemannian metric on $I\!R^m$*

Proof. We only need to prove that at each point $p \in I\!R^m$ that $(\Phi^*\boldsymbol{\gamma})_p$ is an inner product. XXXX □

More examples will be given in the next section.

A simple theorem which follows from the definitions in this section is the following.

Theorem 6.2.7. *Let $\Phi : I\!R^n \to I\!R^m$ be a smooth function and let α, β be differential one forms on the range space $I\!R^m$ and let $g \in C^\infty(I\!R^m)$. Then*

$$\Phi^*(g \cdot \alpha \otimes \beta) = \Phi^* g \cdot (\Phi^*\alpha) \otimes (\Phi^*\beta). \tag{6.37}$$

6.3 Isometries

Let $\boldsymbol{\gamma}$ be a fixed metric tensor on $I\!R^n$.

Definition 6.3.1. A diffeomorphism $\Phi : I\!R^n \to I\!R^n$ is an *isometry* of the metric $\boldsymbol{\gamma}$ if

$$\Phi^*\boldsymbol{\gamma} = \boldsymbol{\gamma}. \tag{6.38}$$

Let's write this out more carefully using the definition of pullback in equation 6.33. We have Φ is an isometry of $\boldsymbol{\gamma}$ if and only if

$$\boldsymbol{\gamma}(X_p, Y_p) = \boldsymbol{\gamma}_(\Phi_{*,p}X_p, \Phi_{*,p}Y_p), \quad for\ all p \in I\!R^n,\ X_p, Y_p \in T_pI\!R^n\,. \tag{6.39}$$

The metric tensor $\boldsymbol{\gamma}$ on the right hand side is evaluated at $\Phi(p$.

Lemma 6.3.2. *Let $\Phi : I\!R^n \to I\!R^n$ be a diffeomorphism, then the following are equivalent:*

1. *Φ is an isometry.*
2. *For all $p \in I\!R^n$, and $1 \leq i, j, \leq n$,*

   $$\boldsymbol{\gamma}(\partial_{x^i}|_p, \partial_{x^j}|_p) = \boldsymbol{\gamma}(\Phi_{*,p}(\partial_{x^i}|_p), \Phi_{*,p}(\partial_{x^j}|_p))\ . \tag{6.40}$$

3. *For all $p \in I\!R^n$ and $X_p \in T_pI\!R^n$, $\boldsymbol{\gamma}(X_p, X_p) = \boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}X_p)$.*

Proof. Clearly 1 implies 2 by equation 6.39. Suppose that 2 holds and $X_p \in T_pI\!R^n$ where $X_p = \sum_{i=1}^n X^i\partial_{x^i}|_p,\ X^i \in I\!R$. Using bilinearity,

$$\begin{aligned}
\boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}X_p) &= \boldsymbol{\gamma}(\Phi_{*,p}\sum_{i=1}^n X^i\partial x^i, \Phi_{*,p}\sum_{j=1}^n X^j\partial_{x^j}) \\
&= \sum_{i,j=1}^n X^iX^j\boldsymbol{\gamma}(\Phi_{*,p}\partial_{x^i}|_p, \Phi_{*,p}\partial_{x^j}|_p) \\
&= \sum_{i,j=1}^n X^iX^j\boldsymbol{\gamma}(\partial_{x^i}|_p, \partial_{x^j}|_p) \quad \text{by1.} \\
&= \boldsymbol{\gamma}(X, X).
\end{aligned}$$

Therefore 2 implies 3.

For 3 to imply 1), let $X_p, Y_p \in T_pI\!R^n$, then by hypothesis

$$\boldsymbol{\gamma}(X_p + Y_p, X_p + Y_p) = \boldsymbol{\gamma}(\Phi_{*,p}(X_p + Y_p), \Phi_{*,p}(X_p + Y_p)). \tag{6.41}$$

Expanding this equation using bilinearity the left hand side of this is

$$\boldsymbol{\gamma}(X_p + Y_p, X_p + Y_p) = \boldsymbol{\gamma}(X_p, X_p) + 2\boldsymbol{\gamma}(X_p, Y_p) + \boldsymbol{\gamma}(Y_p, Y_p) \tag{6.42}$$

while the right hand side of equation 6.41 is

$$\boldsymbol{\gamma}(\Phi_{*,p}(X_p+Y_p), \Phi_{*,p}(X_p+Y_p)) = \boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}X_p)+2\boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}Y_p)+\boldsymbol{\gamma}(\Phi_{*,p}Y_p, \Phi_{*,p}Y_p). \tag{6.43}$$

Substituting equations 6.42 and 6.43 into equation 6.41 we have

$$\boldsymbol{\gamma}(X_p, X_p)+2\boldsymbol{\gamma}(X_p, Y_p)+\boldsymbol{\gamma}(Y_p, Y_p) = \boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}X_p)+2\boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}Y_p)+\boldsymbol{\gamma}(\Phi_{*,p}Y_p, \Phi_{*,p}Y_p). \tag{6.44}$$

Again using the hypothesis 3, $\boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}X_p) = \boldsymbol{\gamma}(X_p, X_p)$, and $\boldsymbol{\gamma}(\Phi_{*,p}Y, \Phi_{*,p}Y) = \boldsymbol{\gamma}(Y, Y)$ in equation 6.44 we are left with

$$2\boldsymbol{\gamma}(\Phi_{*,p}X_p, \Phi_{*,p}Y_p) = 2\boldsymbol{\gamma}(X_p, Y_p),$$

which shows that 3 implies 1 (by equation 6.39). □

The last condition says that for Φ to be an isometry it is necessary and sufficient that $\Phi_{*,p}$ preserves lengths of vectors.

Example 6.3.3. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ be the diffeomorphism,

$$\Phi(x, y) = \left(\frac{1}{\sqrt{2}}(x - y), \frac{1}{\sqrt{2}}(x + y) \right). \tag{6.45}$$

This function is a counter clockwise rotation by $\pi/4$ about the origin in $\mathbb{R}^2$. We compute now what Φ does to tangent vectors. Let $X_p = X^1\partial_x + X^2\partial_y$ at the point $p = (x_0, y_0)$. We find the coefficient of $\Phi_{*,p}X_p$ in the coordinate basis are

$$[\Phi_{*,p}(X_p)] = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} X^1 \\ X^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(X^1 - X^2) \\ \frac{1}{\sqrt{2}}(X^1 + X^2) \end{bmatrix}.$$

The two by two matrix is the Jacobian matrix for Φ at the point p (in this case the point p doesn't show up in evaluating the Jacobian). We see the the coefficients of the image vector, are just the rotated form of the ones we started with in X_p.

XXXXXXXXXXXXX DRAW PICTURE XXXXXXXXXXXXXX

Therefore we can check condition 3 in Lemma 6.3.2 for the Euclidean metric tensor $\boldsymbol{\gamma}_{Eu}$ by computing

$$\begin{aligned}\boldsymbol{\gamma}_{Eu}(\Phi_{*,p}X_p, \Phi_{*,p}X_p) &= \left(\frac{1}{\sqrt{2}}(X^1 - X^2)\right)^2 + \left(\frac{1}{\sqrt{2}}(X^1 + X^2)\right)^2 \\ &= (X^1)^2 + (X^2)^2 \\ &= \boldsymbol{\gamma}_{Eu}(X_p, X_p)\end{aligned}$$

Example 6.3.4. In this next example, consider the metric tensor in $\mathbb{R}^2$ given by

$$\boldsymbol{\gamma} = \frac{1}{1+x^2+y^2}(dx^2 + dy^2) \tag{6.46}$$

We claim the diffeomorphism Φ in equation (6.45) is an isometry for this metric. We first compute,

$$\Phi^* dx = \frac{1}{\sqrt{2}}(dx - dy)\ ,\ \Phi^* dy = \frac{1}{\sqrt{2}}(dx + dy).$$

Then computing,

$$\begin{aligned}\Phi^*\boldsymbol{\gamma} &= \frac{1}{1 + (\frac{1}{\sqrt{2}}(x-y))^2 + (\frac{1}{\sqrt{2}}(x+y))^2}\left(\frac{1}{2}(dx-dy)^2 + \frac{1}{2}(dx-dy)^2\right) \\ &= \frac{1}{1+x^2+y^2}(dx^2 + dy^2).\end{aligned}$$

Therefore Φ satisfies equation 6.38 and is an isometry.

Example 6.3.5. In this next example consider the diffeomorphisms of $\mathbb{R}^2$,

$$\Psi_t(x, y) = (x + t, y), \tag{6.47}$$

where $t \in \mathbb{R}$. We compute

$$\Psi_t^* dx = dx,\ \Psi_t^* dy = dy$$

since t is a constant. For the Euclidean metric $\boldsymbol{\gamma}^E$, we find

$$\Psi_t^*\boldsymbol{\gamma}^E = dx^2 + dy^2.$$

Therefore Ψ in equation (6.47) is an isometry *for all* $t \in I\!R$.

Are Ψ_t isometries for the metric in equation (6.46) in example 6.3.4? We have

$$\Psi_t^*\left(\frac{1}{1+x^2+y^2}(dx^2+dy^2)\right) = \frac{1}{1+(x+t)^2+y^2}(dx^2+dy^2).$$

This will not equal $\boldsymbol{\gamma}$ unless $t = 0$. In which case Ψ_t is just the identity transformation.

Example 6.3.6. In this next example consider the diffeomorphisms of $I\!R^2$,

$$\Psi_t(x,y) = (x\cos t - y\sin t, x\sin t + y\cos t), \tag{6.48}$$

where $t \in [0, 2\pi)$. We compute from equation 6.17,

$$\Psi_t^* dx = \cos t\, dx - \sin t\, dy,\ \ \Psi_t^* dy = \sin t\, dx + \cos t\, dy,$$

since t is a constant. For the Euclidean metric $\boldsymbol{\gamma}^E$ we find

$$\begin{aligned}\Psi_t^*\boldsymbol{\gamma}^E &= (\cos t\, dx - \sin t\, dy)^2 + (\sin t\, dx + \cos t\, dy)^2\\ &= dx^2 + dy^2.\end{aligned}$$

Therefore Ψ_t in equation (6.39) is an isometry *for all* $t \in I\!R$.

What about the metric in equation (6.46) in example 6.3.4. We have

$$\begin{aligned}&\Psi_t^*\left(\frac{1}{1+x^2+y^2}(dx^2+dy^2)\right) =\\ &\frac{1}{1+(x\cos t - y\sin t)^2 + (x\sin t + y\cos t)^2}\left((\cos t\, dx - \sin t\, dy)^2 + ((\sin t\, dx + \cos t\, dy))^2\right)\\ &= \frac{1}{1+x^2+y^2}(dx^2+dy^2)\end{aligned}$$

Therefore Φ_t is an isometry of this metric tensor as well.

Example 6.3.7. In example 5.4.3 we have $U = \{(x,y,z) \in I\!R^3 \mid xz \neq 0\ \}$ with the following metric tensor on U,

$$\boldsymbol{\gamma} = \frac{1}{x^2}dx^2 + \frac{1}{x^2}dy^2 - 2\frac{y}{zx^2}dydz + \left(\frac{y^2}{x^2z^2} + \frac{1}{z^2}\right)dz^2. \tag{6.49}$$

For each $a, c \in I\!R^*$ and let $b \in I\!R$, define the function $\Phi_{(a,b,c)} : U \to U$ by

$$\Phi_{(a,b,c)}(x,y,z) = (u = ax, v = ay + bz, w = cz) \tag{6.50}$$

Therefore (noting that a, b, c are constants) we have

$$\Phi^*_{(a,b,c)}du = a\,dx\,,\ \Phi^*_{(a,b,c)}dv = a\,dy + b\,dz\,,\ \Phi^*_{(a,b,c)}dw = c\,dz\,. \tag{6.51}$$

Using equation 6.51 we find

$$\begin{aligned}
&\Phi^*_{(a,b,c)}\left(\frac{1}{u^2}du^2 + \frac{1}{u^2}dv^2 - 2\frac{v}{wu^2}dvdw + \left(\frac{v^2}{u^2w^2} + \frac{1}{w^2}\right)dw^2\right),\\
&= \frac{1}{ax^2}(adx)^2 + \frac{1}{(ax)^2}(a\,dy + b\,dz)^2 - 2\frac{ay+bz}{cz(ax)^2}(a\,dy + b\,dz)(c\,dz)\\
&\quad + \left(\frac{(ay+bw)^2}{(ax)^2(cz)^2} + \frac{1}{(cz)^2}\right)(c\,dz)^2\\
&= \frac{1}{x^2}dx^2 + \frac{1}{x^2}dy^2 - 2\frac{y}{zx^2}dydz + \left(\frac{y^2}{x^2z^2} + \frac{1}{z^2}\right)dz^2.
\end{aligned}$$

Therefore for each a, b, c the diffeomorphism $\Phi_{(a,b,c)}$ is an isometry of the metric $\boldsymbol{\gamma}$.

Given a metric tensor $\boldsymbol{\gamma}$ the the isometries have a simple algebraic structure.

Theorem 6.3.8. *Let $\boldsymbol{\gamma}$ be a metric tensor in $I\!R^n$. The set of isometries of $\boldsymbol{\gamma}$ form a group with composition of functions as the group operations. This group is called the isometry group of the metric.*

Proof. Let Φ and Ψ be isometries of the metric $\boldsymbol{\gamma}$ and $X_p, Y_p \in T_pI\!R^n$, then

$$\begin{aligned}
\boldsymbol{\gamma}\left((\Phi\circ\Psi)_*X_p, (\Phi\circ\Psi)_*Y_p\right) &= \boldsymbol{\gamma}\left(\Phi_{*,p}\Psi_*X_p, \Phi_{*,p}\Psi_*Y_p\right) && \text{chain rule}\\
&= \boldsymbol{\gamma}\left(\Psi_*X_p, \Psi_*Y_p\right) \quad \Phi && \text{is an isometry 6.39}\\
&= \boldsymbol{\gamma}\left(X_p, Y_p\right) \quad \Psi && \text{is an isometry 6.39}
\end{aligned}$$

Therefore the composition of two isometries is an isometry, and we have a well defined operation for the group. Composition of functions is associative, and so the group operation is associative. The identity element is the identity function. We leave as an exercise to prove that if Φ is an isometry, then Φ^{-1} is also an isometry. □

Lastly we write out in coordinates the isometry condition. Suppose that $\boldsymbol{\gamma} = \sum_{i,j=1}^n g^i j(\mathbf{x}) dx^i dx^j$ is a metric tensor in $I\!R^n$ and that $\Phi : I\!R^n \to I\!R^n$ is a diffeomorphism. Let's expand out the right hand side of condition 2, in Lemma 6.3.2 with $q = \Phi(p)$,

$$
\begin{aligned}
\boldsymbol{\gamma}(\Phi_{*,p}\partial_{x^i}|_p), \Phi_{*,p}\partial_{x^j}|_p) &= \boldsymbol{\gamma}\left(\sum_{k=1}^n \left.\frac{\partial \Phi^k}{\partial x^i}\right|_p \partial_{y^k}|_q, \sum_{l=1}^n \left.\frac{\partial \Phi^l}{\partial x^j}\right|_p \partial_{y^l}|_{q)} \right) \\
&= \sum_{k,l=1}^n \left.\frac{\partial \Phi^k}{\partial x^i}\right|_p \left.\frac{\partial \Phi^l}{\partial x^j}\right|_p \boldsymbol{\gamma}(\partial_{y^k}|_q, \partial_{y^l}|_q) \\
&= \sum_{k,l=1}^n \left.\frac{\partial \Phi^k}{\partial x^i}\right|_p \left.\frac{\partial \Phi^l}{\partial x^j}\right|_p [\boldsymbol{\gamma}_q]_{kl}
\end{aligned}
\tag{6.52}
$$

Since p was arbitrary we can summarize the computation in equation 6.52 with the following lemma which is the component form for a diffeomorphism to be an isometry.

Lemma 6.3.9. *A diffeomorphism* $\Phi : I\!R^n \to I\!R^n$ *is an isometry of the metric tensor* $\boldsymbol{\gamma} = \sum_{i,j=1}^n g^{ij}(\mathbf{x}) dx^i dx^j$ *on* $I\!R^n$ *if and only if*

$$
\sum_{k,l=1}^n \frac{\partial \Phi^k}{\partial x^i} \frac{\partial \Phi^l}{\partial x^j} g_{kl}(\Phi(\mathbf{x})) = g_{ij}(\mathbf{x}). \tag{6.53}
$$

This lemma can be viewed in two ways. First given a diffeomorphism, we can check if it is an isometry of a given metric $\boldsymbol{\gamma}$. The second and more interesting point of view is that equations (6.53) can be viewed as partial differential equations for the function Φ given $\boldsymbol{\gamma}$. These partial differential equation are very non-linear for Φ, but have some very unusual properties and can be integrated in a number of special situations.

Example 6.3.10. Equation 6.53 for the Euclidean metric tensor-field $\boldsymbol{\gamma}^E$ on $I\!R^n$ is

$$
\sum_{k,l=1}^n \frac{\partial \Phi^k}{\partial x^i} \frac{\partial \Phi^l}{\partial x^j} \delta_{kl} = \delta_{ij}. \tag{6.54}
$$

If we differentiate this with respect to x^m we get

$$
\sum_{k,l=1}^n \left(\frac{\partial^2 \Phi^k}{\partial x^i \partial x^m} \frac{\partial \Phi^l}{\partial x^j} + \frac{\partial \Phi^k}{\partial x^i} \frac{\partial^2 \Phi^l}{\partial x^j \partial x^m} \right) \delta_{kl} = 0. \tag{6.55}
$$

Now using equation 6.54 but replacing i with m and then differentiating with respect to x^i we get equation 6.55 with i and m switched,

$$\sum_{k,l=1}^{n}\left(\frac{\partial^2\Phi^k}{\partial x^m\partial x^i}\frac{\partial\Phi^l}{\partial x^j}+\frac{\partial\Phi^k}{\partial x^m}\frac{\partial^2\Phi^l}{\partial x^j\partial x^i}\right)\delta_{kl}=0. \tag{6.56}$$

Do this again with j and m to get,

$$\sum_{k,l=1}^{n}\left(\frac{\partial^2\Phi^k}{\partial x^i\partial x^j}\frac{\partial\Phi^l}{\partial x^m}+\frac{\partial\Phi^k}{\partial x^i}\frac{\partial^2\Phi^l}{\partial x^m\partial x^j}\right)\delta_{kl}=0. \tag{6.57}$$

Now take equation 6.55 plus 6.56 minus 6.57 to get

$$\begin{aligned}0=\sum_{k,l=1}^{n}\Big(&\frac{\partial^2\Phi^k}{\partial x^i\partial x^m}\frac{\partial\Phi^l}{\partial x^j}+\frac{\partial\Phi^k}{\partial x^i}\frac{\partial^2\Phi^l}{\partial x^j\partial x^m}\\&+\frac{\partial^2\Phi^k}{\partial x^m\partial x^i}\frac{\partial\Phi^l}{\partial x^j}+\frac{\partial\Phi^k}{\partial x^m}\frac{\partial^2\Phi^l}{\partial x^j\partial x^i}\\&-\frac{\partial^2\Phi^k}{\partial x^i\partial x^j}\frac{\partial\Phi^l}{\partial x^m}-\frac{\partial\Phi^k}{\partial x^i}\frac{\partial^2\Phi^l}{\partial x^m\partial x^j}\Big)\delta_{kl}.\end{aligned}$$

The second and sixth term cancel, and so do the fourth and fifth, while the first and third term are the same. Therefore this simplifies to

$$0=2\sum_{k,l=1}^{n}\frac{\partial^2\Phi^k}{\partial x^i\partial x^m}\frac{\partial\Phi^l}{\partial x^j}\delta_{kl}.$$

Now the condition Φ is a diffeomorphism implies that Φ_* is invertible (so the Jacobian matrix is invertible), and so

$$0=\frac{\partial^2\Phi^k}{\partial x^i\partial x^m}.$$

This implies Φ is linear and that

$$\Phi(\mathbf{x})=A\mathbf{x}+\mathbf{b}$$

where A is an invertible matrix, and $\mathbf{b}\in I\!R^n$. Finally using condition 6.54, we have

$$A^TA=I$$

and A is an orthogonal matrix.

The method we used to solve these equations is to take the original system of equations and differentiate them to make a larger system for which all possible second order partial derivatives are prescribed. This holds in general for the isometry equations for a diffeomorphism and the equations are what is known as a system of partial differential equations of finite type.

There another way to find isometries without appealing to the equations (6.53) for the isometries. This involves finding what are known as “Killing vectors” and their corresponding flows, which we discuss in the next two chapters. The equations for a Killing vector are linear and often easier to solve than the non-linear equations for the isometries. “Killing vectors” are named after the mathematician Wilhelm Killing (1847-1923).

6.4 Exercises

1. If $T : V \to W$ is a linear transformation, show that $T^t : W^* \to V^*$ is also a linear transformation.

2. Let $\Phi : \mathbb{R}^3 \to \mathbb{R}^2$ be
$$\Phi(x, y, z) = (u = x + y + z, v = xy + xz)$$
and compute
   (a) $\Phi^t \left(2du|_{(4,3)} - dv|_{(4,3)}\right)_{(1,2,1)}$, and
   (b) $\Phi^*(vdu + dv)$.

3. Let $\Phi : U \to \mathbb{R}^3$, $U = \{(\rho, \theta, \phi) \in \mathbb{R}^3 \mid 0 < \rho, 0 < \theta < 2\pi, 0 < \phi < \pi\}$, be
$$\Phi(\rho, \theta, \phi) = (x = \rho\cos\theta\sin\phi, y = \rho\sin\theta\sin\phi, z = \rho\cos\phi)$$
and compute
   (a) $\Phi^*(xdx + ydy + zdz)$,
   (b) $\Phi^*(ydx - xdy)$,
   (c) $\Phi^*(dx^2 + dy^2 + dz^2)$,
   (d) $\Phi^*(df),\ f = x^2 + y^2$.

4. Let $B \in \mathbf{B}(W)$ and $T : V \to W$ injective. Prove that if B is positive -definite then $T^t(B)$ is positive-definite.

5. Let $U = \{(x, y) \mid y > 0\}$ with metric tensor
$$\boldsymbol{\gamma} = \frac{1}{y^2}(dx^2 + dy^2)$$
   (a) Show that for each $t \in \mathbb{R}$, the transformations
$$\psi_t(x, y) \to (e^t x, e^t y)$$
   are isometries.

6. Show that for each $a, b, c \in I\!R$, that $\Phi_{(a,b,c)} : I\!R^3 \to I\!R^3$ given by

$$\Phi_{(a,b,c)}(x, y, z) = (x + a, y + az + b, z + c)$$

is an isometry for the metric tensor $\boldsymbol{\gamma}$ on $I\!R^3$,

$$\boldsymbol{\gamma} = dx^2 + dy^2 - 2x dy dz + (1 + x^2) dz^2 . \tag{6.58}$$

7. For which $a, b, c \in I\!R$ is

$$\Phi_{(a,b,c)}(x, y, z) = (x + a, y + cx + b, z + c)$$

an isometry of the metric tensor 6.58 in the previous problem?

8. By using the pullback, show that every diffeomorphism $\Phi : I\!R^n \to I\!R^n$ of the form

$$\Phi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}, \tag{6.59}$$

where A is an $n \times n$ matrix satisfying $A^T A = I$, and $\mathbf{b} \in I\!R^n$ is an isometry of the Euclidean metric tensor on $I\!R^n$. (Hint in components equation 6.59 is $\Phi^i = \sum_{j=1}^n A^i_j x^j + b^i$.)

9. Complete the proof of Theorem 6.3.8 by showing that if Φ is an isometry of the metric tensor $\boldsymbol{\gamma}$, then Φ^{-1} is also an isometry of $\boldsymbol{\gamma}$.

Chapter 7

Hypersurfaces

7.1 Regular Level Hyper-Surfaces

Let $F \in C^\infty(\mathbb{R}^{n+1})$ and let $c \in \mathbb{R}$. The set of points $S \subset \mathbb{R}^{n+1}$ defined by

$$S = \{\ p \in \mathbb{R}^{n+1} \mid F(p) = c\}$$

is called a regular level surface (or hyper-surface) if the differential $F_* : T_p\mathbb{R}^{n+1} \to T_{F(p)}\mathbb{R}$ is surjective at each point $p \in S$. Let's rewrite this condition in a basis. If $\partial_{x^i}|_p$ and $\partial_u|_q$ are the coordinate basis at the points p and $q = F(p)$ then the matrix representation of F_* is computed using **??** to be

$$[F_*] = (\partial_{x^1}F, \partial_{x^2}F, \ldots, \partial_{x^{n+1}}F)\,. \tag{7.1}$$

Therefore S is a regular level surface if at each point $p \in S$, at least one of the partial derivative of F with respect to x^i does not vanish at that point.

Example 7.1.1. Let $F : \mathbb{R}^{n+1} \to \mathbb{R}$ be the function

$$F(\mathbf{x}) = \mathbf{x}^T\mathbf{x} = \sum_{i=1}^{n+1}(x^i)^2$$

and let $r \in \mathbb{R}^+$. The n sphere of radius r given by

$$S^n_r = \{\ p \in \mathbb{R}^{n+1} \mid F(p) = r^2\ \}.$$

The standard n-sphere denoted by S^n has radius 1.

The sphere S_r^n is a regular level surface. To check this we compute

$$[F_*] = \left(2x^2, 2x^2, \ldots, 2x^{n+1}\right)$$

and note that at any point $p \in S_r^n$ not all $x^1, \ldots, x^{n+1}$ can be zero at the same time (because $r^2 > 0$).

Let $S \subset \mathbb{R}^{n+1}$ be a regular level surface $F = c$. The tangent space $T_pS, p \in S$ is

$$T_pS = \{\ X_p \in T_p\mathbb{R}^{n+1} \mid X_x \in \ker F_*|_p\ \}.$$

Note that since F_* has rank 1, that $\dim T_pS = n$ by the dimension theorem 2.2.7.

Lemma 7.1.2. *If $X_p \in T_p\mathbb{R}^{n+1}$ then $X_p \in T_pS$ if and only if*

$$X_p(F) = 0$$

Proof. Let $\iota \in C^\infty(\mathbb{R})$ be the identity function. We compute $F_*X_p(\iota) = X_p(\iota \circ F) = X_p(F)$. This vanishes if and only if $X_p(F) = 0$. □

Example 7.1.3. Let $r > 0$ and $S_r^2 = \{\ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = r^2\ \}$, which is the 2-sphere in $\mathbb{R}^3$ of radius r. Let $p = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{3}}\right) \in S^2$ (where $r = 1$), and let

$$X_p = \partial_x - \partial_y$$

then

$$X_p(x^2 + y^2 + z^2) = (2x - 2y)|_p = 0.$$

Therefore $X_p \in T_pS^2$.

Let's compute the tangent space $T_pS_r^2$ by finding a basis. In the coordinate basis we have by equation 7.1,

$$[F_*] = (2x, 2y, 2z)$$

where x, y, z satisfy $x^2 + y^2 + z^2 = r^2$. In order to compute the kernel note the following, if $z \neq \pm r$, then

$$\ker[F_*] = \mathrm{span}\{(-y, x, 0), (0, -z, y)\}.$$

Rewriting this in the standard basis we have

$$T_pS_c^2 = \mathrm{span}\{-y\partial_x + x\partial_y, -z\partial_y + y\partial_z\} \quad p \neq (0, 0, \pm r).$$

At the point $p = (0, 0, \pm r)$ we have

$$T_pS_c^2 = \mathrm{span}\{\partial_x, \partial_y\} \quad p = (0, 0, \pm r).$$

Example 7.1.4. Let $z = f(x, y)$, $(x, y) \subset U$ where U is an open set in $\mathbb{R}^2$. As usual we let $F(x, y, z) = z - f(x, y)$, so that the graph $z = f(x, y)$ is written as the level surface $F = 0$. We compute in the coordinate basis

$$[F_*] = (-f_x, -f_y, 1)$$

and so the surface is a regular level surface. At a point $p(x_0, y_0, f(x_0, y_0)) \in S$ we have

$$T_pS = span\{\ \partial_x + f_x(x_0, y_0)\partial_z\ ,\ \partial_y + f_y(x_0, y_0)\partial_z\ \}.$$

Let $\sigma : I \to \mathbb{R}$ be a smooth curve lying on the surface S, which means

$$F \circ \sigma = c.$$

Applying the chain rule to the function $F \circ \sigma : \mathbb{R} \to \mathbb{R}$ (or differentiating with respect to t) gives

$$(F \circ \sigma)_* \partial_t = F_* \sigma_* \partial_t = c_* \partial_t = 0.$$

Therefore $\sigma_* \partial_t \in T_{\sigma(t)}S$. In the next section we will answer the question of whether for every tangent vector $X_p \in T_pS$ there exists a representative curve σ for X_p lying on S.

7.2 Patches and Covers

Let $S \subset \mathbb{R}^{n+1}$ be a regular level surface.

Definition 7.2.1. A coordinate patch (or coordinate chart) on S is a pair (U, ψ) where

1. $U \subset \mathbb{R}^n$ is open,
2. $\psi : U \to \mathbb{R}^{n+1}$ is a smooth injective function,
3. $\psi(U) \subset S$,
4. and ψ_* is injective at every point in U (so ψ is an immersion)

A coordinate patch about a point $p \in S$ is a coordinate patch (U, ψ) with $p \in \psi(U)$. The function ψ provides coordinates on the set $\psi(U) \subset S$.

Example 7.2.2. Recall the regular level surface $S^2_r = \{(x,y,z) \in I\!R^3 \mid x^2 + y^2 + z^2 = r^2,\ r \in I\!R^+\}$. Let $U = (0,2\pi) \times (0,\pi) \subset I\!R^2$, which is clearly open. The function $\psi : U \to I\!R^3$ given by

$$\psi(u,v) = (r\cos u \sin v, r \sin u \sin v, r\cos v) \tag{7.2}$$

is a surface patch on the 2-sphere S^2_r.

Example 7.2.3. Let S be the regular level surface defined by a graph $z = f(x,y)$, $(x,y) \in U$ open. let $\psi : U \to I\!R^3$ be the function

$$\psi(u,v) = (u,v,f(u,v)) \quad (u,v) \in U.$$

The conditions for a patch are easily checked. Every point $(x_0,y_0,z_0) \in S$ is contained in the given patch.

Suppose (U,ψ) is a patch on a regular level surface S. Then $\psi : U \to S$ is a one-to-one immersion. The differential ψ_* is injective by definition and so $\psi_*(T_xU) \subset T_{\psi(x)}I\!R^{n+1}$ is an n-dimensional subspace. However we find even more is true.

Lemma 7.2.4. *The map $\psi_* : T_pU \to T_qS$, where $q = \psi(p)$ is an isomorphism.*

Proof. We have $\psi_*(T_pU)$ and T_qS are n-dimensional subspaces of $T_qI\!R^{n+1}$. If $\psi_*(T_pU) \subset T_pS$ then they are isomorphic. Let $X_p \in T_pU$, we only need to check that $\psi_* X_p \in \ker F_*$. We compute

$$F_*\psi_*X_p = X_p(F \circ \psi).$$

However by the patch conditions $F \circ \psi = c$ and so $X_p(F\circ\psi) = 0$. Therefore Lemma 7.1.2 implies $\psi_*X_p \in T_qS$, and so $psi_*(T_pU) = T_qS$. □

Example 7.2.5. Continuing with example 7.2.2, let $(u,v) \in (0,2\pi) \times (0,\pi)$ we find $\psi_*\partial_u|_{(u,v)} \in T_pS^2$ is

$$\psi_*\partial_u|_{(u,v)} = (-\sin u \sin v \partial_x + \cos u \sin v \partial_y)_{(\cos u \sin v, \sin u \sin v, \cos v)}.$$

Note that one can check that $\psi_*\partial_u \in T_pS^2$ using Lemma 7.1.2.

Example 7.2.6. For the example of a graph in 7.2.3 we compute

$$\psi_*\partial_u|_{(u_0,v_0)} = \partial_x + f_x(u_0,v_0)\partial_z\ ,\ \psi_*\partial_v|_{(u_0,v_0)} = \partial_y + f_y(u_0,v_0)\partial_z.$$

Let (U, ψ) be a patch on a regular level surface S, and let $q \in S$ be a point contained in $\psi(U)$. We now argue that given any $X_q \in T_qS$ there exists a curve $\sigma : I \to S$ with $\dot{\sigma}(0) = X_q$. By Lemma 7.2.4 let $Y_p \in T_pU$ with $\psi_* Y_p = X_q$, which exists and is unique since ψ_* is an isomorphism. Let σ be a representative curve for Y_p in $U \subset I\!R^n$. Then $\psi \circ \sigma$ is a representative curve for X_p. This follows from the chain rule,

$$\frac{d}{dt}\psi \circ \sigma|_{t=0} = \psi_* \sigma_* \left.\frac{d}{dt}\right|_{t=0} = \psi_* Y_p = X_q.$$

A *covering* of a regular level surface is a collection of surface patches $C = (U_\alpha, \psi_\alpha)_{\alpha \in A}$ where

$$S = \bigcup_{\alpha \in A} \psi_\alpha(U_\alpha).$$

In other words every point $p \in S$ is contained in the image of some surface patch.

Example 7.2.7. We continue with $S_r^2 = \{(x, y, z) \in I\!R^3 \mid x^2 + y^2 + z^2 = r^2,\ r > 0\}$. Let $D = \{(u, v) \in I\!R^2 \mid u^2 + v^2 < r^2\}$ which is an open set in $I\!R^2$. The set D with the function $\psi_{z^+} : D \to I\!R^3$ given by

$$\psi_{z^+}(u, v) = (x = u, y = v, z = \sqrt{r^2 - u^2 - v^2})$$

is a surface patch on S_r^2 (the upper hemi-sphere). Likewise the pair (D, ψ_{z^-}) is a surface patch on S_r^2 (the bottom hemi-sphere) where

$$\psi_{z^-}(u, v) = (x = u, y = v, z = -\sqrt{r^2 - u^2 - v^2}).$$

Continuing in the way we construct four more patches all using D and the functions

$$\psi_{x^\pm}(u, v) = (x = \pm\sqrt{r^2 - u^2 - v^2}, y = u, z = v)$$
$$\psi_{y^\pm}(u, v) = (x = u, y = \pm\sqrt{r^2 - u^2 - v^2}, z = v).$$

The collection

$$C = \{(D, \psi_{z^\pm}), (D, \psi_{x^\pm}), (D, \psi_{y^\pm})\ \}$$

is a cover of S_r^2 by coordinate patches.

The fact is regular level surfaces always admit a cover. This follows from the next theorem which we won't prove.

Theorem 7.2.8. *Let $S \subset I\!R^{n+1}$ be a regular level hyper-surface, and let $p \in S$. There exists a exists a surface patch (U, ψ) with $p \in \psi(U)$.*

The proof of this theorem involves the implicit function theorem from advanced calculus, see [12] for the theorem.

Corollary 7.2.9. *Let $S \subset I\!R^{n+1}$ be a regular level hyper-surface. There exists a cover $(U_\alpha, \psi_\alpha), \alpha \in A$ of S.*

7.3 Maps between surfaces

Suppose $S \subset I\!R^{n+1}$ and $\Sigma \subset I\!R^{m+1}$ are two regular level surfaces and that $\Phi : I\!R^n \to I\!R^m$. We'll say that Φ restricts to a smooth map from S to Σ if

$$\Phi(p) \in \Sigma \quad for\ all\ p \in S.$$

Example 7.3.1. Let $S^n \subset I\!R^{n+1}$ be the standard n-sphere, and consider the function $\Phi : I\!R^{n+1} \to I\!R^{n+1}$ given by $\Phi(p) = -p$. The function Φ restricts to a smooth function from $S^n \subset I\!R^{n+1}$ to $S^n \subset I\!R^{n+1}$. More generally let $A \in M_{n+1,n+1}(I\!R)$ where $A^T A = I$. Define the function $\Phi_A : I\!R^{n+1} \to I\!R^{n+1}$ by

$$\Phi_A(\mathbf{x}) = A\mathbf{x}.$$

The function Φ_A is linear, and so smooth. If $\mathbf{x} \in S^n$ (so $\mathbf{x}^T\mathbf{x} = 1$), then

$$[\Phi_A(\mathbf{x})]^T \Phi_A(\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T\mathbf{x} = 1.$$

Therefore Φ_A restricts to a smooth map $\Phi_A : S^n \to S^n$.

Example 7.3.2. Let $S \subset I\!R^3$ be the regular level surface

$$S = \{(u, v, w) \mid 4u^2 + 9v^2 + w^2 = 1\ \}.$$

The function $\Phi : I\!R^3 \to I\!R^3$ given by

$$\Phi(x, y, z) = (2x, 3y, z)$$

restricts to a smooth map from S^2 to Σ.

Example 7.3.3. Let $\Phi : I\!R^3 \to I\!R^5$ be given by

$$\Phi(x, y, z) = (x, y, z, 0, 0),$$

then Φ restricts to a smooth function from S^2 to S^4.

A smooth map $\Phi : S \to \Sigma$ is said to be an immersion if $\Phi_* : T_pS \to T_{\Phi(p)}\Sigma$ is injective for each $p \in S$, and a submersion if Φ_* is surjective for each p. A general notion of a smooth function $\Phi : S \to \Sigma$ which does not necessarily come from a function on the ambient $I\!R$ space is given in more advanced courses.

7.4 More General Surfaces

Another type of surface that is often encountered in multi-variable calculus is a parameterized surface. An example is $S \subset I\!R^3$ given by

$$S = \{(x = s\cos t, y = s\sin t, z = t), \quad s, t \in I\!R^2\}$$

which is known as the helicoid. With this description of S it is unclear whether S is actually a level surface or not. It is possible to define what is known as a parameterized surface, but let's look at the general definition of a regular surface which includes parameterized surfaces and regular level surfaces.

Definition 7.4.1. A regular surface $S \subset I\!R^{n+1}$ is a subset with the following properties,

1. for each point $p \in S$ there exists an open set $U \subset I\!R^n$ and a smooth injective immersion $\psi : U \to I\!R^{n+1}$ such that $p \in \psi(U) \subset S$,
2. and furthermore there exists an open set $V \in I\!R^{n+1}$ such that $\psi(U) = S \cap V$.

One important component of this definition is that S can be covered by surface patches. The idea of a cover is fundamental in the definition of a manifold.

7.5 Metric Tensors on Surfaces

Consider the coordinate patch on the upper half of the unit sphere $S^2 \subset I\!R^3$ in example 7.2.7 with $r = 1$, given by the function $\psi : U \to I\!R^3$ where U is the inside of the unit disk $u^2 + v^2 < 1$, and

$$\psi(u, v) = (\, x = u,\ y = v,\ z = \sqrt{1 - u^2 - v^2}\,). \tag{7.3}$$

We can view the disk $u^2+v^2<1$ as lying in the xy-plane and the image under ψ as the upper part of S^2. Let $\sigma(t)=(x(t),y(t),z(t))$ be a curve on the upper half of the sphere. The curve σ is the image of the curve $\tau(t)=(u(t),v(t))\subset U$ which is the projection of σ into the xy-plane. In particular from equation 7.3 we have $u(t)=x(t), v(t)=y(t)$ and therefore have

$$\sigma(t)=\psi\circ\tau(t)=(u(t),v(t),\sqrt{1-u(t)^2-v(t)^2}) \tag{7.4}$$

Given the curve (7.4) on the surface of the sphere we can compute its arc-length as a curve in $I\!R^3$ using the Euclidean metric γ_{Eu} (see equation **??**),

$$\begin{aligned} L_{Eu}(\sigma) &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2}\,dt \\ &= \int_a^b \sqrt{\left(\frac{du}{dt}\right)^2+\left(\frac{dv}{dt}\right)^2+\left(\frac{1}{\sqrt{1-u(t)^2-v(t)^2}}(u\frac{du}{dt}+v\frac{dv}{dt})\right)^2}\,dt \\ &= \int_a^b \sqrt{\frac{1}{1-u^2-v^2}\left((1-v^2)(\dot u)^2+2uv\dot u\dot v+(1-u^2)(\dot v)^2\right)}\,dt \end{aligned}$$

Note that this is the same arc-length we would have computed for the curve $\tau(t)=(u(t),v(t))$ using the metric tensor

$$\hat\gamma_U=\frac{1}{1-u^2-v^2}\left((1-v^2)du+2uvdudv+(1-u^2)dv^2\right) \tag{7.5}$$

defined on the set U!

Let's look at this problem in general and see where the metric tensor 7.5 comes from.

Suppose $S\subset I\!R^{n+1}$ is a regular level surface. Let γ be a metric tensor on $I\!R^{n+1}$. The metric $\boldsymbol{\gamma}$ induces a metric tensor on S, we denote by $\boldsymbol{\gamma}_S$ as follows. Let $q\in S$, and $X,Y\in T_qS$, then

$$\boldsymbol{\gamma}_S(X,Y)=\boldsymbol{\gamma}(X,Y). \tag{7.6}$$

This is well defined because $T_qS\subset T_qI\!R^{n+1}$. The function $\gamma_S: T_qS\times T_qS\to I\!R$ easily satisfies the properties of bi-linearity, symmetric, positive definite and is an inner product. The algebraic properties are then satisfied for $\boldsymbol{\gamma}_S$ to be a metric tensor are satisfied, but what could smoothness be?

Suppose that $\psi : U \to I\!R^{n+1}$ is a smooth patch on S. We now construct a metric tensor on $U \subset I\!R^n$ which represents $\boldsymbol{\gamma}_S$ as above. We define $\hat{\boldsymbol{\gamma}}_U$ as the metric tensor $\boldsymbol{\gamma}_S$ in the patch (U, ψ) in a point-wise manner as

$$\hat{\boldsymbol{\gamma}}_U(X_p, Y_p) = \boldsymbol{\gamma}_S(\psi_* X_p, \psi_* Y_p), \quad X_p, Y_p \in T_{\mathbf{u}}U,\ p \in U. \tag{7.7}$$

We now claim that is $\hat{\boldsymbol{\gamma}}_U = \psi^* \boldsymbol{\gamma}$! Expanding equation 7.7 using 7.6 we have

$$\hat{\boldsymbol{\gamma}}_U(X_p, Y_p) = \boldsymbol{\gamma}_S(\psi_* X_p, \psi_* Y_p) = \boldsymbol{\gamma}(\psi_* X_p, \psi_* Y_p) \quad X_p, Y_p \in T_p U.$$

Therefore by definition (see 6.33),

$$\hat{\boldsymbol{\gamma}}_U = \psi^* \boldsymbol{\gamma}. \tag{7.8}$$

Finally according to Theorem 6.2.6, $\boldsymbol{\gamma}_U$ is a Riemannian metric tensor on U. We then define $\boldsymbol{\gamma}_S$ to be smooth because $\hat{\boldsymbol{\gamma}}_U$ is smooth on any chart on S.

Example 7.5.1. Using the chart on S^2 from equation (7.3), we find

$$\psi^* dx = du,\ \psi^* dy = dv,\ \psi^* dz = -\frac{u}{\sqrt{1-u^2-v^2}}du - \frac{v}{\sqrt{1-u^2-v^2}}dv.$$

Computing the induced metric using 7.8, $\psi^*(dx^2 + dy^2 + dz^2)$ we get $\hat{\boldsymbol{\gamma}}_U$ in equation (7.5).

Example 7.5.2. Let $z = f(x, y)$ be a surface in $I\!R^3$, an let $\psi : I\!R^2 \to I\!R^3$ be the standard patch $\psi(x, y) = (x, y, z = f(x, y))$. We computed

$$\psi^* \boldsymbol{\gamma}_{Eu} = (1 + f_x^2)dx^2 + 2f_x f_y dxdy + (1 + f_y^2)dy^2,$$

in example 6.2.4 of section 6.2. This is the metric tensor on a surface in $I\!R^3$ given by a graph.

Example 7.5.3. Let $S = \{(w, x, y, z) \mid x^2 + y^2 + z^2 - w^2 = 1\ \}$, and let (U, ψ) be the coordinate patch on S,

$$\psi(t, u, v) = (w = t, x = \sqrt{t^2+1}\cos u \sin v, y = \sqrt{t^2+1}\sin u \sin v, z = \sqrt{t^2+1}\cos v),$$

where $U = \{(t, u, v) \mid t \in I\!R\ u \in (0, 2\pi), v \in (0, \pi)\}$. With the Euclidean metric on $I\!R^4$, the components of the surface metric γ_S in the patch (U, ψ)

using equation 7.8 are computed using

$$\begin{aligned}
\psi^* dx &= \frac{t}{\sqrt{t^2+1}}\cos u \sin v\, dt - \sqrt{t^2+1}\sin u \sin v\, du + \sqrt{t^2+1}\cos u \cos v\, dv,\\
\psi^* dy &= \frac{t}{\sqrt{t^2+1}}\sin u \sin v\, dt - \sqrt{t^2+1}\cos u \sin v\, du + \sqrt{t^2+1}\sin u \cos v\, dv,\\
\psi^* dz &= \frac{t}{\sqrt{t^2+1}}\cos v\, dt + \sqrt{t^2+1}\sin v\, dv\\
\psi^* dw &= dt.
\end{aligned}$$

Therefore

$$\begin{aligned}
\psi^*\boldsymbol{\gamma}_{Eu} &= (\psi^* dx)^2 + (\psi^* dy)^2 + (\psi^* dz)^2 + (\psi^* dw)^2\\
&= \frac{2t^2+1}{t^2+1}dt^2 + (t^2+1)(\sin^2 v\, du^2 + dv^2).
\end{aligned}$$

Remark 7.5.4. If γ is not positive definite then the signature of γ_S will depend on S.

We now come to a very important computation. Suppose that S is a regular level hypersurface, $\boldsymbol{\gamma}$ a metric tensor on $I\!R^{n+1}$ and that the corresponding metric tensor on S is $\boldsymbol{\gamma}_S$. Let (U,ψ) and (V,ϕ) be two coordinate patches on S which satisfy $\psi(U)\cap\phi(V) \neq \{\}$. On each of these open sets $U, V \subset I\!R^n$ let $\hat{\boldsymbol{\gamma}}_U$, and $\hat{\gamma}_V$ be the induced metric tensor as defined by equation (7.6) on the sets U and V respectively. The question is then how are $\hat{\boldsymbol{\gamma}}_U$ and $\hat{\boldsymbol{\gamma}}_V$ related? In other words how are the coordinate forms of the metric on S related at points of S in two different coordinate patches?

Let $W = \psi(U)\cap\phi(V)$ which is a non-empty subset of S, and let $U_0 = \{u \in U \mid \psi(u) \in W\}$, and $V_0 = \{v \in V \mid \psi(v) \in W\}$. The functions ψ and ϕ are injective and so $\psi : U_0 \to W$ and $\phi : V_0 \to W$ are then bijective. Consequently we have the bijection

$$\phi^{-1}\circ\psi : U_0 \to V_0, \tag{7.9}$$

where $\phi^{-1} : W \to V_0$. While the function $\phi^{-1}\circ\psi$ exists, its explicit determination is not always easy.

The functions ψ and ϕ provide two different coordinate systems for the points of S which lie in W, and 7.9 are the *change of coordinate functions* for the points in W.

Example 7.5.5. Let (U, ψ) be the chart on S^2 from equation (7.3) (upper half of sphere), and let (V, ϕ) be the chart

$$\phi(s,t) = (x = s, y = \sqrt{1 - s^2 - t^2}, z = t)$$

where V is the interior of the unit disk $V = \{(s,t) \mid s^2 + t^2 < 1\ \}$. The set W is then set

$$Drawpicture,$$

and

$$U_0 = \{\ (u,v) \mid u^2 + v^2 = 1, v > 0\ \}, \quad V_0 = \{\ (s,t) \mid s^2 + t^2 = 1, t > 0\ \}.$$

In order to compute $\phi^{-1} \circ \psi : U_0 \to V_0$ we use the projection map $\pi(x,y,z) = (s = x, t = z)$ which maps the right hemisphere to V. Therefore

$$\phi^{-1} \circ \psi(u,v) = \pi(u, v, \sqrt{1 - u^2 - v^2}) = (s = u, t = \sqrt{1 - u^2 - v^2}).$$

Example 7.5.6. We now let (U, ψ) be the coordinate patch on S^2 in example 7.2.2 given in equation 7.2 by

$$\psi(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi),\ 0 < \theta < 2\pi,\ 0 < \phi < \pi,$$

and let (V, ζ) be the patch

$$\zeta(u,v) = (u, v, \sqrt{1 - u^2 - v^2}),\ u^2 + v^2 < 1.$$

The overlap on S^2 consists of points in the top half of S^2 minus those with $y = 0, x \geq 0$, and

$$U_0 = \{(\theta,\phi) \mid 0 < \theta < 2\pi,\ 0 < \phi < \frac{\pi}{2},\ V_0 = \{(u,v) \mid u^2 + v^2 < 1\} - \{(u,v) \mid v = 0, 0 < u < 1\}.$$

The function $\lambda : U_0 \to V_0$ is again easily determined using the projection map $\pi(x,y,z) = (x,y)$ to give

$$\lambda(\theta,\phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi),\ (\theta,\phi) \in U_0.$$

We now find the relationship between $\hat{\gamma}_U$ and $\hat{\gamma}_V$. Assume that the change of coordinate functions $\lambda : U_0 \to V_0$

$$\lambda = \phi^{-1} \circ \psi. \tag{7.10}$$

Using this definition we have

$$\psi = \phi \circ \lambda.$$

We claim that

$$\hat{\boldsymbol{\gamma}}_U = \lambda^* \hat{\boldsymbol{\gamma}}_V. \tag{7.11}$$

To check this we find by equation 7.8 that,

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}_U &= \psi^* \boldsymbol{\gamma} \\
&= (\phi \circ \lambda)^* \boldsymbol{\gamma} \\
&= \lambda^* \phi^* \boldsymbol{\gamma} \quad \text{by the chain} - \text{rule} \\
&= \lambda^* \hat{\boldsymbol{\gamma}}_V
\end{aligned}$$

where $\boldsymbol{\gamma}$ is a metric on $I\!R^{n+1}$.

Example 7.5.7. Let's check the formula 7.11 using λ from example 7.5.6. We have from 7.5

$$\hat{\boldsymbol{\gamma}}_{V_0} = \frac{1}{1 - u^2 - v^2} \left((1 - v^2) du + 2uvdudv + (1 - u^2) dv^2 \right)$$

We compute using

$$\lambda^* du = \lambda^* du = -\sin\theta \sin\phi d\theta + \cos\theta \cos\phi d\phi, , \ \lambda^* dv = \cos\theta \sin\phi d\theta + \sin\theta \cos\phi d\phi,$$

$$\begin{aligned}
\lambda^* \hat{\boldsymbol{\gamma}}_{V_0} = &\frac{1 - \sin^2\theta \sin^2\phi}{\cos^2\phi} \left(-\sin\theta \sin\phi d\theta + \cos\theta \cos\phi d\phi \right)^2 \\
&+ \frac{2 \cos\theta \sin\theta \sin^2\phi}{\cos^2\phi} \left(-\sin\theta \sin\phi d\theta + \cos\theta \cos\phi d\phi \right) \left(\cos\theta \sin\phi d\theta + \sin\theta \cos\phi d\phi \right) \\
&+ \frac{1 - \cos^2\theta \sin^2\phi}{\cos^2\phi} \left(\cos\theta \sin\phi d\theta + \sin\theta \cos\phi d\phi \right)^2,
\end{aligned}$$

which expands to (noting that the term with $d\theta d\phi$ is zero after expanding),

$$\begin{aligned}
&\frac{(1 - \sin^2\theta \sin^2\phi) \sin^2\theta \sin^2\phi - 2\cos^2\theta \sin^2\theta \sin^4\phi + (1 - \cos^2\theta \sin^2\phi) \cos^2\theta \sin^2\phi}{\cos^2\phi} d\theta^2 \\
&+ \left((1 - \sin^2\theta \sin^2\phi) \cos^2\theta\phi + 2\cos^2\theta \sin^2\theta \sin^2\phi + (1 - \cos^2\theta \sin^2\phi) \sin^2\theta \right) d\phi^2
\end{aligned}$$

finally giving

$$\lambda^* \hat{\boldsymbol{\gamma}}_{V_0} = \sin^2\phi d\theta^2 + d\phi^2 = \hat{\boldsymbol{\gamma}}_{U_0}.$$

Example 7.5.8. Let's check formula (7.11) using the two patches on S^2 from example (7.5.5) above. Here we have

$$\hat{\boldsymbol{\gamma}}_U = \frac{1}{1-u^2-v^2}\left((1-v^2)du^2 + 2uvdudv + (1-u^2)dv^2\right)$$

$$\boldsymbol{\gamma}_V = \frac{1}{1-s^2-t^2}\left((1-t^2)ds^2 + 2stdsdt + (1-s^2)dt^2\right)$$

using $\lambda(u,v) = (u, \sqrt{1-u^2-v^2})$, then

$$\lambda^* ds = du, \quad \lambda^* dt = -\frac{u}{\sqrt{1-u^2-v^2}}du - \frac{v}{\sqrt{1-u^2-v^2}}dv.$$

Therefore

$$\begin{aligned}
\lambda^*\hat{\boldsymbol{\gamma}}_V &= \frac{1}{v^2}\left((u^2+v^2)du^2 - 2udu(udu+vdv) + \frac{1-u^2}{1-u^2-v^2}(udu+vdv)^2\right)\\
&= \frac{1}{v^2}\left((u^2+v^2) - 2u^2 + \frac{(1-u^2)^2}{1-u^2-v^2}\right)du^2\\
&\quad + \frac{1}{v^2}\left(\frac{2uv(1-u^2)}{1-u^2-v^2} - 2uv\right)dudv + \frac{1-u^2}{1-u^2-v^2}dv^2\\
&= \frac{1}{1-u^2-v^2}\left((1-v^2)du^2 + 2uvdudv + (1-u^2)dv^2\right)
\end{aligned}$$

7.6 Exercises

1. Find a basis for the tangent spaces (as subspaces of the tangent space to the ambient $\mathbb{R}^{n+1}$) to the following hypersurfaces at the indicated points in two different ways. First computing the kernel of F_*, and then by constructing a sufficient number of curves on the level surface which pass through the given point. Check the answers are the same.

   (a) $z^2 - xy = 3, \quad p = (1, 1, 2)$.

   (b) $xw - yz = 1, \quad p = (1, 1, 0, 0)$.

2. Let $S \subset \mathbb{R}^3$ be the set of points

   $$S = \{ \ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 1\} \tag{7.12}$$

   (a) Show that S is a regular level surface, and sketch it.

   (b) Find a basis for T_qS at $q = (1, 1, 1)$ using curves, and derivations.

   (c) Show that $\psi : U \to \mathbb{R}^3$, where $U = (0, 2\pi) \times \mathbb{R}$ given by

   $$\psi(u, v) = (\sqrt{v^2+1}\cos u, \sqrt{v^2+1}\sin u, v) \tag{7.13}$$

   is a smooth patch on S.

   (d) Find $p = (u_0, v_0) \in U$ and $Y_p \in T_pU$ such that $\psi(p) = (1, 1, 1) \in S$ and $\psi_* Y_p = (\partial_x - \partial_y)_q$.

   (e) Find a cover of S by patches of which one is ψ from equation 7.13.

3. Find a cover of S^2_r containing two patches, where one patch is in example 7.2.2.

4. Let $\Sigma \in \mathbb{R}^3$ be the regular level surface

   $$\Sigma = \{ \ (u, v, w) \mid w^2 - u^2 - v^2 = -16\},$$

   and let $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ be

   $$\Phi(x, y, z) = (u = 4x, v = 3z + 5y, w = 5z + 3y).$$

   (a) Show that Φ restricts to a smooth map $\Phi : S \to \Sigma$ where S is the surface in 7.12 from problem 2.

(b) Choose a tangent vector $X_p \in T_pS$, where $p = (1,1,1)$, and compute

$$\Phi_* X_p$$

using curves and derivations. Write your answer a basis for $T_q\sigma$ where $q = (4,8,8)$.

(c) Is Φ an immersion or submersion (of surfaces) ?

5. Let $\psi_1, \psi_2 : I\!R^2 \to I\!R^3$, where

$$\psi^1(u,v) = \left(\frac{2u}{1+u^2+v^2}, \frac{2v}{1+u^2+v^2}, \frac{u^2+v^2-1}{1+u^2+v^2}\right)$$
$$\psi^2(u,v) = \left(\frac{2u}{1+u^2+v^2}, \frac{2v}{1+u^2+v^2}, \frac{1-u^2-v^2}{1+u^2+v^2}\right).$$

Show that this is a covering of the unit sphere S^2 by coordinate patches. These patches are called the stereographic charts on S^2.

6. Let $L_\pm$ be the line in $I\!R^3$ with end point $(0,0,\pm 1)$ and passing through the point $(u,v,0) \in I\!R^3$. Compute the (x,y,z) coordinates of the point on S^2 where $L_\pm$ intersects S^2 in terms of u and v. Relate this to the previous problem.

7. The Veronese map from $\Phi : I\!R^3 \to I\!R^5$ is defined by

$$\Phi(x,y,z) = (xy, xz, yz, \frac{1}{2}x^2 - \frac{1}{2}y^2, \frac{\sqrt{3}}{6}(x^2+y^2-2z^2).$$

(a) Show that Φ restricts to a map $\Phi : S^2 \to S^4_{\sqrt{3}/3}$ from the unit sphere in $I\!R^3$ to the sphere of radius $\sqrt{3}/3$ in $I\!R^5$.

(b) Find basis for $T_p(S^2)$ and $T_q(S^4_{\sqrt{3}/3})$ at $p = (\sqrt{2}/2, 0, \sqrt{2}/2)$ and $q = \Phi(p)$ and compute the push-forward $\Phi_* : T_p(S^2) \to T_q(S^4_{\sqrt{3}/3})$ with respect to your basis.

(c) Show that $\Phi : S^2_1 \to S^4_{\sqrt{3}/3}$ is an immersion.

8. Find the the metric $\boldsymbol{\gamma}_{S^2_r}$ in the chart in example 7.2.2.

9. Let γ_{S^2} be the metric tensor on S^2 induced from the Euclidean metric on $I\!R^3$.

(a) Compute the coordinate form of the metric γ_{S^2} on the two stereographic charts $\psi^{1,2}$ from question 5.

(b) Compute the change of coordinate functions and identify the domain on S^2 where this is valid. (Draw a picture).

(c) Check the overlap formula (7.11) and identify the domain where this is valid.

Chapter 8

Flows, Invariants and the Straightening Lemma

8.1 Flows

In equation (6.47) of example 6.3.5 the family of diffeomorphisms $\Psi_t : \mathbb{R}^2 \to \mathbb{R}^2$ are given

$$\Psi_t(x, y) = (x + t, y)$$

which are parameterized by $t \in \mathbb{R}$. We can rewrite the function Ψ_t as $\Psi : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\Psi(t, (x, y)) = (x + t, y).$$

This function has the properties

1) $\Psi(0, (x, y)) = (x, y)$

2) $\Psi(s, \Psi(t, (x, y))) = \Psi(s, (x + t, y)) = (x + t + s, y) = \Psi(s + t, y)$.

The function Ψ is an example of a flow.

Definition 8.1.1. A (global) *flow* on $\mathbb{R}^n$ is a smooth function $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ which satisfies the two properties

1. $\Psi(0, p) = p, \quad \textit{for all } p \in \mathbb{R}^n$, and

2. $\Psi(s, \Psi(t, p)) = \Psi(s + t, p), \quad \textit{for all } s, t \in \mathbb{R},\ p \in \mathbb{R}^n$.

Another term often used for a flow is a *one parameter group of transformations*. The reason for this terminology is given in chapter 10.

Example 8.1.2. The smooth function $\Psi : I\!R \times I\!R^2 \to I\!R^2$,

$$\Psi(t,(x,y)) = (x\cos t - y\sin t, x\sin t + y\cos t) \tag{8.1}$$

is a flow on $I\!R^2$.

We will mainly be interested in cases where diffeomorphisms in flows Ψ which have geometric meaning as in example 6.3.5. In example 6.3.5 the flow consists of a family of isometries (one for each fixed value of $t \in I\!R$) for the Euclidean metric tensor $\boldsymbol{\gamma}_E$. In example 8.1.2 it is also the case that for each fixed value of $t \in I\!R$ the diffeomorphism Ψ_t is an isometry of Euclidean space.

Flows are closely related to vector-fields. In order to see why, first recall from section 3.1 that one way to represent a tangent vector $X_p \in T_pI\!R^n$ is by a curve $\sigma : I \to I\!R^n$, $p \in I \subset I\!R$, where $\sigma(0) = p$ and $X_p = \dot\sigma(0)$. Therefore one might want to consider a vector-field X as being given by a family of curves $\sigma_{\mathbf{x}}$ defined for each $\mathbf{x} \in I\!R^n$, on a family of intervals $I_{\mathbf{x}}$ each containing 0, and where at each point $\mathbf{x} \in I\!R^n$. Then $X_{\mathbf{x}} = \dot\sigma_{\mathbf{x}}|_{t=0}, \ \textit{for all}\ \mathbf{x} \in I\!R^n$. Flows allow us to do this.

Let $\Psi : I\!R \times I\!R^n \to I\!R^n$ be a flow, and let $p \in I\!R^n$. Fix $\mathbf{x} = p$ in the function Ψ consider the function $\psi_p : I\!R \to I\!R^n$ given by

$$\psi_p(t) = \psi(t,p).$$

The function $\psi_p(t)$ is a curve that has the property $\psi_p(0) = \Psi(0,p) = p$ from property 1 in definition 8.1.1 of flow. We can therefore compute the tangent vector to ψ_p at p. Call this tangent vector X_p and it is

$$X_p = \left.\frac{\partial \Psi(t,p)}{\partial t}\right|_{t=0}. \tag{8.2}$$

Now the point p in this formula was arbitrary and so equation 8.2 defines a vector field X whose value at a fixed point p is (8.2). We can write the derivative in equation 8.2 as a partial derivative (by the definition of partial derivative), to get the vector-field on $X = \sum_{i=1}^n X^i(\mathbf{x})\partial_{x^i}$ with

$$X^i(\mathbf{x}) = \left.\left(\frac{\partial}{\partial t}\Psi(t,\mathbf{x})\right)\right|_{t=0}. \tag{8.3}$$

These are smooth functions (since Ψ is), and so X is a smooth vector field. The vector field X is called the *infinitesimal generator* of the flow Ψ.

Example 8.1.3. Let Ψ be the flow equation (6.47). By equation 8.3 the corresponding vector field is just

$$X = \partial_x .$$

Example 8.1.4. Let Ψ be the flow from equation 8.1, then the corresponding vector field $X = X^1\partial_x + X^2\partial_y$ are computed by equation 8.3 to be

$$\begin{aligned} X^1 &= \left(\frac{\partial}{\partial t}(x\cos t - y \sin t)\right)\bigg|_{t=0} &= -y \\ X^2 &= \left(\frac{\partial}{\partial t}(x\sin t + y \cos t)\right)\bigg|_{t=0} &= x. \end{aligned}$$

Therefore the infinitesimal generator of the flow is

$$X = -y\partial_x + x\partial_y .$$

Example 8.1.5. Let $\Psi : I\!R \times I\!R^3 \to I\!R^3$ be the flow

$$\Psi(t,x,y,z) = (x + ty + \frac{1}{2}t^2 z, y + tz, z).$$

The corresponding infinitesimal generator is

$$X = y\partial_x + z\partial_y .$$

The tangent vector to a flow Ψ has the following property.

Lemma 8.1.6. *For each $t \in I\!R$ we have*

$$\frac{d\psi_p(t)}{dt} = X_{\psi(t,p)} . \tag{8.4}$$

The tangent vector to the curve $\psi_p(t)$ at any point on the curve is the infinitesimal generator at the point $\Psi(t,p)$.

Proof. Let $p_1 = \psi_p(t_1)$ be a point along the curve, and let $s = t - t_1$. Define the curve $\sigma(s) = \psi_p(s + t_1)$, and note

$$\frac{d\sigma(s)}{ds}\bigg|_{s=0} = \frac{d\psi_p(t)}{dt}\bigg|_{t=t_1} . \tag{8.5}$$

We also have property 2 in definition 8.1.1 of flow that

$$\sigma(s) = \Psi(s, \Psi(t_1, p)) = \Psi(s, p_1). \tag{8.6}$$

Therefore by equation 8.5 and 8.6,

$$\frac{d\psi_p(t)}{dt}_{t=t_1} = \left.\frac{d\sigma}{ds}\right|_{s=0} = \left.\left(\frac{d}{ds}\Psi(s, p_1)\right)\right|_{s=0} = X_{p_1}.$$

This proves (8.4). □

We have seen that given a flow Ψ on $I\!R^n$ how it determines a vector-field X on $I\!R^n$, we now consider the converse to this. Given a vector field X on $I\!R^n$ is it the infinitesimal generator of a flow Ψ on $I\!R^n$? Lemma 8.1.6 shows how they would be related. If X is a vector field on $I\!R^n$, and Ψ was a flow, then the curve $\sigma(t) = \Psi(t, p)$ satisfies the tangency condition in Lemma 8.1.6. If the vector field X is given in components by $X = \sum_{i=1}^n X^i \partial_{x^i}$ then σ satisfies the differential equation 8.4,

$$\frac{d\sigma^i}{dt} = X^i(\sigma(t)), \quad \sigma(0) = p.$$

To make this equation look a little more familiar, if we denote $\sigma(t) = (x^1(t), x^2(t), \ldots, x^n(t))$, and $p = (x_0^1, \ldots, x_0^n)$ then

$$\frac{dx^i}{dt} = X^i \quad x^i(0) = x_0^i, 1 \le i \le n. \tag{8.7}$$

Example 8.1.7. Consider the vector field on $I\!R^3$ from Example 8.1.5,

$$X = y\partial_x + z\partial_y.$$

The system of ordinary differential equations (8.7) is

$$\frac{dx}{dt} = y, \ \frac{dy}{dt} = z, \ \frac{dz}{dt} = 0, \quad x(0) = x_0, y(0) = y_0, z(0) = z_0. \tag{8.8}$$

Let's solve these. The last equation for z(and using the initial condition) gives $z = z_0$. Putting this in to the second equation gives $y = y_0 + tz_0$. Then using this in the first equation we get $x = x_0 + ty_0 + \frac{1}{2}t^2 z_0$, and the solution to equation 8.8 is

$$x(t) = x_0 + ty_0 + \frac{1}{2}t^2 z_0, \ y(t) = y_0, +tz_0, \ z(t) = z_0. \tag{8.9}$$

If in this formula we view (x_0, y_0, z_0) on the right-hand side of this equation as variables (x, y, z), then we can write equation

$$\Psi(t,(x,y,z)) = (x(t), y(t), z(t)) = (\ x + ty + \frac{1}{2}t^2 z,\ y,\ +tz, z_0\ ),$$

which is the flow in Example 8.1.5.

What we have done is consider the initial condition (x_0, y_0, z_0) as the initial (or domain) $\mathbb{R}^3$ for a flow Ψ . That is we think of the solution to the differential equations as the function $\Psi : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ where T is the amount of time we "flow" from the initial condition (x_0, y_0, z_0).

This example extends to the general case by the following Theorem.

Theorem 8.1.8. *Let $X = \sum_{i=1}^n X^i(\mathbf{x})\partial_{x^i}$ be a smooth vector field on $\mathbb{R}^n$, such that a solution $\mathbf{x}(t)$ to the differential equations*

$$\frac{dx^i}{dt} = X^i(\mathbf{x}), \quad x^i(0) = x^i_0 \tag{8.10}$$

exists for all $t \in \mathbb{R}$ and all initial conditions $\mathbf{x}_0 \in \mathbb{R}^n$. Then the function $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\Psi(t, \mathbf{x}_0) = \mathbf{x}(t)$$

is a flow. Furthermore, the infinitesimal generator corresponding to this flow is X.

A solution to the differential equations 8.10 for a fixed initial condition $\mathbf{x}_0$ is called an *integral curve* of the vector-field though the point $\mathbf{x}_0$ (see diagram above XXX). The flow Ψ consists of all integral curves.

Remark 8.1.9. *This theorem states that the operation of finding the flow of a vector field and finding the infinitesimal generator of a flow are inverse operations to each other!* (Subject to the condition that the solutions to 8.10 exist for all $t \in \mathbb{R}$).

Example 8.1.10. Find the flow corresponding to the vector field

$$X = (x + 3y)\partial_x + (4x + 2y)\partial_y.$$

According to Theorem 8.1.8 we need to solve the system of differential equations

$$\dot{x} = x + 3y\ , \quad \dot{y} = 4x + 2y$$

with initial conditions $x(0) = x_0, y(0) = y_0$. To do so we find the eigenvectors of the coefficient matrix

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}.$$

Two eigenvectors are

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

We then make a change of variables

$$\begin{bmatrix} x \\ y \end{bmatrix} = u \begin{bmatrix} 1 \\ -1 \end{bmatrix} + v \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Inverting this we have

$$u = \frac{4}{7}x - \frac{3}{7}y \ , \quad v = \frac{1}{7}x + \frac{1}{7}y.$$

Therefore

$$\begin{aligned} \dot{u} &= \frac{4}{7}\dot{x} - \frac{3}{7}\dot{y} = \frac{4}{7}(x + 3y) - \frac{3}{7}(4x + 2y) = -2u, \\ \dot{v} &= \frac{1}{7}(x + 3y) + \frac{1}{7}(4x + 2y) = 5v. \end{aligned}$$

Therefore

$$u = u_0 e^{-2t} \ , \quad v = v_0 e^{5t}.$$

We now back substitute,

$$\begin{aligned} x &= u_0 e^{-2t} + 3v_0 e^{5t} = (\frac{4}{7}x_0 - \frac{3}{7}y_0)e^{-2t} + \frac{3}{7}(x_0 + y_0)e^{5t} \\ y &= -u_0 e^{-2t} + 4v_0 e^{5t} = -(\frac{4}{7}x_0 - \frac{3}{7}y_0)e^{-2t} + \frac{4}{7}(x_0 + y_0)e^{5t} \end{aligned}$$

which after relabeling the initial conditions, gives the flow Ψ.

The system of ordinary differential equations in 8.10 are called autonomous (the right hand side does not depend on t), [7]. Solution to the differential equations (8.10) always exists in some interval about $t = 0$. If this interval can not be extended to the entire set $\mathbb{R}$, then the function Ψ defines what is known as a local flow. Local flows have infinitesimal generators and so a bijection correspondence exists between vector fields and certain local flows. In a more advanced course [4], local flows are known as flows.

Example 8.1.11. Let $X = x^2\partial_x$, then the flow equation 8.10 is

$$\frac{dx}{dt} = x^2, \quad x(0) = x_0.$$

The solution is

$$x(t) = \frac{1}{x_0 - t},$$

and the flow is

$$\Psi(t, x) = \frac{1}{x - t}$$

which is only defined for $t \neq x$.

8.2 Invariants

We begin this section with some ideas that we will study in more detail in the rest of the book. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism. A subset $S \subset \mathbb{R}^n$ is a Φ invariant subset if $\Phi(S) \subset S$, and a function $f \in C^\infty(\mathbb{R}^n)$ is a Φ invariant function if $\Phi^* f = f$. Given these definitions we now define what a flow invariant set and a flow invariant function are.

Let $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ be a flow. If we fix $t = t_0$ in the function Ψ, let $\Psi_{t_0} : \mathbb{R}^n \to \mathbb{R}^n$ be the smooth function

$$\Psi_{t_0}(\mathbf{x}) = \Psi(t_0, \mathbf{x}).$$

Now the function Ψ_{-t_0} has the property

$$\Psi_{-t_0} \circ \Psi_{t_0}(\mathbf{x}) = \Psi(-t_0, \Psi(t_0, \mathbf{x})) = \Psi(-t_0 + t_0, \mathbf{x}) = \Psi(0, \mathbf{x}) = \mathbf{x} = \Psi_{t_0} \circ \Psi_{-t_0}(\mathbf{x}).$$

Therefore Ψ_{t_0} is a diffeomorphism.

Definition 8.2.1. A subset $S \subset \mathbb{R}^n$ is a Ψ *invariant subset* if for all $t \in \mathbb{R}$, S is invariant with respect to every diffeomorphism Ψ_t.

In other words $S \subset \mathbb{R}^n$ is a Ψ invariant set if for all $t \in \mathbb{R}$ and $p \in S$, $\Psi(t, p) \in S$.

Example 8.2.2. For the flow $\Phi(t, (x, y)) = (x \cos t - y \sin t, x \sin t + y \cos t)$ from example 8.1.2, any set

$$S = \{ \ (x, y) \in \mathbb{R}^2 \mid a \leq x^2 + y^2 \leq b\}$$

where $0 \leq a \leq b$, $a, b \in I\!R$, is invariant. If $p = (x_0, y_0) \in S$ then

$$\Psi(t,p) = (x_0 \cos t - y_0 \sin t, x_0 \sin t + y_0 \cos t).$$

The sum of the squares of these entries is $x_0^2 + y_0^2$, and so $\Psi(t,p) \in S$.

Example 8.2.3. Let $\Psi(t,(x,y,x)) = (x + ty + \frac{t^2}{2}z, y + tz, z)$ be the flow from example 8.1.5. The set

$$S = \{ \ (x,y,0) \in I\!R^3 \ \}$$

is invariant. Let $p = (x_0, y_0, 0) \in S$, then

$$\Psi(t,p) = (x_0 + ty_0, y_0, 0).$$

It is easy to find other invariant subsets.

Example 8.2.4. Let Ψ be a flow, and let $S \subset I\!R^n$ be an invariant subset. Then the complement of S in $I\!R^n$, $I\!R^n - S$ is also a Ψ invariant set.

Definition 8.2.5. A function $f : I\!R^n \to I\!R$ is an invariant of the flow Ψ if f is an invariant function with respect to each diffeomorphism Ψ_t.

In other words, $f : I\!R^n \to I\!R$ is Ψ invariant if for all $t \in I\!R$, $\mathbf{x} \in I\!R^n$,

$$f(\mathbf{x}) = f(\Psi(t, \mathbf{x})). \tag{8.11}$$

Example 8.2.6. For the flow Ψ in example 8.1.2, the function $f(x,y) = x^2 + y^2$ is invariant. We check this by equation 8.11,

$$f(\Psi(t,(x,y)) = (x \cos t - y \sin t)^2 + (x \sin t + y \cos t)^2 = x^2 + y^2 = f(x,y).$$

If $S \subset I\!R^n$ is Ψ invariant, the definition of a Ψ invariant function $f : S \to I\!R$ is the same as that for $f : I\!R^n \to I\!R$.

Example 8.2.7. For the flow in example 8.1.5 above, the function $f(x,y,z) = z$ is an invariant. Let $S = \{(x,y,z) \in I\!R^3 \mid z \neq 0\}$, which is a Ψ invariant set. The function $g(x,y,z) = x - \frac{y^2}{2z}$ defined on S is Ψ invariant. We check that g is an invariant using equation 8.11 at points in S,

$$g(\Psi(t,(x,y,z))) = x + ty + \frac{t^2}{2}z - \frac{(y+tz)^2}{z} = x - \frac{y^2}{2z} = g(x,y,z)$$

Let X be a vector field on $\mathbb{R}^n$,

Definition 8.2.8. A function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$X(f) = 0 \tag{8.12}$$

is called an X-invariant function .

Example 8.2.9. For $X = -y\partial_x + x\partial_y$ the function $f(x, y) = x^2 + y^2$ is X-invariant,

$$X(x^2 + y^2) = -y(2x) + x(2y) = 0.$$

Example 8.2.10. For $X = y\partial_x + z\partial_y$ the function $f(x, y, z) = z$, and the function $g(x, y, x) = x - \frac{y^2}{2z}$ defined on $z \neq 0$ are X-invariant.

Example 8.2.11. Let (x, y, p_x, p_y) be coordinates on $\mathbb{R}^4$, and let

$$X = p_x\partial_x + p_y\partial_y - x\partial_{p_x} - y\partial_{p_y}.$$

This is the Hamiltonian vector field for the Hamiltonian $H = p_x^2 + p_y^2 + x^2 + y^2$. The functions

$$f_1 = yp_x - xp_y,\ p_x^2 + x^2,\ p_y^2 + y^2$$

are all X-invariants. The first invariant is the angular momentum.

The invariants in examples 8.1.2 and 8.2.9, and example 8.2.7 and 8.2.10 suggest a relationship between invariants of a flow and invariants of their corresponding infinitesimal generator.

Theorem 8.2.12. *Let Ψ be a flow with infinitesimal generator X. The function f is an invariant of Ψ if and only if $X(f) = 0$.*

Proof. To prove the only if direction suppose f is a Ψ invariant function, and take the derivative condition 8.11 with respect to t and setting $t = 0$ we get

$$\left(\frac{d}{dt} f(\Psi(t, \mathbf{x}))\right)\Big|_{t=0} = \left(\frac{d}{dt} f(\mathbf{x})\right)\Big|_{t=0} = 0$$

Expanding this equation and using equation 8.3 the left hand side becomes

$$\sum_{i=1}^{n} \frac{\partial \Psi^i}{\partial t}\Big|_{t=0} \frac{\partial f}{\partial x^i}\Big|_{\Psi(0,\mathbf{x})} = X(f) = 0$$

Suppose now that $f : I\!R^n \to I\!R$ is a function satisfying $X(f) = 0$. Let Ψ be the flow corresponding to X, and fix $p \in I\!R^n$. Now consider the function $\kappa : I\!R \to I\!R$ given by

$$\kappa(t) = f(\Psi(t, p)).$$

By taking the derivative of this equation and using the chain-rule, equations 8.3 and 8.4 (with $\sigma(t) = \Psi(t, p)$) we get,

$$\frac{d\kappa}{dt} = \sum_{i=1}^{n} \frac{\partial \Psi(t,p)}{dt} \frac{\partial f}{\partial x^i} = X_{\Psi(t,p)}(f) = 0.$$

Therefore κ is a function on $I\!R$ with zero derivative, and so is a constant. This argument did not depend on the original choice of p, and so $f(\Psi(t, \mathbf{x})) = \Psi(\mathbf{x})$ for all $t \in I\!R, x \in I\!R^n$. □

Finally we give an important definition in which the notion of invariants is conveniently described. Let X be a vector-field which is the infinitesimal generator of the flow of Ψ, and define the *Lie derivative of a function* $f \in C^\infty(I\!R^n)$ by,

$$\mathcal{L}_X f = \lim_{t \to 0} \frac{\Psi_p^* f - f(p)}{t} = X_p(f) \tag{8.13}$$

The Lie Derivative of f with respect to X at the point p measure the rate of change of f along the flow. It is given by the directional derivative $X_p(f)$.

Lemma 8.2.13. *The function f is an invariant of the flow Ψ if and only if*

$$\mathcal{L}_X f = 0$$

8.3 Invariants I

In Theorem 8.2.12 it was shown that finding invariants of a flow Ψ and its infinitesimal generator X are the same problems. We will describe two different ways to find these invariants, the first is based on the equation 8.12 $X(f) = 0$, and the second is based on the equation 8.11 $f(\Psi(t, \mathbf{x})) = f(\mathbf{x})$.

Suppose that

$$X = \sum_{i=1}^{n} X^i(\mathbf{x}) \partial_{x^i}$$

is a vector field on $\mathbb{R}^n$. In order to find an invariant function $f(\mathbf{x})$ we need to find a solution to the partial differential equation

$$\sum_{i=1}^{n} X^i(\mathbf{x}) \frac{\partial f}{\partial x^i} = 0 \tag{8.14}$$

The partial differential equation (8.14) is linear, and so the constants are solutions. These turn out to not be useful solutions.

Therefore the first method to finding invariants is to solve the partial differential equation (8.14). This theory is well developed, and in this section will just demonstrate it through a few examples. The method developed in the next section for finding invariants of flows solves (in a sense defined below) the partial differential equations 8.14.

To see how this works, suppose that $X = a(x,y)\partial_x + b(x,y)\partial_y$, and that we are searching for an invariant $f(x,y)$. Consider the ordinary differential equation

$$\frac{dy}{dx} = \frac{b(x,y)}{a(x,y)} \tag{8.15}$$

which is only defined where $a(x,y) \neq 0$. If

$$g(x,y) = c \tag{8.16}$$

is the general solution to equation 8.15, then $g(x,y)$ is an invariant of X. Let's check this in an example.

Example 8.3.1. Let $X = x\partial_x + y\partial_y$. The ode (8.15) is

$$\frac{dy}{dx} = \frac{y}{x}.$$

The general solution is $y = kx$, $k \in \mathbb{R}$ a constant. The function $\frac{y}{x}$ is easily checked to be X invariant. This can easily be generalized for special vector fields in higher dimension. For example

$$X = x\partial_x + y\partial_y + z\partial_z,$$

the functions xy^{-1} and xz^{-1} are invariant. This works because of missing variables in the coefficients of X. If

$$X = (x+y+z)\partial_x + (x+y)\partial_y + (x+z)\partial_z$$

then this would not work.

We now check that g in equation 8.16 is an invariant of X.

Proof. Let $(x_0, y_0) \in I\!R^2$ and let $y = f(x)$ be the solution to the differential equation

$$\frac{dy}{dx} = \frac{b}{a}, \quad y(x_0) = y_0.$$

If $g(x, y) = c_0$ is the general solution, by differentiating $g(x, f(x)) = c_0$ with respect to x and evaluating at x_0 we get

$$\begin{aligned}\left(\frac{d}{dx} g(x, f(x))\right)\bigg|_{x_0} &= \left(\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{df}{dx}\right)\bigg|_{x=x_0, y=y_0} \\ &= \left(\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{b}{a}\right)\bigg|_{x=x_0, y=y_0} \\ &= 0.\end{aligned}$$

Therefore

$$\left(a\frac{\partial g}{\partial x} + b\frac{\partial g}{\partial y}\right)_{x=x_0, y=y_0} = 0$$

and so $X(g) = 0$ at (x_0, y_0), which was arbitrary. □

A function F which is an invariant of the flow Ψ is also called a *constant of the motion*. To see why, fix $p \in I\!R^n$ and consider the curve $\psi_p(t)$. Then for each value $t \in I\!R$,

$$F(\psi_p(t)) = F(\psi(t, p)) = F(p)$$

we get the constant $F(p)$. That is, F is a constant as we move along one of flow curves (the motion).

We now turn to the question of how many invariants of a flow are there? Let $p_0 \in I\!R^n$, a set of C^∞ functions $f^1, f^2, \ldots, f^m$ where $m \leq n$ are *functionally independent* at the point p if the $m \times n$ Jacobian matrix

$$J_p = \left(\frac{\partial f^a}{\partial x^i}\bigg|_p\right)$$

has maximum rank m. In section 3.2 functionally independence was defined when $m = n$.

XXXX NEED SOME MORE INFO ON FUNCTIONALLY INDEPENDENT XXX

The basic theorem about the number of invariants is the following.

Theorem 8.3.2. *Let X be a vector field on $\mathbb{R}^n$, and let $p \in \mathbb{R}^n$. If $X_p \neq 0$, then there exists a non-empty open set $U \subset \mathbb{R}^n$ containing p, and $n-1$ X-invariant functions $f^1, \ldots, f^{n-1}$ which are functionally independent at each point in U. Furthermore if f is an X-invariant function defined at p, then there exists an open set V containing p and a function $F : \mathbb{R}^{n-1} \to \mathbb{R}^n$ such that*

$$f(\mathbf{x}) = F(f^1(\mathbf{x}), f^2(\mathbf{x}), \ldots, f^{n-1}(\mathbf{x})) \quad \mathbf{x} \in V.$$

This fundamental theorem is usually interpreted as saying the general solution to the partial differential equation $X(f) = 0$ is given by an arbitrary function of $n-1$ functionally independent invariants. Therefore when we say we have found all X invariant functions, or the general solution to the partial differential equation $X(f) = 0$, we mean that we have found $n-1$ functionally independent solutions.

Example 8.3.3. We continue with 8.2.10, where

$$X = y\partial_x + z\partial_y,$$

and two functions $u = x - \frac{y^2}{2z}$, $v = z$ are X invariant. The Jacobian is

$$\begin{pmatrix} 1 & -\frac{y}{z} & \frac{y^2}{2z^2} \\ 0 & 0 & 1 \end{pmatrix}$$

and so the two functions are functionally independent everywhere they are defined. By Theorem 8.3.2 the general solution to the partial differential equation

$$y\frac{\partial w}{\partial x} + z\frac{\partial w}{\partial y} = 0$$

is $w = F\left(z, x - \frac{y^2}{2z}\right)$ where $F : \mathbb{R}^2 \to \mathbb{R}$ is arbitrary.

Let $U \subset \mathbb{R}^n$ be an open set and X a vector field X on U. A set $f^1, \ldots, f^{n-1}$ of invariants of X that are functionally independent at each point of U are called a *complete set of invariants of X*.

Given a complete set of invariants $f^1, \ldots, f^{n-1}$, of a vector field X we can use these as local coordinates in a *partial straightening lemma*.

Lemma 8.3.4. *Let $f^1, \ldots, f^{n-1}$ be a complete set of invariants of X, and let $p \in U$. There exists a set of coordinates $y^1, \ldots, y^n$ defined on an open set V containing p such that*

$$X = a(\mathbf{y})\partial_{y^n}$$

Proof. Let f^n be a function such that $f^1, \ldots, f^n$ are functionally independent on some open set V of p. Therefore $y^1 = f^1, \ldots, y^n = f^n$ define coordinates on V. Now apply the change of variable formula for vector-fields in equation 4.26. Let $Y = \sum_{i=1}^n Y^i(\mathbf{y})\partial_{y^i}$ be the vector-field X in the new coordinates. By applying equation 4.27, the first $n-1$ components of the vector-field are

$$\begin{aligned} Y^1 &= Y(y^1) = X(y^1) = X(f^1) = 0, \\ Y^2 &= X(y^2) = X(f^2) = 0, \\ &\vdots, \\ Y^{n-1} &= X(y^{n-1}) = X(f^{n-1}) = 0. \end{aligned}$$

Therefore the vector field Y which is X in the coordinates $y^1, \ldots, y^n$ has the prescribed form. □

Example 8.3.5. We continue with Example 8.2.10, where

$$X = y\partial_x + z\partial_y.$$

The two invariants are $u = x - \frac{y^2}{2z}, v = z$, and these are functionally independent away from $(y,z) = (0,0)$. Take for the third function $w = y$, so that u, v, w are a functionally independent set. We now find the vector field in X these coordinates. All we need is

$$X(w) = z.$$

Therefore in the (u, v, w) coordinates,

$$X = v\partial_w,$$

as expected from Lemma 8.3.4. Note if we take as our third coordinate $w = \frac{y}{z}$ we get

$$X(w) = 1$$

and

$$X = \partial_w.$$

See also example 4.4.6.

8.4 Invariants II

We now look at a method for finding the invariants of a flow $\Psi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$.

We begin with a global description of the solution. Let $\iota : \mathbb{R}^{n-1} \to \mathbb{R}^n$ be a smooth function and define the function $\Phi : \mathbb{R} \times \mathbb{R}^{n-1} \to \mathbb{R}^n$, by

$$\Phi(t, \mathbf{u}) = \Psi(t, \iota(\mathbf{u})) \tag{8.17}$$

and note Φ can also be considered as a function $\Phi : \mathbb{R}^n \to \mathbb{R}^n$.

Definition 8.4.1. The function $\iota : \mathbb{R}^{n-1} \to \mathbb{R}^n$ is a *cross-section* of the flow Ψ if the function Φ is a diffeomorphism.

If $\iota : \mathbb{R}^{n-1} \to \mathbb{R}^n$ is a cross-section of the flow Ψ, we think of the image $K = \iota(\mathbb{R}^{n-1}) \subset \mathbb{R}^n$ as a hyper-surface through which the flow passes.Let $u \in \mathbb{R}^{n-1}$ and let $p = \iota(u)$ then the curve $\psi_p(t)$ passes through p when $t = 0$.
XXXXXXXXXXXXXXX Picture here XXXXXXXXXXXXXXX

Lemma 8.4.2. *Let ι be a cross-section to the flow Ψ. For any point $p = \iota(u) \in K$ the curve $\psi_p(t)$ intersects K only at the point p.*

Proof. Suppose that $p = \iota(u_0)$ and that $\psi_p(t_1) = \iota(u_1)$. This implies

$$\begin{aligned} \psi_p(t_1) &= \Psi(t, p) = \Psi(t_1, \iota(u_0)) \\ &= \Psi(t_1, \iota(\mathbf{u}_0)) \qquad \text{by 8.17.} \end{aligned} \tag{8.18}$$

While by equation 8.17 $\iota(u_1) = \Phi(0, u_1)$. Therefore $\psi_p(t_1) = \iota(u_1)$ implies by equation 8.18

$$\Phi(t_1, u_0) = \Phi(0, u_1).$$

Now Φ is injective so $t_1 = 0, u_1 = u_0$, which proves the lemma. $\square$

A fundamental property of the function Φ is the following.

Lemma 8.4.3. *Suppose $\iota : \mathbb{R}^{n-1} \to \mathbb{R}^n$ is a cross-section, and that $p = \Phi(t_0, u)$. Then*

$$\Psi(t_1, p) = \Phi(t_1 + t_0, u_0).$$

Proof. To prove this we just write out the definitions using equation 8.17,

$$\begin{aligned}\Psi(t_1, p) &= \Psi(t_1, \Phi(t_0, u)) \\ &= \Psi(t_1, \Psi(t_0, \iota(u))) \\ &= \Psi(t_1 + t_0, \iota(u)) \\ &= \Phi(t_1 + t_0, u).\end{aligned}$$

□

This lemma will be critical for what we have to say next. Consider $\Phi^{-1} : \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^{n-1}$ so that we can write the components

$$\begin{aligned}\rho(\mathbf{x}) &= (\Phi^{-1}(\mathbf{x}))^1 \\ f^i(\mathbf{x}) &= (\Phi^{-1}(\mathbf{x}))^{i+1}, \ 1 \le i \le n-1.\end{aligned} \tag{8.19}$$

The functions $\rho, f^i \in C^\infty(\mathbb{R}^n), 1 \le i \le n$.

Using Lemma 8.4.3, we can then deduce the following properties of the functions ρ and f^i.

Lemma 8.4.4. *The function ρ satisfies*

$$\rho(\Psi(t, \mathbf{x})) = \rho(\mathbf{x}) + t \tag{8.20}$$

and the functions f^i satisfy

$$f^i(\Psi(t, \mathbf{x})) = f^i(\mathbf{x})$$

and so f^i are invariants.

Proof. Suppose $p = \Phi(t_0, u)$, so that $\rho(p) = t_0$ and $f^i(p) = u^i$. By Lemma 8.4.3, $\Psi(t, p) = \Phi(t + t_0, u)$. Therefore

$$\Phi^{-1}(\Psi(t, p)) = \Phi^{-1}(\Phi(t + t_0, p)) = (t + t_0, p).$$

By taking the individual components in this equation, and using the fact that p was arbitrary proves the lemma. □

Corollary 8.4.5. *The function $f^i, 1 \le i \le n-1$ in Lemma 8.4.4, are a complete set of functionally independent invariants of the flow Ψ.*

The geometry of the function ρ is particularly interesting. Let $p \in I\!R^n$, and let $(t_0, u) \in I\!R^n$ such that $p = \Phi(t_0, u)$. Then $\rho(p) = t_0$, and

$$\Psi(\rho(p), \iota(\mathbf{u}_0)) = \Psi(t_0, \iota(u)) = \Phi(t_0, u) = p.$$

Therefore the function $\rho(p)$ gives the t-value of the flow required to get from the point u on the cross-section $K = \iota(I\!R^{n-1})$ to the point p. It is also worth noting then that $-\rho(p)$ is the t-value for p that flows p back to $u \in K$. This interpretation helps in determining what to chose for K, and also what Φ^{-1} is. In other words, we need to choose K so that we can flow back to K from any point in $I\!R^n$, this determines ρ which in turn determines Φ^{-1} and hence the invariants.

Example 8.4.6. Consider the flow $\Psi : I\!R \times I\!R^3 \to I\!R^3$

$$\Psi(t, (x, y, z)) = (x \cos t - y \sin t, -x \sin t + y \cos t, z + kt),$$

where $k \neq 0$, (see from problem 1(c) on assignment 8). Let $\iota : I\!R^2 \to I\!R^3$ be

$$\iota(u^1, u^2) = (u^1, u^2, 0).$$

By the discussion above, we need to find a function $\rho : I\!R^3 \to I\!R$ that has the property $\Psi(-\rho(\mathbf{x}), \mathbf{x}) \in K$. Since K consists of point whose z coordinate is 0, we need to solve for $t = -\rho(\mathbf{x})$ such that $z + kt = 0$. Therefore $t = -\frac{z}{k}$. With this choice of $\rho(\mathbf{x}) = \frac{z}{k}$, and we get

$$\Psi(-\rho(\mathbf{x}), \mathbf{x}) = (x \cos \frac{z}{k} + y \sin \frac{z}{k}, -x \sin \frac{z}{k} + y \cos \frac{z}{k}, 0)$$

and so

$$u^1 = x \cos \frac{z}{k} + y \sin \frac{z}{k}, \ u^2 = -x \sin \frac{z}{k} + y \cos \frac{z}{k},$$

are the two invariants.

The argument above for constructing invariants easily generalizes to the case of a flow on an open set $U \subset I\!R^n$, $\Psi : I\!R \times U \to U$ where $\iota : V \to U$ and $V \subset I\!R^{n-1}$ is an open set.

Example 8.4.7. Let $\Psi : I\!R \times I\!R^3$ be the flow from Example 8.1.5,

$$\Psi(t, x, y, z) = (x + ty + \frac{1}{2}t^2 z, y + tz, z).$$

Let $U = \{(x,y,z) \mid z \neq 0\}$, $V = \{(u,v) \mid v \neq 0\}$, and let $\iota : V \to U$ be the function $\iota(u,v) = (u,0,v)$. We fund the function $\rho : U \to I\!R$ by finding the t (flow) amount that maps us from a point $(x,y,z) \in U$ to K whose y component is 0. Thus we solve $y + tz = 0$, which gives $t = -\frac{y}{z}$, and therefore

$$\Psi(-\frac{y}{z}, (x,y,z)) = (x - \frac{y^2}{2z}, 0, z) \tag{8.21}$$

The values of our invariants (u,v) are read off equation 8.21 to be,

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

Example 8.4.8. The rotation in the plane flow is

$$\Psi(t,(x,y)) = (x\cos t - y\sin t, x\sin t + y\cos t).$$

On the set $U = I\!R^2 - (0,0)$ let $\iota : I\!R^+ \to U$ be $\iota(r) = (r,0)$. The function $\rho : I\!R^2 - (0,0) \to I\!R$ is found by solving $x\sin t + y\cos t = 0$ for t. This gives

$$t = -\arctan\frac{y}{x},$$

We then compute $\Psi(\rho(x,y),(x,y))$ to be

$$\Psi(\rho(x,y),(x,y)) = (x\cos(\arctan\frac{y}{x}) + y\sin(\arctan\frac{y}{x}), 0) = (\sqrt{x^2+y^2}, 0).$$

The invariant is then $\sqrt{x^2+y^2}$.

Finally we come to the Straightening Lemma, which is an improvement on Lemma 8.3.4.

Lemma 8.4.9. *Let $\iota : I\!R^{n-1} \to I\!R^n$ be a cross-section, and let $s = \rho(\mathbf{x})$, $y^i = f^i(\mathbf{x}), 1 \leq i \leq n-1$ be the coordinates as defined in equation 8.19. In these coordinates, the infinitesimal generator of the flow Ψ has the form*

$$X = \frac{\partial}{\partial s}.$$

Proof. As in Lemma 8.3.4 the fact that $y^i = f^i, 1 \leq i \leq n-1$ are invariants and the change of variables formula for vector fields 4.27 implies

$$X = a(\mathbf{y}, s)\partial_s, \quad a \in C^\infty(I\!R^n),$$

in the $(y^1, \ldots, y^{n-1}, s)$ coordinates. Now let $p \in I\!R^n$ and let $\psi_p(t)$ be the flow through p. We then apply X to the function $\rho(\mathbf{x})$ at $\mathbf{x}_0$ by computing

$$X_p(\rho) = \frac{d}{dt}\rho(\psi_p(t))|_{t=0}$$

which simplifies on account of (8.20) to give

$$X_p(y^n) = X(\rho)_p = 1.$$

Since the point p was arbitrary this proves the lemma. □

XXXXXXXXXXXX FIX below this XXXXXXXXXXXXXXx

The local version is the following. Let $p \in I\!R^n$, $V \subset I\!R^{n-1}$ and open set and let $\iota : V \to I\!R^n$ be an immersion such that $\mathbf{x}_0 = \iota(\mathbf{u}_0)$. Suppose that the function $\Phi : I\!R \times V \to I\!R^n$ given by

$$\Phi(t, \mathbf{u}) = \Psi(t, \iota(\mathbf{u}))$$

satisfies $(\Phi_*)_{0,\mathbf{u}_0}$ is invertible.

Lemma 8.4.10. *There exists a neighbourhood U of $\mathbf{x}_0$ such that Φ_U is a diffeomorphism. The functions $(\Phi_U^{-1})^i, 2 \leq i \leq n$ are a set of functionally independent invariants of the infinitesimal generator X of Ψ.*

8.5 Exercises

1. Check that each of the following mappings define a flow.

   (a) On $I\!R^2$, $\Psi(t,(x,y)) = (x\cosh t + y\sinh t, x\sinh t + y\cosh t)$.

   (b) On $I\!R^2$, $\Psi(t,(x,y)) = ((yt+x)e^t, y = ye^t))$.

   (c) On $I\!R^3$, $\Psi(t,(x,y,z)) = (x\cos t - y\sin t, -x\sin t + y\cos t, z + kt)$.

2. Plot the curves $\Psi_p(t)$ for a few initial points p in the flows in question 1.

3. Justify the claim in Example 8.2.4 (the complement of an invariant set, is also an invariant set).

4. Find the flows $\Psi : R \times I\!R^n \to I\!R^n$ (for an appropriate n) for each of the following vector fields using the flow equation 8.10. Determine which of the flows are defined for all values of the flow parameter (global flows). For the ones which are global flows, check the flow conditions $\Psi(0,p) = p$, $\Psi_t \circ \Psi_s = \Psi_{(s+t)}$, and equation 8.4.

   (a) $X = \partial_x - \partial_y + 2\partial_z$.

   (b) $X = x\partial_x - y\partial_y + 2z\partial_z$.

   (c) $X = x\partial_y + y\partial_z$.

   (d) $X = x\partial_y + y\partial_x$.

   (e) $X = e^x\partial_x$ (on R).

   (f) $X = (2x - y)\partial_x + (3x - 2y)\partial_y$

5. For each of the global flows in problem 4, find an open set together with a complete set of invariants on this set. Check that your invariants are indeed constant along the flow (constants of motion).

6. For each global flow in problem 4, straighten out the vectors fields on an open set about a point of your choosing.

7. Find a complete set of invariants for each of the following vector fields (Remember that you do not necessarily need to find the flow). Identify an open set where the invariants are functionally independent.

(a) $X = y^3\partial_x - x^3\partial_y$.

(b) $X = yz\partial_x + xz\partial_y + xy\partial_z$.

(c) $X = (x^2 + y^2)\partial_x + 2xy\partial_y$

(d) $X = (y - z)\partial_x + (z - x)\partial_y + (x - y)\partial_z$,

(e) $X = x^2\partial_x + y^2\partial_y + (x + y)z\partial_z$

8. Use the invariants as coordinates to obtain a partial straightening of the vector fields in 7(a) and 7(b).

9. Show that if f and g are invariants for a flow Ψ, then $h(x) = F(f(x), g(x))$ is an invariant, where F is any smooth function of two variables. Repeat this problem if f, g are invariants of a vector field X.

10. Let $X = \sum_{i=1}^n \sum_{j=1}^n A^i_j x^j \partial_i$ be a linear vector field on R^n. Find matrix equations which are necessary and suffice for a linear function $f(x) = \sum_{i=1}^n b_i x^i$ and a quadratic function $g(x) = \sum_{i=1}^n \sum_{j=1}^n b_{ij} x^i x^j$ (here $b_{ij} = b_{ji}$) to be invariants of X.

11. Prove corollary 8.4.5.

Chapter 9

The Lie Bracket and Killing Vectors

9.1 Lie Bracket

Let X and Y be vector fields on $\mathbb{R}^n$. We construct a third vector from X and Y which will be written as $[X, Y]$ and is called the commutator, or bracket of X and Y. The vector field $[X, Y]$ will be defined in two ways, first by its value on smooth functions $f \in C^\infty(\mathbb{R})$, and second a more geometric definition the Lie derivative.

Lemma 9.1.1. *There exists a vector field $[X, Y]$ uniquely defined by*

$$[X, Y](f) = X(Y(f)) - Y(X(f)) \tag{9.1}$$

for all $f \in C^\infty(\mathbb{R}^n)$.

Proof. In order to check that $[X, Y]$ defines a vector-field we need to check that it satisfies the derivation properties. The only difficult one is

$$[X, Y](f\, g) = f[X, Y](g) + g[X, Y](f), \quad f, g \in C^\infty(\mathbb{R}^n). \tag{9.2}$$

Writing out the left side using the definition 9.1 we get

$$\begin{aligned} X(Y(fg)) - Y(X(fg)) &= X(fY(g) + gY(f)) - Y(fX(g)gX(f)) \\ &= fX(Y(g)) + X(f)Y(g) + gX(Y(f)) + X(g)Y(f) \\ &\quad - fY(X(g)) - Y(f)X(g) - gY(X(f)) - Y(g)X(f) \\ &= fX(Y(g)) - fY(X(g)) + gX(Y(f)) - g(X(f)). \end{aligned}$$

This is the right hand side of equation 9.2. □

Let $X = \sum_{i=1}^n X^i(\mathbf{x})\partial_{x^i}$ and $Y = \sum_{i=1}^n Y^i(\mathbf{x})\partial_{x^i}$ be vector-fields on $I\!R^n$. We now work out the components of $[X,Y]$ in the coordinate basis $\{\partial_{x^i}\}_{1\leq i\leq n}$ by applying $[X,Y]$ to the functions x^i to get

$$\begin{aligned}[X,Y](x^i) &= X(Y(x^i)) - Y(X(x^i))\\ &= X(Y^i(\mathbf{x})) - Y(X^i(\mathbf{x})).\end{aligned}$$

Therefore $[X,Y]$ can be written in one of the following ways,

$$\begin{aligned}[X,Y] &= \sum_{i=1}^n X(Y^i(\mathbf{x}))\partial_{x^i} - \sum_{i=1}^n Y(X^i(\mathbf{x}))\partial_{x^i}\\ &= \sum_{i=1}^n \big(X(Y^i(\mathbf{x})) - Y(X^i(\mathbf{x}))\big)\,\partial_{x^i}\\ &= \sum_{i=1}^n \left(\sum_{j=1}^n X^j(\mathbf{x})\frac{\partial Y^i(\mathbf{x})}{\partial x^j} - Y^j(\mathbf{x})\frac{\partial X^i(\mathbf{x})}{\partial x^j}\right)\partial_{x^i}.\end{aligned} \tag{9.3}$$

The first formula in equation 9.3 is often convenient to use.

Example 9.1.2. Let

$$X = xe^y\partial_x + xz\partial_y + \partial_z, Y = x\partial_x + (y+z)\partial_y + e^z\partial_z.$$

then by the first in equation 9.3

$$\begin{aligned}[X,Y] &= X(x)\partial_x + X(y+z)\partial_y + X(e^z)\partial_z - Y(xe^y)\partial_x - Y(xz)\partial_y - Y(1)\partial_z\\ &= xe^y\partial_x + (xz+1)\partial_y + e^z\partial_z - (xe^y + x(y+z)e^y)\partial_x - (xz + xe^z)\partial_y\\ &= -x(y+z)e^y\partial_x(1 - xe^z)\partial_y + e^z\partial_z\,.\end{aligned} \tag{9.4}$$

Some algebraic properties of the operations $[\ ,\ ]$ are the following.

Lemma 9.1.3. *The operation* $[\ ,\ ]$ *satisfies,*

1. $[X,Y] = -[Y,X]$,
2. $[aX+bY,Z] = a[X,Z] + b[Y,Z]$,
3. $[X,aY+bZ] = a[X,Y] + b[X,Z]$,
4. $[X,[Y,Z]] + [Z,[X,Y]] + [Y,[Z,X]] = 0$,

for all vector fields $X, Y, Z,$ *and* $a, b \in \mathbb{R}$.

The proof of these facts is left for the exercises. Property 4 is called the Jacobi-identity.

Definition 9.1.4. A vector space V with a bilinear function $[\ ,\ ] : V \times V \to V$ satisfying

1. $[X, Y] = -[Y, X]$, and
2. $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$,

is a *Lie algebra*.

Note that a Lie algebra is a special type of algebra (see definition 1.6.1). Condition 1 in the definition of a Lie Algeba says that $[\ ,\ ]$ is alternating, or anti-symmetric. Condition 2 is again known as the *Jacobi identity*. Lemma 9.1.3 shows that the vector space of vector fields on $\mathbb{R}^n$ forms a Lie algebra. This vector space (over $\mathbb{R}$) is **not** finite dimensional.

Example 9.1.5. Let

$$X = \partial_x + \partial_y,\ Y = x\partial_x + y\partial_y,\ Z = x^2\partial_x + y^2\partial_y$$

be vector-fields on $\mathbb{R}^2$. We check that $\Gamma = \text{span}\{X, Y, Z\}$ forms a Lie algebra with $[\ ,\ ]$ the operation on vector-fields.

Recall from section 1.6 that a sub-space $W \subset V$ of an algebra V that forms an algebra with the operation from V is called a subalgebra. Lemmas 1.6.6, 1.6.8 provide an easy way to check is a subspace is a subalgebra. We use Lemma 1.6.8 to test that Γ is a subalgebra of all vector fields on $\mathbb{R}^n$ using the basis X, Y, Z. We only need to compute

$$[X, Y] = X, \quad [X, Z] = 2Y, \quad [Y, Z] = Y. \tag{9.5}$$

because of the anti-symmetry of $[\ ,\ ]$. Therefore equation 9.5 implies on account of Lemma 1.6.8 that Γ is a sub-algebra of all the vector-fields on $\mathbb{R}^2$ (and is then by itself a Lie algebra).

Note if we let $W = x^3\partial_x + y^3\partial_y$, then $\Gamma = \text{span}\{X, Y, Z, W\}$ does not form a Lie algebra, because $[Z, W] \notin \Gamma$ and so Lemma 1.6.8 is not satisfied.

Remark 9.1.6. On the classification of Lie algebras of vector-fields in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$. Still to be added XXXX.

We now define the Lie derivative of a vector field Y with respect to X, $\mathcal{L}_X Y$. This agrees with the function $[X, Y]$ but has a geometric definition analogous to the Lie derivative of a function given in equation 8.13. Suppose that Ψ is flow on $\mathbb{R}^n$ with infinitesimal generator X, and that Y is a vector field on $\mathbb{R}^n$. Let $p \in \mathbb{R}^n$. The *Lie derivative* of the vector field Y in the direction X at the point p is the tangent vector

$$(\mathcal{L}_X Y)_p = \lim_{t \to 0} \frac{1}{t} \left((\Psi_{-t})_* Y_{\Psi(t,p)} - Y_p \right). \tag{9.6}$$

This is the analogue to the directional derivative of a function in the direction X, and it measures the rate of change of Y in the direction X. Notice that the point p was arbitrary here, and so $\mathcal{L}_X Y$ defines a tangent vector at every point.

Lemma 9.1.7. *Given vector fields X and Y on $\mathbb{R}^n$,*

$$\mathcal{L}_X Y = [X, Y]. \tag{9.7}$$

Proof. In order to check this formula we fix a point $p \in \mathbb{R}^n$ and show equation 9.7 holds at p. Equation (9.6) written in components in the coordinate basis $\{\partial_{x^i}\}_{1 \le i \le n}$,

$$(\mathcal{L}_X Y)^i_p = \lim_{t \to 0} \frac{1}{t} \left(\sum_{j=1}^n \frac{\partial \Psi^i(-t,p)}{\partial x^j} Y^j(\Psi(t,p)) - Y^i(\Psi(t,p)) + Y^i(\Psi(t,p)) - Y^i(p) \right). \tag{9.8}$$

Using the fact that $\Psi(0, \mathbf{x}) = \mathbf{x}$, we have

$$\partial_{x^i} \Psi^j(0, \mathbf{x}) = \delta^i_j \tag{9.9}$$

and equation 9.8 can be written

$$\begin{aligned}
(\mathcal{L}_X Y)^i_p &= \lim_{t \to 0} \frac{1}{t} \left(\sum_{j=1}^n Y^j(\Psi(t,p)) (\frac{\partial \Psi^i(-t,p)}{\partial x^j} - \delta^i_j + (Y^i(\Psi(t,p)) - Y^i_p) \right) \\
&= \sum_{j=1}^n Y^j(\Psi(0,p)) \left(\frac{\partial}{\partial t} \frac{\partial \Psi^i(-t,p)}{\partial x^j} \right)\Bigg|_{t=0} + \left(\frac{\partial}{\partial t} Y^i(\Psi(t,p)) \right)\Bigg|_{t=0} \\
&= -\sum_{j=1}^n Y^j(p) \left(\frac{\partial X^i(\mathbf{x})}{\partial x^j} \right)\Bigg|_{\mathbf{x}=p} + \sum_{j=1}^n X^j(p) \left(\frac{\partial Y^i(\mathbf{x})}{\partial x^j} \right)\Bigg|_{\mathbf{x}=p}
\end{aligned}$$

where we have switched the order of differentiation in the term with the second derivatives. Comparing this with the components of $[X, Y]$ in the coordinate basis in equation 9.3 proves $[X, Y]_p^i = (\mathcal{L}_X Y)_p^i$. Since p is arbitrary the theorem is proved. □

Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism. A vector field Y on $\mathbb{R}^n$ is Φ-invariant if

$$\Phi_* Y = Y.$$

A vector-field Y is said to be invariant under the flow Ψ if for all diffeomorphisms $\Psi_t,\ t \in \mathbb{R}$,

$$(\Psi_t)_* Y = Y \ . \tag{9.10}$$

Theorem 9.1.8. *A vector field Y on $\mathbb{R}^n$ is invariant under the flow Ψ if and only if*

$$[X, Y] = \mathcal{L}_X Y = 0$$

where X is the infinitesimal generator of Ψ.

Proof. If Y is invariant with respect to the flow Ψ, then substituting equation 9.10 into equation (9.6) which defines the Lie derivative we have

$$\mathcal{L}_X Y = 0.$$

To prove sufficiency XXX (not done). □

Let X be the infinitesimal generator of the flow Ψ and Y the infinitesimal generator of the flow Φ. The skew-symmetric property of $[\ ,\ ]$ (property 1 in Lemma 9.1.3) leads to an obvious corollary.

Corollary 9.1.9. *The vector field Y is invariant under a flow Ψ with if and only X of is invariant under the flow Φ.*

Theorem 9.1.8 and corollary 9.1.9 extend to the following commuting property of the flows,

Theorem 9.1.10. *The flows Ψ and Φ commute*

$$\Phi(s, \Psi(t, \mathbf{x})) = \Psi(t, \Phi(s, \mathbf{x})) \quad \textit{for all } t, s \in \mathbb{R},\ \mathbf{x} \in \mathbb{R}^n \tag{9.11}$$

if and only if

$$[X, Y] = 0.$$

Proof. Need to put the proof in XXXXX □

Another way to write equation 9.11 is that for all $t, s \in I\!R$ the diffeomorphisms commute

$$\Psi_t \circ \Phi_s = \Phi_s \circ \Psi_t.$$

Example 9.1.11. Let $X = x\partial_x + y\partial_y$, and $Y = -y\partial_x + x\partial_y$. We'll show that $[X, Y] = 0$, the flows commute, and Y is invariant under the flow of X.

First using equation 9.3 we have

$$\begin{aligned} &= X(-y)\partial_x + X(x)\partial_y - Y(x)\partial_x - Y(y)\partial_y \\ &= -y\partial_x + x\partial_y + y\partial_x - x\partial_y = 0. \end{aligned}$$

The flow Ψ for X and Φ for Y are easily found to be

$$\begin{aligned} \Psi(t,(x,y)) &= (e^t x, e^t y), \\ \Phi(t(x,y)) &= (x\cos t - y\sin t, x\sin t + y\cos t). \end{aligned} \tag{9.12}$$

We easily compute the two compositions in 9.11 using the flows in 9.12

$$\begin{aligned} \Phi(s, \Psi(t,(x,y)) &= (xe^t\cos s - ye^t\sin s, xe^t\sin s + ye^y\cos s) \\ &= \Psi(t, \Psi(s,(x,y)). \end{aligned}$$

Finally let's check that Y is invariant under the flow of Ψ. We find the components from push-forward formula in equation **??** to be

$$((\Psi_t)_* Y)^i = \sum_{j=1}^n \frac{\partial \Psi^i(t, \mathbf{x})}{\partial x^j} Y^j(\mathbf{x}) = \begin{bmatrix} -ye^t \\ xe^t \end{bmatrix} = Y^i(\Psi(t,x),$$

and so Y is Ψ invariant.

We now look at a generalization of the Straightening Lemma 8.4.9, which simultaneously straightens two commuting vector-fields.

Lemma 9.1.12. *Suppose that the vector fields X and Y are linearly independent at the point p and that $[X, Y] = 0$. Then there exists an open set $U \subset I\!R^n$ with $p \in U$, and coordinates $(y^1, \ldots, y^{n-2}, r, s)$ on U such that*

$$X = \partial_s, \quad Y = \partial_r.$$

Proof. We show two proofs here. The first is based on the Straightening Lemma 8.4.9. Let $(u^i, v)_{1\leq i\leq n-1}$ be the straightening coordinates for X (on an open set V with $p \in V$), so that

$$X = \partial_v. \tag{9.13}$$

The condition $[X, Y] = 0$ implies by equation 9.3,

$$Y = Y^0(\mathbf{u})\partial_v + \sum_{a=1}^{n-1} Y^a(\mathbf{u})\partial_{u^a}, \tag{9.14}$$

and so the coefficients of Y don't depend on v. Let

$$Z = \sum_{a=1}^{n-1} Y^a(\mathbf{u})\partial_{u^i}$$

which does not vanish at p, by the linearly independence condition. Now use the Straightening Lemma 8.4.9 on Z, to find an open set U with $p \in U$, and coordinates on U, $(y^i = f^i(\mathbf{u}), r = f(\mathbf{u}), v = v)_{1\leq i\leq n-2}$, such that

$$Z = \partial_r.$$

With the coordinates (y^i, r, v) the form of the vector field X in 9.13 does not change, and from equation 9.14

$$Y = Y^0(r, y^i)\partial_v + \partial_r.$$

Making the final change of variables

$$s = v - \int Y^0(r, v^i)dr$$

gives the coordinates in the theorem. □

Proof. The second proof is similar to the sequence of steps used in proof 2 of the Straightening Lemma 8.4.9. Suppose Ψ is the flow for X and Φ the flow for Y, and define $H : \mathbb{R}^2 \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$H(s, t, \mathbf{x}) = \Phi(s, \Psi(t, \mathbf{x})). \tag{9.15}$$

Let $\iota : I\!R^{n-2} \to I\!R^n$ be an immersion, such that $F : I\!R \times I\!R \times I\!R^{n-2} \to I\!R$ given by

$$F(s,t,v) = H(s,t,\iota(v)) = \Phi(s, \Psi(t, \iota(v)))$$

is a diffeomorphism. The function ι is a cross section to the joint flows Ψ, Φ. Let $\rho : I\!R^n \to I\!R^2$ be the function $((F^{-1})^1, (F^{-1})^2)$. Exactly as in the proof of Lemma 8.4.4, the functions

$$f^i = (F^{-1})^{i+2}(\mathbf{x}), \quad i = 1, \dots, n-2$$

are invariants of **both** flows, which together with $-\rho(\mathbf{x})$ give the coordinates in the straightening lemma. □

Example 9.1.13. Let $X = -y\partial_x + x\partial_y$, and $Y = x\partial_x + y\partial_y + z\partial_z$, be vector-fields on $I\!R^3$ and $[X,Y] = 0$.

The functions $r = \sqrt{x^2+y^2}, \theta = \arctan \frac{y}{x}, z = z$ provide coordinates on $V = I\!R^2 - \{(x,0) \,|\, x \geq 0\}$ which straighten X. In these coordinates

$$X = \partial_\theta, \quad Y = r\partial_r + z\partial_z.$$

To straighten $r\partial_r + z\partial_z$ we let $s = \log z$ and $v = \frac{z}{r}$ on $U = V - \{z \leq 0\}$, giving

$$Y = \partial_s.$$

Using the construction in the second proof of Lemma 9.1.12, we first need the flows

$$\Psi(t,(x,y,z)) = (x\cos t - y \sin t, x \sin t + y \cos t, z)$$
$$\Phi(t,(x,y,z))(e^t x, e^t y, e^t z).$$

The function H in equation 9.15 is

$$H(s,t,(x,y,z)) = (xe^s \cos t - ye^s \sin t, xe^s \sin t + ye^s \cos t, e^s z).$$

Choosing the cross-section $\iota(v) = (0, v, 1)$, then we compute $-\rho(\mathbf{x})$, which as described in the paragraph above example 8.4.6 as the function taking us to the cross section $(0, v, 1)$ with H. Therefore we solve

$$(xe^s \cos t - ye^s \sin t, xe^s \sin t + ye^s \cos t, e^s z) = (0, v, 1)$$

for s and t. This gives,

$$t = \arctan \frac{y}{x}, \quad s = \log z$$

and resulting in $v = \frac{r}{z}$, with $r = \sqrt{x^2+y^2}$. The simultaneous straightening coordinates are (v, s, t).

9.2 Killing vectors

We have seen in Lemma 8.2.12 of section 8.2 that given a flow Ψ with infinitesimal generator X, that a function $f \in C^\infty(\mathbb{R}^n)$ is an invariant of the flow if and only if $\mathcal{L}_X(f) = X(f) = 0$. In the previous section in Lemma 9.1.8 we had a similar situation that a vector-field Y is an invariant of the flow if and only if $\mathcal{L}_X Y = [X, Y] = 0$. We now find in a similar manner the condition that a metric tensor field be invariant under a flow Ψ in terms of the infinitesimal generator X of the flow.

Let $\boldsymbol{\gamma}$ be a metric tensor on $\mathbb{R}^n$. Recall from section 6.38 that a diffeomorphism $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry if $\Phi^*\boldsymbol{\gamma} = \boldsymbol{\gamma}$. The metric tensor is invariant under a flow Ψ if each diffeomorphism Ψ_t is an isometry (see equation 6.38),

$$\Psi_t^*\boldsymbol{\gamma} = \boldsymbol{\gamma}, \quad \textit{for all } t \in \mathbb{R} \tag{9.16}$$

A flow Ψ which satisfies this equation is also called a one-parameter group of isometries for the metric $\boldsymbol{\gamma}$.

We have the following theorem.

Theorem 9.2.1. *A metric tensor $\boldsymbol{\gamma} = \sum_{i,j=1}^n g_{ij}(\mathbf{x})dx^i dx^j$ on $\mathbb{R}^n$ is invariant under a flow Ψ with infinitesimal generator $X = \sum_{i=1}^n X^i(\mathbf{x})\partial_{x^i}$ if and only if*

$$\sum_{m=1}^n \frac{\partial g_{kl}(\mathbf{x})}{\partial x^m} X^m + g_{km}(\mathbf{x})\frac{\partial X^m(\mathbf{x})}{\partial x^l} + g_{ml}(\mathbf{x})\frac{\partial X^m(\mathbf{x})}{\partial x^k} = 0. \tag{9.17}$$

Proof. We begin by showing this is necessary. Suppose that equation 9.16 holds, then taking the derivative with respect to t at $t = 0$ gives

$$\left(\frac{\partial}{\partial t}\Psi_t^*\boldsymbol{\gamma}\right)\Big|_{t=0} = 0 \tag{9.18}$$

Expanding out the left hand side of this in components using equation 6.35 to compute $\Psi_t^*\boldsymbol{\gamma}_{\Psi(t,\mathbf{x})}$ we have

$$\left(\Psi_t^*\boldsymbol{\gamma}_{\Psi(t,\mathbf{x})}\right)_{kl} = \left(\sum_{i,j=1}^n g_{ij}(\Psi(t,\mathbf{x}))\frac{\partial \Psi^i(t,\mathbf{x})}{\partial x^k}\frac{\partial \Psi^j(t,\mathbf{x})}{\partial x^l}\right).$$

Now taking the derivative of this (and surpressing arguments) gives,
(9.19)

$$\partial_t \left(\Psi_t^* \boldsymbol{\gamma}_{\Psi(t,\mathbf{x})}\right)_{kl} = \partial_t \left(\sum_{i,j=1}^n g_{ij}(\Psi(t,\mathbf{x})) \frac{\partial \Psi^i(t,\mathbf{x})}{\partial x^k} \frac{\partial \Psi^j(t,\mathbf{x})}{\partial x^l} \right)$$
$$= \sum_{i,j,m=1}^n \frac{\partial g_{ij}}{\partial x^m} \frac{\partial \Psi^m}{\partial t} \frac{\partial \Psi^i}{\partial x^k} \frac{\partial \Psi^j}{\partial x^l} + \sum_{i,j=1}^n g_{ij} \left(\frac{\partial^2 \Psi^i}{\partial t \partial x^k} \frac{\partial \Psi^j}{\partial x^l} + \frac{\partial \Psi^i}{\partial x^k} \frac{\partial^2 \Psi^j}{\partial t \partial x^l} \right)$$

Equation 9.19 now needs to be evaluated at $t = 0$, and the first term is

$$(9.20) \quad \sum_{i,j,m=1}^n \frac{\partial g_{ij}}{\partial x^m} \frac{\partial \Psi^m}{\partial t} \frac{\partial \Psi^i}{\partial x^k} \frac{\partial \Psi^j}{\partial x^l} = \sum_{i,j,m=1}^n \frac{\partial g_{ij}}{\partial x^m} X^m \delta^i_k \delta^j_l = \frac{\partial g_{kl}(\mathbf{x})}{\partial x^m} X^m(\mathbf{x})$$

where we have substitute from equations 8.3, and 9.9. Next we evaluate at $t = 0$ in the second derivative term to get,

$$(9.21) \quad \left(\frac{\partial^2 \Psi^i(t,\mathbf{x})}{\partial t \partial x^k} \right)\Bigg|_{t=0} = \frac{\partial}{\partial x^k} \left(\frac{\partial \Psi^i(t,\mathbf{x})}{\partial t} \Bigg|_{t=0} \right) = \frac{\partial}{\partial x^k} X^i(\mathbf{x})$$

where again we have used equation 8.3 for the infinitesimal generator. Finally substituting equations 9.20 and 9.21 into equation 9.19 gives 9.17. □

Equation 9.17 is known as Killing's equation (named after W. Killing, 1847-1923). Note that since $g_{ij} = g_{ji}$ in the Killing equations 9.17, these equations constitute $\frac{n^2(n+1)}{2}$ equations.

Example 9.2.2. Let $X = (x^2 - y^2)\partial_x + 2xy\partial_y$. This is a Killing vector for the metric $ds^2 = y^{-2}(dx^2 + dy^2)$. Fortunately the metric tensor $\boldsymbol{\gamma}$ is diagonal in the coordinate basis $\{dx, dy\}$,and so equations 9.17 only contain four separate equations, given by the index values $k = l = 1, m = 1, 2$, and $k = l = 2, m = 1, 2$. Suppose $X = X^1(x,y)\partial_x + X^2(x,y)\partial_y$, and let's write these equations for X, and then check that we have a solution. For $k = l = 1$

we use $g_{11} = y^{-2}, g_{1m} = 0$, $x^1 = x, x^2 = y$ to compute

$$
\begin{aligned}
&\sum_{m=1}^{2} \frac{\partial g_{11}}{\partial x^m} X^m = -\frac{2}{y^3} X^2 \\
&\sum_{m=1}^{2} g_{1m} \frac{\partial X^m}{\partial x} = \frac{1}{y^2} \frac{\partial X^1}{\partial x} \\
&\sum_{m=1}^{2} g_{m1} \frac{\partial X^m}{\partial x} = \frac{1}{y^2} \frac{\partial X^1}{\partial x}.
\end{aligned} \tag{9.22}
$$

Therefore for $k = l = 1$ equation 9.17 is,

$$
-\frac{2}{y^3} X^2 + \frac{2}{y^2} \frac{\partial X^1}{\partial x} = 0 \tag{9.23}
$$

Similarly for $k = l = 2$ we have

$$
-\frac{2}{y^3} X^2 + \frac{2}{y^2} \frac{\partial X^2}{\partial y} = 0, \tag{9.24}
$$

while for $k = 1, l = 2$ (which as noted above is the same as $k = 2, l = 1$ by symmetry),

$$
\frac{\partial X^2}{\partial x} + \frac{\partial X^1}{\partial y} = 0. \tag{9.25}
$$

To verify X is a Killing vector we only need to check that $X^1 = (x^2 - y^2$ and $X^2 = 2xy$ satisfy the three equation 9.23, 9.25, 9.24.

This example highlights one aspect of the Killing equations 9.17. Given a metric tensor $\boldsymbol{\gamma}$ on $\mathbb{R}^n$, equation 9.17 can be considered a system of $\frac{n^2(n+1)}{2}$ partial differential equations for the n unknown functions $X^i(\mathbf{x})$ of n variables x^i. These partial differential equations are linear in the unknowns X^i, and are very over determined (more equations than unknown functions). Solving these determines all the infinitesimal generators whose flow is by isometries.

Example 9.2.3. Continuing from example 9.2.2 above, we found the partial differential equations for a Killing vector $X = X^1\partial_x + X^2\partial_y$ to be

(9.23,9.25,9.24),

$$
\begin{aligned}
-\frac{2}{y^3}X^2 + \frac{2}{y^2}\frac{\partial X^1}{\partial x} &= 0,\\
-\frac{2}{y^3}X^2 + \frac{2}{y^2}\frac{\partial X^2}{\partial y} &= 0,\\
\frac{\partial X^2}{\partial x} + \frac{\partial X^1}{\partial y} &=0.
\end{aligned}
\tag{9.26}
$$

We first integrate in the second equation in 9.26 to get

$$X_2 = F(x)y,$$

where F is unknown. Using this in the last equation of 9.26 allows us to solve the last equation,

$$X^1 = -\frac{1}{2}y^2\frac{dF}{dx} + G(x).$$

Substituting into the first equation in 9.26 we have

$$\frac{d^F}{dx^2}y^2 - 2\frac{dG}{dx} + 2F(x) = 0.$$

This must hold for all possible y so the coefficient of y^2 must be zero which leads to,

$$F = a + bx, \quad G = c + ax + b\frac{x^2}{2}.$$

Therefore the most general Killing vector is

$$X = c\partial_x + a(x\partial_x + y\partial_y) + \frac{b}{2}((x^2 - y^2)\partial_x + xy\partial_y).$$

In general because the Killing vector equations are over-determined (more equations than unknown functions), they have only the zero vector-field as a solution.

Now in the same way we measure the rate of change of a function or a vector-field along a flow we define the Lie derivative of a metric at the point $p \in I\!R^n$ to be

$$(\mathcal{L}_X\gamma)_p = \lim_{t\to 0}\frac{1}{t}\left(\Psi_t^*\gamma_{\Psi(t,p)} - \gamma_p\right). \tag{9.27}$$

As p varies $\mathcal{L}_X\boldsymbol{\gamma}$ defines a field of bilinear functions on $I\!R^n$. A direct consequence of the computation in Theorem 9.2.1 gives a coordinate formula for $\mathcal{L}_X\boldsymbol{\gamma}$.

Lemma 9.2.4. *The Lie derivative in equation 9.27 is*

(9.28)
$$\mathcal{L}_X\boldsymbol{\gamma} = \sum_{i,j=1}^{n}\left(\sum_{l=1}^{n} g_{il}(\mathbf{x})\frac{\partial X^l(\mathbf{x})}{\partial x^j} + g_{lj}(\mathbf{x})\frac{\partial X^l(\mathbf{x})}{\partial x^i} + X^l(\mathbf{x})\frac{\partial g_{ij}(\mathbf{x})}{\partial x^l}\right) dx^i dx^j.$$

Proof. We observe

$$\mathcal{L}_X\boldsymbol{\gamma} = \lim_{t\to 0}\frac{1}{t}\left(\Psi_t^*\boldsymbol{\gamma}_{\Psi(t,p)} - \boldsymbol{\gamma}_p\right) = \left(\frac{\partial}{\partial t}\Psi_t^*\boldsymbol{\gamma}_{\Psi(t,p)}\right)\bigg|_{t=0},$$

which by the computation in equation 9.19 proves equation 9.28. □

Corollary 9.2.5. *A vector field X is a Killing vector of the metric $\boldsymbol{\gamma}$ if and only if $\mathcal{L}_X\boldsymbol{\gamma} = 0$.*

The equations $L_X\boldsymbol{\gamma} = 0$ are linear equations for the coefficients of X. Therefore any linear combination of Killing vector fields is a Killing vector field. However even more can be said as this last theorem for this section shows.

Theorem 9.2.6. *Let X and Y be Killing vector fields for the metric $\boldsymbol{\gamma}$, then $[X, Y]$ is also a Killing vector field.*

Proof. We will prove this by direct computation. Suppose $X = \sum_{i=1}^n X^i \partial x^i$, and $Y = \sum_{i=1}^n Y^i \partial_{x^i}$ are Killing vectors (with arguments suppressed). We then apply the derivative Y to the Killing equation 9.17 for X giving

$$0 = \sum_{j,m=1}^{n}\left(\frac{\partial^2 g_{kl}}{\partial x^m \partial x^j}X^m + \frac{\partial g_{kl}}{\partial x^m}\frac{\partial X^m}{\partial x^j} + \frac{\partial g_{km}}{\partial x^j}\frac{\partial X^m}{\partial x^l} + g_{km}\frac{\partial^2 X^m}{\partial x^l \partial x^j} + \frac{\partial g_{ml}}{\partial x^j}\frac{\partial X^m}{\partial x^k}\right.$$
$$\left. + g_{ml}\frac{\partial^2 X^m}{\partial x^k \partial x^j}\right) Y^j$$

Switching the role of X and Y in this equation, and taking the difference

gives
(9.29)

$$
\begin{aligned}
0 = \sum_{j,m=1}^{n} \Bigg(& \frac{\partial^2 g_{kl}}{\partial x^m \partial x^j}(X^m Y^j - Y^m X^j) + \frac{\partial g_{kl}}{\partial x^m}\left(Y^j \frac{\partial X^m}{\partial x^j} - X^j \frac{\partial Y^m}{\partial x^j} \right) \\
& + \frac{\partial g_{km}}{Y^j \partial x^j}\left(Y^j \frac{\partial X^m}{\partial x^l} - X^j \frac{\partial Y^m}{\partial x^l} \right) + g_{km}\left(Y^j \frac{\partial^2 X^m}{\partial x^l \partial x^j} - X^j \frac{\partial^2 Y^m}{\partial x^l \partial x^j} \right) \\
& + \frac{\partial g_{ml}}{\partial x^j}\left(Y^j \frac{\partial X^m}{\partial x^k} - X^j \frac{\partial Y^m}{\partial x^k} \right) + g_{ml}\left(Y^j \frac{\partial^2 X^m}{\partial x^k \partial x^j} - X^j \frac{\partial^2 Y^m}{\partial x^k \partial x^j} \right) \Bigg).
\end{aligned}
$$

Now by the equality of mixed partial derivatives, the first term in equation 9.29 satisfies

$$\sum_{j,m=1}^{n} \frac{\partial^2 g_{kl}}{\partial x^m \partial x^j}(X^m Y^j - Y^m X^j) = 0$$

and equation 9.29 is the Killing equation for the vector-field $[X, Y]$. □

The equations $L_X \boldsymbol{\gamma} = 0$ are linear equations for the coefficients of X. Therefore any linear combination of Killing vector fields is a Killing vector field. Combining this with Theorem 9.2.6 above and we have the following.

Corollary 9.2.7. *Let Γ be the vector space of Killing vectors, then Γ with the vector-field bracket $[\ ,\ ]$ is a Lie algebra.*

Remark 9.2.8. An upper bound on the dimension of the Lie algebra of Killing vectors is known to be [**?**]

$$\dim \Gamma \leq \frac{n(n+1)}{2}$$

9.3 Exercises

1. Which set of sets of vector-fields form a Lie algebra as a subalgebra of all vector fields on the given$\mathbb{R}^n$.

   (a) On $\mathbb{R}^2$, $\Gamma = \mathrm{span}\{\partial_x,\ x\partial_x,\ y\partial_y,\quad x^2\partial_x + xy\partial_y\}$.

   (b) On $\mathbb{R}^3$, $\Gamma = \mathrm{span}\{x\partial_y - y\partial_x,\quad x\partial_z,\quad y\partial_z - z\partial_y\}$.

   (c) On $\mathbb{R}$, $\Gamma = \mathrm{span}\{\partial_x,\ x\partial_x,\ x^2\partial_x,\}$.

2. Prove Lemma 9.1.3. Hint: Apply each identity to an arbitrary function f.

3. If $f \in C^\infty(\mathbb{R}^n)$ is an invariant function for the vector fields X and Y on $\mathbb{R}^n$, show that f is an invariant function of $[X, Y]$.

4. Show that if X and Y commute with Z, then $[X, Y]$ commutes with Z.

5. Are there vector fields X, Y, and Z which satisfy

$$[X, Y] = Z, \quad [X, Z] = -X + 2Y, \quad [Y, Z] = X + Y + Z.$$

6. Compute the Lie brackets of the following pairs of vector fields.

   (a) $X = \partial_x + z\partial_y, \quad Y = z\partial_z$.

   (b) $X = \partial_x + y\partial_z, \quad Y = z\partial_z$.

   (c) $X = y^2\partial_x + z^2\partial_y + x^2\partial_z, \quad Y = -y\partial_x + x\partial_y$.

   (d) Do the flows commute?

7. Consider the following pairs X, Y of vector-fields.

   1) On $\mathbb{R}^2$, $X = \partial_x + y\partial_z$, and $Y = \partial_y + x\partial_z$.

   2) On $\mathbb{R}^3$, $X = \partial_x + y\partial_z$, and $Y = \partial_y + x\partial_z$.

   3) On $\mathbb{R}^4$, $X = x\partial_w + z\partial_y$, $Y = w\partial_y + x\partial_z$.

   For each pair let Ψ be the flow of X and Φ the flow of Y. Show that

   (a) $[X, Y] = 0$.

   (b) The flows Ψ and Ψ are commuting flows.

(c) X is Φ invariant.

(d) Y is Ψ invariant.

(e) Find coordinates which simultaneously straighten the vector fields X and Y. (Lemma 9.1.12)

8. Show that the transformation group in problem 1(c) of Assignment 8 is a one-parameter group of isometries of the Euclidean metric in three dimensions. What is the corresponding Killing vector field.

9. Show that the one parameter group of transformation in equation 8.1 are isometries of the metric $\boldsymbol{\gamma}$ in the plane given by

$$\boldsymbol{\gamma} = \frac{1}{1 + x^2 + y^2}(dx^2 + dy^2)$$

10. For the metric tensor on $\mathbb{R}^2$ given by

$$\boldsymbol{\gamma} = \frac{1}{(1 + x^2 + y^2)^2}(dx^2 + dy^2),$$

(a) Show that the vector fields

$$X = x\partial_y - y\partial_x, \quad Y = 2xy\partial_x + (1 - x^2 + y^2)\partial_y$$

are Killing vectors-fields.

(b) Find a third linearly independent Killing vector field.

(c) Check whether or not the span of the three Killing vectors forms a Lie algebra.

11. For the metric tensor on $\mathbb{R}^3$ given by

$$\boldsymbol{\gamma} = dx^2 + dy^2 - 2xdydz + (1 + x^2)dz^2 .$$

(a) show that the vector field

$$X = \partial_x + z\partial_y$$

is a Killing vector for $\boldsymbol{\gamma}$.

(b) Compute the flow Ψ of X and show that Ψ is one-parameter family of isometries of $\boldsymbol{\gamma}$.

Chapter 10

Group actions and Multi-parameter Groups

10.1 Group Actions

Let G be a group (see Definition 12.1.1), and S a set. We begin this section with a definition.

Definition 10.1.1. An action of G on S is a function $\mu : G \times S \to S$ which satisfies

1. $\mu(e, x) = x$ for all $x \in S$,
2. $\mu(a, \mu(b, x)) = \mu(a * b, x)$ for all $a, b \in G$, $x \in S$.

If we fix $a \in G$ in the function μ, then $\mu(a, -) : X \to X$ and denote this function by μ_a. If a^{-1} is the inverse element of a then by properties 1 and 2 in definition 10.1.1

$$\mu_a(\mu(a^{-1}, \mathbf{x})) = \mu(e, \mathbf{x}) = \mathbf{x}\ , \quad for\ all \mathbf{x} \in S.$$

This implies that μ_a is an invertible function, and that is a bijection because $(\mu_a)^{-1} = \mu_{a^{-1}}$.

We will exclusively be interested in the case where $S \subset \mathbb{R}^n$.

Example 10.1.2. Suppose

$$\mu(a, b, x, y) = (ax + by, y)$$

is a group action where $a, b \in I\!R$ are parameters on the group. Determine the maximum value of a, b so that μ defines a group action on $I\!R^2$. Find the group multiplication.

Suppose $(a, b) \in G$ and $(a', b') \in G$. The left side in the group action condition 2 in 10.1.1 is

$$\mu(a, b, \mu(a', b', x, y)) = \mu(a, b, (a'x+b'y, y)) = (a(a'x+b'y)+by, y) = (aa'x+(ab'+b)y, y).$$

If $(a, b) * (a', b') = (c, d)$, then This is supposed to be equal to the right side which is $\mu((a, b) * (a', b'), x, y) = (cx + dy, y)$. Therefore

$$(a, b) * (a', b') = (aa', ab' + b).$$

For this to be a group there exists an identity element $e = (a', b')$ such that

$$(a, b) * (a', b') = (aa', ab' + b) = (a, b).$$

for all $(a, b) \in G$. Therefore $a' = 1, b' = 0$, and we conclude $(a, b) \in I\!R^* \times I\!R$ are the maximum value for the parameters (a, b).

Example 10.1.3. Let $G = I\!R$ with ordinary addition as the group operation. Then a flow $\Psi : I\!R \times I\!R^n \to I\!R^n$ is an action of G on $I\!R^n$ (here $\mu = \Psi$). In fact properties 1 and 2 in definition 8.1.1 of a flow are properties 1 and 2 in definition 10.1.1.

Let $P(S)$ be the permutation group on S. That is

$$P(S) = \{F : S \to S \mid F \text{is a bijection}\}.$$

The group operation on $P(S)$ is composition of functions. We have

Lemma 10.1.4. *Let $\mu : G \times S \to S$, and let $\Phi : G \to P(S)$ be the map $\Phi(g) = \mu_g$. Then Φ is a homomorphism*

If $p \in S$ then the set

$$\mathcal{O}_p = \{\ q \in S \mid q = ap\ a \in G\ \} \tag{10.1}$$

is called the *G-orbit of x*. A group action is said to be transitive if $\mathcal{O}_x = S$. In other words, a group G acts transitively on S if and only if given any fixed point $\mathbf{x}_0 \in S$, and any other point $\mathbf{x} \in S$, there exists $g \in G$ such that $g\mathbf{x}_0 = \mathbf{x}$. If the action is transitive on S, then S is called a homogeneous space, these spaces are very important in more advanced courses.

The *isotropy group* of $p \in S$ is the subset $G_p \subset G$ defined by

$$G_p = \{\ g \in G \mid gp = p\ \}.$$

Example 10.1.5. Let $I\!R$ act on $I\!R^2$ by rotations,

$$\mu(t, x, y) = (x \cos t - y \sin t, x \sin t + y \cos t).$$

(This is the example of a flow we have worked with before.) Let $p = (1, 0)$, then $\mathcal{O}_p = \{\ (x, y) \mid x^2 + y^2 = 1\}$. To find G_p we need to solve

$$(\cos t, \sin t) = (1, 0)$$

and we find $t = 2k\pi, k \in Z$.

If $p = (0, 0)$, then orbit is$\mathcal{O}_p = (0, 0)$, and the isotropy group $G_p = I\!R$, the entire group.

Lemma 10.1.6. *For any $p \in S$, the subset $G_p \subset G$ is a subgroup.*

Proof. To prove this we refer to Lemma 12.1.7. First note that $e \in G_p$ for any $p \in S$ so G_p is not empty. Now suppose $a, b \in G_p$, we then check that $ab \in G_x$. (That is we need to check that $(ab)p = p$.) By the second property of group actions we find

$$(ab)p = a(bp) = ap = p,$$

and therefore $ab \in G_p$. We now need to check if $a \in G_p$, then $a^{-1} \in G_p$. We compute

$$a^{-1}p = a^{-1}(ap) = (a^{-1}a)p = ep = p.$$

Therefore G_p is a subgroup. □

Definition 10.1.7. A group G is an n-parameter (or multi-parameter) group if

- $G \subset I\!R^n$ is an open set,
- the multiplication map $* : G \times G \to G$ is smooth, and
- the map $\iota : G \to G$ with $\iota(a) = a^{-1}$ is smooth.

The rest of this chapter involves the action of multi-parameter groups on sets $S \subset I\!R^n$. This generalizes the notion of flow.

A multi-parameter group is a special case of what are called *Lie groups*. The general definition of Lie group is beyond the scope of this book. All the constructions given here are important and carry over almost directly in the general case of Lie groups and their actions.

XXXXXXXXXXXXXXX EXAMPLES NEED IMPROVEMENT XXXXXXXXXXXX

Example 10.1.8. Let $G_1 = (a, b), a, b \in I\!R$, with the multiplication law

$$(a, b) * (a', b') = (a + a', b + b').$$

Then G_1 is a two-parameter (abelian) group. The properties in definition 10.1.7 are easily verified.

An action of G_1 on $I\!R^2$ is given by

$$(a, b)(x, y) = (x + a, y + b).$$

This group action is transitive. Why? Let (x_1, y_1) and (x_2, y_2) be two points in $I\!R^2$. We can find a group element (a, b) such that

$$(a, b)(x_1, y_1) = (x_2, y_2),$$

and this is $(a = x_2 - x_1, b = y_2 - y_1)$.

Example 10.1.9. Let $G_2 = (a, b), a \in I\!R^* n, b \in I\!R$, with the multiplication law

$$(a', b') * (a, b) = (a'a, b' + b). \qquad (10.2)$$

Then G_2 is a two-parameter (abelian) group. The set $G = I\!R^* \times I\!R \subset I\!R^2$ is open. The identity element $e = (1, 0)$, and the multiplication function in 10.2 is smooth. The map ι in property 3 of definition 10.1.7 is

$$\iota(a, b) = (a^{-1}, -b), \quad a \in I\!R^*, b \in I\!R,$$

and ι is smooth.

An action of the group G_2 on $I\!R^2$ is

$$(a, b)(x, y) = (ax, y + b).$$

This action is not transitive. For example $(0, 1)$ and $(1, 0)$ are not in the same orbit. It is not difficult to check that there are only two orbits in this example.

Example 10.1.10. Let $G_3 = (a, b), a \in I\!R^* n, b \in I\!R$, with the multiplication law

$$(a', b') * (a, b) = (a'a, a'b + b').$$

Then G_3 is a two-parameter group which is not Abelian. The identity is $e = (1, 0)$ and

$$\iota(a, b) = (a^{-1}, -a^{-1}b).$$

The properties in definition 10.1.7 are easily verified.

An action of the group G_3 on $I\!R$ is

$$(a, b)x = (ax + b).$$

This action is transitive. If $x_0 \in I\!R$, then G_{x_0} is found by solving

$$x_0 = ax_0 + b,$$

for b. So the isotropy group is

$$G_{x_0} = (a, (1 - a)x_0).$$

An action of the group G_2 on $I\!R^2$ is

$$(a, b)(x, y) = (ax + by, y),$$

and this action is intransitive.

Example 10.1.11. Consider the transformations of $I\!R^2$

$$\mu((a, b, c), (x, y)) = (ax + by, cy).$$

What would the group multiplication law for (a, b, c) need to be for the function μ to be a group action? What are maximum domains for the parameters a, b, c ?

First we need to find the identity. This would be values for a, b, c giving the identity transformation, and these are $(1, 0, 1)$. Now suppose (a', b', c') is another element of the group, then condition 2 gives

$$\begin{aligned}\rho((a', b', c') * (a, b, c), (x, y)) &= \rho((a', b', c'), (ax + by, cy)) \\ &(a'(ax + by) + b'(cy), c'cy) \\ &(a'ax + (a'b + b'c)y, c'cy).\end{aligned}$$

For this equation to hold for all $x, y \in I\!R^2$ we must have

$$(a', b', c')(a, b, c) = (a'a, a'b + b'c, c'c).$$

Note that for an inverse to exists $c \neq 0$, and $a \neq 0$. Therefore $G = \{\ (a, b, c) \mid a \neq 0, c \neq 0\ \}$ (with the maximum domain for the parameters).

10.2 Infinitesimal Generators

Suppose that $G \subset I\!R^n$ is a multi-parameter group acting on $I\!R^m$. Let $e = (x_0^1, x_0^2, \ldots, x_0^n)$ be the identity element, and denote by $\mathbf{g} = T_eG$ the tangent space at the point e, and let $e_1, e_2, \ldots, e_n$ be a basis for the tangent space. We will now associate to each tangent vector e_i a vector field X_i on $I\!R^m$ which is called the *infinitesimal generator* corresponding to e_i.

Choose $e_i \in T_eG$, and let $\sigma : I \to G$ be a curve with the property that $\sigma(t_0) = e$, and $\dot\sigma(t_0) = e_i$.

Let μ be an action of the multi-parameter group G on $I\!R^n$.

Definition 10.2.1. The infinitesimal generator X_i corresponding to e_i is the vector-field in $I\!R^n$ defined point-wise by,

$$X_p = \frac{d}{dt}\mu(\sigma(t), x)\Big|_{t=t_0}, \quad for\ all\ p \in I\!R^n$$

Example 10.2.2. Consider the group and actions in example 13.5. We have $e = (1, 0)$, and let

$$e_1 = \partial_a, \quad e_2 = \partial_b$$

be a basis for T_eG. The two curves

$$\sigma_1(t) = (t, 0), \quad \sigma_2(t) = (1, t) \tag{10.3}$$

satisfy

$$\sigma_1(1) = e, \dot\sigma_1(1) = e_1,$$
$$\sigma_2(0) = e, \dot\sigma_2(0) = e_2.$$

Using these curves we can find the corresponding infinitesimal generators.

For the action of G_3 on $I\!R$ we find the infinitesimal generator for e_1 is

$$X_1(x) = \frac{d}{dt}(xt)|_{t=1} = (x)_x, \quad X_1 = x\partial_x.$$

The infinitesimal generator corresponding to e_2

$$X_2 = \frac{d}{dt}(x+t)|_{t=0} = 1_x, \quad X_2 = \partial_x.$$

For the action of G_3 on $I\!R^2$ we find the infinitesimal generator corresponding to e_1 is

$$X_1 = x\partial_x,$$

while the generator for e_2 is

$$X_2(x,y) = \frac{d}{dt}(x+ty,y)|_{t=0} = (y,0)_{x,y}, \quad X_2 = y\partial_x$$

Let $X_1, X_2, \ldots, X_n$ be the infinitesimal generators of G corresponding to $e_1, \ldots, e_n$, and let

$$\Gamma = \text{span}\{X_1, \ldots, X_n\}$$

where the span is computed over the real numbers. This is a subspace of the vector space of ALL vector fields on $I\!R^m$.

Theorem 10.2.3. *The vector space* Γ *is a Lie algebra (using the bracket of vector fields), with dimension* $\dim \Gamma$*.*

If $\mathbf{g}$ is a Lie algebra, and $\mathbf{h} \subset \mathbf{g}$ is a subspace which is a Lie algebra with the inherited bracket, then $\mathbf{h}$ is called a subalgebra. The Lie algebra Γ is a subalgebra of the infinite dimensional Lie algebra of all vector fields.

Example 10.2.4. Continuing with example 13.XX, in the first case we have

$$\Gamma = \text{span}\{x\partial_x, \partial_x\},$$

is a two dimensional vector space. To check this forms a Lie algebra we need to only check that $[X_1, X_2] \in \Gamma$. We find

$$[X_1, X_2] = -\partial_x \in \Gamma.$$

In the second case

$$\Gamma = \text{span}\{\ x\partial_x,\ y\partial_y\ \}$$

and Γ is two dimensional. Again we only need to check that $[X_1, X_2] \in \Gamma$, and

$$[x\partial_x, y\partial_x] = -y\partial_x \in \Gamma.$$

Therefore Γ is a two-dimensional Lie algebra.

Given a Lie algebra $\mathbf{g}$ of dimension n, let $e_1, \ldots, e_n$ be a basis for $\mathbf{g}$. Since $[e_i, e_j] \in \Gamma$ there exists $c^k \in I\!R$, such that

$$[e_i, e_j] = \sum_{k=1}^{n} c^k e_k.$$

Since this equation holds for each $1 \leq i, j \leq n$, there exists $c_{ij}^k \in I\!R, 1 \leq i, j, k \leq n$, such that

$$[e_i, e_j] = \sum_{k=1}^{n} c_{ij}^k e_k.$$

These are the *structure constants* of the Lie algebra $\mathbf{g}$ in the basis e_i. Writing the brackets in a table, we get the "multiplication table" for $\mathbf{g}$.

Example 10.2.5. For Γ in Example 13.8 above, the multiplication table is

$$[X_1, X_1] = 0, [X_1, X_2] = -X_2, [X_2, X_1] = X_2, [X_2, X_2] = 0.$$

Of course $[e_i, e_i] = 0$, and $[e_j, e_i] = -[e_i, e_j]$, and so the multiplication table is skew-symmetric.

10.3 Right and Left Invariant Vector Fields

Let $G \subset I\!R^n$ be a multi-parameter group. The group multiplication $\mu : G \times G \to G$ allows us to think of G acting on itself. Therefore we can compute the infinitesimal generators of G acting on itself, and the span of these generators are the *right invariant vector fields on* G.

Example 10.3.1. Consider G_3 from example 10.1.10, where

$$(a', b') * (a, b) = (a'a, a'b + b').$$

Using the curves in equation (10.3),

$$\sigma_1(t) = (t, 0), \quad \sigma_2(t) = (1, t)$$

we compute the corresponding infinitesimal generators,

$$X_{(a,b)} = \frac{d}{dt}\mu(\sigma_1(t), (a, b))|_{t=1} = \frac{d}{dt}(ta, tb)|_{t=1} = (a, b), \quad X = a\partial_a + b\partial_b,$$

and

$$Y_{(a,b)} = \frac{d}{dt}\mu(\sigma_2(t), (a, b))|_{t=0} = \frac{d}{dt}(a, t + b)|_{t=0} = (0, 1), \quad Y = \partial_b.$$

Note that

$$[X, Y] = -Y$$

and so the right invariant vector fields $\mathbf{g} = \text{span}\{X, Y\}$ are a two dimensional Lie algebra.

Theorem 10.3.2. *Let X be an infinitesimal generator of G acting on itself on the left. Then*

$$(R_g)_* X = X.$$

That is X is a **right***-invariant vector-fields. The set of all right invariant vector-fields form an n-dimensional Lie algebra.*

If we now view G as acting on the right of G instead of the left, (so G is an example of a **right** action), we end up with the left invariant vector-fields.

Example 10.3.3. Consider again example 10.1.10

$$(a', b') * (a, b) = (a'a, a'b + b').$$

Using the curves in equation (10.3),

$$\sigma_1(t) = (t, 0), \quad \sigma_2(t) = (1, t)$$

their flows acting on the right are then

$$\mu((a, b), (t, 0)) = (at, b), \quad \mu((a, b), (1, t)) = (a, b + at).$$

The corresponding infinitesimal generators are

$$U_{(a,b)} = \frac{d}{dt}\mu((a, b), (t, 0)|_{t=1} = \frac{d}{dt}(ta, b)|_{t=1} = (a, 0), \quad U = a\partial_a,$$

$$V_{(a,b)} = \frac{d}{dt}\mu((a, b), \sigma_2(t))|_{t=0} = \frac{d}{dt}(a, b + ta)|_{t=0} = (0, 1), \quad V = a\partial_b.$$

Note that

$$[U, V] = V$$

and so the left invariant vector fields $\mathbf{g} = \text{span}\{X, Y\}$ are a two dimensional Lie algebra.

Theorem 10.3.4. *Let X be an infinitesimal generator of G acting on itself on the right. Then*

$$(L_g)_* X = X.$$

That is X is a **left***-invariant vector-fields. The set of all left invariant vector-fields form an n-dimensional Lie algebra.*

If σ^i are curves which give rise to the right invariant vector-fields X_i, and the left invariant vector-fields Y_i then

$$[X_i, X_j] = \sum_{k=1}^{n} C^k_{ij} X_k, \quad [Y_i, Y_j] = -\sum_{k=1}^{n} C^k_{ij} Y_k.$$

It is also a curious fact that

$$[X_i, Y_j] = 0!$$

10.4 Invariant Metric Tensors on Multi-parameter groups

Let X_i be a basis for the left invariant vector-fields on a multi-parameter group G, and let α^i be the algebraic dual defined in equation 5.66. We then define the field of bilinear functions $\boldsymbol{\gamma}_p : T_pG \times T_pG \to I\!R$, by

$$\boldsymbol{\gamma}_p = \sum_{i=1}^{n} \alpha^i_p \otimes \alpha^i_p, \quad for\ all\ p \in G. \tag{10.4}$$

Lemma 10.4.1. *The field of bilinear functions* $\boldsymbol{\gamma}$ *defined in equation 10.4 is a metric tensor on* G*.*

Proof. We □

A metric tensor $\boldsymbol{\gamma}$ for which every diffeomorphism $L_a : G \to G$ is an isometry

$$L_g^* \boldsymbol{\gamma} = \boldsymbol{\gamma}$$

is called a left-invariant metric on G.

Lemma 10.4.2. *The metric tensor in 10.4 is left invariant.*

Proof. We check the isometry condition (property 3 in Lemma 6.3.2)

$$L_a^* \boldsymbol{\gamma}_{ap}(X_p, X_p) = \boldsymbol{\gamma}_p(X_p, X_p) \tag{10.5}$$

equation on the basis $\{X_i(p)\}$ of left invariant vector-fields on G at p. The left side is
(10.6)

$$\begin{aligned} L_a^*\boldsymbol{\gamma}_{ap}(X_i, X_i) &= \boldsymbol{\gamma}_{ap}\left((L_a)_*X_i(p), (L_a)_*X_i(p)\right) && \text{byequation 6.33} \\ &= \boldsymbol{\gamma}_{ap}\left(X_i(ap), X_i(ap)\right) && \text{by left invariance of } X^i \\ &= \sum_{j=1}^{n} \alpha^j_{ap}(X_i(ap))\alpha^j_{ap}(X_i(ap)) && \text{by equation 10.4} \\ &= 1. \end{aligned}$$

The right hand side of equation 10.5 is

$$\boldsymbol{\gamma}_p(X_i(p), X_i(p)) = \sum_{j=1}^{n} \alpha^j_p(X_i(p))\alpha^j_{ap}(X_i(p)) = 1.$$

Therefore equation 10.5 holds for on basis of T_pG for any p, and so L_g is an isometry. □

By Theorem 10.3.2 the infinitesimal generators of the left action are the right invariant vector-fields, which provides the following corollary.

Corollary 10.4.3. *The right invariant metric vector-fields on G are Killing vectors for the metric tensor in 10.4.*

Note the left-right switch in Lemma 10.4.2 and Corollary 10.4.3: The right invariant vector-fields are the Killing vectors for a left invariant metric tensor!

Example 10.4.4. For the group $G = (a, b, c),\ a, c \in I\!R^*, b \in I\!R$ and multiplication

$$\mu((a, b, c), (x, y, z)) = (ax, y = ay + bz, z = cz) \tag{10.7}$$

the identity if $e = (1, 0, 1)$. The left invariant vector fields for the basis $\{\partial_a, \partial_b, \partial_c$ of T_eG, the left invariant vector-fields are computed to be

$$X_1 = x\partial_x,\ X_2 = x\partial_y,\ X_3 = x\partial_z + xyz^{-1}\partial_y\,.$$

The dual frame of differential one-forms (equation 5.66) are

$$\alpha^1 = \frac{1}{x}dx,\ \alpha^2 = \frac{1}{x}(dy - \frac{y}{z}dc),\ \alpha^3 = \frac{1}{z}dz. \tag{10.8}$$

Using the one-forms in equation 10.8 the metric tensor in equation 10.4 is

$$\begin{aligned}\gamma &= \left(\frac{1}{x}dx\right)^2 + \left(\frac{1}{x}(dy - \frac{y}{z}dz)\right)^2 + \left(\frac{1}{z}dz\right)^2 \\ &= \frac{1}{x^2}dx^2 + dy^2 - 2\frac{y}{zx^2}dydz + \left(\frac{y^2}{x^2z^2} + \frac{1}{z^2}\right)dz^2\end{aligned} \tag{10.9}$$

By Lemma 10.4.2 the metric in 10.9 is invariant under $\mu_{(a,b,c)} : I\!R^3 \to I\!R^3$ from equation 10.7. This explains example 6.3.7.

Now a basis for the right invariant vector-fields are

$$Y_1 = x\partial_x + y\partial_y,\ Y_2 = z\partial_y,\ Y_3 = z\partial_z.$$

10.5 Exercises

1. Suppose that each of the following transformations defines a group action. Find the maximum domain for the parameters and group multiplication law $*$. Determine the identity in the group and give the inverse of each element.

   (a) $\bar{x} = ax + by + c, \quad \bar{y} = a^k y$, with parameters (a, b, c).

   (b) $\bar{x} = x + ay + bz, \quad \bar{y} = y + cz, \quad \bar{z} = z$ with parameters (a, b, c)

   (c) $\bar{x} = \exp(\lambda)(x\cos\theta + y\sin\theta), \quad \bar{y} = \exp(\lambda)(-x\sin(\theta) + y\cos\theta), \quad (\lambda, \theta)$, with parameters λ, θ.

   (d) $\bar{x} = ax + by + c, \quad \bar{y} = y + \log(a)$, with parameters (a, b, c).

2. Let $\mu : G \times I\!R^3 \to I\!R^3$ be given by

$$\mu((a, b, c), (x, y, z)) = (ax, bx + ay, z + cx).$$

   (a) Find the group multiplication law in order that μ be a group action.

   (b) Determine the identity element, and the inverse of $(a, b, c) \in G$.

   (c) What are the possible values for the group parameters (a, b, c).

   (d) Are the points $(1, 2, 3)$, $(4, 5, 6)$ on the same orbit? Is this action transitive on $I\!R^3$?

   (e) Compute $G_{(x_0, y_0, z_0)}$ where $x_0 \neq 0$. Do all points in $I\!R^3$ have the same isotropy subgroup?

   (f) Compute the infinitesimal generators of this action.

   (g) Compute a basis for the right invariant vector fields on G, and compute the bracket multiplication table.

   (h) Compute a basis for the left invariant vector fields on G using the same curves as in the previous question. Compute the bracket multiplication table and compare with the right invariant vector fields.

3. Which of the following actions are transitive. Prove you claim.

   (a) Question 1 a.

(b) Question 1 d.

(c) The group $G = SL(2)$ acting on the upper half plan $Im\, z > 0$ by $z \to \dfrac{az+b}{cz+d}$, where

$$SL(2) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2\times 2}(\mathbb{R}) \mid ad - bc = 1 \}.$$

(d) Let $G = O(3)$,

$$O(3) = \{ \ A \in M_{3\times 3}(\mathbb{R}) \mid AA^T = I \ \},$$

act on $R^3 - 0$ by $\mu(A, \mathbf{x}) = A\mathbf{x}$.

4. Find the isotropy subgroup for each of the following actions at the given point.

   (a) Question 1 a $\quad (x_0, y_0) = (2, 3)$ and $(x_0, y_0) = (0, 0)$.

   (b) Question 1 b $\quad (x_0, y_0, z_0) = (1, 2, 3)$ and $(x_0, y_0, z_0) = (1, 1, 0)$, and $(x_0, y_0, z_0) = (1, 0, 0)$

   (c) Question 4 c $\quad z_0 = i$ and $z_0 = i + 1$.

5. Find the right invariant vector fields on the multi-parameter group in Example 13.6.

6. Let
$$ds^2 = (u+v)^{-2}(du^2 + dv^2)$$
be a metric tensor on $U = \{(u, v) \subset \mathbb{R}^2 \mid u + v \neq 0\}$, and let $a \in \mathbb{R}^*, b \in \mathbb{R}$, and let $\Phi_{(a,b)} : \mathbb{R}^2 \to \mathbb{R}^2$ be
$$\Phi_{(a,b)}(u, v) = (au + b, av - b)$$

   (a) Show that Φ is a group action.

   (b) Show that $U \subset \mathbb{R}^2$ is a $\Phi_{(a,b)}$ invariant set for any $a \in \mathbb{R}^*, b \in \mathbb{R}$.

   (c) Show that $\Phi_{(a,b)}$ is an isometry of the metric tensor for any $a \in \mathbb{R}^*, b \in \mathbb{R}$.

   (d) Compute the infinitesimal generators of Φ.

(e) Check that infinitesimal generators vector-fields are Killing vectors.

7. Let $GL(n, I\!R)$ be the group of invertible matrices (See appendix A).

(a) Show that $\rho : GL(n, I\!R) \times I\!R^n \to I\!R^n$

$$\rho(A, \mathbf{x}) = A\mathbf{x} \quad A \in GL(n, I\!R),\ \mathbf{x} \in I\!R^n.$$

is a group action.

(b) Let $M = M_{n\times n}(I\!R)$ be the set of all $n \times n$ real matrices. Is the function $\rho : G \times M \to M$ given by

$$\rho(A, X) = AXA^{-1}, \quad A \in G, X \in M$$

an action of $GL(n, I\!R)$ on $M = R^{n^2}$?

8. Let $G = (a, b, c), a \in I\!R^*, b, c \in I\!R$ be a group with multiplication

$$(a, b, c) * (x, y, z) = (ax, ay + b, az + c).$$

(a) Compute the left and right invariant vector-fields.

(b) Compute the coframe dual to the left invariant vector fields.

(c) Construct the metric tensor as in 10.4 from the left invariant vector fields and check that the right invariant vector fields are Killing vectors.

(d) Find a basis for all Killing vectors fields for the metric $\boldsymbol{\gamma}$ constructed in part 3.

Chapter 11

Computations in Maple

In this final chapter we demonstrate the differential geometry package in Maple.

Load the differential geometry and tensor packages.

```
> with(DifferentialGeometry):  with(Tensor):
```

Initialize a coordinate systems with x,y,z; Label this M.

```
> DGsetup([x,y,z],M):
```

Input the metric (named g here) from example 6.3.7 and check that the transformations in 6.42 are isometries. The expression &t stands for the tensor

```
> g := evalDG(1/x^2*(dx&t dx +dy &t dy - y/z*( dy &t dz + dz
&t dy)) +
(y^2/x^2/z^2+1/z^2)*dz &t dz);
```

$$g := \frac{dx\, dx}{x^2} + \frac{dy\, dy}{x^2} - \frac{y\, dy\, dz}{x^2\, z} - \frac{y\, dz\, dy}{x^2\, z} + \frac{(y^2 + x^2)\, dz\, dz}{x^2\, z^2}$$

```
> Phi1 := Transformation(M, M, [x = x * a, y = a*y+b*z, z = c*z]);
```

$$\Phi 1 := [x = x\, a,\ y = a\, y + b\, z,\ z = c\, z]$$

```
> Pullback(Phi1, g);
```

$$\frac{dx\, dx}{x^2} + \frac{dy\, dy}{x^2} - \frac{y\, dy\, dz}{x^2\, z} - \frac{y\, dz\, dy}{x^2\, z} + \frac{(y^2 + x^2)\, dz\, dz}{x^2\, z^2}$$

Therefore the isometry condition 6.24 is checked.

Check equation 6.43, the pullback of the coordinate differential one-forms

```
> Pullback(Phi1,dx); Pullback(Phi1,dy); Pullback(Phi1,dz);
```

$$a\,dx$$
$$a\,dy + b\,dz$$
$$c\,dz$$

Input the orthonormal frame from equation 5.27.

For example, the vector \partial_x is represented by D_x.

```
> Z1:=evalDG( x*D_x):
> Z2:=evalDG( x*D_y):
> Z3:=evalDG( y*D_y+z*D_z):
```

Now check that Z1,Z2,Z3 for an orthonormal frame

```
> A:=matrix(3,3):
> for i from 1 to 3 do for j from 1 to 3 do
A[i,j]:=(ContractIndices(ContractIndices(g, Z||i,
[[1,1]]),Z||j,[[1,1]])):od:od:
> op(A);
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Create a new coordinate system u,v,w; Label with N.

```
> DGsetup([u,v,w],N):
```

Construct a transformation called $\Phi_2$from N to M.

```
> Phi2:=Transformation(N,M, [x= exp(u),y=v, z=exp(w)]);
```

$$\Phi 2 := [x = e^u,\ y = v,\ z = e^w]$$

Pullback the metric g with Φ_2.

```
> Pullback(Phi2, g);
```

$$du\, du + e^{(-2\,u)}\, dv\, dv - v\, e^{(-2\,u)}\, dv\, dw - v\, e^{(-2\,u)}\, dw\, dv + \left(v^2 + e^{(2\,u)}\right) e^{(-2\,u)}\, dw\, dw$$

Chapter 12

Algebraic Notions

12.1 Groups

A binary operation on a set S is a function $b : S \times S \to S$. So it takes two things in S (binary) and it's output is some other element in S.

Definition 12.1.1. A **group** G is a non-empty set with one binary operation $* : G \times G \to G$ (usually called group multiplication) which satisfies

G1) $(a * b) * c = a * (b * c)$,

G2) there exists $e \in G$ such that $e * a = a = a * e$,

G3) for all $a \in G$ there exists $b \in G$ such that $a * b = e = b * a$,

for all $a, b \in G$. The element b in $G3$) is the inverse of a, often written as a^{-1}.

If $*$ satisfies

G4') $a * b = b * a$

then G is an **Abelian** group.

Example 12.1.2. Let $G = I\!R$ and take for the group operation ordinary addition of real numbers. That is $a * b = a + b$. Then $I\!R$ with $+$ is an Abelian group. The identity element is the number 0.

Example 12.1.3. Let $I\!R^*$ denote the non-zero real numbers. Let $*$ be ordinary multiplication, the $I\!R^*$ with $*$ is an Abelian group. The identity element is the number 1.

Example 12.1.4. Let $G = GL(n, \mathbb{R})$ denote the set of $n \times n$ invertible matrices. For the binary operation $*$ we take ordinary matrix multiplication. That is $A*B = AB$. If $n = 1$ this is example 9.3, but when $n > 1$ this group is **not** Abelian because matrix multiplication is not commutative. The identity element is the identity matrix.

Definition 12.1.5. A subset $H \subset G$ of a group G which is a group with the operation $*$ from G is called a *subgroup* .

Example 12.1.6. Let $G = GL(n, \mathbb{R})$ and let $SL(n, \mathbb{R}) \subset GL(n, \mathbb{R})$ be the matrices with determinant 1,

$$SL(n, \mathbb{R}) = \{ \ A \in GL(n, \mathbb{R}) \mid detA = 1\}.$$

We can ask if this is a group on it's own with the group operation being matrix multiplication. However it is not obvious that matrix multiplication is a binary operation on $SL(n, \mathbb{R})$. That is, if $A, B \in SL(n, \mathbb{R})$ is $AB \in SL(n, \mathbb{R})$? This is easily checked

$$\det AB = \det A \det B = 1.$$

Therefore this is a binary operation on $SL(n, \mathbb{R})$. Another way to phrase this is $SL(n, \mathbb{R}) \subset GL(n, \mathbb{R})$ is closed under the group operation on $GL(n, \mathbb{R})$. For $SL(n, \mathbb{R})$ to be a group every element must have an inverse with determinant 1. If $A \in SL(n, \mathbb{R})$ the A is invertible, and

$$\det A^{-1} = \frac{1}{\det A} = 1.$$

The simple test for a subgroup is the following.

Lemma 12.1.7. *A non-empty subset $H \subset G$ is a subgroup if and only if*

1. *$a, b \in H$ then $a * b \in H$ (closed with respect to $*$),*

2. *if $a \in H$ then $a^{-1} \in H$.*

Note these conditions imply that $e \in H$.

There will be many more examples later.

12.2 Rings

Definition 12.2.1. A **ring** R is a non-empty set with two binary operations $+, \cdot$ which satisfy,

R1) $(a+b)+c = a+(b+c)$,
R2) $a+b = b+a$,
R3) there exits $0 \in R$ such that $a+0 = a$,
R4) for all $a \in R$, there exists $b \in R$ such that $a+b = 0$,
R5) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
R6) $a \cdot (b+c) = a \cdot b + a \cdot c$, and $(a+b) \cdot c = a \cdot c + b \cdot c$,

for all $a, b, c \in R$.

If in addition $\cdot$ satisfies,

R7') $a \cdot b = b \cdot a, \ for \ all a, b \in R$

then R is a **commutative ring**. If there exits $1_R \in R$ satisfying

R8') $1_R \cdot a = a \cdot 1_R = a$

then R is a ring R with multiplicative identity.

Example 12.2.2. Let $R = C^\infty(U)$ where $U \subset \mathbb{R}^n$ is an open set. For $+$ take addition of functions,

$$(f+g)(x) = f(x) + g(x)$$

and multiplication

$$(f \cdot g)(x) = f(x)g(x),$$

where $f, g \in C^\infty(U)$

Remark: A ring R or vector space V considered with ONLY the operation $+$ is an Abelian group.

Index

Bibliography

[1] A. Hyvarinen, J. Karhunen, E. Oja, *Independent Component Analysis*, Wiley Interscience, 2001.

[2] T. Moon, J. Gunther, *Contravariant adaptation on structured matrix spaces*, Elsevier, Signal Processing Volume 82 , Issue 10, 2002, 1389-1410.

[3] C. Balanis, *Antenna Theory*, Wiley, 2005.

[4] W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry, Second Edition* , Academic Press 2002.

[5] J.-F. Cardoso, *Learning in Manifolds: The Case of Source Separation*, Proceedings, Ninth IEEE SP Workshop on Statistical Signal and Array Processing, 1998, 136-139.

[6] S. H. Friedberd, A. J. Insel, L. E. Spence, *Linear Algebra*, Prentice Hall, New York 2002.

[7] T.W. Hungerford, *Abstract Algebra an Introduction, Second Edition*, Brooks/Cole, 1997.

[8] M. W. Hirsch, S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, San Diego, 1974.

[9] P.J. Olver, *Applications of Lie Groups to Differential Equations, Second Edition*, Springer-Verlag, 1993.

[10] B. O'Neill, Elementary Differential Geometry, Revised 2nd Edition, Second Edition, Academic Press, 2006.

[11] A.Z. Petrov, Einstein Spaces, Pergamon Press 1969.

[12] W.R. Wade, *An Introduction to Analysis, Third Edition*, Prentice Hall, New York 2003.

[13] R.M. Wald, *General Relativity*, University Of Chicago Press, Chicago 1984.

Index