Tangent Vectors and Differential Forms

Geometrical Requirements for a Local Description

I. Motivations for New Definitions of Vectors

Let us take the usual electric field as our *a priori* "example" of what a **vector field** is. At every point of space, and also at every time, due to some given configuration of charges, there is an electric field vector, $\vec{E}(\vec{r})$. This vector of course has dimensions of Newtons/Coulomb, quite different from the units of distances, with units of meters. Therefore it must surely "live" in some diffeent "place" than the physical space, where distances are measured. As well, of course, we often need to add two distinct electric field vectors at some particular point, giving the result some new name, such as the "Total" electric field. Therefore, these electric field vectors must be elements of some vector space, where addition is allowed, and well-defined. Moreover, it should be clear that we must have such a vector space at each and every location, i.e., an entire vector space for each value of $\vec{r}$, and for each value of the time t. Therefore we need to generalize the more elementary, geometric notions of a vector, and a vector space, so as to allow these desires, preserving carefully of course the important properties that vector spaces have, namely that one may add vectors and multiply them by scalars, the result being simply new vectors.

There are two reasonably distinct natural ways to do this, which will be described below; we will refer to them as **tangent vectors** and **differential forms**. In either case the result will be an entire n-dimensional vector space adjoined to each point of the n-dimensional space, or more commonly, the "spacetime" that we wish to study. Such an object is then $2n$-dimensional, and quite difficult to "draw," or visualize in a literal sense; the best analogy I know is to imagine something like a porcupine, where the underlying spacetime is the body of the porcupine, each point of which has quills sticking out which we use to visualize the many different vector spaces we have. For some given spacetime, this entire collection of vector spaces is called a *vector bundle*. In addition to trying to understand the vectors, and the vector spaces, themselves, we will also want to understand how to compare vectors of the same kind corresponding to (at least) nearby points of the spacetime; this is accomplished at infinitesimally near points by an object called an *affine connection.* These notes will describe both sorts of vectors, the relationships between them, the role of the affine connection, and also the role of the metric, i.e., the function that tells us about lengths of vectors, if it exists. In order to do that it will also be necessary to say at least a little bit about how to characterize the spacetime itself, in the case where it is not an ordinary "flat" space.

II. Tangent Vectors to Curves

Since a very important use of vectors is to keep track of "directions," we will use the idea of **directional derivatives** along curves, evaluated at any given point, as our generalization for vectors on curved spacetimes. As a concrete example, consider the variations of the temperature in a room, with $T = T(x, y, z) \equiv T(\vec{r})$. We want to determine the change in the temperature as we follow along some curve whose tangent vector is given by $\vec{u}$. The method, from the usual vector-analysis class, is to describe this via

$$\vec{u} \cdot \nabla T(\vec{r}) \quad = \quad u^i \partial_i T(\vec{r}) \quad = \quad u^x \frac{\partial T}{\partial x} + u^y \frac{\partial T}{\partial y} + u^z \frac{\partial T}{\partial z} \quad = \left\{u^i \partial_i\right\} T(\vec{r}) \quad . \tag{1.1}$$

Evaluated at some particular, arbitrary point, P, with coordinates $\vec{r}$, we say that

this gives us the rate at which the temperature is changing as we move away from P in the direction $\vec{u}$.

The direction $\vec{u}$ is the tangent vector to some curve, through some portion of the room, which we can characterize as a neighborhood of some particular point P. We would like to be able to use it to answer questions not only about the temperature but also other quantities which vary throughout the room; it is the operator $\widetilde{u} \equiv u^i \partial_i$, the directional derivative in the direction $\vec{u}$ that allows this. Its components, u^i, with respect to the basis set $\{\partial_x, \partial_y, \partial_z\}$, are the same as the components of $\vec{u}$ with respect to the basis set $\{\hat{x}, \hat{y}, \hat{z}\}$. At the point P, it has numerical values and lies in the vector space—of linear differential operators acting on functions—over that point. At nearby points through which the curve passes, the same is true so that we can say that this actually determines a (local) vector field, since we know the values in some neighborhood of points. We will pick up this notion of first-order differential operators as our generalized notion of vectors. This only requires that we can associate with every point P coordinates in some differentiable way, and that we are allowed to consider (differentiable) functions for them to act on. In mathematical terms, spaces that satisfy requirements like this are usually referred to by the term *manifold*. We should not need to be very careful about our use of this; nonetheless, it is reasonable to at least write down somewhat carefully some of what is meant by these terms.

III. Manifolds, at the lowest-possible degree of precision

A manifold is a set of points such that every point has a neighbourhood that "looks like" some open set of some ordinary, flat, m-dimensional space, i.e., like $\mathbb{R}^m$, for some fixed integer m. The phrase "looks like" means that we may use, locally, a set of m coordinates just as we would in an ordinary flat space, i.e., a piece of $\mathbb{R}^m$, although we must surely allow the choice of coordinates to vary as we move over the manifold. It will be desirable to restrict this varying however so that the coordinates are continuous and invertible functions. At least some of the reasons for this are that we want several different sorts of things to "make sense" on this manifold:

1) curves on the manifold, paramterized by some range of real numbers,
2) functions, from the manifold to the real numbers,
3) mappings, between manifolds,
4) differential forms, i.e., total derivatives of functions,
5) tangent vectors, i.e., first-order differential operators that describe rates of change of functions,
6) surfaces in the manifold, paramterized by some sets of real numbers,
7) more complicated objects, perhaps made from products of two or more of the preceding.

In all these cases, we will want these objects to be "sufficiently continuous" to allow their differentiation, presumably as many times as we think appropriate. **Differentiation itself**, originally, is well-defined only for functions over a field, such as the real numbers, the complex numbers, multiple copies of the real numbers, i.e., $\mathbb{R}^m$, for some m, etc. It is for this reason that we must allow coordinates on a manifold, with some restrictions on continuity and differentiability. Therefore, we now introduce

the general notion of a 'smooth' manifold

Defn. 1. A *chart* for a manifold M is a set of **invertible** mappings, φ_i and associated open neighborhoods, $\{U_1, U_2, \ldots\}$, which together cover all of M. Each such mapping sends its neighborhood into (an open subset of) $\mathbb{R}^m$; i.e., we have

$$\begin{gathered}\forall P \in M\,,\ \exists \text{ within the chart at least one } U \subseteq M \text{ and } \varphi\colon U \to \mathbb{R}^m \\ \text{such that} \quad P \in U, \text{ and } \varphi(P) \equiv \big(\varphi^1(P), \varphi^2(P), \ldots, \varphi^m(P)\big) \in \mathbb{R}^m \quad .\end{gathered} \tag{2.1}$$

The quantities $\{\varphi^k(P)\}$, where k varies from 1 to m, are referred to as the *coordinates for* P relative to the choice of open set U.

We acquire differentiability for functions over our manifold by considering behavior in overlaps of coordinate neighborhoods. For an arbitrary point $P \in M$, consider two distinct U_i's, say U_1 and U_2, both of which contain P. Then, the mapping $\varphi_2 \circ {\varphi_1}^{-1}$ sends $\mathbb{R}^m$ to $\mathbb{R}^m$; we insist that it be differentiable arbitrarily often. [These functions are called *transition functions*.] There are actually different sorts of manifolds, having differing levels of continuity for the transition functions; in our case, **we will insist that all transition functions must be C^∞, and refer to such manifolds as "smooth."**

Defn. 2. A set of charts that contains the largest possible number of charts for a given manifold, M, consistent with the requirement that **all** transition functions be smooth, is called a (maximal) smooth *atlas* for M.

Defn. 3. Any space of points, M which admits a maximal smooth atlas, with a constant value of the dimension m, will be called a *smooth manifold*.

An obvious example to consider for a (2-dimensional) manifold is the sphere, S^2. The usual set of labels we use for the sphere, namely $\{\theta, \varphi\}$, are good labels, but have singularities, concerning invertibility, at both the north and south poles. Therefore, one choice of charts would be to use those coordinates in a neighborhood, say, between $-85°$ latitude and $+85°$ latitude. Then, to define a 'new' north pole in Quito, Ecuador, say, and define a second neighborhood, done with the same way, using the new $\{\theta_Q, \varphi_Q\}$—Q for Quito—treating Quito as the north pole.

On the other hand, this has quite an awkward set of transition functions. We will nonetheless sometimes use them in this way. However, it is also worthwhile to spend a little time looking at a somewhat simpler approach to a "good" set of charts, usually called stereographic projection for the sphere. This example is discussed in some detail in the first appendix to these notes, along with the transition functions for it.

III. Tangent Vector Fields

A (continuous) curve of points on a manifold is an important notion for us. We can think of it as being parametrized by some (real) parameter which varies continuously as the points vary on the manifold. Mathematically this is a mapping from a subset of the real numbers, say W, into our manifold:

$$\Gamma : W \subseteq \mathbb{R} \;\to M \;\Longrightarrow \forall\lambda \in W \subseteq \mathbb{R}\,,\; \Gamma(\lambda) \in M \quad . \tag{3.1}$$

Continuity means that the coordinates of the points are continuous functions of the parameter λ, i.e., for every allowed coordinate system the m different functions $x^i[\Gamma(\lambda)]$ must all be continuous. Note that normally we will not be so "formal" with the notation and will simply write $x^i(\lambda)$ for the functional dependence of the coordinates of the points on the curve.

Using the motivational paragraphs above, we will label the tangent vector to this curve as $\widetilde{u}_{|P} \equiv \frac{d}{d\lambda}_{|P}$, where λ is the choice that has been made for a parameter along that curve. In terms of coordinates on the manifold, this simply gives us the directional derivative operator:

$$\widetilde{u}_{|P} \equiv \frac{d}{d\lambda}\Big|_P = \left(\frac{dx^i[\Gamma(\lambda)]}{d\lambda}\right)\Big|_P \frac{\partial}{\partial x^i} \equiv u^i(P)\,\partial_{x^i} \equiv u^i(P)\,\partial_i\ , \quad \text{for } P \in M\,. \tag{3.2}$$

The $x^i[\Gamma(\lambda)]$ are simply a choice for coordinates of the points along the curve Γ, and we therefore see that the tangent vector has components that are the derivative of those coordinates with respect to that parameter. (The "over-tilde" will be our standard notation to indicate that the object in question is indeed a tangent vector.)

Since these tangent vectors are differential operators they act on functions defined over the manifold. It is therefore appropriate to backup for just a moment and give a definition of smooth functions over a manifold.

Defn. 5. A continuous function over M is one where the coordinate form of that function is continuous for any allowed choice of coordinates. That form is of course a map $\mathbb{R}^m$ to $\mathbb{R}$, which we may think of as some function $f(x^i)$.

For our smooth manifolds we prefer to restrict consideration to smooth functions, i.e., to those which are C^∞ functions.

we denote the set of all C^∞ functions over the manifold by the symbol $\boldsymbol{\mathfrak{F}}$.

At a single point, $P \in M$, a vector $\widetilde{u}_{|P}$ operates on a function, $f \in \boldsymbol{\mathfrak{F}}$, defined in some neighborhood of P, in the expected way. The result is a number which is the value of the directional derivative, in the direction $\widetilde{u}$, at the point P. Therefore the vector at the point P maps functions into numbers. However, if we do this in some neighborhood of P where this vector is defined, then a different number will be generated at each point, giving us therefore a new function in that neighborhood. Therefore, referring to the *tangent vector field*, defined at all neighboring points, this vector field maps functions into other functions, namely their derivatives in the direction $\widetilde{u}$:

$$\begin{aligned} &\text{at any single point, } \widetilde{u}_{|P} : \boldsymbol{\mathfrak{F}}\Big|_{P\in U} \to \mathbb{R} \, , \\ &\text{in a neighborhood of a point, } \widetilde{u} : \boldsymbol{\mathfrak{F}} \to \boldsymbol{\mathfrak{F}} \quad , \\ &\widetilde{u}[f] = u^j(\lambda)\,\frac{\partial}{\partial x^j} f(x^i) = \frac{dx^j(\lambda)}{d\lambda}\,\frac{\partial}{\partial x^j} f(x^i) \, . \end{aligned} \tag{3.3}$$

If one wants to be very pedantic about the details of the mapping, one could write the same thing in very much more detail:

$$\begin{aligned} \widetilde{u}[f] &= \frac{d}{d\lambda} f[\Gamma(\lambda)] = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\{f[\Gamma(\lambda+\epsilon)] - f[\Gamma(\lambda)]\} = \frac{d}{d\lambda}(f\circ\Gamma) \\ &= \frac{d}{d\lambda}(f\circ\varphi^{-1}\circ\varphi\circ\Gamma) = \left(\frac{\partial}{\partial x^i}(f\circ\varphi^{-1})\right)\left\{\frac{d}{d\lambda}(\varphi\circ\Gamma)^i\right\} = \frac{dx^i[\Gamma(\lambda)]}{d\lambda}\,\frac{\partial}{\partial x^i}(f\circ\varphi^{-1})(x^j) \, , \end{aligned} \tag{3.4}$$

where the set $\{x^j\}$ constitute the coordinate choice for a chart labelled by φ, in some neighborhood of the point P, the $x^j[\Gamma(\lambda)]$ are the coordinate presentation of the curve, so that they constitute a set of maps from $W \subseteq \mathbb{R} \to \mathbb{R}^m$ and the $(f \circ \varphi^{-1})(x^j)$ constitute the coordinate presentation of the function f, i.e., the $f \circ \varphi^{-1}$ constitute a map from $\mathbb{R}^m \to \mathbb{R}$.

Abstractly, we summarize the situation by saying that the tangent vectors **at each point**, P, **of the manifold** satisfy the following properties:

1. $\widetilde{u}_{|P} : \boldsymbol{\mathfrak{F}}\Big|_U \to \mathbb{R}\,, \qquad P \in U \subseteq M,$
2. $\widetilde{u}_{|P}(f + \alpha g) = \widetilde{u}_{|P}(f) + \alpha\,\widetilde{u}_{|P}(g)$, for α a constant, —**linearity**
3. $\widetilde{u}_{|P}(fg) = f(P)\,\widetilde{u}_{|P}(g) + \widetilde{u}_{|P}(f)\,g(P)$ —**derivation property**.

Objects with these general properties are often referred to as derivations. It should be clear that these properties are maintained under addition and multiplication by scalars; therefore we may say that our tangent vectors form a vector space of derivations (of functions). We will refer to this vector space of tangent vectors at the point P by the symbol $\mathfrak{T}_P$, and the entirety of all these vector spaces as the *tangent bundle*, $\mathfrak{T} \equiv \mathfrak{T}^1$. Our construction shows us that for each of the vector spaces $\mathfrak{T}_P$, one may treat the partial derivative operators at P, namely $\{\partial_{i_P}$, as a particular choice of basis for that vector space.

Suppose, now, that we consider what happens when one makes a different choice of coordinates for the neighborhood; for instance we consider a transformation from coordinates x^i to some other choice y^s, where we may write either $y^s = y^s(x^i)$ or vice versa, since the transformation is certainly required to be invertible. Then "advanced calculus" tells us that

$$\begin{aligned} \partial_i \equiv \partial_{x^i} &= \left(\frac{\partial y^s}{\partial x^i}\right)\partial_{y^s} \equiv Y_i^{\ s}\,\partial_{y^s}\ , \\ \frac{dx^i}{d\lambda} &= \left(\frac{\partial x^i}{\partial y^s}\right)\frac{dy^s}{d\lambda} \equiv X^i_{\ s}\frac{dy^s}{d\lambda}\ . \end{aligned} \tag{3.5}$$

As the two matrices of partial derivatives are inverse to each other, the form of the tangent vector itself is unchanged by such a change in one's choice of coordinates—as surely seems appropriate. However, it is also clear that the **components** of the vector are changed when one goes from one choice of coordinates to another. Denoting the components with respect to the x^i-basis by u^j and the (new) components, with respect to the (new) y^s-basis, by ${u'}^s$, we have the transformation equation

$$u^i = X^i_s \, {u'}^s \ . \tag{3.6}$$

Since Eqs. (3.5) show us that the components of a tangent vector transform with a matrix that is inverse, or opposite, to the matrix that transforms the basis vectors themselves, the transformation of the components of the tangent vector is said to be *contravariant*. Therefore tangent vectors are sometimes referred to as "contravariant vectors," even though, truthfully, it is their components that transform in this way.

IV. Differential Forms: an Alternative Bundle of Vector Spaces

Every vector space has associated with it a *dual space* of quantities that map the vectors into whatever field of numbers over which the vectors are defined—in our case, either the real or complex numbers. In addition, these mappings are required to preserve the structure of the original vector space; i.e., the mappings should be linear and simply let scalars pass through. More precisely, if ω is such a mapping, $\widetilde{u}$, $\widetilde{v}$ are arbitrary (tangent) vectors and a is a scalar, then we require that

$$\omega(\widetilde{u} + a\,\widetilde{v}) = \omega(\widetilde{u}) + a\,\omega(\widetilde{v})\ . \tag{4.1}$$

Since these are mappings on vectors, it is straightforward to show that they are also vectors. Therefore vector spaces always come in pairs. However, in many of the simpler cases, there is a way to identify the two spaces, so that it is often difficult to see that there are actually two. An example is given by the bra's and ket's in the Hilbert space for quantum mechanics. A second, well-known example is given by the *polar vectors* and *axial vectors* in classical mechanics, say. The ordinary linear momentum is a polar vector, while the angular momentum is a cross product of two polar vectors, and therefore is an axial vector. In classical mechanics vectors are often characterized by their behavior under rotations and translations. For the usual rotations, the two sorts of vectors behave exactly the same; however, for a parity transformation, polar vectors change sign while axial vectors do not. Therefore it is possible to distinguish these two sorts of vectors, but it does take a little effort. It is the generalization of this particular distinction that we are now beginning to discuss.

We want now to describe the dual space for tangent vectors over a manifold. We begin by considering a function, $f \in \mathfrak{F}$, defined in a neighborhood of some point $P \in M$. We will define the **differential form**, df, associated with $f \in \mathfrak{F}$, as a member of the vector space dual to $\mathfrak{T}_P$ by giving its action on tangent vectors. We do this first at a single point,

$$df_{|P}(\widetilde{u}_{|P}) \equiv \widetilde{u}_{|P}(f) = (u^i\,\partial_{x^i} f)_{|P} \equiv (u^i\,f_{,i})_{|P} \quad , \tag{4.2}$$

which does indeed satisfy the requirements of linearity, since $\widetilde{u}$ does. The calculation takes a vector and gives a number. However by doing this in each of the vector spaces over the points of some neighborhood containing P we obtain a function from our tangent vector. Therefore, over a neighborhood we may think of df as mapping tangent vector fields into functions over the manifold,

$$\text{i.e., } df : \mathfrak{T} \to \mathfrak{F}.$$

We can now proceed further along the path of determining the entirety of the dual space by next using our notions from "advanced calculus," concerning *differentials.* Certainly we know that $df = f_x\, dx + f_y\, dy + \ldots$. This shows us that the set $\{dx, dy, \ldots\}$ could be chosen as a basis for this vector space. Using linearity the previous equation then tells us that

$$u^i\, f_{,i} \equiv u^i\, (\partial_i)(f) = \widetilde{u}(f) \equiv df(\widetilde{u}) = f_{,i}\, dx^i(\widetilde{u}) = f_{,i}\, dx^i(u^j \partial_j) = (f_{,i} u^j)\; dx^i(\partial_j)\ . \tag{4.3}$$

First this shows us the action of the basis-quantities, $\{dx^i\}$ on vectors, which is an altogether appropriate action since the x^i are just a rather special set of functions defined over our neighborhood. Taking $\{\partial_j\}$ as a **basis for $\mathfrak{T}$**, we see that $\{dx^i\}$ may easily be taken as a basis for this vector space of differential forms, **and** that it is a reciprocal basis, i.e., it satisfies

$$dx^i(\partial_j) = \delta^i_j \quad . \tag{4.4}$$

This concept allows us to easily generalize the notion of the differential of a function to define the entire vector space, $\Lambda \equiv \Lambda^1$, of all 1-forms. We say that a 1-form is simply any (finite) sum of products of functions and the differentials of functions, locally-defined in each vector space over the points of some neighborhood of any point. For now we will treat the differentials of the coordinates on the manifold as the basis 1-forms for Λ^1. As element of this dual space are supposed to have an action on vectors, mapping them to functions—which become scalar numbers at each point on the manifold, but, of course, in a way which varies "smoothly" from one point to the next, we may use our statements about the reciprocal basis to define the general action of the dual space on the tangent vector space:

- **Dual Action:** Let $\underset{\sim}{\alpha} \in \Lambda^1$ and $\widetilde{u} \in \boldsymbol{\mathfrak{T}}$ be written in terms of their (reciprocal) basis sets as

$$\underset{\sim}{\alpha} = \alpha_i \, dx^i \quad , \qquad \widetilde{u} = u^i \, \partial_i \quad . \tag{4.5a}$$

The mapping $\underset{\sim}{\alpha} : \boldsymbol{\mathfrak{T}} \to \boldsymbol{\mathfrak{F}}$ is defined by

$$\underset{\sim}{\alpha}(\widetilde{u}) = \alpha_i \, u^i \quad , \tag{4.5b}$$

$$\text{so that} \qquad df(\partial_j) = f_{,j} \quad \text{and} \quad \underset{\sim}{\alpha}(\partial_j) = \alpha_j \;\; \text{along with} \;\; dx^i(\widetilde{u}) = u^i \quad . \tag{4.5c}$$

- **Geometrical Description of Differential Forms**

We created tangent vectors via an altogether reasonable attempt to generalize the very intuitive notion of a vector as a directed line segment. What, then, is a reasonable geometrical picture of the elements of our vector space of differential forms? Although I do not want to belabor the creation of such a geometrical picture for 1-forms, there is a standard approach to what it might be. (Much more detail, with many pictures, is given in MTW.)

Recall the usual geometrical notion of the gradient of a function, such as the temperature in a room. The ordinary 3-dimensional vector, $\nabla T(\vec{r})$ defines a vector field throughout the room, which we usually describe by saying that it is the "direction of **greatest change**" of T, at the particular point with coordinates $\vec{r}$. However, we also talk about a somewhat dual construction, namely the surfaces of constant value of the temperature, T, to which the direction ∇T is perpendicular. These are surfaces, much like equipotentials for electrostatics, where the temperature doesn't change as you move from one point to another on them. It is altogether plausible that the surface on which the function does NOT change isthe one that is **perpendicular** to the direction of its greatest change!

We therefore take the analytic idea that *df* is the generalization of *the gradient* of f, i.e., ∇f. As in the case of tangent vectors and directional derivatives, this comparison is reasonable since the two quantities have the same components; only the basis vectors look different. Again, the new basis vectors—for *df*—are defined locally, at each single point P on the manifold. Then we may say that geometrically it corresponds to a local view of the surfaces of constant values for f. Since they are *locally-defined*, over an m-dimensional manifold, surfaces of constant values for some function, say the temperature T, are $m-1$-dimensional surfaces—so-called hypersurfaces since they are only one less dimension than the entire space. The simplest 1-form dT is an algebraic representation of the set of (hyper)-surfaces of constant value for T, at the point in question.

Since the ideal of an infinitesimal, like dT, is surely a local idea, defined, within a neighborhood, but at each single point, these surfaces must live within their own vector space, the vector space of differential forms, with one such space attached at each point of our underlying manifold, Then, in particular, the hypersurfaces dx^i correspond to constant values for the coordinate function x^i, so that they are those hypersurfaces in which all the **other** coordinates are allowed to take all possible values, the value of x^i itself being fixed to one particular value. As a brief example, we look at an ordinary 3-dimensional space and the usual coordinates, $\{x, y, z\}$, so that ∂_x corresponds to the tangent vector to the x-axis, while then the 1-form dx corresponds to the set of 2-surfaces we would describe by saying that they are all parallel to the y, z-plane, for different, fixed values of x.

V. Extension of Differential Forms to the entire Grassmann Algebra

As it turns out, differential forms have many very useful and important properties, quite sufficient to justify the effort required to learn to manipulate and understand them. The vector space can be turned into a *differential algebra* by adding a **product** and a special differential operator into this space; the result is called a *Grassmann algebra*, following Grassmann, who was a high-school teacher in Germany in the 1870's.

- **The Exterior Product, and p-forms**

We define a (**skew-symmetric**) (tensor) product for differential forms, which we denote with the symbol $\wedge$ between the two 1-forms being multiplied together; it is a straightforward generalization of the usual "cross-product" that ordinary 3-dimensional vectors possess. Specifically, the product is such that changing the order changes the sign:

$$dx \wedge dy = -dy \wedge dx \quad , \tag{5.1}$$

so that an expression like $dx \wedge dx$ is just identically zero! The result of this product is no longer a 1-form, i.e., it is no longer a linear combination of the differentials of the coordinates of M. Therefore, we generalize the notion of 1-forms to allow for some additional vector spaces, whose elements we refer to as *p-forms*, where p=$1, \ldots, m$. One could, for instance, first define 2-forms as the vector space of all linear combinations of multiples of the wedge product of 1-forms, requiring this product to be linear and associative over addition. One could then generalize to products of more than two at a time, again requiring associativity. We may also give an approach where we use a basis set of 1-forms to create basis sets for each of these spaces of products of 1-forms, p at a time.

- **p-forms:** For each positive integer, $p \leq m$, we may then define the vector space Λ^p as the set of all possible (finite) linear combinations of multiples, via functions, of its basis set, which is the set of all linearly-independent, non-zero "wedge-products" of p of the $\{dx^i\}$ at once.

Since, for instance, $dx \wedge dy = -dy \wedge dx$, the number of linearly-independent p-forms is reduced by the requirements of skew symmetry. In general the dimension of the space $\Lambda^p M$ of p-forms is just $\binom{m}{p}$. When there are m dimensions to M, then one may have spaces of 1-forms, 2-forms, 3-forms, etc., up to and including m-forms, but no higher. This is because one cannot make objects skew symmetric in the interchange of two adjacent entries in more than m things at once if you only have m things to play with! In addition, notice that there is only one independent dimension to the space of m-forms, e.g., $dx^1 \wedge dx^2 \wedge \ldots \wedge dx^m$. Some particular choice of basis for that 1 dimension is usually referred to as *the volume form* for the space. If we also decide that λ^0 is to be defined as the same as the space of functions, $\mathfrak{F}$, which also has dimension 1, we can see a duality in these various vector spaces, at least in dimensions. For example, $\binom{m}{m-1} = \binom{m}{1}$, and also $\binom{m}{m-2} = \binom{m}{2}$, etc. This is called Hodge duality, and we will return to it later.

Example: If the underlying space is 3-dimensional—just the usual space, $\mathbb{R}^3$, with coordinates $\{x, y, z\}$, then we take $\{dx, dy, dz\}$ as a basis for the space of 1-forms, then the space of 2-forms is the vector space with basis $\{dx \wedge dy, dy \wedge dz, dz \wedge dx\}$, and the space of 3-forms has basis $\{dx \wedge dy \wedge dz\}$. (We don't count both $dx \wedge dy$ and $dy \wedge dx$, since they are not linearly independent; one is just the negative of the other.) As it happens, the space of 2-forms is the same dimension as the space of 1-forms. If we wanted to, we could characterize all this by saying that $\Lambda^2(\mathbb{R}^3)$ is *spanned* by the set $\{dx \wedge dy, dy \wedge dz, dz \wedge dx\}$. This (interesting) equality occurs only for 3 dimensions, and is the source of the existence of the "cross-product" only in 3 dimensions. On the other hand, in 4 dimensions, the space of 1-forms of course has 4 dimensions, being spanned by $\{dx, dy, dz, dt\}$, while the space of 2-forms already has 6 dimensions, since it is spanned by $\{dx \wedge dy, dy \wedge dz, dz \wedge dt, dt \wedge dx, dy \wedge dt, dz \wedge dx\}$.

- **Action of p-forms, on Tangent Vectors**

As a finite linear combination of wedge products of p 1-forms at a time, a p-form may be "fed" p distinct tangent vectors, which will result in the production of a function, i.e., a number at each point. At a more sophisticated level, one could only feed some smaller number of vectors to the p-form, resulting, then, in a form of some lesser order. The first thing to notice is that the skew-symmetry of a p-form arranges it so that it operates on its vectors in a skew-symmetric way. More precisely, **if the p vectors fed to the p-form are not all linearly independent, then the resulting number will simply be zero!**

Of especial interest in physics are 2-forms; therefore, we want to establish, as a formula, how one determines the action of a 2-form on a pair of vectors. A reasonable approach for the action of the 2-form, $\underset{\sim}{\alpha}_1 \wedge \underset{\sim}{\alpha}_2$ on a pair of vectors, $\widetilde{u}_1$,$\widetilde{u}_2 \in \mathcal{T}$, should be given by

$$(\underset{\sim}{\alpha}_1 \wedge \underset{\sim}{\alpha}_2)(\widetilde{u}_1, \widetilde{u}_2) = \begin{vmatrix} \underset{\sim}{\alpha}_1(\widetilde{u}_1) & \underset{\sim}{\alpha}_1(\widetilde{u}_2) \\ \underset{\sim}{\alpha}_2(\widetilde{u}_1) & \underset{\sim}{\alpha}_2(\widetilde{u}_2) \end{vmatrix} = \text{determinant}\left((\underset{\sim}{\alpha}_i(\widetilde{u}_j))\right) \quad . \tag{5.2}$$

Using the structure of a determinant arranges it so that the skew-symmetry is made automatic, as desired. (There are in fact other normalizations in the literature: sometimes a factor of 2! is used to multiply the definition above.) If, now, the 2-form given is more general, the fact that $\underset{\sim}{\omega}^i \wedge \underset{\sim}{\omega}^j$ satisfies the form given in Eq. (5.2) and that $\underset{\sim}{\omega}^i(\widetilde{u}) = u^i$ allow us to write

$$\underset{\sim}{\beta} \equiv \tfrac{1}{2}\beta_{ij}\underset{\sim}{\omega}^i \wedge \underset{\sim}{\omega}^j \implies \underset{\sim}{\beta}(\widetilde{u}, \widetilde{v}) = \beta_{ij}\, u^i v^j \quad . \tag{5.3}$$

For a general p-form, then we acquire the similar formula:

$$\underset{\sim}{\psi} \equiv \frac{1}{p!}\psi_{i_1\ldots i_p}\,\underset{\sim}{\omega}^{i_1}\wedge\ldots\wedge\underset{\sim}{\omega}^{i_p} \implies \underset{\sim}{\psi}(\widetilde{u},\ldots\widetilde{w}) = \psi_{i_1\ldots i_p}\,u^{i_1}\ldots w^{i_p} \quad . \tag{5.4}$$

Notice that I have adopted the notational convention of putting an "under-tilde" on any p-form, analogous to the method that I use to denote 1-forms. In general, we won't bother to denote the value of p, hoping that it's reasonably obvious from context.

V. The Exterior Differential, and p-forms

Since the space of 1-forms begins with the differentials of functions, one should be able to differentiate other things as well. We have now decided that it is reasonable to look at the space of functions as a "zero-th" order space of forms, so that it seems that differentiation takes 0-forms into 1-forms. Therefore, we now want to generalize the notion of differentiation so that it acts on any p-form, with the result being a $(p+1)$-form. Such a generalization should maintain the usual "product rule" for derivatives while also maintaining and creating the skew-symmetry of p-forms which we have just endeavored to create. More precisely, we would like it to have the following sorts of properties:

1. **the exterior derivative** d **is a derivation** so that it handles its arguments according to (some version of) the product rule we know from calculus, i.e., for (scalar) functions, $f, g \in \Lambda^0 \equiv \boldsymbol{\mathfrak{F}}$, we have $d(fg) = f\,dg + (df)\,g$,
2. to preserve skew-symmetry of p-forms, it anticommutes with 1-forms, since our product is anti-symmetric.

This motivation is sufficient to allow us to define the **exterior derivative** as a mapping on any p-form, acting something like a derivative.

$$d\colon \Lambda^p M \longrightarrow \Lambda^{p+1} M \quad \text{such that}$$

$$\begin{cases} df = f_{,i}\,dx^i \equiv \frac{df}{dx^i}\,dx^i & , \quad \text{for functions } f \in \Lambda^0 M, \\ d(\phi_i\,dx^i) = d\phi_i \wedge dx^i \in \Lambda^2 M & , \quad \text{for 1-forms } \phi = \phi_i\,dx^i \in \Lambda^1 M, \\ d(\underset{\sim}{\alpha}\wedge\underset{\sim}{\beta}) = (d\underset{\sim}{\alpha})\wedge\underset{\sim}{\beta} - \underset{\sim}{\alpha}\wedge(d\underset{\sim}{\beta}) & , \quad \text{for } \underset{\sim}{\alpha},\ \underset{\sim}{\beta} \in \Lambda^1 M. \end{cases} \tag{5.5}$$

These properties are sufficient to allow one to prove the following more general form, regarding commutativity, for its action on arbitrary p- and q-forms:

$$\text{for} \quad \underset{\sim}{\lambda} \in \Lambda^p M, \quad \underset{\sim}{\nu} \in \Lambda^q M, \qquad d(\underset{\sim}{\lambda} \wedge \underset{\sim}{\nu}) = (d\underset{\sim}{\lambda}) \wedge \underset{\sim}{\nu} + (-1)^p \underset{\sim}{\lambda} \wedge (d\underset{\sim}{\nu}) \quad . \tag{5.6}$$

Examples:

0. Suppose, now, that $f = f(x, y, z)$ is a function over $\mathbb{R}^3$, then $df \wedge dy$ is a 2-form, which, in the usual basis, can be thought of as having 2 non-zero components, i.e.,

$$\text{over } \mathbb{R}^3, \qquad df \wedge dy = \frac{df}{dx}\, dx \wedge dy + \frac{df}{dz}\, dz \wedge dy \quad \equiv f_x\, dx \wedge dy - f_z\, dy \wedge dz \quad .$$

Notice the following comparisons to ordinary 3-dimensional vector analysis:

1. We may think of the usual 3-dimensional vector, $\vec{v}$, as a 1-form, $\underset{\sim}{\zeta}$, over a space of 3 variables, by identifying the components of the two; then the 2-form $d\,\underset{\sim}{\zeta}$ has 3 components which are just the usual components of the *curl* of $\vec{v}$, i.e., $\nabla \times \vec{v}$;
2. Similarly we may identify the 3 components of the 2-form $\underset{\sim}{\beta}$, over a space of 3 variables, with the components of a vector $\vec{w}$, then the 3-form $d\,\underset{\sim}{\beta}$ has just one component, which is just the *divergence* of $\vec{w}$, i.e.,$\nabla \cdot \vec{w}$.
3. In any number of dimensions, a 2-form is specified by components which (a) have two indices, and (b) are skew-symmetric with respect to the interchange of those two indices. Therefore, we may always use matrix notation to display (or present) the components of a 2-form; in general that matrix will be skew-symmetric. A common example is the electromagnetic field tensor in 4 dimensions, which we will discuss in more detail later.

- **Poincaré's Lemma** Because of the anti-symmetry of the wedge product, we get some slightly unexpected properties of the exterior derivative operator, which actually give it much more power than one might have expected. Firstly, notice that the skew symmetry of wedge products and the fact that partial derivatives commute, i.e., are symmetric with respect to order, make it easy to see that, for any function, f, and indeed for any p-form, $\underset{\sim}{\psi}$, we have

$$d^2 \equiv d(d\,f) = 0\ , \quad d(d\underset{\sim}{\psi}) = 0\ . \tag{5.7}$$

The statement of Poincaré's Lemma given below is a generalization of the (rather more well-known) statement from ordinary 3-dimensional vector analysis that

div curl $\vec{v} \equiv \nabla \cdot \nabla \times \vec{v} = 0$, i.e., identifying $\vec{v}$ with the 1-form $\underset{\sim}{\zeta}$, and utilizing the previous example, we have the equivalence of this with $d\,d\,\underset{\sim}{\zeta} = 0$.

In order to state Poincare's Lemma, it is convenient to first have the following two definitions:

Closed Forms: A p-form, $\underset{\sim}{\alpha} \in \Lambda^p$ is closed $\iff$ $d\,\underset{\sim}{\alpha} = 0 \in \Lambda^{p+1}$;

Exact Forms: A p-form $\underset{\sim}{\beta}$ in Λ^p is exact $\iff$ there is some (p-1)-form, $\underset{\sim}{\gamma} \in \Lambda^{p-1}$ such that $\underset{\sim}{\beta} = d\,\underset{\sim}{\gamma}$.

Poincaré's Lemma tells us that for "nice" regions of a set of variables, the two sorts of forms are equivalent; i.e.,

an exact form is closed, **for all values of the variables**,

a closed form is exact, **inside some sort of appropriate region**.

The first half is easy to prove by a simple computation using skew-symmetry and symmetry, as suggested above. That the converse is also "true" (locally) is a moderately difficult exercise in advanced calculus; nonetheless it is true for proper sorts of neighborhoods. Over large regions of space, the converse may well not be true; for just which forms it is true is the subject of the mathematical theory called cohomology. That theory is used to study the global properties of manifolds.

Notice the following comparisons to ordinary 3-dimensional vector analysis, which may help us to keep the content in mind:

Example: If we think of the usual 3-dimensional vector in terms of a 1-form over a space of 3 variables, as in the examples just above, we can notice that Poincaré's Lemma has several special cases already known to us:

1. If $\nabla \times \vec{v} = 0$, then there exists a (scalar) potential V such that $\vec{v} = \nabla V$ becomes, identifying $\vec{v}$ and $\underset{\sim}{\zeta}$, that $d\,\underset{\sim}{\zeta} = 0$ implies the existence of a 0-form $f \in \Lambda^0 M$ such that $\underset{\sim}{\zeta} = df$;
2. If $\nabla \cdot \vec{w} = 0$, then there exists a vector potential $\vec{h}$ such that $\vec{w} = \nabla \times \vec{h}$ becomes, identifying $\vec{w}$ and the 2-form $\underset{\sim}{\beta}$, that $d\,\underset{\sim}{\beta} = 0$ implies the existence of a 1-form $\underset{\sim}{\eta}$ such that $\underset{\sim}{\beta} = d\,\underset{\sim}{\eta}$.

VI. Choices of Bases for our Vector Spaces

While we have so far only discussed basis sets, for $\mathcal{T}$, and Λ^1, made directly from the coordinates, the discussion of transformations between basis sets, above, could easily lead one to consider other sets of bases by taking linear combinations of the old ones. Quite often there may be very good **physical reasons** for making some particular choice of basis vectors for either $\mathcal{T}$ or Λ^1. In general, we may choose any set of basis vectors for $\mathcal{T}^1$ that are linearly independent and sufficiently many, i.e., m of them. Under those circumstances, let us agree to refer to a general set of basis vectors, for $\mathcal{T}^1$, by the standard symbols $\{\widetilde{e}_a\}_{a=1}^m$, while for Λ^1 we refer to an associated set of basis 1-forms by the symbols $\{\underset{\sim}{\omega}^b\}_{b=1}^m$, the association between the two being determined by the requirement that they remain dual to one another; i.e., they should satisfy an appropriate version of duality, analogous to Eq. (4.3):

$$\text{dual basis sets are defined by} \qquad \underset{\sim}{\omega}^a(\widetilde{e}_b) = \delta^a_b \quad . \tag{6.1}$$

Orthonormal Choices for Bases

When dealing with ordinary physics problems in a very simple space, such as say, this room, we traditionally use orthonormal sets of basis vectors; among other things, it seems that our intuition about the behavior of components of vectors is normally built over the assumption that the basis sets are orthonormal. However, in general on a more sophisticated surface, it is impossible to require **both** that the basis sets be made as partial derivatives with respect to a set of coordinates, **and** that they are orthonormal with respect to some notion of a choice of metric. Many years ago it was thought that the requirement that the basis set be derivatives with respect to a set of coordinates was the most important of these requirements; such basis sets are referred to as *holonomic bases*. Therefore, the difference between covariant and contravariant components was important, and one needed to develop accurate intuitions concerning the two. However, over the last 30 years, physicists have realized that it is not actually more difficult but, instead, somewhat easier to use *non-holonomic* sets of basis vectors. They make it much simpler to provide quick physical interpretations of the results of calculations! Therefore, we need to begin the introduction of such sets, beginning by giving a very simple example, and then criteria for knowing which kind one has.

Example of a non-Holonomic Basis originating with polar coordinates

We consider flat space, in 3-dimensions, but insist on using polar coordinates. Then, using coordinates $\{r, \theta, \varphi\}$, a plausible holonomic basis set for the tangent vectors would be $\{\partial_r, \partial_\theta, \partial_\varphi\}$. However, notice that these basis vectors do not even all have the same dimensions! Therefore, a set of basis vectors for the tangent space that is physically more consistent is given by the following:

$$\{\widetilde{e}_r \equiv \partial_r,\, \widetilde{e}_\theta \equiv \frac{1}{r}\partial_\theta,\, \widetilde{e}_\varphi \equiv \frac{1}{r\sin\theta}\partial_\varphi\} \quad \longleftarrow \quad \text{a non-holonomic basis set for } \boldsymbol{\mathcal{T}} \text{ over } \mathbb{R}^3. \tag{6.2}$$

All the vectors above have the same dimensions, and, as it turns out, for the usual choice of a metric on $\mathbb{R}^3$, they are orthonormal, so that there is no difference between the covariant and contravariant components of a vector, thereby giving them the usual physical interpretation one has when dealing with $\{\hat{r}, \hat{\theta}, \hat{\varphi}\}$.

Criteria for distinguishing holonomicity

The next question that should arise is how one knows, given a particular basis set for $\boldsymbol{\mathcal{T}}$, whether or not it is holonomic. After all, if one re-writes the holonomic set $\{\partial_r, \partial_\theta, \partial_\varphi\}$ in terms of the coordinates $\{x, y, z\}$, then it really will not appear to be holonomic. Therefore, one needs a sure test, which is given by the fact that partial derivatives with respect to the coordinates of a single coordinate system **always commute** one with another. Therefore, in general, for an arbitrary basis set $\{\widetilde{e}_i\}$ of $\boldsymbol{\mathcal{T}}$, we define the commutation coefficients:

$$[\widetilde{e}_i, \widetilde{e}_j] \equiv C_{ij}{}^k\, \widetilde{e}_k \quad \text{where} \quad C_{ij}{}^k = -C_{ji}{}^k \quad . \tag{6.3}$$

Theorem: The commutation coefficients vanish $\Longleftrightarrow$ the basis set is holonomic, with respect to some system of coordinates.

For our example, we find that

$$
\begin{aligned}
[\widetilde{e}_r\,,\widetilde{e}_\theta] = [\partial_r\,,\frac{1}{r}\partial_\theta] = -\frac{1}{r^2}\partial_\theta = -\frac{1}{r}\{\frac{1}{r}\partial_\theta\} = -\frac{1}{r}\widetilde{e}_\theta\,, \quad &\Rightarrow C_{r\theta}{}^{\theta} = -\frac{1}{r}\,,\\
[\widetilde{e}_r\,,\widetilde{e}_\varphi] = [\partial_r\,,\frac{1}{r\sin\theta}\partial_\varphi] = -\frac{1}{r^2\sin\theta}\partial_\varphi = -\frac{1}{r}\{\frac{1}{r\sin\theta}\partial_\varphi\} = -\frac{1}{r}\widetilde{e}_\varphi\,, \quad &\Rightarrow C_{r\varphi}{}^{\varphi} = -\frac{1}{r}\,,\\
[\widetilde{e}_\theta\,,\widetilde{e}_\varphi] = [\frac{1}{r}\partial_\theta\,,\frac{1}{r\sin\theta}\partial_\varphi] = -\frac{\cot\theta}{r}\frac{1}{r\sin\theta}\partial_\varphi = -\frac{\cot\theta}{r}\widetilde{e}_\varphi\,, \quad &\Rightarrow C_{\theta\varphi}{}^{\varphi} = -\frac{\cot\theta}{r}
\end{aligned}
$$

As well, we will later find that the commutation coefficients tell us some useful and interesting things about the structure of the space on which one is studying physics.

A related question is then the structure, or behavior, of a basis for the space of 1-forms, Λ^1. One would reasonably expect that the basis for 1-forms that is dual to a non-holonomic basis for tangent vectors would also be non-holonomic. What would that mean? Certainly, the basis sets we have so far been using for $\Lambda^1 M$ have been the differentials of some coordinate set. Therefore, we expect that a non-holonomic basis would correspond to linear combinations of differentials, which would not be *closed*, i.e., would **not** be differentials of **any** possible system of coordinates. Therefore, a reasonable test for whether a basis set for 1-forms is holonomic would be to calculate the set of their differentials, i.e., $\{d\underset{\sim}{\omega}^a\}$. Such a set would of course be 2-forms, and therefore must be expressible as linear combinations of wedge product of the basis set themselves. As it turns out, these combinations are related to the commutation coefficients already discussed:

$$d\underset{\sim}{\omega}^a \equiv -\tfrac{1}{2}C_{bc}{}^{a}\,\underset{\sim}{\omega}^b\wedge\underset{\sim}{\omega}^c \quad . \tag{6.4}$$

Notice that the set of non-holonomic basis forms for $\mathbb{R}^3$ dual to the polar coordinate ones given in Eq. (6.3) is

$$\{\underset{\sim}{\omega}^r \equiv dr,\ \underset{\sim}{\omega}^\theta \equiv r\,d\theta,\ \underset{\sim}{\omega}^\varphi \equiv r\sin\theta\,d\varphi\} \quad \longleftarrow \text{ a non-holonomic basis set for } \Lambda^1 \text{ over } \mathbb{R}^3. \tag{6.5}$$

One calculates that

$$
\begin{aligned}
d\underset{\approx}{\omega}^r &= d\,dr = 0 \Rightarrow {C_{ij}}^r = 0\,, \\
d\underset{\approx}{\omega}^\theta &= d\,(r\,d\theta) = dr\wedge d\theta = \frac{1}{r}dr\wedge r\,d\theta = \frac{1}{r}\underset{\approx}{\omega}^r\wedge\underset{\approx}{\omega}^\theta \\
&= \frac{1}{r}\left\{\tfrac{1}{2}[\underset{\approx}{\omega}^r\wedge\underset{\approx}{\omega}^\theta - \underset{\approx}{\omega}^\theta\wedge\underset{\approx}{\omega}^r]\right\}\ , \quad \Rightarrow {C_{r\theta}}^\theta = -\frac{1}{r}\,, \\
d\underset{\approx}{\omega}^\varphi &= d\,r\sin\theta\,d\varphi = \sin\theta\,dr\wedge d\varphi + r\cos\theta\,d\theta\wedge d\varphi \\
&= \frac{1}{r}\underset{\approx}{\omega}^r\wedge\underset{\approx}{\omega}^\varphi + \frac{\cos\theta}{r\sin\theta}\underset{\approx}{\omega}^r\wedge\underset{\approx}{\omega}^\varphi\ , \quad \Rightarrow {C_{r\varphi}}^\varphi = -\frac{1}{r}\ ,\ {C_{\theta\varphi}}^\varphi = -\frac{\cot\theta}{r}\ .
\end{aligned}
$$

These are in complete agreement with the ones just calculated above, using tangent vectors; in both cases, of course, those not listed, or determined by skew-symmetry, are exactly zero.

Appendix I:

Stereographic Projection of the Sphere

Stereographic Projection creates a map of the sphere onto a plane on which it sits. We place the sphere so that its South pole sits squarely on an infinite plane, which has $\{x, y\}$ coordinates with origin at the point of contact. Then, we create a mapping between the sphere and the plane by drawing a straight line beginning at the North pole, passing through an arbitrary point on the sphere, which we characterize by P, but also by its values of $\{\theta, \varphi\}$, and then passing on until it strikes the plane below. The values of x and y at this point of striking we will refer to as the values of the coordinates for that point P. Such a straight line, in the envelopping 3-dimensional space, may be written in a parametric form by the 3 equations, as t varies from 0 to 1, the sphere being of **constant diameter** a:

$$\begin{cases} z = at \quad , \\ x = k(1-t) \quad , \\ y = m(1-t) \quad . \end{cases}$$

We then determine the values of the 2 constants, $\{k, m\}$ by relating them to the point on the sphere through which it passes, characterized by $\{\theta, \varphi\}$, and then determine the values of $\{x, y\}$ when the line strikes the plane. After a little algebra, this gives us the result that

$$Z_1 \equiv x_1 + iy_1 = a\cot(\theta/2)\, e^{i\varphi} \quad ,$$

along with the inverse mapping (2.2a)

$$\cot(\theta/2) = \frac{\sqrt{{x_1}^2 + {y_1}^2}}{a} \quad , \quad \tan\varphi = \frac{y_1}{x_1} \quad ,$$

and we have used the subscripts 1 on the coordinates because these correspond to some choice of neighborhood which we call U_1. This mapping gives the South pole the coordinates $(0, 0)$, the equator on the sphere is mapped into a circle on the plane, of radius a, equal to twice the radius of the sphere, etc.. This mapping is well-defined, invertible, continuous, etc., everywhere on the sphere except for the North pole; therefore we may take for our U_1 any open subset of the sphere not containing the North pole, with the coordinate map being given by $(x_1(P), y_1(P))$, as above, in Eq. (A1).

Then, we **must** choose a second neighborhood U_2. We then do the same thing with another plane, now tangent, at its origin, to the North pole, and use lines of projection that begin at the South pole. Following through the same algebra as above, we find that these coordinates, $(x_2(P), y_2(P))$ are given by

$$Z_2 \equiv x_2 + iy_2 = a\tan(\theta/2)\,e^{i\varphi} \quad ,$$

along with the inverse mapping (A2)

$$\tan(\theta/2) = \frac{\sqrt{{x_2}^2 + {y_2}^2}}{a} \quad , \quad \tan\varphi = \frac{y_2}{x_2} \quad .$$

The transition functions are then given by

$$x_2 = \frac{a^2\, x_1}{{x_1}^2 + {y_1}^2} \quad , \quad y_2 = \frac{a^2\, y_1}{{x_1}^2 + {y_1}^2} \quad ,$$

or, more elegantly, simply as

$$Z_2 = \frac{a^2}{\overline{Z_1}} = \frac{a^2\, Z_1}{|Z_1|^2} \quad ,$$

and clearly have all the required continuity properties.