Lecture Notes on Bundles and Connections

Chris Wendl

September 26, 2008

Preface

These notes were written originally as the primary textbook for the second half of MIT's *18.950: Differential Geometry*, Spring 2007. As such, they're aimed at an audience of advanced undergraduates or beginning grad students in math and/or physics, who are expected to be comfortable already with the basic notions of manifolds, vector fields, tensors and differential forms, as covered e.g. in Spivak's *Calculus on Manifolds* [Spi65] or *A Comprehensive Introduction to Differential Geometry*, volume 1 [Spi99]. Though intended as a text for a specific course, the notes treat several topics that go beyond the scope of that course, e.g. principal fiber bundles, symplectic structures and infinite dimensional bundles—these were all considered optional reading for students in 18.950 with the time and interest. There are exercises interspersed throughout, which are for the most part meant not as homework problems, but rather as supplements to the exposition. Some are rather hard, some are almost trivial; the student who doesn't find time to work through them will at least usually benefit from *reading* them.

The main goal here is to give a readable and well-motivated introduction to some of the notions that form the underpinnings of modern differential geometry: vector bundles, fiber bundles, metrics, geodesics, curvature—and in particular the one concept that ties all of these together, that of a *connection*. The focus is thus quite different from that of several popular treatments of differential geometry for undergraduates, e.g. do Carmo's *Differential Geometry of Curves and Surfaces* [dC76]. Unlike such books, we will have fairly little to say here about the theory of distinctly low-dimensional objects, i.e. curves and surfaces in 3-space.[1] We will indeed make heavy use of such objects in examples, for the sake of visualization, but the actual theory will be presented in a way that applies to any number of dimensions. The hope is that the student should thus obtain a good intuitive understanding of concepts that take on fundamental significance in graduate level geometry and geometric research, as well as in theoretical physics.

Connections are a problematic subject for students: one can formulate

[1] The one major exception is the Gauss-Bonnet theorem for surfaces, which is simply too beautiful to ignore.

four or five distinct definitions, each of which has its own advantages and disadvantages, all of which are equivalent, but the beginner will often find it quite difficult to understand why. One of the central themes in these notes is that a connection is not only a *useful* notion, it is also in some sense *necessary*. There are natural questions that one can ask about manifolds, or bundles over manifolds, and the attempt to answer such questions inevitably leads to the first definition of a connection. This definition is natural, but hard to use in practice, so we then go about finding equivalent formulations. Once the basics about connections on vector bundles are well understood, it becomes fairly simple to discuss the most important concepts from Riemannian geometry: geodesics and curvature. Moreover, the alternative definitions turn out to be far more than a matter of mere bookkeeping: while connections serve in Riemannian geometry mainly as a prerequisite for discussing derivatives, they have an altogether different interpretation in quantum field theory, as "wave functions" which can represent particles that mediate the fundamental forces of nature. We will not have space to discuss this subject in any detail, but we will touch upon it.

One more word about physics: it is an unfortunate fact of modern science that mathematicians and physicists may spend a great deal of time thinking about the same concepts, yet still find it difficult to understand each other's work. This is partly a problem of differing (and to some extent irreconcilable) philosophies, but on a more mundane level, it has a lot to do with notation, and that's a problem that can be fixed. Anyone who's ever perused a book on general relativity is familiar with the standard proliferation of upper and lower tensor indices, and the notational shortcut known as the *Einstein summation convention*. Contrary to what some physicists may claim, these notational details are *not* fundamental to the theory of differential geometry: it's possible to get by entirely without them, and many mathematicians do. That said, sometimes it's nice to be able to do computations in coordinates, and for this the physicists' system works beautifully, as long as one is careful enough to keep in mind the geometric meaning behind the indices. We shall therefore use both systems of notation in these notes, and try always to make it clear how one should translate between them.

As of this writing, the notes are still a work in progress (intended eventually to become a book), with occasional updates posted at:

`http://www.math.ethz.ch/~wendl/connections.html`

Questions, comments and corrections are of course welcome!

Chapter 1

Introduction: what is a connection, and why should we care?

Contents

The concept of a connection arises immediately from the following simple question:

> *If M is a manifold, X is a vector field and γ is a smooth path in M, how can we judge whether X is* constant *along γ?*

You can visualize the question as in Figure 1.1: imagine M as a smooth 2-dimensional surface in $\mathbb{R}^3$, with a smooth curve $\gamma \subset M$ and a smooth vector field X tangent to M. For each $p \in M$, $X(p)$ is now a vector in $\mathbb{R}^3$. But one cannot hope in general for X to be constant along γ as a vector in $\mathbb{R}^3$, because the tangent spaces may change—the vector $X(p_0) \in T_{p_0}M \subset \mathbb{R}^3$ will not necessarily be tangent to M at p.

Clearly asking X to be constant in the ambient space is not the right approach. Intuitively, one would think that one should call X constant along γ if its derivative in directions tangent to γ is always zero. Such a directional derivative would have to be computed in some choice of coordinates. The trouble is that, in general, the answer depends on this choice.

Exercise 1.1. In the situation described above, suppose $(x^1, \ldots, x^n)$ are coordinates in a neighborhood of $p_0 \in \gamma \subset M$. We can then express the

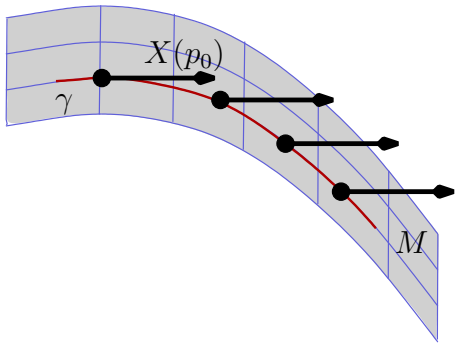

Figure 1.1: The vectors along $\gamma \subset M$ are constant as vectors in $\mathbb{R}^3$, but therefore not tangent to M.

vector field X at points p near p_0 via its component functions $X^j(p)$:

$$X(p) = \sum_j X^j(p) \frac{\partial}{\partial x^j}.$$

Show that for any given tangent vector $v = \sum_i v^i \frac{\partial}{\partial x^i} \in T_{p_0}M$, the relation $dX^j(p_0)v = 0$ is *not* invariant under coordinate changes. In other words, given another coordinate system $(\tilde{x}_1, \ldots, \tilde{x}_n)$ near p_0 such that

$$X(p) = \sum_j \tilde{X}^j(p) \frac{\partial}{\partial \tilde{x}^j},$$

derive the formula

$$d\tilde{X}^i(p_0)v = \sum_j \frac{\partial \tilde{x}^i}{\partial x^j}(p_0) \cdot dX^j(p_0)v + \sum_{j,k} \frac{\partial^2 \tilde{x}^i}{\partial x^j \partial x^k}(p_0) \cdot v^k X^j(p_0),$$

and use it to observe that $d\tilde{X}^j(p_0)v$ and $dX^j(p_0)v$ need not both be zero if one of them is.

What this shows is that there is no naturally defined notion of a "constant" vector field on smooth manifolds in general. In certain situations however, one can easily imagine what constant *should* mean. On the surface $M \subset \mathbb{R}^3$, the following prescription seems natural:

> *X is constant along γ if for every point $p \subset \gamma$, $X(p)$ has the same length and makes the same angle with γ.* (1.1)

Figure 1.2 shows two vector fields along $\gamma \subset M \subset \mathbb{R}^3$ which are constant by this definition. Notice which pieces of structure we've used to make sense of this. On smooth manifolds in general, there is no naturally defined notion

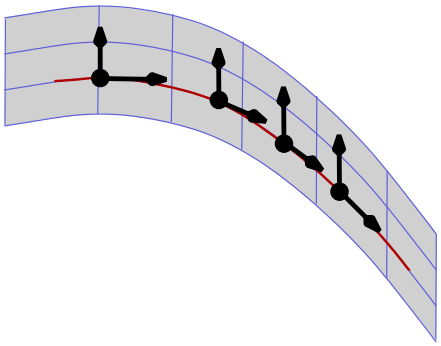

Figure 1.2: Two vector fields along $\gamma \subset M \subset \mathbb{R}^3$ which are constant according to (1.1).

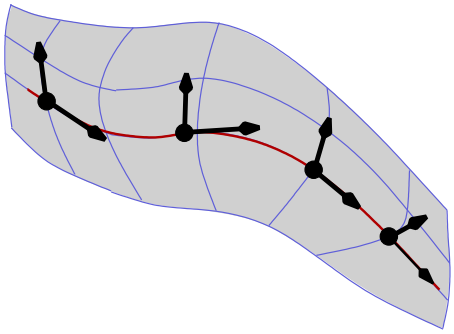

Figure 1.3: After deforming the embedding $M \hookrightarrow \mathbb{R}^3$, the two vector fields of Figure 1.2 no longer appear constant along γ.

of "length" or "angle"—they are not coordinate invariant. Things become simpler when we have an embedding of M into $\mathbb{R}^3$, because it then inherits these notions from the Euclidean structure of $\mathbb{R}^3$. The catch is that things would change if we changed the embedding, i.e. we could deform the surface so that lengths and angles change, and a formerly constant vector field no longer looks constant (Figure 1.3). In this sense our definition depends on the choice of embedding $M \hookrightarrow \mathbb{R}^3$, and cannot be expressed in terms of the intrinsic properties of M. In fact, the embedding into $\mathbb{R}^3$ is not crucial; what's important rather is that we know how to define lengths of tangent vectors and angles between them. There is a standard type of structure that one can add to any manifold so that these notions are well defined: it's called a *Riemannian metric*, and is the starting point of Riemannian geometry.

Our example above illustrates a fact that is sometimes called the *fundamental theorem of Riemannian geometry*: a choice of Riemannian metric uniquely determines a special *connection* on M, i.e. a notion of constant

vector fields along paths.[1] As we will later see, a connection need not always be associated with a metric, and it is often useful to keep this distinction in mind. The most interesting aspects of Riemannian geometry—the ideas of geodesics and curvature—can in fact be defined purely in terms of connections, without referring explicitly to the metric. Moreover, it can sometimes be interesting to consider connections in settings where a metric is either nonexistent or beside the point.

1.1 Motivation from Riemannian geometry

So far this discussion of constant vector fields may seem somewhat academic, so let us see how it relates to some of the more fundamental questions in geometry.

As mentioned above, a *Riemannian metric* is a structure one can add to any smooth manifold so that lengths and angles are well defined for tangent vectors. Clearly a surface M embedded smoothly in $\mathbb{R}^3$ inherits such structure from the embedding, so for the sake of intuitive clarity, we consider only this case in the present discussion. It should already be clear that the same manifold can admit many distinct metrics, e.g. if $M = S^2$, deforming the embedding $S^2 \hookrightarrow \mathbb{R}^3$ may change the metric at each point. Figures 1.5 and 1.6 show two examples of spheres in $\mathbb{R}^3$ with distinct metrics.

It is then natural to ask the following question:

> For a given point $p \in M \subset \mathbb{R}^3$, can a small neighborhood of p be deformed, *without changing the metric*, so that it becomes a *flat* surface in $\mathbb{R}^3$?

If M has this property at the point p, it is called *locally flat* at p. This is indeed a purely *local* question, i.e. it depends only on the metric structure of M in arbitrarily small neighborhoods of a given point. It would be trivial if we left out the detail about not changing the metric: *any* small enough neighborhood of p can be deformed smoothly into a flat surface. The key is to imagine M as a physical surface that can be bent but not stretched, e.g. a piece of paper. Figure 1.6 shows an example of a sphere in $\mathbb{R}^3$ that is locally flat in a certain area—the darkly shaded region—but not everywhere. One can see that the shaded region is locally flat because it looks like a cylinder, and we can imagine taking a piece of paper that's shaped like part of a cylinder and flattening it out, without having to stretch or tear it. In other words, a piece of a cylinder can be flattened

[1] Strictly speaking, the characteristization in (1.1) of the special connection determined by a Riemannian metric is only correct if the path γ is a *geodesic*, i.e. a path that locally minimizes length. We'll formulate the precise definition in Chapter 4.

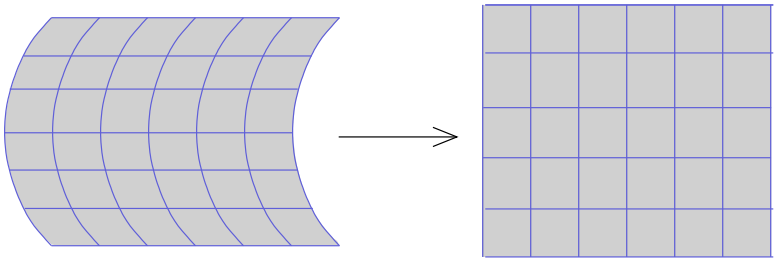

Figure 1.4: A piece of a cylinder can be flattened to a plane without changing any lengths or angles on the surface.

without changing any of the lengths of its tangent vectors or angles between them (Figure 1.4).

This is *not* true of the standard round sphere in $\mathbb{R}^3$. Perhaps you've never held in your hand a piece of paper that's shaped like part of a globe[2], but you can surely imagine that if you did, you could never make it *flat* without breaking or stretching it. The reason for this, as we will later see, is that the round sphere has *positive curvature* at every point. By contrast, a cylinder has *zero curvature*, and is thus locally flat. The statement that a cylinder is in some sense "not curved" may seem jarring at first, but you'll get used to it: the point is that the quantity we're calling curvature should depend only on the metric, and not on the specific way we've chosen to embed the surface in $\mathbb{R}^3$.

So how can curvature be quantified? One answer is to return to the notion of "constant vector fields". Recall from the previous discussion that our choice of metric defines a natural connection on M via (1.1). Consider now a cylinder $Z \subset \mathbb{R}^3$ with a simple closed curve $\gamma \subset Z$, e.g. the square depicted in the darkly shaded region of Figure 1.6. Choose any point $p_0 \in \gamma$ and a tangent vector $v_0 \in T_{p_0}Z$, and imagine extending v to a vector field which is always constant along γ. (The fancy term for this is *parallel transport*.) It's easy to see that if we follow this "parallel" vector field from p_0 once around the path, it will always return to the same initial vector v_0.

The situation on the round sphere $S^2 \subset \mathbb{R}^3$ is quite different. To see this, choose $p_0 \in S^2$ to be a point on the equator, with $v_0 \in T_{p_0}S^2$ pointing along the equator. Let γ be the triangular path depicted in Figure 1.5, moving 90 degrees of longitude along the equator, up along a great circle to the north pole, and down along another great circle to the original point p_0. Extending v_0 as a "parallel" vector field along this path, we see that it remains parallel to the equator on the first leg, thus becoming perpendicular to the great circle as it moves toward the north pole, and

[2]If you know where to buy one, please let me know!

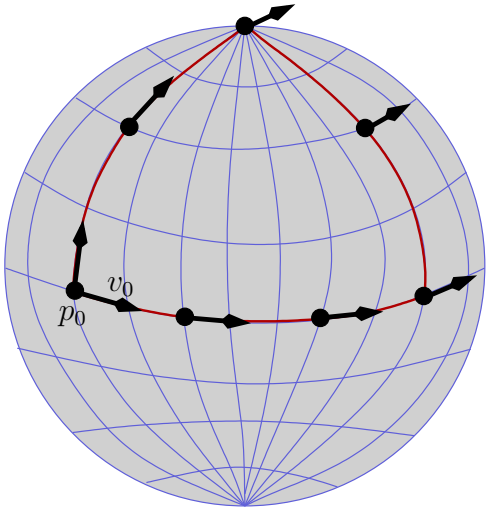

Figure 1.5: The "round" sphere $S^2 \subset \mathbb{R}^3$. Parallel transport of a vector along a closed path leads to a different vector upon return.

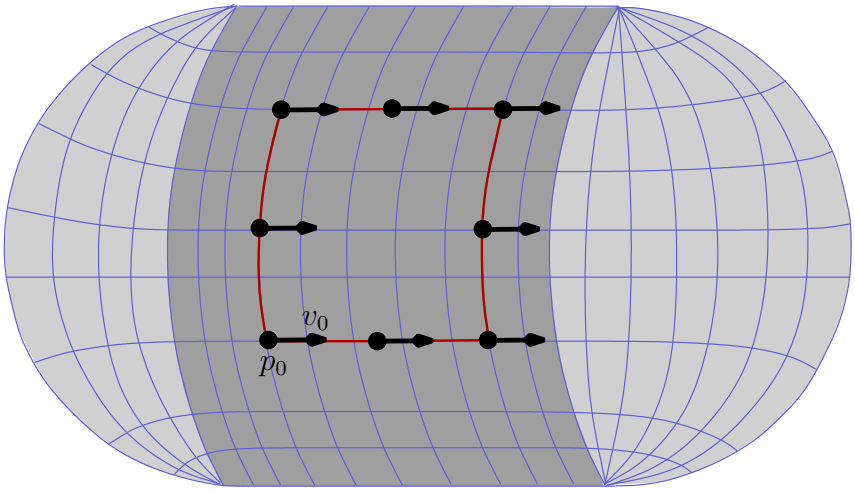

Figure 1.6: A different embedding of S^2 in $\mathbb{R}^3$, so that the darkly shaded region is locally flat. Parallel transport of a vector around a closed path in this region always leads back to the same initial vector.

parallel to the other great circle moving back down. The key observation is that the vector at the end of the path is *different* from the original vector v_0. One can show (and we later will) that this is true for parallel transport along *any* closed path in the round sphere—quantifying this difference for very small paths gives a precise measure of curvature, and shows in particular that the round sphere has *constant positive curvature*. This is not unrelated to the fact that the angles of the "triangle" in Figure 1.5 add up to considerably more than 180 degrees. In the case of the cylinder (or for that matter the shaded region of the sphere in Figure 1.6), the fact that we always return to the same vector around any closed path means that the curvature is zero. We will also see examples later of surfaces with *negative curvature*: the basic picture to keep in mind is the shape of a *saddle*.

Now that we've said a little bit about what curvature is, we take this opportunity to state a rather nontrivial theorem:

Theorem 1.2. *There is no metric on S^2 that is everywhere locally flat.*

This follows from the beautiful *Gauss-Bonnet theorem* for surfaces, which we will discuss later. It relates the integral of the curvature over the whole surface to a topological quantity, its *Euler characteristic*, which in the case of S^2 is positive. This is the reason why Figure 1.6 could not have been drawn so that *every* part of the sphere had zero curvature.

1.2 Motivation from physics

We now switch gears and briefly discuss a quite different situation where connections arise: Quantum Field Theory. It should not be surprising to anyone vaguely familiar with General Relativity that connections play a role in modern physics; Einstein's fundamental idea was that gravitation is, in some sense, not an actual force, but rather a manifestation of the geometry of a certain four-manifold known as "space-time"—more specifically a manifestation of its curvature, and as we've seen above, curvature can be described in terms of connections. A fact which is less familiar in the popular imagination, and admittedly less well understood in general, is that something similar can be said of the other fundamental forces as well—only here the question is not one of connections and curvature on a manifold, but rather on a more general *bundle* over this manifold. We will now discuss this idea in the simplest possible case: Maxwell's electrodynamics.

For this discussion we adopt the notational conventions of relativistic mechanics: thus *space* is $\mathbb{R}^3$, and an *event* is any combination of a point in space $\mathbf{x} = (x^1, x^2, x^3) \in \mathbb{R}^3$ with a *time* $t = x^0 \in \mathbb{R}$; together these form

a point in the 4-dimensional manifold that we call *space-time*,

$$x := (t, \mathbf{x}) = (x^0, x^1, x^2, x^3) \in \mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3.$$

One generally uses *Greek* letters for indices that run from 0 to 3, hence the components of $x \in \mathbb{R}^4$ are called x^μ, $\mu = 0, \ldots, 3$. *Latin* letters on the other hand are reserved for "space-only" components, with an index running from 1 to 3, i.e. the components of $\mathbf{x} \in \mathbb{R}^3$ are x^j, $j = 1, \ldots, 3$. We will employ the *Einstein summation convention* wherever convenient, which means that any pair of matching upper and lower Greek indices implies a summation from 0 to 3: a 1-form can then be written $A = A_\mu \, dx^\mu$, and its action on a vector $X = X^\mu \frac{\partial}{\partial x^\mu}$ given by $A(X) = A_\mu X^\mu$. The convention is explained more fully in Appendix A.

The natural inner product on space-time is the *Minkowski metric*, defined by $\langle X, Y \rangle = \eta_{\mu\nu} X^\mu Y^\nu$, where $\eta_{\mu\nu}$ are the components of the matrix

$$\boldsymbol{\eta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Thus $\langle X, X \rangle = X_\mu X^\mu = (X^0)^2 - \sum_{j=1}^3 (X^j)^2$ is not generally positive—its sign distinguishes between velocities of paths through space-time that move slower ($X_\mu X^\mu > 0$, "timelike") or faster ($X_\mu X^\mu < 0$, "spacelike") than light.

In classical electrodynamics, the electric and magnetic fields are a pair of time-dependent vector fields $\mathbf{E} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ and $\mathbf{B} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ respectively, whose behavior is affected by the presence of a time-dependent *charge density* $\rho : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ and *current density* $\mathbf{j} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$. Specifically, ρ and $\mathbf{j}$ influence $\mathbf{E}$ and $\mathbf{B}$ via *Maxwell's equations*:

$$\begin{aligned} \nabla \times \mathbf{E} + \partial_t \mathbf{B} &= 0 & \qquad \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{B} - \partial_t \mathbf{E} &= \mathbf{j} & \qquad \nabla \cdot \mathbf{E} &= \rho \end{aligned} \tag{1.2}$$

Here the operator $\nabla = (\partial_1, \partial_2, \partial_3)$ involves only the *spatial* partial derivatives, thus for instance

$$\nabla \cdot \mathbf{B} = \partial_1 B^1 + \partial_2 B^2 + \partial_3 B^3$$

and

$$\nabla \times \mathbf{B} = \begin{pmatrix} \partial_2 B^3 - \partial_3 B^2 \\ \partial_3 B^1 - \partial_1 B^3 \\ \partial_1 B^2 - \partial_2 B^1 \end{pmatrix}.$$

Note that Maxwell's equations normally appear with some extra physical constants that we are omitting; by choosing appropriate units, it's possible to set all of these constants equal to 1, so assume this.

The first step in simplifying Maxwell's equations is always to reformulate them in terms of *potentials*. This starts with the observation that since $\nabla \cdot \mathbf{B} = 0$, there exists a time-dependent vector field $\mathbf{A} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, called the *vector potential*, such that $\nabla \times \mathbf{A} = \mathbf{B}$. The existence of such a vector field is a standard fact following from Stokes' theorem. Next we observe that by reversing the order of partial derivatives, $\partial_t \mathbf{B} = \partial_t(\nabla \times \mathbf{A}) = \nabla \times (\partial_t \mathbf{A})$, thus the first of Maxwell's equations says $\nabla \times (\mathbf{E} + \partial_t \mathbf{A}) = 0$. As another application of Stokes' theorem, curl-free vector fields can be written as gradients, thus there exists a real-valued function $V : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$, the *scalar potential*, such that $-\nabla V = \mathbf{E} + \partial_t \mathbf{A}$. The pair of functions V and $\mathbf{A}$ thus determine both fields by

$$\begin{aligned} \mathbf{E} &= -\nabla V - \partial_t \mathbf{A}, \\ \mathbf{B} &= \nabla \times \mathbf{A}. \end{aligned} \tag{1.3}$$

Maxwell's equations can now be rewritten entirely in terms of V and $\mathbf{A}$, and in fact the top two become trivial: they merely express the fact that the curl of a gradient is always zero, as is the divergence of a curl. The bottom two equations of (1.2) give nontrivial PDEs for V and $\mathbf{A}$—an improvement over the original situation, since V and $\mathbf{A}$ consist of *four* unknown real-valued functions rather than six.

We will refrain from writing these equations down for the moment, and focus instead on the way in which the potentials V and $\mathbf{A}$ determine the fields $\mathbf{E}$ and $\mathbf{B}$. In particular, (1.3) allows considerable freedom to change the potentials without changing the fields: $\mathbf{E}$ and $\mathbf{B}$ remain the same if we alter the potentials by a transformation of the form

$$\begin{aligned} V &\mapsto V - \partial_t \theta, \\ \mathbf{A} &\mapsto \mathbf{A} + \nabla \theta \end{aligned} \tag{1.4}$$

for any smooth function $\theta : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$. This type of change is called a *gauge transformation*, and as we will now show, it has considerably more geometric significance than may at first be apparent.

First, we simplify matters one step further by observing that the relation between fields and potentials can be written more elegantly in terms of differential forms. We do this by combining V and $\mathbf{A} = (A^1, A^2, A^3)$ together into a 1-form on $\mathbb{R}^4$:

$$A = A_\mu \, dx^\mu := -V \, dx^0 + A^1 \, dx^1 + A^2 \, dx^2 + A^3 \, dx^3.$$

Denote the exterior derivative of A by

$$F := dA = \partial_\mu A_\nu \, dx^\mu \wedge dx^\nu = (\partial_\mu A_\nu - \partial_\nu A_\mu) \, dx^\mu \otimes dx^\nu,$$

so the components $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ form a matrix of the form

$$\mathbf{F} = \begin{pmatrix} 0 & \partial_t A^1 + \partial_1 V & \partial_t A^2 + \partial_2 V & \partial_t A^3 + \partial_3 V \\ -\partial_1 V - \partial_t A^1 & 0 & \partial_1 A^2 - \partial_2 A^1 & \partial_1 A^3 - \partial_3 A^1 \\ -\partial_2 V - \partial_t A^2 & \partial_2 A^1 - \partial_1 A^2 & 0 & \partial_2 A^3 - \partial_3 A^2 \\ -\partial_3 V - \partial_t A^3 & \partial_3 A^1 - \partial_1 A^3 & \partial_3 A^2 - \partial_2 A^3 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix}.$$

In light of this, we call $F = dA$ the *field strength tensor*. It is unchanged if we add to A any closed (and therefore also exact) 1-form, thus the gauge transformation (1.4) now becomes

$$A \mapsto A + d\theta \tag{1.5}$$

for any smooth function $\theta : \mathbb{R}^4 \to \mathbb{R}$.

To understand the significance of (1.5), we now discuss some basic notions about incorporating "matter fields" into electrodynamics. The reader should be warned that the following ideas occupy a nebulous area somewhere *between* classical and quantum mechanics, and therefore do not serve to describe the actual behavior of either the macroscopic world or elementary particles. What we will explain is in any case a necessary *starting point* for a certain kind of quantum field theory, which leads eventually (after considerably more work) to a description of the fundamental particles and forces of nature.

Matter appears in (1.2) via the charge density ρ and current density $\mathbf{j}$, and classically one thinks of these quantities being defined by a large number of charged point particles with specific positions and velocities in space at specific times. To discuss matter in a quantum context, one must instead think of a charged particle as a "probability wave," described by a time-dependent *wave function*

$$\psi : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{C}.$$

Then one cannot say precisely where the particle is at a given time t, but one can say that the probability of finding the particle within some region $\mathcal{U} \subset \mathbb{R}^3$ at that time is

$$\int_{\mathcal{U}} |\psi(t, \mathbf{x})|^2 \, dx^1 \wedge dx^2 \wedge dx^3. \tag{1.6}$$

One of the principles of quantum mechanics is that the wave function ψ evolves over time according to a PDE, the exact choice of which varies in different versions of the theory. Nonrelativistic quantum mechanics uses

the *Schrödinger equation*; a slightly simpler equation which is also relativistically invariant is the *Klein-Gordon equation*:

$$(-\partial_\mu \partial^\mu + m^2)\psi = -\partial_t^2 \psi + \Delta\psi + m^2\psi = 0, \tag{1.7}$$

where $\Delta := \partial_1^2 + \partial_2^2 + \partial_3^2$ is the spatial Laplacian operator, and $m > 0$ is the mass of the particle.

We said above that the particle represented by ψ has "charge," but what does this mean? If you've seen any intermediate level classical mechanics, you might be familiar with a general principle called *Noether's theorem*: conserved quantities arise naturally from symmetries. The simplest example is the observation that (1.7) is invariant under time and space translations, i.e. coordinate changes that replace x^μ with $x^\mu + a^\mu$ for any constants $a^\mu \in \mathbb{R}$; from this fact one can derive expressions for energy and momentum in terms of ψ, and prove that they are conserved. So we now ask: what symmetry gives rise to the conservation of charge? The answer turns out to depend on the fact that we chose ψ to take values in $\mathbb{C}$ rather than just $\mathbb{R}$: the Klein-Gordon equation and the probability in (1.6) are then both preserved by transformations of the form

$$\psi \mapsto e^{i\theta}\psi \tag{1.8}$$

for any $\theta \in \mathbb{R}$. This observation together with Noether's theorem leads one to define the vector field

$$j^\mu = \mathrm{Im}(\bar{\psi}\ \partial^\mu \psi), \tag{1.9}$$

where Im denotes the *imaginary part*, i.e. $\mathrm{Im}(a + ib) = b$.

Exercise 1.3. Show that if ψ satisfies (1.7), then j^μ satisfies $\partial_\mu j^\mu = 0$. *Hint:* recall for any $z \in \mathbb{C}$, $\mathrm{Im}(z) = \frac{1}{2i}(z - \bar{z})$.

The relation $\partial_\mu j^\mu = 0$ of Exercise 1.3 can be reinterpreted as the so-called *continuity equation*: writing $\rho := j^0$ and $\mathbf{j} := (j^1, j^2, j^3)$, we have

$$\partial_\mu j^\mu = \partial_t \rho + \nabla \cdot \mathbf{j} = 0.$$

Then if $t \in \mathbb{R}$ is a fixed time and $\mathcal{U} \subset \mathbb{R}^3$ is any compact subset with smooth boundary, we can define

$$Q(t, \mathcal{U}) := \int_{\mathcal{U}} \rho(t, \mathbf{x})\ dx^1 \wedge dx^2 \wedge dx^3$$

and apply the divergence theorem to find

$$\frac{d}{dt} Q(t, \mathcal{U}) = -\int_{\mathcal{U}} \nabla \cdot \mathbf{j}\ dx^1 \wedge dx^2 \wedge dx^3 = -\int_{\partial \mathcal{U}} \mathbf{j} \cdot \mathbf{n}.$$

This is precisely the relation that must be satisfied by charge and current densities to imply *conservation of charge*: it indicates that the change in total charge within $\mathcal{U}$ equals minus the amount of current flowing out through $\partial\mathcal{U}$. With this motivation, we interpret $j = (\rho, \mathbf{j})$ as a four-dimensional representation of both charge density ρ and current density $\mathbf{j}$ for the wave function ψ.

With charge and current densities understood, we now see how to plug ψ into Maxwell's equations so that the electromagnetic field is influenced by the wave function of a charged particle—but as yet the evolution equation (1.7) for ψ does not see the electromagnetic field at all. The remedy for this is a fundamentally geometric idea.

Let us first mention how the corresponding issue is dealt with for gravitation. In General Relativity, one begins by considering point particles that move by tracing straight paths through *flat* space-time, i.e. $\mathbb{R}^4$. The crucial geometric observation is that since particles never interact with each other "at a distance" (they may exert forces from a distance, but these forces also take time to *propagate*, rather than acting instantaneously), we need only *locally* assume that space-time looks like $\mathbb{R}^4$: *globally* it could be any smooth 4-dimensional manifold. But with that allowance, it's no longer obvious what "tracing a straight path" means, and one fixes this by introducing a metric (in this case not a Riemannian metric, but something similar), and consequently a connection. We then have a well defined notion of constant vector fields along paths, and can define a path to be "straight" if its velocity vector is constant. Paths with this property are called *geodesics*, and they satisfy a second-order differential equation called the *geodesic equation*, which is then seen as the equation of motion for a point particle in a gravitational field. Likewise, massive particles moving through space-time affect its geometry according to a PDE: this is *Einstein's equation*, which relates curvature to the mass and energy density determined by matter. With this, one has reduced the gravitational field to a purely geometric object.

To understand the electromagnetic field, we will not replace $\mathbb{R}^4$ by a general 4-manifold, although one *can* sometimes do this. We note instead that the symmetry (1.8), which we associated with charge and current, does not act on the ambient space $\mathbb{R}^4$ but rather on the space in which ψ takes its *values*, the complex plane. One can picture ψ by imagining that there is a copy E_x of the plane $\mathbb{C}$ associated to each point $x \in \mathbb{R}^4$, and ψ assigns to x a point in the plane E_x, depicted in Figure 1.7 as a vector based at the origin. Thus far each of the planes E_x is canonically identified with $\mathbb{C}$, but suppose we drop this assumption: suppose the collection $\{E_x\}_{x\in\mathbb{R}^4}$ is simply a set of planes that "vary smoothly" in some sense as x moves over $\mathbb{R}^4$, but without any canonical isomorphisms between E_x and E_y for $x \neq y$. If you like, imagine the planes "moving" smoothly as x varies, as in Figure 1.8.

This is the essence of a *vector bundle*, and the map ψ associating to x a vector in E_x is called a *section* of the bundle. Just like the tangent bundle of a manifold, we now have a collection of vector spaces that vary smoothly, but it is no longer obvious what it means to call a section ψ "constant," or how to define the partial derivatives $\partial_\mu \psi$ which appear in the Klein-Gordon equation. Moreover, the symmetry transformation of (1.8) is now far too restrictive: it implicitly assumes some notion of constant sections, since these are preserved by the transformation $\psi(x) \mapsto e^{i\theta}\psi(x)$. To avoid this implicit assumption, one should allow more general transformations of the form

$$\psi(x) \mapsto e^{i\theta(x)}\psi(x) \tag{1.10}$$

where $\theta : \mathbb{R}^4 \to \mathbb{R}$ is an arbitrary smooth function. This is also called a *gauge transformation*, for reasons that will be clear in a moment; it can be seen as a kind of "coordinate transformation" on the vector bundle. In order to make sure that the laws of physics never depend on our choice of coordinates, we must therefore change (1.7) so as to be invariant under (1.10).

The solution is to choose a connection on the vector bundle and use it to define a new, gauge invariant partial derivative operator, called the *covariant derivative*. We will explain in detail later the various ways that one can define connections on vector bundles and why they make sense; for now we simply give the basic ideas for the present example. Let $A = A_\mu \, dx^\mu$ be a smooth 1-form on $\mathbb{R}^4$ and define the operator

$$\nabla_\mu = \partial_\mu + iA_\mu$$

on complex-valued functions. This will be the covariant derivative: we only have to stipulate that under gauge transformations, the 1-form A must change in the appropriate way to make ∇_μ an invariant operator. This means the following. For any smooth function $\theta : \mathbb{R}^4 \to \mathbb{R}$, let $\tilde{\psi} = e^{i\theta}\psi$; the covariant derivative ∇ will simultaneously change to $\widetilde{\nabla}_\mu = \partial_\mu + i\tilde{A}_\mu$, where $\tilde{A}$ is a new 1-form to be determined by the condition

$$\widetilde{\nabla}_\mu \tilde{\psi} = \widetilde{\nabla_\mu \psi}.$$

We compute:

$$\begin{aligned}
& & (\partial_\mu + i\tilde{A}_\mu)e^{i\theta}\psi &= e^{i\theta}(\partial_\mu + iA_\mu)\psi \\
\iff & & e^{i\theta}(\partial_\mu + i\tilde{A}_\mu)\psi + e^{i\theta}i(\partial_\mu\theta)\psi &= e^{i\theta}(\partial_\mu + iA_\mu)\psi \\
\iff & & e^{i\theta}\left[\partial_\mu + i(\tilde{A}_\mu + \partial_\mu\theta)\right]\psi &= e^{i\theta}(\partial_\mu + iA_\mu)\psi \\
& \iff & \tilde{A}_\mu + \partial_\mu\theta &= A_\mu,
\end{aligned}$$

or in terms of 1-forms, $\tilde{A} + d\theta = A$.

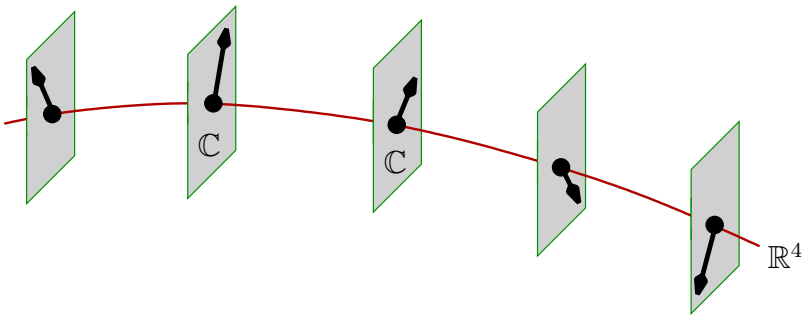

Figure 1.7: A piece of $\mathbb{R}^4$ (depicted as a one-dimensional curve) with a copy of the plane $\mathbb{C}$ attached to each point, and a "section" $\psi : \mathbb{R}^4 \to \mathbb{C}$ depicted by vectors in the corresponding planes.

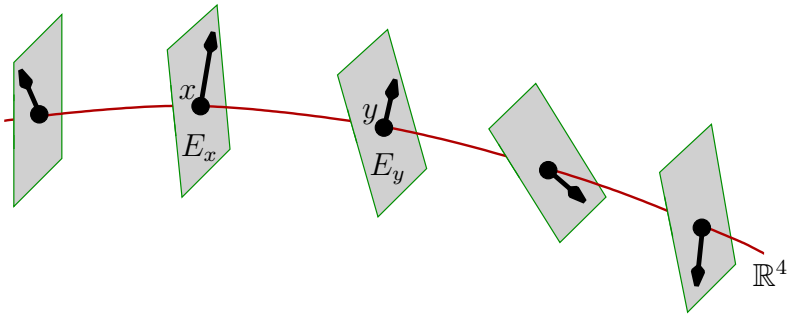

Figure 1.8: A less obviously trivial vector bundle, smoothly associating to each point $x \in \mathbb{R}^4$ a plane E_x. Here the distinct planes are isomorphic vector spaces but there is generally no canonical isomorphism $E_x \to E_y$ for $x \neq y$.

This expression matches the transformation (1.5) of the electromagnetic potential $A = A_\mu\, dx^\mu$ and explains why we've chosen to use the same term, "gauge transformation" for both (1.10) and (1.5). The upshot is that the electromagnetic potential obtains a geometric interpretation as a connection on the vector bundle in which the matter field ψ takes its values. The effect of A on the evolution of ψ is then seen by replacing (1.7) by the *covariant* Klein-Gordon equation:

$$(-\nabla_\mu \nabla^\mu + m^2)\psi = \left[-\partial_\mu \partial^\mu - 2iA^\mu \partial_\mu - i(\partial_\mu A^\mu) + m^2\right] \psi = 0.$$

By construction, this equation is invariant under gauge transformations.

Just as with connections on manifolds, one can analyze parallel transport of vectors in a bundle around closed paths and use this to define a notion of *curvature* for the bundle: this curvature turns out to be zero if and only if $dA = 0$. In other words, *the field strength tensor is the curvature!* And Maxwell's equations now play the role for electrodynamics that Einstein's equation does in gravitation, relating the curvature of our connection to the charge and current distributions defined by ψ.

This discussion has been necessarily vague in two respects: mathematically, because we've not yet defined terms such as "vector bundle" and "connection" which will be needed to make everything precise. This will be fixed in subsequent chapters. We should at least mention that in modern mathematics, gauge theory plays an important role quite independent of its physical interpretation. Indeed, beginning with the work of Donaldson in the early 1980's, it turned out that spaces of connections on vector bundles satisfying certain PDEs have nice geometric properties which can be used to define invariants in differential topology, notably for three- and four-manifolds. This is a large topic which is far beyond the scope of these notes, but the interested reader is referred to the book of Donaldson and Kronheimer [DK90] for an introduction.

As mentioned above, the physics in this discussion has also been rather vague, partly because we're not describing a real theory, but more of an intermediate step between the classical and quantum theories. The ideas are nevertheless important, and they generalize nicely to describe other fundamental forces, not just electrodynamics. What we've been discussing is called *abelian gauge theory*, owing to the fact that the transformations (1.8) can be described by elements of an abelian (i.e. commutative) group, the unitary group $U(1)$. A natural generalization is to consider *nonabelian* gauge theories, in which transformations come from nonabelian groups such as $SU(3) \times SU(2) \times U(1)$: the latter is in fact quite important in the standard model of elementary particle physics, which unifies electrodynamics with the so-called *weak* and *strong* forces. We will have little room to say more on this subject in these notes (besides making more precise what has already been said), but refer the reader to [GS87] or [Tic99] for further details.

References

[dC76] M. P. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1976. Translated from the Portuguese.

[DK90] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1990. ; Oxford Science Publications.

[GS87] M. Göckeler and T. Schücker, *Differential geometry, gauge theories, and gravity*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge, 1987.

[Spi65] M. Spivak, *Calculus on manifolds*: *A modern approach to classical theorems of advanced calculus*, W. A. Benjamin, Inc., New York-Amsterdam, 1965.

[Spi99] ———, *A comprehensive introduction to differential geometry*, 3rd ed., Vol. 1, Publish or Perish Inc., Houston, TX, 1999.

[Tic99] R. Ticciati, *Quantum field theory for mathematicians*, Encyclopedia of Mathematics and its Applications, vol. 72, Cambridge University Press, Cambridge, 1999.