Topics in Mathematical Physics

Guo Chuan Thiang

Lecture notes for a course at Peking University, Fall 2022. Please email me for feedback, and if you would like to use some of this material in teaching.

Contents

1 On fields and frames

The complex number field $\mathbb{C}$ has a 1-dimensional underlying vector space (or inner product/Hilbert space). Implicitly, it also has a distinguished basis $\{1\}$. A general 1-dimensional Hilbert space E has no distinguished basis, and there is a U(1)-phase freedom in choosing a normalized basis vector.

More generally, let E be an abstract n-dimensional Hilbert space. Then each choice of orthonormal basis gives a unitary identification $E \cong \mathbb{C}^n$. Consider the set $\mathrm{Fr}(E)$ of orthonormal bases $\{e_1, \ldots, e_n\}$ ("frames of E"). There

is a right action of $\mathrm{U}(n)$ on $\mathrm{Fr}(E)$,

$$e'_j = \sum_k e_k u_{kj}, \qquad u = (u_{kj}) \in \mathrm{U}(n). \tag{1.1}$$

This action is simply transitive, so $\mathrm{Fr}(E)$ is a $\mathrm{U}(n)$-*torsor*. This means that *after* we fix a reference basis $\{e_j\}_{j=1}^n$, then any other basis $\{e'_j\}_{j=1}^n$ can be labelled by the unique unitary matrix satisfying (1.1). But there is no canonical reference basis, therefore no canonical identification $\mathrm{Fr}(E) \cong \mathrm{U}(n)$.

Similarly, if E is a real inner product space, we have $\mathrm{Fr}(E) \cong \mathrm{O}(n)$. If there is no inner product, and we just want to describe the set of bases, we have $\mathrm{Fr}(E) \cong \mathrm{GL}(n)$.

Q : Why do we care about distinguishing identifications $E \cong \mathbb{C}^n$?

An n-component function $X \to \mathbb{C}^n$ defined on some initial base space X has, conceptually, a single "external target space" $\mathbb{C}^n$. All points of X get mapped into the same $\mathbb{C}^n$ by the function. Also, we are allowed to start from any other space Y and consider functions $Y \to \mathbb{C}^n$ mapping into this same $\mathbb{C}^n$. So the target $\mathbb{C}^n$ has no special relation to the domain X or Y.

The concept of a *field* is subtly but dramatically different. At each point $x \in X$, a field has a value *localized at* x, and not at some other point x'. That is, each point x has its own "local target space" E_x. In total, there is a "bundle" $E = \sqcup_{x \in X} E_x$, equipped with a natural projection map $\pi : E \to X$. A field is a map

$$v : X \to E, \qquad \text{such that} \quad v(x) \in E_x, \ \forall x \in X,$$

also called a *section* of the bundle $E \to X$. Equivalently, v satisfies $\pi \circ v = \mathrm{id}_X$.

A priori, the target spaces $E_x, E_{x'}$ at different points are not identified. For example, if E_x is a vector space, then a choice of basis (frame) for E_x does not instantaneously determine basis choices for the other $E_{x'}$.

Now, if X and the E_x are not just sets, but topological spaces, then we have "local groupings" of the points of X coming from the topology (the open sets). So for an open set $U \subset X$, we can at least demand that the $E_x, x \in U$ are "continuously identifiable". Formally, we ask for the existence of a homeomorphism

$$\pi^{-1}(U) \equiv \bigsqcup_{x \in U} E_x \overset{f}{\cong} U \times F,$$

such that $\pi(p) = \pi_U(f(p))$. Here, F is some "reference" fibre, and $U \times F$ is a reference "trivial bundle" with π_U the projection onto the first component. This condition is called "local trivializability". The observant reader might notice that we have not yet given $\bigsqcup_{x \in U} E_x$ a topology, so what we have just described does not really make sense! Indeed, in physics, one often comes across informal "definitions" like "...the tangent bundle of a manifold is the (disjoint) union of all its tangent spaces...". However, the global topology of E is actually extremely important, and we will spend some time understanding how, e.g., the topology on the tangent bundle of E is defined.

Now, if the base space X is a manifold, then we not only have a notion of "neighbourhood", but we also have a notion of "infinitesimally approaching a point" in various directions. Then there is scope for specifying how $E_x, E_{x'}$ should be compared, or "smoothly connected" to each other. Such relationships between the various E_x constitute an *extra* piece of geometric data, called a *connection on* E. In physics, connections are often called *gauge fields*. However, beware that this terminology can be confusing, because a connection is not a field in the sense of "section of some bundle over X". Rather, we have to first fix a basis convention for the $E_x, x \in U$ (i.e., refer to a local trivialization), then the connection will be represented as a Lie-algebra valued 1-form on X. This local representation is called a local gauge potential/field in physics. A key point is that the basis convention is unphysical. So the important information in a gauge potential does not lie in its values (which are convention-dependent), but in those properties which are convention-independent, and therefore intrinsic to the connection data.

Example 1.1. In Euclidean 3-space $\mathbf{R}^3$, each point $x \in \mathbf{R}^3$ has a 3-dimensional vector space of "direction vectors" based at x. A vector field is sometimes thought of as a "three-component function" $v = (v_1, v_2, v_3) : \mathbf{R}^3 \to \mathbb{R}^3$. The components $v_i(x)$, however, only make sense when a basis choice $\{e_1(x), e_2(x), e_2(x)\}$ for the vector space attached to x has been specified; then $v(x) = \sum_{i=1}^3 v_i(x) e_i(x)$. We have to do this at every x. This detail is often neglected because Euclidean space comes with an action of the translation group $\mathbb{R}^3$, so a basis choice at one point may be "carried over" to every other point, obtaining a global parallelism.

For a general manifold X, a consistent identification of the tangent vector spaces over all points of X may not even be possible! A familiar example is $X = S^2$ embedded in $\mathbf{R}^3$ (think of the surface of the earth). Each $x \in S^2$ has a 2-dimensional tangent plane $T_x S^2 \subset \mathbb{R}^3$ attached to it. These tangent

planes vary with x. Altogether, the tangent bundle is the subset

$$TS^2 = \bigsqcup_{x \in S^2} T_x S^2 \subset S^2 \times \mathbb{R}^3,$$

and this subset can be shown to be a manifold in its own right. Informally, the tangent spaces $T_x S^2 \subset \mathbb{R}^3$ depend smoothly on x, and the entire collection TS^2 is called a smooth *vector bundle*. However, TS^2 is not diffeomorphic to a Cartesian product $S^2 \times \mathbb{R}^2$ — this is a standard result in differential topology. Consequently, we cannot reduce a tangent vector field over S^2 into a "two-component map" $v : S^2 \to \mathbb{R}^2$, no matter how we try to choose bases for the $T_x S^2$. At best, we can regard v as a "three-component map" $S^2 \to \mathbb{R}^3$ subject to the constraint that

$$v(x) \in T_x S^2, \qquad \forall\, x \in S^2.$$

This latter description is still rather unsatisfactory, because it depends on an extrinsic embedding of S^2 into Euclidean 3-space.

Example 1.2. Consider the unit circle S^1 in the plane, labelled by its angular coordinate $\theta \in [0, 2\pi]/_{0 \sim 2\pi}$. The normal space at $\theta \in S^1$ is the line

$$N_\theta S^1 = \{\lambda(\cos\theta, \sin\theta) \,:\, \lambda \in \mathbb{R}\} \subset \mathbb{R}^2.$$

In total, the normal bundle is the subset

$$NS^1 = \bigsqcup_{\theta \in S^1} N_\theta S^1 \subset S^1 \times \mathbb{R}^2.$$

There is no problem identifying NS^1 with $S^1 \times \mathbb{R}$, via

$$(\theta, \lambda(\cos\theta, \sin\theta)) \mapsto (\theta, \lambda).$$

Essentially, we have chosen the basis vector $(\cos\theta, \sin\theta)$ for $N_\theta S^1 \subset \mathbb{R}^2$, and this choice is well-defined as θ is varied around all of S^1. So we could discuss normal vector fields over S^1 as though they were simply real-valued functions.

Now consider the quotient of S^1 by the antipodal map — this is called the real projective line $\mathbb{RP}^1$, and it is again a circle, but parametrized by $\theta \in [0, \pi]/_{0 \sim \pi}$. Attached to each $\theta \in \mathbb{RP}^1$ is the line $N_\theta S^1 \subset \mathbb{R}^2$ as before, but now we think of $N_\theta S^1$ as the "tautological line" over θ. This assignment

of lines is consistent at $\theta = 0 \sim \pi$, since $N_0 S^1$ and $N_\pi S^1$ are the same line. Altogether we have constructed the "tautological line bundle" over $\mathbb{RP}^1$,

$$\mathcal{L}^{\mathbb{R}} := \bigsqcup_{\theta \in \mathbb{RP}^1} \underbrace{N_\theta S^1}_{\mathcal{L}^{\mathbb{R}}_\theta} \subset \mathbb{RP}^1 \times \mathbb{R}^2.$$

However, we no longer have a globally well-defined choice of *bases* for the lines $\mathcal{L}^{\mathbb{R}}_\theta$ as θ is varied over $\mathbb{RP}^1$. To see this, start with the basis vector $(1,0) \in \mathcal{L}^{\mathbb{R}}_0 = N_0 S^1$. For continuity, we have to choose the basis vector $(\cos\theta, \sin\theta) \in \mathcal{L}^{\mathbb{R}}_\theta = N_\theta S^1$ at $\theta \in [0, \pi)$. So upon reaching $\theta = \pi$, the basis vector becomes $(-1, 0) \in \mathcal{L}^{\mathbb{R}}_\pi$. This has a $-$ sign mismatch compared to the starting choice $(1,0) \in \mathcal{L}^{\mathbb{R}}_0$.

The above-mentioned sign problem is an example of *holonomy*. It is a symptom of the fact that $\mathcal{L}^{\mathbb{R}}$ is *not* diffeomorphic to $\mathbb{RP}^1 \times \mathbb{R}$.

Hopefully, we have now appreciated the difference between a function $X \to \mathbb{C}^n$ and a section $v : X \to E$ of a vector bundle $E \to X$. Only after an identification $E \cong X \times \mathbb{C}^n$ has been made, can a section be considered as a function,

$$\begin{aligned} X \to \quad & E \quad \overset{\cong}{\leftrightarrow} X \times \mathbb{C} \\ x \mapsto & \, v(x) \leftrightarrow (x, f(x)), \qquad\qquad f : X \to \mathbb{C}^n. \end{aligned}$$

So the section becomes the graph of some function $f : X \to \mathbb{C}^n$. To complicate matters further, the examples of TS^2 and $\mathcal{L}^{\mathbb{R}}$ show that such an identification is often impossible in the first place. So in general, we can expect a section to be describable as functions $f : U \to \mathbb{C}^n$ only on local patches U of X.

From the physical perspective, this "promotion" of functions to fields in not only a matter of pedantry, but has fundamental conceptual importance. One expects physical laws to obey locality principles; for example, we do not expect that basis conventions at $E_x, E_{x'}$ are instantaneously identified when x, x' are far apart. The ideas of vector bundles, sections, and connections are consistent with such locality considerations. These allow for manifestly gauge-invariant formulations of the physical laws. Certain quantities may depend (either in their definitions, computation, or for convenience) on references to gauge choices, and we must always remember to pay special attention to those quantities which are *gauge-independent*.

Remark. In the case where $E_x \cong \mathbb{R}$, the freedom to choose a normalized basis vector is labelled by a discrete set $\mathrm{O}(1) \cong \mathbb{Z}_2$. This may be thought of as the choice of orientation on E_x. So there are no "local gauge freedoms" in the sense that continuity in X forces the E_x to be identified with $\mathbb{R}$ in the same way for all x in a local patch of X. This rigidity is a reason why gauge-theory ideas hardly appear when studying classical differential equations governing real-valued scalar systems (temperature, pressure, etc.) Subtleties arise only when the E_x happen to be globally arranged in the manner of $\mathcal{L}^{\mathbb{R}}$. Then, for example, one cannot give global meaning to a quantity like "height above the central circle of a Möbius band".

In contrast, once we have E_x being a complex vector space, then we will naturally encounter the local gauge freedoms. This occurs for the differential equations occurring in quantum mechanics. While at first, these gauge freedoms seem like irritating details, they are now understood to be an indispensable part of modern physics!

1.1 Tangent bundle and other bundles

Historically, the development of general relativity (perhaps better described as "Einstein's theory of gravitation") was motivated by an idea of satisfying a "principle of general covariance". Informally, the laws of physics are "the same" in any "reference frame". What does this mean precisely? I do not have a good answer, but we shall see that the language of differentiable manifolds encodes such an idea.

One of the most important roles of manifolds is that it formalizes calculus (e.g. derivatives, integration) in a coordinate-independent way. Put in another way, a consistent calculus "exists out there in space", whether or not a human mind has decided to write down symbols like $\frac{\partial}{\partial x}$.

A manifold X is locally diffeomorphic to some (open subset of) Euclidean space, and a choice of such a local diffeomorphism provides local coordinates for points on a local patch U of the manifold. There are plenty of choices of local coordinates, but none is canonically given as part of the manifold data. For *functions* $f : X \to Y$ between two manifolds, we have a coordinate-independent notion of the derivative of f — it is a map on the tangent bundles $df : TX \to TY$, taking the tangent space T_xX linearly to the tangent space $T_{f(x)}Y$. Local coordinates for $U \in X$ also lead to the coordinate vector fields

$\frac{\partial}{\partial x^i}$ defined over U. The $\left.\frac{\partial}{\partial x^i}\right|_x, i = 1, \ldots, n$ provide a basis for each of the tangent spaces $T_x X, x \in U$, identifying each of them with $\mathbb{R}^n$. The derivative df_x is then expressed as a matrix of partial derivatives. But df itself, as well as rules like the chain rule, is independent of such choices.

So a smooth manifold X automatically comes with its tangent bundle TX, and tangent vector fields over X are sections of the tangent bundle TX. Tangent vector fields have long been known to arise in nature, the classical electric and magnetic vector fields being basic examples.

For a long time, geometers agonized over the problem of relating tangent vectors attached to different points of X, and therefore the issue of differentiating vector fields. One observation was that of path dependence.

Example 1.3. For example, consider on the unit 2-sphere $X = S^2$, a geodesic triangle (i.e. following great circles) with first vertex on the North pole, and the other two vertices on the equator. Start from a horizontal tangent vector, say $(1, 0, 0)$ at the north pole. Intuitively, we may parallel transport it southwards along a longitude line $y^2 + z^2 = 1$, until we reach the equator. As we move eastwards along the equator, however, $(1, 0, 0)$ no longer lies in the tangent plane at of the base point. Rather, after moving through a polar angle of θ along the equator, the parallel transported tangent vector would becomes $(\cos\theta, \sin\theta, 0)$. Then parallel transport back to the north pole results in the final tangent vector $(\cos\theta, \sin\theta, 0)$.

In the above example, the space of orthonormal frames of E_x is $\mathrm{O}(2)$, and we acquired a *holonomy* of $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{O}(2)$ after completing a loop on S^2 with solid angle θ.

We observe that we have implicitly used the round metric on S^2, inherited from the background Euclidean metric, in the parallel transport prescription. First, the lengths of the tangent vectors are preserved throughout the process. Second, the parallel transported vector maintains the same angle with the tangent to the path. Formally, we have used the parallel transport associated to the canonical *Levi–Civita connection* for the round metric on S^2. When dealing with Riemannian manifolds, this distinguished connection is usually implicitly used to define covariant differentiation of tangent vector fields.

General vector bundles, not canonically arising from the base manifold alone, are harder to find in nature, which perhaps explains why it was quite a conceptual leap to investigate them, and develop a theory of connections on

them. Once we entertain the idea that the valuation spaces E_x may be independent from the smooth structure of X, we may go further and consider the more general notion of a *fibre bundle*. In a fibre bundle, all the E_x are identifiable with a reference object F (the "typical fibre", generally some topological space or manifold), but not canonically so.

Historically ($\sim$ 1918), the first "geometrically unusual" fibre bundle was introduced by Hermann Weyl, and it had $E_x \cong \mathbb{R}_+ = (0, \infty)$ as typical fibre. A positive number in E_x was meant to be a "length scale" at x (hence the terminology "gauge"). The multiplicative group $(\mathbb{R}_{>0}, \times)$ acts as a "change-of-scale" on each E_x. This idea of Weyl was quite radical — he was thinking of a generalization of (pseudo-)Riemannian geometry, where there is a bundle of possible local gauges, without any distinguished global section/length scale. He wanted to relate the $(\mathbb{R}_{>0}, \times)$ freedom in the local gauges with a similar freedom in choosing electromagnetic potentials.

Weyl's idea was rejected by Einstein, Pauli etc. on empirical grounds, and it remained a beautiful piece of pure mathematics. As it turns out, Weyl was ahead of his time. As *quantum theory* developed and increased in mathematical precision, it became apparent that it is the quantum mechanical wave-function, rather than the metric tensor field of general relativity, which should be coupled with the electromagnetic potential in Weyl's theory. The "wave-function" has a U(1)-phase indeterminacy, while the "group of length scales" is $(\mathbb{R}_{>0}, \times)$. These two groups are "infinitesimally the same", in the sense of sharing the same Lie algebra. But the difference in their *global topology* (one is compact, the other is simply-connected) has important consequences.

2 Geometry and gauge in quantum theory

By itself, gauge theory is a very geometrical subject. It involves the differential geometry of fibre bundles and connections. Even in physics, its historical development started out in relation to general relativity (Einstein's geometrization of gravitation), as mentioned earlier. Unfortunately, starting directly with the formal mathematical treatment can greatly obscure the physical motivations; conversely, the usual physical presentation tends to obscure the natural geometrical meaning.

With a lot of hindsight, we shall start the discussion of gauge theoretic ideas from general considerations of the mathematical structure of quantum mechanics. One reason to do this is to emphasize the importance of geometry

in quantum theory from the beginning. Another reason is to avoid giving the impression that "all fundamental physics is geometry". Rather, the worlds of geometry–topology and functional analysis–spectral theory intersect in a very deep way, via operators appearing in quantum theory – this is *index theory*.

Again, due to the historical development of index theory (Atiyah–Singer), it appears to be an esoteric subject which the majority of physicists and mathematicians can safely ignore. This is quite untrue — index theoretic phenomena can already be seen in fairly elementary toy examples, although such a perspective is not easily found in existing textbooks. I particularly highlight the phenomenon of *topological phases of matter* in modern condensed matter physics, as an area which can benefit greatly from a more geometric and universal perspective afforded by index theory.

2.1 "Wavefunctions" in quantum mechanics

In quantum mechanics, a wavefunction ψ ("quantum state") is usually taken to be a function $\psi : X \to \mathbb{C}$ on a measure space (X, μ) with unit L^2-norm. One of the most important feature of general quantum theory is that a wavefunction is not itself observable. Only real-valued *transition probabilities* between quantum states,

$$|\langle\psi|\psi'\rangle|^2 = |\langle\psi|\psi'\rangle_{L^2(X)}|^2 = \Big|\int_X \overline{\psi(x)}\psi'(x)\,d\mu(x)\Big|^2 \tag{2.1}$$

are. This probabilistic interpretation of $|\langle\psi|\psi'\rangle|^2$ is called the *Born rule*. The number $\langle\psi|\psi'\rangle$ is usually called a *transition amplitude*.

The absolute value in Eq. (2.1) tells us that we may independently multiply ψ and ψ' by some U(1) phases,

$$\psi \mapsto \alpha \cdot \psi, \quad \psi' \mapsto \beta \cdot \psi, \qquad \alpha, \beta \in \mathrm{U}(1),$$

without affecting the transition probability. So more accurately, quantum states are really elements of a *projective* Hilbert space.

This redundancy of the global U(1) phase in ψ is a general feature of quantum mechanics. So far, the information of X has not entered yet. Now observe that at each point $x \in X$, we are actually free to multiply the $\psi(x)$ and $\psi'(x)$ by a *common* U(1)-valued phase *function* $u(x)$, without changing the transition amplitude in Eq. (2.1). This follows from the pointwise sesquilinearity of the L^2-inner product. It appears that the phase information contained in ψ,

i.e., $\frac{\psi}{|\psi|} : X \to \mathrm{U}(1)$, is not directly observable. That is, ψ is overdetermined as a complex-valued function.

How should we understand ψ "modulo phase redundancy"? If we make the phase information totally redundant, and work just with $|\psi|$, then we do get a real-valued scalar function: the *probability density* of ψ,

$$\rho_\psi : X \to \mathbb{R}, \qquad \rho_\psi(x) = |\psi(x)|^2,$$

which of course integrates to 1,

$$\int_X |\psi(x)|^2 \, d\mu = ||\psi||^2_{L^2(X)} = 1,$$

as required for a probabilistic interpretation of ψ. Specifically, fix any measurable subset $Z \subset X$. Then the integral $\int_Z \rho_\psi$ is the probability of finding the particle in the set Z, when the particle is in the quantum state ψ. Indeed, for a single state ψ, these probabilities do not require the phase information.

However, the transition amplitude between two states ψ, ψ' involves the *relative* phase information between $\psi(x)$ and $\psi'(x)$,

$$\langle \psi | \psi' \rangle_{L^2(X)} = \int_X \overline{\psi(x)} \psi'(x) \, d\mu(x) \neq \int_X \overline{|\psi(x)|} |\psi'(x)| \, d\mu(x) = \langle |\psi| \, | \, |\psi'| \rangle_{L^2(X)}.$$

Thus, "local U(1) phase information" is not completely redundant. In particular, *relative* phase information is *essential* in quantum mechanics. The redundancy is, more precisely, in a simultaneous redefining all wavefunctions ψ by any phase function $u : X \to \mathrm{U}(1)$,

$$\psi \rightsquigarrow u \cdot \psi.$$

This operation is an abelian gauge transformation.

The upshot is that we should not think of ψ as having complex-number values $\psi(x) \in \mathbb{C}$. Instead, each $\psi(x)$ is an element of a 1-dimensional Hilbert space E_x, where E_x is "attached to x". The E_x does not come with any canonical choice of basis vector, so $\psi(x)$ does not have a canonical phase-value. Nevertheless, at each x, the relative phase between $\psi(x), \psi'(x) \in E_x$ is well-defined, whatever the choice of basis for E_x.

So we are precisely in the situation discussed in Section 1. We have a family $E = \sqcup_{x \in X} E_x$ of 1-dimensional Hilbert spaces, and a wavefunction ψ is really a "section" $\psi : X \to E$ such that $\psi(x) \in E_x$ for all $x \in X$. Formally,

we write $\psi \in L^2(X;E)$. To turn ψ into a complex-number-valued function, we need a reference global trivialization $h : E \xrightarrow{\cong} X \times \mathbb{C}$, i.e. a choice of basis for every E_x. Composing ψ with the trivialization gives the map

$$X \to E \to X \times \mathbb{C}, \qquad x \overset{\psi}{\mapsto} \psi(x) \overset{h}{\mapsto} (x, f(x)),$$

where $f \in L^2(X)$ is now some function, representing ψ with respect to the trivialization h.

Let us now consider the operation of pointwise multiplication by some phase function $u : X \to \mathrm{U}(1)$. This operation gives another possible way of trivializing E,

$$E \xrightarrow{h} X \times \mathbb{C} \xrightarrow{(\mathrm{id},u)} X \times \mathbb{C}.$$

Basically, we are "changing the gauge", and correspondingly, the section ψ is now represented by the modified function $u \cdot f$. Notice that the gauge transformation (multiplying everything pointwise by u) is implemented unitarily on the Hilbert space $L^2(X)$, as any symmetry must be in quantum mechanics.

So we see that transition amplitudes $\langle \psi | \psi' \rangle_{L^2(X)}$ are *gauge-invariant*, thus observable, quantities.

2.1.1 Differentiating quantum states

Usually, X is a smooth manifold. Then we would like $E = \sqcup_{x \in X} E_x$ to be a smooth family of 1-dimensional Hilbert spaces. Formally, we require E to be a manifold, which admits a diffeomorphism $E \cong X \times \mathbb{C}$ that restricts to isomorphisms $E_x \cong \mathbb{C}$. In this case, E is called a smooth, trivializable Hermitian line bundle over X. A particular choice of identification $E \cong X \times \mathbb{C}$ is called a *trivialization* of E, and it is an unphysical extra piece of data.

Now, we need to take derivatives of ψ, e.g., in the Schrödinger differential equation governing the time-evolution of ψ. This requires a prescription for comparing the vector spaces E_x and $E_{x'}$ at infinitesimally nearby points x, x' — a *connection* ∇ on E. This data provides a notion of "covariant differentiation" of sections of E. To describe a connection concretely, we might decide to pick a trivialization $E \cong X \times \mathbb{C}$, and compare ∇ with the "trivial connection" ∇_0 corresponding to the usual derivatives $\frac{\partial}{\partial x^j}$ of functions. Then ∇ becomes the covariant derivatives $\frac{\partial}{\partial x^j} - iA_j$, where A_j are the real-valued components of a "gauge potential". But we must not attach too much meaning to the A_j, since they require reference to a trivialization ("gauge-dependence"). As we

will learn, certain aspects of A_j are nevertheless intrinsic, e.g., its curvature, holonomy, etc.

But what determines the connection on E? The section ψ is, typically, supposed to be the "electron field". As it turns out, A_j is identifiable with the magnetic vector potential, and its curvature ("curl") is identifiable with the magnetic field that the electron is subject to. Now, in classical electromagnetic theory by itself, one learns that the potentials A_j are "unphysical", or "fictitious", or "redundant" data, and only the electric/magnetic fields they determine are physical. But in quantum theory, the A_j have a more important role in specifying connection data for the quantum line bundle E. More to the point, we must be careful about what precise aspect of A_j is essential/physical in quantum theory. The answer is not "the potential A_j is measurable" (literally false); rather, the connection represented by the A_j is quantum mechanically measurable.

Here, it is not just the curvature of the connection which can be measured (this is just the classical magnetic field). More pertinently, the holonomy of the connection ("geometric phase") along a path can be measured, in the so-called Aharonov–Bohm effect, via interference experiments. This put an end to a long confusing debate on the physical status of potentials.

Remark. In introductory courses, it is often said that electromagnetism is the motivating and simplest example of a "U(1)-gauge theory". This is supposed to be due to a "gauge freedom" to modify the potentials,

$$A_j \rightsquigarrow \tilde{A}_j := A_j + \partial_j \Lambda,$$

where Λ is any smooth real-valued function on X, without changing the curl of (A_1, A_2, A_3), i.e., the magnetic field. In this setting, there is certainly a "freedom", but the reason for using the word "gauge" is completely mysterious.

For example, as far as I know, there is nothing which determines whether the A_j are gauge potentials for a U(1) or $(\mathbb{R}_{>0}, \times)$ principal bundle (jumping ahead quite a bit). More plainly, the "local symmetry group" could be either one, as far as I know. As mentioned, Weyl initially thought of using $(\mathbb{R}_{>0}, \times)$. In any case, one just obtains a reformulation of electromagnetism, and the core gauge theory ideas do not really enter.

On the other hand, quantum theory inherently demands that E is associated to a U(1) principal bundle. The fact that the connection on E can be identified with the magnetic potential is an empirical finding.

When we change A_j to $\tilde{A}_j$, thus $\partial_j - iA_j$ is changed to $\partial_j - i\tilde{A}_j$, physicists observed that we can "compensate" by simultaneously replacing

$$\psi \rightsquigarrow \tilde{\psi} := e^{i\Lambda} \cdot \psi. \tag{2.2}$$

What is really happening is: when we apply $e^{i\Lambda}$ to change the gauge, this effects the description of ψ according to Eq. (2.2), and also effects

$$e^{i\Lambda}(\partial_j - iA_j)e^{-i\Lambda} = e^{i\Lambda}e^{-i\Lambda}(\partial_j - i\partial_j\Lambda - iA_j) = \partial_j - i\tilde{A}_j$$

for the operation of taking derivatives. That is, the differentiation of the electron field is also *described* in a modified way, when we change gauge.

Remark. In the seminal paper, Elektron und Gravitation. I. Z. Phys. 56, 330 (1929), H. Weyl proclaims:

...In my opinion the origin and necessity for the electromagnetic field is in the following...From the arbitrariness of the gauge-factor in ψ appears the necessity of introducing the electromagnetic potential...

To paraphrase, *geometric* considerations in *quantum* theory lie at the heart of what is nowadays called the "gauge principle".

Remark. There is another important setting in which connections on vector bundles appear in quantum theory. This is when we have quantum systems *parametrized* by some other manifold M (of control parameters). There is a notion of "adiabatic parallel transport" of quantum states, which leads to the notion of "Berry phase", or "geometric phase" in physics. This was soon clarified by Barry Simon to be the holonomy of a certain connection on a line bundle over M.

All of the above discussion generalizes to *vector bundles* where E_x is n-dimensional. We can also have E_x being real inner product spaces. For instance, the tangent spaces of a Riemannian manifold have such a structure. In the real case, the gauge transformations would be $\mathrm{O}(n)$-valued, and there is often occasion to restrict to $\mathrm{SO}(n)$-valued ones (i.e. orientation-preserving ones).

Finally, the global topology of vector bundles is extremely interesting and important. This studies phenomena that arise when E is not globally trivializable, i.e., no homeomorphism/diffeomorphism $E \cong X \times \mathbb{C}^n$ is available. Only local trivializations over $U \subset X$ exist. Without a global trivialization, there is no notion of globally-defined wavefunctions in the first place! Similarly, the

formulation of gauge transformations, connections, and so on, becomes more subtle.

We saw such global issues in the example of TS^2. Actually, S^2 is a complex manifold, and the tangent spaces T_xS^2 can be thought of as 1-dimensional complex vector spaces. The Dirac monopole involves, roughly speaking, "half of TS^2". Namely, there is another complex line bundle $\mathcal{L}^{\mathbb{C}}$ over S^2, called the Hopf line bundle, which is not diffeomorphic to $S^2 \times \mathbb{C}$. In $\mathcal{L}^{\mathbb{C}}$, we can never have access to genuine wavefunctions defined on all of S^2. We only have "wavesections", which can only be locally turned into functions by means of local trivializations. Similarly, a connection on $\mathcal{L}^{\mathbb{C}}$ can only be locally described by gauge potentials. We will discuss the Dirac monopole and $\mathcal{L}^{\mathbb{C}}$ in more detail later on.

2.2 Time evolution in quantum mechanics

Remark. We will discuss unbounded Hilbert space operators and their spectrum, and some mention of their technicalities is unavoidable. If you are unfamiliar with these (important!) analytic issues, there will be a more detailed discussion later on.

Quantum mechanics on a Riemannian manifold (X, g) is usually done on a Hilbert space such as $L^2(X)$, with the measure induced by the metric g. We suppress the line bundle E for simplicity of discussion. The classical symmetries of (X, g) are the isometries, which actually form a Lie group (a result of Myers–Steenrod). You may think of a Lie group as a "smooth group", which is meant to act on a smooth manifold.

On $L^2(X)$, the isometry group of X has a natural unitary representation via pullback,

$$(T_s \cdot \psi)(x) = \psi(s^{-1} \cdot x), \qquad s \in \mathrm{Isom}(X), \psi \in L^2(X).$$

Symmetry operations, such as the above, should preserve transition amplitude, Eq. (2.1). This is why they must be represented unitarily. For example, we also saw that gauge transformations of multiplication by $u : X \to \mathrm{U}(1)$ are unitary. (Actually, antiunitary symmetries are allowed as well.)

Next, we discuss time evolution. The equation of motion for quantum states ψ is the *Schrödinger "wave" equation*

$$i\frac{\partial}{\partial t}\psi = H\psi,$$

where H is a self-adjoint operator called the *Hamiltonian*, and ψ belongs to its domain (a dense linear subspace of the Hilbert space $L^2(X)$). Self-adjoint operators are infinite-dimensional versions of Hermitian matrices.

For a free particle on X, the Hamiltonian operator would be $H_{\text{free}} = -\nabla^2$, the Laplace operator on X. This is the square of the momentum operator $-i\hbar\frac{d}{dx}$, just like the classical kinetic energy is $\frac{\text{momentum}^2}{2\times\text{mass}}$. Hereafter, we ignore the $\frac{1}{2m}$ factor and the Planck constant $\hbar$ for notational convenience. There are also the *Schrödinger operators* $H = -\nabla^2 + V$, where V is the operator of multiplication by some (real-valued) potential function on X.

There are two major questions in the study of Schrödinger operators. First, does $H = -\nabla^2 + V$ really make sense as a self-adjoint operator? If so, then the spectral theorem applies, $H = \int_{\mathbb{R}} \lambda \, d\mu(\lambda)$, where μ is the *spectral measure*

of H. This is the infinite-dimensional version of unitary diagonalization. If λ is a (real) eigenvalue of H with eigenfunction ψ (in the domain of H), then the Schrödinger equation simplifies to

$$i\frac{d}{dt}\psi = \lambda\psi \;\Rightarrow \psi(x,t) = e^{-i\lambda t}\psi(x).$$

So an eigenfunction evolves in time in a very simple way — it acquires a *dynamical phase factor* which rotates at a rate dependent on its eigenvalue λ (physically, the energy).

The spectrum of the linear operator H generalizes the notion of eigenvalues of matrices. Self-adjointness of H implies that the spectrum of H is a subset of the real line. Furthermore, it allows one to make sense of functions of H (the functional calculus). A self-adjoint H exponentiates to a strongly-continuous 1-parameter family of unitary time-evolution operators,

$$U_t = e^{-iHt}, \qquad t \in \mathbb{R}.$$

Here, strong continuity means that for each fixed state ψ, the map $t \mapsto ||U_t\psi||$ is continuous. But the continuity is not generally uniform in the choice of ψ. Conversely, any strongly-continuous time-evolution group is generated by a unique self-adjoint operator. This correspondence between unitary time-evolution groups and self-adjoint operators is of fundamental importance in quantum mechanics, and is a theorem due to M. Stone.

In principle, knowing the spectral decomposition of H, thus its time-evolution group, amounts to solving the Schrödinger equation for $\psi_t(x) = \psi(x,t)$ for any time t, when the initial $t = 0$ quantum state $\psi_0 = \psi(x,0)$ is given:

$$\psi_t = U_t\psi_0.$$

So the second major question is: what can we say about the spectrum of H? The spectrum is exactly solvable only for some simple toy models (e.g. harmonic oscillator, Landau operator).

General quantum mechanics also allows for the Hamiltonian H to be a first-order differential operator, and this is especially important in *relativistic quantum mechanics*. The key example is the *Dirac operator*, which is roughly a "square root of Laplacian". The Schrödinger equation with H being the Dirac operator (instead of Laplacian) is sometimes called a *Dirac equation*. Importantly, the Dirac operator acts on square-integrable sections of *spinor bundles* (a special kind of vector bundle), rather than line bundles.

3 Dirac operators in one dimension

3.1 The many faces of $\mathbb{R}$

Write $\mathbb{R}$ for the 1-dimensional vector space, and let $\mathbf{R}$ denote the underlying affine space. Once an origin is chosen for $\mathbf{R}$, each point of $\mathbf{R}$ is uniquely labelled by the element $x \in \mathbb{R}$ which translates the origin to that point. So we get a global coordinate function x on the manifold $\mathbf{R}$.

There is also a metric on $\mathbf{R}$ given by the Euclidean distance $|x - x'|$. Then the additive group $(\mathbb{R}, +)$ acts isometrically on the Riemannian manifold $\mathbf{R}$ by translations.

The 1-form dx is translation invariant, orients $\mathbf{R}$, and defines a translation-invariant measure on $\mathbf{R}$ (the Lebesgue measure in this case). When we talk about integrating "functions" f on $\mathbf{R}$, we are actually integrating the 1-form $f\,dx$ over the manifold $\mathbf{R}$.

So $G = (\mathbb{R}, +)$ is a Lie group of orientation-preserving isometries on the oriented Riemannian manifold $(\mathbf{R}, dx^2)$. It is convenient to view the Lie group $(\mathbb{R}, +)$ as a subgroup of invertible matrices,

$$G = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \,:\, s \in \mathbb{R} \right\} \cong (\mathbb{R}, +).$$

Then its Lie algebra $\mathfrak{g}$ is the vector space

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & s \\ 0 & 0 \end{pmatrix} \,:\, s \in \mathbb{R} \right\}$$

equipped with the trivial commutator (Lie bracket). The Lie group and Lie algebra are related by the exponential map $\mathfrak{g} \to G$,

$$\exp \begin{pmatrix} 0 & s \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

Each Lie group element (labelled by s) is represented on $\mathbf{R}$ as the isometry T_s of translation-by-s. Each Lie algebra element $\begin{pmatrix} 0 & s \\ 0 & 0 \end{pmatrix}$ is represented by the *Killing vector field* $s\partial_x$ on $\mathbf{R}$. That is, the derivative operator is the "infinitesimal version" of translations; symbolically,

$$\exp(s\partial_x) = T_s. \tag{3.1}$$

The classical interpretation of Eq. (3.1) is that the Killing vector field $s\partial_x$ exponentiates to an "isometric flow" $T_s : x \mapsto x + s$ of points on the manifold $\mathbf{R}$.

In quantum mechanics, we care about wavefunctions $\psi \in L^2(\mathbb{R})$ rather than individual points. In this setting, the Lie group of translation symmetries is unitarily represented by the operator $(T_s\psi)(x) = \psi(x - s)$. This is called the (left) *regular representation* of $\mathbb{R}$ on $L^2(\mathbb{R})$. As for ∂_x, it is formally a skew-adjoint operator with respect to the L^2-inner product (verify this using integration-by-parts). It is customary to consider the formally *self-adjoint* operator $P = -i\partial_x$ instead. The self-adjoint P is called the *momentum operator* on $\mathbf{R}$, and we have $\exp(-isP) = T_s$.

The point of this discussion is to list down some of the ingredients used in quantum theory. Some of these are

- Number line, spectrum of self-adjoint operator lies in $\mathbb{R}$;
- Vector/affine space, coordinates;
- Riemannian manifold, isometry group;
- Hilbert L^2-space with respect to Riemannian measure;
- Representation of symmetry group (e.g. isometries) and/or Lie algebra.

When working in Euclidean space, the symbol $\mathbb{R}$ appears in all of the above roles, which unfortunately obscures the importance of the geometric data in the setup of a quantum mechanical problem.

3.2 Dirac operators on complete 1-manifolds

3.2.1 Dirac operator on oriented Euclidean line

Fix an origin, and write Q for the "position operator" of multiplication by the coordinate x,

$$(Q \cdot \psi)(x) = x\psi(x), \qquad \psi \in L^2(\mathbb{R}).$$

Of course, Q is an unbounded operator, which would take L^2-functions outside of L^2. So we should actually restrict its domain to the subspace

$$\mathrm{Dom}(Q) = \{\psi \in L^2(\mathbb{R}) \,:\, Q \cdot \psi \in L^2(\mathbb{R})\}.$$

It may be shown that Q is self-adjoint on this domain. The spectrum of Q, denoted $\sigma(Q)$, is the whole real line $\mathbb{R}$. For example, an "eigenfunction" of Q with eigenvalue λ is the delta "function" concentrated at λ.

A simple calculation (integrate by parts) shows that $P = -i\frac{d}{dx}$ is formally self-adjoint on the compactly-supported smooth functions,

$$\langle \psi | P\psi' \rangle_{L^2(X)} = \langle P\psi | \psi' \rangle_{L^2(X)}, \qquad \psi, \psi' \in C_c^\infty(\mathbb{R}).$$

Actually, P is *essentially self-adjoint* on $C_c^\infty(\mathbb{R})$, meaning that its closure (again denoted P) is a genuine self-adjoint operator. The precise domain of self-adjointness is a *Sobolev space* (usually denoted $H^1(\mathbf{R})$), which we will study in the future. We simply mention that non-smooth functions in $L^2(\mathbf{R})$, such as $e^{-|x|}$, are allowed, and the derivative is interpreted as a *weak* derivative (distributional sense).

From now on, we write P for the self-adjoint operator $-i\frac{d}{dx}$. As it turns out, P is the precisely the (right-handed) *Dirac operator* on the Riemannian (spin) manifold $X = \mathbf{R}$. Informally, a Dirac operator is a first order differential operator which squares to the Laplace operator, up to lower order terms. Certainly, $P^2 = -\nabla^2$.

Now, think of P as the quantum mechanical Hamiltonian operator entering the Schrödinger wave equation,

$$i\frac{\partial}{\partial t}\psi = P\psi.$$

The time evolution group generated by P is

$$e^{-itP} = e^{-t\partial_x} = T_t,$$

i.e., the group of translations. So the time evolution of $\psi = \psi(x,t) \equiv \psi_t(x)$ is simply

$$\psi_t = T_t\psi_0.$$

Thus any wavefunction simply propagates to the right at unit speed (the speed of light, if physical units are restored). There is also a left-handed Dirac operator, $-P$, which propagates quantum states to the left.

Remark. This relativistic property was one of the original motivations for Dirac's introduction of the concept of Dirac operators. In 3 spatial + 1 time dimensions, the massive Dirac equation requires ψ to be a *Dirac spinor field* with *four* components.

To obtain the spectrum of P, we can perform a Fourier transform $L^2(\mathbb{R}) \to L^2(\hat{\mathbb{R}})$. The reason for the notation $\hat{\mathbb{R}}$ is because it refers not to the abelian translation group $\mathbb{R}$, but its *Pontryagin dual* abelian group $\hat{\mathbb{R}} = \mathrm{Hom}(\mathbb{R}, \mathrm{U}(1))$, also known as the *momentum space*. Namely, the real number $p \in \hat{\mathbb{R}}$ labels the character (irreducible representation)

$$\chi_p : x \mapsto e^{ipx},$$

and the characters themselves obey a composition law, $\chi_{p_1}\chi_{p_2} = \chi_{p_1+p_2}$. On $\hat{\mathbb{R}}$, there is again the Lebesgue measure, defining the Hilbert space $L^2(\hat{\mathbb{R}})$.

The Fourier transform

$$\begin{aligned} \mathcal{F} : L^2(\mathbb{R}) &\to L^2(\hat{\mathbb{R}}) \\ \psi \mapsto \hat{\psi} \qquad \hat{\psi}(p) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ipx}\psi(x)\,dx \end{aligned}$$

is unitary (we take this for granted for now), and it conjugates P into the operator of multiplication by the coordinate $p \in \hat{\mathbb{R}}$. So the spectrum of P is the whole real line. In fact, the plane wave $x \mapsto e^{ipx}$ is an "eigenfunction" of P corresponding to the "eigenvalue" p. If you know some spectral theory, this is basically the spectral theorem applied to the operator $P = -i\frac{d}{dx}$. The Fourier transform is the unitary "change of basis" which "diagonalizes" P into the operator of multiplication by its "eigenvalues".

Now we have a pair of self-adjoint operators (Q, P), and they are easily seen to satisfy the *canonical commutation relation*

$$[Q, P] = i.$$

A famous theorem of Stone–von Neumann says that any pair of self-adjoint operators Q', P' on a Hilbert space satisfying $[Q', P'] = i$ (more precisely, the exponentiated version $e^{isQ}e^{itP} = e^{ist}e^{itP}e^{isQ}$ called the *Weyl relations*), must be unitarily equivalent to the standard pair (Q, P).

Here is an interesting example: Let

$$Q' = Q, \qquad P' = P + A, \qquad A \in C^\infty(\mathbf{R}; \mathbb{R})$$

where A is (the multiplication operator by) any smooth real-valued[1] "potential function".

[1] If A is bounded, then $P + A$ has the same domain of self-adjointness as P; otherwise a bit more care is needed to understand what $P + A$ means.

Clearly A commutes with Q, so that we still have

$$[Q', P'] \equiv [Q, P + A] = [Q, A] = i.$$

(More precisely, (Q', P') can be shown to satisfy the Weyl relations.) Therefore, by the Stone–von Neumann theorem, there must be a unitary operator U on $L^2(\mathbb{R})$, such that

$$UQU^{-1} = Q' = Q, \qquad UPU^{-1} = P' = P + A.$$

The required U is nothing but the gauge transformation of multiplication by the function

$$u : \mathbb{R} \to \mathrm{U}(1), \qquad u(x) = e^{-i\int_{x_0}^{x} A}.$$

Here, the basepoint x_0 is arbitrary, and only leads to an overall phase factor in u. You may recognize u as the integrating factor for the first-order differential operator $-i\frac{d}{dx} + A(x)$.

In the reverse direction, suppose U is the gauge transformation by a U(1)-valued function $u = e^{-if}$ with smooth real-valued f. Write $A = f'$ for the derivative of f. Then U will conjugate $P = -i\frac{d}{dx}$ into the operator

$$UPU^{-1} = U(-i\frac{d}{dx})U^{-1} = -i(\frac{d}{dx} + iA) =: -i\nabla_A = P + A.$$

Conclusion. Any of the operators $P + A$ represents "the" Dirac operator on the Euclidean line $\mathbf{R}$. We say that A is "pure gauge" in the sense that it can be set to zero by working in an appropriate gauge.

3.2.2 Dirac operators on a circle

The Euclidean line $\mathbf{R}$ is topologically trivial in the sense that it is contractible to a point. This property allows any $A\,dx$ to be expressible as $A = df$ for some well-defined smooth function f on $\mathbf{R}$, thereby giving the u which gauge-transforms it to zero.

Let us now work over a compact 1-dimensional Riemannian manifold X. Parametrizing by path-length, we see that X is actually isometric to a circle of some radius r in the Euclidean plane. Without loss of generality, we study the momentum operator $P = -i\frac{d}{d\theta}$ on the unit circle $X = S^1$, oriented counter-clockwise, and parametrized by $\theta \in [0, 2\pi]/_{0\sim 2\pi}$.

The operator P is self-adjoint on the Sobolev space $H^1(S^1)$, which we can think of as comprising those functions ψ whose Fourier transform $(\psi_n)_{n\in\mathbb{Z}}$ is such that $(n\psi_n)_{n\in\mathbb{Z}}$ is square-summable. The eigenvalues and eigenfunctions of P are easily verified to be

$$\psi_n : e^{i\theta} \mapsto e^{in\theta}, \qquad P\psi_n = n\psi_n.$$

So the spectrum of P is $\sigma(P) = \mathbb{Z}$.

So far the story seems quite elementary and uninteresting. However, now observe that in declaring the initial domain as the *function* space $C^\infty(S^1)$, we have committed to a common choice of basis for the copy of $\mathbb{C}$ above each $e^{i\theta} \in S^1$. For $m \in \mathbb{Z}$, let us perform the gauge transformation of multiplication by the smooth function $u_m : e^{i\theta} \mapsto e^{-im\theta}$. This converts P to the operator

$$P^{(m)} := u_m P u_m^{-1} = u_m(-i\frac{d}{d\theta})u_m^{-1} = -i\frac{d}{d\theta} + m = P + m.$$

The eigenfunctions of $P^{(m)}$ are still φ_n, but the eigenvalue of φ_n is now shifted to $n + m$. We still have $\sigma(P^{(m)}) = \mathbb{Z}$, independently of the value of m, which is consistent with the unitarity of the gauge transformation. Therefore any integer potential term $m \in \mathbb{Z}$ is "pure gauge", and can be transformed away.

What about the operator $P^{(\xi)} := -i\frac{d}{d\theta} + \xi$ for a potential term $\xi \in (0, 1)$? Obviously, the spectrum becomes shifted by ξ,

$$\sigma(P^{(\xi)}) = \mathbb{Z} + \xi \neq \sigma(P).$$

So P_ξ must be unitarily *in*equivalent (in particular, gauge-inequivalent) to P. In fact, the operators $P_\xi, \xi \in [0, 1)$ must all be mutually gauge-inequivalent! Here, the real value of ξ is not gauge-invariant due to the possibility of modifying it by an integer m through the gauge transformation u_m. But $e^{2\pi i\xi} \in \mathrm{U}(1)$ (equivalently, ξ mod $\mathbb{Z}$) is gauge-invariant (check this!); it is the holonomy of a U(1)-connection on the line bundle $S^1 \times \mathbb{C}$.

So we see that there are many gauge-equivalence classes of connections on the same (trivial) line bundle. Each such connection defines a *twisted Dirac operator* $P + \xi$. Incidentally, the family $\{P + \xi\}_{\xi\in[0,1]}$ provides an example of *spectral flow*, see Fig. 1, about which we will have much more to say.

3.2.3 Aharonov–Bohm effect

Now, $\xi\, d\theta$ is supposed to be the magnetic potential 1-form (The difference and relation between vector fields and differential 1-forms will be clarified in later

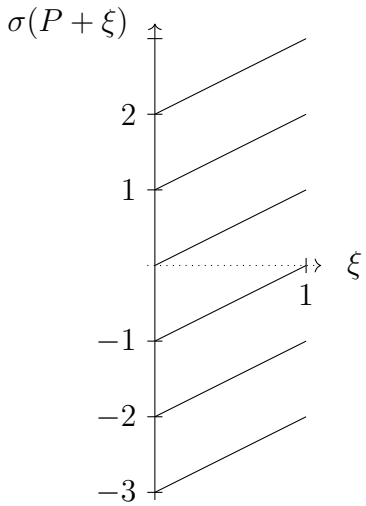

Figure 1: Spectral flow of family of Dirac operators on S^1, as the holonomy parameter $\xi \in [0,1]/_{0\sim 1}$ is increased in a loop.

lectures). Here, we are picturing S^1 to lie on the horizontal plane of $\mathbb{R}^3$, and we work in cylindrical coordinates (r, θ, z). This gauge potential has vanishing curvature, $d(\xi\, d\theta)) = 0$, corresponding to zero magnetic field. *A key point is that $d\theta$ is only defined on $\mathbb{R}^3$ with the z-axis removed*, so the vanishing of the magnetic field holds only for the region $r > 0$.

The operator P_ξ is the Hamiltonian operator for a Dirac particle on a circle, subject to zero magnetic field. Not quite! Remember that the relevant data for the quantum theory is not the magnetic field, but the connection, or alternatively, the gauge potential $\xi\, d\theta$. The magnetic field just happens to be one of the intrinsic features of the connection (namely, its curvature). So P_θ is the Dirac operator on the circle with respect to this connection data.

Still, there is another point of confusion. How is this situation different from that of the Euclidean line, where the gauge potential $A\, dx$ could always be transformed away by an appropriate gauge transformation? After all, isn't $d\theta$ precisely the differential of the coordinate function θ on the circle? Why can't we just apply the gauge transformation $u_\xi(\theta) = e^{-i\xi\theta}$ as before?

We come to the second key point: *θ is not a genuine globally defined function* on all of S^1. The *topology* of S^1, specifically the failure of simply-connectedness, prevents us writing $d\theta$ as the derivative of a function on S^1 —

the gauge potential $\xi d\theta$ is not an *exact* differential form. The candidate "gauge transformation" u_ξ is necessarily discontinuous at one point. Multiplication by u_ξ is a unitary transformation, so it would seem that conjugating P by u_ξ should still produce "$P + \xi$". This is not true: we must be extremely careful with the domains of self-adjointness for *unbounded* operators on Hilbert space, as explained in Section 3.3.4.

Remark. In cartesian coordinates, the 1-form $\xi\, d\theta$ is (exercise)

$$\xi\, d\theta = \frac{\xi}{x^2 + y^2}(-y\, dx + x\, dy),$$

which corresponds to the vector potential $\mathbf{A}(x, y, z) = \frac{\xi}{x^2+y^2}(-y, x, 0)$. The curl, which is the magnetic vector field, is easily seen to be $\mathbf{B} = \nabla \wedge \mathbf{A} = 0$ (away from z-axis). In practice, we consider a very thin cylinder around the z axis, of small radius $r_0 \ll 1$, say. By Stokes' theorem, the line integral of $\mathbf{A}$ along any loop $\mathcal{C}$ around the thin cylinder is

$$\oint_{\mathcal{C}} \mathbf{A} \cdot d\mathbf{x} = \oint_{S^1} \mathbf{A} \cdot d\mathbf{x} = 2\pi\xi = \int_{B_1} \mathbf{B} \cdot d\mathbf{S} = \int_{B_{r_0}} \mathbf{B} \cdot d\mathbf{S},$$

where the last surface integral is the *magnetic flux* through the cross section of the cylinder. This is because the support of the magnetic field is contained in this cylinder. Classically, the value of $\xi \in \mathbb{R}$ is relevant and measurable only inside this cylinder, via the magnetic flux.

The term $\xi\, d\theta$ has a different status in quantum theory — as a gauge potential representing the U(1)-phase connection data. Away from the small cylinder, this gauge potential is non-vanishing, and *intrinsically* so — there is no U(1)-valued gauge transformation which can get rid of $\xi\, d\theta$.

Consider the following operator on S^1,

$$D_\xi = \begin{pmatrix} P + \xi & 0 \\ 0 & -(P + \xi) \end{pmatrix} = \begin{pmatrix} -i\frac{d}{d\theta} + \xi & 0 \\ 0 & i\frac{d}{d\theta} - \xi \end{pmatrix}.$$

This is a toy Hamiltonian operator which will illustrate the so-called Aharonov–Bohm effect.

The self-adjoint operator D_ξ governs the time-evolution of a two-component wavefunction $\psi = \binom{\psi_1}{\psi_2}$. Let the initial wave function be

$$\psi(\theta; 0) \equiv \psi(\theta; t = 0) = \begin{pmatrix} \psi_1(\theta; 0) \\ \psi_2(\theta; 0) \end{pmatrix},$$

where $\theta \mapsto \psi_i(\theta; 0)$ is some localized function on S^1, with center θ_0, say. According to the Schrödinger equation with Hamiltonian D_ξ, the first component ψ_1 propagates anticlockwise, evolving in time as (Exercise)

$$\psi_1(\theta; t) = e^{i \int_\theta^{\theta - t} \xi \, d\theta} \psi_1(\theta - t; 0) = e^{-it\xi} \psi_1(\theta - t; 0).$$

Similarly, ψ_2 evolves clockwise as

$$\psi_2(\theta; t) = e^{it\xi} \psi_1(\theta + t; 0).$$

So after time $t = \pi$, the localized ψ_1, ψ_2 meet again at the antipodal point $\theta_0 + \pi$. The total wavefunction at $t = \pi$ is

$$\psi(\theta; \pi) = \begin{pmatrix} e^{-\pi i \xi} \psi_1(\theta - \pi; 0) \\ e^{\pi i \xi} \psi_2(\theta + \pi, 0) \end{pmatrix} = e^{-i\pi\xi} \begin{pmatrix} \psi_1(\theta - \pi; 0) \\ e^{2\pi i \xi} \psi_2(\theta - \pi; 0) \end{pmatrix},$$

and is centered at $\theta_0 + \pi$. This final wavefunction is almost the same as the original one, except for a *relative* phase factor $e^{2\pi i \xi}$ between the two components (and the propagation from θ_0 to $\theta + \pi$ of course). Such a relative phase is observable in interference experiments. Geometrically, this phase factor is the holonomy of the connection 1-form $\xi \, d\theta$ around S^1.

Remark. The Aharonov–Bohm phase shift can also be derived for Schrödinger particles, but it is somewhat harder to make analytically precise without some handwaving.

In any case, the significance of the effect is this: despite the classical magnetic field being zero everywhere along the path being sampled by the wavefunction, the magnetic potential, correctly understood as a gauge potential, is physically meaningful in *quantum* mechanics. Furthermore, there are many "inequivalent quantum theories" on S^1, despite each of them being classically the same (zero magnetic field).

3.2.4 Anomaly

Returning to the real line $\mathbf{R}$, a simple consequence of $P + A$ being unitarily equivalent to P, is that $P + A$ always has spectrum being the entire $\mathbb{R}$. Similarly, there is a "spectrally rigid" left-handed Dirac operator, represented as $-(P + A)$. These two Dirac operators are sometimes called *chiral Dirac operators*, or *Weyl operators*.

On its own, a chiral Dirac operator becomes problematic in *quantum field theory*. This is a very complicated subject, so let us take a simplified perspective without all that baggage. Now, we have seen that we should think of $P + A$ as "the same operator", regardless of A, since A is really an artifact of gauge choices. For simplicity of discussion, take A to be a real constant function. Then replacing P by $P + A$ shifts the energy by A, so we would appear to lose any canonical meaning of "0-energy". In quantum field theory, or "second quantization", this leads to the problem of defining the "vacuum" in a gauge-invariant way.

Now consider the direct sum, $\not{D}_A = (P + A) \oplus -(P + A)$, of a left-handed and a right-handed Dirac operator, which will still have full spectrum in $\mathbb{R}$. However, the spectral gaplessness is no longer rigid. Specifically, we can add an off-diagonal "mass term",

$$\begin{aligned} \not{D}_A + M := \begin{pmatrix} P + A & m \\ m & -(P + A) \end{pmatrix} &\xrightarrow{\text{gauge}} \begin{pmatrix} P & m \\ m & -P \end{pmatrix} \\ &\xrightarrow{\text{Fourier}} \underbrace{\begin{pmatrix} p & m \\ m & -p \end{pmatrix}}_{\text{eigenvalues } \pm\sqrt{p^2+m^2}}, \qquad m \in \mathbb{R},\ p \in \hat{\mathbb{R}}. \end{aligned}$$

So the spectrum of $\not{D}_A + M$ is

$$\sigma(\not{D}_A + M) = \bigcup_{p \in \hat{\mathbb{R}}} \{-\sqrt{p^2 + m^2}, +\sqrt{p^2 + m^2}\} = (-\infty, -|m|] \ \cup \ [|m|, \infty).$$

Independently of A, a spectral gap $(-|m|, |m|)$ opens up as long as $m \neq 0$. The "vacuum level" can now be set to 0 without ambiguity. The operator $\not{D}_A + M$ is called the *massive Dirac operator* with mass m.

3.3 Dirac operators on incomplete 1-manifolds

3.3.1 Momentum operator on a half-line?

Consider the half-Euclidean line $\mathbb{R}_+ = (0, \infty)$. Note that $\mathbb{R}_+$ is diffeomorphic to $\mathbb{R}$, but they are not isometric. A crucial difference between $\mathbb{R}$ and $\mathbb{R}_+$ is that the former is complete, but the latter is not. This means that geodesics in $\mathbb{R}_+$ do not extend indefinitely, but will "hit the boundary" $x = 0$.

This incompleteness causes severe problems for the definition of a self-adjoint momentum operator on $L^2(\mathbb{R}_+)$. In a naïve construction, we will start with $P = -i\frac{d}{dx}$ defined on $C_c^\infty(0, \infty)$, then extend to a suitable domain of self-adjointness. Suppose this is possible, then we would obtain a unitary time-evolution group $U_t = e^{-iPt}$ on $L^2(\mathbb{R}_+)$. This time-evolution applies, in particular, to $\psi \in C_c^\infty(0, \infty)$, and we know that U_t will just translate such a ψ by a distance t. However, for $t \ll 0$, we would have ψ hitting the boundary and "disappearing", violating the unitary and hence the conservation of probability. So, the assumption that P could be made self-adjoint must be incorrect! Let us see what the problem is.

First, a linear operator L with domain $\mathrm{Dom}(L)$ is *formally self-adjoint* (or *symmetric*), if

$$\langle L\varphi|\psi\rangle = \langle \varphi|L\psi\rangle, \qquad \forall\, \varphi, \psi \in \mathrm{Dom}(L).$$

For example, $P = -i\frac{d}{dx}$ with $\mathrm{Dom}(P) = C_c^\infty(\mathbb{R}_+)$ is formally self-adjoint. As we will learn later, this is not enough to guarantee that its spectrum is real.

Generally, the domain of the adjoint operator L^* is defined to be the set of $\varphi \in L^2(\mathbb{R}_+)$ such that there exists some $\eta \in L^2(\mathbb{R}_+)$ with

$$\langle \eta|\psi\rangle = \langle \varphi|L\psi\rangle, \qquad \forall \psi \in C_c^\infty(\mathbb{R}_+) = \mathrm{Dom}(L).$$

Then for such φ, we define $L^*\varphi = \eta$. This is the precise meaning of the equation defining the adjoint operator,

$$\langle L^*\varphi|\psi\rangle = \langle \varphi|L\psi\rangle, \qquad \forall \psi \in \mathrm{Dom}(L),\ \varphi \in \mathrm{Dom}(L^*).$$

Notice that the larger the domain of L, the smaller the domain of L^*.

Now take L to be the formally self-adjoint P. Then P^* will be an *extension* of P to a larger domain. For example, $\mathrm{Dom}(P^*)$ contains functions $\varphi \in C_c([0, \infty))$ *without* Dirichlet condition at $x = 0$:

$$\langle P^*\varphi|\psi\rangle = \int_0^\infty \overline{-i\varphi'(x)}\psi(x) = -\int_0^\infty i\overline{\varphi(x)}\psi'(x) = \langle \varphi|P\psi\rangle, \qquad \forall \psi \in C_c^\infty(\mathbb{R}_+).$$

No boundary term appears in the integration-by-parts because $\psi \in \mathrm{Dom}(P)$ always has $\psi(0) = 0$.

On the other hand, P^* has no chance of being formally self-adjoint: take any $\psi \in \mathrm{Dom}(P^*)$ with $\psi(0) \neq 0$. Then a short calculation shows that

$$\langle P^*\psi|\psi\rangle - \langle\psi|P^*\psi\rangle = -i|\psi(0)|^2 \neq 0.$$

In fact, this calculation shows that we cannot enlarge $\mathrm{Dom}(P)$ by any non-Dirichlet ψ, otherwise formal self-adjointness will immediately be violated.

There is an impasse: P is forced to have Dirichlet domain, so its adjoint will have non-Dirichlet functions. So there is simply no way of picking domains such that $P = P^*$.

Let us stress that $P \neq P^*$ at the level of unequal domains, is not merely a technical issue. For example, $\varphi(x) = e^{-x} \in \mathrm{Dom}(P^*)$, and is clearly it is an eigenfunction,

$$P^*\varphi = i\varphi,$$

with *imaginary* eigenvalue!

3.3.2 Dirac operators on half-line?

The situation can be rectified by taking

$$\not{D} = \begin{pmatrix} P & 0 \\ 0 & -P \end{pmatrix}, \qquad \mathrm{Dom}(\not{D}) = C_c^\infty(\mathbb{R}_+) \oplus C_c^\infty(\mathbb{R}_+).$$

As before, the adjoint operator $\not{D}^*$ has domain without the Dirichlet condition. This domain is too big for formal self-adjointness to hold, so we need to cut it down. Let us only allow $\psi = \binom{\psi_1}{\psi_2}$ which satisfy a boundary condition,

$$\psi_2(0) = \alpha\psi_1(0), \qquad \alpha \in \mathrm{U}(1).$$

Intuitively, α^{-1} is the phase-shift encountered when a left-moving ψ_1 is reflected at $x = 0$ and turned into a right-moving ψ_2. Indeed, imposing this boundary condition and completing the function space to the Sobolev space $H^1(\mathbb{R}_+)$ (subject to boundary condition) gives a genuine self-adjoint operator.

The conclusion is that there is a *family* of self-adjoint $\not{D}_\alpha = P \oplus -P$, parametrized by $\alpha \in \mathrm{U}(1)$. Let us mention that this *relative* phase shift parameter α is gauge-invariant, so we could have replaced P by any $P + A$ in the above discussion.

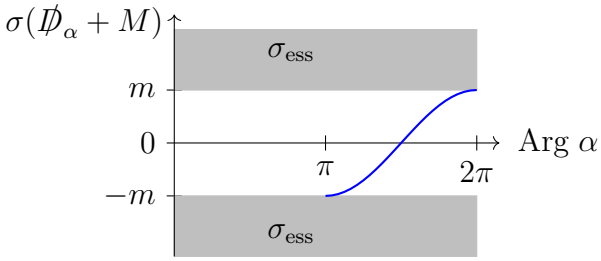

Figure 2: Spectral flow phenomenon for massive Dirac operators on a half-line, parametrized by reflection boundary condition $\alpha \in \mathrm{U}(1)$.

In a later lecture, we will learn about the rigorous and systematic theory of self-adjoint extensions, following von Neumann. Our more immediate task is to figure out what the spectral meaning of the parameter α is (if any).

3.3.3 Massive Dirac operators on half-line

We also have the *massive Dirac operators* on the half-line,

$$\not{D}_\alpha + M = \begin{pmatrix} -i\frac{d}{dx} & m \\ m & i\frac{d}{dx} \end{pmatrix}, \qquad \mathrm{Dom}(\not{D}_\alpha + M) = \{\psi \in H^1(\mathbb{R}_+)^{\oplus 2} \,:\, \psi_2(0) = \alpha\psi_1(0)\},$$

As with the operator $\not{D} + M$ on the full line $\mathbf{R}$, it may be shown that the spectrum of $\not{D}_\alpha + M$ on $\mathbb{R}_+$ still has an "essential part" having a spectral gap,

$$\sigma_{\text{ess}}(\not{D}_\alpha + M) = (-\infty, -|m|] \cup [|m|, \infty).$$

However, extra discrete eigenvalues may now appear within the interval $(-|m|, |m|)$.

An explicit calculation reveals that for certain values of α, an eigenvalue will appear, and as α is varied, this eigenvalue flows from the negative essential spectrum to the positive essential spectrum, see Fig. 2 (Exercise). This is an example of *spectral flow* of eigenvalues, as the parameter α is varied around a loop once.

Generally speaking, non-trivial spectral flow of eigenvalues is only possible if there is an "infinite number" of negative spectra and positive spectra as "sources" and "sinks" for discrete eigenvalues. At the beginning and end of the loop, there are "infinitely many negative energy states". Yet, we saw that one negative-energy state is "lost" during the process, yielding the mysterious equation

$$\infty_{\text{before}} - \infty_{\text{after}} = 1.$$

Much of *index theory* is concerned with making sense of $\infty - \infty$. In fact, we will learn that spectral flow is *topological* — this makes it very resistant to perturbations.

3.3.4 Momentum/Dirac operator on a bounded interval

Let us now consider X to be a bounded interval, say $X = (0, 2\pi)$. The initial domain for $-i\frac{d}{dx}$ would be $C_c^\infty(0, 2\pi)$, i.e., Dirichlet conditions at $x = 0$ and $x = 1$ are imposed. As before, the adjoint $(-i\frac{d}{dx})^*$ will have a larger domain comprising non-Dirichlet functions. We make a compromise and consider the following domain

$$\{\psi \in C^\infty(0, 2\pi) \,:\, \psi(2\pi) = e^{2\pi i\xi}\psi(0)\}, \qquad e^{2\pi i\xi} \in \mathrm{U}(1), \tag{3.2}$$

labelled by a *quasiperiodicity* parameter $e^{2\pi i\xi}$. It can be shown that the closure of $-i\frac{d}{dx}$ on such a domain is a self-adjoint operator.

Intuitively, "what leaves $x = 2\pi$ returns at $x = 0$", with a specified phase shift. This is consistent with unitary time evolution. In physics, this sort of condition is sometimes called a *twisted* boundary condition, or *Bloch wave condition*, and occurs in the so-called Floquet–Bloch theory of periodic Schrödinger operators.

Relation to Dirac operators on circle. Recall the Dirac operators $P^{(\xi)} = -i\frac{d}{d\theta} + \xi$ on the circle S^1, where $\xi \in \mathbb{R}$. For $\xi \notin \mathbb{Z}$, we contemplated applying the "singular gauge transformation",

$$u_\xi : e^{i\theta} \mapsto e^{-i\xi\theta}, \qquad \xi \in \mathbb{R} \setminus \mathbb{Z},$$

to convert P to $P + \xi$. Here, u_ξ is discontinuous, but it is still well-defined as a bounded measurable function, and it does act as a unitary operator on $L^2(S^1)$. Naïvely, we might expect

$$P^{(\xi)} := u_\xi P u_\xi^{-1} \overset{?}{=} P + \xi \;\; \Rightarrow \;\; \sigma(P^{(\xi)}) = \xi + \mathbb{Z} \overset{\text{contradiction}}{\neq} \mathbb{Z} = \sigma(P).$$

What went wrong?

The key point is that the discontinuous u_ξ does not preserve the domain $C^\infty(S^1)$. So $P \equiv P^{(0)}$ is actually unitarily equivalent to the operator

$$P^{(\xi)} = u_\xi P u_\xi^{-1}, \qquad \mathrm{Dom}(P^{(\xi)}) = u_\xi \cdot C^\infty(S^1) \neq C^\infty(S^1),$$

acting on a domain of *discontinous* functions. On the other hand, $P + \xi$ is a genuine differential operator,

$$P + \xi = -i\frac{d}{d\theta} + \xi, \qquad \mathrm{Dom}(P + \xi) = \mathrm{Dom}(P) = C^\infty(S^1),$$

on the smooth functions on S^1. So

$$P + \xi \neq P^{(\xi)} \cong_{u_\xi} P.$$

That is, $P + \xi$ and P are actually *not* unitarily gauge-equivalent, unless ξ happens to be an integer.

Once we introduce a point of discontinuity, we might as well disconnect the circle at $\theta = 0$, and work over the interval $(0, 2\pi)$. Thus we write

$$\tilde{P}^{(\xi)} := -i\frac{d}{dx}, \qquad \mathrm{Dom}(\tilde{P}^{(\xi)}) = \{\psi \in C^\infty(0, 2\pi) \,:\, \psi(2\pi) = e^{2\pi i\xi}\psi(0)\},$$

These are the quasiperiodicity conditions that we saw in Eq. (3.2). Now the gauge transformation u_ξ is a genuinely smooth map $(0, 2\pi) \ni \theta \mapsto e^{-i\xi\theta} \in \mathrm{U}(1)$, and it implements

$$u_\xi \tilde{P}^{(\xi)} u_\xi^{-1} = -i\frac{d}{dx} + \xi, \qquad \mathrm{Dom}(u_\xi \tilde{P}^{(\xi)} u_\xi^{-1}) = \{\psi \in C^\infty(0, 2\pi) \,:\, \psi(2\pi) = \psi(0)\}. \tag{3.3}$$

That is,

$$u_\xi \tilde{P}^{(\xi)} u_\xi^{-1} = \tilde{P}^{(0)} + \xi,$$

where $\tilde{P}^{(0)}$ has *periodic* boundary conditions. Therefore,

$$\sigma(\tilde{P}^{(\xi)}) = \sigma(\tilde{P}^{(0)} + \xi) = \xi + \mathbb{Z}.$$

Since $\tilde{P}^{(0)} + \xi$ acts on periodic functions, we may identify it with the Dirac operator $P + \xi$ on the circle.

3.3.5 Gauge-invariance of quasiperiodicity?

The "non-local" nature of a quasiperiodic boundary condition should make one rather nervous. After all, it involves comparing *values* at two different positions, which may be extremely far apart!

Indeed, the quasiperiodicity label $e^{2\pi i\xi}$ is *not* gauge invariant; a gauge transformation $u \equiv u(x)$ will modify it by an extra $\frac{u(2\pi)}{u(0)}$ factor. For example, consider $\tilde{P}^{(\xi)} + A$ for some smooth real-valued A on $(0, 2\pi)$. Define $\xi_A \in \mathbb{R}$ by $e^{2\pi i\xi_A} = e^{\int_0^{2\pi} A}$. Then a simple calculation shows that

$$\tilde{P}^{(\xi)} + A \xrightarrow{e^{i\int_0^x A}} \tilde{P}^{(\xi+\xi_A)} \xrightarrow{u_{\xi+\xi_A}} \tilde{P}^{(0)} + \xi + \xi_A. \tag{3.4}$$

The three operators in Eq. (3.4) are different representations of "*the Dirac operator on* $(0, 2\pi)$ *coupled to a connection with holonomy* $e^{2\pi i(\xi+\xi_A)}$", with respect to different gauge choices. Only the combined phase $e^{2\pi i(\xi+\xi_A)}$ is gauge invariant.

Remark (Optional). What about the *quasimomentum* label in solid-state physics (surely a meaningful quantity)? Is it actually the same thing as quasiperiodicity? (See the notion of magnetic translations in the Assignment 1.)

Remark. Altogether, there is a circle's worth of gauge-inequivalent twisted Dirac operators S^1 (or on $(0, 2\pi)$), labelled by $e^{2\pi i\xi} \in \mathrm{U}(1)$. The space $\mathbb{R}/\mathbb{Z}$ is called the the *moduli space* of (flat) connections on S^1. The moduli space and the manifold S^1 are both circles, but do not mix them up! There are many occasions where one has to consider a family of problems parametrized by the moduli space. (In Floquet–Bloch theory, the so-called Brillouin zone is basically the moduli space, and S^1 is the unit cell.)

3.4 Dirac operators and SUSY quantum mechanics

The *simple harmonic oscillator* Hamiltonian is

$$H_{\mathrm{SHO}} = -\frac{d^2}{dx^2} + x^2.$$

Let us consider the first-order operator

$$D_{\mathrm{SHO}} = \begin{pmatrix} 0 & -\frac{d}{dx} + x \\ \frac{d}{dx} + x & 0 \end{pmatrix} = \begin{pmatrix} 0 & a^\dagger \\ a & 0 \end{pmatrix}.$$

Now observe that

$$D_{\mathrm{SHO}}^2 = \begin{pmatrix} a^\dagger a & 0 \\ 0 & aa^\dagger \end{pmatrix} = \begin{pmatrix} H_{\mathrm{SHO}} - 1 & 0 \\ 0 & H_{\mathrm{SHO}} + 1 \end{pmatrix} \tag{3.5}$$

holds on $C_0^\infty(\mathbb{R})$, or even the Schwartz functions $\mathscr{S}(\mathbb{R})$ (smooth functions whose derivatives decay faster than any inverse power of $|x|$). It can be shown [17] that D_{SHO} and H_{SHO} are essentially self-adjoint on $\mathscr{S}(\mathbb{R})$. We use the same symbols for their closures to self-adjoint operators, and write a^* instead of $a^\dagger$ for the genuine (rather than formal) adjoint of a.

The operators a^*a and aa^* are positive-definite, so their spectrum is a subset of $\mathbb{R}_{\geq 0}$. Quite generally, the non-zero spectrum of a^*a and aa^* coincide (this is the key lemma in supersymmetric QM). From Eq. (3.5), we even have $aa^* = a^*a + 2$, so $aa^* \geq 2$. So the interval $(0, 2)$ is a spectral gap for aa^*, thus also for a^*a. Then $(2, 4)$ is a spectral gap for $a^*a + 2 = aa^*$. By induction, the spectrum of a^*a lies within the discrete set $2\mathbb{N}$, so it comprises eigenvalues. The operator a^* is a "raising operator" in the sense that if ψ_n is an eigenfunction of a^*a with eigenvalue $2n$, then

$$(a^*a)(a^*\psi_n) = a^*(aa^*\psi_n) = a^*(a^*a\psi_n + 2\psi_n) = a^*((2n+2)\psi_n) = 2(n+1)(a^*\psi_n).$$

It is easy to check that $a^*\psi_n$ is nonzero. Thus $a^*\psi_n$ is an eigenfunction of a^*a with eigenvalue $2(n + 1)$.

Of course, we need an initial eigenfunction to start off a non-vacuous chain of eigenfunctions. The lowest possible eigenvalue is 0, and it is evident that the kernel of a^*a is spanned by the Gaussian $\psi_0(x) = e^{-x^2/2}$. This is the well-known ground state of H_{SHO}. We see that the lowest eigenspace of H_{SHO} is the same thing as the kernel a^*a, which is itself the kernel of the Dirac-type operator D_{SHO}.

In the language of supersymmetric quantum mechanics, the odd operator D_{SHO} is a *supercharge* for the positive operator Eq. (3.5). The latter has two components (a "bosonic" and a "fermionic" one), and its ground state is intimately related to the bosonic/fermionic kernel of the supercharge. A remarkable point is that the *supersymmetric kernel*, meaning the difference in dimensions of the bosonic and fermionic kernels, is very stable — mathematically, this is the stability of the *index*.

3.5 Outlook for higher-dimensional manifolds

1D manifolds X have the special property that any connection (on any line/vector bundle over X) is *flat*. Since there is no curvature, all connections are the same at the classical level (where curvature $\sim$ magnetic field strength). Yet there

can be gauge-inequivalent connections when X is topologically non-trivial, with quantum mechanical consequences (e.g. AB-effect).

For 2D manifolds X, the possibility of *curved* connections arises. Even if X is a boring contractible space like $\mathbb{R}^2$, curvature of the connection has dramatic consequences for the spectrum of Dirac/Laplace operators on X. The prime example of this is the *Landau quantization* of the magnetic Laplacian spectrum, which is key to the famous *quantum Hall effect* (several Nobel prizes).

If $X = S^2$, it is possible that the line bundle over S^2 is topologically non-trivializable (much like the tangent bundle of S^2), a famous example being Dirac's *magnetic monopole*. In this case, there is simply no way to present a section ψ as a global defined wavefunction, and the gauge potentials A are at best *locally defined* terms contingent on *local gauge choices*.

This concludes our informal discussion of gauge-theoretic matters in quantum theory. To fully understand these matters rigorously, we need to develop the theory of fibre bundles, connections and curvature.

4 Smooth manifolds

We start with a working definition of a smooth manifold: it is a subspace of some Euclidean space which, near each of its points, is locally diffeomorphic to (some open subset) of $\mathbb{R}^n$. Caveat: How the manifold is embedded in a Euclidean space, is *not* part of its data!

Differential calculus of maps between manifolds can be set up, by working locally and referring to the usual multivariable calculus in Euclidean spaces. Integral calculus also makes sense, by invoking "partition-of-unity" and summing up local results.

For functions $f : \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces, partial derivatives $\frac{\partial}{\partial x^i}$ are defined once we pick some coordinates (e.g. Cartesian, polar, spherical, cylindrical...), and they allow us to speak about quantities like $\frac{\partial f^i}{\partial x^j}\big|_x$ representing the j-th partial derivative of the i-th component of f at the point x. We may arrange these partial derivatives neatly into a matrix, the Jacobian matrix, and apply the appropriate transformation rules when (smoothly) changing coordinates. Here, we stress that coordinates and coordinate changes may only be defined on some open subset of X (e.g. Cartesian to polar coordinates is only defined away from a branch cut). The consistency of the transformation rules tells us that the Jacobian matrix of partial derivatives is simply a particular coordinate-dependent expression of *the* derivative of f. The latter, denoted df_x, is meant to be the "best linear approximation" of f at x,

$$f(x+v) = f(x) + df_x(v) + \dots .$$

But how should we make sense of this idea in a coordinate-independent way? (Many commonly used coordinate systems are not linearly related to one another...)

On a smooth manifold X, each local diffeomorphism of X with $\mathbb{R}^n$ serves as *local coordinates*, and is valid only within the domain of the coordinate chart $U \subset X$. We must at least have a fiducial collection of local coordinate charts (called an "atlas") that suffices to cover all points of X. These charts must also be smoothly compatible with each other (C^∞ coordinate changes on overlaps). However, a key conceptual point is that there are no preferred coordinates! Rather, *all* local coordinates which are smoothly compatible with the fiducial coordinate charts are allowed. The mathematical expression of this requirement is "smooth structure $\leftrightarrow$ maximal C^∞-atlas".

Now for the abstract mathematical definitions:

Definition 1. A *topological n-manifold* is a second-countable, Hausdorff topological space X, such that every point $x \in X$ has an open neighbourhood homeomorphic to an open subset of $\mathbf{R}^n$.

Thus X admits an open cover $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ together with homeomorphisms

$$\varphi_\alpha : U_\alpha \to \varphi_\alpha(U_\alpha) \subset \mathbf{R}^n$$

called local coordinates (or a *chart*). So $\varphi_\alpha = (x^1_\alpha, \dots x^n_\alpha)$ is an n-tuple of *coordinate functions*

$$x^i_\alpha : U_\alpha \to \mathbb{R}, \qquad i = 1, \dots, n,$$

used to provide numerical labels for the points in U_α. The subscript α will usually be dropped unless ambiguity arises.

Every pair of charts has a change-of-coordinates homeomorphism on the (possibly empty) overlap,

$$\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta) \to \varphi_\alpha(U_\alpha \cap U_\beta).$$

A (smooth) *atlas* is a collection of charts $\{U_\alpha, \varphi_\alpha\}_{\alpha\in\mathcal{I}}$ such that $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ covers X, and has all change-of-coordinates functions being C^∞ (infinitely differentiable). Extra charts can be admitted to an atlas provided the change-of-coordinates functions remain C^∞. A maximal atlas is called a *smooth structure* for X. So there is no distinguished sub-atlas (no distinguished coordinate choices), although a minimal fiducial one may be used for convenience when we want to describe a manifold to a colleague.

In particular, translated coordinate charts can always be used (i.e. shifting of origin), so we regularly say "coordinate chart centred at $x \in X$" to mean local coordinates φ for a neighbourhood of x, with the property that $\varphi(x) = 0$.

Definition 2. A smooth n-manifold is a topological n-manifold with a smooth structure.

By default, we will use "manifold" to mean "smooth manifold".

Example 4.1. The most simple example is the vector space $\mathbb{R}$, with the single fiducial chart $(\mathbb{R}, \mathrm{id})$; similarly for $\mathbb{R}^n$. These are the manifolds on which general manifolds are modelled. The identity chart can be composed with a

translation by any $s \in \mathbb{R}^n$, to obtain another compatible chart. No particular chart is distinguished amongst these translated charts. So as a manifold, $\mathbb{R}^n$ has no particular distinguished point as its origin, and we sometimes write $\mathbf{R}^n$ to emphasize its role as a manifold as opposed to a vector space.

Unless otherwise specified, we will simply use "manifold" to refer to smooth manifolds. The technical second-countable and Hausdorff conditions preclude some pathological examples, and in practice, ensure that an abstract manifold can be smoothly embedded into Euclidean space (Whitney embedding theorem), and admits partitions-of-unity subordinate to any open cover. (You can find these technical definitions in any comprehensive textbook on differentiable manifolds, such as [11].) Embedding an abstract smooth manifold into a background Euclidean space allows for a concrete visualization of its abstract tangent spaces as hyperplanes in the background. Partitions-of-unity allow local constructions to be smoothly globalized.

4.1 Tangent vectors

Let $f : X \to Y$ be a map of manifolds. Then near each point x, we may describe f in terms of coordinates (U, φ) and (V, ψ) near x and $f(x)$, respectively,

$$\underbrace{\psi \circ f \circ \varphi^{-1}}_{\check{f}} : \varphi(U \cap f^{-1}(V)) \to \psi(V).$$

So $\check{f}$ is the *coordinate representation* of f as an ordinary map between (open subsets of) Euclidean spaces, and we know what it means for $\check{f}$ to be smooth, from elementary calculus.

Definition 3. A map $f : X \to Y$ is *smooth* if there are atlases for X and Y such that all coordinate representations of f are smooth. If f has a smooth inverse, then we say that f implements a *diffeomorphism* between X and Y.

Note that once f is established to be smooth with respect to some atlases, the C^∞ compatibility condition will automatically guarantee smoothness in *any* coordinate charts in the maximal atlases. So smoothness is a coordinate-independent concept, even if we only checked it on some (rather than arbitrary) coordinate charts.

Example 4.2. For any chart (U, φ) in a (smooth) maximal atlas, φ is a diffeomorphism onto its image.

Example 4.3. When $Y = \mathbb{R}$, the set of smooth functions $X \to \mathbb{R}$ is denoted $C^\infty(X)$. Here, we actually treat the target $\mathbb{R}$ as a number field, not just a manifold. So $C^\infty(X)$ is an *algebra*, equipped with the compatible operations of pointwise addition and multiplication of functions.

Let $\gamma : (a, b) \to X$ be a smooth map, also known as a *curve* in X, parametrized by $t \in (a, b)$. Let $x = \gamma(t)$ be a point on the image of the curve. Then at x, we can consider the rate at which a given function $f \in C^\infty(X)$ changes along the curve: just take the ordinary derivative of the single-variable function $f \circ \gamma : (a, b) \to \mathbb{R}$. We think of the curve γ as assigning the rate-of-change (at time t) to each smooth function,

$$\gamma'(t) : C^\infty(X) \to \mathbb{R}, \qquad \gamma'(t)(f) := \frac{d(f \circ \gamma)}{ds}\Big|_{s=t}.$$

The assignment $\gamma'(t)$ is called the *velocity vector* of γ at the point $x = \gamma(t)$. So the velocity vector operates on functions; the number $\gamma'(t)(f)$ is the *derivative of f along γ* evaluated at the point $x = \gamma(t)$.

Obviously, $\gamma'(t)$ is a linear assignment (for each t). Furthermore, by elementary calculus, $\gamma'(t)$ satisfies a *Leibniz rule at $x = \gamma(t)$*,

$$\begin{aligned}
\gamma'(t)(f \cdot g) &= \frac{d((f \cdot g) \circ \gamma)}{ds}\Big|_{s=t} \\
&= \frac{d((f \circ \gamma) \cdot (g \circ \gamma))}{ds}\Big|_{s=t} \\
&= \frac{d(f \circ \gamma)}{ds}\Big|_{s=t} \cdot g(x) + f(x) \cdot \frac{d(g \circ \gamma)}{ds}\Big|_{s=t} \\
&= \gamma'(t)(f) \cdot g(x) + f(x) \cdot \gamma'(t)(g), \qquad \forall\, f, g \in C^\infty(X). \quad (4.1)
\end{aligned}$$

Formally, a map $v : C^\infty(X) \to \mathbb{R}$ is called a *derivation of the algebra $C^\infty(X)$ at the point x*, if it is linear and satisfies the Leibniz rule at x,

$$v(f \cdot g) = v(f) \cdot g(x) + f(x) \cdot v(g), \qquad f, g \in C^\infty(X).$$

Thus Eq. (4.1) says that the velocity vector $\gamma'(t)$ is a derivation of $C^\infty(X)$ at the point $x = \gamma(t)$.

If v_1, v_2 are derivations at x, and $\lambda \in \mathbb{R}$, then it is easy to see that $(\lambda v_1 + v_2)(f) := \lambda v_1(f) + v_2(f)$ defines another derivation. This motivates the following:

Definition 4. The *tangent space* to X at a point $x \in X$, denoted T_xX, is the vector space of derivations of $C^\infty(X)$ at x.

This definition captures the abstract algebraic aspect (Leibniz rule) of a derivative, and is manifestly coordinate-independent. We will see that all abstract derivations at x are in fact geometric: any derivation at x is of the form $\gamma'(t)$ for some (non-unique) smooth curve γ through x. Evidently, many distinct curves can give rise to the same derivation, so for the purposes of differentiating functions, we only distinguish two curves when they define distinct derivations.

Let x be a point in some open subset U of X. Since U is itself a manifold, we also have T_xU. The value of a derivation applied to f depends only on the values of f in some arbitrarily small open neighbourhood of x, so it is actually enough to define T_xX as the space of derivations of $C^\infty(U)$ at x, i.e. $T_xX = T_xU$. In particular, if a derivation at x is specified by a curve, we only need to consider the part of the curve lying within a coordinate chart around x.

For $i = 1, \ldots, n$, let $x^i : U \to \mathbb{R}$ be the coordinate functions for a chart (U, φ) centered at x. Consider the i-th coordinate curve through x, i.e.,

$$\begin{aligned} \gamma_i : (-\epsilon, \epsilon) &\to X \\ t &\mapsto \varphi^{-1}(0, \ldots, 0, \underbrace{t}_{i\text{-th}}, 0, \ldots, 0). \end{aligned}$$

(We take any sufficiently small $\epsilon > 0$ to ensure that the curve lies within the domain U of the chart.) This gives a set of *coordinate tangent vectors* at x,

$$\partial_i|_x := (\gamma_i)'(0), \qquad i = 1, \ldots, n.$$

Remark. If the chart is not centered at x, then we use the curves

$$\gamma_i : t \mapsto \varphi^{-1}(\varphi(x) + (0, \ldots, 0, t, 0, \ldots, 0))$$

to define the coordinate tangent vectors at x.

Actually, the derivations $\partial_i|_{x'}$ are also defined at every other point $x' \in U$ in the same manner. Thus a coordinate chart (U, φ) gives rise to *coordinate vector fields*

$$\partial_i : x' \mapsto \partial_i|_{x'}, \qquad x' \in U, \;\; i = 1, \ldots, n.$$

Writing $\iota_i : (-\epsilon, \epsilon) \to \mathbb{R}^n$ for the inclusion $t \mapsto (0, \dots, 0, \underbrace{t}_{i\text{-th}}, 0, \dots, 0)$, we untangle the definition of $\partial_i|_x$ using the ordinary chain rule:

$$\begin{aligned}\partial_i|_x(f) \equiv \frac{d(f \circ \gamma_i)}{ds}\Big|_{s=0} &= \sum_{j=1}^{n} \frac{\partial(f \circ \varphi^{-1})}{\partial x^j}\Big|_{\varphi(x)} \cdot \frac{d(\iota_i)^j}{ds}\Big|_{s=0} \\ &= \frac{\partial(f \circ \varphi^{-1})}{\partial x^i}\Big|_{\varphi(x)}, \qquad f \in C^\infty(U). \qquad (4.2)\end{aligned}$$

Thus $\partial_i|_x(f)$ simply means: take the i-th partial derivative of the coordinate representation $f \circ \varphi^{-1}$ of f, at the point $\varphi(x)$.

Exercise 4.1. Let $v \in T_xX$ be an arbitrary derivation at x. Show that v is a linear combination,

$$v = \sum_{i=1}^{n} v^i \partial_i|_x,$$

where $v^i := v(x^i) \in \mathbb{R}$. Also show that $\{\partial_i|_x\}_{i=1,\dots,n}$ is a linearly independent set of derivations, thus it is a basis for T_xX.

Exercise 4.1 shows that the coordinate vector fields ∂_i actually provide, for every point $x \in U$, a basis for the tangent space T_xX.

In fact, given $v \in T_xX$, we can easily construct an explicit geometric curve γ whose velocity vector at x is equal to v. Assume, without loss of generality, that the coordinates of x are $(0, \dots, 0)$, and write v^i for the components of v with respect to the basis $\{\partial_i|_x\}_{i=1,\dots,n}$. Then the curve

$$\gamma : (-\epsilon, \epsilon) \to U \subset X, \qquad t \mapsto \varphi^{-1}(tv^1, \dots, tv^n).$$

has velocity vector $\gamma'(0) = v$.

Physics conventions. Physicists usually think of tangent vectors as a list of components with certain behaviour under change of coordinates.

Exercise 4.2. Let (U, φ) and $(\tilde{U}, \tilde{\varphi})$ be two charts, $x^i, \tilde{x}^j$ be their respective coordinate functions, and $\partial_i|_x, \tilde{\partial}_j|_x$ the respective coordinate tangent vectors, $i, j = 1, \dots, n$. Let $x \in U \cap \tilde{U}$. Show that

$$\partial_i|_x = \frac{\partial \tilde{x}^j}{\partial x^i}\Big|_{\varphi(x)} \tilde{\partial}_j|_x. \qquad \text{(summation convention)}$$

Above, we use the very handy *Einstein summation convention* of summing over repeated indices when they appear both as a superscript and as a subscript.

Note that $\left(\frac{\partial \tilde{x}^j}{\partial x^i}\Big|_{\varphi(x)}\right)_{ji}$ is precisely the Jacobian matrix for the change-of-coordinates map $\tilde{\varphi} \circ \varphi^{-1}$. Since we may expand a general $v \in T_x X$ as

$$v = v^i \partial_i|_x \qquad \text{or} \qquad v = \tilde{v}^j \tilde{\partial}_j|_x,$$

we obtain the *contravariant transformation rule* for the components of tangent vectors,

$$\tilde{v}^j = \frac{\partial \tilde{x}^j}{\partial x^i}\Big|_{\varphi(x)} v^i. \tag{4.3}$$

Example 4.4. Let X denote an n-dimensional real vector space, and also the underlying manifold. Fix a point x in X. Given any vector $\mathrm{v} \in X$, we can consider the curve $t \mapsto x + t\mathrm{v}$. The velocity vector of this curve at x is some element in $T_x X$, which we denote by v_x. The map

$$X \to T_x X, \qquad \mathrm{v} \mapsto v_x$$

can be checked to be a linear isomorphism. Thus X is canonically isomorphic to $T_x X$, and this is true for all points $x \in X$. In simpler terms, for a vector space considered as a manifold, all of the tangent spaces are canonically isomorphic to each other. This is why we can freely "parallel transport" tangent vectors of Euclidean space from one basepoint to another, without much problem. A general manifold, however, does not have such a privileged identification of its various tangent spaces!

4.2 Derivatives of smooth maps

Definition 5. Let $f : X \to Y$ be a smooth map. Its *derivative at* $x \in X$ is the map

$$df_x : T_x X \to T_{f(x)} Y$$

defined by

$$df_x(v)(g) = v(g \circ f), \qquad v \in T_x X,\ g \in C^\infty(Y). \tag{4.4}$$

Exercise 4.3. Check the following:

- $df_x(v)$ is indeed a derivation of $C^\infty(Y)$ at $f(x)$.

- df_x is a linear map.
- The chain rule is satisfied: if $X \xrightarrow{f} Y \xrightarrow{g} Z$, then $d(g \circ f)_x = dg_{f(x)} \circ df_x$.
- $d(\mathrm{id}_X)_x = \mathrm{id}_{T_xX}$.
- If $f : X \to Y$ is a diffeomorphism, then df_x is a linear isomorphism whose inverse is $d(f^{-1})_{f(x)}$.

Example 4.5. Let $\gamma : (-\epsilon, \epsilon) \to X$ be a smooth curve, and write $x = \gamma(0)$. So the derivation $\frac{d}{dt}\big|_{t=0}$ is a coordinate tangent vector to the interval $(-\epsilon, \epsilon)$ at the point $t = 0$. We have

$$(d\gamma)_0 \left(\frac{d}{dt}\Big|_{t=0} \right) (g) = \frac{d}{dt}\Big|_{t=0} (g \circ \gamma) \equiv \gamma'(0)(g), \qquad \forall g \in C^\infty(X).$$

Thus the velocity vector $\gamma'(0) \in T_xX$ is just the derivative of the map γ applied to $\frac{d}{dt}\big|_{t=0}$.

Example 4.6. Let $f : X \to \mathbb{R}$ be a smooth function, viewed as a smooth map $X \to \mathbf{R}$ of manifolds. Then $df_x : T_xX \to T_{f(x)}\mathbf{R} = \mathbb{R}$ is a linear functional, also known as a *cotangent/dual vector* at x.

Definition 6. Let X be a manifold. The *cotangent space at* $x \in X$ is the space of linear functionals $T_xX \to \mathbb{R}$.

Exercise 4.4. Let $x^i : U \to \mathbb{R}, i = 1, \ldots, n$ be local coordinate functions for a chart around x. Show that $(dx^i)_x, i = 1, \ldots, n$ are linearly independent cotangent vectors at x.

Returning to a general smooth map $f : X \to Y$, let us make the description of df_x more concrete by working in a chart (U, φ) around x and a chart (V, ψ) around $f(x)$. Write x^i and $\tilde{x}^i$ for the respective coordinates, and

$$\check{f} := \psi \circ f \circ \varphi^{-1} = (\check{f}^1, \ldots, \check{f}^m)$$

for the coordinate representation of f (here, $m = \dim Y$). Since df_x is linear,

it suffices to see what it does to the basis of coordinate tangent vectors,

$$\begin{aligned}
df_x\left(\partial_i|_x\right)(g) &= \left(\partial_i|_x\right)(g\circ f) \\
&= \frac{\partial(g\circ f\circ\varphi^{-1})}{\partial x^i}\Big|_{\varphi(x)} \qquad \text{(Eq. (4.2))} \\
&= \frac{\partial(g\circ\psi^{-1})}{\partial\tilde{x}^j}\Big|_{\psi(f(x))}\frac{\partial(\psi\circ f\circ\varphi^{-1})^j}{\partial x^i}\Big|_{\varphi(x)} \\
&= \left(\left(\tilde{\partial}_j|_{f(x)}\right)(g)\right)\cdot\frac{\partial\check{f}^j}{\partial x^i}\Big|_{\varphi(x)} \\
&= \left(\frac{\partial\check{f}^j}{\partial x^i}\Big|_{\varphi(x)}\left(\tilde{\partial}_j|_{f(x)}\right)\right)(g), \qquad g\in C^\infty(Y)
\end{aligned}$$

So for a general $v = v^i\partial_i|_x$, we have

$$df_x(v) = v^i\frac{\partial\check{f}^j}{\partial x^i}\Big|_{\varphi(x)}\left(\tilde{\partial}_j|_{f(x)}\right).$$

The components of the tangent vector $df_x(v)$ with respect to the basis $\{\tilde{\partial}_j|_{f(x)}\}_{j=1,\dots,m}$ are

$$\left(df_x(v)\right)^j = \frac{\partial\check{f}^j}{\partial x^i}\Big|_{\varphi(x)}v^i, \qquad j=1,\dots,m.$$

In other words, with respect to the bases of coordinate tangent vectors, the linear map df_x is represented as the Jacobian matrix of $\check{f}$,

$$df_x \leftrightarrow J(\check{f})|_{\varphi(x)} := \left(\frac{\partial\check{f}^j}{\partial x^i}\Big|_{\varphi(x)}\right)_{ji}. \tag{4.5}$$

Example 4.7. Suppose f has coordinate representation $\check{f}$ being a linear map. Then the Jacobian of $\check{f}$ is $\check{f}$ itself.

Earlier, we explained how any tangent vector v at x may be obtained as $\gamma'(0)$ for some curve γ with $\gamma(0)=x$. Actually, $df_x(v)$ is simply represented by the curve $f\circ\gamma$ in Y:

Proposition 4.1. *Let $f: X\to Y$ be a smooth map, $v\in T_xX$, and γ be any curve in X with $\gamma(0)=x$ and $\gamma'(0)=v$. Then*

$$df_x(v) = (f\circ\gamma)'(0).$$

Proof. By definition, for any $g \in C^\infty(Y)$, we have

$$(f \circ \gamma)'(0)(g) = \frac{d}{dt}\Big|_{t=0}(g \circ f \circ \gamma)(t) = \gamma'(0)(g \circ f) = v(g \circ f) = df_x(v)(g).$$

□

4.3 Immersions, submersions, embeddings

Definition 7. A smooth map $f : X \to Y$ is a *immersion* (resp. *submersion*) *at* $x \in X$, if df_x is injective (resp. surjective). If f is an immersion at every $x \in X$, it is simply called an *immersion*; likewise for submersions.

Remark 1. The following important facts may be found in standard texts on smooth manifolds, e.g. [6, 11].

- If df_x is an isomorphism, then the *Inverse Function Theorem* says that $f : X \to Y$ is actually a *local diffeomorphism* at x, i.e., there is an open neighbourhood U of x such that $f : U \to f(U)$ is a diffeomorphism.
- Write $\dim X = n, \dim Y = m$. If f is an immersion at x, then in some coordinate charts centred at x and $f(x)$, it is represented as a canonical immersion $(x^1, \ldots, x^n) \mapsto (x^1, \ldots, x^n, 0, \ldots, 0)$.
- Similarly, if f is a submersion at x, then it has a coordinate representation as a standard projection $(x^1, \ldots, x^m, x^{m+1}, \ldots, x^n) \mapsto (x^1, \ldots, x^m)$.

The following Lemma will be useful for us.

Lemma 4.2. *Let $\pi : E \to X$ be a surjective submersion. Then a map $f : X \to Y$ is smooth iff $f \circ \pi : E \to Y$ is smooth.*

Proof. Suppose $f \circ \pi$ is smooth. Let $x \in X$ and pick a point $p \in \pi^{-1}(x)$. Since π is a submersion at p, it has the form $(x^1, \ldots, x^m, x^{m+1}, \ldots, x^n) \mapsto (x^1, \ldots, x^m)$ in some coordinate charts (V, ψ) centred at p and (U, φ) centred at x. We may then define the smooth map $s : U \to E$,

$$s = \psi^{-1} \circ \big((x^1, \ldots, x^m) \mapsto (x^1, \ldots, x^m, 0, \ldots, 0)\big) \circ \varphi,$$

which satisfies

$$s(x) = p, \qquad \pi \circ s = \mathrm{id}_U.$$

Then

$$f|_U = f|_U \circ \mathrm{id}_U = f \circ \pi \circ s$$

is the composition of the smooth maps $f \circ \pi$ and s, thus smooth. Since $x \in X$ is arbitrary, f is smooth.

If f is smooth, obviously its composition with the smooth π is smooth. $\square$

Remark. In topology, a *quotient* map is a surjective map $\pi : E \to X$ such that U is open in X iff $\pi^{-1}(U)$ is open in E. Then one shows $f : X \to Y$ is continuous iff $f \circ \pi : E \to Y$ is continuous. A surjective submersion is the smooth analogue of a quotient map.

In the following, we consider $k \leq n$ and $\mathbb{R}^k \subset \mathbb{R}^n$ as the standard slice with last $n - k$ coordinates being zero.

Definition 8. A k-dimensional *embedded/regular submanifold* (or simply *submanifold*) of an n-manifold X is a subspace $Z \subset X$ such that every point $z \in Z$ is contained in a chart (U, φ) of X with the property that $\varphi(U \cap Z) = \varphi(U) \cap \mathbb{R}^k$.

As its name suggests, a submanifold $Z \subset X$ is a manifold in its own right. Specifically, we take the charts $\varphi|_{U \cap Z} : U \cap Z \mapsto \mathbb{R}^k$, and check that they make Z into a k-manifold.

For example, any open subset U of a manifold X is a submanifold.

Definition 9. A smooth immersion $f : Z \to X$ which is also a homeomorphism onto its image $f(Z)$ (given the subspace topology), is called a (smooth) *embedding*.

Exercise 4.5. The inclusion map of a submanifold is an embedding.

Conversely, the image $f(Z)$ of an embedding $f : Z \to X$ is a submanifold of X, and $f : Z \to f(Z)$ is a diffeomorphism. We omit the proof (see [11] §5, [13] §5.6), and just mention an important example.

Exercise 4.6. Let $f : X \to Y$ be any smooth map. Its *graph* is the subset

$$\Gamma_f := \{(x, y) \in X \times Y \,:\, y = f(x)\}.$$

Then the map

$$X \to X \times Y, \qquad x \mapsto (x, f(x))$$

is an embedding, so X is diffeomorphic to its graph.

Smoothness of maps behaves well with respect to restriction to submanifolds:

Exercise 4.7. Let $f : X \to Y$ be a smooth map. Show that if X' is a submanifold of X, then $f|_{X'}$ remains smooth. Similarly, show that if Y' is a submanifold of Y such that $f(X) \subset Y$, then $f : X \to Y'$ remains smooth.

Definition 10. If $f : X \to Y$ is a smooth map, then $y \in Y$ is a *regular value* of f if f is a submersion at all points in the (possibly empty) preimage $f^{-1}(y)$.

Proposition 4.3. *Let y be a regular value of a smooth map $f : X \to Y$. Then $f^{-1}(y)$ is a submanifold of X with codimension equal to* $\dim Y$.

Proof. Exercise. □

More generally:

Definition 11. A submanifold $Z \subset Y$ is *transverse* to $f : X \to Y$ if

$$\mathrm{Image}(df_x) + T_{f(x)}(Z) = T_{f(x)}Y, \qquad \forall\, x \in f^{-1}(Z).$$

Theorem 4.4. *Let $Z \subset Y$ be a submanifold transverse to a map $f : X \to Y$. Then its preimage $f^{-1}(Z)$ is a submanifold of X, with*

$$\operatorname{codim} f^{-1}(Z) = \operatorname{codim} Z. \tag{4.6}$$

Proof. See §5 of [6], Theorem 6.30 of [11]. □

Such results involving transversality constitute an important part of *differential topology*, you may learn more about this subject from [6, 12]. The importance for us is the following corollary:

Proposition 4.5. *Let $\pi : E \to X$ be a surjective submersion. Then for any submanifold $Z \subset X$, the preimage $\pi^{-1}(Z)$ is a submanifold of E with the same codimension.*

Proof. The stated assumptions imply that π is transverse to any submanifold $Z \subset X$, whence the conclusion follows from (4.6). □

5 Tangent bundle and vector bundles

5.1 Tangent bundle

As a set, the *tangent bundle* of a manifold X is a disjoint union

$$TX = \bigsqcup_{x \in X} T_x X,$$

with a projection map

$$\pi : TX \to X, \qquad (x, v_x) \mapsto x.$$

We can give TX the structure of a smooth $2n$-manifold, such that $\pi : TX \to X$ is smooth. First, we already have coordinate charts $(U_\alpha, \varphi_\alpha), \alpha \in \mathcal{I}$ covering X. For each $\alpha \in \mathcal{I}$, consider the "local tangent bundle",

$$TU_\alpha := \pi^{-1}(U_\alpha) = \bigsqcup_{x \in U_\alpha} T_x X.$$

For any $x \in U_\alpha$ and any $v_x \in T_x X$, we have the expansion $v_x = v_x^i \partial_i|_x$ in terms of coordinate tangent vectors. So define the injective maps

$$\begin{aligned} \tilde{\varphi}_\alpha : TU_\alpha &\to \varphi_\alpha(U_\alpha) \times \mathbb{R}^n \subset \mathbb{R}^{2n} \\ (x, v_x) &\mapsto (\varphi_\alpha(x); v_x^1, \ldots, v_x^n). \end{aligned}$$

Clearly, the $TU_\alpha, \alpha \in \mathcal{I}$ cover TX, and we would like to use the $(TU_\alpha, \tilde{\varphi}_\alpha)$ as charts for TX. To make sense of this, we first need to topologize TX, by taking the basis of open sets to be

$$\tilde{\varphi}_\alpha^{-1}(V), \qquad V \text{ open in } \mathbb{R}^{2n}, \alpha \in \mathcal{I}.$$

One checks that this is indeed a basis for a topology; the verification of Hausdorffness and second-countability are omitted. The charts $\tilde{\varphi}_\alpha$ are now homeomorphisms, and TX is a topological manifold. The maps $\tilde{\varphi}_\alpha \circ \tilde{\varphi}_\beta^{-1}$ are easily seen to be C^∞ — the base coordinates transform smoothly since X is a manifold, while the vector components transform according to the smooth formula Eq. (4.3). Therefore we have constructed a smooth atlas for TX, turning it into a smooth $(2n)$-manifold. With respect to the charts $(TU_\alpha, \tilde{\varphi}_\alpha)$ and $(U_\alpha, \varphi_\alpha)$, the projection map π is simply projection onto the first factor, thus smooth.

Remark. If we have a smooth embedding $X \hookrightarrow \mathbf{R}^N$, then we may construct TX as a subset of $X \times \mathbb{R}^N$. The advantage is that the topology is directly inherited from $X \times \mathbb{R}^N$. A disadvantage is that we have to check that this is a canonical construction, independent of choice of embedding.

Observe that the maps

$$\Phi_\alpha : TU_\alpha \xrightarrow{\tilde{\varphi}_\alpha} \varphi_\alpha(U_\alpha) \times \mathbb{R}^n \xrightarrow{\varphi_\alpha^{-1} \times \mathrm{id}} U_\alpha \times \mathbb{R}^n$$

are diffeomorphisms which identify TU_α as a cartesian product — these are called local trivializations of TX, and turn it into a vector bundle in the sense of Definition 13.

5.1.1 Tangent vector fields as derivations

Definition 12. A *tangent vector field* over X is a smooth map $v : X \to TX$ such that $\pi(v(x)) = x$, $\forall x \in X$. We denote the space of tangent vector fields by $\mathfrak{X}(X)$.

Let $v : X \to TX$ be a not-necessarily-smooth tangent vector field over X. So $\pi \circ v = \mathrm{id}_X$, but v may not be smooth. Pointwise in X, we have the derivation v_x of $C^\infty(X)$ at x. In total, we get a function,

$$v(f) : X \to \mathbb{R}, \qquad v(f)(x) := v_x(f). \tag{5.1}$$

Exercise 5.1. Check that the following are equivalent:

- v is a smooth tangent vector field.
- For all $f \in C^\infty(X)$, the function $v(f)$ defined by Eq. (5.1) lies in $C^\infty(X)$.

Exercise 5.1 implies that a smooth vector field v defines a *derivation* on the algebra $C^\infty(X)$, i.e.,

- $v : C^\infty(X) \to C^\infty(X), f \mapsto v(f)$ is linear;
- $v(fg) = fv(g) + v(f)g, \qquad f, g \in C^\infty(X)$.

Remark 2. Conversely, if D is any derivation on $C^\infty(X)$, then we obtain a pointwise derivation $D_x \in T_xX$ at each x by the formula

$$D_x(f) := (Df)(x), \qquad f \in C^\infty(X).$$

The assignment $x \mapsto D_x$ is smooth, due to Exercise 5.1.

5.2 General vector bundles

The tangent bundle TX is a special case of the following:

Definition 13. A (real) *vector bundle* $\pi : E \to X$ *of rank* n is a smooth surjective map, such that there exists an open cover $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ of X and diffeomorphisms ("*local trivializations*")

$$\Phi_\alpha : E|_{U_\alpha} := \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$$

such that

- $\pi_{U_\alpha} \circ \Phi_\alpha(p) = \pi(p)$ for all $p \in E|_{U_\alpha}$, where π_{U_α} denotes the projection $U_\alpha \times \mathbb{R}^n \to U_\alpha$,
- For any two local trivializations $(U_\alpha, \Phi_\alpha), (U_\beta, \Phi_\beta)$, we have

$$\Phi_\beta \circ \Phi_\alpha^{-1}(x, \mathbf{v}) = (x, g_{\beta\alpha} \cdot \mathbf{v}), \qquad x \in U_\alpha \cap U_\beta, \mathbf{v} \in \mathbb{R}^n, \tag{5.2}$$

  with smooth *transition functions* $g_{\beta\alpha} : U_\alpha \cap U_\beta \to \mathrm{GL}(n, \mathbb{R})$.

If there is a Φ_α with $U_\alpha = X$, it is called a *(global) trivialization* of E, and E is said to be *trivializable*. A *section* of E, also called a smooth *vector field*, is a smooth map $s : X \to E$ such that $\pi \circ s = \mathrm{id}_X$. The space of sections on E is denoted $\Gamma(E)$. If s is only defined on an open subset $U \subset X$, it is called a *local section* over U. *Complex* vector bundles are defined in the same way, except that $\mathbb{R}^n$ is replaced by $\mathbb{C}^n$, and $g_{\beta\alpha}$ takes values in $\mathrm{GL}(n, \mathbb{C})$.

If $v \in \Gamma(E)$ is a vector field, then in a local trivialization (U, Φ), we have

$$\Phi(v(x)) = (x, \mathbf{v}(x)), \qquad x \in U,$$

for some smooth function $\mathbf{v} \equiv (v^1, \ldots, v^n) : U \to \mathbb{R}^n$. The v^i are the *local component functions* of v.

Remark. For each $x \in X$, the fibre $E_x := \pi^{-1}(x)$ inherits a vector space structure through its identification with $\mathbb{R}^n$ via a(ny) local trivialization. In particular, each E_x has a zero element, and it makes sense to ask whether a section s has $s(x) = 0$ or $s(x) \neq 0$. Clearly, if E is trivializable, then it admits a nowhere-vanishing section $s : X \to E$.

Remark. In an equivalent approach, the fibres E_x are *a priori* vector spaces, and instead of having the smooth compatibility condition (5.2) in the definition, one requires the local trivializations Φ_α to restrict fibrewise to linear isomorphisms $E_x \xrightarrow{\cong} \{x\} \times \mathbb{R}^n$. Then it is proved that the compatibility condition by smooth $\mathrm{GL}(n)$-valued transition functions holds (details can be found in Lemma 10.5 of [11]).

Remark. A specific choice of trivializing cover $\{U_\alpha, \Phi_\alpha\}_{\alpha \in \mathcal{I}}$ is not part of the data of a vector bundle, only the availability of such a cover is. In practice, we may specify a fiducial set of local trivializations, then pass to a maximal trivializing cover compatible with this fiducial set.

Remark. The space of sections, $\Gamma(E)$, is a module over the ring $C^\infty(X, \mathbb{K})$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ is the ground field of the fibres.

Exercise 5.2. Show that the projection map of a vector bundle is a submersion.

Definition 14. Let $\pi : E \to X$ be a rank-n vector bundle over X, and $U \subset X$ an open subset. An n-tuple of local sections $(s_1, \ldots, s_n)$ is called a *local frame for E over U*, if for each $x \in U$, $\{s_i(x)\}_{i=1,\ldots,n}$ is a basis for E_x.

A local trivialization $\Phi : \pi^{-1}(U) \to U \times \mathbb{R}^n$ determines a local frame for E over U; set $s_i(x) = \Phi^{-1}(x, e_i)$, where e_i is the i-th standard basis vector for $\mathbb{R}^n$. Conversely:

Exercise 5.3. Let $(s_1, \ldots, s_n)$ be a local frame for E over U. Define the map

$$\Psi : U \times \mathbb{R}^n \to \pi^{-1}(U)$$
$$(x, (v^1, \ldots, v^n)) \mapsto \sum_{i=1}^n v^i s_i(x).$$

Check that Ψ is a diffeomorphism, thus Ψ^{-1} is a local trivialization of E.

5.3 Tautological line bundle over $\mathbb{RP}^1$

The *real projective line* $\mathbb{RP}^1$ is the set of one-dimensional subspaces ℓ (i.e. lines) in $\mathbb{R}^2$. On $v = (x, y) \in \mathbb{R}^2 \setminus \{0\}$, define the equivalence relation $[v] = [v']$ iff v, v' lie in the same line. Clearly, $\mathbb{RP}^1$ is precisely the quotient space under this equivalence,

$$q : \mathbb{R}^2 \setminus \{0\} \to \mathbb{RP}^1, \qquad v \mapsto [v].$$

We equip $\mathbb{RP}^1$ with the quotient topology; that is, $U \subset \mathbb{RP}^1$ is open iff $q^{-1}(U)$ is open in $\mathbb{R}^2 \setminus \{0\}$.

Let ℓ_{H} be the line spanned by $(1,0)$ and ℓ_{V} be the line spanned by $(0,1)$. Define the subsets

$$U_1 = \mathbb{RP}^1 \setminus \{\ell_{\mathrm{V}}\}, \qquad U_2 = \mathbb{RP}^1 \setminus \{\ell_{\mathrm{H}}\}.$$

Then $\{U_1, U_2\}$ is an open cover for $\mathbb{RP}^1$ (Why?). Define the maps

$$\begin{aligned} \varphi_1 &: U_1 \to \mathbb{R}, & [(x,y)] &\mapsto y/x, \\ \varphi_2 &: U_2 \to \mathbb{R}, & [(x,y)] &\mapsto x/y, \end{aligned}$$

which are well-defined homeomorphisms (why?). The overlapping region is

$$U_{12} \equiv U_1 \cap U_2 = \mathbb{RP}^1 \setminus \{\ell_{\mathrm{V}}, \ell_{\mathrm{H}}\},$$

and clearly $\varphi_i(U_{12}) = \mathbb{R} \setminus \{0\} \equiv \mathbb{R}^\times$ for $i = 1, 2$. The change-of-coordinates map is

$$\varphi_2 \circ \varphi_1^{-1} : \mathbb{R}^\times \to \mathbb{R}^\times, \qquad \lambda \mapsto \lambda^{-1},$$

which is smooth. Thus $\mathbb{RP}^1$ is a smooth manifold.

The *tautological line bundle* $\pi : \mathcal{L}^{\mathbb{R}} \to \mathbb{RP}^1$ sits inside the product $\mathbb{RP}^1 \times \mathbb{R}^2$ as the following subset,

$$\mathcal{L}^{\mathbb{R}} = \{(\ell; v) \in \mathbb{RP}^1 \times \mathbb{R}^2 \,|\, v \in \ell\},$$

and it is equipped with the projection map,

$$\pi : \mathcal{L}^{\mathbb{R}} \to \mathbb{RP}^1, \qquad (\ell; v) \mapsto \ell.$$

Define the continuous bijective maps

$$\begin{aligned} \Phi_1 &: \pi^{-1}(U_1) \to U_1 \times \mathbb{R}, & (\ell; (x,y)) &\mapsto \left(\ell; x\sqrt{1 + (\varphi_1(\ell))^2}\right), \\ \Phi_2 &: \pi^{-1}(U_2) \to U_2 \times \mathbb{R}, & (\ell; (x,y)) &\mapsto \left(\ell; y\sqrt{1 + (\varphi_2(\ell))^2}\right), \end{aligned}$$

which respect the projection maps to U_i. Their inverses are

$$\begin{aligned} &\Phi_i^{-1} : U_1 \times \mathbb{R} \to \pi^{-1}(U_i) \subset \mathcal{L}^{\mathbb{R}} \subset \mathbb{RP}^1 \times \mathbb{R}^2 \\ &(\ell; \mu) \mapsto \begin{cases} \left(\ell; \mu\left(\frac{1}{\sqrt{1+(\varphi_1(\ell))^2}}, \frac{\varphi_1(\ell)}{\sqrt{1+(\varphi_1(\ell))^2}}\right)\right), & i = 1, \\ \left(\ell; \mu\left(\frac{\varphi_2(\ell)}{\sqrt{1+(\varphi_2(\ell))^2}}, \frac{1}{\sqrt{1+(\varphi_2(\ell))^2}}\right)\right), & i = 2. \end{cases} \end{aligned}$$

and are also continuous. We may compute (exercise)

$$\begin{aligned}\Phi_2 \circ \Phi_1^{-1} : U_{12} \times \mathbb{R} &\to U_{12} \times \mathbb{R} \\ (\ell; \mu) &\mapsto \big(\ell; \operatorname{sgn}(\varphi_2(\ell))\mu\big).\end{aligned}$$

Note that $U_{12} \xrightarrow{\varphi_2} \mathbb{R}^\times \xrightarrow{\operatorname{sgn}} \mathrm{O}(1)$ is continuous (even smooth), so the above map is smooth. Then $(\varphi_1 \times \mathrm{id}_\mathbb{R}) \circ \Phi_1$ and $(\varphi_2 \circ \mathrm{id}_\mathbb{R}) \circ \Phi_2$ may be regarded as defining coordinate charts for $\mathcal{L}^\mathbb{R}$ as a smooth manifold. With this smooth structure, $\pi : \mathcal{L}^\mathbb{R} \to \mathbb{RP}^1$ is smooth. Also, Φ_1, Φ_2 are diffeomorphisms, and they serve as local trivializations of $\mathcal{L}^\mathbb{R}$, with smooth transition function being $\operatorname{sgn} \circ \varphi_2 : U_{12} \to \mathrm{O}(1)$.

Non-trivializability. We have constructed $\pi : \mathcal{L}^\mathbb{R} \to \mathbb{RP}^1$ as a rank-1 vector bundle. However, it does not admit any nowhere-vanishing sections (Exercise), and it is therefore non-trivializable. So any globally-defined $s \in \Gamma(\mathcal{L}^\mathbb{R})$ will always have some zeroes, and it is not hard to imagine some significant consequences for defining differential operators on such a space.

Remark. Our local trivializations of $\mathcal{L}^\mathbb{R}$ appear somewhat more complicated than necessary, but they have the nice property that the lengths of fibre vectors are preserved. For example, if $\ell \in U_1$, then

$$(x, y) \in \ell \;\Rightarrow\; |x\sqrt{1 + (\varphi_1(\ell))^2}|^2 = |x\sqrt{1 + (y/x)^2}|^2 = x^2 + y^2 = ||(x, y)||^2;$$

similarly if $\ell \in U_2$. This property is reflected in the fact that the transition functions are $\mathrm{O}(1)$-valued instead of being $\mathrm{GL}(1, \mathbb{R})$-valued. So $\mathcal{L}^\mathbb{R}$ is actually an unorientable Euclidean line bundle (this concept will be defined in a later lecture).

Remark. It is not hard to show that $\mathbb{RP}^1$ is diffeomorphic to a circle S^1. Furthermore, there is a natural metric on $\mathbb{RP}^1$, given on a pair of lines by the angle between them. A little thought will show that $\mathbb{RP}^1$ is "half of a unit circle", having circumference π.

5.4 Morphisms of vector bundles

The tangent bundle TX comes *canonically* with the manifold X. To understand this further, we must discuss morphisms between vector bundles.

Definition 15. A *morphism of vector bundles* from $\pi : E \to X$ to $\pi' : E' \to X'$ is a smooth map $F : E \to E'$, such that there exists a smooth map $f : X \to X'$ making the following diagram commute

$$\begin{tikzcd} E \arrow[r, "F"] \arrow[d, "\pi"'] & E' \arrow[d, "\pi'"] \\ X \arrow[r, "f"] & X' \end{tikzcd}, \tag{5.3}$$

and which restricts to linear maps $F : E_x \to E'_{f(x)}$ on fibres.

Example 5.1. As a simple example, let $U \subset X$, and consider $E|_U := \pi^{-1}(U)$, which is still a vector bundle. Then the inclusion map $F : E|_U \to E$ is a vector bundle morphism, by taking $f : U \to X$ to be the inclusion.

5.4.1 Derivative as morphism of tangent bundles

We had already encountered the pointwise derivative of a smooth map $f : X \to Y$ (Eq. (4.4)). This is a linear map $df_x : T_xX \to T_{f(x)}Y$. Globally, the *(total) derivative* of f is

$$df : TX \to TY,$$
$$\bigsqcup_{x\in X} df_x : \bigsqcup_{x\in X} T_xX \to \bigsqcup_{x\in X} T_{f(x)}Y \subset \bigsqcup_{y\in Y} T_yY.$$

Let $\check{f} = \psi \circ f \circ \phi^{-1}$ be the representation of f with respect to some charts (U, φ) and (V, ψ) around x and $f(x)$ respectively. Recall from Eq. (4.5) that df_x is represented as the Jacobian matrix $J(\check{f})|_{\varphi(x)}$ acting on the components v^i of a vector $v \in T_xX$. Therefore, in the corresponding charts (TU, Φ) and (TV, Ψ) for the tangent bundles, we have df represented as

$$\Psi \circ df \circ \Phi^{-1} : (\underbrace{x^1, \ldots, x^n}_{\mathbf{x}}; \underbrace{v^1 \ldots, v^n}_{\mathbf{v}}) \mapsto (\check{f}(\mathbf{x}); J(\check{f})|_{\mathbf{x}}(\mathbf{v})),$$

which is smooth, since $\check{f}$ and $J(\check{f})$ are both smooth. If follows that $df : TX \to TY$ is a smooth map.

Therefore, df is a morphism of tangent vector bundles, in the sense of Definition 15. It also follows from Exercise 4.3 that the derivative is *functorial*, in the sense that:

- Chain Rule: If $f : X \to Y$ and $g : Y \to Z$ are smooth maps, then $d(g \circ f) = dg \circ df : TX \to TZ$;
- $d(\mathrm{id}_X) = \mathrm{id}_{TX}$.

In particular, if $f : X \to Y$ is a diffeomorphism, then so is $df : TX \to TY$.

6 Principal G-bundles

6.1 Lie groups

Definition 16. A Lie group is a manifold G together with a group structure such that

$$G \times G \to G, \qquad (g_1, g_2) \mapsto g_1 g_2^{-1}$$

is smooth.

Thus the group operations (inversion and composition) are smooth maps. It follows easily that the operations of left and right multiplication on G by any fixed $g \in G$,

$$L_g : g' \mapsto gg', \qquad R_g : g' \to g'g, \qquad g' \in G,$$

are self-diffeomorphisms of G.

The basic example of a Lie group is $\mathrm{GL}(n)$ (over $\mathbb{R}$ or $\mathbb{C}$). It is usually considered as an open (why?) submanifold of the vector space of matrices $\mathrm{M}_n(\mathbb{R})$ or $\mathrm{M}_n(\mathbb{C})$. The tangent spaces at each $g \in \mathrm{GL}(n)$ are the same as those when g is considered as element of $\mathrm{M}_n(\mathbb{K})$. Thus the tangent bundle of $\mathrm{GL}(n)$ is identified with

$$T\mathrm{GL}(n) \cong \mathrm{GL}(n) \times \underbrace{\mathrm{M_n}(\mathbb{K})}_{\cong \mathbb{K}^{n^2}}.$$

Similarly for the orientation-preserving subgroup $\mathrm{GL}(n, \mathbb{R})^+$ (i.e. positive determinant matrices). Smoothness of the group operations is deduced from their algebraic nature in terms of the matrix entries. For an n-dimensional vector space V, the invertible linear maps $\mathrm{GL}(V)$ are identified with $\mathrm{GL}(n)$ by picking a basis for V, and any two such choices are related by a conjugation.

The subgroups $\mathrm{O}(n), \mathrm{SO}(n), \mathrm{U}(n), \mathrm{SU}(n)$ are also Lie groups, but of lower dimension. This is usually shown by the following exercise:

Exercise 6.1. Consider $\mathrm{M}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ and the subspace $\mathrm{Sym}(n)$ of symmetric matrices. Check that 1_n is a regular value for the map

$$f : \mathrm{M}_n(\mathbb{R}) \to \mathrm{Sym}(n), \qquad A \mapsto A^{\mathrm{t}} A.$$

Consequently, $\mathrm{O}(n) = f^{-1}(1_n)$ is actually a submanifold of $\mathrm{M}_n(\mathbb{R})$ (also of $\mathrm{GL}(n, \mathbb{R})$), due to Prop. 4.3. It follows (by Exercise 4.7) that $\mathrm{O}(n)$ is a Lie subgroup of $\mathrm{GL}(n, \mathbb{R})$. Similarly, $\mathrm{SO}(n)$ and $\mathrm{U}(n), \mathrm{SU}(n)$ are Lie subgroups of $\mathrm{GL}(n, \mathbb{R})^+$ and $\mathrm{GL}(n, \mathbb{C})$ respectively.

6.2 Idea of principal bundles

In a vector bundle, the vector space E_x attached to $x \in X$ does not come with a canonical basis ("frame at x"). A frame, or "gauge", at x is a choice of isomorphism

$$\begin{aligned} \mathbf{e} : \mathbb{K}^n &\xrightarrow{\cong} E_x \\ (0, \ldots, 0, \underbrace{1}_{i\text{-th}}, 0, \ldots, 0) &\mapsto e_i \in E_x, \qquad i = 1, \ldots, n. \end{aligned}$$

We will sometimes write $\mathbf{e} = \{e_1, \ldots, e_n\}$.

The set $\mathrm{Fr}(E_x)$ of frames at x has an important structure. Namely, it has a right action by $G = \mathrm{GL}(n)$, denoted

$$\mathrm{Fr}(E_x) \times G \to \mathrm{Fr}(E_x), \qquad (\mathbf{e}, g) \mapsto \mathbf{e} \cdot g.$$

Explicitly, if $\mathbf{e} = \{e_1, \ldots, e_n\}$, then

$$\mathbf{e} \cdot g = \left\{ \sum_{i=1}^n e_i g^i{}_1, \ldots, \sum_{i=1}^n e_i g^i{}_n \right\}, \qquad g = (g^i{}_j) \in G = \mathrm{GL}(n).$$

Symbolically, the frame $\mathbf{e}$ is written as a row vector, and it is multiplied by the $n \times n$ matrix g to get the row vector for the new frame $\mathbf{e} \cdot g$.

The map

$$\begin{aligned} \mathrm{Fr}(E_x) \times G &\to \mathrm{Fr}(E_x) \times \mathrm{Fr}(E_x) \\ (\mathbf{e}, g) &\mapsto (\mathbf{e}, \mathbf{e} \cdot g) \end{aligned} \tag{6.1}$$

is a bijection. Formally, this says that $\mathrm{Fr}(E_x)$ is a *G-torsor* — it has a free and transitive action of G, implementing change-of-frame. Informally, this is "gauge freedom" of "gauge symmetry" at x.

Let $\mathbf{e}_{\mathrm{ref}}$ be a reference basis/frame (or "gauge choice"). Fixing $\mathbf{e}$ to be $\mathbf{e}_{\mathrm{ref}}$ in Eq. (6.1), we obtain a $\mathbf{e}_{\mathrm{ref}}$-dependent bijection

$$\begin{aligned} G &\longleftrightarrow \mathrm{Fr}(E_x) \\ g &\longleftrightarrow \mathbf{e}_{\mathrm{ref}} \cdot g, \end{aligned}$$

turning $\mathrm{Fr}(E_x)$ into a group. Effectively, we declare $\mathbf{e}_{\mathrm{ref}}$ to be "identity frame", and the frame space $\mathrm{Fr}(E_x)$ becomes identified with $\mathrm{Fr}(\mathbb{K}^n) = G$.

If the E_x come with an inner product, then we can restrict to orthonormal frames. The above discussion still holds, with G replaced by $\mathrm{O}(n)$ or $\mathrm{U}(n)$.

6.2.1 Trivializable principal bundles

Consider a trivial*ized* vector bundle $E = X \times \mathbb{K}^n$. Its (orthonormal) *frame bundle* is then the product manifold

$$\mathrm{Fr}(X \times \mathbb{K}^n) = \bigsqcup_{x \in X} \mathrm{Fr}(\mathbb{K}^n) = X \times G, \qquad G = \mathrm{GL(n)} \text{ (or } \mathrm{U}(n) \text{ or } \mathrm{O}(n)).$$

When equipped with the projection map $\pi : X \times G \to X$ and the right G-action,

$$(x, g) \cdot g' = (x, gg'),$$

$\mathrm{Fr}(X \times \mathbb{K}^n)$ is an example of a trivial*ized* principal G-bundle.

Next, consider a trivial*izable* vector bundle $E \to X$. The frame bundle of E is

$$\mathrm{Fr}(E) = \bigsqcup_{x \in X} \mathrm{Fr}(E_x), \tag{6.2}$$

which has, set-theoretically, the projection $\pi : \mathrm{Fr}(E) \to X$, as well as a right G-action on each fibre $\mathrm{Fr}(E_x)$. Under a smooth trivializing map $\Phi : E \xrightarrow{\cong} X \times \mathbb{K}^n$, each E_x gets identified with $\mathbb{K}^n$. This is equivalent to picking a reference frame in $\mathrm{Fr}(E_x)$, and therefore an identification $\mathrm{Fr}(E_x) \cong G$, for each $x \in X$. Thus we get a bijective map

$$\tilde{\Phi} : \mathrm{Fr}(E) \to X \times G,$$

which respects the right G-actions. Use $\tilde{\Phi}$ to transfer the topology and smooth structure of $X \times G$ to $\mathrm{Fr}(E)$.

This frame bundle $\pi : \mathrm{Fr}(E) \to X$ is an example of a trivial*izable* principal G-bundle. If we had used some other trivialization $\Phi' : E \xrightarrow{\cong} X \times \mathbb{K}^n$ of the vector bundle E, then we would obtain another trivialization $\mathrm{Fr}(E) \cong X \times G$ of the frame bundle. The smooth structure on $\mathrm{Fr}(E)$, bundle projection, and G action on $\mathrm{Fr}(E)$ do not depend on which trivialization we started with. Indeed, no preferred trivialization is specified for either E or its frame bundle $\mathrm{Fr}(E)$.

6.3 General principal bundle

Definition 17. Let G be a Lie group. The *trivial principal G-bundle* over a manifold X is $\pi_X : X \times G \to X$, $(x, g) \mapsto x$, equipped with the right G-action,

$$(x, h) \cdot g = (x, hg), \qquad x \in X,\ h, g \in G.$$

A *principal G-bundle* over a manifold X is a manifold P with a smooth surjection $\pi : P \to X$ and a smooth right G-action (by diffeomorphisms),

$$P \times G \to P, \qquad (p, g) \mapsto p \cdot g,$$

such that

- the right G-action restricts to fibres,

$$\pi(p \cdot g) = \pi(p), \qquad p \in P,\ g \in G;$$

- every $x \in X$ is contained in some open neighbourhood U with $P|_U := \pi^{-1}(U)$ being trivializable, i.e., there exists a diffeomorphism

$$\Phi : \pi^{-1}(U) \to U \times G$$

  such that

$$\begin{aligned} \pi_U \circ \Phi &= \pi, \\ \Phi(p \cdot g) &= \Phi(p) \cdot g. \end{aligned}$$

As with Exercise 5.2, the projection map π of a principal G-bundle can be checked to be a submersion.

Example 6.1. Recall Example 1.2 of the real line bundle $\mathcal{L}^{\mathbb{R}} \to \mathbb{RP}^1$. In the fibre above each $\ell \in \mathbb{RP}^1$, there are two unit-length vectors. Let us consider the subset of $\mathcal{L}^{\mathbb{R}}$ comprising the unit-length vectors. This is the (unit) *sphere bundle* $S(\mathcal{L}^{\mathbb{R}})$ of $\mathcal{L}^{\mathbb{R}}$.

The fibre of $S(\mathcal{L}^{\mathbb{R}})$ lying above $\ell \in \mathbb{RP}^1$ comprises the two unit vectors in the line ℓ. Think of these unit vectors as the two possible choices of orthonormal frames for the line ℓ, with no preference for either one. Globally, $S(\mathcal{L}^{\mathbb{R}})$ is the bundle of orthonormal frames for $\mathcal{L}^{\mathbb{R}}$. The group $\mathrm{O}(1) \cong \mathbb{Z}_2$ acts on $S(\mathcal{L}^{\mathbb{R}})$ by swapping the two frames at each $\ell \in \mathbb{RP}^1$. This gives $S(\mathcal{L}^{\mathbb{R}})$ the structure of a principal $\mathrm{O}(1) \cong \mathbb{Z}_2$-bundle over $\mathbb{RP}^1$ (Exercise).

Interestingly, $S(\mathcal{L}^{\mathbb{R}})$ is not simply two disconnected copies of $\mathbb{RP}^1$. Rather, the total space $S(\mathcal{L}^{\mathbb{R}})$ can be identified with the unit circle S^1, as a *connected* double cover of $\mathbb{RP}^1$.

Example 6.2. Example 6.1 has a direct complex counterpart, $\pi : \mathcal{L}^{\mathbb{C}} \to \mathbb{CP}^1$. Here, $\mathbb{CP}^1$ is the space of one-dimensional complex subspaces in $\mathbb{C}^2$, and $\mathcal{L}^{\mathbb{C}}$ is the complex tautological line bundle over $\mathbb{CP}^1$, constructed in the same way as $\mathbb{RP}^1$.

For each $\ell \in \mathbb{CP}^1$, we can consider only the unit complex vectors inside the corresponding fibre of $\mathbb{C}$, and this is interpreted as the space of orthonormal frames of ℓ, which is a U(1)-torsor. The sphere bundle of $\mathcal{L}^{\mathbb{C}}$ is thus the bundle of orthonormal frames for the line bundle $\mathcal{L}^{\mathbb{C}}$. The group U(1) acts on this frame bundle by "rotating frames" at each ℓ.

This frame bundle can be shown to be diffeomorphic to S^3. So we get an interesting principal U(1)-bundle $\pi : S^3 \to \mathbb{CP}^1$. It is well-known that $\mathbb{CP}^1 \cong S^2$, and π is actually the Hopf fibration, which we will study in detail in Section 7.1.

Some motivation. Now, $S^3 \ncong S^2 \times \mathrm{U}(1)$, which is mathematically interesting. But from the physical viewpoint, why would we need to consider general abstract "twisted frame bundles"? Historically, this non-trivial principal bundle arose when Dirac was considering the idea of a magnetic monopole (albeit still not found in nature). It was not until many decades later, that Dirac's monopole was understood in bundle-theoretic and algebraic topology terms, and generalizations with non-Abelian G grew in importance in physics. Another reason is *spin*, particularly its mysterious relationship with Riemannian geometry. Dirac also played a key role in the integration of spin into quantum theory, through his famous Dirac operator invented in 1928. Again, it was not until many decades later, that Atiyah, Bott, Singer and others uncovered the rich geometric-analytic structures underlying spin.

Definition 18. A morphism between two principal G-bundles $P_1 \to X_1$ and $P_2 \to X_2$, is a smooth map $F : P_1 \to P_2$ which is G-equivariant, i.e.,

$$F(p \cdot g) = F(p) \cdot g, \qquad p \in P_1,\ g \in G.$$

Two principal G-bundles P_1, P_2 over X are *isomorphic* if there is a diffeomorphism $F : P_1 \to P_2$ covering the identity id_X.

Exercise 6.2. Check that when P is a principal G-bundle (Definition 17), then the following hold:

- For any $p \in P$, its orbit under the right G-action is precisely the fibre $\pi^{-1}(\pi(p))$.

- The right G-action σ is free (every non-identity element $g \in G$ has no fixed points).

If F is a morphism of principal G-bundles (Definition 18), then

- For each $p \in P$, F restricts to a diffeomorphism between the fibre containing p and the fibre containing $F(p)$; thus F projects down to a well-defined map $f : X_1 \to X_2$, and we say that F *covers* f:

$$\begin{array}{ccc} P_1 & \xrightarrow{F} & P_2 \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle \pi_2} \\ X_1 & \xrightarrow{f} & X_2 \end{array} . \tag{6.3}$$

- The induced $f : X_1 \to X_2$ is smooth (recall Lemma 4.2).
- If f is a diffeomorphism, then so is F, in which case $F^{-1} : P_2 \to P_1$ is also a morphism of principal G-bundles.

Example 6.3. Let $\pi : P \to X$ be a principal G-bundle, and $Z \subset X$ be a submanifold. The subset $P|_Z := \pi^{-1}(Z)$ is a submanifold of P (Prop. 4.5), and is a principal G bundle by itself. The inclusion $P|_Z \hookrightarrow P$ is easily seen to a principal G-bundle morphism, covering the inclusion $Z \hookrightarrow X$.

Definition 19. Let $\pi : P \to X$ be a principal G-bundle. A *local gauge* over an open neighbourhood $U \subset X$ is a smooth section over U, i.e., a smooth map $s : U \to P$ such that $\pi \circ s = \mathrm{id}_U$.

Why emphasize "local" in "local gauge"? If there exists a global trivializing diffeomorphism $\Phi : P \to X \times G$ to the trivial principal G-bundle, then P is said to be (globally) *trivializable*. In this case, we can indeed find a global gauge, i.e. a section $s : X \to P$ defined on all of X. However, non-trivializable principle bundles exist, so there may simply be no such thing as a global gauge!

Let $s : U \to P$ be a local gauge. Then at each $x \in U$, any fibre point $p \in \pi^{-1}(x)$ has the form $s(x) \cdot g$ for some unique label $g = g(p) \in G$. This is because the G-action is free and transitive on each fibre $\pi^{-1}(x)$. Therefore a local gauge s determines a map

$$\Phi_s : P|_U \equiv \pi^{-1}(U) \to U \times G, \qquad (s(x) \cdot g) \mapsto (x, g). \tag{6.4}$$

The role of a local gauge s is to provide a "field of reference frames" over U, so that we may *label* any other frame $p \in P|_U$ by the group element $g = g(p)$ which "rotates" the reference frame $s(\pi(p))$ to p.

Exercise 6.3. Show that the map Φ_s is precisely a trivialization of $P|_U$. The G-equivariance and bijectivity are quite straightforward. The diffeomorphism property is more involved, and requires some general smooth manifold and Lie group theory (Lemma 4.2.7 of [7]).

Given a trivialization $\Phi : P|_U \to U \times G$, we can transport the trivial section $x \mapsto (x, e)$ of $U \times G$ back to $P|_U$, by taking

$$\begin{aligned} s_\Phi : U &\to P|_U \\ x &\mapsto \Phi^{-1}(x, e). \end{aligned} \tag{6.5}$$

It is easily checked that $s_{\Phi_s} = s$ and $\Phi_{s_\Phi} = \Phi$. That is, local sections/gauges over U are in 1-1 correspondence with local trivializations over U.

6.4 Transition functions

Definition 20. A *trivializing cover* for a principal G-bundle $P \to X$, is an open cover $\{U_\alpha\}_{\alpha \in \mathcal{I}}$ of X, together with local trivializations

$$\Phi_\alpha : P|_{U_\alpha} \to U_\alpha \times G, \qquad \alpha \in \mathcal{I}.$$

On a double overlap $U_{\beta\alpha} := U_\beta \cap U_\alpha$, there are apparently two different ways of viewing $P|_{U_{\beta\alpha}}$ as a product bundle $U_{\beta\alpha} \times G$. Specifically, for each $x \in U_{\alpha\beta}$, we have two different labelling conventions for each frame $p \in \pi^{-1}(x)$,

$$\Phi_\alpha(p) = (x, \phi_\alpha(p)) \quad \text{or} \quad \Phi_\beta(p) = (x, \phi_\beta(p)),$$

where ϕ_α, ϕ_β are G-valued labelling maps. So we must have

$$(x, \phi_\beta(p)) = (x, g_{\beta\alpha}(p)\phi_\alpha(p)) \tag{6.6}$$

for some smooth map $g_{\beta\alpha} = \phi_\beta \phi_\alpha^{-1} : P|_{U_{\beta\alpha}} \to G$ (the "relabelling rule"). It is fairly straightforward to see that $g_{\beta\alpha}$ depends only on the basepoint x, due to the compatibility of Φ_α, Φ_β with the fibrewise G-actions. Thus the two labelling conventions ϕ_α, ϕ_β are related by

$$\phi_\beta(p) = g_{\beta\alpha}(x)\phi_\alpha(p),$$

for some *transition function* $g_{\beta\alpha} : U_{\beta\alpha} \to G$.

On triple overlaps, associativity (of group multiplication) implies that transition functions must satisfy the *cocycle condition*,

$$g_{\alpha\beta} g_{\beta\gamma} g_{\gamma\alpha} = 1 \qquad \text{on } U_{\alpha\beta\gamma} := U_\alpha \cap U_\beta \cap U_\gamma, \qquad \alpha, \beta, \gamma \in \mathcal{I}. \tag{6.7}$$

Remark 3. It is important to realize that the transition functions $g_{\alpha\beta}, \alpha, \beta \in \mathcal{I}$, refer to a choice of trivializing cover $(U_\alpha, \Phi_\alpha)_{\alpha\in\mathcal{I}}$ for P. Consider modifying each $\Phi_\alpha = (\pi, \phi_\alpha)$ by picking smooth functions $\lambda_\alpha : U_\alpha \to G$ and defining

$$\Phi'_\alpha = (\pi, \underbrace{\lambda_\alpha^{-1} \cdot \phi_\alpha}_{\phi'_\alpha}).$$

The transition functions obtained by using $(U_\alpha, \Phi'_\alpha)_{\alpha\in\mathcal{I}}$ are then

$$g'_{\alpha\beta} = \phi'_\alpha (\phi'_\beta)^{-1} = \lambda_\alpha^{-1} \cdot \phi_\alpha \phi_\beta^{-1} \cdot \lambda_\beta = \lambda_\alpha^{-1} g_{\alpha\beta} \lambda_\beta.$$

So P only determines transition functions up to some relabellings ("local changes of gauge"). The relation (6.8) says that the transition functions $\{g'_{\alpha\beta}\}$ and $\{g_{\alpha\beta}\}$ differ by a *coboundary*.

More generally, suppose there is another principal G-bundle $P' \to X$, with transition functions $g_{\alpha'\beta'}, \alpha', \beta' \in \mathcal{I}'$ with respect to another trivializing cover $\{U_{\alpha'}, \Phi_{\alpha'}\}_{\alpha'\in\mathcal{I}'}$. To compare the two sets of transition functions, we consider the open cover

$$\{U_\alpha \cap U_{\alpha'}\}_{\alpha\in\mathcal{I}, \alpha'\in\mathcal{I}'}.$$

Every open set in this cover is contained in some open set of $\{U_\alpha\}_{\alpha\in\mathcal{I}}$; likewise for $\{U_{\alpha'}\}_{\alpha'\in\mathcal{I}'}$. We say that the new cover is a *common refinement* of the previous two covers. The unprimed transition functions can be restricted to open sets of the new cover; likewise for the primed transition functions. Now we can compare the two sets of restricted transition functions, since they are both defined with respect to the same open cover of X.

Lemma 6.1. *Let $P \to X$ and $P' \to X$ be two principal G-bundles, with respective transition functions $g_{\alpha\beta}$ and $g'_{\alpha\beta}$ defined with respect to the same open cover $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ of X. Then P and P' are isomorphic iff there exist smooth functions $\lambda_\alpha : U_\alpha \to G$ such that*

$$g'_{\alpha\beta}(x) = \lambda_\alpha(x)^{-1} g_{\alpha\beta}(x) \lambda_\beta(x), \qquad x \in U_{\alpha\beta}, \alpha, \beta \in \mathcal{I}. \tag{6.8}$$

Proof. Exercise. □

6.4.1 Principal bundle via transition functions

Intuitively, transition functions provide the instructions for building a global principal G-bundle out of a collection of local trivial G-bundles $U_\alpha \times G$.

So suppose we are given smooth functions $g_{\alpha\beta} : U_{\alpha\beta} \to G$ satisfying the cocycle condition, Eq. (6.7), with respect to some open cover $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ of X. This is called a (smooth) *Čech 1-cocycle*. Note that the following are also automatically satisfied,

$$g_{\alpha\alpha} = 1, \qquad g_{\alpha\beta} = g_{\beta\alpha}^{-1}, \qquad \alpha, \beta \in \mathcal{I}.$$

With the Čech cocycle $g_{\alpha\beta}$, we can construct a principal G-bundle as follows (for full details, see §3.2 of [14], Prop. 5.2 of [9]).

- Give $\mathcal{I}$ the discrete topology, and take the subset

$$\begin{aligned} T &:= \{(x, g, \alpha) \in X \times G \times \mathcal{I} \,:\, x \in U_\alpha\} \\ &= \bigsqcup_{\alpha\in\mathcal{I}} U_\alpha \times G \times \{\alpha\}, \end{aligned}$$

  which is open. (Why?)

- "Gluing" via $g_{\alpha\beta}$ means to take equivalence classes

$$(x, g, \alpha) \sim (x', g', \beta) \quad \text{iff} \quad x = x' \;\text{ and }\; g' = g_{\beta\alpha} g,$$

  which makes sense because of the 1-cocycle condition, Eq. (6.7), (Exercise). So take P to be the set of equivalence classes, with the quotient topology, and define the projection map to be $\pi[x, g, \alpha] = x$, and the right G-action to be $[x, g, \alpha] \cdot h = [x, gh, \alpha]$. Check well-definedness of π and the G-action, that P is a manifold, and that the G-action is smooth.

- Local trivializations of P are given by

$$\Phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times G, \qquad [x, g, \alpha] \mapsto (x, g), \qquad \alpha \in \mathcal{I}.$$

  (Check diffeomorphism and G-equivariance properties).

- Inspect that, by construction, the transition functions between the local trivializations of P are precisely $g_{\alpha\beta}$.

So if the $g_{\alpha\beta}$ came from some trivializing cover for P_{initial}, then the principal bundle reconstructed out of $g_{\alpha\beta}$ is isomorphic to P_{initial}. Due to Lemma 6.1, passing to a refinement and modifying the transition functions as in Eq. (6.8) will not change the isomorphism class of the reconstructed bundle. So we see that principal G-bundles over X are classified, up to isomorphism, by equivalence classes of Čech 1-cocycles $g_{\alpha\beta}$ on X. This is known as the *Čech cohomology* classification.

6.5 Frame bundle of a vector bundle

Let us return to the frame bundle of a general, possibly non-trivializable, vector bundle E. As a set, $\mathrm{Fr}(E)$ is again Eq. (6.2), with the same projection and fibrewise right G-action, where $G = \mathrm{GL}(n)$. We explain why this frame bundle has a principal G-bundle structure.

Let $(U_\alpha, \Phi_\alpha)_{\alpha\in\mathcal{I}}$ be a trivializing cover for E, with smooth transition functions $g_{\alpha\beta}$. Each local trivialization $\Phi_\alpha : E|_{U_\alpha} \cong U_\alpha \times \mathbb{K}^n$ gives a corresponding trivialization of the local frame bundle,

$$\tilde{\Phi}_\alpha : \mathrm{Fr}(E|_{U_\alpha}) \to U_\alpha \times G, \qquad G = \mathrm{GL}(n),$$

as in Section 6.2.1. Here, the $\tilde{\Phi}_\alpha$ are G-equivariant and respect the projection maps to the base. Each $(x, g) \in (U_\alpha \cap U_\beta) \times G$ is a basis $\{g_1, \ldots, g_n\}$ of $\mathbb{K}^n$ at x, where the g_i denote the columns of g. Then $\tilde{\Phi}_\alpha^{-1}(x, g)$ is some frame $\{e_{1,x}, \ldots, e_{n,x}\}$ of E_x. If we switch to the β local trivialization of E, then this frame would be mapped by $\tilde{\Phi}_\beta$ to the basis $\{g_{\beta\alpha}(x)g_1, \ldots, g_{\beta\alpha}(x)g_1\}$ of $\mathbb{K}^n$ at x, i.e., the element $(x, g_{\beta\alpha}(x) \cdot g) \in U_\beta \times G$. Thus, we have

$$\begin{aligned} \tilde{\Phi}_\beta \circ \tilde{\Phi}_\alpha^{-1} : U_{\beta\alpha} \times G &\to U_{\beta\alpha} \times G \\ (x, g) &\mapsto (x, g_{\beta\alpha}(x) \cdot g), \end{aligned}$$

with the same smooth transition functions $g_{\beta\alpha}$.

We still lack the smooth manifold structure on $\mathrm{Fr}(E)$, compatible with the smooth structures on the local frame bundles $\mathrm{Fr}(E|_{U_\alpha})$. For this, we first topologize $\mathrm{Fr}(E)$ by declaring $V \subset \mathrm{Fr}(E)$ to be open iff $V \cap \mathrm{Fr}(E|_{U_\alpha})$ is open for all $\alpha \in \mathcal{I}$. The manifold charts of all the $\mathrm{Fr}(E|_{U_\alpha}) \subset \mathrm{Fr}(E)$ combine to make $\mathrm{Fr}(E)$ a topological manifold, with smoothness of coordinate changes inherited from that on E.

Remark. The vector bundle transition functions $g_{\alpha\beta}$ give a Čech 1-cocycle. From this, we can also abstractly construct a principal G-bundle over X, following the prescription of Section 6.4.1. This produces an isomorphic bundle to the more concrete frame bundle described above, due to Lemma 6.1.

6.6 Gauge transformations

6.6.1 Local gauge transformations

Earlier, we saw that a local trivialization over U_α is the same thing as a local section/gauge $s_\alpha : U_\alpha \to P|_{U_\alpha}$. So specifying a trivializing cover is the same thing as equipping an open cover $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ with local sections s_α. Explicitly, the local trivialization associated with s_α is

$$\Phi_{s_\alpha} : (\underbrace{s_\alpha(x)\cdot g}_{p\in P|_{U_\alpha}}) \mapsto (\underbrace{x}_{\pi(p)}, \underbrace{g}_{\phi_\alpha(p)}).$$

So if we have two local sections/gauges, $s_\alpha : U_\alpha \to P$ and $s_\beta : U_\beta \to P$, there are two ways of writing $p \in \pi^{-1}(x)$ for points $x \in U_{\alpha\beta}$ on the overlap,

$$s_\alpha(x)\cdot\phi_\alpha(p) = p = s_\beta(x)\cdot\phi_\beta(p).$$

Thus, the *change* of frame labelling when switching local gauges is precisely implemented by the transition function,

$$\phi_\beta(p)\phi_\alpha^{-1}(p) \equiv g_{\beta\alpha}(x).$$

We could also think of the transition function as implementing the change of section/gauge,

$$s_\alpha(x) = s_\beta(x)\cdot\phi_\beta(p)\phi_\alpha^{-1}(p) = s_\beta(x)\cdot g_{\beta\alpha}(x). \tag{6.9}$$

In this capacity, the map $g_{\beta\alpha} : U_{\beta\alpha} \to G$ is referred to as a *local gauge transformation* (from the β to the α gauge). We will discuss this local viewpoint further in Section 6.6.3.

6.6.2 Global gauge group

The *global* and *gauge-independent* notion of a gauge transformation is:

Definition 21. A *gauge transformation* of a principal G bundle $P \to X$ is an automorphism of P (as a principal G-bundle) which covers the identity map id_X. The group $\mathcal{G}(P)$ of gauge transformations is called the *gauge group* of P.

Example 6.4. Take $X = \mathrm{pt}$, then P is just a G-torsor. A gauge transformation is a self-map $F : P \to P$ which respects its structure as a G-torsor. So F satisfies

$$F(p \cdot g) = F(p) \cdot g, \qquad \forall g \in G, p \in P. \tag{6.10}$$

In other words, the G-relations between frames are preserved under the transformation F.

Now, the target frames $F(p)$ must have the form

$$F : p \mapsto p \cdot \sigma_F(p), \tag{6.11}$$

for some uniquely-defined $\sigma_F(p) \in G$. That is, F must be implemented by right-multiplication by some smooth map $\sigma_F : P \to G$. Putting Eq. (6.11) into Eq. (6.10) gives

$$F(p \cdot g) = F(p) \cdot g = (p \cdot \sigma_F(p)) \cdot g = (p \cdot g) \cdot \underbrace{(g^{-1}\sigma_F(p)g)}_{=\sigma_F(p \cdot g)}. \tag{6.12}$$

Therefore, the required $\sigma_F : P \to G$ is a G-equivariant map, where the action on the target G is by conjugation.

This looks a little complicated, so consider a situation where we had a reference frame $p_0 \in P$, but wanted to switch to a shifted reference frame $p_0 \cdot h$, where $h \in G$. *We do not simply shift all the other frames by h.* Instead, the correct gauge transformation F_h must be defined on a general frame $p = p_0 \cdot g$ by the formula

$$F_h(p) = F(p_0 \cdot g) = F_h(p_0) \cdot g = (p_0 \cdot h) \cdot g = p \cdot (g^{-1}hg).$$

Specifying gauge group. For P over a general base space X, a gauge transformation F still respects each fibre as a G-torsor. The calculations in Example 6.4 carry over in the same way, leading to Eq. (6.12). That is, the gauge group is naturally identified with the following space of equivariant G-valued maps on P,

$$\mathcal{G}(P) \cong \mathrm{Map}(P, G)^G := \{\sigma : P \to G \text{ smooth} \,:\, \sigma(p \cdot g) = g^{-1}\sigma(p)g, \;\; \forall p \in P, g \in G\},$$

acting on P by multiplication on the right.

Exercise 6.4. Check that $\mathrm{Map}(P,G)^G$ is a group under pointwise multiplication, and that the correspondence

$$\mathcal{G}(P) \to \mathrm{Map}(P,G)^G, \qquad F \mapsto \sigma_F,$$

where σ_F is defined by Eq. (6.11), is indeed a group isomorphism.

Abelian gauge transformations.

Example 6.5. If G is Abelian, then

$$\mathrm{Map}(P,G)^G = \{\sigma : P \to G \text{ smooth } : \sigma(p \cdot g) = \sigma(p) \ \ \forall p \in P,\, g \in G\}.$$

Such a map is constant on the fibres, and descends to X. So we have a canonical isomorphism

$$\mathrm{Map}(X,G) \to \mathrm{Map}(P,G)^G \cong \mathcal{G}(P), \qquad \tau \mapsto \tau \circ \pi.$$

Example 6.6. Consider the principal bundle $P \cong S^1 \times \mathrm{U}(1)$, with S^1 the unit circle. For $n \in \mathbb{Z}$, the maps $g_n : e^{i\theta} \mapsto e^{in\theta}$ are elements of $\mathrm{Map}(S^1, \mathrm{U}(1))$, and they implement gauge transformations of P. The g_n with different values of n are not smoothly deformable to each other within $\mathrm{Map}(S^1, \mathrm{U}(1))$. Actually, any $g \in \mathrm{Map}(S^1, \mathrm{U}(1))$ is smoothly deformable to some g_n, so the gauge transformations of P are sorted into deformation classes labelled by $n \in \mathbb{Z}$ (which is the winding number of g). Those with winding number $n \neq 0$ are called *large gauge transformations* since they are not deformable to the identity transformation.

6.6.3 Local viewpoint

Let $P|_U$ be trivialized by a local section $s : U \to P|_U$. Let $\sigma \in \mathrm{Map}(P|_U, G)^G$ be a gauge transformation of $P|_U$ ("local gauge transformation"). Then the composition $\tau(\sigma) := \sigma \circ s$ is an element of $\mathrm{Map}(U,G)$. We may then use $\tau(\sigma)$ to change the local section/gauge to

$$s \cdot \tau(\sigma) : U \to P|_U.$$

We check that for $\sigma_1, \sigma_2 \in \mathrm{Map}(P|_U, G)^G$, we have

$$\tau(\sigma_1) \cdot \tau(\sigma_2)(x) = (\sigma_1 \circ s(x)) \cdot (\sigma_2 \circ s(x)) = (\sigma_1 \cdot \sigma_2) \circ s(x) = \tau(\sigma_1 \cdot \sigma_2)(x),$$

so

$$\tau : \mathrm{Map}(P|_U, G)^G \to \mathrm{Map}(U, G), \qquad \sigma \mapsto \tau(\sigma)$$

is a group homomorphism.

In reverse, let $f \in \mathrm{Map}(U, G)$. Define a map

$$\eta : \mathrm{Map}(U, G) \to \mathrm{Map}(P|_U, G)^G$$

by

$$\eta(f)(s(x) \cdot g) = g^{-1} f(x) g, \qquad x \in U,\ g \in G.$$

(Equivariance of $\eta(f)$ is clear from this formula.) Given $f_1, f_2 \in \mathrm{Map}(U, G)$, we check that

$$\begin{aligned}\eta(f_1 \cdot f_2)(s(x) \cdot g) = g^{-1}(f_1(x) f_2(x))g &= (g^{-1} f_1(x) g)(g^{-1} f_2(x) g)) \\ &= \eta(f_1)(s(x) \cdot g)\eta(f_2)(s(x) \cdot g) \\ &= (\eta(f_1) \cdot \eta(f_2))(s(x) \cdot g),\end{aligned}$$

therefore $\eta(f_1 \cdot f_2)$ and $\eta(f_1) \cdot \eta(f_2)$ coincide.

One may verify that τ and η are inverse maps. That is, they implement group isomorphisms

$$\mathrm{Map}(P|_U, G)^G \longleftrightarrow \mathrm{Map}(U, G).$$

This is why, in physics, *local* gauge transformations are often simply regarded as maps from the base U to the symmetry group G.

However, the identification of the local gauge transformations over U with $\mathrm{Map}(U, G)$ uses a choice of reference local gauge s. So, for instance, gauge transformations of non-trivializable P cannot simply be described as $\mathrm{Map}(X, G)$ (in the case of non-Abelian G).

What *are* gauge transformations? It is hard to give an answer that will satisfy all practitioners. Here is a possible one. As we will learn later, the gauge group is supposed to act on some other spaces of objects, such as connections on P, vector bundles associated to P, etc. If we "mod out" by the action of the gauge group, we are really considering P only as an object *up to isomorphism*. We had been thinking of P as an abstract frame bundle, and each element of P as some frame located at some point $x \in X$. But if we ultimately only care about the isomorphism class $[P]$ of P, then the

frames are not actually being regarded as things-in-themselves. Rather, only the abstract *inter-frame relations* are considered important. *Gauge-invariant* quantities are those which are intrinsic to this relational structure, as encoded by $[P]$.

7 Preview of Spin

7.1 Hopf bundle over S^2

Consider S^3 as the set of unit vectors in $\mathbb{C}^2$,

$$S^3 = \{(w, z) \in \mathbb{C}^2 \,:\, |w|^2 + |z|^2 = 1\}.$$

The *Hopf fibration* is the smooth surjective map to the unit 2-sphere,

$$\begin{aligned} \pi : S^3 &\to S^2 \subset \mathbb{R}^3 = \mathbb{C} \times \mathbb{R} \\ (w, z) &\mapsto (2w\bar{z}, |w|^2 - |z|^2). \end{aligned}$$

On S^3, there is a smooth right action of $e^{i\theta} \in \mathrm{U}(1)$ by scalar multiplication, $(w, z) \cdot e^{i\theta} = (we^{i\theta}, ze^{i\theta})$. It is easy to check that

$$\pi(w_1, z_1) = \pi(w_2, z_2) \quad \text{iff} \quad (w_2, z_2) = (w_1, z_1) \cdot e^{i\theta} \text{ for some } e^{i\theta} \in \mathrm{U}(1).$$

So π is precisely the quotient map under the action of U(1).

The north and south poles of S^2 are

$$N = ((0, 0), 1), \qquad S = ((0, 0), -1),$$

and we shall consider the open cover

$$U_\alpha = S^2 \setminus \{N\}, \qquad U_\beta = S^2 \setminus \{S\},$$

with preimages

$$\pi^{-1}(U_\alpha) = \{(w, z) \in S^3 \,:\, z \neq 0\}, \qquad \pi^{-1}(U_\beta) = \{(w, z) \in S^3 \,:\, w \neq 0\}.$$

Define the maps

$$\begin{aligned} \Phi_\alpha : \pi^{-1}(U_\alpha) &\to U_\alpha \times \mathrm{U}(1) \\ (w, z) &\mapsto (\pi(w, z); z/|z|) \end{aligned} \qquad \begin{aligned} \Phi_\beta : \pi^{-1}(U_\beta) &\to U_\beta \times \mathrm{U}(1) \\ (w, z) &\mapsto (\pi(w, z); w/|w|) \end{aligned}$$

These provide local trivializations to trivial principal U(1)-bundles. So $\pi : S^3 \to S^2$ is a principal U(1)-bundle, with transition function being

$$g_{\beta\alpha} : U_\alpha \cap U_\beta \to \mathrm{U}(1), \qquad (2w\bar{z}, |w|^2 - |z|^2) \mapsto \frac{w/|w|}{z/|z|} = \exp(i\mathrm{Arg}(w\bar{z})).$$

Note that the overlap $U_\alpha \cap U_\beta \subset S^2$ retracts to the equatorial circle $S^1 \subset S^2$, and it is easy to see that the transition function $g_{\beta\alpha}|_{S^1} : S^1 \to \mathrm{U}(1)$ has winding number 1. This topologically obstructs the extension of $g_{\beta\alpha}$ to include the north and/or south poles. The significance of this obstruction is as follows. Suppose there is a global trivialization $\tilde{\Phi} : S^3 \cong \mathbb{CP}^1 \times \mathrm{U}(1)$. Then there are transition functions $\tilde{\Phi}_\alpha \circ \tilde{\Phi}^{-1}$ and $\tilde{\Phi}_\beta \circ \tilde{\Phi}^{-1}$, both of which must have winding number zero when restricted to the equator, since they can be extended to the south pole and the north pole respectively. This would contradict the non-trivial winding number of $g_{\beta\alpha} = \tilde{\Phi}_\beta \circ \tilde{\Phi}_\alpha^{-1}$. Thus S^3 is an example of a *non-trivializable* principal U(1)-bundle.

Remark. There is also the *quaternionic* Hopf fibration $\pi : S^7 \to S^4$. Here, S^7 is regarded as the unit-length quaternions in $\mathbb{H}^2$, and there is a right "scalar multiplication" on S^7 by the unit quaternion group $\mathrm{Sp}(1) \cong S^3$. The quotient manifold S^4 is identified with the quaternionic projective space $\mathbb{HP}^1$. As with the complex Hopf fibration, there are local trivializations over the upper and lower hemispheres of $\mathbb{HP}^1$, with transition function on the equator $S^3 \subset \mathbb{HP}^1$ being a map $g_{\beta\alpha} : S^3 \to \mathrm{Sp}(1)$ of degree 1. Thus $S^7 \to S^4$ is a non-trivializable principal Sp(1)-bundle. The interested reader is invited to work out the details.

7.2 Orthonormal frame bundle of S^2

The Euclidean vector space $\mathbb{R}^d$ has a standard inner product,

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{d} x^i y^i, \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

Recall that each of its tangent spaces is itself canonically isomorphic to $\mathbb{R}^d$, and we give them the same inner product. This means that we consider Euclidean space as a Riemannian manifold, whose tangent bundle $\mathbb{R}^d \times \mathbb{R}^d$ is equipped with the standard Euclidean metric on each fibre $\{\mathbf{x}\} \times \mathbb{R}^d$. Furthermore, there is an orientation given by the standard basis.

Take the unit 2-sphere $S^2 \subset \mathbb{R}^3$, whose points are labelled by a unit 3-vector $\mathbf{a}$. The tangent plane at $\mathbf{a}$ is

$$T_{\mathbf{a}}\mathbb{R}^3 = \{\mathbf{v} \in \mathbb{R}^3 \,:\, \mathbf{v} \cdot \mathbf{a} = 0\}.$$

In total, the tangent bundle TS^2 is the subset

$$TS^2 = \{(\mathbf{a}, \mathbf{v}) \in S^2 \times \mathbb{R}^3 \,:\, \mathbf{v} \cdot \mathbf{a} = 0\},$$

and it is a rank-2 Euclidean vector bundle over S^2. Actually, we can orient each tangent space by declaring a basis $\{\mathbf{u}, \mathbf{v}\}$ of $T_{\mathbf{a}}S^2$ to be positively oriented if $\{\mathbf{u}, \mathbf{v}, \mathbf{a}\}$ is positively oriented for $\mathbb{R}^3$. So S^2 is an oriented Riemannian manifold.

We can consider the oriented orthonormal frame bundle of S^2, denoted $\pi : \mathrm{Fr}^{\mathrm{SO}}(S^2) \to S^2$. By general theory to be developed later, this would be a principal SO(2)-bundle, with the SO(2) action being rotation of oriented frames for each tangent plane. For now, let us obtain this structure directly.

Write each element of $\mathrm{Fr}^{\mathrm{SO}}(S^2)$ as

$$(\{\mathbf{u}, \mathbf{v}\}; \mathbf{a}), \qquad \mathbf{a} \in S^2,\ \{\mathbf{u}, \mathbf{v}\} \text{ oriented orthonormal basis for } T_{\mathbf{a}}S^2.$$

The rotation of frames by $g_0 \in \mathrm{SO}(2)$ is

$$(\{\mathbf{u}, \mathbf{v}\}; \mathbf{a}) \cdot g_0 = (\{\mathbf{u}, \mathbf{v}\} \cdot g_0\,;\, \mathbf{a}).$$

Now, observe that $\{\mathbf{u}, \mathbf{v}, \mathbf{a}\}$ is a positively-oriented orthonormal basis for $\mathbb{R}^3$, so the matrix $(\mathbf{u}\ \mathbf{v}\ \mathbf{a})$ is an SO(3) matrix. Thus, we have an identification

$$\mathrm{Fr}^{\mathrm{SO}}(S^2) \cong \mathrm{SO}(3), \qquad (\{\mathbf{u}, \mathbf{v}\}; \mathbf{a}) \leftrightarrow (\mathbf{u}\ \mathbf{v}\ \mathbf{a}), \tag{7.1}$$

with π being the projection onto the last component $\mathbf{a} \in S^2$. Under this identification, the action of SO(2) becomes

$$\big((\mathbf{u}\ \mathbf{v})g_0\ \ \mathbf{a}\big) = (\mathbf{u}\ \mathbf{v}\ \mathbf{a}) \begin{pmatrix} g_0 & 0 \\ 0 & 1 \end{pmatrix},$$

i.e., right multiplication by the subgroup

$$G_0 = \left\{ \begin{pmatrix} g_0 & 0 \\ 0 & 1 \end{pmatrix},\ g_0 \in \mathrm{SO}(2) \right\}.$$

Thus the principal SO(2)-bundle structure of $\mathrm{Fr}^{\mathrm{SO}}(S^2) \to S^2$ is identified with the principal G_0-bundle structure of SO(3).

7.3 Bundle of spin frames over S^2

To recapitulate, we have the principal SO(2)-bundle

$$\pi : \mathrm{SO}(3) \cong \mathrm{Fr}^{\mathrm{SO}}(S^2) \to S^2,$$

as well as the principal U(1)-bundle

$$\pi : S^3 \to S^2.$$

Actually, we shall use the dual U(1)-bundle

$$\bar{\pi} : S^3 \to S^2, \qquad (w, z) \mapsto (2\bar{w}z, |w|^2 - |z|^2),$$

which has clutching transition function $S^1 \to \mathrm{U}(1)$ having winding -1 instead of $+1$.

Now, S^3 is the manifold underlying the Lie group SU(2),

$$S^3 \cong \mathrm{SU}(2), \qquad (w, z) \longleftrightarrow \begin{pmatrix} w & -\bar{z} \\ z & \bar{w} \end{pmatrix}.$$

Under this identification, the corresponding U(1)-action on SU(2) is given by right multiplication by the subgroup $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$.

We shall make use of the famous $2:1$ group homomorphism $F : \mathrm{SU}(2) \to \mathrm{SO}(3)$ (exercise), indicated in the following commutative diagram:

$$\begin{tikzcd} S^3 \cong \mathrm{SU}(2) \arrow[r, "F"] \arrow[dr, "\bar{\pi}"'] & \mathrm{SO}(3) \arrow[d, "\pi"] \\ & S^2 \subset \mathbb{C} \times \mathbb{R} \end{tikzcd} \qquad \begin{tikzcd} \begin{pmatrix} w & -\bar{z} \\ z & \bar{w} \end{pmatrix} \arrow[r, "F"] \arrow[dr, "\bar{\pi}"'] & \begin{pmatrix} \mathrm{Re}(w^2-\bar{z}^2) & \mathrm{Im}(w^2+\bar{z}^2) & 2\,\mathrm{Re}(\bar{w}z) \\ -\mathrm{Im}(w^2-\bar{z}^2) & \mathrm{Re}(w^2+\bar{z}^2) & 2\,\mathrm{Im}(\bar{w}z) \\ -2\,\mathrm{Re}(wz) & -2\,\mathrm{Im}(wz) & |w|^2-|z|^2 \end{pmatrix} \arrow[d, "\pi"] \\ & (2\bar{w}z, |w|^2 - |z|^2) \end{tikzcd} . \tag{7.2}$$

The map F restricts to a double-covering,

$$\chi : \mathrm{U}(1) \to \mathrm{SO}(2), \qquad \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \mapsto \begin{pmatrix} \cos(2\theta) & \sin(2\theta) & 0 \\ -\sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In its capacity as a double-covering of SO(2), the group U(1) is called the *spin group*, Spin(2), and we say that $\bar{\pi} : S^3 \cong \mathrm{SU}(2) \to S^2$ is a principal Spin(2)-bundle. To stress this principal bundle structure, we write $S^3 \cong \mathrm{Fr}^{\mathrm{Spin}}(S^2)$.

The commuting diagram Eq. (7.2) may be summarized as

$$F(S \cdot g) = F(s) \cdot \chi(g), \qquad \bar{\pi}(S) = \pi(F(S)).$$

The key observation is that we have "upgraded" the oriented orthonormal frame bundle $\mathrm{Fr}^{\mathrm{SO}}(S^2)$ to some "spin frame bundle" $\mathrm{Fr}^{\mathrm{Spin}}(S^2)$, which has "twice as many frames" at each point. Such an "upgrade" is mathematically called a *spin structure* on S^2 (see Section 16.1).

Why do we stress the label Spin(2) instead of U(1)? Why the word "Spin"? First, the double cover $\mathrm{Spin}(2) \to \mathrm{SO}(2)$ means that completing a full loop in the orthonormal frame space SO(2) only corresponds to completing *half* a loop in Spin(2). So we do not get back to where we started from, but rather acquire a -1 mismatch, ending up with the "opposite spin frame" to the original one. This is precisely the characteristic feature of *fermionic* fields in physics.

A fermion field is not like the classical geometry vector/tensor fields. On an n-dimensional oriented Riemannian manifold X, the space of "spin frames" at each point is labelled by the "spin rotation group" $\mathrm{Spin}(n)$ rather than the classical rotation group $\mathrm{SO}(n)$. Globally, there is a spin frame bundle $\mathrm{Fr}^{\mathrm{Spin}}(X)$ over X.

In general, there exist *topological* obstructions to defining the spin frame bundle globally, and there may not be a unique choice of spin structure. The classification of spin structures on X is *not* the same thing as the classification of principal $\mathrm{Spin}(n)$-bundles over X — the doubling relation between a spin frame bundle and the oriented orthonormal frame bundle of X is a defining ingredient for the notion of spin and spinors. We also mention that Spin(2) is not the same as the U(1) appearing in quantum mechanics or electromagnetism discussed in the introduction.

Incidentally, for $n = 3$, the double cover $\mathrm{SU}(2) \to \mathrm{SO}(3)$ exhibits $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$ as the space of "spin rotations" at a point on a 3-dimensional manifold.

8 Tensors and differential forms

8.1 Commutator of vector fields

Over a coordinate chart U, the coordinate tangent vector fields ∂_i provide a commuting family of derivations of $C^\infty(U)$. This expresses the convenient notion that partial derivatives commute. Unfortunately, ∂_i is seldom globally defined.

For general vector fields $u, v \in \mathfrak{X}(X)$, we will acquire a commutator

$$[u, v] : u \circ v - v \circ u$$

whenever we swap their order of operation,

$$u(v(f)) = v(u(f)) + [u, v](f).$$

The commutator of two derivations turns out to be another derivation (Exercise), so $[u, v]$ is also a vector field. Formally, $[\cdot, \cdot]$ is a *Lie bracket* operation on $\mathfrak{X}(X) \times \mathfrak{X}(X)$, satisfying the following properties

$$\begin{aligned} [\lambda u + v, w] &= \lambda[u, w] + [v, w] \\ [v, w] &= -[w, v] \\ 0 &= [u, [v, w]] + [v, [w, u]] + [w, [u, v]], \qquad u, v, w \in \mathfrak{X}(X),\ \lambda \in \mathbb{R}. \end{aligned} \tag{8.1}$$

Thus $\mathfrak{X}(X)$ is an infinite-dimensional *Lie algebra*.

Exercise 8.1. Verify that

$$[f \cdot u, g \cdot v] = fg \cdot [u, v] + f(u(g)) \cdot v - g(v(f)) \cdot u, \qquad v, w \in \mathfrak{X}(X), f, g \in C^\infty(X).$$

So the commutator is not C^∞-linear in either entry. Consequently, $[\cdot, \cdot]$ is not a "tensorial" operation — it does not make sense for individual tangent vectors. Nevertheless, the Lie bracket is *natural*:

Exercise 8.2. Let $u, v \in \mathfrak{X}(X)$, and $f : X \to Y$ be a diffeomorphism. Show that

$$f_*[u, v] = [f_* u, f_* v].$$

8.2 Differential 1-forms

Recall Definition 6 of the cotangent spaces $T_x^*X := (T_xX)^*$. From doing Exercise 4.4, we learn that the basis $\{(dx^i)_x\}$ for T_x^*X is dual to the basis of coordinate tangent vectors,

$$(dx^i)_x(\partial_j|_x) = \delta^i_j.$$

A general $\omega \in T_x^*X$ can therefore be expanded as $\omega = \omega_i(dx^i)_x$.

Exercise 8.3. Show that the components $\omega_i = \omega\big(\partial_i|_x\big)$ transform *covariantly* as

$$\omega_i = \frac{\partial \tilde{x}^j}{\partial x^i}\Big|_{\varphi(x)} \tilde{\omega}_j \qquad \text{or} \qquad \tilde{\omega}_j = \frac{\partial x^i}{\partial \tilde{x}^j}\Big|_{\psi(x)} \omega_i \tag{8.2}$$

under a change of coordinates $x^i \to \tilde{x}^j$. (Cf. contravariant transformation property for the components of tangent vectors, Eq. (4.3).)

Definition 22. A (differential) 1-form on a manifold X is an assignment of cotangent vectors, $\omega : x \mapsto \omega_x \in T_x^*X$, such that its component functions in some (thus any) coordinate chart are smooth. The space of 1-forms on X is denoted $\Omega^1(X)$.

Example 8.1. Over a coordinate chart U, we have the *coordinate 1-forms* $(dx^i), i = 1, \ldots, n$, dual to the coordinate tangent vector fields ∂_i.

Example 8.2. We can cover the unit circle S^1 with two charts, with angular coordinates $\theta \in (0, 2\pi)$ and $\tilde{\theta} \in (-\pi, \pi)$. The change-of-coordinates map is the identity map on the upper half-circle, and the shift of -2π on the lower half-circle. The coordinate vector fields $\frac{\partial}{\partial\theta}$ and $\frac{\partial}{\partial\tilde{\theta}}$ coincide on the overlap, and we usually just abuse notation and write $\frac{\partial}{\partial\theta}$ for the globally defined tangent vector field over S^1. Similarly the "coordinate" 1-form $d\theta$ is globally defined over S^1.

The pairing of 1-forms with vector fields works as follows. Let $\omega \in \Omega^1(X)$ and $v \in \mathfrak{X}(X)$. Then we can define the function $\omega(v) : X \to \mathbb{R}$ to be

$$\omega(v)(x) = \omega_x(v_x) \in \mathbb{R}.$$

Expanding in an arbitrary coordinate basis,

$$\omega_x(v_x) = \sum_{i,j=1}^n \omega_i(x)(dx^i)_x\left(v^j(x)\partial_j|_x\right) = \sum_{i,j=1}^n \omega_i(x)v^j(x)\delta^i_j = \sum_{i=1}^n \omega_i(x)v^i(x),$$

which depends smoothly on x, since the components ω_i and v^i depend smoothly on x. Thus we may regard a 1-form as a map $\omega : \mathfrak{X}(X) \to C^\infty(X)$.

Exercise 8.4. Show that a 1-form ω is $C^\infty(X)$-linear, in the sense that

$$\omega(f \cdot v) = f \cdot \omega(v), \qquad f \in C^\infty(X),\ v \in \mathfrak{X}(X).$$

Conversely, show that any $C^\infty(X)$-linear map $\omega : \mathfrak{X}(X) \to C^\infty(X)$ must be a 1-form.

Example 8.3. Let us revisit the derivative df of a *function* $f \in C^\infty(X)$. So $f : X \to \mathbf{R}$ as a map of manifolds, and $df : TX \to T\mathbf{R} = \mathbf{R} \times \mathbb{R}$. Given any vector field $v \in \mathfrak{X}(X)$, we have

$$df(v)(x) := df_x(v_x) : T_xX \to T_{f(x)}\mathbf{R} = \mathbb{R},$$

so $df(v)$ is a $\mathbb{R}$-valued function on X, which depends smoothly on x, since

$$df_x(v_x) = v_x(f) = v(f)(x),$$

and the right-hand side depends smoothly on x (Exercise 5.1). Thus we can regard df as a 1-form,

$$\begin{aligned} df : \mathfrak{X}(X) &\to C^\infty(X) \\ v &\mapsto v(f). \end{aligned}$$

Cf. Example 4.6, where we saw that df_x was a cotangent vector at x. If you prefer, write in local coordinates,

$$v(f) = \frac{\partial \check{f}}{\partial x^i} v^i = \underbrace{\left(\frac{\partial \check{f}}{\partial x^i} dx^i \right)}_{df} \underbrace{(v^j \partial_j)}_{v}.$$

Although we might be used to thinking of the n-tuple $\frac{\partial \check{f}}{\partial x^i}$ as a "gradient vector field", it is perhaps better to think of it as a 1-form, which eats up various tangent vectors to give directional derivatives of f.

Remark. In the same way as for TX, we can construct the *cotangent bundle* T^*X as a $(2n)$-manifold. Alternatively, when we study TX as fibre bundles later, we will see that T^*X may be viewed as the vector bundle with transition functions being the transpose of those of TX. Then 1-forms are sections of the cotangent bundle T^*X.

8.3 Pullbacks and push-forwards

It is usual to write $\Omega^0(X) := C^\infty(X)$, and refer to smooth functions as "0-forms". If $g \in \Omega^0(Y)$, then it may be pre-composed with a smooth map $f : X \to Y$ to obtain a smooth 0-form

$$f^*(g) \in \Omega^0(X), \qquad f^*(g) := g \circ f.$$

This is called the *pullback* of g by the map f.

Similarly, let $\omega \in \Omega^1(Y)$. Then there is a *pullback 1-form* $f^*\omega \in \Omega^1(X)$, defined by

$$(f^*\omega)(v)(x) = \omega_{f(x)}(df_x(v_x)), \qquad v \in \mathfrak{X}(X). \tag{8.3}$$

We should verify that $f^*\omega(v)$ is indeed a smooth function of x. This can be shown from the following properties of the pullback:

Exercise 8.5. Let $\omega \in \Omega^1(Y)$ be a 1-form, and let $f : X \to Y$ and $g \in C^\infty(Y)$ be smooth maps. Verify the following formulae:

- $f^*(g \cdot \omega) = (f^*(g)) \cdot f^*\omega$.
- $f^*(dg) = d(f^*(g))$.

Exercise 8.6. Let $f_1 : X \to Y$ and $f_2 : Y \to Z$ be smooth maps. Check that

$$(f_2 \circ f_1)^* = f_1^* \circ f_2^*$$

when applied to $\Omega^1(Z)$ or to $\Omega^0(Z)$.

Given a smooth map $f : X \to Y$, we could try to "push-forward" a vector field on X to become a vector field on Y. Indeed, we have the derivative map $df : TX \to TY$, built out of the pointwise assignments $df_x : T_xX \to T_{f(x)}Y$. However, because f may not be a bijection, df may not map sections of TX to sections of TY. Nevertheless, if f is a diffeomorphism, then we can use f to push forward a vector field to another vector field.

Exercise 8.7. Let $v \in \mathfrak{X}(X)$ and $f : X \to Y$ be a diffeomorphism. Define f_*v by

$$\begin{aligned} f_*v &:= df \circ v \circ f^{-1} : Y \to TY \\ (f_*v)_y &= df_{f^{-1}(y)}(v_{f^{-1}(y)}) \ \in T_yY, \qquad y \in Y. \end{aligned} \tag{8.4}$$

Check that f_*v is indeed a smooth section of TY, i.e., a smooth vector field over Y.

8.4 2-forms and Riemannian metrics

The dual space V^* of a (real) vector space can be generalized to the space of *bilinear* maps (or forms) $V \times V \to \mathbb{R}$. A bilinear form b is *non-degenerate* if

$$b(v_1, v_2) = 0 \text{ for all } v_2 \in V \text{ implies } v_1 = 0.$$

In this case, the bilinear form induces an isomorphism

$$\flat : V \to V^*, \qquad v \mapsto (u \mapsto b(u, v)),$$

with inverse denoted $\sharp : V^* \to V$ (this should not be confused with the same notation used for so-called fundamental vector fields later).

A symmetric (resp. antisymmetric) bilinear form b satisfies $b(v_1, v_2) = b(v_2, v_1)$ (resp. $b(v_1, v_2) = -b(v_2, v_1)$) for all $v_1, v_2 \in V$. (Anti)symmetric forms constitute a vector space.

Example 8.4. An *inner product* on a real vector space V is a symmetric bilinear form g which is *positive definite* ($g(v, v) \geq 0$ with equality iff $v = 0$). This is automatically non-degenerate. There is an induced inner product on V^*, denoted with the same symbol, and given by $g(\eta, \omega) = g(\eta^\sharp, \omega^\sharp)$.

Example 8.5. If $\eta, \omega \in V^*$, their *tensor product* is the bilinear form

$$\eta \otimes \omega : (u, v) \mapsto \eta(u)\omega(v).$$

If $\{e_1, \ldots, e_n\}$ is a basis for V and $\{e^1, \ldots e^n\}$ the dual basis for V^*, then $e^i \otimes e^j$ is a basis for the bilinear forms on V (Exercise).

Example 8.6. We can symmetrize $\eta \otimes \omega$ by taking $\frac{1}{2}(\eta \otimes \omega + \omega \otimes \eta)$. Similarly, we can take the antisymmetrized tensor product, also called the *wedge product*,

$$\eta \wedge \omega := \eta \otimes \omega - \omega \otimes \eta. \tag{8.5}$$

The space of skewsymmetric bilinear forms is denoted $\Lambda^2(V)$, and it has dimension $\binom{n}{2}$ (Exercise).

On a manifold X, we consider a family of bilinear forms $b_x : T_xX \times T_xX \to \mathbb{R}, x \in X$ to be smooth, if in all coordinate charts (of an atlas) $(U_\alpha, \varphi_\alpha)$ of X, we have

$$x \mapsto b_x(\partial_i|_x, \partial_j|_x), \qquad x \in U_\alpha,$$

being smooth, where ∂_i are the coordinate tangent vector fields. We will simply refer to a smooth family of bilinear forms as a "bilinear form on X". Notice

that the wedge product of two one-forms η, ω can be defined by the pointwise wedge product of Eq. (8.5), and this produces an antisymmetric bilinear form.

As in Exercise 8.4, we can also define a bilinear form on X as a $C^\infty(X)$-bilinear map

$$\mathfrak{X}(X) \times \mathfrak{X}(X) \to C^\infty(X).$$

Definition 23. A (smooth) positive-definite symmetric bilinear form g on X is called a *Riemannian metric*. An antisymmetric bilinear form ω on X is called a (differential) *2-form*. The space of 2-forms on X is denoted $\Omega^2(X)$.

Remark. If we think of T_xX as a linearized version of X near x, then a Riemannian metric provides the notion of lengths and angles on T_xX. A 2-form prescribes the "oriented area" of the parallelogram spanned by tangent vectors u, v (in that order). There may not be any relationship between a given Riemannian metric and a given 2-form.

Just as 1-forms can be pulled back under a map $f : Y \to X$, we can also pull back bilinear forms,

$$(f^*b)_y(u_y, v_y) := b_{f(y)}((df)_y(u_y), (df)_y(v_y)), \qquad u_y, v_y \in T_yY. \tag{8.6}$$

It is easy to see that Riemannian metrics get pulled back to Riemannian metrics, and 2-forms get pulled back to 2-forms. Furthermore, the wedge product is compatible with pullback,

$$f^*(\eta \wedge \omega) = f^*\eta \wedge f^*\omega. \tag{8.7}$$

Example 8.7. On $\mathbb{R}^N$ as a manifold, all its tangent spaces are canonically $\mathbb{R}^N$, and they can be given the standard inner product. This gives the standard Euclidean space as a Riemannian manifold. If X is a submanifold of $\mathbb{R}^N$, then the Euclidean space metric restricts to a Riemannian metric on X.

8.5 Differential k-forms and exterior derivative

Now, recall from Example 8.3 that d turns 0-forms (functions) into 1-forms. This is actually just one part of the *exterior calculus* on general differential forms. Let us see how this works on 1-forms.

Definition 24. Let $\omega \in \Omega^1(X)$ be a 1-form on X. Its *exterior derivative* $d\omega$ is the 2-form defined by

$$d\omega(u, v) := u(\omega(v)) - v(\omega(u)) - \omega([u, v]). \tag{8.8}$$

Exercise 8.8. Check that $d\omega$ as defined above is indeed a 2-form. (Exercise 8.1 would be helpful.)

Exercise 8.9. Check that the exterior derivative $d : \Omega^1(X) \to \Omega^2(X)$ satisfies: For $\lambda \in \mathbb{R}, \eta, \omega \in \Omega^1(X)$, $f \in C^\infty(X)$,

- $\mathbb{R}$-linearity: $d(\eta + \lambda\omega) = d\eta + \lambda d\omega$;
- Leibniz rule: $d(f \cdot \omega) = f \cdot d\omega + df \wedge \omega$;
- Chain complex: $d(df) = 0$;
- Coordinate formula: In a coordinate chart (U, φ), so that $\omega|_U = \sum_{i=1}^n \omega_i dx^i$ for some $\omega_i \in C^\infty(U)$, we have

$$d(\sum_{i=1}^n \omega_i \, dx^i) = \sum_{i=1}^n d\omega_i \wedge dx^i = \sum_{i,j=1}^n \frac{\partial \omega_i}{\partial x^j} dx^j \wedge dx^i. \tag{8.9}$$

- Naturality: If $h : Y \to X$ is smooth, then $h^*(d\omega) = d(h^*\omega)$.

One can go beyond bilinearity, and study *multilinear maps* $V \times \ldots \times V \to \mathbb{R}$ with k-arguments. Taking $V = T_x X$ and requiring smooth dependence on x, we have the $C^\infty(X)$-multilinear forms,

$$\mathfrak{X}(X) \times \ldots \times \mathfrak{X}(X) \to C^\infty(X).$$

The subspace of totally antisymmetric ones (minus sign acquired upon exchanging any pair of arguments) are called differential *k-forms*, and is denoted $\Omega^k(X)$. Intuitively, at each x, a k-form assigns an "oriented volume" to each parallelepiped spanned by k given tangent vectors.

The *wedge product* of a k-form η and an l-form ω may be defined as

$$\eta \wedge \omega(v_1, \ldots, v_{k+l}) = \frac{1}{k!\, l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma) \eta(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \omega(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}),$$

where $v_1, \ldots v_{k+l} \in \mathfrak{X}(X)$, and S_r is the symmetric group (i.e. permutations) on r objects. The pullback of k-forms is defined in the same way as Eq. (8.6), and it is a straightforward exercise to check that pullback commutes with wedge product, Eq. (8.7).

The exterior derivative extends to $\mathbb{R}$-linear maps $d : \Omega^k(X) \to \Omega^{k+1}(X)$, and it can be axiomatically characterized as the unique $\mathbb{R}$-linear maps satisfying the conditions

- df is the usual differential for $f \in \Omega^0(X) \equiv C^\infty(X)$;
- $d(\eta \wedge \omega) = d\eta \wedge \omega + (-1)^p(\eta \wedge d\omega)$ when η is a p-form;
- $d^2 = 0$.

In particular, Definition 24 of $d\omega$ for a 1-form ω coincides with this abstract characterization, as we verified in Exercise 8.9. For a general k form, the local coordinate formula for $d\eta$, analogous to Eq. (8.9), may be likewise derived.

Exercise 8.10. Show that the exterior derivative of a 2-form $\eta \in \Omega^2(X)$ can be expressed as

$$\begin{aligned} d\eta(u,v,w) = {} & u(\eta(v,w)) + v(\eta(w,u)) + w(\eta(u,v)) \\ & - \eta([u,v],w) - \eta([v,w]u) - \eta([w,u],v), \qquad u,v,w \in \mathfrak{X}(X). \end{aligned} \tag{8.10}$$

9 Integral curves and flows

9.1 Integral curves of vector fields

Every tangent vector at x has a geometric representative as a curve through x. For a tangent vector field, we might expect that there is a smooth family of curves on X, whose velocity vectors reproduce the given vector field.

Definition 25. Let v be a vector field on a manifold X. An *integral curve* of v is a smooth curve $\gamma : (a, b) \to X$ such that

$$\gamma'(t) = v_{\gamma(t)}, \qquad t \in (a, b).$$

Example 9.1. Consider $\mathbf{R}^2$ with (x^1, x^2) coordinates, and global vector field $v = \partial_2$. The integral curves are vertical lines. For example, $\gamma(t) = (x_0, y_0+t), t \in \mathbb{R}$ has, for each $f \in C^\infty(\mathbf{R}^2)$,

$$\gamma'(t)(f) = \frac{d(f \circ \gamma)}{ds}\Big|_{s=t} = \frac{\partial f}{\partial x^2}\Big|_{\gamma(t)} = \partial_2|_{\gamma(t)}(f) = v_{\gamma(t)}(f).$$

So the velocity vector $\gamma'(t)$ equals the vector field $v_{\gamma(t)}$ for all $t \in \mathbb{R}$.

Example 9.2. In standard global (x^1, x^2) coordinates for $\mathbf{R}^2$, take the vector field $v = x^1\partial_2 - x^2\partial_1$. It is convenient to work with (local) polar coordinates (r, θ) on $U = \mathbf{R}^2 \setminus \{(x, 0) : x \leq 0\}$. Check that on U, we have $v = \partial_\theta$ (Exercise). The curves

$$\gamma(t) = (r_0, \theta_0 + t), \qquad r_0 > 0,\ \theta_0 + t \in (-\pi, \pi)$$

have the property that $\gamma'(t) = \partial_\theta|_{(r_0,\theta_0+t)} = v_{\gamma(t)}$. Actually, these curves can smoothly extended to $t \in \mathbb{R}$ (write them in Cartesian coordinates).

Within a local coordinate chart, the problem of finding integral curves reduces to that of solving first-order ODEs. In a coordinate chart (U, φ), write $v = \sum_{i=1}^n v^i\partial_i$ with $v^i \equiv v^i(x)$ the component functions. A curve γ lying within U has the form

$$\gamma(t) = \varphi^{-1}(\gamma^i(t), \ldots, \gamma^n(t))$$

for functions $\gamma^i : (a, b) \to \mathbb{R}$. For $f \in C^\infty(X)$, we have

$$\gamma'(t)(f) = \frac{d(f \circ \gamma)}{ds}\Big|_{s=t} = \sum_{i=1}^n \frac{\partial(f \circ \varphi^{-1})}{\partial x^i}\Big|_{\gamma(t)} \frac{d\gamma^i}{ds}\Big|_{s=t} \equiv \sum_{i=1}^n \underbrace{\frac{d\gamma^i}{ds}\Big|_{s=t}}_{\dot\gamma^i(t)} \partial_i|_{\gamma(t)}(f).$$

So the velocity vector of γ at time t is

$$\gamma'(t) = \sum_{i=1}^{n} \dot{\gamma}^i(t)\partial_i|_{\gamma(t)} \overset{\text{need}}{=} \sum_{i=1}^{n} v^i(\gamma(t))\partial_i|_{\gamma(t)}.$$

Comparing the coefficients, we have the ODEs

$$\dot{\gamma}^i(t) = v^i(\varphi^{-1}(\gamma^1(t), \ldots, \gamma^n(t))), \qquad i = 1, \ldots, n, \tag{9.1}$$

for the coordinate functions γ^i of the curve. Requiring the curve to pass through a prescribed point $x_0 \in U$ at time t_0, is the same thing as setting the initial conditions

$$(\gamma^1(t_0), \ldots, \gamma^n(t_0)) = \varphi(x_0). \tag{9.2}$$

ODE theory $\Rightarrow$ local existence. There always exists some time interval containing t_0, during which the initial value problem (9.1)-(9.2) has a (unique, smooth) solution. See Theorem D.1 of [11] for details of the ODE theory justifying this fact.

Proposition 9.1. *Let v be a vector field on a manifold X. For each point $x \in X$, there exists $\epsilon > 0$ and an integral curve $\gamma : (-\epsilon, \epsilon) \to X$ starting from x, i.e., $\gamma(0) = x$.*

Generally, we may not be able to extend an integral curve to one which is defined for all $t \in \mathbb{R}$.

Exercise 9.1. Find an integral curve of the vector field $((x^1)^2 + (x^2)^2)\partial_1$ on $\mathbf{R}^2$, passing through the point $(1, 0)$. Check that this integral curve cannot be extended to all of $\mathbb{R}$.

If γ is an integral curve of a vector field v, we have (basically by definition),

$$v_{\gamma(t)}(f) = \lim_{t' \to 0} \frac{f(\gamma(t + t')) - f(\gamma(t))}{t'}.$$

So a vector field v differentiates functions along its integral curves.

9.2 Flows

Definition 26. A *global flow* on a manifold X is a smooth $\mathbb{R}$-action on X, i.e., a smooth map

$$\theta : \mathbb{R} \times X \to X$$

satisfying

$$\theta(t, (\theta(s, x)) = \theta(t + s, x), \qquad \theta(0, x) = x, \qquad x \in X,\ s, t \in \mathbb{R}.$$

For each $t \in \mathbb{R}$, a flow determines the smooth map

$$\theta_t : X \to X, \qquad x \mapsto \theta(t, x).$$

These θ_t satisfy a group law when composed,

$$\theta_0 = \mathrm{id}_X, \qquad \theta_t \circ \theta_s = \theta_{s+t}, \qquad s, t \in \mathbb{R},$$

so they are diffeomorphisms.

Starting from an initial point $x \in X$, a flow provides a smooth curve

$$\gamma^{(x)} : \mathbb{R} \to X, \qquad t \mapsto \theta_t(x), \tag{9.3}$$

and therefore a tangent vector at x,

$$v_x = (\gamma^{(x)})'(0) \ \in T_x X. \tag{9.4}$$

Definition 27. The tangent vector assignment, Eq. (9.4), associated to a flow on X is called the *infinitesimal generator* of the flow.

Proposition 9.2. *Let θ be a global flow on X. Its infinitesimal generator is a smooth vector field, and the $\gamma^{(x)}$ of Eq.* (9.4) *are its integral curves.*

Proof. We may check that v is a derivation of $C^\infty(X)$ (exercise), so it is a smooth vector field (Remark 2).

Next, consider the curve $\gamma^{(x)}$ starting at a given $x \in X$. We need to check the integral curve condition, $v_{\gamma^{(x)}(t)} = (\gamma^{(x)})'(t)$, $t \in \mathbb{R}$.

Fix $t \in \mathbb{R}$ and write $x' := \gamma^{(x)}(t) = \theta_t(x)$. We have

$$\gamma^{(x')}(s) \equiv \theta_s(x') = \theta_s \circ \theta_t(x) = \theta_{s+t}(x) = \gamma^{(x)}(s + t).$$

So for any $f \in C^\infty(X)$,

$$\begin{aligned} v_{\gamma^{(x)}(t)}(f) \equiv v_{x'}(f) \equiv (\gamma^{(x')})'(0)(f) &= \frac{d(f \circ \gamma^{(x')}(s))}{ds}\Big|_{s=0} \\ &= \frac{d(f \circ \gamma^{(x)}(s + t))}{d(s + t)}\Big|_{s+t=t} \\ &= (\gamma^{(x)})'(t)(f), \end{aligned}$$

as required. □

On $\mathbf{R}^2$, the flows given by vertical translation by t and anticlockwise rotation by an angle t, respectively give rise to the vector fields of Examples 9.1, 9.2. However, the example of Exercise 9.1 shows that a general vector field might not be associated to any globally defined flow.

So in general, we have to consider *local flows* θ defined only on some open subset $\mathcal{D} \subset \mathbb{R} \times X$ called the *flow domain.* At time t, the flow map θ_t is only defined on a space-time subset

$$X_t := \{x \in X \,:\, (t, x) \in \mathcal{D}\}.$$

So from an initial point x, the curve $\theta(\cdot, x)$ only makes sense for some open time interval $\mathcal{D}^{(x)}$ containing 0. This is enough to make sense of the tangent vectors v_x in Eq. (9.4), and the notion of infinitesimal generator.

Theorem 9.3. *Let v be a vector field on X. There is a unique maximal flow $\theta : \mathcal{D} \to X$ with infinitesimal generator being v. The curve $\gamma^{(x)} : \mathcal{D}^{(x)} \to X$ is the unique maximal integral curve of v starting at x. For each $t \in \mathbb{R}$, the maps*

$$\theta_t : X_t \to X_{-t}$$

are well-defined diffeomorphisms (of possibly empty open subsets of X), with $\theta_t^{-1} = \theta_{-t}$. This (maximal) flow is called the flow generated by v.

A full proof may be found in [11], Theorem 9.12. On some subset of X, the θ_t satisfy $\theta_t \circ \theta_s = \theta_{s+t}$. So a vector field is sometimes said to generate a "local 1-parameter group of local diffeomorphisms".

Definition 28. A vector field on a manifold is *complete* if it generates a global flow; equivalently, all its (maximal) integral curves are defined for all $t \in \mathbb{R}$.

Here is a sufficient criterion for v to be complete.

Lemma 9.4 (Uniform time Lemma)**.** *Let v be a vector field on a manifold X, and θ be its flow. Suppose there exists $\delta > 0$ such that for all x, the curves $\gamma^{(x)}$ are defined at least for $t \in (-\delta, \delta)$. Then v is complete.*

Proof. Suppose, for a contradiction, that v has some integral curve $\gamma^{(x)}$ which can only be defined for $t \in (-\delta, t_{\max})$ with $\delta \leq t_{\max} < \infty$. Define $x' =$

$\gamma^{(x)}(t_{\max} - \frac{\delta}{2})$. By hypothesis, the integral curve $\gamma^{(x')}$ starting at x' is at least defined for $t \in (-\delta, \delta)$. We try extending $\gamma^{(x)}$ as follows,

$$\gamma^{(x)}(t) = \begin{cases} \gamma^{(x)}(t), & t \in (-\delta, t_{\max}), \\ \gamma^{(x')}(t - t_{\max} + \frac{\delta}{2}), & t \in (t_{\max} - \frac{3\delta}{2}, t_{\max} + \frac{\delta}{2}). \end{cases}$$

This is well-defined on the overlapping time range, due to

$$\begin{aligned} \gamma^{(x')}(t - t_{\max} + \frac{\delta}{2}) = \theta_{t - t_{\max} + \frac{\delta}{2}}(x') &= \theta_{t - t_{\max} + \frac{\delta}{2}} \circ \theta_{t_{\max} - \frac{\delta}{2}}(x) \\ &= \theta_t(x) = \gamma^{(x)}(t). \end{aligned}$$

The extended $\gamma^{(x)}$ remains an integral curve starting at x, but it has now been defined until $t_{\max} + \frac{\delta}{2} > t_{\max}$. □

9.3 Lie derivative

Let v, w be vector fields, and θ be the (local) flow generated by v. It might appear that we can simply define the "derivative of w along v", as

$$L_v(w)(x) := \lim_{t \to 0} \frac{(\theta_{-t})_*(w_{\theta_t(x)}) - w_x}{t}. \tag{9.5}$$

In other words, we use the flow of v to transport the tangent vectors $w_{\theta_t(x)}$ back to the basepoint x. After this transportation, the tangent vectors belong to the same tangent space, so the right side of Eq. (9.5) makes sense.

In fact, one can compute (exercise) that this "derivative" is given by taking the commutator with v,

$$[v, w]_x = \lim_{t \to 0} \frac{(\theta_{-t})_*(w_{\theta_t(x)}) - w_x}{t}.$$

Definition 29. Let $v, w \in \mathfrak{X}(X)$. The *Lie derivative* of w along v is the vector field

$$L_v(w) := [v, w].$$

However, suppose we only have a curve with velocity vector v_x, and we wanted to define the rate of change of the vector field w along this curve. The Lie derivative cannot do this for us, since a *vector field* is required to generate a local flow for the transport of tangent vectors.

10 Geometry of Lie group actions

10.1 Lie algebra of left-invariant vector fields

Definition 30. A vector field v on a Lie group G is *left-invariant* if

$$(L_g)_* v = v, \qquad \forall g \in G.$$

Left-invariance is preserved by taking linear combinations, and even more is true:

Proposition 10.1. *The commutator of two left-invariant vector fields on a Lie group is left-invariant.*

Proof. For left-invariant vector fields v, w on a Lie group G, we have $(L_g)_* v = v$ and $(L_g)_* w = w$ for all $g \in G$. So using Exercise 8.2,

$$(L_g)_*[v, w] = [(L_g)_* v, (L_g)_* w] = [v, w].$$

□

Recall from Eq. (8.1) that $\mathfrak{X}(X)$ is a Lie algebra under the Lie bracket $[\cdot, \cdot]$. Prop. 10.1 says that the left-invariant vector fields on X form a Lie subalgebra of $\mathfrak{X}(X)$.

Definition 31. The *Lie algebra* of a Lie group G, denoted $\mathfrak{g}$, is the vector space of left-invariant vector fields on G, with Lie bracket being the commutator of vector fields.

Theorem 10.2. *Let $\mathfrak{g}$ be the Lie algebra of a Lie group G. Evaluation at the identity element $e \in G$,*

$$\varepsilon : \mathfrak{g} \to T_e G, \qquad v \mapsto v_e,$$

is a linear isomorphism.

Proof. The linearity and injectivity of ε is straightforward. For surjectivity, pick $\xi \in T_e G$, and look for a left-invariant vector field v such that $\varepsilon(v) \equiv v_e = \xi$. Left-invariance means that the only possible candidate has

$$v_g = ((L_g)_* v)_g = (dL_g)_e(v_e) = (dL_g)_e(\xi), \qquad g \in G.$$

From the group property of G, it follows that this tangent vector assignment

$$v : g \mapsto (dL_g)_e(\xi) \tag{10.1}$$

is left-invariant, and has $v_e = (dL_e)_e(\xi) = d(\mathrm{id}_G)_e(\xi) = \xi$.

It remains to check that the v constructed above is a smooth tangent vector field. By Exercise 5.1, we need to check that for any $f \in C^\infty(G)$, the function

$$\begin{aligned} v(f) : G &\to \mathbb{R} \\ g &\mapsto v_g(f) \end{aligned}$$

is smooth. To this end, let $\gamma : (-\delta, \delta) \to G$ be a curve in G with $\gamma(0) = e$ and $\gamma'(0) = \xi$. Then we compute

$$v(f)(g) \equiv v_g(f) = ((dL_g)_e(\xi))\,(f) = \xi\,(f \circ L_g) = \frac{d(f \circ L_g \circ \gamma)}{ds}\Big|_{s=0}. \tag{10.2}$$

We regard $f \circ L_g \circ \gamma$ as a function

$$\eta : (-\delta, \delta) \times G \to \mathbb{R}, \qquad (s, g) \mapsto f \circ L_g \circ \gamma(s),$$

which is smooth. Then the right-hand-side of Eq. (10.2) equals the partial derivative $\frac{\partial \eta}{\partial s}\Big|_{(s,g)=(0,g)}$, which is therefore smoothly dependent on $g \in G$. This shows that $v(f) \in C^\infty(G)$ as required. □

Thus there is only a finite-dimensional vector space of left-invariant vector fields on G, with, $\dim \mathfrak{g} = \dim T_eG = \dim G$. Notice that T_eG inherits a Lie bracket operation from its identification with $\mathfrak{g}$.

Corollary 10.3. *A Lie group is* parallelizable*, i.e. its tangent bundle is trivializable.*

Proof. Pick any basis $\{\xi_1, \ldots, \xi_n\}$ for T_eG. Then we obtain left-invariant vector fields $v^{(\xi_i)} \in \mathfrak{g}, i = 1, \ldots, \dim G$ by Theorem 10.2, Eq. (10.1), and they are linearly independent at every $g \in G$. So we obtain a global frame for TG, which determines a global trivialization by Exercise 5.3. □

Example 10.1. $\mathrm{GL}(n, \mathbb{R})$ is an open submanifold of the vector space $\mathrm{M}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$. So $T_e(\mathrm{GL}(n))$ is naturally isomorphic to $\mathrm{M}_n(\mathbb{R})$ as a vector space. On the one hand, the matrix commutator provides a Lie bracket for $T_e(\mathrm{GL}(n))$.

On the other hand, Theorem 10.2 tells that $T_e(\mathrm{GL}(n))$ is canonically identified with the left-invariant vector fields on $\mathrm{GL}(n,\mathbb{R})$, so we get another Lie bracket on $T_e(\mathrm{GL}(n))$ using the commutator of vector fields. These two natural Lie brackets on $T_e(\mathrm{GL}(n))$ coincide (Exercise).

Thus it is customary to write $\mathfrak{gl}(n,\mathbb{R}) = \mathrm{M}_n(\mathbb{R})$, and $\mathfrak{gl}(n,\mathbb{C}) = \mathrm{M}_n(\mathbb{C})$ for these matrix Lie algebras. If you prefer, we can also work with $\mathrm{GL}(V)$ for a general vector space V without a preferred basis. Then $\mathfrak{gl}(V) = \mathrm{End}(V)$, the linear maps on V, with commutator as Lie bracket.

Definition 32. A *Lie group homomorphism* $f : G \to H$ is a smooth group homomorphism, while a *Lie group homomorphism* $\mathfrak{g} \to \mathfrak{h}$ is a bracket-preserving linear map.

Exercise 10.1. Let $f : G \to H$ be a Lie group homomorphism. Regard $(df)_e : T_eG \to T_eH$ as a linear map $\mathfrak{g} \to \mathfrak{h}$ through the identification of Theorem 10.2. Show that $(df)_e$ is a Lie algebra homomorphism.

The above induced homomorphism is usually denoted $f_* : \mathfrak{g} \to \mathfrak{h}$.

Exercise 10.2. Show that if G is an abelian Lie group, then its Lie algebra $\mathfrak{g}$ of left-invariant vector fields has trivial Lie bracket, $[u, v] = 0$ for all $u, v \in \mathfrak{g}$.

10.2 Matrix Lie group and algebra examples

Example 10.2. The matrix Lie subgroups $\mathrm{O}(n), \mathrm{SO}(n), \mathrm{U}(n), \mathrm{SU}(n)$ have their corresponding Lie algebras being Lie subalgebras of $\mathfrak{gl}(n,\mathbb{R})$ or $\mathfrak{gl}(n,\mathbb{C})$, i.e., certain linear subspaces of matrices equipped with the matrix commutator.

Example 10.3. As an exercise, we had calculated the tangent space of $\mathrm{O}(n)$ at the identity element, thus also that of $\mathrm{SO}(n)$, as the space of real antisymmetric matrices, and this is identified with $\mathfrak{o}(n) \cong \mathfrak{so}(n)$ via Theorem 10.2. A similar calculation shows that $(\mathfrak{s})\mathfrak{u}(n)$ is the space of (traceless) antihermitian complex $n \times n$ matrices.

Remark. Non-isomorphic Lie groups can have isomorphic Lie algebras. For example, the Lie algebra of $\mathrm{GL}(n,\mathbb{R})^+$ (the positive determinant subgroup) is the same as that of $\mathrm{GL}(n,\mathbb{R})$; similarly, $\mathrm{SO}(n)$ and $\mathrm{O}(n)$ have the same Lie

algebra. A more interesting example is $\mathfrak{su}(2) \cong \mathfrak{so}(3)$. Take

$$\begin{aligned}
\mathfrak{su}(2) &= \operatorname{span}_{\mathbb{R}} \left\{ \frac{1}{2i}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \frac{1}{2i}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \frac{1}{2i}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\} \\
\mathfrak{so}(3) &= \operatorname{span}_{\mathbb{R}} \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}. \qquad (10.3)
\end{aligned}$$

A direct calculation shows that the linear map taking the above basis vectors to each other is a Lie algebra isomorphism (Exercise).

Remark. Even if G is known to be a matrix Lie group, $G \subset \mathrm{GL}(n)$, it is often represented on some other vector space V other than $\mathbb{C}^n$. For example, there is a distinguished representation of G on the vector space $\mathfrak{g}$ (see Section 10.3). If G is represented as a subgroup of $\mathrm{GL}(V)$, then $\mathfrak{g}$ will be represented as a Lie subalgebra of $\mathrm{End}(V)$.

One should always keep in mind the geometric meaning of $\mathfrak{g}$ as left-invariant vector fields on G. The description as the tangent space at the identity matrix, is largely for convenience of calculations.

10.3 Adjoint representation

Conjugation of G by any element $g \in G$,

$$C_g : G \to G, \qquad g' \mapsto g g' g^{-1},$$

is a group automorphism of G. In fact, C_g is a diffeomorphism, thus a *Lie group automorphism.* Therefore, there is an induced automorphism of $\mathfrak{g} \cong T_e G$ (Exercise 10.1),

$$\mathrm{Ad}_g := (C_g)_* = (dC_g)_e : \mathfrak{g} \to \mathfrak{g}.$$

So $\mathrm{Ad}_g \in \mathrm{GL}(\mathfrak{g})$ for each $g \in G$. Since

$$\mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2} = \mathrm{Ad}_{g_1 g_2},$$

the map

$$\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g}), \qquad g \mapsto \mathrm{Ad}_g,$$

is itself a Lie group homomorphism. (Verification of smoothness of Ad is omitted.) This homomorphism is called the *adjoint representation* of G on its Lie algebra $\mathfrak{g}$.

Example 10.4. Let $G = \mathrm{GL}(n, \mathbb{R})$, which has $\mathfrak{gl}(n, \mathbb{R}) \cong T_e\mathrm{GL}(n, \mathbb{R}) = \mathrm{M}_n(\mathbb{R})$. The conjugation operation C_g is linear in the matrix entries (which are serving as the coordinates). So the adjoint representation $\mathrm{Ad} : \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(\mathrm{M}_n(\mathbb{R}))$ is given by

$$\mathrm{Ad}_g \equiv (dC_g)_e : A \mapsto gAg^{-1}, \qquad A \in \mathrm{M}_n(\mathbb{R}). \tag{10.4}$$

As for matrix Lie groups $G \subset \mathrm{GL}(n, \mathbb{K})$, their Lie algebras are realized as (real) linear subspaces inside $T_e\mathrm{GL}(n, \mathbb{K}) = \mathrm{M}_n(\mathbb{K})$. Restrict the adjoint representation of $\mathrm{GL}(n)$, Eq. (10.4), to the subgroup G, and let it act only on the corresponding tangent subspace $\mathfrak{g} \subset \mathrm{M}_n(\mathbb{K})$. This gives the adjoint representation of G on $\mathfrak{g}$.

Exercise 10.3. A basis for $\mathfrak{so}(3)$ was given in Eq. (10.3). In this basis, the adjoint representation of $g \in \mathrm{SO}(3)$ is given by the matrix g itself.

10.4 Exponential map $\mathfrak{g} \to G$

Proposition 10.4. *Left-invariant vector fields on Lie groups are complete.*

Proof. Let $\theta : \mathcal{D} \to G$ be the flow of a left-invariant vector field v on the Lie group G. At the identity element $e \in G$, the integral curve $t \mapsto \gamma^{(e)}(t) = \theta(t, e)$ is defined for some time interval $t \in (-\delta, \delta)$. By left-invariance, for any $g \in G$, the curve $L_g \circ \gamma^{(e)}$ is the integral curve $\gamma^{(g)}$ starting at g, and it is defined for $t \in (-\delta, \delta)$. By the uniform time Lemma 9.4, v is complete. □

The completeness of left-invariant v on G has a very important consequence. Let γ_v be the integral curve of v which starts from e at time $t = 0$; this curve is defined for all times $t \in \mathbb{R}$. Pick some other point $\gamma_v(s), s \in \mathbb{R}$ lying on this curve. We have two ways of obtaining an integral curve of v starting form $\gamma_v(s)$:

- $t \mapsto (L_{\gamma_v(s)} \circ \gamma_v)(t) = \gamma_v(s) \cdot \gamma_v(t)$;
- $t \mapsto \gamma_v(s + t)$.

These two curves must coincide, so $\gamma_v(s) \cdot \gamma_v(t) = \gamma(s + t)$ for all $s, t \in \mathbb{R}$. Thus the integral curve $\gamma_v : \mathbb{R} \to G$ is necessarily a *Lie group homomorphism.*

Definition 33. A Lie group homomorphism $\gamma : \mathbb{R} \to G$ is called a *one-parameter subgroup of* G. In particular, for a left-invariant vector field $v \in \mathfrak{g}$, the integral curve γ_v starting at e is called the *one-parameter subgroup of* G *generated by* v.

Theorem 10.5. *One-parameter subgroups of a Lie group* G *are in bijection with left-invariant vector fields on* G*, via taking the maximal integral curve starting at* e*.*

Proof. We have just seen how a left-invariant vector field generates a one-parameter subgroup of G.

In reverse, let $\gamma : \mathbb{R} \to G$ be an arbitrary one-parameter subgroup of G. Note that $\gamma(0) = e$. Let us investigate the tangent vectors to the curve γ, which are

$$\gamma'(s) = (d\gamma)_s \left(\frac{d}{dt}\Big|_{t=s} \right) \in T_{\gamma(s)}G, \qquad s \in \mathbb{R}.$$

Write L_s for the translation on $\mathbb{R}$ by s. The homomorphism property $\gamma(s+t) = \gamma(s)\gamma(t)$ can be rewritten as $\gamma \circ L_s = L_{\gamma(s)} \circ \gamma$, and this implies

$$d\gamma \circ dL_s = dL_{\gamma(s)} \circ d\gamma, \qquad s \in \mathbb{R}.$$

Then

$$\begin{aligned} \gamma'(s) = (d\gamma)_s \left(\frac{d}{dt}\Big|_{t=s} \right) &= d\gamma_s \left((dL_s)_0 \left(\frac{d}{dt}\Big|_{t=0} \right) \right) \\ &= (dL_{\gamma(s)})_e \underbrace{\left(d\gamma_0 \left(\frac{d}{dt}\Big|_{t=0} \right) \right)}_{\equiv \gamma'(0) \in T_eG}. \end{aligned} \tag{10.5}$$

By Theorem 10.2, the tangent vector $\gamma'(0) \in T_eG$ corresponds to a left-invariant vector field v, whose tangent vectors along the curve are, in particular, given by

$$v_{\gamma(s)} = (dL_{\gamma(s)})_e(\gamma'(0)) \overset{\text{Eq. (10.5)}}{=} \gamma'(s), \qquad s \in \mathbb{R}.$$

This says that γ is recovered as the (maximal) integral curve of the left-invariant vector field v, starting at e. $\square$

Definition 34. For a Lie group G with Lie algebra $\mathfrak{g}$, the *exponential map* is defined as

$$\exp : \mathfrak{g} \to G, \qquad v \mapsto \gamma_v(1),$$

where $\gamma_v : \mathbb{R} \to G$ is the 1-parameter subgroup of G generated by $v \in \mathfrak{g}$.

Both v and sv generate the same subgroup of G, but as parametrized subgroups, the latter has time parameter t rescaled to st. Then it follows that

$$\exp(sv) = \gamma_{sv}(1) = \gamma_v(s) \quad \Rightarrow \quad \exp(\,\cdot\, v)'(0) = \gamma_v'(0) = v,$$

and the expected symbolic properties of exp hold,

$$\exp((s+t)v) = \exp(sv)\exp(tv), \qquad \exp(v)^{-1} = \exp(-v), \qquad v \in \mathfrak{g},\ s,t \in \mathbb{R}.$$

Example 10.5. Recall that the Lie group $\mathrm{GL}(n,\mathbb{R})$ has $\mathfrak{gl}(n,\mathbb{R}) \cong \mathrm{M}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$. The matrix elements X^{ij} play the role of coordinates, so there are coordinate vector fields $\frac{\partial}{\partial X^{ij}}$. An element $A \in \mathrm{M}_n(\mathbb{R})$ is regarded as the tangent vector

$$A \leftrightarrow \sum_{i,j=1}^{n} A^{ij} \frac{\partial}{\partial X^{ij}}\bigg|_e \in \mathfrak{gl}(n,\mathbb{R}) = T_e(\mathrm{GL}(n,\mathbb{R})),$$

and it also corresponds to the left-invariant vector field

$$v^{(A)} : g \mapsto (dL_g)_e(A) = (gA)^{ij} \frac{\partial}{\partial X^{ij}}\bigg|_g, \qquad g \in \mathrm{GL}(n,\mathbb{R}).$$

Here we used the fact that L_g (matrix multiplication by g) is linear in the coordinates X^{ij}, so dL_g is also represented as matrix-multiplication-by-g.

Theorem 10.5 says that the one-parameter subgroup $\gamma : \mathbb{R} \to \mathrm{GL}(n,\mathbb{R})$ generated by A is the integral curve for $v^{(A)}$ starting at the identity matrix. Explicitly,

$$\gamma(0) = e = 1_n, \qquad \gamma'(s) = v^{(A)}_{\gamma(s)} = (\gamma(s)A)^{ij} \frac{\partial}{\partial X^{ij}}\bigg|_{\gamma(s)}.$$

On the right-side, the integral curve condition is exhibited as a collection of ODEs for the matrix elements of $\gamma(s)$, subject to the initial condition on the left side. These ODEs are concisely written as the matrix equations

$$\gamma(0) = 1_n, \qquad \gamma'(s) = \gamma(s)A.$$

The solution is provided by the matrix exponential,

$$\gamma(s) = \sum_{k=0}^{\infty} \frac{1}{k!}(sA)^k = 1_n + sA + \frac{1}{2}(sA)^2 + \dots,$$

where the series on the right converges to an element of $\mathrm{GL}(n, \mathbb{R})$, and depends smoothly on s (Exercise). Similarly for the complex case. Setting $s = 1$, we get the *matrix exponential map* $\exp : \mathfrak{gl}(n) \to \mathrm{GL}(n)$.

Proposition 10.6. *The flow $\theta^{(v)}$ on G generated by $v \in \mathfrak{g}$ is implemented by right-multiplication by* $\exp(tv)$*,*

$$\theta^{(v)}(t, g) = g \cdot \exp(tv) \equiv R_{\exp(tv)}(g). \tag{10.6}$$

Proof. Since $v \in \mathfrak{g}$ is complete (Prop. 10.4), it generates a global flow $\theta^{(v)}$ on G. By definition, at each $g \in G$, the curve $t \mapsto \theta^{(v)}(t, g)$ has velocity vector being $v_g = dL_g(v_e)$. Since the curve $t \mapsto (L_g \circ \gamma_v)(t)$ has

$$(L_g \circ \gamma_v)'(0) = (dL_g)_e((\gamma_v)'(0)) = dL_g(v_e),$$

we learn that

$$\theta^{(v)}(t, g) = L_g(\gamma_v(t)) = L_g(\exp(tv)) = g \cdot \exp(tv).$$

□

Example 10.6. On the Lie group $\mathbb{R}$, a left invariant vector field is ∂_x, and it generates the global flow $\theta(t, x) = x + t$. In this case, $R_{\exp(t\partial_x)}$ is translation-by-t.

Exercise 10.4. Let $f : G \to H$ be a Lie group homomorphism, and $f_* : \mathfrak{g} \to \mathfrak{h}$ be the induced Lie group homomorphism. Show that the exponential map is natural, in the sense that the following diagram commutes:

$$\begin{tikzcd} \mathfrak{g} \arrow[r, "f_*"] \arrow[d, "\exp"'] & \mathfrak{h} \arrow[d, "\exp"] \\ G \arrow[r, "f"] & H \end{tikzcd}.$$

Proposition 10.7. *Let $f : G \to H$ be a Lie group homomorphism. Then adjoint representations of G and H are compatible with f in the sense that*

$$f_* \circ \mathrm{Ad}_g = \mathrm{Ad}_{f(g)} \circ f_*, \qquad g \in G.$$

Proof. Let $g \in G$ and $v \in T_eG \cong \mathfrak{g}$. From Exercise 10.4, we have

$$f(g\,\exp(tv)\,g^{-1}) = f(g)f(\exp(tv))f(g)^{-1} = f(g)\exp(tf_*(v))f(g)^{-1},$$

equivalently,

$$f \circ C_g \circ \exp(tv) = C_{f(g)} \circ \exp(tf_*(v)).$$

Differentiate at $t = 0$ to obtain

$$f_* \circ \mathrm{Ad}_g(v) = \mathrm{Ad}_{f(g)} \circ f_*(v).$$

□

Example 10.7. The "group-of-scales" considered by H. Weyl is the matrix Lie group $G = \mathbb{R}_{>0} = ((0,\infty),\times) = \mathrm{GL}(1,\mathbb{R})^+$. The Lie algebra is $\mathfrak{g} = \mathrm{M}_1(\mathbb{R})$ with trivial commutator. The exponential map is the usual one, $t \mapsto e^t$, and this is a diffeomorphism.

Note that $G = (\mathbb{R}_{>0},\times)$ is isomorphic to the additive Lie group $(\mathbb{R},+)$ (by the exponential, again). We had exhibited $(\mathbb{R},+)$ as the matrix subgroup $\left\{\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix} : x \in \mathbb{R}\right\}$ in Section 3.1, and its Lie algebra sits inside $\mathfrak{gl}(2,\mathbb{R}) = \mathrm{M}_2(\mathbb{R})$ as $\left\{\begin{pmatrix}0 & s\\ 0 & 0\end{pmatrix} : s \in \mathbb{R}\right\}$. In this embedding, the exponential map is the matrix exponential. We have basically provided two different representations of the same abstract Lie group (and Lie algebra).

Now consider the matrix Lie group $H = \mathrm{U}(1) \subset \mathrm{GL}(1,\mathbb{C})$, which has Lie algebra $\mathfrak{h} = i\mathbb{R} \subset \mathrm{M}_1(\mathbb{C})$ with trivial commutator (Exercise). Clearly $\mathfrak{g} \cong \mathfrak{h}$ as Lie algebras. However, G and H are non-isomorphic, and even non-homeomorphic. So the exponential map depends on which Lie group one is referring to.

10.5 Fundamental vector fields on G-manifolds

Let G act smoothly on another manifold P on the right (by diffeomorphisms), denoted by the smooth map

$$\sigma : P \times G \to P, \qquad (p,g) \mapsto p \cdot g.$$

We call P a *G-space*.

Fix $v \in \mathfrak{g}$, and restrict attention to the action of its 1-parameter subgroup $\gamma_v : \mathbb{R} \to G$,

$$\begin{aligned} \sigma : P \times \mathbb{R} &\to P \\ (p, t) &\mapsto p \cdot \underbrace{\exp(tv)}_{\gamma_v(t)}. \end{aligned}$$

As with the $P = G$ case studied earlier, this right-multiplication action describes a global flow on P. Its infinitesimal generator (see Eq. (9.4)) is some vector field on P, which we call $v^\sharp$. Explicitly,

$$v^\sharp_p = \underbrace{(p \cdot \exp(\,\cdot\, v))'(0)}_{=\text{“}\frac{d}{dt}(p\cdot\exp(tv))\big|_{t=0}\text{”}} \in T_pP. \tag{10.7}$$

Notation: For a curve $\gamma(\cdot)$ in P, we shall often denote its velocity vector $\gamma'(0) \in T_{\gamma(0)}P$ at $t = 0$ by the suggestive expression $\frac{d\gamma(t)}{dt}\big|_{t=0}$.

Definition 35. Let P be a G-space. The linear map

$$\begin{aligned} \mathfrak{g} &\to \mathfrak{X}(P) \\ v &\mapsto v^\sharp, \end{aligned} \tag{10.8}$$

defined by Eq. (10.7), is called the *infinitesimal generator* of the G-action on P. Vector fields on P induced in this way are called *fundamental vector fields.*

Example 10.8. Let $P = G$, with G acting on itself by right multiplication. If $v \in \mathfrak{g}$ is a left-invariant vector field, then at any $g' \in G$,

$$v^\sharp_{g'} = \frac{d}{dt}(g' \cdot \exp(tv))\Big|_{t=0} = (L_{g'})_* \left(\frac{d}{dt}(\exp(tv))\Big|_{t=0} \right) = (L_{g'})_*(v_e) = v_{g'}.$$

So in this case, $v^\sharp = v$ itself.

It is instructive to fix $p \in P$, and define the *orbit map* based at p,

$$\mathcal{O}_p : G \to P, \qquad g \mapsto p \cdot g.$$

We can then rewrite $v^\sharp_p$ as

$$v^\sharp_p = (\mathcal{O}_p \circ \gamma_v)'(0) = (d\mathcal{O}_p)_e((\gamma_v)'(0)) \equiv (d\mathcal{O}_p)(v_e). \tag{10.9}$$

Clearly $\mathcal{O}_{p\cdot g} = \mathcal{O}_p \circ L_g$. Then

$$v^\sharp_{p\cdot g} = (d\mathcal{O}_{p\cdot g})_e(v_e) = (d(\mathcal{O}_p \circ L_g))_e(v_e) = (d\mathcal{O}_p)_g((dL_g)_e(v_e)) = (d\mathcal{O}_p)_g(v_g),$$

where the last equality follows from left-invariance of v. Thus for all $g \in G$, $d\mathcal{O}_p$ "pushes forward" v_g to $v^\sharp_{p\cdot g}$. One says that v and $v^\sharp$ are $\mathcal{O}_p$-related. By the naturality of Lie brackets (see Prop. 8.30 of [11] for details), we also have $[v, w]$ being $\mathcal{O}_p$-related to $[v^\sharp, w^\sharp]$. This means, in particular, that

$$[v^\sharp, w^\sharp]_p = (d\mathcal{O}_p)_e([v, w]_e) = ([v, w]^\sharp)_p.$$

We obtain this equality at every other $p' \in P$, by applying the orbit map $\mathcal{O}_{p'}$. So $[v, w]^\sharp = [v^\sharp, w^\sharp]$. We arrive at the following result:

Proposition 10.8. *The infinitesimal generator of a group action of G on an manifold P, Eq.* (10.8)*, is a Lie algebra homomorphism.*

Remark. Think of G as being represented on P as a smooth family of diffeomorphisms. Infinitesimally, $\mathfrak{g}$ is represented as "infinitesimal diffeomorphisms" of P, namely, the fundamental vector fields on P.

Example 10.9. Let $S^2 \subset \mathbb{R}^3$ be the unit sphere centered at the origin, so its points are labelled by a unit vector (n_x, n_y, n_z) thought of as a row vector. The group SO(3) acts on row vectors by right-multiplication, and this restricts to an action on S^2. Consider the $\mathfrak{so}(3)$ element $J_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. A calculation of the matrix exponential shows that J_z generates the 1-parameter subgroup

$$\exp(tJ_z) = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and this subgroup effects

$$(n_x, n_y, n_z) \cdot \exp(tJ_z) = (n_x \cos t + n_y \sin t, -n_x \sin t + n_y \cos t, n_z).$$

This is a clockwise rotation by angle t in the horizontal plane. The fundamental vector field $(J_z)^\sharp$ generates flow along the constant-latitude orbits, and it vanishes at the north and south poles.

Write R_g for the diffeomorphism $p \mapsto p \cdot g$. The following Proposition expresses the sense in which fundamental vector fields are G-equivariant.

Proposition 10.9. *Let $v^\sharp$ be the fundamental vector field on P induced by $v \in \mathfrak{g}$. Then*

$$(R_g)_*(v^\sharp) = (\mathrm{Ad}_{g^{-1}}(v))^\sharp, \qquad \forall g \in G.$$

Proof. We compute, at any $p \in P$,

$$\begin{aligned}
((R_g)_*(v^\sharp))_p &= (d(R_g))_{p\cdot g^{-1}}((v^\sharp)_{p\cdot g^{-1}}) \\
&= (d(R_g))_{p\cdot g^{-1}} \frac{d}{dt}\Big|_{t=0} (p \cdot g^{-1} \exp(tv)) \\
&= \frac{d}{dt}\Big|_{t=0} (p \cdot g^{-1} \exp(tv) g) \\
&= \frac{d}{dt}\Big|_{t=0} (p \cdot C_{g^{-1}}(\exp(tv))) \\
&\overset{\text{Ex. 10.4}}{=} \frac{d}{dt}\Big|_{t=0} (p \cdot \exp(t\mathrm{Ad}_{g^{-1}}(v))) \\
&= (\mathrm{Ad}_{g^{-1}}(v))^\sharp_p.
\end{aligned}$$

□

10.6 Maurer–Cartan form

A 1-form $\omega \in \Omega^1(X)$ is a smooth collection of linear functionals $\omega_x : T_xX \to \mathbb{R}, x \in X$. More generally, we can consider 1-forms $\omega \in \Omega^1(X, V)$ which are valued in a fixed vector space V. This means a smooth assignment of linear maps, $\omega_x : T_xX \to V$. So ω takes a vector field to a V-valued function, in a $C^\infty(X)$-linear way. The smoothness means that upon picking any basis for V, ω becomes an n-tuple of ordinary $\mathbb{R}$-valued 1-forms. More invariantly, for any $\lambda \in V^*$, the $\mathbb{R}$-valued 1-form $\lambda \circ \omega$ is smooth. We can pull back V-valued 1-forms in the same way as Eq. (8.3), and the pullback still has the properties found in Exercise 8.5.

We are primarily interested in the case $V = \mathfrak{g}$.

Definition 36. The Maurer–Cartan form on a Lie group is a $T_eG \cong \mathfrak{g}$-valued 1-form $\Theta \in \Omega^1(G, \mathfrak{g})$, defined by

$$\Theta_g(v_g) := (L_{g^{-1}})_*(v_g), \qquad v_g \in T_gG. \tag{10.10}$$

The idea is use T_eG as a "reference tangent space", and utilize the group action L to canonically convert tangent vectors at general g to tangent vectors at e. Doing this to an arbitrary vector field on G will turn it into a $\mathfrak{g}$-valued function — this is what Θ is doing for us.

Exercise 10.5. Show that the Maurer–Cartan form Θ on a Lie group G satisfies left-invariance,

$$(L_g)^*\Theta = \Theta, \qquad \forall g \in G,$$

fixes fundamental vector fields on G,

$$\Theta(v) = v, \qquad v \in \mathfrak{g}, \tag{10.11}$$

and satisfies right-equivariance,

$$(R_g)^*\Theta = \mathrm{Ad}_{g^{-1}} \circ \Theta. \tag{10.12}$$

For $G = \mathrm{GL}(n)$, the Maurer–Cartan form is usually written as

$$\Theta = g^{-1}dg.$$

What this means is that at each point $g' \in \mathrm{GL}(n, \mathbb{R})$, we have the matrix $(g')^{-1}$ multiplied into the matrix of coordinate 1-forms at g',

$$(g^{-1}dg)_{g'} = (g')^{-1} \cdot \begin{pmatrix} (dX^{11})_{g'} & \cdots & (dX^{1n})_{g'} \\ \vdots & \ddots & \vdots \\ (dX^{n1})_{g'} & \cdots & (dX^{nn})_{g'} \end{pmatrix}, \tag{10.13}$$

thus $g^{-1}dg$ is a $\mathrm{M}_n(\mathbb{R}) = \mathfrak{gl}(n, \mathbb{R})$-valued 1-form. One checks that $g^{-1}dg$ is indeed the Maurer–Cartan form of $\mathrm{GL}(n)$ according to Definition 36.

For $\mathrm{GL}(n, \mathbb{C})$, the coordinate X^{ij} is replaced by a complex coordinate Z^{ij}, which should be thought of as a pair of real coordinates, $Z^{ij} = X^{ij} + iY^{ij}$. The Maurer–Cartan form has the same formula, with dX^{ij} replaced by dZ^{ij}.

For a matrix Lie group $G \subset \mathrm{GL}(n)$, the Maurer–Cartan form is still given by the same formula, but the 1-forms dX^{ij} are restricted to act on TG.

Exercise 10.6. Work out the Maurer–Cartan form for the matrix Lie groups $\mathrm{U}(1)$ and $\mathrm{SU}(2)$.

Remark. When G acts on a manifold P, the condition Eq. (10.12) still makes sense, as does Eq. (10.11) with $\Theta(v^\sharp)$ in place of $\Theta(v)$. As we will see, these constitute the defining conditions for a connection on a principal G-bundle P.

10.7 Derivative of Lie group action

For later use, we record the following calculation:

Proposition 10.10. *Let $\sigma : P \times G \to P$ denote the action of a Lie group G on a manifold P. Then*

$$d\sigma_{(p,g)}(v,\xi) = d(R_g)_p(v) + (\Theta(\xi))^\sharp_{p\cdot g}, \qquad (v,\xi) \in T_pP \oplus T_gG, \tag{10.14}$$

where $R_g : P \to P$ denotes right-multiplication-by-g, and Θ is the Maurer–Cartan form on G.

Proof. First consider $(v,0) \in T_pP \oplus T_gG$. We have v being represented by a curve γ in P with $\gamma(0) = p$ and $\gamma'(0) = v$. So $(v,0)$ is represented by the curve $t \mapsto (\gamma(t), g)$ in $P \times G$. By Prop. 4.1,

$$d\sigma_{(p,g)}(v,0) = \frac{d(\sigma(\gamma(t),g))}{dt}\Big|_{t=0} = \frac{d(R_g \circ \gamma(t))}{dt}\Big|_{t=0} = d(R_g)_p(\gamma'(0)) = d(R_g)_p(v).$$

Next, consider $(0,\xi) \in T_p \oplus T_gG$. We have

$$\xi = (L_g)_*(L_{g^{-1}})_*(\xi) = (L_g)_*(\Theta(\xi)) = \frac{d(g\exp(t\Theta(\xi)))}{dt}\Big|_{t=0}.$$

Then

$$\begin{aligned} d\sigma_{(p,g)}(0,\xi) &= \frac{d(p\cdot g\exp(t\Theta(\xi)))}{dt}\Big|_{t=0} \\ &= (\Theta(\xi))^\sharp_{p\cdot g}, \end{aligned}$$

by definition of the fundamental vector field associated to $\Theta(\xi) \in \mathfrak{g}$. $\square$

11 Connections on principal bundles and their curvature

We are finally ready to relate frames attached to different points.

11.1 Canonical vertical bundle

Let $\pi : P \to X$ be a principal G-bundle. Each $p \in P$ lies in some fibre $P_{\pi(p)}$ diffeomorphic to G. So we may consider the *vertical tangent space* at p,

$$V_p := T_p P_{\pi(p)},$$

which has dimension $\dim G$. A vertical tangent vector at p is geometrically represented by some curve through p lying within its fibre. Such a curve becomes a constant curve upon application of π. This means that

$$V_p \subset \ker d\pi_p.$$

Because π is a submersion, every $d\pi_p$ is surjective, so

$$\dim \ker d\pi_p = \dim P - \dim X = \dim G.$$

Thus $V_p = \ker d\pi_p$.

Globally, the *vertical bundle* of P sits inside the full tangent bundle of P as the subset

$$VP := \ker(d\pi : TP \to TX) = \bigsqcup_{p \in P} V_p \ \subset TP.$$

To see that $\pi : VP \to X$ is indeed a subbundle of TP, we use the fundamental vector fields $v^\sharp$ on P associated to $v \in \mathfrak{g} \cong T_e G$, induced by the G-action on P (Definition 35). For each $p \in P$, the orbit map

$$\mathcal{O}_p : G \to P_{\pi(p)}, \qquad g \mapsto p \cdot g,$$

pushes v to $v_p^\sharp$ (Eq. (10.9)),

$$\begin{aligned} (d\mathcal{O}_p)_e : \overbrace{T_e G}^{\mathfrak{g}} &\longrightarrow T_p P_{\pi(p)} \equiv V_p \\ v &\mapsto v_p^\sharp. \end{aligned} \tag{11.1}$$

Since P is a principal bundle, $\mathcal{O}_p$ is a diffeomorphism, so $(d\mathcal{O}_p)_e$ is actually a linear isomorphism. Thus Eq. (11.1) gives a canonical way to identify each V_p with the "reference" $\mathfrak{g}$, generalizing the Maurer–Cartan form.

Remark 4. Any basis of $\mathfrak{g}$ leads, via Eq. (11.1), to $(\dim G)$ smooth vector fields which span V_p at every $p \in P$. This framing property shows that VP is a subbundle of TP (see Lemma 10.32 of [11]). In fact, there is a canonical vector bundle isomorphism,

$$\begin{aligned} P \times \mathfrak{g} &\to VP \\ (p, v) &\mapsto v_p^\sharp. \end{aligned}$$

So VP is trivializable, in the same way that a Lie group is (Corollary 10.3).

We also observe that a gauge transformation of P takes vertical tangent vectors to vertical tangent vectors, and therefore leaves VP invariant.

11.2 Connection as horizontal bundle

Let us proceed to make sense of the "bundle of horizontal directions in P". There is no canonical meaning of "horizontal", so an extra piece of geometric data must be supplied.

Definition 37. A *connection* on a principal G-bundle P is a choice of G-invariant *horizontal bundle/distribution* HP on P, i.e., HP is a subbundle of TX such that

$$\begin{aligned} H_p \oplus V_p &= T_pP, \\ (R_g)_* H_p &= H_{pg}, \qquad\qquad \forall p \in P,\ g \in G. \end{aligned}$$

The first condition is quite an obvious one. The second condition is imposed because all points in a fibre are supposed to be treated on an equal footing. The term *principal connection* is sometimes used to stress this second property.

Once P is equipped with a connection, then

$$d\pi_p : T_pP = H_p \oplus V_p \to T_{\pi(p)}X$$

restricts to an isomorphism $H_p \to T_{\pi(p)}X$. Given $\check{v}_x \in T_xX$ and a fibre point $p \in \pi^{-1}(x)$, the corresponding vector in H_p under this isomorphism is called the *horizontal lift* of $\check{v}_x$ at p.

Remark. In general, there are many vectors $v_p = (v_p^{\mathrm{H}}, v_p^{\mathrm{V}}) \in H_p \oplus V_p = T_pP$ which *lift* $\check{v}_x$ in the sense that $d\pi_p(v_p) = \check{v}_x$. Nevertheless, any such lift has the same horizontal component, namely, the horizontal lift of $\check{v}_x$ at p.

Definition 38. With respect to a connection on a principal G-bundle $\pi : P \to X$, the *horizontal lift* of $\check{v} \in \mathfrak{X}(X)$ is the unique vector field $v^{\mathrm{H}} \in \mathfrak{X}(P)$, such that at every $p \in P$, the vector $(v^{\mathrm{H}})_p \in H_p$ is the horizontal lift of $\check{v}_{\pi(p)}$ at p.

Exercise 11.1. Show that for any $\check{v} \in \mathfrak{X}(X)$, its horizontal lift is a G-invariant vector field on P, i.e., $(R_g)_* v^{\mathrm{H}} = v^{\mathrm{H}}$ for all $g \in G$.

Exercise 11.2. Let u be a G-invariant vector field on P, and $v^\sharp$ be a fundamental vector field on P (thus $v^\sharp$ is vertical). Show that $[u, v^\sharp] = 0$.

11.3 Connection as 1-form

While Definition 37 directly captures the geometric essence of the connection concept, it is not so convenient for calculations, so we seek a more algebraic formulation.

Given a connection, we are able to uniquely decompose a general tangent vector $v_p \in T_pP$ into its vertical and horizontal components,

$$v_p = v_p^{\mathrm{H}} + v_p^{\mathrm{V}}, \qquad v_p^{\mathrm{H}} \in H_p,\ v_p^{\mathrm{V}} \in V_p.$$

We would like to define a map which projects each v_p onto its vertical component. Although the various V_p are not the "same" vector space, we have seen how to identify every V_p with $\mathfrak{g}$, using Eq. (11.1). So we can define the $\mathfrak{g}$-valued 1-form ω on P,

$$\begin{aligned} \omega_p : T_pP \to\ & V_p \overset{(d\mathcal{O}_p)_e^{-1}}{\to} \mathfrak{g} \\ v_p \mapsto\ & v_p^{\mathrm{V}} \quad \mapsto \quad u, \end{aligned} \tag{11.2}$$

where u is such that $u_p^\sharp = v_p^{\mathrm{V}}$.

Remark. Near any $p \in P$, we can find some neighbourhood U on which TU is the direct sum of a vertical bundle and horizontal bundle. Since ω just projects onto the vertical part (which is trivialized to $U \times \mathfrak{g}$ according to Remark 4), we see that ω is smooth.

The 1-form ω defined in Eq. (11.2) has the following properties:

- ω takes a (vertical) fundamental vector field $u^\sharp$ back to u,

$$\omega_p(u_p^\sharp) = u, \qquad \forall p \in P, u \in \mathfrak{g}. \tag{11.3}$$

- ω is G-equivariant,

$$R_g^*\omega = \mathrm{Ad}_{g^{-1}} \circ \omega, \qquad \forall g \in G. \tag{11.4}$$

The equivariance in Eq. (11.4) may be deduced by observing that, by construction, ω annihilates the G-invariant horizontal bundle (Exercise).

Our second definition of a connection is:

Definition 39. A connection on a principal G bundle P is a $\mathfrak{g}$-valued 1-form on P, which satisfies Eq.(11.3)-(11.4).

Relation between the two definitions of connection. From Definition 39 of a connection, we define the horizontal subspaces of T_pP to be $\ker(\omega_p), p \in P$. Conditions (11.3)-(11.4) ensure that ω is "vertically canonical", and that the kernels of ω_p assemble into a G-invariant "horizontal distribution". This recovers the notion of a connection in the sense of Definition 37.

Remark. Intuitively, $p \mapsto p \cdot g$ corresponds to "rotating" a reference frame by g. Then the G-symmetry of frames with respect to p as reference, needs to be conjugated by g^{-1} when $p \cdot g$ is used as reference. Accordingly, the labelling of frame-wise (i.e. vertical) directions by $\mathfrak{g}$ needs to be adjusted by applying $\mathrm{Ad}_{g^{-1}}$.

Example 11.1. Consider $P = G$ as a principal G-bundle over a point. In Example 10.8, we saw that $v^\sharp = v$, $v \in \mathfrak{g}$. So we see from Exercise 10.5 that the Maurer–Cartan form Θ is actually a connection on G.

Example 11.2. The *trivial connection* on $X \times G$ is $\omega_{\mathrm{triv}} = \mathrm{pr}_2^*\Theta$, where Θ is the Maurer–Cartan form on G. Here, pr_2 denotes projection on to the G factor. Thus ω_{triv} restricts to the Maurer–Cartan form on each fibre $\{x\} \times G$, while the horizontal subspaces are specified as $H_{(x,g)} = T_xX \oplus 0$.

Let P be a trivializable principal G-bundle over X, with trivialization Ψ written as

$$\begin{aligned} \Psi : P &\xrightarrow{\cong} X \times G \\ p &\mapsto (\pi(p), \psi(p)). \end{aligned}$$

Then

$$\psi^*\Theta = (\mathrm{pr}_2 \circ \Psi)^*(\Theta) = \Psi^*(\mathrm{pr}_2^*\Theta) = \Psi^*(\omega_{\mathrm{triv}})$$

is a connection on P (exercise).

Without reference to a trivialization, there is no meaning to "trivial connection" on P. Furthermore, if P is not trivializable, there is no notion of "trivial connection" in the first place.

Example 11.3. Let $P = S^1 \times \mathrm{U}(1)$, so $\mathfrak{g} = i\mathbb{R}$. Let $d\theta, d\varphi$ be the respective coordinate 1-forms on S^1 and $\mathrm{U}(1)$ (Example 8.2). The Maurer–Cartan form on $\mathrm{U}(1)$ is $id\varphi$. For $i \in \mathfrak{g}$, the corresponding fundamental vector field on P is $\frac{\partial}{\partial\varphi}$ (Exercise). Fix any $ik \in \mathfrak{g}$, and consider the $\mathfrak{g}$-valued 1-form $\omega^{(k)} := ik\, d\theta + id\varphi$. Then $\omega^{(k)}$ is a connection 1-form on P. The horizontal subspaces defined by $\omega^{(k)}$ are spanned by $\partial_\theta - k\partial_\varphi$.

Example 11.4. If G is a discrete group, then $\mathfrak{g} = 0$, and the only available connection 1-form is the zero-valued 1-form. As Example 6.1 of $S(\mathcal{L}^{\mathbb{R}})$ shows, it is possible to have vanishing connection 1-form, with globally non-trivial "horizontal parallel transport".

Exercise 11.3. Let ω, ω' be two connections on a principal G bundle P. Show that $\omega - \omega'$ is a *horizontal* 1-form in the sense that it annihilates all vertical vectors,

$$(\omega - \omega')_p(v_p) = 0, \qquad \forall v_p \in V_p,\ p \in P.$$

11.4 Local description of connection: gauge potentials

It is useful to have a description of a connection as an object living on X, rather than on P.

Definition 40. Let ω be a connection on a principal G-bundle P (Definition 39). Let $s : U \to P$ be a local section/gauge over an open subset $U \subset X$. Then $\omega^{(s)} := s^*\omega$ is a $\mathfrak{g}$-valued 1-form on U, called a *local gauge potential*.

Example 11.5. On the trivial principal bundle $S^1 \times \mathrm{U}(1)$, consider the connection 1-form $\omega = ikd\theta + id\varphi$ of Example 11.3. Pulling back by the trivializing section $s_0 : e^{i\theta} \mapsto (e^{i\theta}, 1)$, we get the gauge potential $s_0^*\omega = ikd\theta$. For $n \in \mathbb{Z}$, apply the large gauge transformation $g_n \in \mathrm{Map}(S^1, \mathrm{U}(1))$, $g_n(e^{i\theta}) = e^{in\theta}$, which converts s_0 to the section $s_n = s_0 \cdot g_n$. Check that ω is now represented by the gauge potential $s_n^*\omega = i(k+n)d\theta$ (Exercise).

Theorem 11.1. *Let $\omega \in \Omega^1(P, \mathfrak{g})$ be a connection on a principal G bundle P. Let s_α, s_β be local sections of P defined over open subsets U_α, U_β respectively, related by the local gauge transformation $g_{\alpha\beta} : U_{\alpha\beta} \to G$ (Eq. (6.9)). Write $\omega^{(s_\alpha)}, \omega^{(s_\beta)}$ for the respective local gauge potentials. On the overlap $U_{\alpha\beta}$, they are related by*

$$\omega^{(s_\beta)} = \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \circ \omega^{(s_\alpha)} + g_{\alpha\beta}^*(\Theta), \tag{11.5}$$

where Θ is the Maurer–Cartan form on G.

Proof. We have $s_\beta(x) = \sigma(s_\alpha(x), g_{\alpha\beta}(x)), x \in U_{\alpha\beta}$. Thus $s_\beta = \sigma \circ (s_\alpha, g_{\alpha\beta})$, and its derivative is

$$d(s_\beta)_x = d\sigma_{(s_\alpha(x), g_{\alpha\beta}(x))} \circ (d(s_\alpha)_x, d(g_{\alpha\beta})_x), \qquad x \in U_{\alpha\beta}. \tag{11.6}$$

We computed $d\sigma$ in Prop. 10.10, with the formula Eq. (10.14) recalled below,

$$d\sigma_{(p,g)}(v, \xi) = d(R_g)_p(v) + (\Theta(\xi))^\sharp_{p \cdot g}.$$

Substitute this into Eq. (11.6) and evaluate on some $\eta \in T_x X$,

$$\begin{aligned} d(s_\beta)_x(\eta) &= d(R_{g_{\alpha\beta}(x)})_{s_\alpha(x)}(d(s_\alpha)_x(\eta)) \ + \ (\Theta(d(g_{\alpha\beta})_x(\eta)))^\sharp_{s_\beta(x)} \\ &= d(R_{g_{\alpha\beta}(x)})_{s_\alpha(x)}(d(s_\alpha)_x(\eta)) \ + \ ((g_{\alpha\beta}^*\Theta)_x(\eta))^\sharp_{s_\beta(x)}. \end{aligned} \tag{11.7}$$

For convenience, we drop the reference to x. Apply ω to the first term of Eq. (11.7),

$$\begin{aligned} \omega\left(d(R_{g_{\alpha\beta}})_{s_\alpha}(d(s_\alpha)(\eta))\right) &= (R_{g_{\alpha\beta}}^*\omega)(d(s_\alpha)(\eta)) \\ &= \mathrm{Ad}_{g_{\alpha\beta}^{-1}}\big(\omega(d(s_\alpha)(\eta))\big) \qquad \text{(Eq. (11.4))} \\ &= \mathrm{Ad}_{g_{\alpha\beta}^{-1}}(s_\alpha^*\omega(\eta)) \\ &= \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \circ \omega^{(s_\alpha)}(\eta). \end{aligned}$$

For the second term, we get

$$\omega((g_{\alpha\beta}^*\Theta)(\eta))^\sharp_{s_\beta} \overset{\text{Eq. (11.3)}}{=} (g_{\alpha\beta}^*\Theta)(\eta)$$

In total,

$$\underbrace{\omega(d(s_\beta)(\eta))}_{=s_\beta^*\omega(\eta) \equiv \omega^{(s_\beta)}(\eta)} = (\mathrm{Ad}_{g_{\alpha\beta}^{-1}} \circ \omega^{(s_\alpha)} + g_{\alpha\beta}^*\Theta)(\eta),$$

which is the desired transformation law, Eq. (11.5). □

Remark. When G is abelian, the transformation law, Eq. (11.5), simplifies to $\omega^{(s_\beta)} = \omega^{(s_\alpha)} + g^*_{\alpha\beta}(\Theta)$.

We may think of Eq. (11.5) as the "action" of local gauge transformations $g_{\alpha\beta}$ on local gauge potentials, effecting $\omega^{(s_\alpha)} \rightsquigarrow \omega^{(s_\beta)}$.

Exercise 11.4. Let $s_\gamma : U_\gamma \to P$ be a third local section. So we have the local gauge transformation $g_{\beta\gamma} : U_{\beta\gamma} \to G$ relating s_β to s_γ, as well as $g_{\alpha\gamma} : U_{\alpha\gamma} \to G$ relating s_α to s_γ. Show that on the triple overlap $U_{\alpha\beta\gamma}$, applying $g_{\alpha\gamma}$ to a local gauge potential $\omega^{(s_\alpha)}$ gives the same result as applying $g_{\alpha\beta}$ followed by $g_{\beta\gamma}$.

Example 11.6. In Example 11.5, the section s_n is obtained from s_0 by applying $g_{0n} : S^1 \to G$, $e^{i\theta} \mapsto e^{in\theta}$. The transformation law, Eq. (11.5), correctly gives

$$\omega^{(s_n)} = \omega^{(s_0)} + g^*_{0n}(id\varphi) = ikd\theta + ind\theta.$$

Local versus global. We emphasize that a gauge potential is a *local and gauge-dependent* description of a connection. Since P is generally not trivializable, we need a collection of locally defined gauge potentials $\omega^{(s_\alpha)}$ with respect to some trivializing cover, in order completely capture the connection data.

Consider a trivializing cover of P specified by local sections $s_\alpha, \alpha \in \mathcal{I}$. Suppose we have a collection of $\mathfrak{g}$-valued local 1-forms $\mathcal{A}_\alpha \in \Omega^1(U_\alpha, \mathfrak{g})$, such that the transformation law, Eq. (11.5) holds on every overlap $U_{\alpha\beta}$,

$$\mathcal{A}_\beta = \mathrm{Ad}_{g^{-1}_{\alpha\beta}} \circ \mathcal{A}_\alpha + g^*_{\alpha\beta}(\Theta),$$

where $g_{\alpha\beta}$ are the local gauge transformations converting s_α to s_β. Then it may be shown that there exists a *unique* connection ω on P such that $\mathcal{A}_\alpha = s^*_\alpha\omega$. This "algebraic" method of defining connections via local gauge potentials and checking (or declaring) mutual compatibility, is usually utilized in physics. However, the global geometrical meaning of the connection becomes rather obscured.

Finally, we remark on the terminology "gauge field". Often, a local gauge potential is called a "gauge field on X". A somewhat better choice is to call a connection on P a gauge field. Actually, it is possible to show that the *difference* between two connections is a section of the so-called *adjoint vector bundle* over X associated to P. So the space of connections could be identified as a space of fields over X, but only *affinely* (there is no canonical "trivial" or "zero" connection).

11.5 Curvature of connections

The exterior derivative of a k-form generalizes in the obvious way to V-valued k-forms, giving rise to V-valued $(k+1)$-forms. Namely, we can expand a V-valued 1-form into $\mathbb{R}$-valued k-forms with respect to any basis for V. Then take the usual exterior derivative of each component, and sum the result up.

Definition 41. Let $\omega \in \Omega^1(P, \mathfrak{g})$ be a connection 1-form on a principal G-bundle P. Its *curvature* is the $\mathfrak{g}$-valued 2-form

$$\Omega(u, v) = d\omega(u^{\mathrm{H}}, v^{\mathrm{H}}), \qquad u, v \in \mathfrak{X}(P). \tag{11.8}$$

Recall that $(\cdot)^{\mathrm{H}}$ denotes the *horizontal part* of a vector field on P, as determined by the connection ω. The expression $d\omega(u, v)$ would generally involve three terms, as in Eq. (8.8). But on the right side of Eq. (11.8), we only put in horizontal vector fields. Because ω annihilates horizontal vector fields, we are left with only a single term,

$$\Omega(u, v) = d\omega(u^{\mathrm{H}}, v^{\mathrm{H}}) = -\omega([u^{\mathrm{H}}, v^{\mathrm{H}}]).$$

Intuitively: Starting from p, we imagine moving a small distance horizontally along u^{H}, then horizontally along v^{H}. Now do this in the opposite order. If the horizontal distribution defined by ω is "curved", we might not end up at the same point! Infinitesimally, the endpoint mismatch is an infinitesimal vertical offset along the final fibre, i.e., a vertical tangent vector (identified with a Lie algebra element as usual); the curvature form Ω measures such mismatches.

Definition 42. Let ω be a connection on a principal G-bundle P. The *horizontal part* η^{H} of a k-form $\eta \in \Omega^k(P, V)$ is defined to be

$$\eta^{\mathrm{H}}(v_1, \ldots, v_k) := \eta(v_1^{\mathrm{H}}, \ldots, v_k^{\mathrm{H}}), \qquad v_i \in \mathfrak{X}(P),$$

and the *exterior covariant derivative* of a k-form η with respect to ω is defined to be

$$D^\omega \eta := (d\eta)^{\mathrm{H}}.$$

In particular, the curvature of ω, Eq. (11.8), can be written as

$$\Omega = D^\omega(\omega),$$

i.e., the exterior covariant derivative of ω with respect to itself.

Recall that a connection 1-form is G-equivariant (Eq. (11.4)). Similarly:

Proposition 11.2. *The curvature Ω of a connection ω is G-equivariant,*

$$R_g^*\Omega = \mathrm{Ad}_{g^{-1}} \circ \Omega, \qquad \forall\, g \in G.$$

Proof. Because $(R_g)_*$ respects the horizontal-vertical splitting of tangent spaces, we have

$$(R_g)_* v^{\mathrm{H}} = ((R_g)_* v)^{\mathrm{H}}, \qquad v \in \mathfrak{X}(P).$$

So for any $u, v \in \mathfrak{X}(P)$, we have

$$\begin{aligned}
(R_g^*\Omega)(u,v) &= \Omega((R_g)_* u, (R_g)_* v) \\
&= d\omega(((R_g)_* u)^{\mathrm{H}}, ((R_g)_* v)^{\mathrm{H}}) \\
&= d\omega((R_g)_* u^{\mathrm{H}}, (R_g)_* v^{\mathrm{H}}) \\
&= (R_g^*(d\omega))(u^{\mathrm{H}}, v^{\mathrm{H}}) \\
&= d(R_g^*\omega)(u^{\mathrm{H}}, v^{\mathrm{H}}) \\
&= d(\mathrm{Ad}_{g^{-1}} \circ \omega)(u^{\mathrm{H}}, v^{\mathrm{H}}) \\
&= \mathrm{Ad}_{g^{-1}} \circ (d\omega)(u^{\mathrm{H}}, v^{\mathrm{H}}) = \mathrm{Ad}_{g^{-1}} \circ \Omega(u,v).
\end{aligned}$$

□

Definition 43. The *wedge product* of $\eta, \omega \in \Omega^1(X, \mathfrak{g})$ is the $\mathfrak{g}$-valued 2-form defined by

$$[\eta, \omega](u,v)(x) = [\eta(u)(x), \omega(v)(x)] - [\eta(v)(x), \omega(u)(x)], \qquad u, v \in \mathfrak{X}(X),$$

where $[\cdot,\cdot]$ on the right side denotes the Lie bracket in $\mathfrak{g}$.

This is like an ordinary wedge product, but instead of pointwise multiplying $\eta(u)$ and $\omega(v)$, we take the pointwise Lie bracket. In the same way, we can take wedge products of $\mathfrak{g}$-valued k-forms and l-forms.

Theorem 11.3 (Cartan structure equation)**.** *The curvature of a connection 1-form $\omega \in \Omega^1(P, \mathfrak{g})$ can be expressed as*

$$\Omega = d\omega + \frac{1}{2}[\omega, \omega]. \tag{11.9}$$

Proof. Exercise. It is helpful to split the calculation of $\Omega(u,v)$ into three cases: u, v both horizontal, u, v both vertical, u vertical and v horizontal. □

Remark. In the particular case where $P = G$ and $\omega = \Theta$ is the Maurer–Cartan form, there are no horizontal directions at all. So $d\Theta + \frac{1}{2}[\Theta, \Theta] = 0$.

Exercise 11.5. Using the Cartan structure equation and Eq. (8.10), verify the *Bianchi indentity*: $D^\omega \Omega \equiv (d\Omega)^{\mathrm{H}} = 0$ for the curvature Ω of any connection ω.

11.5.1 Local field strength

As with the local gauge potentials, we define:

Definition 44. Let Ω be the curvature of a connection ω on a principal G-bundle P. Let $s : U \to P$ be a local section of P. Then $s^*\Omega \in \Omega^2(U, \mathfrak{g})$ is called the *local field strength* of ω.

Because pullback is compatible with wedge product and the exterior derivative, it is straightforward to verify the local form of the structure equations,

$$s^*\Omega = d\omega^{(s)} + \frac{1}{2}[\omega^{(s)}, \omega^{(s)}]. \tag{11.10}$$

The local field strengths are gauge-dependent, so let us work out their transformation law under local gauge transformations.

Theorem 11.4. *Let Ω be the curvature of a connection ω on a principal G-bundle. Let $s_\alpha : U_\alpha \to P$ and $s_\beta : U_\beta \to P$ be two local gauges/sections, related on the overlap $U_{\alpha\beta}$ by the local gauge transform $g_{\alpha\beta} : U_{\alpha\beta} \to G$. On $U_{\alpha\beta}$, the local field strengths are related by*

$$s_\beta^*\Omega = \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \circ s_\alpha^*\Omega. \tag{11.11}$$

Proof. We need to compute

$$s_\beta^*\Omega(u, v) = \Omega(ds_\beta(u), ds_\beta(v)), \qquad u, v \in \mathfrak{X}(U).$$

While proving the transformation law for the local gauge potentials, Theorem 11.1, we derived Eq. (11.7),

$$d(s_\beta)_x(\eta) = d(R_{g_{\alpha\beta}(x)})_{s_\alpha(x)} \circ d(s_\alpha)_x(\eta) + ((g_{\alpha\beta}^*\Theta)_x(\eta))^\sharp_{s_\beta(x)}, \qquad \eta \in T_xX.$$

The second term on the right is vertical, so it does not contribute to $\Omega(\cdot,\cdot)$. Thus for any $u_x, v_x \in T_xX$,

$$\begin{aligned}
(s_\beta^*\Omega)_x(u_x, v_x) &= \Omega_{s_\beta(x)}(d(R_{g_{\alpha\beta}(x)})_{s_\alpha(x)} \circ d(s_\alpha)_x(u_x), d(R_{g_{\alpha\beta}(x)})_{s_\alpha(x)} \circ d(s_\alpha(x))(v_x)) \\
&= (R^*_{g_{\alpha\beta}(x)}\Omega)_{s_\alpha(x)}(d(s_\alpha)_x(u_x), d(s_\alpha)_x(v_x)) \\
&= (\mathrm{Ad}_{g^{-1}_{\alpha\beta}(x)} \circ \Omega)_{s_\alpha(x)}(d(s_\alpha)_x(u_x), d(s_\alpha)_x(v_x)) \\
&= \mathrm{Ad}_{g^{-1}_{\alpha\beta}(x)} \circ (s_\alpha^*\Omega)_x(u_x, v_x),
\end{aligned}$$

as required for Eq. (11.11). □

11.6 Physics notation

The physicists' notation for the local gauge potentials and local field strengths are

$$\mathcal{A} = s^*\omega, \qquad \mathcal{F} = s^*\Omega.$$

where reference to a local section s (or local trivialization) is usually left implicit. It is seldom mentioned that these potentials arise from a globally defined object (the connection on P). So one only has the local structure equation, Eq. (11.10),

$$\mathcal{F} = d\mathcal{A} + \frac{1}{2}[\mathcal{A}, \mathcal{A}], \tag{11.12}$$

which is usually taken as the *definition* of the (local) curvature 2-form.

Furthermore, one often works in a coordinate chart (U, φ), and expands $\mathcal{A}$ in terms of coordinate 1-forms. So $\mathcal{A}$ (more precisely, $(\varphi^{-1})^*\mathcal{A}$) is written as

$$\mathcal{A} = \sum_i^n \mathcal{A}_i dx^i,$$

where each component $\mathcal{A}_i$ is a smooth $\mathfrak{g}$-valued function on $\varphi(U)$. Similarly, $\mathcal{F} = \sum_{i,j=1}^n \mathcal{F}_{ij} dx^i \wedge dx^j$, where (Exercise)

$$\mathcal{F}_{ij} = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i + [\mathcal{A}_i, \mathcal{A}_j]$$

is the coordinate-dependent expression of Eq. (11.12).

Next, G is usually identified as a matrix Lie group, $G \subset \mathrm{GL}(n, \mathbb{R})$, so $\Theta = g^{-1}dg$ (see Eq. (10.13)). Then the second term in the gauge transformation

law, Eq. (11.5), is

$$(g^*_{\alpha\beta}(\Theta))_x = g^{-1}_{\alpha\beta}(x) \cdot \begin{pmatrix} d(g^{11}_{\alpha\beta})_x & \cdots & d(g^{1n}_{\alpha\beta})_x \\ \vdots & \ddots & \vdots \\ d(g^{n1}_{\alpha\beta})_x & \cdots & d(g^{nn}_{\alpha\beta})_x \end{pmatrix},$$

with $g^{ij}_{\alpha\beta} = X^{ij} \circ g_{\alpha\beta} : U \to \mathbb{R}$ being the ij-th matrix entry of the transition function $g_{\alpha\beta}$. It is common practice to simply write $g = g_{\alpha\beta} : U_{\alpha\beta} \to G$ for the local gauge transformation, and condense the above expression into "$g^{-1}dg$". The gauge transformation rules are then

$$\begin{aligned} \mathcal{A} &\rightsquigarrow \mathcal{A}' = g^{-1}\mathcal{A}g + g^{-1}dg \\ \mathcal{F} &\rightsquigarrow \mathcal{F}' = g^{-1}\mathcal{F}g. \end{aligned} \tag{11.13}$$

Remark. Sometimes, g is replaced by g^{-1} to get a left action of local gauge transformations, and there is a factor of i relating physicists' "matrix Lie algebras" and the mathematicians' Lie algebras. The matrix identity $dg = -g(dg^{-1})g$ may be used to write $-dg^{-1}g$ instead of $g^{-1}dg$.

11.7 Globally gauge-transformed connections and curvatures

We saw how the local gauge potentials and field strengths transform under a change of local gauge (i.e. choosing a different local section), Eq. (11.5)-(11.11). Before we discuss the global picture, we make a general observation.

Pullback connection. Let η be a G-equivariant $\mathfrak{g}$-valued k-form on a principal G-bundle P_2,

$$R^*_g\eta = \mathrm{Ad}_{g^{-1}} \circ \eta, \qquad \forall g \in G.$$

If $F : P_1 \to P_2$ is a morphism of principal G-bundles, then $F^*\eta$ is a G-equivariant k-form on P_1, because of

$$R^*_g(F^*\eta) = (F \circ R_g)^*\eta = (R_g \circ F)^*\eta = F^*(R^*_g\eta) = F^*(\mathrm{Ad}_{g^{-1}} \circ \eta) = \mathrm{Ad}_{g^{-1}} \circ (F^*\eta).$$

Thus, for a connection ω on P_2, the pullback $F^*\omega$ remains G-equivariant.

Furthermore, if $v^\sharp$ is a fundamental vector field on P_1, then

$$\begin{aligned}(dF)_p(v_p^\sharp) &= (dF)_p\left(\frac{d(p\cdot\exp(tv))}{dt}\Big|_{t=0}\right)\\ &= \frac{d(F(R_{\exp(tv)}(p)))}{dt}\Big|_{t=0}\\ &= \frac{d(R_{\exp(tv)}(F(p)))}{dt}\Big|_{t=0} = v^\sharp_{F(p)},\end{aligned}$$

so that

$$(F^*\omega)_p(v_p^\sharp) = \omega_{F(p)}((dF)_p(v_p^\sharp)) = \omega_{F(p)}(v^\sharp_{F(p)}) = v, \qquad \forall\, p\in P_1.$$

Thus $F^*\omega$ is a connection 1-form on P_1.

In particular, gauge transformations of P act on connections on P by pull-back.

Theorem 11.5. *Let $\omega\in\Omega^1(P,\mathfrak{g})$ be a connection on a principal G-bundle P, and let $F\in\mathcal{G}(P)$ be a gauge transformation, implemented by right multiplication by an equivariant map $\sigma_F\in\mathrm{Map}(P,G)^G$ as in Eq.* (6.11)*. Then*

$$F^*\omega = \mathrm{Ad}_{\sigma_F^{-1}}\circ\omega + \sigma_F^*\Theta. \tag{11.14}$$

Proof. The argument is similar to that in Theorem 11.1, and is omitted. □

Remark. A significant portion of gauge theory is concerned with understanding the "moduli space" of connections on P, after quotienting out by the above action of gauge transformations.

Exercise 11.6. Recall the connection 1-form $\omega = i(kd\theta + d\varphi)$ on $P = S^1\times \mathrm{U}(1)$. For $n\in\mathbb{Z}$, the map $F_n : (e^{i\theta}, e^{i\varphi})\mapsto(e^{i\theta}, e^{i\varphi+n\theta})$ is a (large) gauge transformation of P. Work out what $F_n^*\omega$ is.

For the curvature 2-form, the global version of Theorem 11.4 is

Theorem 11.6. *Let $F\in\mathcal{G}(P)$ be a gauge transformation of a principal G-bundle P. Let ω be a connection 1-form on P, and let $F^*\omega$ be its gauge transform (Eq.* (11.14)*). Write Ω^ω and $\Omega^{F^*\omega}$ for the respective curvature 2-forms. Then*

$$\Omega^{F^*\omega} = F^*\Omega^\omega = \mathrm{Ad}_{\sigma_F^{-1}}\circ\Omega^\omega.$$

Proof. For the first equality, we use the structure equations, and commutativity of pullback with wedge product,

$$\begin{aligned} F^*\Omega^\omega \equiv F^*(D^\omega\omega) &= F^*\left(d\omega + \frac{1}{2}[\omega,\omega]\right) \\ &= d(F^*\omega) + \frac{1}{2}[F^*\omega, F^*\omega] \\ &= D^{F^*\omega}(F^*\omega) \equiv \Omega^{F^*\omega}. \end{aligned}$$

For the second equality, let $u, v \in \mathfrak{X}(P)$. In the expression

$$F^*\Omega^\omega(u,v) = \Omega^\omega\left(F_*u, F_*v\right), \tag{11.15}$$

we may discard any vertical vectors occurring in the arguments on the right. To understand $F_* = dF$, we note that $F : P \to P$ may be written as the composition

$$p \mapsto (p, \sigma_F^{-1}(p)) \overset{\sigma}{\mapsto} p \cdot \sigma_F(p),$$

where $\sigma : P \times G \to P$ is the group action map. The derivative of σ was found in Eq. (10.14), and discarding vertical terms, we simply have

$$d\sigma_{(p,g)}(v_p, \xi_g) \text{ “}=\text{” } (dR_g)_p(v_p), \qquad v_p \in T_pP,\ \xi_g \in T_gG.$$

Substitute this into Eq. (11.15),

$$\begin{aligned} (F^*\Omega^\omega)_p(u_p, v_p) &= \Omega^\omega_{p\cdot\sigma_F(p)}\left((dR_{\sigma_F(p)})_p(u_p), (dR_{\sigma_F(p)})_p(v_p)\right) \\ &= \mathrm{Ad}_{\sigma_F^{-1}(p)} \circ \Omega^\omega_p\left(u_p, v_p\right) \end{aligned}$$

where the last line is due to Ω being G-equivariant. Thus

$$F^*\Omega^\omega = \mathrm{Ad}_{\sigma_F^{-1}} \circ \Omega^\omega.$$

□

11.7.1 Abelian G

Notice that when G is abelian, the curvature is simply $\Omega = d\omega$, which is linear in ω. In this case, the curvature is also invariant under gauge transformations. In particular, the local field strengths $s^*\omega$ are independent of the choice of local gauge/section s. This can be explicitly seen from Eq. (11.11). Thus we

may pick any collection of local sections, sufficient to cover all of X, and obtain a *globally* well-defined field strength $\mathcal{F} \in \Omega^2(X, \mathfrak{g})$.

Furthermore, the field strength is a *closed* 2-form, $d\mathcal{F} = 0$, because we may compute, using any local gauge, that

$$d\mathcal{F} = d(s^*\Omega) = d(d(\omega^{(s)})) + \frac{1}{2} d \underbrace{[\omega^{(s)}, \omega^{(s)}]}_{\text{abelian} \Rightarrow 0} = d^2(\omega^{(s)}) = 0.$$

However, it is important to remember that "$\mathcal{F} = d(\omega^{(s)})$", or "$\mathcal{F} = d\mathcal{A}$" in physics notation, holds only over the subset that s is defined on. One says that $\mathcal{F}$ is locally exact, but not globally *exact*. The failure of a closed 2-form on X (such as $\mathcal{F}$) to be globally exact, is measured by the *second cohomology group* of X. The latter is a topological invariant of X.

12 Fibre bundles and structure groups

Some motivation from physics. In classical geometry, principal G-bundles usually arise as a bundle of frames, for some concrete vector bundle of interest, such as a tangent bundle. In (quantum) physics, the principal G-bundle with connection encodes an abstract notion of "*local* gauge symmetry", or "inter-frame relationships", both at a single point of X, and between neighbouring points. The principal bundle can be realised as the frame bundle of some *associated vector bundle*, and here there is considerable freedom to (unitarily) *represent* G as transformations of some reference vector space. Different representations are labelled by "charges". The associated vector bundle is also viewed abstractly, as a *fibre bundle with structure group* G, and it is determined up to isomorphism.

Sections of the above-mentioned vector bundle are typically thought of as *matter fields* (electron, proton, etc.). The principal bundle connection induces a *covariant derivative* on the associated vector bundle, thereby allowing for a gauge-independent notion of differentiation for the matter fields. The connection itself could be allowed to vary — it then becomes a "dynamical gauge field interacting with matter fields". So we often need to consider the entire space of connections rather than a fixed prescribed one. Although the space of connections is (affinely) an infinite-dimensional linear space, in the end, everything is done modulo gauge transformations. This is because we cannot measure the section/connection as-it-is, but only the gauge-invariant aspects.

The above discussion can be taken as the "gauge principle"; compare the situation in general quantum mechanics. It strongly constrains the variety of admissible physical theories of matter and interactions.

12.1 Vector=[frame, components]

Let E be a vector space of dimension n. With respect to a frame/basis $\mathbf{e} = \{e_1, \ldots, e_n\}$, a vector $v \in E$ may be expanded as

$$v = \sum_{a=1}^{n} e_a v^a,$$

where $\mathbf{v} = (v^1, \ldots, v^n) \in \mathbb{K}^n$ is the n-tuple of components representing v. We could choose to expand v with respect to another frame $\mathbf{e}' = \mathbf{e} \cdot g$, $g \in \mathrm{GL}(n)$,

$$v = \sum_{b=1}^{n} e'_b (v')^b = \sum_{a,b=1}^{n} e_a g^a{}_b (v')^b = \sum_{a=1}^{n} e_a \underbrace{\sum_{b=1}^{n} g^a{}_b (v')^b}_{v^a}.$$

The new n-tuple of components $\mathbf{v}' = (v'^1, \ldots, v'^n)$ is related to the old one by

$$\mathbf{v}' = g^{-1} \cdot \mathbf{v} \qquad \text{with respect to} \qquad \mathbf{e}' = \mathbf{e} \cdot g. \tag{12.1}$$

Neither the choice of frame $\mathbf{e}$ nor the components $\mathbf{v}$ is intrinsic to $v \in E_x$. Only the *combination* $(\mathbf{e}, \mathbf{v})$ is intrinsic,

$$v = \sum_{a=1}^{n} e_a v^a = \sum_{a=1}^{n} e'_a (v')^a.$$

So, according to Eq. (12.1), we should make the identifications

$$(\mathbf{e}, \mathbf{v}) \sim (\mathbf{e} \cdot g, g^{-1} \cdot \mathbf{v}), \qquad \forall g \in G. \tag{12.2}$$

In words: A vector $v \in E$ may be be described as "n-tuple with respect to a frame" in many different ways. Without a preferred frame, v is simply the *equivalence class of any such description.*

Formally, this operation of taking equivalence classes is denoted

$$E = \mathrm{Fr}(E) \times_\rho \mathbb{K}^n := (\mathrm{Fr}(E) \times \mathbb{K}^n)/_{\sim \text{Eq. (12.2)}},$$

where $\rho : G \to \mathrm{GL}(n)$ makes explicit that G is to act on $\mathbb{K}^n$ in the defining representation. It should be clear that the equivalence classes inherit a well-defined vector space structure from $\mathbb{K}^n$.

12.2 Associated fibre bundles

Let $\pi : P \to X$ be a principal G-bundle over X. Let ρ be a left G-action on a manifold F (by diffeomorphisms). This action will be denoted

$$g \cdot \xi := \rho(g)(\xi), \qquad g \in G, \xi \in F,$$

with ρ suppressed unless emphasis is necessary.

Prompted by Eq. (12.2), we consider the set $P \times F$, equipped with the right G-action,

$$(p, \xi) \cdot g := (p \cdot g, g^{-1} \cdot \xi).$$

Then pass to the set of equivalence classes modulo this G-action,

$$P \times_\rho F := \{(p, \xi) \in P \times F \,:\, (p \cdot g, g^{-1} \cdot \xi)\}. \tag{12.3}$$

Notice that the projection map

$$\begin{aligned} \pi_\rho : P \times_\rho F &\to X \\ [p, \xi] &\mapsto \pi(p) \end{aligned}$$

is well-defined, since $\pi(p \cdot g) = \pi(p)$ for all $g \in G$. Give $P \times_\rho F$ the quotient topology, then $\pi_\rho : P \times_\rho F \to X$ is continuous (exercise).

Exercise 12.1. For $x \in X$, show that

$$\pi_\rho^{-1}(x) = \{[p, \xi] \,:\, \xi \in F\}$$

where p can be chosen to be any point in $P_x = \pi^{-1}(x)$.

Thus every fibre $\pi_\rho^{-1}(x)$ is in bijection with the *typical fibre* F. But such an identification requires choosing, non-canonically, a reference frame $p \in P_x$. Next, we sketch how these fibres assemble into a smooth "fibre bundle" (See Definition 46 later).

Let $x \in X$, and $\Phi : \pi^{-1}(U) \to U \times G$ be a local trivialization of P around x. Equivalently, Φ^{-1} corresponds to a local gauge

$$s_\Phi : U \to P, \qquad x \mapsto \Phi^{-1}(x, 1).$$

We pair up this local gauge with F-valued components to get a continuous "trivialization" of $\pi_\rho^{-1}(U)$,

$$\begin{aligned} (\Phi^{(\rho)})^{-1} : U \times F &\to \pi_\rho^{-1}(U) \\ (x, \xi) &\mapsto [s_\Phi(x), \xi] \end{aligned}$$

Since x was arbitrary, we see that $P \times_\rho F$ is built up from "locally trivial pieces" $\pi_\rho^{-1}(U) \cong U \times F$. As in the case of the tangent bundle construction, $P \times_\rho F$ is a topological manifold, and establish its smooth structure as follows.

From a trivializing cover $\{U_\alpha, \Phi_\alpha\}_{\alpha\in\mathcal{I}}$ of P, we obtain a corresponding continuous trivializing cover $\{U_\alpha, \Phi_\alpha^{(\rho)}\}_{\alpha\in\mathcal{I}}$ of $P\times_\rho F$. Each $U_\alpha \times F$ is a smooth manifold, so $(\Phi_\alpha^{(\rho)})^{-1}(U_\alpha \times F)$ inherits a smooth structure. Over $U_{\alpha\beta}$, the coordinate charts coming from $\Phi_\alpha^{(\rho)}$ and from $\Phi_\beta^{(\rho)}$ are compatible, due to the following.

Exercise 12.2. Let $(U_\alpha, \Phi_\alpha^{(\rho)})$, $(U_\beta, \Phi_\beta^{(\rho)})$ be two (continuous) local trivializations for $P \times_\rho F$. Show that on the overlap $U_{\alpha\beta} \times F$, we have

$$\begin{aligned} \Phi_\beta^{(\rho)} \circ (\Phi_\alpha^{(\rho)})^{-1} : U_{\alpha\beta} \times F &\to U_{\alpha\beta} \times F \\ (x, \xi) &\mapsto (x, \rho(g_{\beta\alpha}(x))(\xi)), \end{aligned} \tag{12.4}$$

where $g_{\beta\alpha} : U_{\alpha\beta} \to G$ are the smooth transition functions for P.

Definition 45. $\pi_\rho : P \times_\rho F \to X$, with the above smooth structure, is called the *fibre bundle associated with* $\pi : P \to X$*, via the action* ρ *on* F, or simply an *associated fibre bundle.*

Example 12.1. The frame bundle $\mathrm{Fr}(E)$ of a vector bundle E is a principal $\mathrm{GL}(n)$-bundle. Let ρ be the defining action of G as matrix multiplication on $\mathbb{K}^n$. Then $\mathrm{Fr}(E) \times_\rho \mathbb{K}^n$ is a vector bundle. It is isomorphic to E as follows: if $\mathbf{e}_x \in \mathrm{Fr}(E_x)$ is a frame at x, then $[\mathbf{e}_x, \xi] \in \mathrm{Fr}(E) \times_\rho \mathbb{K}^n$ is identified with $\mathbf{e}_x(\xi) \in E_x$.

In the associated fibre bundle construction, observe that the transition functions are G-valued, so not all diffeomorphisms of F will be used. So $P \times_\rho F$ is a "bundle of F" with extra structure:

Definition 46. A *fibre bundle* with *typical fibre* F on which the *structure Lie group* G acts (on the left by diffeomorphisms), is a smooth surjective map $\pi : E \to X$, such that every $x \in X$ lies in an open neighbourhood U with $E|_U := \pi^{-1}(U)$ locally trivializable: there is a diffeomorphism

$$\Phi : \pi^{-1}(U) \to U \times F \quad \text{such that} \quad \pi(\Phi^{-1}(x, \xi)) = x, \qquad x \in U,\, \xi \in F.$$

Furthermore, for each pair $(U_\alpha, \Phi_\alpha), (U_\beta, \Phi_\beta)$ of local trivializations, we have

$$\Phi_\beta \circ \Phi_\alpha^{-1}(x, \xi) = (x, g_{\beta\alpha}(x) \cdot \xi), \qquad x \in U_{\alpha\beta}, \xi \in F,$$

with smooth $g_{\beta\alpha} : U_{\alpha\beta} \to G$, called the G-valued *transition functions.*

Example 12.2. As we saw in Eq. (6.6), a principal G-bundle P is a fibre bundle with typical fibre G, and structure group G acting on the typical fibre G by left multiplication ("frame relabelling"). Beware that this left multiplication only makes sense in a local trivialization, and there is no global left G-action on P.

Example 12.3. A rank-n vector bundle is a fibre bundle, with structure group $\mathrm{GL}(n)$ acting on the typical fibre $\mathbb{K}^n$.

Remark. As usual, a fibre bundle is allowed the maximal compatible collection of local trivializations, in the sense of all G-valued transition functions remaining smooth. For example, the associated fibre bundle does not inherit preferred local trivializations from those used in its construction.

12.3 Operations on vector bundles

On vector spaces E, F, we have algebraic operations,

- Duals E^*;
- Direct sums;
- (Symmetrized/antisymmetrized) tensor products;
- $\mathrm{Hom}(E, F) \cong F \otimes E^*$;
- Complexification/realification;
- Complex conjugate.

These algebraic operations generalize to vector bundles E, F over X.

For example, for the direct sum $E \oplus F$ of a rank-m and a rank n-vector bundle, we would take each $(E \oplus F)_x$ to be $E_x \oplus F_x$. Then use local trivializations $E|_U \cong U \times \mathbb{K}^m$ and $F|_U \cong U \times \mathbb{K}^n$ to construct $(E \oplus F)|_U \cong U \times (\mathbb{K}^m \oplus \mathbb{K}^m)$, and topologize $E \oplus F$ as we did for the tangent bundle in Section 5.1. The resulting transition functions will be the direct product of those for E and those for F, so they are smooth. More abstractly, we can also construct $E \oplus F$ from the direct product $\mathrm{GL}(m) \times \mathrm{GL}(n)$-valued transition functions, by the gluing construction of Section 6.4.1.

Similarly, the cotangent bundle T^*X can be constructed by taking the transitions functions $(g_{\alpha\beta}^{-1})^{\mathrm{t}}$, where $g_{\alpha\beta}$ are the ones for TX. We can construct T^*X from the tangent frame bundle by taking $\mathrm{GL}(n)$ to act on $\mathbb{R}^n$ in

the contragedient representation $(g^{-1})^{\mathrm{t}}$. Similarly for tensor product bundles. In particular, any manifold X automatically comes with a bundle $\Lambda^k(X)$ of differential k-forms, and $\Omega^k(X) = \Gamma(\Lambda^k(X))$.

A common procedure is to tensor $\Lambda^k(X)$ with an auxiliary vector bundle E, where E has structure group $G' \subset \mathrm{GL}(\dim E)$ independent from that of TX. Elements of

$$\Gamma(\Lambda^k(X) \otimes E) = \Omega^k(X, E)$$

are called *1-forms twisted by E*.

12.3.1 Gauge transformation of vector bundles

Of particular importance is the so-called *endomorphism bundle* $\mathrm{End}(E) := E \otimes E^*$ of E. The terminology arises because sections $T \in \Gamma(\mathrm{End}(E))$ act on sections $v \in \Gamma(E)$ in the obvious way,

$$T \cdot v(x) = T(x)(v(x)),$$

and $T \cdot (fv) = fT(v)$, for all $f \in C^\infty(X)$. Elements of $\Gamma(\mathrm{End}(E))$ are called bundle endomorphisms. They can be added linearly, and be composed. There is an identity bundle endomorphism, and the invertible bundle endomorphisms form a group, denoted $\mathrm{GL}(E)$.

Now, if E has some extra structure (e.g. metric, orientation), and is regarded as a fibre bundle with reduced structure group $G \subset \mathrm{GL}(n)$, then we can restrict to those sections $T \in \Gamma(\mathrm{End}(E))$ for which each $T(x)$ acts on E_x as an element of G. Such a bundle endomorphism T is also called a *gauge transformation of E*, where E regarded as a vector bundle with structure group G. These restricted bundle endomorphisms form a subgroup, denoted $G(E)$.

Now let $F \in \mathcal{G}(P)$ be a gauge transformation of P, represented as right-multiplication by the equivariant map $\sigma_F : P \to G$. On any associated vector bundle, there is a corresponding gauge transformation. For $[p, \xi] \in E = P \times_\rho F$, define

$$F \cdot [p, \xi] := [p \cdot \sigma_F(p), \xi] = [p, \rho(\sigma_F(p)) \cdot \xi], \tag{12.5}$$

One checks that (exercise)

- Eq. (12.5) defines a bundle endomorphism of T_F of E as a vector bundle with structure group $\rho(G)$.

- If ρ is a faithful representation (i.e. injective), then $F \mapsto T_F$ is a group isomorphism $\mathcal{G}(P) \to G(E)$.

Question: When do we need to reduce the structure group of a vector bundle from $\mathrm{GL}(n)$ to some subgroup?

12.4 Vector bundle metrics

If E is a real vector bundle, then a *Euclidean bundle metric* is a section

$$\langle \cdot, \cdot \rangle_E \in \Gamma(E^* \otimes E^*)$$

which restricts at each $x \in X$ to an inner product $\langle \cdot, \cdot \rangle_x$ on E_x. Then E is called a *Euclidean vector bundle*. The inner product of two sections is then a real-valued function,

$$\langle \psi, \tilde{\psi} \rangle_E \in C^\infty(X).$$

For example, a Riemannian metric on X is a Euclidean bundle metric on TX.

If E is a complex vector bundle, then a *Hermitian bundle metric* is a section

$$\langle \cdot, \cdot \rangle \in \Gamma(\overline{E}^* \otimes E^*)$$

which restricts at each $x \in X$ to a Hermitian inner product on E_x. Then E is called a *Hermitian vector bundle*.

12.5 Restricted frame bundles

Let $\pi : E \to X$ be a rank-n vector bundle. The full frame bundle $\mathrm{Fr}(E)$ has the structure of a principal $\mathrm{GL}(n)$-bundle.

Metrics. Suppose E is a Euclidean vector bundle. Then we may restrict attention to orthonormal frames. Pointwise, orthonormal frames form an $\mathrm{O}(n)$-torsor, instead of a $\mathrm{GL}(n)$-torsor. If we only allow local trivializations for E such that $E_x \cong \mathbb{K}^n$ preserve inner products, then the transition functions will be $\mathrm{O}(n)$-valued. So the structure group of E gets reduced from $\mathrm{GL}(n)$ to $\mathrm{O}(n)$. The corresponding *orthonormal frame bundle* $\mathrm{Fr}^{\mathrm{O}}(E)$ is a principal $\mathrm{O}(n)$-bundle inside the full frame bundle.

Similarly, if E is a Hermitian vector bundle, the orthonormal (or unitary) frame bundle is a principal $\mathrm{U}(n)$-bundle, denoted $\mathrm{Fr}^{\mathrm{U}}(E)$.

Orientation. Two bases for a real vector space E are said to be equivalent if the change-of-basis matrix has positive/negative determinant. An *orientation* of E is a choice of equivalence classes of bases. The set of orientations is a $\mathbb{Z}_2$-torsor (there is no canonical "trivial" orientation).

For E a real vector bundle, we can assign a pointwise orientation to every E_x. A local frame $\{e_1, \ldots, e_n\}$ over $U \subset X$ is *positively oriented* if $\{e_1(x), \ldots, e_n(x)\}$ has the same orientation as the given pointwise orientation for all $x \in U$. The pointwise orientation is globally *smooth* if every point of X lies in the domain of an oriented local frame; in this case, E is said to be an *orientable* vector bundle, and a choice of smooth pointwise orientation is simply called an *orientation* on E.

Once E has been given an orientation, we can decide to use only the positively oriented local frames as the local trivializations. Then the transition functions will be $\mathrm{GL}(n, \mathbb{R})^+$-valued. We obtain the *oriented frame bundle* $\mathrm{Fr}^+(E)$ as a principal $\mathrm{GL}(n, \mathbb{R})^+$-bundle, with $\mathrm{GL}(n, \mathbb{R})^+$-valued transition functions.

If E has a bundle metric, there is a further reduction to (oriented) orthonormal frames, and we have the *(oriented) orthonormal frame bundle* $\mathrm{Fr}^{(\mathrm{S})\mathrm{O}}(E)$ as a principal $(\mathrm{S})\mathrm{O}(n)$-bundle.

Remark. If $E = TX$ is the tangent bundle of a (oriented, Riemannian) manifold X, then the above frame bundles are usually referred to the various frame bundles of X, with the vector bundle TX being implicit.

Spin? As we will learn later, there can be a *further* "reduction" from $\mathrm{SO}(n)$ to a so-called $\mathrm{Spin}(n)$ group. This does not naturally take place on a vector/tensor bundle over X. Instead, the natural starting point is the principal "spin frame bundle", to which the vector bundle of spinors is associated.

12.6 Reduction of structure group

Definition 47. Let $\phi : H \to G$ be a Lie group homomorphism, and $\pi' : P' \to X$ be a principal G-bundle. A map $F : P \to P'$, where $\pi : P \to X$ is a principal H-bundle, is called a *ϕ-reduction of P'* if

- $\pi' \circ F = \pi$,
- $F(p \cdot h) = F(p) \cdot \phi(h), \qquad h \in H, p \in P.$

Remark. The map F restricts to a map $P_x \to P'_x$ of fibres. Picking basepoints $p \in P_x$ and $F(p) \in P'_x$ gives identifications $P_x \cong H$ and $P'_x \cong G$. Then $F|_{P_x}$ is identified with the homomorphism $\phi : H \to G$.

An id_G-reduction is just a principal G-bundle isomorphism. When ϕ is the embedding of a Lie subgroup $H \subset G$, we simply call P a *H-reduction of P'*. The image $F(P)$ is then a principal H-subbundle inside P'.

Example 12.4. When an orientable real vector bundle E is given an orientation, the oriented frame bundle is a principal $\mathrm{GL}(n, \mathbb{R})^+$ subbundle of the full frame bundle. We say that the former is a $\mathrm{GL}(n, \mathbb{R})^+$-reduction of the frame bundle. Similarly, if a bundle metric is given, the oriented orthonormal frame bundle is a further $\mathrm{SO}(n)$-reduction of the frame bundle.

Let P be a principal G-bundle, and consider an associated vector bundle $E = P \times_\rho V$. Suppose V is equipped with a G-invariant inner product,

$$\langle \xi, \zeta \rangle_V = \langle \rho(g)\xi, \rho(g)\zeta \rangle_V, \qquad \xi, \zeta \in V,\ g \in G.$$

In other words, the representation ρ of G on the typical fibre V is unitary/orthogonal. Then the vector bundle $E = P \times_\rho V$ acquires the bundle metric

$$\langle [p, \xi], [p, \zeta] \rangle_E := \langle \xi, \zeta \rangle_V. \tag{12.6}$$

The structure group of E may then be reduced from $\mathrm{GL}(V)$ to $\rho(G) \subset \mathrm{U}(V)$ (or $\rho(G) \subset \mathrm{O}(V)$).

12.7 Scalar fields

What if the structure group of E is reduced to the trivial group? This means that the allowed trivializations have trivial transition functions. A little thought reveals that there is just a single global trivialization $E \cong X \times V$, with all others simply being the restriction to subsets $U \subset X$. Then sections of E are just V-valued functions that we were all familiar with before learning about bundles. So why describe functions in such a bundle-theoretic way?

In general, we really want to distinguish functions/scalar fields (as above) from, e.g., vector fields, k-form fields, spin-$\frac{1}{2}, 1, \frac{3}{2}, \ldots$ fields, etc. The point is that all of these fields are sections of vector bundles associated to a common frame bundle intrinsic to the (typically Riemannian) geometry of X. The distinguishing factor is the *representation* of the structure group G on the typical vector space.

In Riemannian geometry, we would take $\mathrm{Fr}^{\mathrm{SO}}(X)$, and $\mathrm{SO}(n)$ would be represented in the defining way, or in some tensor representation. In quantum theory, we have $\mathrm{Fr}^{\mathrm{Spin}}(X)$ (defined in Section 16.1), and for $\dim X = 3$, we have the unitary representations of Spin(3) labelled by half-integers. In relativistic (quantum) theory, the Lorentz-orthonormal frame bundle $\mathrm{Fr}^{\mathrm{SO}(1,3)}(X)$, or the $\mathrm{Spin}(1,3)$ version, is used. The various irreducible representations of $\mathrm{Spin}(1,3)$ give a notion of elementary particle fields over the semi-Riemannian spacetime X.

In any of the above situations, we may take G to be represented trivially on V. Then the associated vector bundle is a trivial vector bundle (not just trivializable). Its sections are *scalar fields*, which are now just functions since only one global trivialization is usable. In physics, one might refer to scalar fields as "spin-0" fields, and talk about "spinless particle" wavefunctions. So (relativistic) frame rotation has no effect on the values/components of a scalar field at all.

Indeed, Schrödinger's quantum wavefunction theory ($\sim$1926) was formulated before it was established that electrons have "spin angular momentum". So it described "spinless electrons" — such Schrödinger operators abound in physics models when electron spin effects can be ignored.

Later on, the geometric understanding of spin improved, triggered in part by Dirac's equation for the spin-ful electron field and its consistent but strange-looking transformation behaviour under (relativistic) rotations. Over time, it became appreciated that general quantum wavefunctions were not really scalar fields/functions but sections of vector bundles associated to certain non-trivial spin-representations. Then it becomes mandatory to introduce connections, or more precisely, the induced covariant differentiation on the quantum wave-sections.

13 Parallel transport and covariant differentiation

13.1 Parallel transport of frames

Definition 48. Let P be a principal G-bundle over X with a connection ω. Let $\gamma : [0,1] \to X$ be a curve in X. A curve $\tilde{\gamma} : [0,1] \to P$ is a *horizontal lift* of γ if $\pi \circ \tilde{\gamma} = \gamma$ and its velocity vectors $\tilde{\gamma}'(t)$ are horizontal for all $t \in [0,1]$.

Note: At an end point, the velocity vector is defined by the one-sided derivative.

Theorem 13.1. *Let $\pi : P \to X$ be a principal G-bundle with connection, and let $\gamma : [0,1] \to X$. For each $p \in \pi^{-1}(\gamma(0))$ in the initial fibre, there exists a unique* horizontal lift *$\tilde{\gamma}_p : [0,1] \to P$, satisfying $\tilde{\gamma}_p(0) = p$.*

Proof. [Optional.] First, assume that γ lies within U with $P|_U$ trivializable. Fix any section $s : U \to P$ with $s(\gamma(0)) = p$. This gives a reference lift $s \circ \gamma$ of γ. A general (non-horizontal) lift has the form $\tilde{\gamma}_p(t) = (s \circ \gamma)(t) \cdot g(t)$ for some smooth $g : [0,1] \to G$ with $g(0) = e$. We need to solve for the appropriate g such that $\tilde{\gamma}$ is *horizontal*.

For convenience, we leave the initial point p implicit. Writing $\sigma : P \times G \to P$ for the G-action, we consider $\tilde{\gamma}$ as a composition of maps,

$$\tilde{\gamma} : t \mapsto \sigma((s \circ \gamma)(t), g(t)).$$

As in Eq. (11.7) we have

$$\tilde{\gamma}'(t) = d(R_{g(t)})_{(s\circ\gamma)(t)}((s \circ \gamma)'(t)) + \left(g^*\Theta \left(\frac{d}{dt'}\Big|_{t'=t} \right) \right)^\sharp_{\tilde{\gamma}(t)},$$

and horizontal-ness means that it is annihilated by the connection 1-form ω,

$$\begin{aligned} 0 \overset{\text{need}}{=} \omega_{\tilde{\gamma}(t)}(\tilde{\gamma}'(t)) &= (R^*_{g(t)}\omega)_{(s\circ\gamma)(t)}((s \circ \gamma)'(t)) + g^*\Theta \left(\frac{d}{dt'}\Big|_{t'=t} \right) \\ &= \mathrm{Ad}_{g(t)^{-1}} \circ \omega((s \circ \gamma)(t)) + (L_{g(t)^{-1}})_*(g'(t)) \\ &= (L_{g(t)^{-1}})_*\big((R_{g(t)})_*\omega((s \circ \gamma)'(t)) + g'(t)\big). \end{aligned}$$

For convenience, we write $\beta : t \mapsto \omega((s \circ \gamma)'(t))$, which is a fixed curve in $\mathfrak{g} = T_e G$. The above vanishing condition now reads

$$(R_{g(t)})_*(\beta(t)) + g'(t) = 0, \tag{13.1}$$

which is an ODE for the function $g : [0,1] \to G$ with initial condition $g(0) = e$.

To solve this ODE, we define a vector field v on $G \times [0,1]$ by

$$v_{(h,t)} = \Big(- (R_h)_*(\beta(t)), \frac{\partial}{\partial t'}\Big|_{t'=t} \Big).$$

Then, by construction, an integral curve of v starting at $(e, 0)$ will have the form $t \mapsto (g_1(t), t)$, where g_1 satisfies the ODE, Eq. (13.1), and the initial condition. However, g_1 may only be defined for some small time interval $[0, \delta_1] \subset [0,1]$. So we consider another integral curve $t \mapsto (g_2(t), t + \delta_1)$ of v, starting from the point (e, δ_1); then $t \mapsto g_2(t - \delta_1)$ also also satisfies Eq. (13.1) for $t \in [\delta_1, \delta_2]$. We modify this second curve to $t \mapsto g_1(\delta_1) \cdot g_2(t - \delta_1)$, so that it matches up with g_1 at time $t = \delta_1$. Repeat this extension process until we get a solution to Eq. (13.1) for $t \in [0,1]$. Here, compactness of $\{e\} \times [0,1]$ ensures that there is a minimal $\delta > 0$ which works no matter which t we start an integral curve at, so the extension process terminates after a finite number of steps.

The above procedure uniquely produces a horizontal lift in the case where γ lies in one trivializing chart. In general, we can use compactness of the curve to argue that there is a finite subdivision of $[0,1]$ into subintervals, such that each subcurve lies within a single chart. Note that when switching chart during the extension process, we need to convert to a different section/trivialization, but the (partial) horizontal lift does not depend on the choice of section.

□

Remark. Why is lifting a curve γ in X harder than lifting a vector field? Could we not take the velocity vectors along γ, extend it arbitrary to a vector field on X, then use the horizontally lifted vector field on P? The issue is whether we can integrate the resulting vector field into a curve $\tilde{\gamma}_p$ above all of γ. This is what we had to prove above.

Theorem 13.1 allows us to make the following definition:

Definition 49. Let $\pi : P \to X$ be a principal G-bundle with a connection ω, and let $\gamma : [0,1] \to X$ be a curve. The *parallel transport along* γ with respect

to ω is the map

$$\begin{aligned} \tau_\gamma \equiv \tau_\gamma^\omega : P_{\gamma(0)} &\to P_{\gamma(1)} \\ p &\mapsto \tilde{\gamma}_p(1), \end{aligned}$$

where $\tilde{\gamma}_p$ is the unique horizontal lift of γ which starts at $p \in P_{\gamma(0)}$.

Parallel transport is G-equivariant:

Proposition 13.2. *Let ω be a connection on a principal G-bundle $P \to X$. Let τ_γ^ω be the parallel transport map along a curve γ. Then*

$$\tau_\gamma^\omega(p \cdot g) = (\tau_\gamma^\omega(p)) \cdot g, \qquad p \in P_{\gamma(0)},\ g \in G.$$

Proof. Exercise. □

Exercise 13.1. Show that $\tau_\gamma \equiv \tau_\gamma^\omega$ has the following properties:

- It is independent of the parametrization of the curve γ.
- For any path γ obtained as the smooth concatenation of two paths $\gamma_1, \gamma_2 : [0,1] \to X$, i.e.,

  $$\gamma(t) = \begin{cases} \gamma_1(2t), & t \in [0, \frac{1}{2}], \\ \gamma_2(2t-1), & t \in [\frac{1}{2}, 1], \end{cases}$$

  the equation $\tau_\gamma = \tau_{\gamma_2} \circ \tau_{\gamma_1}$ holds.
- For the flipped path $\gamma_{\text{flip}} : t \mapsto \gamma(1-t)$, the parallel transport satisfies $\tau_{\gamma_{\text{flip}}} = \tau_\gamma^{-1}$.

Parallel transport also transforms nicely under gauge transformations:

Proposition 13.3. *Let $F \in \mathcal{G}(P)$ be a gauge transformation, and let $F^*\omega$ be the transformed connection (Eq.* (11.14)*). Then*

$$\tau_\gamma^{F^*\omega} = F^{-1} \circ \tau_\gamma^\omega \circ F.$$

Proof. Let $p \in P$ be given. Write $q = F(p) = p \cdot \sigma_F(p)$, and let $\tilde{\gamma}_p$ be the ω-horizontal lift of γ starting at q. Our candidate $(F^*\omega)$-horizontal lift is

$$\Gamma = F^{-1} \circ \tilde{\gamma}_q.$$

We check that Γ is indeed $(F^*\omega)$-horizontal,

$$(F^*\omega)_{\Gamma(t)}(\Gamma'(0)) = \omega_{\tilde{\gamma}_q(t)}\left(dF_{\Gamma(t)} \circ dF^{-1}_{\tilde{\gamma}_q(t)}(\tilde{\gamma}_q'(t))\right) = \omega_{\tilde{\gamma}_q(t)}(\tilde{\gamma}_q(t)) = 0,$$

where the last equality follows from $\tilde{\gamma}_q$ being ω-horizontal. The initial point of Γ is

$$\Gamma(0) = \tilde{\gamma}_q(0) \cdot \sigma_F^{-1}(\tilde{\gamma}_q(0)) = p \cdot \sigma_F(p) \cdot \underbrace{\sigma_F^{-1}(p \cdot F(p))}_{\sigma_F^{-1}(p)} = p.$$

Thus

$$\tau_\gamma^{F^*\omega}(p) = \Gamma(1) = F^{-1} \circ \tilde{\gamma}_q(1) = F^{-1} \circ \tau_\gamma^\omega \circ F(p).$$

□

13.2 Holonomy of parallel transport

Let us restrict attention to *loops* in X, based at a fixed point x, i.e., $\gamma(1) = \gamma(0) = x$. Then the parallel transport map along γ becomes a self-map of the fibre at x,

$$\tau_\gamma^\omega : P_x \to P_x,$$

called the *holonomy of ω around γ*. It is convenient to think of the holonomy around γ as being implemented by right-multiplication by some map $g_\gamma : P_x \to G$,

$$\tau_\gamma^\omega(p) = p \cdot g_\gamma(p).$$

By Prop. 13.2,

$$\tau_\gamma^\omega(p \cdot g) = \tau_\gamma^\omega(p) \cdot g = p \cdot g_\gamma(p) \cdot g = (p \cdot g) \cdot \underbrace{(g^{-1} g_\gamma(p) g)}_{g_\gamma(p \cdot g)}.$$

Thus, the holonomy map gets conjugated by g^{-1} when we adjust the fibre basepoint by g.

Next, consider loops as maps $\gamma : S^1 \to X$ without a preferred basepoint. The holonomy around the loop will depend on the choice of start/end point

for the loop (as well as the initial fibre point above it). But the ambiguity is just a conjugation by the parallel transport along the part of the loop joining the two choices of start/end points. So, up to conjugacy, the holonomy of ω around a loop γ is well-defined without reference to basepoints.

Definition 50. Fix a point $p \in P_x$. Given a connection ω, its *holonomy group* at p is defined as

$$\mathrm{Hol}_p(\omega) := \{g \in G \,:\, \exists \text{ loop } \gamma \text{ based at } x \text{ with } \tau_\gamma^\omega(p) = p \cdot g\}$$

It follows from Exercise 13.1 that $\mathrm{Hol}_p(\omega)$ is indeed a group, and it is easy to see that

$$\mathrm{Hol}_{p\cdot g}(\omega) = g^{-1}\mathrm{Hol}_p(\omega)g, \qquad g \in G.$$

Prop. 13.3 shows that the conjugacy class of the holonomy along a loop is invariant under gauge transformations of the connection. To utilize this, we pass to a representation $\rho : G \to \mathrm{GL}(n)$, then we can take the trace of $\rho(g_\gamma(p))$ to get a *gauge-invariant* of the loop γ (any fibre point p above any loop point x can be used). Such a quantity is called a *Wilson loop* in physics, and sometimes simply called "the holonomy". Note that if G is abelian, then the conjugations are trivial, and the holonomy itself (not just the conjugacy class) is gauge-invariant.

13.3 Covariant derivatives on associated vector bundles

We have seen that a connection ω on a principal G-bundle $P \to X$ defines parallel transport in P along curves in X. This automatically defines a corresponding parallel transport in *any* associated vector bundle.

Definition 51. Let $P \to X$ be a principal G-bundle with connection ω, and $E = P \times_\rho V$ be an associated vector bundle. Given a curve $\gamma : [0,1] \to X$, the map

$$\begin{aligned} \tau_\gamma^{E,\omega} : E_{\gamma(0)} &\to E_{\gamma(1)} \\ [p,\xi] &\mapsto [\tau_\gamma^\omega(p), \xi] \end{aligned}$$

is called the *parallel transport in E, along γ, with respect to ω.*

It is easy to see that $\tau_\gamma^{E,\omega}$ is well-defined. Suppose $[p,\xi] = [p',\xi']$ in E, so

$$[p',\xi'] = [p \cdot g, g^{-1}\xi]$$

for some $g \in G$. Then

$$[\tau_\gamma^\omega(p'), \xi'] = [\tau_\gamma^\omega(p \cdot g), g^{-1} \cdot \xi] \overset{\text{Prop. 13.2}}{=} [\tau_\gamma^\omega(p) \cdot g, g^{-1} \cdot \xi] = [\tau_\gamma^\omega(p), \xi].$$

Basically, once we know how to parallel transport the frames, then the components ξ just go along for the "parallel ride".

Notice that $\tau_\gamma^{E,\omega}$ provides a linear identification of the initial fibre $E_{\gamma(0)}$ with the final fibre $E_{\gamma(1)}$. *Now we are finally ready to differentiate sections of E.* Let $v_x \in T_xX$ be a tangent vector to the base manifold at some point $x \in X$. Let $\gamma : (-\epsilon, \epsilon) \to X$ be any curve such that $\gamma(0) = x$ and $\gamma'(0) = v_x$. For $t \in (-\epsilon, \epsilon)$, we write γ_t for the curve γ restricted to the time interval $[0, t]$.

Definition 52. Let $P \to X$ be principal G-bundle with connection ω, and let E be an associated vector bundle. The *covariant (directional) derivative* of a section $\psi \in \Gamma(E)$ along a tangent vector $v_x \in T_xX$ is

$$\nabla_{v_x}^\omega \psi := \frac{d}{dt}\Big|_{t=0} (\tau_{\gamma_t}^{E,\omega})^{-1}(\psi(\gamma(t))) \ \in E_x, \tag{13.2}$$

where γ is any representative curve for v_x.

In words: we parallel transport all the $\psi(\gamma(t))$ back to the same fibre E_x, then take the usual t-derivative. The key feature of the covariant derivative is its manifest *gauge-independence* — we do not have to pick any gauge to convert the section ψ into a V-valued function.

Let us check that Definition 52 only depends on the velocity vector v_x, not the choice of representative curve γ. Pick any local gauge $s : U \to P$, and write

$$\psi(x') = [s(x'), \xi(x')], \qquad x' \in U,$$

for some function $\xi : U \to F$ representing ψ. The local gauge gives a reference curve $s \circ \gamma$ in P, which can be parallel transported to a curve r in the initial fibre P_x,

$$r(t) := (\tau_{\gamma_t}^\omega)^{-1}((s \circ \gamma)(t)) \in P_x.$$

The curve r has the form $r(t) = s(x) \cdot g(t)$ for a unique curve $g \equiv g(t)$ in G

satisfying $g(0) = e$. With ψ locally represented as $[s, \xi]$, we rewrite

$$\begin{aligned}
(\tau^{E,\omega}_{\gamma_t})^{-1}(\psi(\gamma(t))) &= (\tau^{E,\omega}_{\gamma_t})^{-1}[(s \circ \gamma)(t), \xi(\gamma(t))] \\
&= (\tau^{E,\omega}_{\gamma_t})^{-1}[\tau^{\omega}_{\gamma_t}(r(t)), \xi(\gamma(t))] \\
&= [r(t), \xi(\gamma(t))] && \text{(Defn. 51)} \\
&= [s(x) \cdot g(t), \xi(\gamma(t))] \\
&= [s(x), \rho(g(t)) \cdot \xi(\gamma(t))],
\end{aligned}$$

where $\rho : G \to \mathrm{GL}(n)$ is the representation defining the associated bundle E. Taking the t-derivative at $t = 0$,

$$\nabla^{\omega}_{v_x}\psi = [s(x), \rho(e) \cdot d\xi_x(v_x) + d\rho_e(\overbrace{g'(0)}^{\in T_eG = \mathfrak{g}}) \cdot \xi(x)]. \tag{13.3}$$

(Here, for the V-valued function ξ, the directional derivative $d\xi_x(v_x)$ is obtained as $\frac{d\xi(\gamma(t))}{dt}|_{t=0}$ with γ a curve representing v_x.)

We now calculate

$$\begin{aligned}
g'(0) \overset{\text{connection}}{=} \omega((s(x) \cdot g)'(0)) &= \omega\left(\overbrace{\frac{d}{dt}\Big|_{t=0} \tau^{\omega}_{\gamma_t}(s(x))}^{\text{horizontal}} + \frac{d}{dt}\Big|_{t=0}(s(x) \cdot g(t))\right) \\
&= \omega\left(\frac{d}{dt}\Big|_{t=0} \tau^{\omega}_{\gamma_t}(s(x)) \cdot g(t)\right) \\
&= \omega\left(\frac{d}{dt}\Big|_{t=0} \tau^{\omega}_{\gamma_t}(s(x) \cdot g(t))\right) \\
&= \omega((s \circ \gamma)'(0)) \\
&= s^*\omega(v_x).
\end{aligned}$$

So Eq. (13.3) becomes the formula

$$\nabla^{\omega}_{v_x}\psi = [s(x), d\xi_x(v_x) + d\rho_e(s^*\omega(v_x)) \cdot \xi(x)],$$

which does not refer to the curve γ.

It also makes sense to write

$$(\nabla^{\omega}_{v}\psi)(x) := \nabla^{\omega}_{v_x}\psi, \qquad \psi \in \Gamma(E), v \in \mathfrak{X}(X),$$

because the dependence on v_x is smooth. Thus each vector field $v \in \mathfrak{X}(X)$ determines a map of sections of E,

$$\nabla^\omega_v : \Gamma(E) \to \Gamma(E).$$

We record the above calculations:

Proposition 13.4. *With respect to a local gauge $s : U \to P$, the* covariant derivative ∇^ω_v *along a vector field $v \in \mathfrak{X}(X)$ has the formula*

$$\nabla^\omega_v \psi = [s, d\xi(v) + d\rho_e(s^*\omega(v)) \cdot \xi], \qquad \psi \in \Gamma(E), \tag{13.4}$$

where ξ is the V-valued component function of ψ with respect to the gauge s.

For the first term in Eq. (13.4), $d\xi$ is a V-valued 1-form, so $d\xi(v)$ is the ordinary directional derivative of the V-valued function ξ along v. The local gauge potential appears as an extra second term $s^*\omega(v)$ acting linearly on ξ.

Finally, it is customary to consider $v \in \mathfrak{X}(X)$ itself as an input variable for the covariant derivative:

Definition 53. Let $P \to X$ be a principal G-bundle with connection ω, and $E = P \times_\rho V$ be an associated vector bundle. With respect to ω, the *covariant derivative* is the map $\nabla^\omega : \Gamma(E) \to \Omega^1(X, E)$ defined by

$$(\nabla^\omega \psi(v))(x) \equiv (\nabla^\omega_v \psi)(x) := \nabla^\omega_{v_x} \psi, \qquad \psi \in \Gamma(E), v \in \mathfrak{X}(X).$$

Proposition 13.5. *Let $P \to X$ be a principal G-bundle with connection ω. On any associated vector bundle E (with fibre a vector space over $\mathbb{K}$), the covariant derivative ∇^ω has the following properties.*
For all $\psi \in \Gamma(E)$, $u, v \in \mathfrak{X}(X)$, $f \in C^\infty(X)$, $h \in C^\infty(X, \mathbb{K})$,

- ∇^ω *is $\mathbb{K}$-linear in ψ;*
- $\nabla^\omega_{u+fv}\psi = \nabla^\omega_u \psi + f \cdot \nabla^\omega_v \psi$;
- $\nabla^\omega_v(h \cdot \psi) = v(h) \cdot \psi + h \cdot \nabla^\omega_v \psi$.

Proof. Exercise. □

Note: When $\mathbb{K} = \mathbb{C}$, the tangent vector field v is a derivation of the function algebra $C^\infty(X, \mathbb{K})$ in the same way as in the $\mathbb{K} = \mathbb{R}$ case).

From covariant derivative to parallel transport. Sometimes, one begins with a vector bundle $E \to X$, and *defines* a *covariant derivative on* E to be a map $\nabla : \Gamma(E) \to \Omega^1(X, E)$ satisfying the algebraic properties listed in Prop. 13.5. Such a ∇ axiomatically allows differentiation of sections in a gauge-independent way. The notion of path-lifting, parallel transport, and holonomy also follows from such a starting point.

However, the information of the structure group $G \subset \mathrm{GL}(n)$ (e.g. due to a bundle metric) is not yet involved. If we want the parallel transport/covariant differentiation to preserve this extra structure, then extra conditions must be imposed on ∇. Some examples are discussed in Section 13.6-13.7.

13.4 Physics notation

Set $\eta_a = (0, \ldots, 0, \underbrace{1}_{a\text{-th}}, 0, \ldots, 0)$ where $a = 1, \ldots, \mathrm{rank}(E)$. So the $e_a := [s, \eta_a]$ provide a basis of local sections of E, whose component functions are constant (with respect to the local gauge s). The covariant derivative of e_a simplifies to

$$\nabla^\omega_v e_a = [s, (d\rho)_e(s^*\omega(v)) \cdot \eta_a].$$

Suppose the local gauge is defined on a coordinate chart U (restrict the domain otherwise). Then we could choose $v = \partial_i$ to be a local coordinate vector field, and obtain

$$\nabla^\omega_i e_a := \nabla^\omega_{\partial_i} e_a = [s, (d\rho)_e(s^*\omega(\partial_i)) \cdot \eta_a]$$

Recall the physicists' notation for the local gauge potential $s^*\omega = \mathcal{A}_j dx^j$, where $\mathcal{A}_j = s^*\omega(\partial_j)$ is a $\mathfrak{g}$-valued function over U. To simplify notation, one suppresses the Lie algebra representation $d\rho_e$ on V, so

$$\nabla^\omega_j e_a = [s, \mathcal{A}_j \cdot \eta_a] = [s, \mathcal{A}^b{}_{ja}\eta_b] = \sum_{b=1}^{\mathrm{rank}\, E} \mathcal{A}^b{}_{ja} e_b$$

for the *connection (matrix) coefficients* $\mathcal{A}^b{}_{ja}$.

A general local section of E is expanded as $\psi = \psi^a e_a$. Its j-th covariant derivative is, by the third Leibniz property in Prop. 13.5,

$$\begin{aligned} \nabla^\omega_j \psi = \nabla^\omega_j(\psi^a e_a) &= (\partial_j \psi^a) e_a + \psi^a \nabla^\omega_j e_a \\ &\text{“}=\text{”}\ ((\partial_j + \mathcal{A}_j)\psi)^a e_a \end{aligned}$$

In the last line, one pretends that ψ is a V-valued function $(\psi^1, \ldots, \psi^{\operatorname{rank} E})$ on which the matrix $\mathcal{A}_j = \mathcal{A}^b{}_{ja}$ acts. Finally, reference to ω and the local frame $\{e_1, \ldots, e_{\operatorname{rank} E}\}$ is usually suppressed, and there is usually an i discrepancy between physicists' and mathematicians' matrix Lie algebra conventions, so

$$\text{“ } \nabla_j \;=\; \partial_j - i\mathcal{A}_j \text{ ”} \quad \text{acting on “}V\text{-valued functions”} \;\; \psi.$$

$G = \mathrm{U}(1)$ **Abelian example.** Let P be a principal $\mathrm{U}(1)$-bundle over X with connection ω, and $\rho : \mathrm{U}(1) \to \mathrm{U}(1)$ be the defining representation on $\mathbb{C}$. So $E = P \times_\rho \mathbb{C}$ is a Hermitian line bundle over X. Under $d\rho_e$, we have $\mathfrak{u}(1) = i\mathbb{R}$ acting on $\mathbb{C}$ by multiplication.

So the local gauge potential is represented on E as $d\rho_e(s^*\omega) = -i \sum_{j=1}^n \mathcal{A}_j \, dx^j$ for some functions $\mathcal{A}_j : X \to \mathbb{R}$. Then the j-th "momentum operator" becomes

$$-i\nabla_j = -i\partial_j - \mathcal{A}_j$$

Now generalize the representation to $\rho^{(q)} : \mathrm{U}(1) \to \mathrm{U}(1)$, $e^{i\varphi} \mapsto e^{iq\varphi}$, where q must be some integer, called the (electric) charge. Then $d\rho_e^{(q)} = q \cdot 1_{\mathbb{C}}$, and we will arrive at the formula

$$\text{“ } \nabla_j \psi = (\partial_j - iq\mathcal{A}_j)\psi \text{ ”}.$$

Note that if $\mathrm{U}(1)$ is replaced by the noncompact Lie group $(\mathbb{R}, +)$, which has the same Lie algebra as $\mathrm{U}(1)$, then $\rho_q : \mathbb{R} \to \mathrm{U}(1)$, $\lambda \mapsto e^{iq\lambda}$, could be used as a unitary representation of the Lie group, whether or not q is integral. So it is actually important to specify the Lie group and its representation, and not just the Lie algebra.

We now understand that ψ is actually a section of a line bundle. The $\partial_j - iqA_j$ is just a particular local-gauge-dependent way of writing the covariant derivative of sections.

The local gauge potential $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$ is also called a magnetic vector potential. That is, after using the Riemannian metric to convert $\mathcal{A}$ to a local vector field, the curl of $\mathcal{A}$ is the magnetic vector field that we are familiar with. Actually, in three spatial dimensions, the magnetic field is better understood as a differential 2-form $\mathcal{F}$ using Hodge duality (induced by the Riemannian metric). Then the relationship between the vector potential and the magnetic field is simply $d\mathcal{A} = \mathcal{F}$, and we have seen that the latter is independent of gauge because $G = \mathrm{U}(1)$ is abelian.

Classically, $\mathcal{A}$, or even the gauge-independent connection ω, is not detectable. This is because it is genuinely a redundant mathematical device for efficiently solving Maxwell's equations for the classical electromagnetic fields. However, we learned from the Aharonov–Bohm effect that $\mathcal{A}$ (or ω) does measurably affect the quantum mechanical evolution of a quantum wave-section. This is because ω is playing a fundamental role in the Schrödinger equation of motion, by specifying what it means to take derivatives. Nevertheless, only the gauge-invariant *holonomy* of ω is actually measured, in accordance with the gauge principle.

13.5 Curvature of covariant derivative

Definition 54. Let ∇ be a covariant derivative on a vector bundle $E \to X$. The *curvature* F^∇ of ∇ is defined by the equation

$$F^\nabla(u,v)\psi = \nabla_u\nabla_v\psi - \nabla_v\nabla_u\psi - \nabla_{[u,v]}\psi, \qquad u,v \in \mathfrak{X}(X),\ \psi \in \Gamma(E).$$

It may be verified that F^∇ is an element of $\Omega^2(X, \mathrm{End}(E))$ (Exercise). In particular, it is $C^\infty(X)$-linear in its arguments.

Exercise 13.2. Let Ω be the curvature of a connection ω on a principal G-bundle, E be an associated vector bundle, and $\nabla = \nabla^\omega$ be the induced covariant derivative on E. In a local gauge s over U, the curvature of ∇ is related to the local field strength $s^*\Omega$ by the formula

$$F^\nabla(u,v)\psi = [s, d\rho_e(s^*\Omega(u,v))(\xi)], \qquad u,v \in \mathfrak{X}(X), \psi \in \Gamma(E). \tag{13.5}$$

13.6 Metric connections

Let E be a vector bundle. A connection ω on its frame bundle is sometimes called a *linear connection*, and it gives rise to a covariant derivative ∇^ω on E.

Suppose E is a Euclidean vector bundle. By definition, a *metric connection* on E is a connection ω on the orthonormal frame bundle $\mathrm{Fr}^{\mathrm{O}}(E)$. Recall that $E \cong \mathrm{Fr}^{\mathrm{O}}(E) \times_{\rho_{\mathrm{std}}} \mathbb{R}^n$ where ρ_{std} is the defining representation of $\mathrm{O}(n)$ on $\mathbb{R}^n$. Furthermore, as in Eq. (12.6), the bundle metric is recovered through the formula

$$\langle [p,\xi], [p,\zeta] \rangle_E = \langle \xi, \zeta \rangle_{\mathbb{R}^n}.$$

The metric connection defines parallel transport of *orthonormal frames*, so the induced parallel transport on E will preserve inner products.

The same is true for Hermitian vector bundles, with a metric connection being a connection on $\mathrm{Fr}^{\mathrm{U}}(E)$. In this case, the terminology *unitary connection* is often adopted.

Proposition 13.6. *Let E be a Euclidean or Hermitian vector bundle over X. If ω is a metric connection, then the covariant derivative ∇^ω satisfies*

$$v(\langle\psi,\varphi\rangle_E) = \langle\nabla^\omega_v\psi,\varphi\rangle_E + \langle\psi,\nabla^\omega_v\varphi\rangle_E, \qquad \psi,\varphi\in\Gamma(E),\ v\in\mathfrak{X}(X). \tag{13.6}$$

Proof. At each x, we calculate in a local gauge s, replacing ψ,φ by their n-component functions ξ,ζ. Using Eq. (13.4), we have

$$\langle\nabla^\omega_v\psi,\varphi\rangle_E + \langle\psi,\nabla^\omega_v\varphi\rangle_E = \langle d\xi(v) + s^*\omega(v)\cdot\xi,\zeta\rangle_{\mathbb{K}^n} + \langle\xi, d\zeta(v) + s^*\omega(v)\cdot\zeta\rangle_{\mathbb{K}^n}.$$

Here $s^*\omega(v)\in\mathfrak{o}(n)$ is represented as a skew-symmetric matrix function acting on $\mathbb{R}^n$ (or skew-Hermitian matrix function on $\mathbb{C}^n$ in the complex case). So

$$\langle s^*\omega(v)\cdot\xi,\zeta\rangle_{\mathbb{K}^n} + \langle\xi, s^*\omega(v)\cdot\zeta\rangle_{\mathbb{K}^n} = 0.$$

Now we are left with

$$\begin{aligned}\langle\nabla^\omega_v\psi,\varphi\rangle_E + \langle\psi,\nabla^\omega_v\varphi\rangle_E &= \langle d\xi(v),\zeta\rangle_{\mathbb{K}^n} + \langle\xi, d\zeta(v)\rangle_{\mathbb{K}^n}\\ &= \sum_{i=1}^n (d\overline{\xi^i}(v)\zeta^i + \overline{\xi^i}d\zeta^i(v))\\ &= \sum_{i=1}^n v(\overline{\xi^i}\zeta^i)\\ &= v(\langle\xi,\zeta\rangle_{\mathbb{K}^n}) = v(\langle\psi,\varphi\rangle_E).\end{aligned}$$

□

Remark. If we start with a covariant derivative ∇ on a Euclidean/Hermitian vector bundle E, it is said to be *metric-compatible* if Eq. (13.6) is satisfied.

13.7 Levi-Civita connection of Riemannian manifold

As a standard important example, take the tangent bundle $E = TX$ of a Riemannian manifold X. The sections ψ,φ of E are tangent vector fields, and the Riemannian metric is a bundle metric, usually denoted $g(\cdot,\cdot)$ (Apologies

for overusing the symbol g). Suppressing the reference to ω, Eq. (13.6) for a metric connection on TX becomes the condition

$$u(g(v,w)) = g(\nabla_u v, w) + g(v, \nabla_u w), \qquad u, v, w \in \mathfrak{X}(X). \tag{13.7}$$

Furthermore, it makes sense to ask for an extra *torison-free* condition,

$$\nabla_u v - \nabla_v u - [u, v] = 0, \qquad u, v \in \mathfrak{X}(X). \tag{13.8}$$

A fundamental result in Riemannian geometry states that there is a unique covariant derivative on TX, called the *Levi–Civita connection* ∇^{LC}, which is metric compatible and torsion-free, i.e., Eq. (13.7)-(13.8) are satisfied. You may find a proof in any Riemannian geometry textbook. The *covariant derivative* on a Riemannian manifold (X, g) means ∇^{LC}, unless otherwise stated.

13.7.1 Levi–Civita connection on submanifold

Write $\partial_i, i = 1, \ldots, N$ for the standard global coordinate vector fields on the standard Riemannian manifold $\mathbb{R}^N$. The Levi–Civita connection on $\mathbb{R}^N$ is the trivial connection,

$$\nabla_u^{\mathrm{triv}}(v) = \sum_{i=1}^{N} u(v^i)\partial_i, \qquad u, v \in \mathfrak{X}(\mathbb{R}^N).$$

Let X be a submanifold of $\mathbb{R}^N$, and equip it with the restricted Riemannian metric. For each $x \in X$, there is an orthogonal projection $p_x : T_x\mathbb{R}^N = \mathbb{R}^N \to T_xX$. Globally, define the map

$$p : X \times \mathbb{R}^N \to TX, \qquad (x, \xi) \mapsto (x, p_x(\xi)).$$

A vector field v on the submanifold X can still be expanded as

$$v = \sum_{i=1}^{N} v^i \partial_i, \qquad v^i \in C^\infty(X).$$

For $u, v \in \mathfrak{X}(X)$, define the connection $\nabla : \mathfrak{X}(X) \to \Omega^1(X) \otimes \mathfrak{X}(X)$ to be

$$\nabla_u(v) = p\left(\sum_{i=1}^{N} u(v^i)\partial_i\right).$$

It can be checked that the above ∇ is the Levi–Civita connection for X (Exercise).

13.7.2 Levi–Civita connection on orthonormal frame bundle

The Levi–Civita ∇^{LC} can be viewed as the covariant derivative on TX obtained from some connection ω^{LC} on the *orthonormal* frame bundle $\mathrm{Fr}^{\mathrm{O}}(X)$.

Let us work in a local *orthonormal* frame field $\mathbf{e} = \{e_i\}_{i=1,\ldots,n}$ over $U \subset X$. Thus $\mathbf{e}$ is a local section of $\mathrm{Fr}^{\mathrm{O}}(X)$ over U. (In the physics literature, $\mathbf{e}$ is called a *tetrad* or *vielbein*, sometimes written $e_i^{\ \mu}\,\partial_\mu$). So each e_i is a normalized local section of TX, and its covariant derivative along any $v \in \mathfrak{X}(X)$ admits an expansion

$$\nabla^{\mathrm{LC}}_v e_i = \sum_{j=1}^{n} \omega_{ij}(v) e_j,$$

with the ω_{ij} being local 1-forms on U given by the formula

$$\omega_{ij}(\cdot) := g(\nabla^{\mathrm{LC}}_{(\cdot)} e_i, e_j). \tag{13.9}$$

Because ∇^{LC} is metric-compatible, Eq. (13.6) of Prop. 13.6 says that

$$\begin{aligned} 0 = v(\underbrace{g(e_i, e_j)}_{\delta_{ij}}) &= g(\nabla^{\mathrm{LC}}_v e_i, e_j) + g(e_i, \nabla^{\mathrm{LC}}_v e_j) \\ &= \omega_{ij}(v) + \omega_{ji}(v). \end{aligned}$$

Thus the local 1-forms ω_{ij} are antisymmetric in the indices i, j. In fact, recalling that $\mathfrak{so}(n)$ are the antisymmetric matrices, these ω_{ij} are the matrix entries of the local gauge potential $\mathbf{e}^*\omega^{\mathrm{LC}} \in \Omega^1(U, \mathfrak{so}(n))$, with respect to the local orthonormal frame field $\mathbf{e}$. For historical reasons, ω_{ij} is sometimes called the "spin connection", but we shall reserve this terminology for a more precise object later on.

Lemma 13.7. *At any point x of a Riemannian manifold, there exists a local orthonormal tangent frame field $\mathbf{e} = \{e_1, \ldots, e_n\}$, such that $\nabla^{\mathrm{LC}} e_i(x) = 0$ for all $i = 1, \ldots, n$.*

Proof. Take an open subset U around x, on which the orthonormal frame bundle is trivializable. It may be assumed that U is a coordinate chart. Use the spherical polar coordinates to obtain a radial curve joining x to x' for any given $x' \in U$. Pick any orthonormal frame $\mathbf{e}_x = \{e_{1,x}, \cdots, e_{n,x}\}$ at x, and parallel transport it to x' along the radial curve. This gives an orthonormal frame field $\mathbf{e} = \{e_1, \ldots, e_n\}$ over U. For any $k = 1, \ldots, n$, there is a radial curve

with initial tangent vector being $e_{k,x}$. Because the $e_{i,x'}$ are parallel transported from x along such radial curves, we have

$$\nabla^{\mathrm{LC}}_{e_k} e_i(x) = \nabla^{\mathrm{LC}}_{e_{k,x}} e_i = 0, \qquad \forall i, k = 1, \ldots, n.$$

by the parallel transport definition of ∇^{LC}, Eq. (13.2). Then $\nabla^{\mathrm{LC}} e_i = 0$ holds at x. □

Remark. Note that the vanishing of $\nabla^{\mathrm{LC}} e_i$ is only guaranteed at x, not at any other point $x' \in U$. Furthermore, the frame field $\mathbf{e}$ will not generally comprise coordinate tangent vector fields for any choice of coordinates on U. Rather, $\mathbf{e}$ is a convenient device for certain calculations that we will do later.

13.8 Berry connection

Consider a trivialized bundle of finite-dimensional Hilbert spaces, $X \times \mathbb{C}^N$. This has a trivial (unitary) connection ∇^{triv}, given by regarding a section as an N-component function, and taking the usual derivatives of each component along tangent vectors (e.g. partial derivatives ∂_i, if coordinates on X are chosen). Let $\iota : E \hookrightarrow X \times \mathbb{C}^N$ be the inclusion of a Hermitian subbundle E of rank $n < N$. So there is a smooth family p of projection matrices, $p(x) = p(x)^* = p(x)^2$, such that each E_x is the range of $p(x)$. The *Berry connection on the subbundle* E is

$$\nabla^{\mathrm{Berry}} = p \circ \nabla^{\mathrm{triv}} \circ \iota. \tag{13.10}$$

Exercise 13.3. Check that ∇^{Berry} defines a covariant derivative on E, and is metric-compatible in the sense of Eq. (13.6).

Remark. The Berry connection occurs very frequently in quantum mechanics. One has a *fixed* ambient finite-dimensional Hilbert space $\mathbb{C}^N$, and a control parameter space X for Hamiltonian operators $H(x)$. One may follow the lowest n-eigenvalues of $H(x)$ as x is varied. Assuming these low eigenvalues never cross the $(n+1)$-th eigenvalue, their eigenspaces define an eigenbundle $E \subset X \times \mathbb{C}^N$. This E is a Hermitian vector bundle with structure group $\mathrm{U}(n)$ (no orthonormal eigenbases are preferred) and has a Berry connection, Eq. (13.10).

Remark (Optional). It is possible to generalize the Berry connection to ambient bundles with infinite-dimensional Hilbert space fibres. This is routinely done

in Floquet–Bloch theory in solid-state physics. An important subtlety, however, is that there is no canonical trivialization on the ambient Hilbert space bundle, and therefore no canonical "Berry connection" on an eigen-subbundle. Rather, there is a *family* of Berry connections, parametrized by choices of origin (implicit in defining the Floquet–Bloch transform), as explained in the note [arXiv:1706.01149]. Therefore, certain apparently gauge-invariant quantities of a single "Berry connection", e.g. holonomy/geometric phase, may not actually be physical. More precisely, such quantities usually only make sense *relative* to some origin (e.g. polarizations).

14 Clifford algebras

Motivation. The Laplace operator on $\mathbb{C}$-valued functions on $\mathbf{R}^n$ is $\Delta = \sum_{j=1}^n (-i\partial_j)^2$. The operator $\sum_{j=1}^n -i\partial_j$ looks like a first-order square root of Δ, but we quickly find that we cannot get rid of the mixed derivatives $\partial_j\partial_k$.

Classical differential equations did not provide demand for such a square root. In *relativistic* quantum theory, a wave equation needs to be first-order in both time and space coordinates. In 1928, Dirac proposed a completely new class of differential operators to address this difficulty. His original construction involved $\mathbb{C}^4$-valued "spinors" rather than ordinary functions.

The basic idea is already apparent from the Dirac operator on Euclidean $\mathbf{R}^2$, which can be written as

$$D = -i \begin{pmatrix} 0 & \partial_x - i\partial_y \\ \partial_x + i\partial_y & 0 \end{pmatrix} \equiv -i \begin{pmatrix} 0 & \partial \\ \bar{\partial} & 0 \end{pmatrix}, \tag{14.1}$$

and does satisfy $D^2 = \Delta \otimes \mathbf{1}_2$. By using a *matrix* version of partial derivatives, we have made the mixed partial derivatives disappear.

More invariantly, the $e_{i,x} = \partial_i|_x$ are orthonormal basis vectors for the tangent spaces T_xX at each x (inner product from Riemannian metric). We simply imposed the formal algebraic condition

$$e_{i,x}e_{j,x} = -e_{j,x}e_{i,x}, \tag{14.2}$$

turning T_xX into a *noncommutative algebra*.

If one only ever works in flat Euclidean space, then this sort of algebraic construction is more or less sufficient. Under the influence of general relativity, in the late 1920s to early 1930s, physicists and mathematicians debated on how to make sense of spinors and Dirac operators on curved spaces. The problem, even locally, is very subtle, particularly the correct notion of parallel transport/covariant derivative of spinor fields, as compared to ordinary vector fields. Furthermore, global well-definedness of spinors is not automatic. We will eventually learn about the intrinsic geometric meaning of spinors and Dirac operators, using the differential geometric tools introduced earlier. But before that, we must first become familiar with the pointwise, algebraic story, which is already not trivial.

14.1 Algebra jargon

We restrict to $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

- A (unital) $\mathbb{K}$-algebra is a $\mathbb{K}$-vector space $\mathscr{A}$ with an extra bilinear associative product $\mathscr{A} \times \mathscr{A} \to \mathscr{A}$ (and a unit element $1_{\mathscr{A}}$ for the product). So, for $a, b, c \in \mathscr{A}$ and $\lambda \in \mathbb{K}$, we have e.g.,

$$(\lambda a + b)c = \lambda(ac) + bc.$$

- A (unital) morphism of algebras is a linear map which respects the product (and identity elements). A (unital) *subalgebra* of $\mathscr{A}$ is a subspace of $\mathscr{A}$ which is closed under the product (and contains the unit).

- A *left ideal* in $\mathscr{A}$ is a subspace $\mathfrak{I} \subset \mathscr{A}$ which absorbs left multiplication,

$$a \in \mathscr{A}, b \in \mathfrak{I} \;\Rightarrow\; ab \in \mathfrak{I}.$$

  Similarly for *right ideals*. A (two-sided) *ideal* is simultaneously a left ideal and a right ideal. The algebra structure descends to the quotient space $\mathcal{A}/\mathfrak{I}$.

- A *quadratic vector space* (V, q) is a finite-dimensional $\mathbb{K}$-vector space V equipped with a quadratic form $q : V \to \mathbb{K}$, i.e.,

$$q(\lambda v) = \lambda^2 q(v), \qquad \lambda \in \mathbb{K}, v \in V,$$

  with $q(u + v) - q(u) - q(v)$ bilinear in u, v. The associated symmetric bilinear form (also denoted q) is

$$q(u, v) := \frac{1}{2}(q(u + v) - q(u) - q(v)), \qquad u, v \in V, \tag{14.3}$$

  and satisfies $q(v, v) = q(v)$. For most of our purposes, it suffices to think of $q(\cdot, \cdot)$ as a real inner product and $q(\cdot)$ as the squared-norm.

- The *tensor algebra* of V is

$$\mathcal{T}(V) = \bigoplus_{k=0}^{\infty} V^{\otimes k} = \mathbb{K} \oplus V \oplus (V \otimes V) \oplus \dots$$

  Let $\mathfrak{I}_0$ be the ideal generated by the elements $v \otimes v$, where $v \in V$. The *exterior algebra* of V is the quotient algebra

$$\Lambda(V) := \mathcal{T}(V)/\mathfrak{I}_0.$$

Note that $u \otimes v + v \otimes u = (u+v) \otimes (u+v) - u \otimes u - v \otimes v$, so symmetric 2-tensors will automatically vanish in $\Lambda(V)$. A little thought shows that the grading on $\mathcal{T}(V)$ (given by the number of tensor factors) remains well-defined on $\Lambda(V)$, i.e.,

$$\Lambda(V) = \bigoplus_{k=0}^{\infty} \Lambda^k(V)$$

The tensor product descends from $\mathcal{T}(V)$ to $\Lambda(V)$, whence it is denoted $\omega \wedge \eta$. One may check that if $\omega \in \Lambda^k(V), \eta \in \Lambda^l(V)$, then

$$\omega \wedge \eta \in \Lambda^{k+l}(V), \qquad \omega \wedge \eta = (-1)^{kl} \eta \wedge \omega.$$

We have $\Lambda^0(V) = \mathbb{K}$ and $\Lambda^1(V) = V$. If $\{\mathsf{e}_i\}_{i=1,\ldots,n}$ is a basis for V, then

$$\{\mathsf{e}_{i_1} \wedge \ldots \mathsf{e}_{i_k} \,:\, 1 \leq i_1 < \ldots < i_k \leq n\}$$

is a basis for $\Lambda^k(V)$, so that $\dim \Lambda^k(V) = \binom{n}{k}$. In particular, $\Lambda^n(V)$ is one-dimensional, and there are no higher exterior powers. So $\Lambda(V)$ has vector space dimension 2^n.

Note: For $V = T_x^* X$, we recover the exterior algebra of differential forms at x.

14.2 Clifford algebra definition

Definition 55. Let (V, q) be a quadratic space, and $\mathcal{J}_q$ be the ideal in $\mathcal{T}(V)$ generated by elements of the form $v \otimes v + q(v)1$. The *Clifford algebra* is the quotient

$$Cl(V, q) := \mathcal{T}(V)/\mathfrak{J}_q.$$

In the Clifford algebra, we will generally write uv, v^2, instead of $u \otimes v, v \otimes v$. There is an inclusion $\iota_V : V \hookrightarrow \mathcal{T}(V) \to Cl(V, q)$ (Exercise). For $q \equiv 0$, we recover the exterior algebra, in which the square of a vector vanishes, $v \wedge v = 0$. For general $q \not\equiv 0$, we have $v^2 = -q(v)$ being a non-zero scalar in $Cl(V, q)$.

In terms of the symmetric bilinear form $q(\cdot, \cdot)$, the vector elements of $Cl(V, q)$ satisfy the anticommutation rule

$$\begin{aligned} uv + vu = (u+v)^2 - u^2 - v^2 \quad &= \quad -q(u+v) + q(u)^2 + q(v)^2 \\ &\overset{\text{Eq. (14.3)}}{=} -2q(u, v), \qquad u, v \in V. \end{aligned} \tag{14.4}$$

Proposition 14.1. *$Cl(V,q)$ is the unique $\mathbb{K}$-algebra with the property: given a linear map $f : V \to \mathscr{A}$ into an associative $\mathbb{K}$-algebra $\mathcal{A}$ satisfying $f(v)^2 = -q(v)1_{\mathscr{A}}$ for all $v \in V$, there exists a unique extension of f to an algebra homomorphism, $\hat{f} : Cl(V,q) \to \mathscr{A}$. Diagrammatically:*

$$\begin{tikzcd} V \arrow[r, "\iota_V"] \arrow[dr, "f"'] & Cl(V,q) \arrow[d, "\exists!\,\hat{f}"] \\ & \mathscr{A} \end{tikzcd} .$$

Sketch. There is an obvious extension of f to the tensor algebra,

$$F : \mathcal{T}(V) \to \mathscr{A}, \qquad v_1 \otimes \cdots \otimes v_n \mapsto f(v_1) \otimes \cdots \otimes f(v_n).$$

In particular, $F(v \otimes v + q(v)) = f(v)^2 + q(v)1_{\mathscr{A}} = 0$ for any $v \in V$, so F annihilates the ideal $\mathfrak{I}_q$. So F descends to a homomorphism

$$\hat{f} : \mathcal{T}(V)/\mathfrak{I}_q \equiv Cl(V,q) \to \mathscr{A},$$

with $\otimes$ replaced by $\wedge$. Proof of uniqueness of $\hat{f}$ is omitted. $\square$

Suppose we have quadratic form preserving linear maps $(V,q) \xrightarrow{f} (V',q') \xrightarrow{g} (V'',q'')$. Then Proposition 14.1 implies that we obtain a commuting diagram

$$\begin{tikzcd} V \arrow[r, "\iota_V"] \arrow[d, "f"'] \arrow[dd, bend right=60, "g\circ f"'] & Cl(V,q) \arrow[d, "\widehat{\iota_{V'}\circ f}"] \arrow[dd, bend left=60, "\widehat{\iota_{V''}\circ g\circ f}"] \\ V' \arrow[r, "\iota_{V'}"] \arrow[d, "g"'] & Cl(V',q') \arrow[d, "\widehat{\iota_{V''}\circ g}"] \\ V'' \arrow[r, "\iota_{V''}"] & Cl(V'',q'') \end{tikzcd} .$$

Therefore the promotion of (V,q) to $Cl(V,q)$ is functorial — q-preserving linear maps may be consistently upgraded into Clifford algebra homomorphisms. In particular, orthogonal transformations of (V,q) induce *Bogoliubov automorphisms* of $Cl(V,q)$.

14.2.1 Clifford superalgebra

Beware that the $\mathbb{Z}$-grading no longer makes sense on $Cl(V,q)$. For example, the 2-tensor $v \otimes v$ and the 0-tensor $-q(v)$ represent the same element $vv = -q(v)$ in the Clifford algebra. Nevertheless, $Cl(V,q)$ does retain an important $\mathbb{Z}_2$-grading, as we now explain.

Proposition 14.1 applies, in particular, to the inversion map $\alpha : v \mapsto -v$. The induced *parity automorphism* of $Cl(V,q)$, again denoted α, satisfies $\alpha^2 = \mathrm{id}$. So we have a linear splitting

$$Cl(V,q) = Cl^0(V,q) \oplus Cl^1(V,q),$$

according to the $+1$ ("even parity") or -1 ("odd parity") eigenspace of α. Since α is a homomorphism, it follows that

$$Cl^i(V,q) \cdot Cl^j(V,q) \subset Cl^{i+j \bmod 2}(V,q).$$

This exhibits $Cl(V,q)$ as a $\mathbb{Z}_2$-*graded algebra*, or *superalgebra*. We often require $Cl(V,q)$ to be represented as linear operators on vector space W, such that:

- $W = W^0 \oplus W^1$ is $\mathbb{Z}_2$-graded;
- $Cl^0(V,q)$ acts as grading-preserving ("even"/"diagonal") operators, while $Cl^1(V,q)$ acts as grading-reversing ("odd"/"off-diagonal") operators.

14.2.2 Linear basis for $Cl(V,q)$

Let (V,q) be a real quadratic space, with non-degenerate q, i.e., $q(v) = 0 \Rightarrow v = 0$. Sylvester's law of inertia says that there exists some basis $\{\mathsf{e}_i\}_{i=1,\ldots,n}$ of V such that

$$q(v) = \sum_{i=1}^{n} \lambda_i (v^i)^2, \qquad \lambda_i = \pm 1,$$

where v^i are the components of v in that basis. Furthermore, the number of i with $\lambda_i = -1$ is independent of the basis. Such a basis is *orthonormal* with respect to q, in the sense that

$$q(\mathsf{e}_i) = \pm 1, \qquad q(\mathsf{e}_i, \mathsf{e}_j) = 0, \quad i \neq j,$$

and provides a very convenient set of generators for the Clifford algebra. For instance, by Eq. (14.4), the basis elements mutually anticommute,

$$\mathsf{e}_i \mathsf{e}_j + \mathsf{e}_j \mathsf{e}_i = 0, \qquad i \neq j,$$

cf. the motivating Eq. (14.2).

Multi-index notation: I denotes a sequence of indices $1 \leq i_1 < \ldots < i_k \leq n$ with length $|I| = k$, and

$$\mathsf{e}_I := \mathsf{e}_{i_1} \ldots \mathsf{e}_{i_k} \in Cl(V,q), \qquad \mathsf{e}_\emptyset = 1.$$

Proposition 14.2. *$Cl(V,q)$ has dimension $\sum_{k=0}^n \binom{n}{k} = 2^n$, with a basis given by e_I where I runs over all multi-indices. Each of $Cl^0(V,q)$ and $Cl^1(V,q)$ has dimension 2^{n-1}, except for the case $n = 0$, where $Cl(V,q) = Cl^0(V,q) = \mathbb{K}$.*

Proof. Exercise. □

Remark. Using Prop. 14.2, we deduce that there is a *linear* identification $Cl(V,q) \to \Lambda^\bullet(V)$, by converting e_I into $\mathsf{e}_{i_1} \wedge \ldots \wedge \mathsf{e}_{i_k}$. However, this is not an *algebra* isomorphism, since e.g., $\mathsf{e}_1\mathsf{e}_1 = -q(\mathsf{e}_1) = -1$ whereas $\mathsf{e}_1 \wedge \mathsf{e}_1 = 0$.

Exercise 14.1. Let $\eta \in Cl^0(V,q)$, so

$$\eta = \sum_{|I| \text{ even}} \eta^I \mathsf{e}_I,$$

with respect to some orthonormal basis $\{\mathsf{e}_1, \ldots, \mathsf{e}_n\}$ of (V,q). Check that the "vacuum coefficient" $\eta^\emptyset$ does not depend on the choice of orthonormal basis.

14.2.3 Complexification

The complex field $\mathbb{C}$ can also be viewed as a real 2-dimensional vector space.

Given a real vector space V, the process of extending scalar multiplication from $\lambda \in \mathbb{R}$ to $\lambda \in \mathbb{C}$ is called *complexification*, and is formally written

$$V^{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}.$$

On the above tensor product of $\mathbb{R}$-vector spaces, we can scalar multiply by complex numbers,

$$(v\lambda \otimes \mu)\nu = (v \otimes \lambda\mu)\nu = v \otimes \lambda\mu\nu, \qquad v \in V, \lambda \in \mathbb{R},\ \mu, \nu \in \mathbb{C},$$

so we have produced a $\mathbb{C}$-vector space. We will just write $v \otimes \mu \equiv v\mu$.

Given a real quadratic space (V,q), the quadratic form extends to a quadratic form $q^{\mathbb{C}}$ on $V^{\mathbb{C}}$, by taking $q^{\mathbb{C}}(v\mu) = q(v)\mu^2 \in \mathbb{C}$. So $(V^{\mathbb{C}}, q^{\mathbb{C}})$ is a complex quadratic space. Similarly, the Clifford algebra can be complexified, $Cl(V,q) \otimes_{\mathbb{R}} \mathbb{C}$, with the algebra product extended complex bilinearly, e.g.,

$$(u \otimes \mu)(v \otimes \nu) = uv \otimes \mu\nu.$$

It may be shown, using Prop. 14.1, that there is a $\mathbb{C}$-algebra isomorphism

$$\begin{aligned} Cl(V^{\mathbb{C}}, q^{\mathbb{C}}) &\cong Cl(V,q) \otimes_{\mathbb{R}} \mathbb{C} \\ (u\mu)(v\nu) &\leftrightarrow (uv) \otimes \mu\nu, \qquad u, v \in V,\ \mu, \nu \in \mathbb{C}. \end{aligned}$$

14.3 Clifford algebras $Cl_n, \mathbb{C}l_n$

14.3.1 Real Clifford algebras

Let us restrict attention to real (V, q) with *positive-definite* q. This means that (V, q) is a real inner product space. A choice of orthonormal basis $\{\mathsf{e}_i\}_{i=1,\dots,n}$ for V identifies (V, q) with $(\mathbb{R}^n, q_{\text{standard}})$, and induces an isomorphism of the Clifford algebras. So it suffies to study

$$Cl_n := Cl(\mathbb{R}^n, q_{\text{standard}}).$$

$\underline{Cl_0}$. We have $Cl_0 = \mathbb{R}$ as a real algebra.

$\underline{Cl_1}$. A linear basis for Cl_1 is $\{1, \mathsf{e}_1\}$, where $\mathsf{e}_1^2 = -1$. That is, e_1 is a square root of -1. So $Cl_1 \cong \mathbb{C}$ regarded as an $\mathbb{R}$-algebra. We could represent this on the $\mathbb{Z}_2$-graded vector space $\mathbb{R} \oplus \mathbb{R}$, with $\mathsf{e}_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

$\underline{Cl_2}$. A linear basis for Cl_2 is $\{1, \mathsf{e}_1\mathsf{e}_2, \mathsf{e}_1, \mathsf{e}_2\}$, where $\mathsf{e}_1, \mathsf{e}_2$ are odd, anticommuting square roots of -1. The even element $\mathsf{e}_1\mathsf{e}_2$ is also a square root of -1, anticommuting with e_1 and e_2. So Cl_2 is isomorphic to the quaternions $\mathbb{H} = \operatorname{span}_{\mathbb{R}}\{1, I, J, K\}$ as $\mathbb{R}$-algebras.

Exercise 14.2. Show that there is an (ungraded) algebra isomorphism $Cl^0_{n+1} \cong Cl_n$. For instance, if $\{\mathsf{e}_i\}_{i=1,\dots,n+1}$ is an orthonormal basis for $\mathbb{R}^{n+1}$, consider the map

$$f : \mathbb{R}^n \to Cl^0_{n+1}, \qquad \mathsf{e}_i \mapsto \mathsf{e}_{n+1}\mathsf{e}_i, \qquad i = 1, \dots, n.$$

14.3.2 Complex Clifford algebras

We write $\mathbb{C}l_n := Cl_n \otimes_{\mathbb{R}} \mathbb{C}$.

Volume/chirality element. Suppose $\mathbb{R}^n$ is *oriented*, and let $\{\mathsf{e}_1, \dots, \mathsf{e}_n\}$ be a positively-oriented orthonormal basis. The *volume/chirality element* of $\mathbb{C}l_n$ is

$$\omega^{\mathbb{C}} := i^{\lfloor \frac{n+1}{2} \rfloor} \mathsf{e}_1 \cdots \mathsf{e}_n.$$

Exercise 14.3. Check that the chirality element $\omega^{\mathbb{C}}$ has the following properties:

- It is independent of the choice of oriented orthonormal basis.
- $(\omega^{\mathbb{C}})^2 = 1$.
- When n is odd, $\omega^{\mathbb{C}}$ is central in $\mathbb{C}l_n$ (commutes with everything). When n is even, only $\mathbb{C}l_n^0$ commutes with $\omega^{\mathbb{C}}$.

The chirality element provides two complementary idempotents,

$$\pi^{\pm} = \frac{1 \pm \omega^{\mathbb{C}}}{2}, \qquad \pi^+\pi^- = \pi^-\pi^+ = 0.$$

Chiral subalgebras, n odd case. The idempotents $\pi^{\pm}$ commute with all of $\mathbb{C}l_n$. Accordingly, we get a splitting of $\mathbb{C}l_n$ into *chiral subalgebras*,

$$\mathbb{C}l_n = \mathbb{C}l_n^+ \oplus \mathbb{C}l_n^-, \qquad \mathbb{C}l_n^{\pm} := \pi^{\pm}\mathbb{C}l_n = \mathbb{C}l_n\pi^{\pm}.$$

Furthermore, $\alpha(\omega^{\mathbb{C}}) = -\omega^{\mathbb{C}}$ in this case, so we have a "chirality swapping" isomorphism,

$$\alpha : \mathbb{C}l_n^{\pm} \to \mathbb{C}l_n^{\mp}. \tag{14.5}$$

<u>$\mathbb{C}l_0$</u>**.** This is $\mathbb{C}$ as a $\mathbb{C}$-algebra.

<u>$\mathbb{C}l_1$</u>**.** We have $\mathbb{C}l_1 = Cl_1 \otimes_{\mathbb{R}} \mathbb{C}$. There is an isomorphism of *ungraded* $\mathbb{C}$-algebras,

$$\mathbb{C} \oplus \mathbb{C} \to \mathbb{C}l_1, \qquad \begin{cases} (1,0) \mapsto \frac{1\otimes 1 + \mathsf{e}_1 \otimes i}{2} = \pi^+, \\ (0,1) \mapsto \frac{1\otimes 1 - \mathsf{e}_1 \otimes i}{2} = \pi^-. \end{cases}$$

Recall that $i\mathsf{e}_1$ is the chirality element. So the factors in the decomposition $\mathbb{C} \oplus \mathbb{C} \cong \mathbb{C}l_1$ are the chiral subalgebras $\mathbb{C}l_1^+ = \pi^+\mathbb{C}l_1$ and $\mathbb{C}l_1^- = \pi^-\mathbb{C}l_1$ respectively. Note that

$$\mathsf{e}_1(\pi^{\pm}u) = \frac{\mathsf{e}_1 \mp i}{2}u = \mp i\frac{1 \pm i\mathsf{e}_1}{2}u = \mp i(\pi^{\pm}u), \qquad u \in \mathbb{C}l_1.$$

So e_1 is identified with $\mp i$ when $\mathbb{C}l_1^{\pm}$ is identified with $\mathbb{C}$.

There are two inequivalent 1-dimensional (complex) representations of $\mathbb{C}l_1$, depending on which chiral subalgebra is represented faithfully, and which one is represented trivially.

<u>$\mathbb{C}l_2$</u>**.** We can represent $\mathbb{C}l_2$ on the vector space $\mathbb{C}^2$ via

$$\mathsf{e}_1 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \qquad \mathsf{e}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Then $i\mathsf{e}_1\mathsf{e}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the chirality element, giving the representation space a grading, $\mathbb{C}^2 = \mathbb{C} \oplus \mathbb{C}$, with respect to which $\mathsf{e}_1, \mathsf{e}_2$ are odd operators. We also see that $\{1, \mathsf{e}_1\mathsf{e}_2, \mathsf{e}_1, \mathsf{e}_2\}$ span $\mathrm{M}_2(\mathbb{C})$. So $\mathbb{C}l_2 \cong \mathrm{M}_2(\mathbb{C})$.

We do not need to work out the rest of $\mathbb{C}l_n$ separately, due to the following "algebraic Bott periodicity".

Proposition 14.3. *For each $n \geq 0$, there is a isomorphism of $\mathbb{Z}_2$-graded algebras $\mathbb{C}l_{n+2} \cong \mathbb{C}l_n \otimes \mathbb{C}l_2 \cong \mathbb{C}l_n \otimes \mathrm{M}_2(\mathbb{C})$.*

Proof. In $\mathbb{C}l_n$, we have odd, mutually anticommuting elements $\mathsf{e}_i, i = 1, \ldots, n$, each squaring to -1. Similarly, in $\mathbb{C}l_2 \cong \mathrm{M}_2(\mathbb{C})$, we have $\tilde{\mathsf{e}}_1, \tilde{\mathsf{e}}_2$. In $\mathbb{C}l_{n+2}$ we have odd, mutually anticommuting $\varepsilon_i, i = i, \ldots, n+2$, each squaring to -1. Then, for example, the correspondence

$$\begin{aligned} \varepsilon_i &\leftrightarrow \mathsf{e}_i \otimes \sqrt{-1}\tilde{\mathsf{e}}_1\tilde{\mathsf{e}}_2, & i &= 1, \ldots, n, \\ \varepsilon_{n+j} &\leftrightarrow 1 \otimes \tilde{\mathsf{e}}_j & j &= 1, 2, \end{aligned}$$

gives an identification $\mathbb{C}l_{n+2} \cong \mathbb{C}l_n \otimes \mathbb{C}l_2$. □

Remark. There are other algebraic periodicity results for the real Clifford algebras, including those with mixed-signature quadratic forms q. Details may be found in, e.g. §1.4 of [10], and are important when dealing with special involutive operations like charge-conjugation, time-reversal etc.

15 Spin groups and representations

15.1 Spin groups

Let $Cl_n^\times$ denote the multiplicative group of invertible elements in Cl_n. This contains, in particular, all the nonzero vectors, since

$$v \cdot \frac{v}{-q(v)} = 1, \qquad v \in \mathbb{R}^n \setminus \{0\}.$$

Definition 56. The *twisted adjoint representation* of the group $Cl_n^\times$ on Cl_n is

$$\begin{aligned} \widetilde{\mathrm{Ad}} : Cl_n^\times &\to \mathrm{GL}(Cl_n) \\ y &\mapsto \widetilde{\mathrm{Ad}}_y, \qquad \widetilde{\mathrm{Ad}}_y(\cdot) := \alpha(y)(\cdot)y^{-1}, \end{aligned}$$

where α is the parity automorphism.

The case where $y \in \mathbb{R}^n \setminus \{0\}$ is instructive.

Lemma 15.1. *For $y \in \mathbb{R}^n \setminus \{0\} \subset Cl_n^\times$, the action $\widetilde{\mathrm{Ad}}_y$ restricts to the vector part of Cl_n,*

$$\widetilde{\mathrm{Ad}}_y(v) = v - 2\frac{\langle y|v\rangle}{||y||^2}y, \qquad v \in \mathbb{R}^n \subset Cl_n.$$

Thus $\widetilde{\mathrm{Ad}}_y$ reflects vectors through the hyperplane orthogonal to y.

Proof. For $y \in \mathbb{R}^n \setminus \{0\}$, we have $\alpha(y) = -y$, $y^{-1} = \frac{y}{-q(y)} = -\frac{y}{||y||^2}$, as well as

$$yv + vy = -2\langle y|v\rangle, \qquad v \in \mathbb{R}^n \subset Cl_n.$$

Then

$$\begin{aligned} ||y||^2\widetilde{\mathrm{Ad}}_y(v) = yv(y^{-1}(-||y||^2)) &= yvy \\ &= y(-yv - 2\langle y|v\rangle) = ||y||^2 v - 2\langle y|v\rangle y. \end{aligned}$$

□

In Lemma 15.1, scaling y by any non-zero number leads to the same reflection. Also, a classical result of Cartan–Dieudonné says that any element of $\mathrm{O}(n)$ is a product of reflections-through-hyperplanes. These lead us to consider the following subgroups of $Cl_n^\times$.

Definition 57. The *(S)pin group* is the subgroup of $Cl_n^\times$ generated by the unit vectors in $\mathbb{R}^n$,

$$\mathrm{Pin}(n) = \{y_1 \cdots y_k \,:\, k \in \mathbb{N},\ y_i \in S^{n-1} \subset \mathbb{R}^n\},$$

$$\mathrm{Spin}(n) = \mathrm{Pin}(n) \cap Cl_n^0 = \{y_1 \cdots y_{2k} \,:\, k \in \mathbb{N},\ y_i \in S^{n-1} \subset \mathbb{R}^n\}.$$

By construction, there is an orthogonal representation of $\mathrm{Pin}(n)$,

$$\begin{aligned} \widetilde{\mathrm{Ad}} : \mathrm{Pin}(n) &\to \mathrm{O}(n) \\ y_1 \cdots y_k &\mapsto \widetilde{\mathrm{Ad}}_{y_1} \cdots \widetilde{\mathrm{Ad}}_{y_k}. \end{aligned}$$

This restricts to a *special* orthogonal representation of $\mathrm{Spin}(n)$. All elements of $\mathrm{Spin}(n)$ are even-parity, so $\widetilde{\mathrm{Ad}}$ reduces to the usual adjoint/conjugation action $\mathrm{Ad}_\varphi(\cdot) = \varphi(\cdot)\varphi^{-1}$ for $\varphi \in \mathrm{Spin}(n)$. Thus we have a homomorphism

$$\begin{aligned} \chi = \widetilde{\mathrm{Ad}} : \mathrm{Spin}(n) &\to \mathrm{SO}(n) \\ y_1 \cdots y_{2k} &\mapsto \widetilde{\mathrm{Ad}}_{y_1} \cdots \widetilde{\mathrm{Ad}}_{y_{2k}} = \mathrm{Ad}_{y_1 y_2} \cdots \mathrm{Ad}_{y_{2k-1} y_{2k}}. \end{aligned}$$

Proposition 15.2. *For $n \geq 1$, there are short exact sequences of Lie groups*

$$\begin{aligned} 0 \to \{\pm 1\} \to \mathrm{Spin}(n) &\overset{\chi = \widetilde{\mathrm{Ad}}}{\to} \mathrm{SO}(n) \to 0, \\ 0 \to \{\pm 1\} \to \mathrm{Pin}(n) &\overset{\widetilde{\mathrm{Ad}}}{\to} \mathrm{O}(n) \to 0. \end{aligned}$$

Proof. Let $\varphi \in \ker \widetilde{\mathrm{Ad}} \subset \mathrm{Pin}(n)$, so $\alpha(\varphi) v \varphi^{-1} = v$ for all $v \in \mathbb{R}^n$. Decompose $\varphi = \varphi_0 + \varphi_1$ into its even and odd parts, so

$$\varphi_0 v = v \varphi_0, \qquad \varphi_1 v = -v \varphi_1, \qquad \forall v \in V. \tag{15.1}$$

Let $\{\mathsf{e}_i\}_{i=1,\ldots,n}$ be the standard orthonormal basis for $\mathbb{R}^n$. Then φ_0 can be expanded as a sum of products of even numbers of the e_i. Whenever e_1 appears in one of these products, it can be successively moved to the front of the product, at the expense of introducing signs and/or lowering the degree by 2 (if e_1 appears twice). So $\varphi_0 = a_0 + \mathsf{e}_1 a_1$ for some even-parity a_0 and some odd-parity a_1, neither of which contains any e_1 factor. Putting $\varphi_0 = a_0 + \mathsf{e}_1 a_1$ and $v = \mathsf{e}_1$ into Eq. (15.1) gives

$$\mathsf{e}_1 a_0 + \mathsf{e}_1^2 a_1 = \mathsf{e}_1 \varphi_0 = \varphi_0 \mathsf{e}_1 = a_0 \mathsf{e}_1 + \mathsf{e}_1 a_1 \mathsf{e}_1 = \mathsf{e}_1 a_0 - \mathsf{e}_1^2 a_1 \ \Rightarrow\ a_1 = 0.$$

Therefore $\varphi_0 = a_0$ and does not contain any e_1 term. Inductively, we deduce that φ_0 does not contain $\mathsf{e}_2, \ldots, \mathsf{e}_n$, and is simply a scalar. A similar argument shows that φ_1 does not contain any $\mathsf{e}_1, \ldots, \mathsf{e}_n$ term, so $\varphi_1 = 0$.

Therefore, φ can only be a scalar, which we know is obtained as a product of unit vectors, $\varphi = y_1 \cdots y_k$. Now, there is a transposition map $(\cdot)^{\mathrm{t}} : Cl_n \to Cl_n$, given by reversing the order of products — this is well-defined because the ideal $\mathfrak{J}_q$ is preserved under reversal of tensor product order. So consider

$$\varphi^2 = \varphi^{\mathrm{t}}\varphi = y_k \cdots y_1 y_1 \cdots y_k = (-1)^k q(y_1) \cdots q(y_k) = \pm 1.$$

Thus $\varphi = \pm 1$, and -1 is attained since $-1 = e_1^2 \in \mathrm{Spin}(n) \subset \mathrm{Pin}(n)$.

Surjectivity of $\widetilde{\mathrm{Ad}}$ uses the Cartan–Dieudonné result. Then (S)pin(n) is a closed subgroup of $Cl_n^\times$. It may be shown that $Cl_n^\times$ is open in the vector space Cl_n (general finite-dimensional algebra result), so it is a Lie group, and therefore, so is (S)pin(n). □

Proposition 15.3. *For $n \geq 2$, the spin group* $\mathrm{Spin}(n)$ *is connected.*

Proof. Each element of $\mathrm{Spin}(n)$ is connected to either $+1$ or -1 (because $\mathrm{SO}(n)$ is connected; proof omitted). So we just need to check that 1 and -1 are connected within $\mathrm{Spin}(n)$. For orthonormal $\mathsf{e}_1, \mathsf{e}_2$, consider

$$\gamma(t) = -\cos(\pi t) - \sin(\pi t)\mathsf{e}_1\mathsf{e}_2, \qquad t \in [0, 1].$$

This is a path in Cl_n joining $\gamma(0) = -1$ to $\gamma(1) = 1$. We can write

$$\gamma(t) = (\cos(\pi t/2)\mathsf{e}_1 + \sin(\pi t/2)\mathsf{e}_2)\,(\cos(\pi t/2)\mathsf{e}_1 - \sin(\pi t/2)\mathsf{e}_2)\,,$$

which is a product of two unit vectors, so $\gamma(t) \in \mathrm{Spin}(n)$. □

Remark. For $n \geq 3$, it may be shown that $\mathrm{SO}(n)$ has fundamental group (i.e. homotopy classes of loops) being $\mathbb{Z}_2$, whereas $\mathrm{Spin}(n)$ is simply-connected (i.e. it has trivial fundamental group).

15.1.1 Examples of (s)pin groups

<u>Spin(1)</u>**.** This is the two element group $\{\pm 1\}$.

Spin(2). It is not hard to deduce the parametrization by a circle coordinate,

$$\mathrm{Spin}(2) = \{a + b\mathsf{e}_1\mathsf{e}_2 \,:\, a^2 + b^2 = 1\} = \{\cos(\theta) + \sin(\theta)\mathsf{e}_1\mathsf{e}_2 \,:\, -\pi \le \theta \le \pi\}, \tag{15.2}$$

with group inversion given by taking $\theta \mapsto -\theta$. Let $\varphi = \cos(\theta) + \sin(\theta)\mathsf{e}_1\mathsf{e}_2$. Then we may calculate

$$\begin{aligned}\widetilde{\mathrm{Ad}}_\varphi(\mathsf{e}_1) &= \varphi\mathsf{e}_1\varphi^{-1} = \cos(2\theta)\mathsf{e}_1 + \sin(2\theta)\mathsf{e}_2\\ \widetilde{\mathrm{Ad}}_\varphi(\mathsf{e}_2) &= \varphi\mathsf{e}_2\varphi^{-1} = -\sin(2\theta)\mathsf{e}_1 + \cos(2\theta)\mathsf{e}_2.\end{aligned}$$

So

$$\begin{aligned}\chi = \widetilde{\mathrm{Ad}} : \mathrm{Spin}(2) &\to \mathrm{SO}(2)\\ \cos(\theta) + \sin(\theta)\mathsf{e}_1\mathsf{e}_2 &\mapsto \begin{pmatrix}\cos(2\theta) & -\sin(2\theta)\\ \sin(2\theta) & \cos(2\theta)\end{pmatrix}\end{aligned} \tag{15.3}$$

is a double-covering of circle groups.

Spin(3). This group is contained in Cl_3^0. Recall from Exercise 14.2 that Cl_3^0 can be identified with the Clifford algebra on two generators, $\mathsf{e}_3\mathsf{e}_1, \mathsf{e}_3\mathsf{e}_2$. Specifically,

$$\underbrace{Cl_3^0}_{\cong Cl_2 \cong \mathbb{H}} = \mathrm{span}_{\mathbb{R}}\{1, \underbrace{\mathsf{e}_1\mathsf{e}_2}_{I}, \underbrace{\mathsf{e}_3\mathsf{e}_1}_{J}, \underbrace{\mathsf{e}_3\mathsf{e}_2}_{K}\}. \tag{15.4}$$

Let $\mathbb{H}^1$ be the quaternions, viewed as a 1-dimensional "vector space" with "scalar multiplication" by the "field" $\mathbb{H}$ on the right. There is an "inner product" on $\mathbb{H}^1$, given by $(q_1, q_2) = \overline{q_1}q_2$, where the conjugate of $q = a + bI + cJ + dK$ is defined as $\overline{q} = a - bI - cJ - dK$. Next, $Cl_2 \cong \mathbb{H}$ is taken to act on $\mathbb{H}^1$ on the left, as an algebra of quaternion-linear operators. Within $\mathbb{C}l_2$, there is a group of operators which preserve the inner product on $\mathbb{H}^1$, called the *symplectic group* Sp(1). It is easily checked that

$$\mathrm{Sp}(1) = \{q \in \mathbb{H} \,:\, \overline{q}q = 1\}$$

comprises the unit quaternions. We will show that $\mathrm{Sp}(1) \subset Cl_2$ coincides with $\mathrm{Spin}(3) \subset Cl_3^0$ under the identification $Cl_2 \cong Cl_3^0$ of Eq. (15.4).

Let $y = \sum_{i=1}^3 a_i\mathsf{e}_i$ be a unit vector in $\mathbb{R}^3$. Then $\mathsf{e}_3 y = a_1\mathsf{e}_3\mathsf{e}_1 + a_2\mathsf{e}_3\mathsf{e}_2 - a_3$ is identified with the unit quaternion $-a_3 1 + a_1 J + a_2 K$. Similarly for $y\mathsf{e}_3$. By definition, elements of Spin(3) have the form

$$y_1 \cdots y_{2k} = (y_1\mathsf{e}_3)(\mathsf{e}_3 y_2)\cdots(y_{2k-1}\mathsf{e}_3)(\mathsf{e}_3 y_{2k}), \qquad |y_i| = 1.$$

So they are identified with products of unit quaternions, i.e. Sp(1) elements. Therefore, the identification $Cl_3^0 \cong Cl_2$ restricts to an injective homomorphism $\mathrm{Spin}(3) \hookrightarrow \mathrm{Sp}(1)$ of Lie groups. Since Sp(1) comprises the unit quaternions in $\mathbb{H}^1 \cong \mathbb{R}^4$, it is S^3 as a manifold, thus a connected Lie group of the same dimension as Spin(3). It may then be shown from general Lie theory that $\mathrm{Spin}(3) \cong \mathrm{Sp}(1)$ (Exercise).

It is also useful to regard $\mathbb{H}^1$ as $\mathbb{C}^2$, via $(a+bI+cJ+dK) \leftrightarrow \binom{a-ib}{c+id} \equiv \binom{w}{z}$, with right-multiplication by $-I$ corresponding to complex scalar multiplication by i. Then left multiplication by a $Cl_3^0 \cong \mathbb{H}$ algebra element $(a+bI+cJ+dK)$ is represented by the complex 2×2 matrix

$$\begin{pmatrix} a-ib & -c+id \\ c+id & a+ib \end{pmatrix} \equiv \begin{pmatrix} w & -\bar{z} \\ z & \bar{w} \end{pmatrix}. \tag{15.5}$$

Taking $q \mapsto \bar{q}$ corresponds to taking Hermitian conjugate; unit quaternions correspond to SU(2) matrices. We obtain a further accidental isomorphism $\mathrm{Spin}(3) \cong \mathrm{Sp}(1) \cong \mathrm{SU}(2)$.

15.2 Spin Lie algebra

The Lie group $Cl_n^\times$ is open in the vector space Cl_n, so its Lie algebra $\mathfrak{cl}_n^\times$ (tangent space at the identity) is identified with Cl_n, with commutator as Lie bracket. The pin and spin groups are Lie subgroups of $Cl_n^\times$, so their Lie algebras sit inside Cl_n. Let us understand the Lie algebra $\mathfrak{spin}(n)$ more explicitly.

We know that $\mathfrak{spin}(n) \subset \mathfrak{cl}_n^\times = Cl_n$ comprises the tangent vectors to the submanifold $\mathrm{Spin}(n) \subset Cl_n^\times$ at the identity element. The curves

$$\gamma_{(ij)} : t \mapsto \cos(t) + \sin(t)\mathsf{e}_i\mathsf{e}_j, \qquad 1 \le i < j \le n,$$

start at the identity, and we saw from the proof of Prop. 15.3 that these curves lie in $\mathrm{Spin}(n)$. At $t = 0$, the velocity vector of $\gamma_{(ij)}$ is $\mathsf{e}_i\mathsf{e}_j$. Therefore, $\mathfrak{spin}(n)$ contains all the bivectors $\mathsf{e}_i\mathsf{e}_j$. By counting dimensions, we see that $\mathfrak{spin}(n) = \mathrm{span}_{\mathbb{R}}\{\mathsf{e}_i\mathsf{e}_j \,:\, 1 \le i < j \le n\}$.

Next, let us compare the Lie algebras $\mathfrak{spin}(n)$ and $\mathfrak{so}(n)$. Recall that latter is the vector space of skew-symmetric operators on $\mathbb{R}^n$ equipped with the commutator. It will be convenient to identify $\Lambda^2(\mathbb{R}^n)$ with $\mathfrak{so}(n)$ as vector spaces, by regarding each wedge product $u \wedge v$ as the skew-symmetric operator

$$(u \wedge v) : w \mapsto \langle u|w\rangle v - \langle v|w\rangle u, \qquad u, v, w \in \mathbb{R}^n.$$

For example, $\mathsf{e}_1\mathsf{e}_2$ corresponds to the matrix

$$E_{12} = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

The surjective Lie group homomorphism

$$\chi = \mathrm{Ad} : \mathrm{Spin}(n) \to \mathrm{SO}(n) \subset \mathrm{GL}(n)$$

induces a Lie algebra isomorphism

$$\begin{aligned} \chi_* : \mathfrak{spin}(n) &\to \mathfrak{so}(n) \subset \mathrm{End}(\mathbb{R}^n) \\ \mathsf{e}_i\mathsf{e}_j &\mapsto \frac{d(\mathrm{Ad}_{\gamma_{(ij)}})}{dt}\Big|_{t=0}. \end{aligned}$$

We wish to understand what the last expression (a skew-endomorphism) does to $w \in \mathbb{R}^n$. A preliminary observation is

$$\frac{d\gamma_{(ij)}^{-1}(t)}{dt}\Big|_{t=0} = -\frac{d\gamma_{(ij)}(t)}{dt}\Big|_{t=0} = -\mathsf{e}_i\mathsf{e}_j.$$

Dropping the (ij)-subscripts for convenience, we have

$$\begin{aligned} \chi_*(\mathsf{e}_i\mathsf{e}_j)(w) &= \frac{d(\mathrm{Ad}_\gamma(w))}{dt} \\ &= \frac{d}{dt}\Big|_{t=0} \big(\gamma(t)w\gamma^{-1}(t)\big) \\ &= \gamma'(0)w\gamma^{-1}(0) + \gamma(0)w(\gamma^{-1})'(0) \\ &= \mathsf{e}_i\mathsf{e}_j w - w\mathsf{e}_i\mathsf{e}_j \\ &= \mathsf{e}_i\mathsf{e}_j w + (\mathsf{e}_i w + 2\langle \mathsf{e}_i | w\rangle)\mathsf{e}_j \\ &= \mathsf{e}_i\mathsf{e}_j w - \mathsf{e}_i\mathsf{e}_j w - 2\langle \mathsf{e}_j | w\rangle \mathsf{e}_i + 2\langle \mathsf{e}_i | w\rangle \mathsf{e}_j \\ &= 2(\mathsf{e}_i \wedge \mathsf{e}_j)(w). \end{aligned}$$

Therefore, the preimage of $\mathsf{e}_i \wedge \mathsf{e}_j$ is

$$\chi_*^{-1}(\mathsf{e}_i \wedge \mathsf{e}_j) = \frac{1}{2}\mathsf{e}_i\mathsf{e}_j = \frac{1}{4}[\mathsf{e}_i, \mathsf{e}_j].$$

Proposition 15.4. *Under the double-covering homomorphism* $\chi : \mathrm{Spin}(n) \to \mathrm{SO}(n)$*, the induced Lie algebra isomorphism* $\chi_* : \mathfrak{spin}(n) \to \mathfrak{so}(n) \cong \Lambda^2(\mathbb{R}^n)$ *is given on basis elements* $\mathsf{e}_i\mathsf{e}_j \in \mathfrak{spin}(n)$ *by the formula*

$$\chi_*(\mathsf{e}_i\mathsf{e}_j) = 2\mathsf{e}_i \wedge \mathsf{e}_j, \qquad 1 \le i < j \le n.$$

Consequently, for $u, v \in \mathbb{R}^n$*, we have*

$$\chi_*^{-1}(u \wedge v) = \frac{1}{4}[u, v].$$

Example 15.1. Elements of Spin(2) were described in Eq. (15.2). The Lie algebra $\mathfrak{spin}(2)$ is spanned by $\mathsf{e}_1\mathsf{e}_2$. For SO(2), the Lie algebra $\mathfrak{so}(2)$ is spanned by $\mathsf{e}_1 \wedge \mathsf{e}_2 \sim \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Although $\mathfrak{spin}(2)$ and $\mathfrak{so}(2)$ are both 1-dimensional Lie algebras with trivial Lie bracket, the specific isomorphism χ_* actually takes $\mathsf{e}_1\mathsf{e}_2 \mapsto 2\mathsf{e}_1 \wedge \mathsf{e}_2$.

Example 15.2. In Section 15.1.1, Eq. (15.4)-(15.5), we provided a 2×2 complex matrix representation of $Cl_3^0 \cong \mathbb{H}$. In particular, the basis $\{\mathsf{e}_1\mathsf{e}_2, \mathsf{e}_2\mathsf{e}_3, \mathsf{e}_3\mathsf{e}_1\}$ for $\mathfrak{spin}(3) \subset Cl_3^0$ is represented as

$$\underbrace{\mathsf{e}_2\mathsf{e}_3}_{-K} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \underbrace{\mathsf{e}_3\mathsf{e}_1}_{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \underbrace{\mathsf{e}_1\mathsf{e}_2}_{I} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{15.6}$$

Up to a factor of $\frac{i\hbar}{2}$, these are the spin matrices $\Sigma_x, \Sigma_y, \Sigma_z$ used in physics.

15.3 Spin representations

Spin groups are supposed to act as frame changes for some vector space. In quantum theory, these vector spaces are finite-dimensional complex Hilbert spaces, so we need $\mathrm{Spin}(n)$ to be represented unitarily on such a Hilbert space. Of particular importance are "spinors", which are constructed from from Clifford algebra representations.

In Section 14.3.2, we learned that $\mathbb{C}l_0 \cong \mathbb{C}$ and $\mathbb{C}l_1 \cong \mathbb{C} \oplus \mathbb{C}$. Proposition 14.3 says that

$$\mathbb{C}l_n \cong \begin{cases} \mathrm{M}_{2^{\frac{n}{2}}}(\mathbb{C}), & n \text{ even}, \\ \mathrm{M}_{2^{\frac{n-1}{2}}}(\mathbb{C}) \oplus \mathrm{M}_{2^{\frac{n-1}{2}}}(\mathbb{C}), & n \text{ odd}. \end{cases}$$

We state a few representation-theoretic results without proof (they are not particularly difficult.) First, any finite-dimensional representation of $M_m(\mathbb{C})$ or $M_m(\mathbb{C}) \oplus M_m(\mathbb{C})$ is completely reducible, in the sense of splitting into a direct sum of irreducible ones.

Next, a matrix algebra $M_m(\mathbb{C})$ has very simple representation theory: up to equivalence, its only irreducible representation (over $\mathbb{C}$) is the defining one on $\mathbb{C}^m$. As for $M_m(\mathbb{C}) \oplus M_m(\mathbb{C})$, it has two inequivalent irreducible representations: the representation space is $\mathbb{C}^m$ as before, except that one of the two factors of $M_m(\mathbb{C})$ acts as 0. So $\mathbb{C}l_n$ has a unique irreducible representation for n even, and two inequivalent irreducible representations for n odd. In the latter case, how should we distinguish the two possibilities?

Odd n. If we have an irreducible representation of $\rho : \mathbb{C}l_n \to \mathrm{End}(S)$ on some vector space S, then $\rho(\omega^{\mathbb{C}})^2 = 1$, so we get a splitting $S = S^+ \oplus S^-$ according to the $\pm$ eigenspaces of $\rho(\omega^{\mathbb{C}})$. Since $\rho(\omega^{\mathbb{C}})$ commutes with all the representative operators of $\mathbb{C}l_n$, the representation ρ can be reduced to either eigenspace, S^+ or S^-. By irreducibility, either $S = S^+$ or $S = S^-$. Thus we deduce that the irreducible representations of $\mathbb{C}l_n$ are distinguished by whether $\rho(\omega^{\mathbb{C}})$ acts as $+1$ or -1. So we could have a "left-handed" or "right-handed" representation space.

Example 15.3. For $\mathbb{C}l_3$, we could choose to irreducibly represent the e_i as

$$\mathsf{e}_1 = \pm \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \mathsf{e}_2 = \pm \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathsf{e}_3 = \pm \mathsf{e}_1 \mathsf{e}_2 = \pm \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{15.7}$$

This gives the left-handed (resp. right-handed) irreducible representation space S^+ (resp. S^-) for $\mathbb{C}l_3$.

Even n. There is a unique (up to equivalence) irreducible representation of $\mathbb{C}l_n$.

Proposition 15.5. *Let $n \geq 2$ be even, and let $\rho : \mathbb{C}l_n \to \mathrm{End}(S)$ be an irreducible representation. Write $S = S^+ \oplus S^-$ according to the $\pm$ eigenspaces of $\rho(\omega^{\mathbb{C}})$. The restriction of ρ to the even subalgebra $\mathbb{C}l_n^0$ can be reduced to either S^+ or S^-, and these give inequivalent irreducible representations of $\mathbb{C}l_n^0 \cong \mathbb{C}l_{n-1}$.*

Proof. The action of $\mathbb{C}l_n^0$ commutes with that of $\omega^{\mathbb{C}}$. So the reduction of the $\mathbb{C}l_n^0$ action to S^+ (or to S^-) is well-defined. Recall that the isomorphism

$\mathbb{C}l_n^0 \cong \mathbb{C}l_{n-1}$ can be implemented by taking, say, $\tilde{\mathsf{e}}_i = \mathsf{e}_n\mathsf{e}_i$, $i = 1, \ldots, n-1$. Then the chirality element for $\mathbb{C}l_n^0 \cong \mathbb{C}l_{n-1}$ would be

$$\omega_{n-1}^{\mathbb{C}} \sim \tilde{\mathsf{e}}_1 \cdots \tilde{\mathsf{e}}_{n-1} \sim (\mathsf{e}_n\mathsf{e}_1) \cdots (\mathsf{e}_n\mathsf{e}_{n-1}) \sim (\mathsf{e}_1 \cdots \mathsf{e}_{n-1}\mathsf{e}_n)\mathsf{e}_n^{n-2} \sim \mathsf{e}_1 \cdots \mathsf{e}_n \sim \omega_n^{\mathbb{C}},$$

where we have ignored various factors of $i^{(\cdot)}$. So, up to a possible sign, the action of $\omega_{n-1}^{\mathbb{C}} \in \mathbb{C}l_n^0$ on $S^{\pm}$ is precisely the action of $\omega_n^{\mathbb{C}}$ on $S^{\pm}$, which is just a sign ± 1 by definition. As $n-1$ is odd, we know from the previous case that this sign determines which irreducible $\mathbb{C}l_{n-1} \cong \mathbb{C}l_n^0$ we are in. $\square$

Example 15.4. For $\mathbb{C}l_2 \cong \mathrm{M}_2(\mathbb{C})$ take $S = \mathbb{C}^2$, and

$$\mathsf{e}_1 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad \mathsf{e}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \omega^{\mathbb{C}} = i\mathsf{e}_1\mathsf{e}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

So $S = S^+ \oplus S^- = \mathbb{C} \oplus \mathbb{C}$.

Definition 58. Let $\mathbb{C}l_n \to \mathrm{End}(S)$ be an irreducible complex representation of $\mathbb{C}l_n$. By restricting this representation to $\mathrm{Spin}(n) \subset \mathbb{C}l_n^0$, we obtain a homomorphism

$$\delta_n : \mathrm{Spin}(n) \to \mathrm{GL}(S),$$

called the *(complex) spin representation* of $\mathrm{Spin}(n)$. Elements of S are called *spinors*. With $S = S^+ \oplus S^-$ (even n case) and $S = S^{\pm}$ (odd n case), elements of $S^{\pm}$ are called *left-handed spinors* and *right-handed spinors* respectively.

Proposition 15.6. *For odd n, the spin representation δ_n of $\mathrm{Spin}(n)$ is independent of which of the two irreducible representations of $\mathbb{C}l_n$ is used. Furthermore, δ_n is an irreducible representation of $\mathrm{Spin}(n)$ in this case.*

For even n, the spin representation δ_n is reducible, and splits into $\delta_n = \delta_n^+ \oplus \delta_n^-$ according to the $\pm$-eigenspaces of $\omega^{\mathbb{C}}$. The representations δ_+ and δ_- are inequivalent irreducible representations of $\mathrm{Spin}(n)$.

Proof. **Odd n.** The parity automorphism α preserves $\mathbb{C}l_n^0$, but swaps $\mathbb{C}l_n^+$ with $\mathbb{C}l_n^-$, as we saw in Eq. (14.5). Thus $\mathbb{C}l_n^0$ is the "diagonal" subalgebra,

$$\mathbb{C}l_n^0 = \{(\varphi, \alpha(\varphi)) \in \mathbb{C}l_n^+ \oplus \mathbb{C}l_n^-\} \subset \mathbb{C}l_n.$$

If we choose the irreducible representation of $\mathbb{C}l_n$ for which $\mathbb{C}l_n^{\pm}$ acts nontrivially, we obtain a representation $\rho^{\pm}$ of $\mathbb{C}l_n^0$, and the latter is irreducible.

Observe that ρ^+ and ρ^- are equivalent representations of $\mathbb{C}l_n^0$, with α implementing the equivalence. In particular, their restrictions to $\mathrm{Spin}(n) \subset \mathbb{C}l_n^0$ are equivalent, so we simply write δ_n for either one. We observe that $\mathrm{Spin}(n)$ contains a linear basis $\{\mathsf{e}_I\}_{|I| \text{ even}}$ for $\mathbb{C}l_n^0$, so δ_n is also irreducible.

Even n**.** We saw from Prop. 15.5 that S splits into $S^+ \oplus S^-$, and that $\mathbb{C}l_n^0$ acts within S^+ or S^- separately. These two options provide inequivalent irreducible representations of $\mathbb{C}l_n^0$, therefore we also get inequivalent representations $\delta^\pm$ of $\mathrm{Spin}(n) \subset \mathbb{C}l_n^0$ on $S^\pm$. □

Example 15.5. We have $\mathrm{Spin}(1) = \{\pm 1\} \subset \mathbb{C}l_1^0 = \mathbb{C}$. The spin representation on $S^\pm = \mathbb{C}$ is the sign representation.

Example 15.6. Recall Example 15.4. The elements $\cos(\theta) + \sin(\theta)\mathsf{e}_1\mathsf{e}_2$ of $\mathrm{Spin}(2)$ are represented on $S^\pm \cong \mathbb{C}$ by the operator of multiplication by $\cos(\theta) \mp i\sin(\theta) = e^{\mp i\theta}$. So $\mathrm{Spin}(2)$ is represented as $\mathrm{U}(S^\pm) = \mathrm{U}(1)$ in two inequivalent ways. This makes clear that we should not be too sloppy about identifying $\mathrm{Spin}(2)$ and $\mathrm{U}(1)$.

Example 15.7. In Example 15.3, we described the irreducible representations of $\mathbb{C}l_3$ on $S^\pm \cong \mathbb{C}^2$, with $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ represented as in Eq. (15.7). Because the volume/chirality element is $\omega^{\mathbb{C}} = -\mathsf{e}_1\mathsf{e}_2\mathsf{e}_3$ is represented as $\pm\mathbf{1}_2$, we actually also have

$$\mathsf{e}_2\mathsf{e}_3 = \pm\mathsf{e}_1, \quad \mathsf{e}_3\mathsf{e}_1 = \pm\mathsf{e}_2, \quad \mathsf{e}_1\mathsf{e}_2 = \pm\mathsf{e}_3 \tag{15.8}$$

in this representation space. So Cl_3^0 is represented on $S^\pm$ by the same matrices found in Eq. (15.6). In particular, we obtain the spin representation of $\mathrm{Spin}(3) \subset Cl_3^0$ on $S^\pm$ as the $\mathrm{SU}(2)$ matrices found in Eq. (15.5).

Remark. In quantum mechanics in three spatial dimensions, it is usual to present the notion of spin as directly coming from the matrix Lie group $\mathrm{SU}(2)$, acting on "spinors" in $\mathbb{C}^2$. Actually, the relevant Lie group is, a priori, $\mathrm{Spin}(3)$ equipped with a double-covering map onto $\mathrm{SO}(3)$. The isomorphism $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$ is accidental.

Remark. The Clifford-algebraic adjoint/conjugation representation

$$\chi = \widetilde{\mathrm{Ad}} : \mathrm{Spin}(3) \to \mathrm{SO}(\mathbb{R}^3)$$

can be computed in the basis $\{\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\}$, so each element of $\mathrm{Spin}(3) \cong \mathrm{SU}(2)$ matrix becomes an $\mathrm{SO}(3)$ matrix, which can be verified to be given by Eq. (7.2) (exercise). Indeed, this is where that explicit double-covering formula came from.

Because of Eq. (15.8), we have coincidentally also computed the Lie-theoretic adjoint representation

$$\mathrm{Ad} : \underbrace{\mathrm{Spin}(3)}_{\mathrm{SU}(2)} \to \mathrm{GL}(\underbrace{\mathfrak{spin}(3)}_{\mathfrak{su}(2)}).$$

In general, the double-cover $\chi : \mathrm{Spin}(n) \to \mathrm{SO}(n)$ is not related to the Lie-theoretic $\mathrm{Ad} : \mathrm{Spin}(n) \to \mathrm{GL}(\mathfrak{spin}(n))$; the representation spaces do not have the same dimension.

15.3.1 Inner product on $S^{\pm}$

For $n = 1, 2$, irreducible representations of $\mathbb{C}l_n$ were provided in Section 14.3.2. With respect to the standard inner product on $\mathbb{C}^{2^{\lfloor \frac{n}{2} \rfloor}}$ the elements $\mathsf{e}_1, \mathsf{e}_2$ are explicitly unitary, therefore also skew-adjoint since they square to -1. In the $n = 2$ case, we have $S = S^+ \oplus S^-$ according to the chirality element $\omega^{\mathbb{C}}$.

The irreducible representations of $\mathbb{C}l_n, n > 2$ can be iteratively obtained, following the periodicity result of Prop. 14.3. The $\mathbb{C}l_3$ case was discussed above. Generally, there will be a Hermitian inner product on the representation space, such that the $\mathsf{e}_i, i = 1, \ldots, n$ are unitary, and are odd operators in the even n case. More abstractly, one can use an averaging argument to deduce the existence of such an inner product (Exercise).

Now, every vector $y \in V$ is a real linear combination of the e_i, so as a Clifford algebra element, it is also represented skew-adjointly. In particular, a unit vector y will be represented unitarily. The spin group is generated by products of such unit vectors, so $\mathrm{Spin}(n)$ is *unitarily* represented in the spin representation. In the even n case, the action of $\mathrm{Spin}(n)$ may be reduced to each graded component $S^{\pm}$.

Proposition 15.7. *The spin representation δ_n is compatible with Clifford multiplication by vectors of $\mathbb{R}^n$ on S, in the sense that for all $g \in \mathrm{Spin}(n), y \in \mathbb{R}^n \subset \mathbb{C}l_n, \xi \in S$,*

$$\delta_n(g) y \delta_n(g^{-1})(\xi) = (\chi(g)y) \cdot \xi. \tag{15.9}$$

Proof. Both $\delta_n(g)$ and the Clifford action $y \cdot (-)$ come from the Clifford algebra representation $\rho : \mathbb{C}l_n \to \mathrm{End}(S)$. So we have

$$\begin{aligned} \delta_n(g) y \delta_n(g^{-1})(\xi) &= \rho(g)\rho(y)\rho(g^{-1})(\xi) \\ &= \rho(gyg^{-1}) \cdot (\xi) \\ &= \rho(\chi(g)(y))(\xi) = (\chi(g)y) \cdot \xi. \end{aligned}$$

□

Remark (Physics notation). If $\mathsf{e}_1, \dots \mathsf{e}_n$ is an oriented orthonormal basis for $\mathbb{R}^n$, then the representative matrices $\gamma_i := \rho(\mathsf{e}_i)$ are called *Dirac* or *gamma matrices*. So an n-tuple $(y^1, \dots, y^n)$ corresponds to a linear combination $y^i\gamma_i$ of gamma matrices. Physicists usually think of $\chi(g) \sim R$ as an $\mathrm{SO}(n)$ rotation matrix (or $\mathrm{SO}(3,1)$ Lorentz) matrix $R^j{}_i$, and write the equality Eq. (15.9) as a "transformation law" for gamma matrices,

$$\Lambda(R)\gamma_i\Lambda(R)^{-1} = \gamma_j R^j{}_i. \tag{15.10}$$

Typically, the indices on γ_i are also raised by the bilinear form q (often a Lorentz metric with mixed signature $-+++$), and one writes $\{\Gamma^i, \Gamma^j\} = 2g^{ij}$ for the (Clifford) anticommutation relations. Here, there is usually another sign discrepency for the square of Γ^i.

Because of Eq. (15.10), the gamma matrices are sometimes said to "transform as a tangent (four-)vector". Later, we will uncover the geometric meaning of the gamma matrices more clearly (Remark 5).

16 Spinor geometry

16.1 Spin structures and spin connection

Recall the double-covering homomorphism, $\chi : \mathrm{Spin}(n) \to \mathrm{SO}(n)$, $n \geq 1$. For a Riemannian manifold, we would like to upgrade this pointwise "frame doubling" to a "principal frame bundle doubling".

Definition 59. Let X be an oriented Riemannian manifold. A *spin structure* is a χ-reduction, $F : \mathrm{Fr}^{\mathrm{Spin}}(X) \to \mathrm{Fr}^{\mathrm{SO}}(X)$, in the sense of (Definition 47). An oriented Riemannian manifold with a spin structure is called a *spin manifold*.

Some remarks:

- Thus, a spin structure is the data of a principal $\mathrm{Spin}(n)$-bundle which "double covers" the $\mathrm{SO}(n)$-bundle of oriented orthonormal tangent frames, while respecting the bundle projections. The same doubling notion makes sense for Euclidean vector bundles, but we will not require this generalization.

- [Optional information.] Not all oriented Riemannian manifolds admit spin structures — there is a topological obstruction called the second Stiefel–Whitney class. So sometimes one encounters the term "spinnable" manifold. If X is spinnable, the choice of spin structure may not be unique — it is labelled by the cohomology group $H^1(X, \mathbb{Z}_2)$.

- Upon fixing a basepoint $p \in P_x$, the restriction $F|_{P_x} : P_x \to \mathrm{Fr}^{\mathrm{SO}}(E_x)$ is identified with $\chi : \mathrm{Spin}(n) \to \mathrm{SO}(n)$.

- At each point, an orthonormal frame lifts to two possible "spin frames". As usual, neither spin frame is canonically preferred, but they are related to each other by the action of $-1 \in \mathrm{Spin}(n)$. Since there are actually more spin frames than orthonormal frames, we sometimes call $\mathrm{Fr}^{\mathrm{Spin}}(E)$ an "equivariant lift" of $\mathrm{Fr}^{\mathrm{SO}}(E)$, instead of a "reduction".

- Let $\mathbf{e} : U \to \mathrm{Fr}^{\mathrm{SO}}(X)$ be a local oriented orthonormal tangent frame field, and let U' be a contractible open subset of $U \subset X$. Then $\mathbf{e}(U')$ is a contractible open subset in $\mathrm{Fr}^{\mathrm{SO}}(X)$, and its preimage under F comprises two disjoint copies of $\mathbf{e}(U')$. These are the choices of "spin-frame fields" $\tilde{\mathbf{e}}$ lifting $\mathbf{e}$ over U'. Note that a lift may not exist over all of U.

Example 16.1. In Section 7, we constructed a spin structure on the round 2-sphere S^2.

Example 16.2. On S^1, we have $\mathrm{Fr}^{\mathrm{SO}}(S^1) = S^1 \times \{1\}$. There are two possible spin structures: $\mathrm{Fr}^{\mathrm{Spin}}(S^1)$ is either disconnected, $S^1 \times \mathbb{Z}_2$, or it is the connected double cover of S^1 (cf. $S(\mathcal{L}^{\mathbb{R}})$). These spin structures are respectively called *Neveu–Schwarz* and *Ramond*, in the theoretical physics literature.

The Levi–Civita connection on $\mathrm{Fr}^{\mathrm{SO}}$ has a canonical lift to $\mathrm{Fr}^{\mathrm{Spin}}$:

Definition 60. Let X be a spin manifold, so we have

$$F : \mathrm{Fr}^{\mathrm{Spin}}(X) \to \mathrm{Fr}^{\mathrm{SO}}(X).$$

The *spin connection* on $\mathrm{Fr}^{\mathrm{Spin}}(X)$ is the $\mathfrak{spin}(n)$-valued 1-form

$$\omega^{\mathrm{Spin}} := \chi_*^{-1} \circ F^* \omega^{\mathrm{LC}}.$$

Proposition 16.1. *The spin connection of Definition 60 is indeed a connection on the* $\mathrm{Spin}(n)$*-principal bundle* $\mathrm{Fr}^{\mathrm{Spin}}(X)$*.*

Proof. Let $u \in \mathfrak{spin}(n)$ and $u^\sharp$ be the corresponding fundamental vector field on $\mathrm{Fr}^{\mathrm{Spin}}(X)$. The first requirement for ω^{Spin} to be a connection on $\mathrm{Fr}^{\mathrm{Spin}}(X)$ is

$$(\chi_*^{-1} \circ F^* \omega^{\mathrm{LC}})(u_p^\sharp) = u, \qquad p \in \mathrm{Fr}^{\mathrm{Spin}}(X). \tag{16.1}$$

This condition is verified by the following computation. First,

$$\begin{aligned}
dF_p(u_p^\sharp) &= dF_p\left(\frac{d}{dt}\Big|_{t=0} p \cdot \exp(tu)\right) \\
&= \frac{d}{dt}\Big|_{t=0} F(p \cdot \exp(tu)) \\
&= \frac{d}{dt}\Big|_{t=0} F(p) \cdot \chi(\exp(tu)) && \text{(spin structure definition)} \\
&= \frac{d}{dt}\Big|_{t=0} F(p) \cdot \exp(t\chi_* u) && \text{(Ex. 10.4)} \\
&= (\chi_* u)^\sharp_{F(p)}.
\end{aligned}$$

Thus

$$(F^*\omega^{\mathrm{LC}})_p(u_p^\sharp) = \omega^{\mathrm{LC}}_{F(p)}(dF_p(u_p^\sharp)) = \omega^{\mathrm{LC}}_{F(p)}((\chi_* u)^\sharp_{F(p)}) = \chi_* u,$$

and applying χ_*^{-1} gives Eq. (16.1).

Next, we check the equivariance condition, Eq. (11.4), for ω^{Spin}. For any $g \in \mathrm{Spin}(n)$,

$$\begin{aligned} R_g^*(\chi_*^{-1} \circ F^*\omega^{\mathrm{LC}}) &= \chi_*^{-1} \circ (F \circ R_g)^*(\omega^{\mathrm{LC}}) \\ &= \chi_*^{-1} \circ (R_{\chi(g)} \circ F)^*(\omega^{\mathrm{LC}}) \\ &= \chi_*^{-1} \circ F^*(R_{\chi(g)}(\omega^{\mathrm{LC}})) \\ &= \chi_*^{-1} \circ F^*(\mathrm{Ad}_{\chi(g)^{-1}}(\omega^{\mathrm{LC}})) \\ &= \chi_*^{-1} \circ \mathrm{Ad}_{\chi(g)^{-1}}(F^*\omega^{\mathrm{LC}}) \\ &\overset{\text{Prop. 10.7}}{=} \mathrm{Ad}_{g^{-1}} \circ \chi_*^{-1}(F^*\omega^{\mathrm{LC}}), \end{aligned}$$

□

Local spin connection. In Section 13.7.2, we saw that a local oriented orthonormal frame $\mathbf{e}$ allows us to express ω^{LC} as a local $\mathfrak{so}(n)$-valued 1-form, with $(\mathbf{e}^*\omega^{\mathrm{LC}})_{ji} = \omega_{ij}$. Recall the standard basis $\mathsf{e}_i \wedge \mathsf{e}_j$ for $\mathfrak{so}(n)$. (Note the tricky transpositions: as a matrix, $\mathsf{e}_i \wedge \mathsf{e}_j$ has ji-entry being $+1$.) So the local form of the Levi–Civita connection is

$$\mathbf{e}^*\omega^{\mathrm{LC}} = \sum_{1 \le i < j \le n} \omega_{ij} \mathsf{e}_i \wedge \mathsf{e}_j = \frac{1}{2} \sum_{i,j=1}^{n} \omega_{ij} \mathsf{e}_i \wedge \mathsf{e}_j. \tag{16.2}$$

Now let $\tilde{\mathbf{e}}$ be either one of the local "spin frames" which lifts $\mathbf{e}$. (Here, we work over a small enough part of X such that the lift exists.) Then the local gauge potential for ω^{Spin} will be

$$\begin{aligned} \tilde{\mathbf{e}}^*(\omega^{\mathrm{Spin}}) = \tilde{\mathbf{e}}^*(\chi_*^{-1} \circ F^*\omega^{\mathrm{LC}}) &= \chi_*^{-1} \circ (\tilde{\mathbf{e}}^* F^*\omega^{\mathrm{LC}}) \\ &= \chi_*^{-1} \circ ((F \circ \tilde{\mathbf{e}})^*\omega^{\mathrm{LC}}) = \chi_*^{-1}(\mathbf{e}^*\omega^{\mathrm{LC}}). \end{aligned} \tag{16.3}$$

Recalling the formula for χ_* from Prop. 15.4, this means that the spin connection has the local expression

$$\tilde{\mathbf{e}}^*(\omega^{\mathrm{Spin}}) = \frac{1}{4} \sum_{i,j=1}^{n} \omega_{ij} \mathsf{e}_i \mathsf{e}_j. \tag{16.4}$$

The extra factor of $\frac{1}{2}$ in Eq. (16.4), compared to Eq. (16.2), is very important.

16.2 Example: spin structure and connection on S^2

In Section 7.3, we had constructed

$$\bar{\pi} : S^3 \to S^2, \qquad (w, z) \mapsto (2\bar{w}z, |w|^2 - |z|^2),$$

as a principal U(1)-bundle with right U(1)-action,

$$(w, z) \cdot \exp i\theta = (we^{i\theta}, ze^{i\theta}).$$

Let

$$\phi : \mathrm{U}(1) \to \mathrm{SO}(2), \qquad e^{i\theta} \mapsto \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{pmatrix}.$$

Then the map $F : S^3 \to \mathrm{Fr}^{\mathrm{SO}}(S^2)$, Eq. (7.2), was shown to be a ϕ-reduction of $\mathrm{Fr}^{\mathrm{SO}}(S^2)$ to the U(1)-principal bundle S^3.

Now recall the spin double covering map, Eq. (15.3),

$$\chi : \mathrm{Spin}(2) \to \mathrm{SO}(2), \qquad \underbrace{\cos\theta + \sin\theta \mathsf{e}_1\mathsf{e}_2}_{\exp(\theta\mathsf{e}_1\mathsf{e}_2)} \mapsto \begin{pmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix}.$$

If we make the identification $\mathrm{Spin}(2) \cong \overline{\mathrm{U}(1)}$, then ϕ is identified with χ, and we have precisely exhibited $S^3 \to S^2$ a spin structure for S^2. That is, we regard $S^3 = \mathrm{Fr}^{\mathrm{Spin}}(S^2)$, with Spin(2)-action

$$(w, z) \cdot \exp(\theta\mathsf{e}_1\mathsf{e}_2) = (w, z) \cdot e^{-i\theta} = (we^{-i\theta}, ze^{-i\theta}).$$

The following $\mathfrak{spin}(2)$-valued 1-form defines a connection on $S^3 = \mathrm{Fr}^{\mathrm{Spin}}(S^2)$,

$$\omega_{(w,z)}(\eta, \zeta) = -\mathsf{e}_1\mathsf{e}_2 \operatorname{Im}(\bar{w}\eta + \bar{z}\zeta), \qquad (\eta, \zeta) \in T_{(w,z)}S^3 \subset T_{(w,z)}\mathbb{C}^2 = \mathbb{C}^2.$$

With a few identifications, ω is the *magnetic monopole* U(1)-connection explored in Assignment 3. We shall verify that ω is precisely the spin connection.

In polar coordinates (ϑ, φ) for S^2,

$$(2\bar{w}z, |w|^2 - |z|^2) = (\sin\vartheta e^{i\varphi}, \cos\vartheta).$$

The following is a local spin frame field,

$$\begin{aligned} \tilde{\mathbf{e}} : S^2 &\to S^3 = \mathrm{Fr}^{\mathrm{Spin}}(S^2) \\ (\vartheta, \varphi) &\mapsto \left(\cos\frac{\vartheta}{2}e^{-i\varphi/2}, \sin\frac{\vartheta}{2}e^{i\varphi/2}\right), \end{aligned}$$

since $\bar{\pi} \circ \tilde{\mathbf{e}}(\vartheta, \varphi) = (\vartheta, \varphi)$ is easily verified. The local gauge potential for ω can be computed to be

$$\tilde{\mathbf{e}}^* \omega = \frac{\mathsf{e}_1 \mathsf{e}_2}{2} \cos \vartheta \, d\varphi.$$

Now, it may be checked that $F \circ \tilde{\mathbf{e}} = \mathbf{e}$, where $\mathbf{e}$ is the following orthonormal tangent frame field,

$$\mathbf{e} : (\vartheta, \varphi) \mapsto \begin{pmatrix} \cos\vartheta \cos\varphi & -\sin\varphi & \sin\vartheta\cos\varphi \\ \cos\vartheta\sin\varphi & \cos\varphi & \sin\vartheta\sin\varphi \\ -\sin\vartheta & 0 & \cos\vartheta \end{pmatrix} \in \mathrm{SO}(3).$$

We recall that the last column of the above SO(3) matrix gives the cartesian coordinates (of $(\vartheta, \varphi) \in S^2$), while the first two columns constitute the orthonormal frame at that point. So $\mathbf{e} = \{e_1, e_2\}$ at (ϑ, φ) is

$$\begin{aligned} e_1 &= \cos\vartheta\cos\varphi\, \partial_{x^1} + \cos\vartheta\sin\varphi\, \partial_{x^2} - \sin\vartheta\, \partial_{x^3} = \partial_\theta = e_\vartheta, \\ e_2 &= -\sin\varphi\, \partial_{x^1} + \cos\varphi\, \partial_{x^2} = \frac{1}{\sin\vartheta}\partial_\varphi = e_\varphi, \end{aligned}$$

i.e., the normalized polar coordinate frame field on S^2. In this frame, we may compute the local expression for the Levi–Civita connection to be

$$\omega_{12}^{\mathrm{LC}}(\vartheta, \varphi) = -\omega_{21}^{\mathrm{LC}}(\vartheta, \varphi) = \cot\vartheta\, e_\varphi^* = \cos\vartheta\, d\varphi.$$

So in the lifted spin frame $\tilde{\mathbf{e}}$, the spin connection has local expression

$$\tilde{\mathbf{e}}(\omega^{\mathrm{Spin}}) \overset{\text{Eq. (16.4)}}{=} \frac{1}{4}(\mathsf{e}_1\mathsf{e}_2\omega_{12}^{\mathrm{LC}} + \mathsf{e}_2\mathsf{e}_1\omega_{21}^{\mathrm{LC}}) = \frac{\mathsf{e}_1\mathsf{e}_2}{2}\cos\vartheta\, d\varphi,$$

which is the same as $\tilde{\mathbf{e}}^*\omega$ found above.

To summarize: The spin frame bundle for S^2 with spin connection can be identified with the magnetic monopole U(1)-bundle.

16.3 Spinor bundles and Clifford multiplication

We have seen that a spin manifold has a "spin-frame bundle", equipped with a canonical spin connection. We use quotation marks because we have not yet described the vector bundles whose "spin frames" are organized according to this "spin-frame bundle".

Definition 61. Let X be a spin manifold. The *spinor bundle* $\mathcal{S}$ over X is the associated Hermitian vector bundle

$$\mathcal{S} = \mathrm{Fr}^{\mathrm{Spin}}(X) \times_{\delta_n} S,$$

where $\delta_n : \mathrm{Spin}(n) \to S$ is the (unitary) spin representation (Definition 58).

Remark. Write $\eta_a, a = 1, \ldots, 2^{\lfloor \frac{n}{2} \rfloor}$ for the a-th basis vector of $S = \mathbb{C}^{2^{\lfloor \frac{n}{2} \rfloor}}$, and let $\mathbf{e}_x \in \mathrm{Fr}^{\mathrm{Spin}}(X)_x$. Then the fibre $\mathcal{S}_x$ is spanned by the orthonormal basis

$$\psi_{a,x} := [\mathbf{e}_x, \eta_a], \qquad a = 1, \ldots, 2^{\lfloor \frac{n}{2} \rfloor}.$$

Similarly, if $\mathbf{e} : U \to \mathrm{Fr}^{\mathrm{Spin}}(X)$ is a local section, then

$$\mathbf{e} = (\psi_1, \ldots, \psi_{2^{\lfloor \frac{n}{2} \rfloor}}), \qquad \psi_a = [\mathbf{e}, \eta_a], \qquad a = 1, \ldots, 2^{\lfloor \frac{n}{2} \rfloor},$$

is a local field of (orthonormal) spin frames for $\mathcal{S}|_U$.

Clifford multiplication. Recall that

$$TX \cong \mathrm{Fr}^{\mathrm{SO}}(X) \times_{\rho_{\mathrm{std}}} \mathbb{R}^n,$$
$$\mathcal{S} \cong \mathrm{Fr}^{\mathrm{Spin}}(X) \times_{\delta_n} S,$$

where ρ_{std} is the defining representation of $\mathrm{SO}(n)$ on $\mathbb{R}^n$ (which will generally be suppressed in our notation). Furthermore, each fibre $\mathcal{S}_x$ of the spinor bundle carries an irreducible Clifford algebra representation ρ, so we can promote tangent vectors of T_xX to operators acting on $\mathcal{S}_x$, as follows.

Definition 62. *Clifford multiplication* is the vector bundle homomorphism

$$\mu : TX \otimes \mathcal{S} \to \mathcal{S},$$

defined for $v_x \in T_xX, \psi_x \in \mathcal{S}_x$, $x \in X$, by the formula

$$v_x \cdot \psi_x \equiv \mu(\underbrace{[\mathbf{e}_x, \mathbf{v}]}_{=v_x} \otimes \underbrace{[\tilde{\mathbf{e}}_x, \xi]}_{=\psi_x}) := [\tilde{\mathbf{e}}_x, \underbrace{\mathbf{v} \cdot \xi}_{\text{Clifford}}], \tag{16.5}$$

where $\mathbf{e}_x \in \mathrm{Fr}^{\mathrm{SO}}(X)$ is any oriented orthonormal frame at x, $\mathbf{v}$ is the n-tuple of components of v_x in this frame, and $\tilde{\mathbf{e}}_x \in \mathrm{Fr}^{\mathrm{Spin}}(X)$ is a spin frame which lifts $\mathbf{e}_x$.

Proposition 16.2. *Clifford multiplication is well-defined. For even dimensional X, the spinor bundle splits canonically into an orthogonal direct sum $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$, according to the induced Clifford action of the chirality element $\omega^{\mathbb{C}}$. Furthermore, Clifford multiplication by TX is odd in the sense of mapping $\mathcal{S}^\pm$ to $\mathcal{S}^\mp$.*

Proof. For any $g \in \mathrm{Spin}(n)$, we could have changed the tangent frame and spin frame,

$$v_x = [\mathbf{e}_x \cdot \chi(g), \chi(g^{-1})\mathbf{v}], \qquad \psi_x = [\tilde{\mathbf{e}}_x \cdot g, \delta_n(g^{-1})\xi].$$

In this alternative representation, Clifford multiplication would be

$$\begin{aligned}
\mu(v_x \otimes \psi_x) &= [\mathbf{e}_x \cdot \chi(g), \chi(g^{-1})\mathbf{v}] \cdot [\tilde{\mathbf{e}}_x \cdot g, \delta_n(g^{-1})\xi] \\
&= [\mathbf{e}_x \cdot \chi(g), \chi(g^{-1})\mathbf{v}] \cdot [\widetilde{\mathbf{e}_x \cdot \chi(g)}, \delta_n(g^{-1})\xi] && \text{(spin structure)} \\
&= [\widetilde{\mathbf{e}_x \cdot \chi(g)}, (\chi(g^{-1})\mathbf{v}) \cdot (\delta_n(g^{-1})\xi)] && \text{(new Clifford multiplication)} \\
&= [\tilde{\mathbf{e}}_x \cdot g, \delta_n(g^{-1})(\mathbf{v} \cdot \xi)] && \text{(Prop. 15.7)} \\
&= [\tilde{\mathbf{e}}_x, \mathbf{v} \cdot \xi], && \text{(associated spinor bundle)}
\end{aligned}$$

which is the same as what we had in the original frames, Eq. (16.5).

The proofs of the remaining statements are left as an exercise. □

Globally, we may think of Clifford multiplication as defining an action of a tangent vector field $v \in \mathfrak{X}(X) = \Gamma(TX)$ on spinor fields $\psi \in \Gamma(\mathcal{S})$. Recall that at the beginning of Section 14, we had motivated Clifford algebras because we wanted to promote tangent vectors to "matrix-derivatives". Later on, we will complete this process by providing the (covariant) derivative aspect.

Remark 5 (Local picture of Clifford multiplication). Let $\mathbf{e} = \{e_1, \ldots, e_n\} : U \to \mathrm{Fr}^{\mathrm{SO}}(X)$ be a local oriented orthonormal frame field, and $\tilde{\mathbf{e}}$ be a lift to a local spin frame field. Then v is given locally by $[\mathbf{e}, \mathbf{v}]$ where $\mathbf{v}$ is some $\mathbb{R}^n$-valued function over U. Similarly, ψ is locally given by $[\pm\tilde{\mathbf{e}}, \pm\xi]$ where ξ is a $\mathbb{C}^{2^{\lfloor \frac{n}{2} \rfloor}}$-valued function over U. Then Clifford multiplication (Definition 62) is locally implemented as

$$\xi(x) \mapsto \mathbf{v}(x) \cdot \xi(x), \qquad x \in U.$$

In particular, take $\mathbf{v}(x) = \mathsf{e}_i$ for all $x \in U$, so that $[\mathbf{e}, \mathsf{e}_i]$ is the i-th vector field e_i of the orthonormal frame field. Clifford multiplication by e_i effects

$\xi(x) \mapsto \mathsf{e}_i \cdot \xi(x)$; in physicists' gamma-matrix notation,

$$\xi \mapsto \gamma_i \cdot \xi.$$

We should keep in mind that the local frames $\mathbf{e}$ and $\tilde{\mathbf{e}}$ are implicit.

In introductory treatments, one works on flat Euclidean or Minkowski space, and uses *global* orthonormal coordinate vector fields. In curved space(time), a deeper geometric understanding is required. The gamma matrices should be understood as the *local, gauge-dependent* implementers of Clifford multiplication by the local orthonormal frame elements. We are free to change *local* orthonormal/spin frames ("local rotation/Lorentz invariance"), and this will cause the local gamma matrices to transform according to Eq. (15.9), or the physicist version Eq. (15.10).

Remark (Physics terminology). In the even n case, sections of $\mathcal{S}$ are called *Dirac spinor fields*, while sections of $\mathcal{S}^{\pm}$ are called *Weyl spinor fields*. The "Weyl spinor subbundles" $\mathcal{S}^{\pm}$ could also be obtained directly as

$$\mathcal{S}^{\pm} \cong \mathrm{Fr}^{\mathrm{Spin}}(X) \times_{\delta_n^+} S^{\pm}.$$

In the odd n case, there are actually two possible Clifford multiplications, depending on which irreducible Clifford representation (left or right-handed) is used. In this case, sections of $\mathcal{S}$ are also called Weyl spinor fields.

16.4 Spinor bundle covariant derivative

By the general constructions of Section 13.3, the associated spinor bundle $\mathcal{S}$ inherits a spin-covariant derivative $\nabla^{\mathrm{Spin}} := \nabla^{\omega^{\mathrm{Spin}}}$ from the spin connection. Furthermore, this is automatically a unitary connection, because the spin group is represented unitarily.

The general local formula for induced covariant derivatives was given in Eq. (13.4). For example, in a local oriented orthonormal frame $\mathbf{e} = \{e_1, \dots, e_n\}$, the Levi–Civita connection is described by $\mathbf{e}^*\omega^{\mathrm{LC}} = \sum_{i,j=1}^n \frac{1}{2}\omega_{ij}\mathsf{e}_i \wedge \mathsf{e}_j$, Eq. (16.2), where the 1-forms ω_{ij} are given by Eq. (13.9). For the spin connection, we have, in a lifted local spin frame $\tilde{\mathbf{e}}$, the formula

$$\begin{aligned}
\nabla_u^{\mathrm{Spin}}\psi &= \nabla_u^{\mathrm{Spin}}[\tilde{\mathbf{e}}, \xi] \\
&= [\tilde{\mathbf{e}}, d\xi(u) + (d\delta_n)_e(\tilde{\mathbf{e}}^*\omega^{\mathrm{Spin}}(u)) \cdot \xi] \\
&\overset{\text{Eq. (16.4)}}{=} [\tilde{\mathbf{e}}, d\xi(u) + \frac{1}{4}\omega_{ij}(u)\mathsf{e}_i\mathsf{e}_j \cdot \xi], \qquad u \in \mathfrak{X}(X).
\end{aligned} \tag{16.6}$$

In the last line, we use the fact that the spin representation δ_n comes from a Clifford algebra representation on the spinor space S, and likewise for the the induced Lie algebra representation of $\mathfrak{spin}(n)$. Keeping the spin frame implicit, and writing $\omega_{kij} := \omega_{ij}(\partial_k)$, the "$k$-th spin-covariant derivative" is

$$\text{“ } \nabla_k^{\text{Spin}} = \partial_k + \frac{1}{4}\omega_{kij}\mathsf{e}_i\mathsf{e}_j \text{ ”}, \tag{16.7}$$

acting on S-valued functions $\xi : U \to S$.

Write $\tilde{\mathbf{e}} = (\tilde{e}_1, \ldots, \tilde{e}_a, \ldots, \tilde{e}_{2^{\lfloor \frac{n}{2} \rfloor}})$ for the spin frame field. Each constituent spinor field $\tilde{e}_a = [\tilde{\mathbf{e}}, \eta_a]$ has constant component functions $\eta_a = (0, \ldots, 0, \underbrace{1}_{a\text{-th}}, 0, \ldots, 0)$.

The spin covariant derivative of $\tilde{e}_a$ is

$$\begin{aligned}
\nabla^{\text{Spin}}\tilde{e}_a &= [\tilde{\mathbf{e}}, \frac{1}{4}\omega_{ij}\mathsf{e}_i\mathsf{e}_j \cdot \eta_a] && \text{(sum over } i, j) \\
&= \frac{1}{4}\omega_{ij}[\mathbf{e}, \mathsf{e}_i] \cdot [\mathbf{e}, \mathsf{e}_j] \cdot [\tilde{\mathbf{e}}, \eta_a] \\
&= \frac{1}{4}\omega_{ij} e_i \cdot e_j \cdot \tilde{e}_a, && \omega_{ij} \overset{\text{Eq. (13.9)}}{=} g(\nabla^{\text{LC}} e_i, e_j).
\end{aligned} \tag{16.8}$$

Proposition 16.3. *The connections $\nabla^{\text{Spin}}, \nabla^{\text{LC}}$ and Clifford multiplication are Leibniz rule compatible, in the sense that for any $u, v \in \mathfrak{X}(X)$ and $\psi \in \Gamma(\mathcal{S})$,*

$$\nabla_u^{\text{Spin}}(v \cdot \psi) = (\nabla_u^{\text{LC}} v) \cdot \psi + v \cdot \nabla_u^{\text{Spin}}\psi. \tag{16.9}$$

Proof. It is possible to verify this directly from the formula Eq. (16.8), by using general properties of covariant derivatives.

We sketch a more conceptual argument. Let $A \in \mathfrak{spin}(n)$ and $h = \exp(tA) \in \text{Spin}(n)$. For any vector $\mathbf{v} \in \mathbb{R}^n \subset \mathbb{C}l_n$ and spinor $\xi \in S$,

$$h(\mathbf{v} \cdot \xi) = (h\mathbf{v}h^{-1}) \cdot (h\xi) \overset{\text{Eq. (15.9)}}{=} (\chi(h)(\mathbf{v})) \cdot (h\xi).$$

Differentiating at $t = 0$ gives, with the usual Leibniz rule,

$$A(\mathbf{v} \cdot \xi) = (\chi_*(A)\mathbf{v}) \cdot \xi + \mathbf{v} \cdot (A\xi). \tag{16.10}$$

This is essentially what is happening when we compute $\nabla_u^{\text{Spin}}(v \cdot \psi)$ in a spin frame, Eq. (16.6). The local $\mathfrak{spin}(n)$-valued connection acts on the local spinor representing $(v \cdot \psi)$, and this splits into the sum of two Clifford multiplications by vectors, as in Eq. (16.10). The latter is the local version of Eq. (16.9). □

Proposition 16.4. *When n is even, ∇^{Spin} preserves the splitting $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$.*

Proof. This can be shown with the help of Prop. (16.3), and is left as an exercise. □

16.5 Spin–Riemannian-curvature relations

Let $\mathcal{R} \in \Omega^2(X, \mathrm{End}(\mathcal{S}))$ be the curvature of ∇^{Spin} (Definition 54). Let us work out the relationship between $\mathcal{R}$ and the *Riemann curvature tensor* $R \in \Omega^2(X, \mathrm{End}(TX))$ of ∇^{LC}. As these are tensors, we can study their relationship in a local orthonormal frame $\mathbf{e}$ over $U \subset X$, with spin frame lift $\tilde{\mathbf{e}}$.

As ω^{LC} is a connection on a principal $\mathrm{SO}(n)$-bundle, have the local Cartan structure equation,

$$\mathbf{e}^*\Omega^{\mathrm{LC}} = d\omega^{(\mathbf{e})} + \frac{1}{2}[\omega^{(\mathbf{e})}, \omega^{(\mathbf{e})}], \qquad \omega^{(\mathbf{e})} := \mathbf{e}^*\omega^{\mathrm{LC}} \in \Omega^1(U, \mathfrak{so}(n)).$$

For $u, v \in \mathfrak{X}(X)$, the $\mathfrak{so}(n)$-valued function $\mathbf{e}^*\Omega^{\mathrm{LC}}(u, v)$ expresses the curvature endomorphism $R(u, v)$ on $TX|_U$ with respect to the local frame $\mathbf{e}$. Explicitly,

$$\mathbf{e}^*\Omega^{\mathrm{LC}}(u, v) = \sum_{i<j} g(R(u, v)e_i, e_j)\mathsf{e}_i \wedge \mathsf{e}_j. \tag{16.11}$$

For the spin connection, the gauge potential is (Eq. (16.3))

$$\tilde{\omega}^{(\tilde{\mathbf{e}})} := \tilde{\mathbf{e}}^*\omega^{\mathrm{Spin}} = \chi_*^{-1}(\mathbf{e}^*\omega^{\mathrm{LC}}) = \chi_*^{-1}(\omega^{(\mathbf{e})}).$$

Since χ_* is a Lie algebra isomorphism, the local field strength of ω^{Spin} is

$$\tilde{\mathbf{e}}^*\Omega^{\mathrm{Spin}} = d\tilde{\omega}^{(\tilde{\mathbf{e}})} + \frac{1}{2}[\tilde{\omega}^{(\tilde{\mathbf{e}})}, \tilde{\omega}^{(\tilde{\mathbf{e}})}] = \chi_*^{-1}(d\omega^{(\mathbf{e})} + \frac{1}{2}[\omega^{(\mathbf{e})}, \omega^{(\mathbf{e})}]) = \chi_*^{-1}(\mathbf{e}^*\Omega^{\mathrm{LC}}),$$

and from Eq. (16.11),

$$\tilde{\mathbf{e}}^*\Omega^{\mathrm{Spin}}(u, v) = \frac{1}{2}\sum_{i<j} g(R(u, v)e_i, e_j)\mathsf{e}_i\mathsf{e}_j, \qquad u, v \in \mathfrak{X}(X).$$

The above $\mathfrak{spin}(n)$-valued function represents the spin curvature endomorphism $\mathcal{R}(u, v)$ with respect to the spin frame $\tilde{\mathbf{e}}$. Here, $\mathsf{e}_i\mathsf{e}_j$ acts by Clifford multiplication. Because $\mathcal{R}(u, v)$ is $C^\infty(X)$-linear, its action on a general section $\psi \in \Gamma(\mathcal{S})$ is

$$\mathcal{R}(u, v)\psi = \frac{1}{4}\sum_{i,j=1}^{n} g(R(u, v)e_i, e_j)e_i \cdot e_j \cdot \psi, \qquad \psi \in \Gamma(\mathcal{S}|_U),\ u, v \in \mathfrak{X}(X). \tag{16.12}$$

Eq. (16.12) is the basic identity relating the spin curvature endomorphism $\mathcal{R}(\cdot, \cdot)$ to the Riemann $R(\cdot, \cdot)$. To progress further, we first recall some general properties of R.

Lemma 16.5. *For any $u, v, w, z \in \mathfrak{X}(X)$,*

$$R(u,v)w + R(v,w)u + R(w,u)v = 0. \qquad \text{(First Bianchi identity)}$$
$$g(R(u,v)w, z) = g(R(w,z)u, v). \qquad \text{(Swap symmetry)}$$

Proof. By $C^\infty(X)$-linearity in all arguments, it suffices to assume that u, v, w, z are coordinate vector fields, i.e., their Lie brackets vanish. Recall that $\nabla = \nabla^{\mathrm{LC}}$ is torsion-free, Eq. (13.8), so $\nabla_u v - \nabla_v u = [u, v] = 0$. Then

$$\begin{aligned} R(u,v)w + R(v,w)u + R(w,u)v = \nabla_u\nabla_v w - \nabla_v\nabla_u w + \nabla_v\nabla_w u - \nabla_w\nabla_v u \ = 0. \\ + \nabla_w\nabla_u v - \nabla_u\nabla_w v \end{aligned}$$

Next, note that R is a skew-symmetric endomorphism-valued 2 form, so $g(R(u,v)w, z)$ is antisymmetric in the u, v arguments as well as the w, z arguments. Then

$$\begin{aligned} 2g(R(u,v)w, z) &= -g(R(v,u)w, z) - g(R(u,v)z, w) \\ &\overset{\text{Bianchi}}{=} g(R(u,w)v, z) + g(R(w,v)u, z) + g(R(v,z)u, w) + g(R(z,u)v, w) \\ &\overset{\text{antisymmetries}}{=} g(R(w,u)z, v) + g(R(u,z)w, v) + g(R(z,v)w, u) + g(R(v,w)z, u) \\ &\overset{\text{Bianchi}}{=} -g(R(z,w)u, v) - g(R(w,z)v, u) = 2g(R(w,z)u, v). \end{aligned}$$

□

Ricci and scalar curvature. For each pair of vector fields v, w, we have a tangent bundle endomorphism, $(\cdot) \mapsto R(\cdot, v)w$, whose trace is denoted $\mathrm{Ric}(v, w)$. If we use an orthonormal frame to compute this trace, we get

$$\mathrm{Ric}(v,w) \overset{\text{local}}{=} \sum_{i=1}^n g(R(e_i, v)w, e_i) \overset{\text{skew}}{=} \sum_{i=1}^n g(R(v, e_i)e_i, w).$$

$\mathrm{Ric}(\cdot,\cdot)$ is called the Ricci tensor. Contracting the Ricci tensor using the metric gives the Riemannian *scalar curvature* function,

$$\mathrm{Sc} = \sum_{i,j=1}^n g^{ij}\mathrm{Ric}(e_i, e_j) \overset{g^{ij}=\delta^{ij}}{=} \sum_{j=1}^n \mathrm{Ric}(e_j, e_j) = \sum_{i,j=1}^n g(R(e_j, e_i)e_i, e_j). \quad (16.13)$$

For surfaces ($n = 2$), Sc is twice of the classical Gaussian curvature κ.

Returning to Eq. (16.12), we would like to move the Clifford multiplications over to the spin side, and contract the arguments. This results in the spin–Riemann *scalar curvature* relationship.

Lemma 16.6. *Let* $\mathbf{e} = \{e_1, \dots, e_n\}$ *be a local orthonormal frame for the Riemannian spin manifold* (X, g)*. Then*

$$\underbrace{\sum_{i,j=1}^{n} e_i \cdot e_j \cdot \mathcal{R}(e_i, e_j)}_{\text{“spin scalar curvature”}} = \frac{1}{2} \underbrace{\sum_{i,j=1}^{n} g(R(e_i, e_j)e_j, e_i)}_{\text{Riemann scalar curvature}} \equiv \frac{1}{2}\,\mathrm{Sc}.$$

Proof. Let us write $R_{lkij} := g(R(e_i, e_j)e_k, e_l)$, so Lemma 16.5 reads

$$R_{lkij} + R_{lijk} + R_{ljki} = 0, \tag{16.14}$$

$$R_{lkij} = R_{ijlk}. \tag{16.15}$$

We rewrite the sum in question in terms of R_{ijkl},

$$\sum_{i,j=1}^{n} e_i \cdot e_j \cdot \mathcal{R}(e_i, e_j) \overset{\text{Eq. (16.12)}}{=} \frac{1}{4} \sum_{i,j,k,l=1}^{n} R_{lkij} e_i \cdot e_j \cdot e_k \cdot e_l.$$

For the terms with distinct i, j, k, we may rearrange the Clifford multiplications and relabel indices to get

$$\sum_{i,j,k \text{ distinct}} R_{lkij}\, e_i \cdot e_j \cdot e_k = \frac{1}{3} \sum_{i,j,k} \underbrace{(R_{lkij} + R_{lijk} + R_{ljki})}_{=0 \text{ by Eq. (16.14)}} e_i \cdot e_j \cdot e_k = 0.$$

The $i = j$ terms contribute 0, by antisymmetry of $R(\cdot, \cdot)$. So we are left with

$$\begin{aligned}
\sum_{i,j}^{n} e_i \cdot e_j \cdot \mathcal{R}(e_i, e_j) &= \frac{1}{4} \left(\sum_{i=k,j,l} + \sum_{i,j=k,l} \right) R_{lkij}\, e_i \cdot e_j \cdot e_k \cdot e_l \\
&= \frac{1}{4} \sum_{i,j,l} (R_{liij}\, e_j \cdot e_l - R_{ljij}\, e_i \cdot e_l) \\
&= -\frac{1}{2} \sum_{i,j,l} R_{ljij}\, e_i \cdot e_l && \text{(relabel indices)} \\
&= -\frac{1}{2} \sum_{i,j} R_{ijij} \underbrace{e_i \cdot e_i}_{-1} && \text{(Eq. (16.15))} \\
&= \frac{1}{2} \sum_{i,j=1}^{n} g(R(e_i, e_j)e_j, e_i) \overset{\text{Eq. (16.13)}}{=} \frac{1}{2}\,\mathrm{Sc}.
\end{aligned}$$

□

17 Dirac operators

17.1 Atiyah–Singer–Dirac operator

Definition 63. The *Atiyah–Singer (spin) Dirac operator* on a spin manifold X is the first-order differential operator on the spinor bundle $\mathcal{S}$ defined by

$$\begin{aligned} \not{D} = \not{D}^{(X)} : \Gamma(\mathcal{S}) &\to \Gamma(\mathcal{S}) \\ \psi &\mapsto \sum_{j=1}^{n} e_j \cdot \nabla^{\mathrm{Spin}}_{e_j} \psi \end{aligned}$$

where $\{e_1, \ldots, e_n\}$ is any local oriented orthonormal frame, ∇^{Spin} is the spin covariant derivative,and $\cdot$ denotes Clifford multiplication.

To emphasize that $\not{D}$ does not depend on the local frame choice, it is sometimes formulated in a slicker way, as

$$\not{D} : \Gamma(\mathcal{S}) \overset{\nabla^{\mathrm{Spin}}}{\to} \Gamma(T^*X \otimes \mathcal{S}) \overset{\sharp}{\to} \Gamma(TX \otimes \mathcal{S}) \overset{\mu}{\to} \Gamma(\mathcal{S})$$

where $\sharp$ is the isomorphism $T^*X \cong TX$ given by the Riemannian metric g, and μ is Clifford multiplication (Definition 62).

Remark. Dirac originally used an orthonormal coordinate frame (over Minkowski space), when he wrote down his equation in 1928. In physics, one often prefers to use a *coordinate* frame $\{\partial_1, \ldots, \partial_n\}$ and writes $\not{D} = -i\gamma^\mu \nabla^{\mathrm{Spin}}_\mu$. Usually, the ∂_μ cannot be chosen to be orthonormal, so one only has an expansion in terms of "abstract" orthonormal frame vectors, $\partial_\mu = E^j_\mu e_j$, with the coefficient functions E^j_μ called the *vielbein*. Then $-i\gamma_\mu = -iE^j_\mu \gamma_j$ acts by Clifford multiplication, and

$$\not{D} = -i\gamma^\mu \nabla^{\mathrm{Spin}}_\mu \equiv -ig^{\mu\nu}\gamma_\nu \nabla^{\mathrm{Spin}}_{\partial_\mu} = g^{\mu\nu} E^k_\nu E^j_\mu e_k \cdot \nabla^{\mathrm{Spin}}_{e_j} = g(e^k, e^j) e_k \cdot \nabla^{\mathrm{Spin}}_{e_j} = e^j \cdot \nabla^{\mathrm{Spin}}_{e_j},$$

as above. Nevertheless, ∇^{Spin} is more naturally described in terms of orthonormal frames, and many calculations become simpler.

17.1.1 Even dimensional case

When n is even, the Dirac operator is an odd operator with respect to the canonical splitting $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$, due to Prop. 16.2-16.4. So we may write

$$\not{D} = \begin{pmatrix} 0 & \not{D}^- \\ \not{D}^+ & 0 \end{pmatrix}.$$

$n = 2$ **Examples.** In Example 15.4-15.6, we saw that the spin representation of Spin(2) decomposes into $\delta_2^+ \oplus \delta_2^-$, each of which is one-dimensional.

Example 17.1. On Euclidean $\mathbf{R}^2$, the orthonormal frame bundle $\mathbf{R}^2 \times \mathrm{SO}(2)$ is trivialized by the global coordinate frame $e_1 = \partial_x, e_2 = \partial_y$. Likewise, the spin frame bundle is trivialized, $\mathbf{R}^2 \times \mathrm{Spin}(2)$, with trivial spin connection. The associated spinor bundle $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$ has a flat trivialization, i.e., $\mathcal{S}^\pm \cong \mathbf{R}^2 \times \mathbb{C}$ with $\nabla^{\mathrm{Spin}}_{e_1}, \nabla^{\mathrm{Spin}}_{e_2}$ simply being ∂_x, ∂_y. Then the Dirac operator in this trivialization is Eq. (14.1).

Example 17.2. We worked out the spin frame bundle and spin connection for S^2 explicitly in Section 16.2. Use the local orthonormal frame $\mathbf{e} = \{\partial_\vartheta, \frac{1}{\sin\vartheta}\partial_\varphi\}$, and its spin lift $\tilde{\mathbf{e}}$ given there. We saw that $\tilde{\mathbf{e}}^*\omega^{\mathrm{Spin}} = \frac{\mathsf{e}_1\mathsf{e}_2}{2}\cos\vartheta\, d\varphi$, and in the spin representation, we have

$$\mathsf{e}_1 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \qquad \mathsf{e}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

So on the spinor bundle, in the *local* frame $\tilde{\mathbf{e}}$, the Dirac operator acts on $\mathbb{C} \oplus \mathbb{C}$-valued functions as

$$\not{D}^{(S^2)} = -i \begin{pmatrix} 0 & \partial_\vartheta + \frac{\cot\vartheta}{2} - \frac{i}{\sin\vartheta}\partial_\varphi \\ \partial_\vartheta + \frac{\cot\vartheta}{2} + \frac{i}{\sin\vartheta}\partial_\varphi & 0 \end{pmatrix}. \tag{17.1}$$

A point of clarification: Although the full spinor bundle $\mathcal{S}$ can be shown to be a trivializable vector bundle, the individual $\mathcal{S}^+$ and $\mathcal{S}^-$ are not trivializable. The spin frame $\tilde{\mathbf{e}}$ is only defined away from a line of longitude, and similarly for the local trivializations of the spinor bundle. *This means that Eq.* (17.1) *must not be mistaken for a differential operator on smooth* $\mathbb{C}\oplus\mathbb{C}$*-valued functions over* S^2. At best, it acts on smooth functions over the coordinate domain $(0,\pi) \times (0, 2\pi) \subset \mathbf{R}^2$, subject to certain boundary conditions. (Test your understanding of the material by figuring out what these conditions are!) Conceptually, $\not{D}^{(S^2)} : \Gamma(\mathcal{S}^\pm) \to \Gamma(\mathcal{S}^\mp)$ acts on *sections*, and its eigenspinors are likewise *sections* of $\mathcal{S}$. The gauge-independent perspective of $\not{D}$ will allow us to understand its spectrum directly from the geometric data (which, unlike the gauge choice, is not a fictitious input). This type of idea will be explored in Section 17.3.

Exercise 17.1. Equation (17.1) appears to be formally non-self-adjoint, due to the $-\frac{i}{2}\cot\vartheta$ term. Work out the Riemannian volume form vol_g (see Section 17.4.2) in the (ϑ, φ) coordinates. Verify that Eq. (17.1) is formally self-adjoint, with respect to the L^2-inner product relative to vol_g.

17.1.2 Odd dimensional case

When n is odd, recall that there are two possible spinor bundles, the left-handed $\mathcal{S}^+$ or the right-handed $\mathcal{S}^-$.

Example 17.3. Let $X \cong \mathbf{R}^1$ be the (flat) Euclidean line, with global coordinate x and orthonormal frame $\{\partial_x\}$. The spinor bundle is $\mathcal{S}^\pm = \mathbf{R}^1 \times S^\pm$, where $S^\pm = \mathbb{C}$ but with e_1 acting as $\mp i$, and the spin connection coefficients are zero in this trivialization. So the Dirac operator is

$$\not{D}^{(\mathbf{R}^1)} = \mathsf{e}_1 \cdot \frac{d}{dx} = \mp i \frac{d}{dx},$$

acting on sections regarded as $\mathbb{C}$-valued functions.

Similarly, on the unit circle $X = S^1$ with global frame $\{\partial_\theta\}$, choose $\mathrm{Fr}^{\mathrm{Spin}}(S^1) = S^1 \times \mathbb{Z}_2$. Then the spinor bundle is trivial, and the Dirac operator is $\not{D}^{(S^1)} = \mp i \frac{d}{d\theta}$. We had already encountered these 1D Dirac operators in Section 3. Later we will see how to twist $\not{D}^{(S^1)}$ by tensoring with a Hermitian line bundle with connection.

Exercise: Describe the Dirac operator on S^1 when the spin structure is instead the connected double cover of S^1.

17.1.3 Relativistic version (Optional)

Physicists say "$1 + n$" dimensions to indicate that they are working in a Lorentzian spacetime manifold with n spatial dimensions (usually $n = 3$). The metric g is semi-Riemannian, meaning that on each tangent space it has mixed signature $- + ++$. A "timelike" coordinate tangent vector has negative length and is usually written ∂_0, while the "spacelike" coordinate tangent vectors are written ∂_i as usual. Let us illustrate the (local) Dirac operator construction in 1+1 dimensions, with flat metric.

On $V \cong \mathbb{R}^2$ with $(-,+)$-signature quadratic form q, pick any orthonormal basis $\{\mathsf{e}_0, \mathsf{e}_1\}$, where $q(\mathsf{e}_0) = -1$ and $q(\mathsf{e}_1) = +1$. So as Clifford algebra elements, we have $\mathsf{e}_0^2 = 1$ while $\mathsf{e}_1^2 = -1$. When we complexify, we still have $Cl(V,q) \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C}l_2 \cong M_2(\mathbb{C})$, which has the usual Clifford representation on $\mathbb{C}^2$; for example,

$$\mathsf{e}_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathsf{e}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \omega^{\mathbb{C}} = \mathsf{e}_1\mathsf{e}_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The *Dirac equation* with mass $m \geq 0$ is

$$(i\mathsf{e}_0\partial_0 + i\mathsf{e}_1\partial_1 \mp m)\psi = 0.$$

The sign in front of the mass term is indeterminate (another torsor!). The main point is that the anticommutation of $\mathsf{e}_0, \mathsf{e}_1$ leads to

$$(i\mathsf{e}_0\partial_0 + i\mathsf{e}_1\partial_1 + m)(i\mathsf{e}_0\partial_0 + i\mathsf{e}_1\partial_1 - m) = (-\partial_0^2 + \partial_1^2) - m^2,$$

which upon writing

$$E = -i\partial_0, \quad \mathbf{p} = -i\partial_1,$$

is the relativistic energy-momentum-mass relation, $E^2 - \mathbf{p}^2 = m^2$.

To recast the relativistic Dirac equation into the (apparently) non-relativistic Schrödinger equation form with a self-adjoint Hamiltonian operator, one multiplies throughout by e_0, to get $i\partial_0\psi = (i\mathsf{e}_1\mathsf{e}_0\partial_1 + m\mathsf{e}_0)\psi$, or after restoring $\partial_0 = \partial_t, \partial_1 = \partial_x$,

$$i\frac{\partial}{\partial t} = \underbrace{\begin{pmatrix} -i\frac{\partial}{\partial x} & m \\ m & i\frac{\partial}{\partial x} \end{pmatrix}}_{\text{massive Dirac Hamiltonian}} .$$

Notice that $m = 0$ is special — the Dirac Hamiltonian splits into independent right-handed and left-handed parts, $\mp i\partial_x$, according to the chirality operator $\omega^{\mathbb{C}}$. Physicists call these the (massless) *Weyl Hamiltonians*.

Remark. We have $\mathrm{SO}(1,1)$ being a *noncompact* Lie group with two connected components, coordinatized as $\begin{pmatrix} \pm\cosh\theta & \sinh\theta \\ \sinh\theta & \pm\cosh\theta \end{pmatrix}, \theta \in \mathbb{R}$. Similarly, $\mathrm{SO}(1,n)$ is non-compact and not connected. Usually one considers the double-cover $\mathrm{Spin}^+(1,n)$ of the component connected to the identity. A more important complication is that $\mathrm{Spin}^+(1,n)$ will not be unitarily represented on the (finite-dimensional) spinor space S. Instead, one defines a certain "Dirac conjugate" of a spinor, $\bar{\psi}$, such that $\bar{\psi}\psi \in \mathbb{C}$ is $\mathrm{Spin}^+(1,n)$-invariant. Why this makes sense physically is a tricky story in quantum field theory, which we do not pursue further.

17.2 Laplacians on vector bundles with connection

In flat Euclidean space, we have a global orthonormal frame, and the Dirac operator is easily seen to square to the Laplacian tensored with an identity

matrix. The generalization of $\not{D}$ to curved manifolds then required a deeper understanding of its geometric ingredients. But what is the fate of "$\not{D}^2$ = Laplacian"? What does "Laplacian" mean on a general vector bundle?

Definition 64 (Connection Laplacian)**.** Let X be a Riemannian manifold, and $E \to X$ be a vector bundle with covariant derivative ∇. For a pair of tangent vector fields $u, v \in \mathfrak{X}(X)$, the *second (covariant) derivative* on $\Gamma(E)$ is defined to be

$$\nabla^2_{u,v} = \nabla_u \nabla_v - \nabla_{\nabla^{\mathrm{LC}}_u v}.$$

The *connection Laplacian* on $\Gamma(E)$ is defined to be the "contracted second covariant derivative",

$$\Delta \equiv \Delta^E := -\sum_{j=1}^{n} \nabla^2_{e_j,e_j}, \tag{17.2}$$

where $\{e_j\}_{j=1,\dots,n}$ is a(ny) local orthonormal tangent frame. If E is a Hermitian/Euclidean vector bundle and ∇ is a metric connection, we write $\nabla^*\nabla$ in place of Δ.

Regarding the notation $\nabla^*\nabla$, this is meant to indicate that the connection Laplacian can also be obtained by composing ∇ with its adjoint operator (see Prop. 17.3).

Remark. Why does Eq. (17.2) make sense? First, note that the order of u, v in $\nabla^2_{\cdot,\cdot}$ matters, and this dependence is measured by the curvature,

$$\begin{aligned}
\nabla^2_{u,v} - \nabla^2_{v,u} &= \nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{\nabla^{\mathrm{LC}}_u v - \nabla^{\mathrm{LC}}_v u} \\
&= \nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]} && (\nabla^{\mathrm{LC}} \text{ torsion-free}) \\
&= F^{\nabla}(u, v). && (17.3)
\end{aligned}$$

Using this, we may deduce that $\nabla^2_{u,v}$ is tensorial in the arguments u, v (Exercise). So in Eq. (17.2), if we replace e_j by $e'_j = \sum_k e_k h_{kj}$ with h an $\mathrm{O}(n)$-valued transformation, then

$$\sum_{j=1}^{n} \nabla^2_{e'_j,e'_j} = \sum_{j,k,l=1}^{n} h_{kj} h_{lj} \nabla^2_{e_k,e_l} = \sum_{k,l=1}^{n} \underbrace{(hh^{\mathrm{t}})_{kl}}_{\delta_{kl}} \nabla^2_{e_k,e_l} = \sum_{k=1}^{n} \nabla^2_{e_k,e_k}.$$

Sometimes, the connection Laplacian is defined invariantly as

$$\Delta^E = -\mathrm{Tr}_g(\nabla^2_{\cdot,\cdot}),$$

where Tr_g is the trace with respect to the Riemannian metric g. If coordinate tangent frames are preferred, then we can also write

$$\Delta^E = -g^{ij}(\nabla^2_{\partial_i,\partial_j}).$$

Example 17.4. Take $E = X \times \mathbb{K}$ with the trivial connection, so ψ may be regarded as an ordinary scalar field $X \to \mathbb{K}$. In this case, the connection Laplacian gives one of several equivalent definitions of the Laplace–Beltrami operator Δ acting on functions.

Example 17.5. Let ∇ be a unitary connection on a Hermitian line bundle over X. Then $\Delta = \nabla^*\nabla$ is the *magnetic Laplacian*, or *Landau Hamiltonian*, with the curvature 2-form F^∇ being the magnetic field. This describes a spinless electron field coupled to the "magnetic vector potential" ∇. We will study this operator in detail in Section 18.

Example 17.6. For $\nabla = \nabla^{\mathrm{Spin}}$ the spin covariant derivative on a spinor bundle, $\Delta^{\mathcal{S}} = \nabla^*\nabla$ is the *spinor Laplacian*. When $\dim X$ is even, Prop. 16.4 shows that ∇^{Spin} can be restricted to $\Gamma(\mathcal{S}^+)$ and $\Gamma(\mathcal{S}^-)$ separately. So there are actually two "'reduced" spinor Laplacians, $\Delta^{\mathcal{S}^+}$ and $\Delta^{\mathcal{S}^-}$.

17.3 Schrödinger–Peres–Lichnerowicz identity

On a spin manifold, we also have the "Dirac Laplacian" $\not{D}^2$ acting on $\Gamma(\mathcal{S})$, *which does not come from the connection Laplacian construction.* In particular, it does not generally coincide with the connection Laplacian $\Delta^{\mathcal{S}}$ for the spin covariant derivative.

Theorem 17.1 (Lichnerowicz, 1963)**.** *Let X be a spin manifold, and $\nabla = \nabla^{\mathrm{Spin}}$ be the spin covariant derivative on its spinor bundle. Then*

$$\not{D}^2 - \Delta^{\mathcal{S}} = \frac{\mathrm{Sc}}{4}, \tag{17.4}$$

where Sc *is the scalar curvature function on X.*

Proof. At each fixed $x \in X$, we compute in a local oriented orthonormal frame

$\{e_1, \dots, e_n\}$ satisfying $\nabla^{\mathrm{LC}} e_i(x) = 0$ (Lemma 13.7),

$$\begin{aligned}
\not{D}^2 &= \sum_{j,k=1}^{n} e_j \cdot \nabla_{e_j}(e_k \cdot \nabla_{e_k}) \\
&= \sum_{j,k=1}^{n} e_j \cdot e_k \cdot \nabla_{e_j} \nabla_{e_k} && \text{(Prop. 16.3)} \\
&= -\sum_{j=1}^{n} \nabla_{e_j} \nabla_{e_j} + \sum_{1 \le j \ne k \le n} e_j \cdot e_k \cdot \nabla^2_{e_j, e_k} \\
&= \Delta^{\mathcal{S}} + \frac{1}{2} \sum_{1 \le j \ne k \le n} e_j \cdot e_k \cdot (\nabla^2_{e_j, e_k} - \nabla^2_{e_k, e_j}) && \text{(antisymmetry)} \\
&= \Delta^{\mathcal{S}} + \frac{1}{2} \sum_{j,k=1}^{n} e_j \cdot e_k \cdot \mathcal{R}(e_j, e_k) && \text{(Eq. (17.3))} \\
&= \Delta^{\mathcal{S}} + \frac{\mathrm{Sc}}{4}. && \text{(Lemma 16.6)}
\end{aligned}$$

□

Next, let us "twist" $\mathcal{S}$ by tensoring it with an auxiliary Hermitian line bundle $\mathcal{L}$ with unitary connection $\nabla^{\mathcal{L}}$. Then $\mathcal{S} \otimes \mathcal{L}$ is called a *twisted spinor bundle*, and it is equipped with a *tensor product connection/covariant derivative*,

$$\nabla^{\mathrm{Spin},\mathcal{L}}(\psi \otimes \varphi) = (\nabla^{\mathrm{Spin}} \psi) \otimes \varphi + \psi \otimes (\nabla^{\mathcal{L}} \varphi), \qquad \psi \in \Gamma(\mathcal{S}),\ \varphi \in \Gamma(\mathcal{L}).$$

As in Prop. 16.3, $\nabla^{\mathrm{Spin},\mathcal{L}}$ is compatible with ∇^{LC} and Clifford multiplication (on the $\mathcal{S}$ factor).

The *twisted Dirac operator* on the twisted spinor bundle is defined as

$$\not{D}_{\mathcal{L}} = \not{D}^{(X)}_{\mathcal{L}} = \sum_{j=1}^{n} e_j \cdot \nabla^{\mathrm{Spin},\mathcal{L}}_{e_j},$$

with $\{e_1, \dots, e_n\}$ a(ny) local oriented orthonormal frame.

Corollary 17.2. *Let X be a spin manifold, and ∇^{Spin} be the spin covariant derivative on its spinor bundle $\mathcal{S}$. Let $\mathcal{L}$ be a Hermitian line bundle with unitary connection $\nabla^{\mathcal{L}}$, and let $\nabla^{\mathrm{Spin},\mathcal{L}}$ be the tensor product connection on*

$\mathcal{S} \otimes \mathcal{L}$. *The twisted spinor Laplacian* $\Delta^{\mathcal{S}\otimes\mathcal{L}}$ *and the twisted Dirac Laplacian* $\not{D}_{\mathcal{L}}^2$ *are related by the geometric identity*

$$\not{D}_{\mathcal{L}}^2 - \Delta^{\mathcal{S}\otimes\mathcal{L}} = \frac{\mathrm{Sc}}{4} + \frac{1}{2}F^{\nabla^{\mathcal{L}}}, \tag{17.5}$$

where $F^{\nabla^{\mathcal{L}}}$ *is the curvature of* $\nabla^{\mathcal{L}}$ *acting by Clifford multiplication,*

$$F^{\nabla^{\mathcal{L}}}(\psi \otimes \varphi) = \sum_{j,k=1}^{n} F^{\nabla^{\mathcal{L}}}(e_j, e_k) e_j \cdot e_k \cdot \psi \otimes \varphi.$$

Proof. The calculation is mostly identical to that in Theorem 17.1,

$$\begin{aligned}\not{D}_{\mathcal{L}}^2(\psi \otimes \varphi) &= \Delta^{\mathcal{S}\otimes\mathcal{L}}(\psi \otimes \varphi) + \frac{1}{2}\sum_{j,k=1}^{n} e_j \cdot e_k \cdot (\mathcal{R}(e_j, e_k)\psi \otimes \varphi + \psi \otimes F^{\nabla^{\mathcal{L}}}(e_j, e_k)\varphi) \\ &= \left(\Delta^{\mathcal{S}\otimes\mathcal{L}} + \frac{\mathrm{Sc}}{4} + \frac{1}{2}\sum_{j,k=1}^{n} F^{\nabla^{\mathcal{L}}}(e_j, e_k) e_j \cdot e_k\right)(\psi \otimes \varphi),\end{aligned}$$

where in the last line, we used the fact that $F^{\nabla^{\mathcal{L}}}(e_j, e_k)$ is a $\mathbb{C}$-valued function, so it can be commuted through the tensor product. □

Historical note. Schrödinger had in fact already derived a Lorentzian signature version of Eq. (17.5) with a mass term in 1932. In 1962, A. Peres rederived a Lorentzian version, with no mass term, as did Lichnerowicz independently.

Remark (Optional). The operator $\not{D}_{\mathcal{L}}$ is a special case of a so-called Spin^c-Dirac operator. The latter requires X to be a Spin^c-manifold, which is a slight weakening of the Spin condition.

17.4 Formal self-adjointness of Dirac operators and Laplacians

17.4.1 Integration on manifolds

To integrate a function $f : X \to \mathbb{K}$ on an n-manifold X, we might try doing so in a coordinate chart (U, φ_α),

$$\text{“}\int_U f\text{”} = \int_{\varphi_\alpha(U)} (f \circ \varphi_\alpha^{-1})\, dx^1 \dots dx^n \equiv \int_{\varphi_\alpha(U)} \check{f}_\alpha\, dx^1 \dots dx^n,$$

and integrate the Euclidean space function $\check{f}_\alpha = f \circ \varphi_\alpha^{-1}$. However, if we switch to another coordinate chart (U, φ_β) with coordinates $(y^1, \ldots, y^n)$, and write $T = \varphi_\alpha \circ \varphi_\beta^{-1}$ for the change-of-coordinate function, then we get an inconsistency,

$$\begin{aligned}\int_{\varphi_\alpha(U)} \check{f}_\alpha \, dx^1 \ldots dx^n &= \int_{\varphi_\beta(U)} \underbrace{(f \circ \varphi_\alpha^{-1} \circ T)}_{f \circ \varphi_\beta^{-1} = \check{f}_\beta} \cdot |\det(dT)| \cdot dy^1 \ldots dy^n \\ &\neq \int_{\varphi_\beta(U)} \check{f}_\beta \, dy^1 \ldots dy^n.\end{aligned}$$

For coordinate n-forms, we have the transformation law (exercise)

$$dx^1 \wedge \ldots \wedge dx^n|_x = \det(dT)|_{\varphi_\beta(x)} \, dy^1 \wedge \ldots \wedge dy^n|_x.$$

If X is *oriented*, we define integration of a *compactly-supported* n-form $\eta = f \, dx^1 \wedge \ldots \wedge dx^n$ as

$$\int_U \eta := \int_{\varphi_\alpha(U)} \check{f}_\alpha \, dx^1 \ldots dx^n, \tag{17.6}$$

with respect to any positively-oriented coordinate chart (U, φ_α). Because

$$\eta|_x = f(x) \, dx^1 \wedge \ldots \wedge dx^n|_x = f(x) \, \det(dT)|_{\varphi_\beta(x)} dy^1 \wedge \ldots \wedge dy^n|_x,$$

if we switch to the (positively-oriented) y^j coordinates instead, we would get

$$\begin{aligned}\int_U \eta &\equiv \int_U f \, \det(dT) \, dy^1 \wedge \ldots \wedge dy^n \\ &= \int_{\varphi_\beta(U)} \check{f}_\beta \det(dT) \, dy^1 \ldots dy^n \\ &= \int_{\varphi_\alpha(U)} \check{f}_\alpha \, \det(dT) \det(dT)^{-1} \, dx^1 \ldots dx^n \\ &= \int_{\varphi_\alpha(U)} \check{f}_\alpha \, dx^1 \ldots dx^n,\end{aligned}$$

consistent with Eq. (17.6). Thus local integration of n-forms is coordinate-independent.

To integrate a compactly-supported n-form over all of X, we pick a partition-of-unity $\{\rho_\alpha\}_{\alpha \in \mathcal{I}}$ subordinate to any open cover $\{U_\alpha\}_{\alpha \in \mathcal{I}}$ by coordinate charts, then define

$$\int_X \eta := \sum_{\alpha \in \mathcal{I}} \int_{U_\alpha} \rho_\alpha \cdot \eta.$$

17.4.2 Riemannian volume form

To make sense of the function space $L^2(X)$, we need an L^2-type inner product, $\langle f|h\rangle_{L^2(X)} =$ "$\int_X \overline{f}h$". But we had just seen that the right side does not make coordinate-invariant sense, unless we have a prescription for converting functions to n-forms. For example, we can pick some nowhere-vanishing n-form, called vol, and define $\langle f|h\rangle_{L^2(X)} = \int_X \overline{f}h \,\mathrm{vol}$ (this requires X to be orientable). Such a "volume form" is canonically provided when X is a Riemannian manifold.

Definition 65. Let (X, g) be an oriented Riemannian n-manifold. Its *(Riemannian) volume form* is given in a positively-oriented coordinate chart as

$$\mathrm{vol} \equiv \mathrm{vol}_{X,g} := \sqrt{\det g}\, dx^1 \wedge \ldots \wedge dx^n,$$

where $g = \sum_{ij=1}^n g_{ij} dx^i \otimes dx^j$ and $\det g := \det(g_{ij})_{ij}$. The Hilbert space $L^2(X, g)$ is obtained by completing the space of compactly-supported smooth functions with respect to the L^2-inner product

$$\langle f|h\rangle_{L^2(X,g)} := \int_X \overline{f}h \,\mathrm{vol}.$$

It is straightforward to check that vol is independent of the coordinate chart, so it is indeed a well-defined nowhere-vanishing n-form on X (exercise). We may also regard vol as defining a measure μ_g on X, so that we may integrate in the measure-theoretic sense. In this latter sense, $L^2(X, g)$ comprises the (equivalence classes of) μ_g-square-integrable measurable functions. We often drop explicit reference to g and just write $L^2(X)$.

Now suppose E is a Hermitian vector bundle over (X, g). Then for two smooth sections ψ, φ, we have $\langle \psi, \varphi\rangle_E$ being a $C^\infty(X)$ function given by the pointwise inner products. In particular, for compactly-supported sections, we can integrate to obtain the $\mathbb{C}$-valued inner product

$$\langle \psi|\varphi\rangle_{L^2(X;E)} = \int_X \langle \psi, \varphi\rangle_E \,\mathrm{vol},$$

and complete this to a Hilbert space $L^2(X; E)$ of square-integrable sections.

17.4.3 Formal self-adjointness

Proposition 17.3. *Let ∇ be a unitary connection on a Hermitian vector bundle E over an oriented Riemannian manifold X. The connection Laplacian Δ^E is formally self-adjoint on the compactly-supported smooth sections of E.*

Proof. Fix any $x \in X$, and work in a local orthonormal tangent frame with $\nabla^{\mathrm{LC}} e_i = 0$ at x. For compactly supported $\psi, \tilde{\psi} \in \Gamma(E)$, we calculate that at x,

$$\begin{aligned}
\langle \Delta^E \psi, \tilde{\psi} \rangle_E &= -\sum_{i=1}^n \langle \nabla_{e_i} \nabla_{e_i} \psi, \tilde{\psi} \rangle_E && (\nabla^{\mathrm{LC}} e_i(x) = 0) \\
&= \sum_{i=1}^n \Big(-e_i \langle \nabla_{e_i} \psi, \tilde{\psi} \rangle_E + \langle \nabla_{e_i} \psi, \nabla_{e_i} \tilde{\psi} \rangle_E \Big) && (\text{metric connection (13.6)})
\end{aligned}$$

The first sum is the divergence of a certain tangent vector field (this is independent of the choice of frame), so it integrates to zero due to $\partial X = \emptyset$. [We omit the Riemannian geometry treatment of div, grad, curl and Stokes' theorem.] The second sum can be written as the inner product of the E-valued 1-forms $\nabla\psi, \nabla\tilde{\psi}$ at x. Altogether, we have

$$\int_X \langle \Delta^E \psi, \tilde{\psi} \rangle_E \,\mathrm{vol} = \int_X \langle \nabla\psi, \nabla\tilde{\psi} \rangle_{T^*X \otimes E} \,\mathrm{vol}.$$

A similar argument gives

$$\int_X \langle \psi, \Delta^E \tilde{\psi} \rangle_E \,\mathrm{vol} = \int_X \langle \nabla\psi, \nabla\tilde{\psi} \rangle_{T^*X \otimes E} \,\mathrm{vol},$$

showing that Δ^E is formally self-adjoint. □

The above calculation also shows that Δ^E is formally the composition of $\nabla : \Gamma(E) \to \Omega^1(X; E)$ with its formal adjoint $\nabla^* : \Omega^1(X; E) \to \Gamma(E)$, justifying the notation $\Delta^E = \nabla^* \nabla$.

Notice that classical geometric derivatives like d and ∇ turn sections into section-valued 1-forms. We need to compose it with the adjoint operator to return to the original space of sections. Upon doing so, we obtain a formally self-adjoint operator, and after making it genuinely self-adjoint, the spectral problem makes sense. This is why Laplace-type operators are very common in

geometric analysis and classical physics. From this viewpoint, the *first-order* Dirac operator achieves the remarkable feat of being formally self-adjoint *on its own.*

Proposition 17.4. *The twisted Dirac operator $\not{D}_{\mathcal{L}}$ is formally self-adjoint on compactly supported smooth sections of $\mathcal{S} \otimes \mathcal{L}$.*

Proof. Write $\nabla = \nabla^{\mathrm{Spin},\mathcal{L}}$ for convenience. Similar to the proof of Prop. 17.3, we calculate at x,

$$\begin{aligned}
\langle \not{D}_{\mathcal{L}}\psi, \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}} &= \sum_{i=1}^{n} \langle e_i \cdot \nabla_{e_i}\psi, \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}} \\
&= \sum_{i=1}^{n} -\langle \nabla_{e_i}\psi, e_i \cdot \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}} && (e_i \cdot \text{ skew}) \\
&= \sum_{i=1}^{n} \Big(-e_i\langle \psi, e_i \cdot \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}} + \langle \psi, \nabla_{e_i} e_i \cdot \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}}\Big) && (\text{metric connection}) \\
&= \sum_{i=1}^{n} -e_i\langle \psi, e_i \cdot \tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}} \\
&\quad + \underbrace{\sum_{i=1}^{n} \langle \psi, e_i \cdot \nabla_{e_i}\tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}}}_{\langle \psi, \not{D}_{\mathcal{L}}\tilde{\psi}\rangle_{\mathcal{S}\otimes\mathcal{L}}}. && (\text{compatibility and } \nabla^{\mathrm{LC}} e_i = 0)
\end{aligned}$$

The first sum is a divergence of some compactly-supported vector field (details omitted), so its integral over X vanishes.

□

Remark. We will often encounter manifolds such as the half-line $(0, \infty)$. When considered as a subset of $\mathbb{R}$, the manifold $(0, \infty)$ has a "boundary" at 0, and we often talk about "boundary conditions". For the formal self-adjointness results, the compact-support condition for sections over $(0, \infty)$ means, in particular, that the sections are "supported away from the boundary", thus they vanish there.

18 Quantum Hall Effect

The quantum-mechanical spinless electron on a two-dimensional sample (embedded in a laboratory Euclidean $\mathbf{R}^3$), and subjected to a magnetic field, exhibits astonishing behaviour known as the *quantum Hall effect* (QHE). In the recent literature, it has become quite popular to call the QHE a "topological phase" phenomenon. Unfortunately, this terminology is not accurate, and often leads a newcomer (especially the enthusiastic topologist) to imagine that the effect has something to do with, e.g., the genus of the sample. Instead, it has much more to do with geometry and analysis, and stability under deformation of this data. The latter deformation-invariance is loosely thought of as "topology" in physics.

18.1 Landau Hamiltonian on a surface

On an oriented Riemannian 3-manifold (such as $\mathbf{R}^3$), there is a (Hodge) duality between tangent vector fields and 2-forms. It is customary to think of a classical "magnetic field $\mathbf{B}$" as a 2-form $\mathcal{B}$, and the divergence-free condition, $\nabla \cdot \mathbf{B} = 0$, corresponds to $d\mathcal{B} = 0$.

Let X be a connected, embedded 2-submanifold in $\mathbf{R}^3$, assumed to be oriented. Let g be the restricted Riemannian metric on X, so (X, g) has a Riemannian volume 2-form $\mathrm{vol} \equiv \mathrm{vol}_{X,g}$. The restriction of the magnetic 2-form $\mathcal{B}$ to X, which we still denote by $\mathcal{B}$, can therefore be written as

$$\mathcal{B} = B \cdot \mathrm{vol}, \qquad B \in C^\infty(X; \mathbb{R}).$$

Note that $\mathcal{B}$ is automatically closed, since X is two-dimensional. Also, the function B represents the *intrinsic* magnetic field strength felt by an electron confined to X (and unaware of the embedding of X in $\mathbf{R}^3$). On a 2-sphere, for instance, an intrinsically constant magnetic vector field would point outwards perpendicularly to the sphere, and this is not extrinsically constant.

By itself, the free spinless electron dynamics on X is governed by the Laplace–Beltrami operator Δ on (scalar) functions $X \to \mathbb{C}$. As we have learned, the magnetic field is (i-times) the curvature of a unitary connection ∇ on a Hermitian line bundle $\mathcal{L} \to X$,

$$F^\nabla = -i\mathcal{B} = -iB \cdot \mathrm{vol}.$$

In physics, the formally self-adjoint magnetic Laplacian $\Delta^{\mathcal{L}} = \nabla^*\nabla$ on $\mathcal{L}$ is called the Landau Hamiltonian, and it governs the spinless quantum-mechanical electron subjected to the magnetic field $\mathcal{B}$.

The Landau Hamiltonian depends not only on the classical magnetic field $\mathcal{B}$, but on the connection ∇. Two simplifications often occur, rendering the choice of ∇ unimportant:

1. The line bundle $\mathcal{L}$ is trivializable, so we can find some global U(1)-frame such that ∇ is represented as a global $\mathfrak{u}(1)$-valued 1-form $\mathcal{A}$ satisfying

$$d\mathcal{A} = F^{\nabla} = -i\mathcal{B}.$$

   This happens, e.g., if the 2-manifold X is non-compact so its second cohomology group vanishes (Omitted; this group classifies the complex line bundles over X.)

2. Suppose $\tilde{\nabla}$ is another connection on the trivializable $\mathcal{L}$ with the same curvature, $F^{\tilde{\nabla}} = F^{\nabla}$. Let $\tilde{\mathcal{A}}$ be the representative $\mathfrak{u}(1)$-valued 1-form (with respect to the global gauge from earlier). Then

$$d(\tilde{\mathcal{A}} - \mathcal{A}) = F^{\tilde{\nabla}} - F^{\nabla} = 0.$$

   Thus $\eta = -i(\tilde{\mathcal{A}} - \mathcal{A})$ is a closed 1-form. If X is simply-connected, then its first cohomology group vanishes, and this implies that $\eta = d\Lambda$ for some 0-form Λ (i.e. function). So in this case,

$$\tilde{\mathcal{A}} = \mathcal{A} + i\eta = \mathcal{A} + e^{-i\Lambda}d(e^{i\Lambda}),$$

   showing that there is a globally-defined gauge transformation which converts ∇ into $\tilde{\nabla}$. Thus all connections on $\mathcal{L}$ with the same curvature are actually gauge equivalent. Up to unitary equivalence, the Landau Hamiltonian depends only on the curvature form.

If the above simplifications occur, we will denote the Landau Hamiltonian and the line bundle $\mathcal{L}$ by

$$H_B := \Delta^{\mathcal{L}}, \qquad \mathcal{L} \equiv \mathcal{L}_B,$$

which, up to unitary gauge equivalence, is unambiguous.

Remark. Recall that condition 2 fails in the Aharonov–Bohm effect.

Experimentally, the quantum Hall effect refers to the quantization of the measured values of a macroscopic transport quantity, called the *transverse Hall conductance*. A priori, this conductance is real-valued, and "quantization" means that only *integer* multiples of some "indivisible" conductance are observed. Furthermore, the quantization effect persists under various kinds of perturbations.

In order to non-vacuously *explain* how this quantization occurs, we cannot *assume* from the beginning that it is correlated with an integer-valued topological invariant. Instead, the *spectrum* of the Landau Hamiltonian is supposed to determine the aforementioned Hall conductance. The non-trivial work is to explain and prove how the quantization occurs, and why it is stable under a large variety of perturbations. One aspect of the stability concerns something called *Anderson localization* of the spectrum under the introduction of disorder, but we will not discuss this.

18.2 Euclidean plane Landau Hamiltonian

Consider the Euclidean plane $X = \mathbf{R}^2$ with coordinates (x, y), Riemannian volume form $\mathrm{vol} = dx \wedge dy$, and let $B = B(x, y) = b$ be a nonzero constant function ($b \in \mathbb{R}$). This describes a spinless electron on the plane, subject to an external magnetic field of *uniform* strength b perpendicular to the plane. This is the highly-simplified geometry for a typical quantum Hall sample.

Since the plane is contractible, there is a global gauge such that

$$\mathcal{A} = -ibx\, dy \qquad \Rightarrow \quad d\mathcal{A} = F^{\nabla} = -ib \cdot \mathrm{vol}.$$

This is called *Landau gauge* in physics. In this gauge, sections ψ of $\mathcal{L}_b$ are identified as functions $X \to \mathbb{C}$, and the covariant derivatives $\nabla_{\partial_x}, \nabla_{\partial_y}$ act as

$$\nabla_{\partial_x} = \partial_x, \qquad \nabla_{\partial_y} = \partial_y - ibx.$$

The *Landau Hamiltonian in Landau gauge* is the operator

$$H_b = -\nabla^2_{\partial_x,\partial_x} - \nabla^2_{\partial_y,\partial_y} = -\nabla_{\partial_x}\nabla_{\partial_x} - \nabla_{\partial_y}\nabla_{\partial_y} = -\partial_x^2 - (\partial_y - ibx)^2. \quad (18.1)$$

18.2.1 Spectrum of Euclidean Landau Hamiltonian.

Direct method. The operator H_b can be shown to be essentially self-adjoint on compactly-supported smooth functions, and we close it to a self-adjoint

operator (see Section 19.3.1). This operator is a remarkable example of an *exactly solvable* Hamiltonian. In fact, its spectrum was already computed by Landau in 1930, long before the discovery of the quantum Hall effect. Let us outline the calculation.

First, observe that H_b is translation invariant in the y-direction. So we may perform a Fourier transform, converting ∂_y into ip_y,

$$H_b = \int_{p_y \in \widehat{\mathbb{R}}}^{\oplus} H_b(p_y), \qquad H_b(p_y) = -\frac{d^2}{dx^2} + (bx - p_y)^2.$$

For each fixed $p_y \in \widehat{\mathbb{R}}$, we have a simple harmonic oscillator centered at $x = \frac{p_y}{b}$. So all the $H_b(p_y)$ are isospectral, and it suffices to work out the spectrum for the $p_y = 0$ case. By a scaling substitution, $\tilde{x} = |b|^{1/2}x$, we get

$$H_b(0) = |b| \left(-\frac{d^2}{d\tilde{x}^2} + \tilde{x}^2 \right) = |b| H_{\mathrm{SHO}},$$

where H_{SHO} is the simple harmonic oscillator Hamiltonian. In Section 3.4, we had already found the spectrum of H_{SHO} to be $2\mathbb{N} + 1$.

Therefore, we conclude that

$$\sigma(H_b) = (2\mathbb{N} + 1)|b|.$$

The eigenvalue $(2n + 1)|b|$ is called the n-th *Landau level*, and it is *infinitely degenerate*. Contrast this with the $b = 0$ case, where the ordinary Laplacian $H_0 = -(\partial_x^2 + \partial_y^2)$ has continuous spectrum $[0, \infty)$. The dramatic qualitative change of the spectrum induced by the mangetic field is called *Landau quantization*, and was first calculated in 1930 by Landau, half a century before the *unexpected* experimental discovery of the quantum Hall effect by von Klitzing!

Geometric, gauge-invariant method. On the (flat) Euclidean plane, the spinor bundle $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$ has $\mathcal{S}^\pm$ being trivializable line bundles with flat connection, $F^{\nabla^{\mathrm{Spin}}} = 0$. Therefore, each twisted spinor line bundle $\mathcal{S}^\pm \otimes \mathcal{L}_b$ has connection with curvature $-ib \cdot \mathrm{vol}$ coming purely from $\mathcal{L}_b$. In other words, the connection Laplacian on $\mathcal{S} \otimes \mathcal{L}_b$ is (unitarily gauge-equivalent to) two copies of a Landau Hamiltonian,

$$\Delta^{\mathcal{S} \otimes \mathcal{L}_b} = \begin{pmatrix} \Delta^{\mathcal{S}^+ \otimes \mathcal{L}_b} & 0 \\ 0 & \Delta^{\mathcal{S}^- \otimes \mathcal{L}_b} \end{pmatrix} \cong \begin{pmatrix} \Delta^{\mathcal{L}_b} & 0 \\ 0 & \Delta^{\mathcal{L}_b} \end{pmatrix} = \begin{pmatrix} H_b & 0 \\ 0 & H_b \end{pmatrix}.$$

On $\mathcal{S} \otimes \mathcal{L}_b$, we also have the formally self-adjoint twisted Dirac operator,

$$\not{D}_{\mathcal{L}_b} = \begin{pmatrix} 0 & \not{D}_- \\ \not{D}_+ & 0 \end{pmatrix}, \qquad \not{D}_- = \not{D}_+^*.$$

Apply the Lichnerowicz identity, Corollary 17.2, with $e_1 = \partial_x, e_2 = \partial_y$, and $\mathrm{Sc} = 0$,

$$\begin{aligned} \not{D}_{\mathcal{L}_b}^2 = \begin{pmatrix} \not{D}_+^* \not{D}_+ & 0 \\ 0 & \not{D}_+ \not{D}_+^* \end{pmatrix} &= \Delta^{\mathcal{S}\otimes\mathcal{L}_b} - ib\,\mathrm{vol}(e_1, e_2) e_1 \cdot e_2 \cdot \\ &= \begin{pmatrix} H_b - b & 0 \\ 0 & H_b + b \end{pmatrix}. \end{aligned} \tag{18.2}$$

Here, we recall that $ie_1 \cdot e_2 \cdot$ acts as ± 1 on $\mathcal{S}^\pm$.

Before proceeding, we mention that $\not{D}_{\mathcal{L}_b}, \not{D}_{\mathcal{L}_b}^2$ and $\Delta^{\mathcal{S}\otimes\mathcal{L}_b}$ are essentially self-adjoint on the compactly-supported smooth sections (see Section 19.3.1), so that Eq. (18.2) holds at the level of self-adjoint operators. We may now run the same inductive supersymmetric argument used in Section 3.4.

Suppose $b > 0$. We have $H_b - b \geq 0$, so $H_b + b \geq 2b$, and $(0, 2b)$ is a spectral gap for the bottom right operator. By spectral supersymmetry, the same gap appears for the top left operator $H_b - b$. So $H_b + b$ has spectral gap $(2b, 4b)$, and again this gap appears for $H_b - b$. Inductively, we see that $\sigma(H_b - b) \subset 2b\mathbb{N}$. If $\ker(H_b - b) = 0$, then $H_b + b$ does not have $2b$ in its spectrum, and neither does $H_b - b$, and inductively, we would deduce that $H_b - b$ has empty spectrum (which is not possible). So $\ker(H_b - b)$ must be nontrivial, and inductively, we get

$$\sigma(H_b - b) = 2b\mathbb{N} \;\Leftrightarrow\; \sigma(H_b) = (2\mathbb{N} + 1)b.$$

The $b < 0$ case is similarly handled, by rewriting $H_b \pm b$ as $H_b \mp |b|$.

Remark (Optional.). The zero-eigenspace $\ker(H_b - b)$ plays a crucial role. We have

$$\begin{aligned} \ker(H_b - b) = \ker((H_b - b) \oplus (H_b + b)) &= \ker(\not{D}_{\mathcal{L}_b}^2) \\ &= \ker(\not{D}_{\mathcal{L}_b}) = \begin{cases} \ker(\not{D}_+), & b > 0, \\ \ker(\not{D}_-), & b < 0. \end{cases} \end{aligned}$$

We may define the "index" of $\not{D}_{\mathcal{L}_b}$ to be the formal difference

$$\text{“ Index}(\not{D}_{\mathcal{L}_b}) = \dim\ker(\not{D}_+) - \dim\ker(\not{D}_-)\text{”}.$$

This index does not immediately make sense as an integer number, because the kernel dimensions are infinite. However, it is possible to understand this "Dirac index" as an object in *operator K-theory*. Furthermore, this K-theoretic index is extremely stable against *geometric* perturbations. Substantially more mathematical background is needed to understand this, but we will attempt to illustrate the idea by considering the non-geometrically idealized Landau Hamiltonian.

18.3 Qualitative spectrum of general Landau Hamiltonian

As mentioned, the real interest in the quantum Hall effect is its stability under *geometric* perturbations. The idealized, homogeneous H_b allows one to bring in Fourier transform methods and define Chern number "topological invariants". But this approach only makes sense for a fictitious setup, whereas we are actually interested in the non-idealized setup. For example, it is impossible to confine electrons to a perfectly flat 2D surface, or to generate a perfectly uniform magnetic field. *A posteriori*, we learn from experiments that the quantum Hall effect is somehow able to ignore such imperfections (up to a certain point). This does not mean that we should *a priori* throw away imperfections in the theory. Indeed, imperfections, or disorder, are actually needed to explain important aspects of the effect.

At first glance, the spectral problem for H_B on a general surface (X, g) and a general magnetic 2-form $-iB\cdot\mathrm{vol}_{X,g}$ looks impossibly complicated. Certainly, we cannot exactly solve for its spectrum with present-day mathematics. *The non-trivial work is justify rigorously, why H_b is enough to predict the qualitatively relevant part of the true H_B spectrum.*

18.3.1 Lichnerowicz identity for Landau Hamiltonians

Consider a noncompact, simply-connected, oriented Riemannian surface (X, g). Under these conditions, a Hermitian line bundle $\mathcal{L} \to X$ is trivializable, and the curvature of its connection ∇ is a complete gauge invariant. Furthermore, a unique spin structure exists.

So let the curvature $F^\nabla = -iB \cdot \mathrm{vol}$ of $\mathcal{L} = \mathcal{L}_B$ be given. We also need the curvature of the spinor line bundles $\mathcal{S}^\pm$. For this, recall Eq. (16.12) for the spin curvature 2-form $\mathcal{R}$ in terms of the Riemann tensor. In two dimensions,

we only need to understand the endomorphism $\mathcal{R}(e_1, e_2)$ for $\{e_1, e_2\}$ a local oriented orthonormal frame,

$$\begin{aligned}
\mathcal{R}(e_1, e_2) &= \frac{1}{4}\sum_{i,j=1}^{n} g(R(e_1, e_2)e_i, e_j)e_i \cdot e_j\cdot \\
&= \frac{1}{4}\left(g(R(e_1, e_2)e_1, e_2)e_1 \cdot e_2 \cdot \ + g(R(e_1, e_2)e_2, e_1)e_2 \cdot e_1\cdot\right) \\
&= \frac{1}{4}\left(-g(R(e_1, e_2)e_2, e_1)e_1 \cdot e_2 \cdot \ - g(R(e_2, e_1)e_1, e_2)e_1 \cdot e_2\cdot\right) \\
&= -\frac{1}{4}\sum_{i,j=1}^{n} g(R(e_j, e_i)e_i, e_j)e_1 \cdot e_2\cdot = -\frac{\mathrm{Sc}}{4}e_1 \cdot e_2 \cdot \ .
\end{aligned}$$

Since $e_1 \cdot e_2\cdot$ acts as $\mp i$ on $\mathcal{S}^{\pm}$, we deduce that $\mathcal{R}(e_1, e_2) = \pm\frac{i\,\mathrm{Sc}}{4}$ on $\mathcal{S}^{\pm}$, and this calculation is independent of the orthonormal frame. So the curvature of $\mathcal{S}^{\pm}$ is $\pm\frac{i\,\mathrm{Sc}}{4} \cdot \mathrm{vol}$, and that of $\mathcal{S}^{\pm} \otimes \mathcal{L}_B$ is $-i(\mp\frac{\mathrm{Sc}}{4} + B)\mathrm{vol}$.

Putting these calculations into the Lichnerowicz identity, Eq. (17.5),

$$\begin{aligned}
(\not{D}^{(X,g)}_{\mathcal{L}_B})^2 &= \Delta^{\mathcal{S}\otimes\mathcal{L}_B} + \frac{\mathrm{Sc}}{4} - iBe_1 \cdot e_2\cdot \\
&= \begin{pmatrix} H_{B-\frac{\mathrm{Sc}}{4}} + \frac{\mathrm{Sc}}{4} - B & 0 \\ 0 & H_{B+\frac{\mathrm{Sc}}{4}} + \frac{\mathrm{Sc}}{4} + B \end{pmatrix}
\end{aligned}$$

Since we are typically given a magnetic field strength B rather than $B - \frac{\mathrm{Sc}}{4}$, the following equivalent identities are useful,

$$\begin{aligned}
(\not{D}^{(X,g)}_{\mathcal{L}_{B+\frac{\mathrm{Sc}}{4}}})^2 &= \begin{pmatrix} H_B - B & 0 \\ 0 & H_{B+\frac{\mathrm{Sc}}{2}} + \frac{\mathrm{Sc}}{2} + B \end{pmatrix}, \\
(\not{D}^{(X,g)}_{\mathcal{L}_{B-\frac{\mathrm{Sc}}{4}}})^2 &= \begin{pmatrix} H_{B-\frac{\mathrm{Sc}}{2}} + \frac{\mathrm{Sc}}{2} - B & 0 \\ 0 & H_B + B \end{pmatrix}.
\end{aligned} \tag{18.3}$$

Now, the key point is the following. We can consider the true magnetic field strength B to be a constant b, up to some perturbation B_{pert} with $\sup_{x\in X} |B_{\mathrm{pert}}(x)| \ll |b|$. If the scalar curvature of X is also much smaller than $|b|$, then the spectral gap argument of Section 18.2.1 still works, although the deduced gap will be smaller. For instance, $H_{B+\frac{\mathrm{Sc}}{2}} + \frac{\mathrm{Sc}}{2} + B$ will remain strictly positive, so that $H_B - B$ retains some spectral gap above 0, e.g., $(0, c)$, where

c is somewhat smaller than $|b|$. Adding back the function B, we deduce that H_B has an isolated band of spectrum around $|b|$.

Generally, the effect of non-constant curvatures B and Sc is to spread out the Landau levels into "Landau bands".

Remark (Optional). In general, the spectral subspace of the lowest Landau band is identified with that of the Dirac operator around 0 energy. Its "large-scale dimension" is measured by a certain K-theoretic Dirac index, which occurs in quantized units. This index is a modern generalization of the classical Fredholm index (which is a genuine integer counting kernel dimensions). The Dirac index is stable (in the appropriate sense) against geometric perturbations (in B and Sc). Unlike the standard Fredholm index of elliptic (e.g. Dirac) operators on compact manifolds, the Dirac index on noncompact manifolds is *not a topological invariant*. It is in fact possible to puncture many holes in X, drastically changing its topology, without changing the index.

Remark (Optional). The idealized Landau Hamiltonian with constant $B = b$ on a hyperbolic plane (constant negative scalar curvature) had been studied previously by Comtet–Houston. There, the scalar curvature needs to be overcome by a sufficiently large field b before any isolated Landau levels can form in the spectrum. This is already very different from the (idealized) Euclidean plane case.

18.4 Some quantum Hall physics (Optional)

We have explained the geometric origin of the formation of Landau levels/bands in the spectrum of magnetic Laplacians. As we outlined, each spectrally isolated Landau band has a Dirac index. In the idealized Euclidean plane setting, a *completely filled* Landau level is understood by physicists to contribute one unit of Hall conductance σ_{xy},

$$n \text{ filled Landau levels } \leftrightarrow \; n \text{ units of Hall conductance.} \tag{18.4}$$

Regarding the "filling" of Landau levels: we are not actually studying a single electron, but a whole density of independent electrons occupying X. It suffices to understand the spectrum of H_b (describing a single electron), and then construct the many-electron wavefunction by a certain "second quantization" procedure. The electrons will "fill up" the eigenspaces of H_b one by one, starting from the lowest energy one.

Each Landau level is able to admit a certain density of electron states. So the total electron density divided by this Landau level capacity gives the filling factor ν. When physical units are restored to the discussion, it turns out that the Landau level capacity is proportional to the magnetic field strength $|b|$. So by controlling b, the filling factor ν may be continuously varied.

Eq. (18.4) is rather tautological (integer = integer), and possibly useless, since it only describes very specific filling factors. The key experimental surprise is that σ_{xy} *remains an integer even when* ν *is significantly varied around integer values*,

$$(\approx n \text{ filled Landau levels}) \leftrightarrow n \text{ units of Hall conductance}$$

The key to understanding the "forced quantization" of σ_{xy} is realizing that the true experimental setting is "dirty". The electron actually experiences some electrostatic potential V coming, e.g., from the atomic lattice constituting the material sample, and more generally, impurities. So we really need to study the spectrum of the magnetic Schrödinger operator $H_B + V$, where again, V is assumed to be small but nonzero, and may even be random (i.e. basically indeterminate to the experimentalist). The effect of this "dirt" is to modify the spectrum of H_B slightly, by introducing lots of new *localized* eigenstates with energies near the idealized Landau levels. In this regard, the B_{pert} term in $B = b + B_{\text{pert}}$ can also be thought of as "geometrical dirt". In contrast to the *delocalized* "eigenstates" constituting the clean Landau bands, the localized eigenstates do not contribute to the Hall conductance.

Now we can understand what happens when ν is not exactly integral — the "fractional remainder" occupies the part of the spectrum contributed by localized eigenstates. Nevertheless, the core of the Landau band is protected from becoming completely localized, because of its non-trivial index (the index cannot abruptly jump to zero).

Edge states. Recall that the "geometric" calculation of $\sigma(H_B)$ only works under the assumption of essential self-adjointness. If X has a boundary, and appropriate boundary conditions (e.g. Dirichlet) are imposed on the Landau Hamiltonian, then the formal Dirac square root will not actually exist as a self-adjoint operator. So we can no longer deduce the spectral gaps by the geometric-algebraic method. In fact, the Dirichlet Landau Hamiltonian H_b on the Euclidean half-plane can be shown to have *unbroken spectrum* $[|b|, \infty)$. The "bulk spectral gaps" $((2n+1)|b|, (2n+3)|b|), n \in \mathbb{N}$ for H_b get completely

filled up upon the introduction of a boundary, and this continuous spectrum is actually "unbreakable" in some sense! Here, we find another example where issues of self-adjointness cannot be ignored, and are instead central to the question at hand.

Conclusion. The quantum Hall effect is a remarkable *spectral* phenomenon for *quantum mechanical* electrons subject to the *combined geometry of a 2D sample and a gauge field*. The qualitative phenomenon is turns out to be insensitive to the fine details of this geometry. Furthermore, a dramatic spectral gap-filling phenomenon occurs when a boundary is introduced. To explain this "topological phase phenomenon", one cannot begin with a theory with only topological ingredients.

19 Unbounded operators on Hilbert space

19.1 Recap of Hilbert space theory

We take for granted the L^p spaces of complex-valued functions on measure spaces X, e.g., that the p-norm $||\psi||_p = \left(\int_X |\psi(x)|^p\right)^{1/p}$ defines a complete metric, as does the L^2-inner product, $\langle\psi|\varphi\rangle = \int_X \overline{\psi(x)}\varphi(x)$. So $L^p(X)$ is a Banach space, and $L^2(X)$ is a Hilbert space.

On a general Hilbert space $\mathcal{H}$, each $\psi \in \mathcal{H}$ determines a functional $\langle\psi| : \varphi \mapsto \langle\psi|\varphi\rangle$, whose norm makes sense, $||\langle\psi|\,|| := \sup_{||\varphi||=1}|\langle\psi|\varphi\rangle|$. The *Riesz representation theorem* says that the correspondence $\psi \leftrightarrow \langle\psi|$ is antilinear, norm-preserving, and exhausts all continuous linear functionals on $\mathcal{H}$. So the dual space $\mathcal{H}^*$ can be turned into a Hilbert space in its own right, and is anti-isomorphic to $\mathcal{H}$. This is the justification behind the ket-bra duality in Dirac's notation for the inner product.

A Hilbert space admits orthonormal bases, whose cardinality is an invariant. (Note: *convergent* infinite linear combinations are allowed!) A linear operator $T : \mathcal{H} \to \mathcal{H}$ is continuous iff it is bounded in the *operator norm*,

$$||T||_{\text{op}} := \sup_{||\psi||=1} ||T\psi||,$$

and we write $\mathcal{B}(\mathcal{H})$ for the normed linear space of bounded linear operators on $\mathcal{H}$. For $T \in \mathcal{B}(\mathcal{H})$, its *adjoint* $T^* \in \mathcal{B}(\mathcal{H})$ is well-defined by the condition

$$\langle T^*\psi|\varphi\rangle = \langle\psi|T\varphi\rangle, \qquad \forall\, \psi, \varphi \in \mathcal{H}, \tag{19.1}$$

and $(T^*)^* = T$ holds. If $T : \mathcal{H}_1 \to \mathcal{H}_2$ is a bounded linear operator between different Hilbert spaces, the adjoint $T^* : \mathcal{H}_2 \to \mathcal{H}_1$ is defined similarly.

If a linear subspace $\mathcal{H}_0 \subset \mathcal{H}$ happens to be *closed*, then it is itself a Hilbert space. The orthogonal complement $\mathcal{H}_0^\perp$ is automatically closed. In fact, the closure of any linear subspace $\mathcal{H}_0 \subset \mathcal{H}$ is obtained by taking the double-complement, $\overline{\mathcal{H}_0} = (\mathcal{H}_0^\perp)^\perp$.

For closed $\mathcal{H}_0$, there is a uniquely defined projection operator, $p = p^2 = p^*$, whose range is $\mathcal{H}_0$; the complementary projection $1 - p$ has range $\mathcal{H}_0^\perp$.

The spectrum of $T \in \mathcal{B}(\mathcal{H})$ is defined to be

$$\sigma(T) := \{\lambda \in \mathbb{C} \,:\, T - \lambda \text{ not invertible in } \mathcal{B}(\mathcal{H})\}.$$

Of particular interest are unitary operators, $U^* = U^{-1}$, and self-adjoint operators $H = H^*$. These are examples of normal operators, $T^*T = TT^*$, for which the *spectral theorem* holds: there exists a unitary transformation V such that

$$VTV^{-1} = \int_{\mathbb{C}} \lambda \, d\mu(\lambda), \tag{19.2}$$

where μ is the projection operator-valued *spectral measure* of T. This "unitary diagonalizability" is a deep result requiring core theorems of functional analysis, and the precise meaning of μ is quite involved. In the special case where $\mathcal{H}$ admits a basis of *eigenvectors* for T, the spectral theorem reduces to the "discrete spectral decomposition"

$$T = \sum_{i=i}^{\infty} \lambda_i |\psi_i\rangle\langle\psi_i|, \tag{19.3}$$

where λ_i are the eigenvalues and ψ_i the corresponding normalized L^2-eigenvectors.

However, many bounded self-adjoint operators (such as multiplication by bounded real-valued function on L^2) do not have any L^2-eigenvectors at all, so Eq. (19.3) does not make sense! The spectrum generally has "continuous" parts which are not accounted for by normalizable eigenfunctions, so Eq. (19.3) is needed. This is a significant departure from linear *algebra*, and the validity of spectral theory hinges on the implicit operator topologies in the Hilbert/Banach space. In fact, in quantum mechanics practice, one wants to distinguish true eigenvectors ("bound states") and approximate ones coming from the "continuous" part of the spectrum ("scattering states").

19.2 Unbounded operators

Many realistic and important operators coming from quantum mechanics are not bounded/continuous. For example, $\frac{d}{d\theta}$ can be defined on $C^\infty(S^1)$, but the derivative can blow up (e.g. apply it to $\psi_n(\theta) = \exp in\theta$, and take $n \to \infty$).

A bounded operator T can completely specified by how it acts on any orthonormal basis; the extension to arbitrary Hilbert space elements is then given by linearity and continuity. This prescription fails for an unbounded (equivalently, discontinuous) operator T.

Closure. For example, $\frac{d}{d\theta} : C^\infty(S^1) \to C^\infty(S^1) \subset L^2(S^1)$ initially makes sense, but we cannot try to "continuously extend" the meaning of $\frac{d}{d\theta}$ to all L^2 functions. So derivative operators T are always unbounded, and can only be defined on some dense linear subspace (such a subspace is not closed in $\mathcal{H}$!),

$$T : \underbrace{\mathrm{Dom}(T)}_{\subsetneq \mathcal{H}} \to \mathcal{H}.$$

Usually, we start by defining a differential operator T on some nice space of smooth L^2 functions. We may attempt to enlarge the domain of T, thereby obtaining an *extension* of T (the range must still land in $\mathcal{H}$, of course). For example, we could consider the *graph* of T,

$$\Gamma_T = \{(\psi, T\psi) \,:\, \psi \in \mathrm{Dom}(T)\} \subset \mathcal{H} \oplus \mathcal{H},$$

and take its closure $\overline{\Gamma_T}$ in $\mathcal{H} \oplus \mathcal{H}$. We say that T is *closed* if its graph is a closed subset of $\mathcal{H} \oplus \mathcal{H}$. There is a complication: there is no guarantee that $\overline{\Gamma_T}$ will still be the graph of some linear operator (extending T).

Adjoint. For a linear operator to make sense as a quantum mechanical Hamiltonian operator, it must be self-adjoint (reality of the spectrum follows from spectral theory of such operators). For unbounded operators, the bounded operator definition of adjoint, Eq. (19.1), does not make sense. The correct generalization which takes $\mathrm{Dom}(T)$ into account is the following.

Definition 66. Let $T : \mathrm{Dom}(T) \to \mathcal{H}$ be a densely-defined linear operator. Define $\mathrm{Dom}(T^*)$ to be the set of $\varphi \in \mathcal{H}$ such that there exists (a unique) $\eta_\varphi \in \mathcal{H}$ obeying the condition

$$\langle \eta_\varphi | \psi \rangle = \langle \varphi | T\psi \rangle, \qquad \forall \psi \in \mathrm{Dom}(T).$$

Then the *adjoint* operator $T^* : \mathrm{Dom}(T^*) \to \mathcal{H}$ is defined as

$$T^*\varphi := \eta_\varphi, \qquad \varphi \in \mathrm{Dom}(T^*).$$

By design, the adjoint's domain is the maximal one for which the adjointness condition makes sense *on the respective domains* of T and T^*,

$$\langle T^*\varphi | \psi \rangle = \langle \eta_\varphi | \psi \rangle = \langle \varphi | T\psi \rangle, \qquad \forall \psi \in \mathrm{Dom}(T),\ \varphi \in \mathrm{Dom}(T^*). \tag{19.4}$$

Notice that the larger $\mathrm{Dom}(T)$ is, the smaller $\mathrm{Dom}(T^*)$ will be.

Lemma 19.1. *Let T be a densely-defined operator on a Hilbert space $\mathcal{H}$. Then* $\mathrm{Range}(T)^\perp = \ker(T^*)$*, and thus*

$$\mathcal{H} = \ker(T^*) \oplus \overline{\mathrm{Range}(T)}. \tag{19.5}$$

Proof. For any $\varphi \in \ker(T^*)$ and $f = T\psi \in \mathrm{Range}(T)$,

$$\langle \varphi | f \rangle = \langle \varphi | T\psi \rangle = \langle T^*\varphi | \psi \rangle = 0.$$

Thus

$$\ker(T^*) \subset \mathrm{Range}(T)^\perp.$$

Conversely, suppose $f \in \mathrm{Range}(T)^\perp$, so $\langle f | T\psi \rangle = 0 = \langle 0 | \psi \rangle$ for all $\psi \in \mathrm{Dom}(T)$. By definition, $f \in \mathrm{Dom}(T^*)$ with $T^*f = 0$, i.e., $f \in \ker(T^*)$. Thus $\mathrm{Range}(T)^\perp = \ker(T^*)$, and the orthogonal decomposition, Eq. (19.5), follows. □

19.2.1 Self-adjointness

Now that we know what adjoint operators are, we can ponder about self-adjointness. By default, we only consider densely-defined operators from now on. First, we say that H is *symmetric*, or *formally self-adjoint*, if

$$\langle H\varphi | \psi \rangle = \langle \varphi | H\psi \rangle, \qquad \forall \varphi, \psi \in \mathrm{Dom}(H).$$

By definition, the true adjoint H^* will generally have a larger domain than H. But note that H^* is generally *not* symmetric, because its domain is too large. What we seek is a genuine equality $H = H^*$, *including equality of their domains*, in which case, we say that H is *self-adjoint*.

So the issue of self-adjointness is the issue of finding an appropriate enlargement of $\mathrm{Dom}(H)$ to $\mathrm{Dom}(\tilde{H})$,

$$\mathrm{Dom}(H) \subset \underbrace{\mathrm{Dom}(\tilde{H})}_{?} = \mathrm{Dom}(\tilde{H}^*) \subset \mathrm{Dom}(H^*).$$

Of course we should also specify how the extended operator $\tilde{H}$ acts on the extended domain.

As a preliminary step, we state, without proof, some general facts:

- For a symmetric H, the graph-closure $\overline{\Gamma_H}$ remains the graph of another symmetric operator, denoted $\overline{H}$. The latter operator is called the *closure* of H, and $\Gamma_{\overline{H}} = \overline{\Gamma_H}$.

- H^* is automatically closed (i.e. its graph is closed), and $(H^*)^* = \overline{H}$.
- $(\overline{H})^* = (H^*)^{**} = \overline{H^*} = H^*$, so we may as well assume that H is a *closed* symmetric operator to begin with.

If we are fortunate, H is already defined on a natural domain such that the closure $\bar{H}$ is self-adjoint. In this case, H is said to be *essentially self-adjoint*. Unfortunately, $\bar{H}$ will generally fail the following self-adjointness criterion.

Proposition 19.2. *A closed symmetric operator H on $\mathcal{H}$ is self-adjoint iff* $\ker(H^* \pm i) = 0$ *iff* $\mathrm{Range}(H \pm i) = \mathcal{H}$.

Proof. Evidently, a self-adjoint $H = H^*$ cannot have $\pm i$ eigenvalues, otherwise

$$H^*\psi = H\psi = i\psi \;\Rightarrow\; i\langle\psi|\psi\rangle = \langle\psi|H\psi\rangle = \langle H\psi|\psi\rangle = -i\langle\psi|\psi\rangle.$$

Now assume H^* has no $\pm i$-eigenvalue. If $0 \neq \psi \in \mathrm{Range}(H \pm i)^\perp$, then for all $\varphi \in \mathrm{Dom}(H)$,

$$\langle\psi|(H \pm i)\varphi\rangle = 0 \;\Rightarrow\; \psi \in \mathrm{Dom}(H^*) \text{ with } 0 = (H^* \mp i)\psi,$$

contradicting the assumption. Thus $\mathrm{Range}(H \pm i)$ must be dense in $\mathcal{H}$. In fact, $\mathrm{Range}(H \pm i)$ is closed (this is proved in Eq. (20.1) later), so $\mathrm{Range}(H \pm i) = \mathcal{H}$.

Finally, assume $\mathrm{Range}(H \pm i) = \mathcal{H}$. So for any $\varphi \in \mathrm{Dom}(H^*)$, we can write $(H^* \mp i)\varphi = (H \mp i)\psi_\pm$ for some $\psi_\mp \in \mathrm{Dom}(H)$. Since $\mathrm{Dom}(H) \subset \mathrm{Dom}(H^*)$, we actually have

$$\begin{aligned}(H^* \mp i)(\varphi - \psi_\mp) = 0 \;&\Leftrightarrow\; \varphi - \psi_\mp \in \ker(H^* \mp i)\\ &\Leftrightarrow\; \varphi - \psi_\mp \in \mathrm{Range}(H \pm i)^\perp = 0 \quad \text{(Lemma 19.1)}.\end{aligned}$$

Thus $\varphi = \psi_\mp \in \mathrm{Dom}(H)$, i.e., $\mathrm{Dom}(H^*) \subset \mathrm{Dom}(H)$. So H and H^* actually have the same domains, meaning that H is self-adjoint. □

In Section 20, we will learn how to find the self-adjoint extensions for $\overline{H}$. But we first discuss the motivating setting of differential operators.

19.3 First-order differential operators

Let us begin with first-order differential operators D acting on $C^\infty(U;\mathbb{C}^N)$, where U is some open domain $U \subset \mathbb{R}^n$. So

$$D\zeta = \left(\sum_{j=1}^n A^j\partial_j + B\right)\zeta, \qquad \zeta \in C^\infty(U;\mathbb{C}^N), \tag{19.6}$$

where A^j, B are some smooth $N \times N$ matrix-valued functions on U. The Fourier transform of D entails the replacement of $-i\partial_j$ by a momentum-space variable η_j. If A^j, B were constant over $U = \mathbb{R}^n$, then we could take the Fourier transform,

$$\widehat{D}_\eta\hat{\zeta}_\eta = \left(i\sum_{j=1}^n A^j\eta_j + B\right)\hat{\zeta}_\eta, \qquad \hat{\zeta}_\eta \in \mathbb{C}^N,\ \eta = (\eta_1,\ldots,\eta_n).$$

That is, $\widehat{D}$ becomes a family of $N\times N$ matrices parametrized by $\eta \in \widehat{\mathbb{R}}^n$. Here, it is useful to think of η as the cotangent vector $\eta_j dx^j$, and the first-order part of $\widehat{D}$ as a "symbol map",

$$\sigma_D : \eta \mapsto i\eta(\sum_{j=1}^n A^j\partial_j) = i\sum_{j=1}^n A^j\eta_j \ \in \mathrm{End}(\mathbb{C}^N).$$

Of course, we often do not have constant coefficients A^j, B. Moreover, we are interested in first-order differential operators D acting on sections of Hermitian vector bundles E over Riemannian manifolds X. This is a linear map

$$D : \Gamma(E) \to \Gamma(E)$$

such that each point $x \in X$ lies in a neighbourhood U on which D has a local representation as a first-order operator, as in Eq. (19.6). So the function ζ in Eq. (19.6) is to be regarded as a local representation of ψ, while the A^j, B are local matrix representations of sections $\tilde{A}^j, \tilde{B}$ of $\mathrm{End}(E)$. Changing the coordinate chart and/or local trivialization of E will change the matrix functions A^j, B, so it does not really make sense to say that the latter are "constant".

Symbol and ellipticity. In particular, we have to remember basepoints $x \in X$, and consider $\partial_j|_x \in T_xX$, as well as $\eta_x \in T_x^*X$. This means that the "Fourier transform" of D should really be an object parametrized by T^*X. Therefore, we define the (principal) *symbol* of D to be the bundle map

$$\begin{aligned} \sigma_D : T^*X &\to \mathrm{End}(E) \\ \eta &\mapsto i\sum_{j=1}^{n} \tilde{A}^j(x)\eta_j, \qquad \eta = \eta_j\, dx^j|_x \in T_x^*X. \end{aligned}$$

We say that D is *elliptic* if its (principal) symbol $\sigma_D(\cdot)$ is invertible for all non-zero cotangent vectors.

Remark. To see that σ_D is independent of the choice of coordinate cotangent basis $\{dx^j\}_{j=1,\dots,n}$, write $\eta = df|_x$ for some smooth function f. Then the calculation

$$([iD, f]\psi)(x) = \left(i\sum_{j=1}^{n} \tilde{A}^j(x) \underbrace{\partial_j|_x f}_{\eta_j} \right) \psi(x) = \sigma_D(\eta)\psi(x), \qquad \psi \in \Gamma(E),$$

provides an alternative coordinate-free description of $\sigma_D(\eta)$.

Domains. The compactly-supported smooth sections, $\Gamma_c^\infty(E)$, serves as a natural initial domain for D, which is dense in $L^2(X;E)$. Suppose $D = H$ is a symmetric (first-order) differential operator on $\Gamma_c^\infty(E)$.

- The closure $\overline{H}$ is a closed symmetric operator, and $\mathrm{Dom}(\overline{H})$ is called the *minimal domain* for D.
- At the other extreme, the *maximal* domain for H is $\mathrm{Dom}(H^*)$. Concretely, $\mathrm{Dom}(H^*)$ comprises those $\psi \in L^2(X;E)$ such that $D\psi$, taken in the distributional sense (see Section 20.3), is square-integrable. There are no further constraints on this maximal domain, because the integration-by-parts and compact support of $\mathrm{Dom}(H)$ ensure that the adjoint relation, Eq. (19.4), holds.

Example 19.1. By construction, the (twisted) Dirac operator has (principal) symbol

$$\sigma_{\not{D}}(\eta) = i\sum_{j=1}^{n} \eta_j e_j \cdot = i\eta\,\cdot,$$

where $\eta = \eta_j e^j$ is expanded in a local orthonormal cotangent frame; the orthonormal frame elements e_j act as endomorphisms of the spinors by Clifford multiplication. Whenever $||\eta|| \neq 0$, the endomorphism $i\eta \cdot$ squares to $||\eta||^2$, so it is invertible; thus $\not{D}$ is elliptic. Our calculation in Prop. 17.4 shows that $\not{D}$ is symmetric on the initial domain.

Remark. There is a notion of higher-order differential operators acting from sections of one bundle to another, and the principal symbol is similarly defined. For example, a Laplacian has principal symbol being scalar, $\sigma_\Delta(\eta) = ||\eta||^2$, and is thus elliptic.

19.3.1 Essential self-adjointness

In the theory of elliptic PDEs, there is a general principle of *elliptic regularity*, whereby weak (distributional) solutions ψ to $D\psi = f$, for smooth given f, boundary conditions on ψ, coefficients of the differential operator D, etc., are actually smooth ("strong solutions"). See Example 20.5, for a concrete instance of this.

To illustrate an application of this principle, consider a *closed* manifold X (meaning compact and without boundary), and a *symmetric* differential operator H of any order. On the minimal domain, $\overline{H}$ cannot have any $\pm i$-eigenvalues, otherwise we would have

$$\mp i||\psi||^2 = \langle H\psi|\psi\rangle = \langle\psi|H\psi\rangle = \pm i||\psi||^2.$$

On the maximal domain, H^* might have some $\pm i$-eigenvalues. A priori, a $\pm i$-eigenfunction in the enlarged domain, $\mathrm{Dom}(H^*)$, would not be smooth. But if H is *elliptic*, then such an eigenfunction must actually be smooth, so it already belongs to the minimal domain, contradicting our earlier assertion. Thus $\ker(H^* \pm i) = 0$, and we conclude from Prop. 19.2 that $\overline{H}$ is self-adjoint. That is, elliptic symmetric operators (e.g., Dirac, Laplace) on closed manifolds are essentially self-adjoint on the domain of smooth sections.

In general, the question of essential self-adjointness is very difficult. One sufficent criterion for *first-order* symmetric D is:

- The Riemannian metric on X is complete. (This is automatic for closed manifolds.)

In this case, all powers of D are also essentially self-adjoint (on $\Gamma_c^\infty(E)$). This result requires substantial analytic techniques ("mollifiers"), which we will not develop. A reference for these facts is [8].

There are also criteria for essential self-adjointness of symmetric H which are of second-order and/or elliptic [17, 3], and these usually depend on the behaviour of lower-order parts of H (e.g. potential terms), not just the principal symbol of H or geometric properties of X alone.

One approach to dealing with the self-adjointness issue is to avoid it altogether, by always working in situations where essential self-adjointness is automatic (e.g. closed manifolds). Unfortunately, this "black box" attitude is not enough to handle many situations of interest to us.

Instead, we shall investigate some "bad" examples, involving non-complete manifolds, where essential self-adjointness fails. For example, X may have "boundaries" that compactly-supported interior signals propagated by H can reach in finite time. Then different self-adjoint extensions of H are obtained by imposing different boundary conditions on $\mathrm{Dom}(H^*)$. It is also possible that no self-adjoint boundary conditions exist at all.

20 Self-adjoint extensions

20.1 Spectrum of symmetric operators

Definition 67. Let T be a closed operator on $\mathcal{H}$. Its *resolvent set* is

$$\rho(T) := \{\lambda \in \mathbb{C} \,:\, (T-\lambda)^{-1} : \mathcal{H} \to \mathrm{Dom}(T) \text{ exists and is bounded}\}.$$

The bounded operator

$$R_\lambda(T) := (T-\lambda)^{-1}, \qquad \lambda \in \rho(T),$$

is called the *resolvent of T at λ*. The *spectrum* of T is

$$\sigma(T) = \mathbb{C} \setminus \rho(T).$$

The simplest example of a spectral value is an eigenvalue: If $(T-\lambda)\psi = 0$ for some $0 \neq \psi \in \mathrm{Dom}(T)$, then $T-\lambda$ is not injective thus $(T-\lambda)^{-1}$ cannot exist. But not all spectral values are eigenvalues!

The spectrum/resolvent set has some basic topological features. For example, it may be shown that $\rho(T)$ is an open subset of $\mathbb{C}$ (possibly empty), and that $R_\lambda(T)$ depends analytically on $\lambda \in \rho(T)$ (see [16] §8 for a proof). Furthermore, from the closed graph theorem in functional analysis, the spectrum of a non-closed T is all of $\mathbb{C}$. This is why we investigate spectral questions for closed operators (or implicitly take the operator closure otherwise).

Proposition 20.1. *Let H be a closed symmetric operator. Then,*

- $\dim\ker(H^* - \lambda)$ *is constant for λ in the open upper half-plane of $\mathbb{C}$;*
- $\dim\ker(H^* - \lambda)$ *is constant for λ in the open lower half-plane of $\mathbb{C}$;*
- *$\sigma(H)$ is either the closed upper-half plane, closed lower-half plane, the entire complex plane, or a subset of $\mathbb{R}$. The last case occurs iff H is genuinely self-adjoint, iff* $\dim\ker(H^* \pm i) = 0$.

Proof. (Optional.) Let $\lambda = a + ib \in \mathbb{C} \setminus \mathbb{R}$, so $b \neq 0$. Since H is symmetric, for all $\psi \in \mathrm{Dom}(H)$,

$$\begin{aligned} ||(H-\bar{\lambda})\psi||^2 &= ||(H-a)\psi||^2 - ib\langle\psi|(H-a)\rangle\psi + ib\langle(H-a)\psi|\psi\rangle + b^2||\psi||^2 \\ &\geq b^2||\psi||^2. \end{aligned} \tag{20.1}$$

If $\lim_{n\to\infty}(H-\bar{\lambda})\psi_n$ exists (with $\psi_n \in \mathrm{Dom}(H)$), then the above inequality means that $\psi_n \to \psi$ converges as well. Since H is assumed to be closed, $(\psi, (H-\bar{\lambda})\psi)$ lies in the graph of $H-\bar{\lambda}$, with $(H-\bar{\lambda})\psi = \lim_{n\to\infty}(H-\bar{\lambda})\psi_n$. Therefore, $\mathrm{Range}(H-\bar{\lambda})$ is a closed subspace of $\mathcal{H}$. By Eq. (19.5),

$$\ker(H^*-\lambda)^\perp = \mathrm{Range}(H-\bar{\lambda}). \tag{20.2}$$

The next step is to show that a sufficiently small change of λ does not modify $\dim\ker(H^*-\lambda)$.

So consider a small perturbation $\lambda+\eta$, with $|\eta| \ll |b|$ so that $\lambda+\eta$ remains in the same upper/lower half-plane. Take any unit vector $\varphi \in \ker(H^*-(\lambda+\eta))$, and suppose $\varphi \in \ker(H^*-\lambda)^\perp$ as well. By Eq. (20.2), we may write $\varphi = (H-\overline{\lambda})\psi$ for some $\psi \in \mathrm{Dom}(H)$, and we have

$$0 = \langle \underbrace{(H^*-(\lambda+\eta))\varphi}_{0} | \psi\rangle = \langle \varphi | \underbrace{(H-\overline{\lambda})\psi}_{\varphi}\rangle - \overline{\eta}\langle\varphi|\psi\rangle = \underbrace{||\varphi||^2}_{1} - \overline{\eta}\langle\varphi|\psi\rangle. \tag{20.3}$$

By Eq. (20.1), we also have

$$||\psi|| \le ||(H-\overline{\lambda})\psi||/|b| = ||\varphi||/|b| = 1/|b| \ll /|\eta|,$$

which is too small for the right-side of Eq. (20.3) to vanish (Cauchy–Schwarz). This contradiction shows that

$$\ker(H^*-(\lambda+\eta)) \cap \ker(H^*-\lambda)^\perp = 0, \qquad \forall\, |\eta| \ll |b|.$$

It follows (exercise) that

$$\dim\ker(H^*-(\lambda+\eta)) \le \dim\ker(H^*-\lambda).$$

So by taking sufficiently small $|\eta|$ and swapping the roles of $\lambda+\eta$ and λ, we see that the dimension of $\ker(H^*-\lambda)$ is locally constant as a function of λ in the upper/lower half-plane. By connectedness, $\ker(H^*-\lambda)$ is constant on the upper/lower half-planes.

Note that Eq. (20.1) implies that $H-\bar{\lambda}$ is injective, with $(H-\bar{\lambda})^{-1} : \mathrm{Range}(H-\bar{\lambda}) \to \mathrm{Dom}(H)$ bounded. By Eq. (20.2), this inverse is defined on all of $\mathcal{H}$ iff $0 = \dim\ker(H^*-\lambda)$. So either $\sigma(H)$ or $\rho(H)$ contains the open upper half-plane; similarly for the lower half-plane. Since $\sigma(H)$ is closed, it has to be one of the four given options.

The fourth option, $\sigma(H) \subset \mathbb{R}$, occurs iff $\dim\ker(H^*-\lambda) = 0$ whenever $\lambda \in \mathbb{C}\setminus\mathbb{R}$, iff $\dim\ker(H^*\pm i) = 0$, iff H is self-adjoint, by Prop. 19.2. $\square$

Proposition 20.1 shows that the obstructions to self-adjointness are the following:

Definition 68. Let H be a (closed) symmetric operator on $\mathcal{H}$. Its *deficiency subspaces* and (possibly infinite) *deficiency indices* are

$$\mathcal{H}_\pm := \ker(H^* \mp i), \qquad n_\pm := \dim \mathcal{H}_\pm.$$

20.2 von Neumann theory of self-adjoint extensions

Let H be closed symmetric, then a self-adjoint extension of H is, in particular, a closed symmetric extension. Let us first look for closed symmetric extensions of H.

Introduce the *graph inner product*,

$$(\varphi, \psi)_H := \langle \varphi | \psi \rangle + \langle H^* \varphi | H^* \psi \rangle, \qquad \varphi, \psi \in \mathrm{Dom}(H^*).$$

Note that this defines a *stronger* topology and *different* notion of orthogonality than those coming from the Hilbert space inner product; but these notions can only be applied to $\mathrm{Dom}(H^*) \subset \mathcal{H}$. The terminology "$H$-closed, H-orthogonal" etc., will refer to the graph inner product. A linear subspace $L \subset \mathrm{Dom}(H^*)$ is *H-symmetric* if

$$[\psi, \varphi]_H := \langle H^* \varphi | \psi \rangle - \langle \varphi | H^* \psi \rangle = 0, \qquad \forall \varphi, \psi \in L.$$

For example, $\mathrm{Dom}(H)$ is H-closed-and-symmetric.

The importance of these definitions is the following. Suppose $\tilde{H}$ is a extension of H. Then in terms of domains, we have the hierarchy

$$H \subset \tilde{H} \subset \tilde{H}^* \subset H^*.$$

It is not hard to check that $\tilde{H}$ is closed symmetric iff $\mathrm{Dom}(\tilde{H})$ is H-closed-and-symmetric. So closed symmetric extensions $\tilde{H}$ are obtained by restricting H^* to H-closed-and-symmetric subspaces $L \subset \mathrm{Dom}(H^*)$.

Lemma 20.2. *Let H be a closed symmetric operator. There is an H-orthogonal decomposition*

$$\mathrm{Dom}(H^*) = \mathrm{Dom}(H) \oplus_H \mathcal{H}_+ \oplus_H \mathcal{H}_-.$$

Furthermore, the H-closed-and-symmetric subspaces of $\mathcal{H}_+ \oplus_H \mathcal{H}_-$ are in one-to-one correspondence with the H-closed-and-symmetric subspaces lying between $\mathrm{Dom}(H)$ *and* $\mathrm{Dom}(H^*)$*, via*

$$\mathcal{L} \leftrightarrow \mathrm{Dom}(H) \oplus_H \mathcal{L}.$$

Proof. This involves a tedious direct check, which we omit (A full proof can be found in pp. 138 of [17]). □

Theorem 20.3. *Let H be a closed symmetric operator. Its closed symmetric extensions are in one-to-one correspondence with partial isometries $U : \mathcal{H}_+ \to \mathcal{H}_-$.*

Proof. Let $\tilde{H}$ be a closed symmetric extension of H. By Lemma 20.2 and the discussion above it, we have $\mathrm{Dom}(\tilde{H}) = \mathrm{Dom}(H) \oplus_H \mathcal{L}$ for some uniquely determined H-closed-and-symmetric subspace $\mathcal{L} \subset \mathcal{H}_+ \oplus_H \mathcal{H}_-$. Consider any $\varphi = \varphi_+ + \varphi_- \in \mathcal{L}$, then

$$\begin{aligned} 0 = [\varphi, \varphi]_H &\equiv \langle \overbrace{H^*\varphi}^{i\varphi_+ - i\varphi_-} | \varphi \rangle - \langle \varphi | H^* \varphi \rangle \qquad (H\text{-symmetry of } \mathcal{L}) \\ &= -2i\langle \varphi_+ | \varphi_+ \rangle + 2i \langle \varphi_- | \varphi_- \rangle, \end{aligned}$$

so $||\varphi_+|| = ||\varphi_-||$. Thus there is a well-defined isometry $\mathcal{L} \cap \mathcal{H}_+$ to $\mathcal{L} \cap \mathcal{H}_-$, taking $\varphi_+ \mapsto \varphi_-$. Extending-by-zero gives a partial isometry $U : \mathcal{H}_+ \to \mathcal{H}_-$ with initial space $I(U) = \mathcal{L} \cap \mathcal{H}_+$.

Conversely, starting form a partial isometry $U : \mathcal{H}_+ \to \mathcal{H}_-$ with initial space $I(U)$, define

$$\mathrm{Dom}(\tilde{H}_U) := \{\psi + \varphi_+ + U\varphi_+ \,:\, \psi \in \mathrm{Dom}(H), \varphi_+ \in I(U)\}, \tag{20.4}$$

$$\tilde{H}_U(\psi + \varphi_+ + U\varphi_+) := H^*(\psi + \varphi_+ + U\varphi_+) = H\psi + i\varphi_+ - iU\varphi_+. \tag{20.5}$$

A little thought shows that $\mathrm{Dom}(\tilde{H}_U)$ is an H-closed-and-symmetric subspace of $\mathrm{Dom}(H^*)$, so $\tilde{H}_U$ is a closed symmetric extension of H by Lemma 20.2. □

In the construction of $\tilde{H}_U$, only pairs $(\varphi_+, U\varphi_+) \in \mathcal{H}_+ \oplus \mathcal{H}_-$ appear in the domain. Since φ_+ is an i-eigenvector of H^* while $U\varphi_+$ is a $(-i)$-eigenvector of H^*, the pair $(\varphi_+, U\varphi_+)$ no longer contributes to the $\pm i$-eigenspaces. In this way, the deficiency subspaces of H^* are "killed off in pairs". It should therefore be clear that we can completely kill the deficiency subspaces by choosing a *unitary* $U : \mathcal{H}_+ \to \mathcal{H}_-$. Of course, such a complete pairing can only be achieved iff $n_+ = n_-$ (possibly infinite).

Corollary 20.4 (von Neumann theory of self-adjoint extensions)**.** *Let H be a closed symmetric operator with deficiency subspaces $\mathcal{H}_\pm$ and deficiency indices (n_+, n_-).*

- *H is self-adjoint iff $n_+ = n_- = 0$.*
- *H admits self-adjoint extensions iff $n_+ = n_-$ (including the case where both are infinite); the self-adjoint extensions are parametrized by unitary maps $U : \mathcal{H}_+ \to \mathcal{H}_-$, according to the formulae in Eq.* (20.4)-(20.5).

Remark. In case $n_+ = n_- = n$, the self-adjoint extensions of H are parametrized by $\mathrm{U}(n)$, but not canonically so. This is because the identification of unitary maps $\mathcal{H}_+ \to \mathcal{H}_-$ with $\mathrm{U}(n)$ requires choosing orthonormal bases for $\mathcal{H}_\pm$.

20.3 Weak derivatives

Let $f : \Omega \to \mathbb{C}$ be a differentiable function, with Ω an open subset of $\mathbb{R}^n$. We usually think of $\partial_k f$ as another function. There are many situations where we have to deal with functions f which are not differentiable, but only approximated (in some specified sense) by some sequence of differentiable functions f_n. It is then useful to have a weaker notion of "derivative", which can be applied to such non-differentiable f.

The idea is to consider "$\partial_k f$" not as a function, but as a *distribution*. Generally speaking, a distribution is a linear functional on a space of test functions. We will take $C_c^\infty(\Omega)$ to be the test functions.

Some distributions are represented by actual functions. For instance, consider the space of *locally integrable* functions $L^1_{\mathrm{loc}}(\Omega)$. This means integrability over any compact subset of Ω. If $f \in L^1_{\mathrm{loc}}(\Omega)$, then it defines a linear functional on $C_c^\infty(\Omega)$ via the formula

$$f : \varphi \mapsto \int_\Omega f\varphi, \qquad \varphi \in C_c^\infty(\Omega).$$

Notice that when a function f is serving as a distribution, we only need its values almost-everywhere.

Some distributions are not representable by any (almost-everywhere defined) function, the most famous one being the delta distribution δ_{x_0}, $x_0 \in \Omega$,

$$\delta_{x_0} : \varphi \mapsto \varphi(x_0) = \text{“} \int_\Omega \delta_{x_0}(x)\varphi(x) \text{”}, \qquad \varphi \in C_c^\infty(\Omega).$$

The quotation marks indicate that δ_{x_0} is not actually a function, and its "integration against φ" is just a formal expression.

If f is classically continuously differentiable so that $\partial_k f$ is a genuine continuous function, then $\partial_k f$ would be the following distribution,

$$\partial_k f : \varphi \mapsto \int_\Omega (\partial_k f)\varphi = -\int_\Omega f \cdot \partial_k \varphi.$$

In the last equality, we used integration-by-parts and compact support of φ. If $f \in L^1_{\text{loc}}(\Omega)$, then "$\partial_k f$" may not exist as a function, but it makes sense as a distribution,

$$\text{“}\,\partial_k f\,\text{”} : \varphi \mapsto -\int_\Omega f \cdot \partial_k \varphi, \qquad \varphi \in C_c^\infty(\Omega). \tag{20.6}$$

By construction, Eq. (20.6) generalizes the classical derivative.

Example 20.1. Let $f : (-1, 1) \to \mathbb{C}$ be the discontinuous function,

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

For any $\varphi \in C_c^\infty(-1, 1)$, we have

$$-\int_{-1}^{1} f\varphi' = -\int_0^1 \varphi' = \varphi(0).$$

Thus, the distributional derivative of f is the delta distribution δ_0 at $x = 0$.

Example 20.2. Let $f : (-1, 1) \to \mathbb{C}$ be the function $f(x) = |x|$, which is only almost-everywhere differentiable. It is straightforward to check that the distributional derivative of f is the sign function.

Example 20.2 is slightly better-behaved than Example 20.1, in the sense that the distributional derivative is represented by a (almost-everywhere defined) function in L^1_{loc}. In such cases, we use the term *weak derivative* for emphasis.

20.3.1 Absolutely continuous functions

A function $\psi : [a, b] \to \mathbb{C}$ is *absolutely continuous* if, given any $\epsilon > 0$, there exists $\delta > 0$ such that for any finite collection of disjoint intervals $(x_i, x_i') \subset (a, b)$ with total length less than δ, one has $\sum_i |\psi(x_i') - \psi(x_i)| < \epsilon$. Absolutely continuous functions are, in particular, uniformly continuous. We write $\text{AC}[a, b]$ for the set of absolutely continuous functions on $[a, b]$.

The Fundamental Theorem of (Lebesgue integral) Calculus is:

Theorem 20.5. *Let $\psi \in \mathrm{AC}[a,b]$. Then ψ is almost everywhere differentiable, and its almost-everywhere defined derivative ψ' satisfies*

$$\psi' \in L^1(a,b), \qquad \psi(x) = \int_a^x \psi'(\tilde{x})\, d\tilde{x}.$$

Conversely, let $f \in L^1(a,b)$. Then $F(x) := \int_a^x f(\tilde{x})\, d\tilde{x}$ is absolutely continuous, with $F' = f$ almost everywhere.

Suppose $\psi, \varphi \in \mathrm{AC}[a,b]$, then $\psi\varphi \in \mathrm{AC}[a,b]$ as well, and $(\psi\varphi)' = \psi'\varphi + \psi\varphi'$ almost everywhere, and is L^1. Integrating over $[a,b]$ gives the integration-by-parts rule,

$$\int_a^b \psi\varphi' = \psi(b)\varphi(b) - \psi(a)\varphi(a) - \int_a^b \psi'\varphi.$$

Remark. If $\psi \in \mathrm{AC}[a,b]$, is its almost-everywhere defined (classical) derivative the same thing as its distributional/weak derivative? By definition,

$$\psi'^{,\mathrm{dist}} : \varphi \mapsto -\int_a^b \psi\varphi' = \int_a^b \psi'\varphi, \qquad \varphi \in C_c^\infty(a,b),$$

where we used integration-by-parts (φ is AC, and vanishes at $x = a, b$). Thus ψ', considered as a distribution, coincides with $\psi'^{,\mathrm{dist}}$. Because of this, we refer to the almost-everywhere defined derivative $\psi' \in L^1(a,b)$ as the weak derivative of $\psi \in \mathrm{AC}[0,1]$.

Now suppose $\psi \in L^1(a,b)$ has weak derivative $\psi' \in L^1(a,b)$. Then the indefinite integral Ψ of ψ' is absolutely continuous. Because Ψ and ψ have the same weak derivative, they coincide (almost everywhere) up to a constant. So as an L^1 equivalence class, ψ may be identified with its absolutely continuous representative.

20.4 Example: momentum operator on an interval

For convenience, we will also refer to $-i\frac{d}{dx}$ as a derivative operator.

Example 20.3. Let us consider the weak derivative operator $T = -i\frac{d}{dx}$ on

$$\mathrm{Dom}(T) = \{\psi \in \mathrm{AC}[a,b] \,:\, \psi' \in L^2(a,b)\}.$$

We have

$$\mathrm{AC}[a,b] \subset L^2(a,b) \subset L^1(a,b),$$

(e.g. by Hölder's inequality), so by Remark 20.3.1, we can actually write

$$\mathrm{Dom}(T) = \{\psi \in L^2(a,b) \,:\, \psi' \in L^2(a,b)\} =: W^{1,2}(a,b),$$

Here, $W^{1,2}(a,b)$ is an example of a *Sobolev space*, equipped with the *Sobolev norm*,

$$||\psi||_{1,2} = \left(||\psi||_{L^2}^2 + ||\psi'||_{L^2}^2\right)^{1/2}.$$

Sometimes, $W^{1,2}$ is written H^1, but we use the former notation, since H is already being used for "Hamiltonian".

As an exercise in analysis, we may check that T is a closed operator. In fact, it is the closure of the classical derivative operator $-i\frac{d}{dx}$ defined initially on $C^\infty[a,b]$.

Let us compute the spectrum of T. For each $\lambda \in \mathbb{C}$, the smooth function $e^{i\lambda(\cdot)}$ belongs to $\mathrm{Dom}(T)$. Furthermore, it is an eigenfunction of T with eigenvalue λ,

$$T(e^{i\lambda x}) = \lambda\, e^{i\lambda x}.$$

Thus the spectrum of T is the entire complex plane, and T is definitely not self-adjoint.

Example 20.4. Let H be the classical derivative operator $-i\frac{d}{dx}$ on the initial domain

$$\mathrm{Dom}(H) = C_c^\infty(a,b).$$

So functions in $\mathrm{Dom}(H)$ vanish at the end points. Similar to Example 20.3, the closure $\overline{H}$ is the weak derivative operator on the domain

$$\mathrm{Dom}(\overline{H}) = \{\psi \in W^{1,2}(a,b) \,:\, \psi(a) = 0 = \psi(b)\}.$$

Integration-by-parts shows that $\overline{H}$ is a symmetric operator: for $\psi, \varphi \in \mathrm{Dom}(\overline{H}) \subset \mathrm{AC}[a,b]$,

$$\langle \psi | \overline{H}\varphi \rangle = \int_a^b \overline{\psi}(-i\varphi') = \int_a^b \overline{-i\psi'}\varphi = \langle \overline{H}\psi | \varphi \rangle.$$

Now consider the adjoint operator $H^* = \overline{H}^*$. Suppose $\psi \in \mathrm{Dom}(H^*)$. By definition, there exists some $H^*\psi \in L^2(a,b)$ such that

$$\langle H^*\psi | \varphi \rangle = \langle \psi | H\varphi \rangle, \qquad \forall \varphi \in \mathrm{Dom}(H).$$

Equivalently,

$$\int_a^b \overline{H^*\psi}\varphi = -i\int_a^b \overline{\psi}\varphi' = -\int_a^b \overline{-i\psi}\varphi', \qquad \forall \varphi \in C_c^\infty(a,b).$$

The right side is the distributional derivative of $\overline{-i\psi}$ applied to φ. The left side is the distribution represented by the L^2-function $\overline{H^*\psi}$. Thus the distributional/weak derivative of ψ is precisely $H^*\psi \in L^2(a,b)$, and we see that $\psi \in W^{1,2}(a,b)$.

Conversely, suppose $\psi \in W^{1,2}(a,b)$, so ψ is AC with L^2 weak derivative. When integrating by parts, the condition $\psi(a) = 0 = \psi(b)$ for $\psi \in \mathrm{Dom}(H)$ ensures that no boundary terms will appear. So ψ is allowed in $\mathrm{Dom}(H^*)$ with $H^*\psi$ being its weak derivative.

In summary, H^* is the weak derivative $-i\frac{d}{dx}$ on the maximal domain

$$\mathrm{Dom}(H^*) = W^{1,2}(a,b),$$

which is the operator T from Example 20.3.

Example 20.5. For simplicity, we shall consider the interval $(-\pi,\pi)$. Example 20.4 shows that $\overline{H} = -i\frac{d}{dx}$ is symmetric but not self-adjoint on

$$\mathrm{Dom}(\overline{H}) = \{\psi \in W^{1,2}(-\pi,\pi) \,:\, \psi(-\pi) = 0 = \psi(\pi)\} =: W_0^{1,2}(-\pi,\pi).$$

Let us use the von Neumann theory to find the self-adjoint extensions of $\overline{H}$.

First, we work out the deficiency subspaces $\mathcal{H}_\pm = \ker(H^* \mp i)$. Suppose $\psi \in \mathrm{Dom}(H^*)$ lies in $\mathcal{H}_+$. So ψ is absolutely continuous, and satisfies

$$-i\psi' - i\psi = 0. \tag{20.7}$$

Then ψ' is also absolutely continuous. Differentiate Eq. (20.7) again, and deduce similarly that ψ'' is absolutely continuous. Inductively, ψ is actually smooth, and satisfies $\psi' = -\psi$ (ordinary derivative). Similarly for $\mathcal{H}_-$. Therefore,

$$\mathcal{H}_\pm = \mathrm{span}\{x \mapsto e^{\mp x}\},$$

and the deficiency indices are $(1,1)$. So there is a family of self-adjoint extensions of $\overline{H}$, parametrized by the maps $\mathcal{H}_+ \to \mathcal{H}_-$. Noting that $\varphi_+ : x \mapsto e^{-x}$ and $\varphi_- : x \mapsto e^{x}$ have the same norm on $L^2(-\pi,\pi)$, we see that unitary maps $U_\alpha : \mathcal{H}_+ \to \mathcal{H}_-$ are simply labelled by a phase factor $\alpha \in \mathrm{U}(1)$,

$$U_\alpha : \varphi_+ \mapsto \alpha\varphi_-.$$

Correspondingly, the self-adjoint extensions H_α of $\overline{H}$ are the restrictions of $H^* = -i\frac{d}{dx}$ (the weak derivative) to

$$\mathrm{Dom}(H_\alpha) = \{\psi + \beta(\varphi_+ + \alpha\varphi_-) \,:\, \psi \in W_0^{1,2}(-\pi,\pi),\, \beta \in \mathbb{C}\}.$$

Observe that if $f = \psi + \beta(\varphi_+ + \alpha\varphi_-) \in \mathrm{Dom}(H_\alpha)$, then

$$\begin{aligned} f(-\pi) &= \beta(e^{\pi} + \alpha e^{-\pi}), \\ f(\pi) &= \beta(e^{-\pi} + \alpha e^{\pi}), \end{aligned}$$

so $f(\pi) = \gamma_\alpha f(-\pi)$ where

$$\gamma_\alpha = \frac{e^{-\pi} + \alpha e^{\pi}}{e^{\pi} + \alpha e^{-\pi}} = \overline{\alpha}\frac{1 + \alpha e^{2\pi}}{1 + \overline{\alpha} e^{2\pi}} \in \mathrm{U}(1).$$

Thus the parameter α corresponds to the quasi-periodicity phase parameter γ_α.

We had already found the operators H_α of Example 20.5, in Section 3.3.4, by less rigorous methods. The precise domain comprises not just the smooth functions obeying a quasiperiodicity condition, but also absolutely continuous ones (with square-integrable weak derivative).

20.4.1 Momentum operator on a circle

The self-adjoint operator $H_{\alpha=1}$ acts on the domain of *periodic* $W^{1,2}$ functions. Writing $S^1 = [-\pi, \pi]/_{-\pi\sim\pi}$, this domain is equivalently

$$\{\psi \in L^2(S^1) \,:\, \psi' \in L^2(S^1)\} \equiv W^{1,2}(S^1),$$

a Sobolev space on the circle.

Note that although $L^2(S^1) \cong L^2(-\pi, \pi)$ as Hilbert spaces, to make sense of the weak derivative on S^1, we need to work with overlapping coordinate charts, not just on $(-\pi, \pi)$. For example, the initial domain $C_c^\infty(S^1) = C^\infty(S^1)$ is not the same as the initial domain $C_c^\infty(-\pi, \pi)$. The adjoint of the classical derivative operator $-i\frac{d}{dx}$ on $C^\infty(S^1)$ is the weak derivative operator on $W^{1,2}(S^1)$. On S^1, the existence of the weak derivative over S^1 already implies continuity at $\theta = \pi$.

To summarize, the momentum operator $-i\frac{d}{d\theta}$ on the circle is already essentially self-adjoint on $C^\infty(S^1)$. Its closure D is the weak derivative on the domain

$$\mathrm{Dom}(D) = W^{1,2}(S^1).$$

The function space $W^{1,2}(S^1)$ has a concrete description in terms of the Fourier transform.

There is a Hilbert space isomorphism

$$\begin{aligned} \ell^2(\mathbb{Z}) &\to L^2(S^1) \\ \delta_m &\mapsto w_m, \qquad\qquad w_m(\theta) = e^{im\theta}/\sqrt{2\pi}. \end{aligned}$$

Here, $\delta_m = (\ldots, 0, \underbrace{1}_{m\text{-th}}, 0 \ldots)$, and the assignment $\delta_m \mapsto w_m$ is a map of orthonormal basis elements. A general element $\psi \in L^2(S^1)$ is expandable as (an L^2-convergent sum)

$$\psi = \sum_{m\in\mathbb{Z}} \widehat{\psi}(m) w_m,$$

where the Fourier coefficients are computed as

$$\widehat{\psi}(m) = \langle w_m | \psi \rangle_{L^2(S^1)} = \frac{1}{\sqrt{2\pi}} \int_{S^1} e^{-im\theta} \psi(\theta)\, d\theta.$$

If $\psi \in \mathrm{Dom}(D) = W^{1,2}(S^1)$, then the Fourier coefficients of $D\psi = -i\psi'$ are

$$\widehat{D\psi}(m) = \langle w_m | -i\psi' \rangle_{L^2(S^1)} = \langle -iw_m' | \psi \rangle_{L^2(S^1)} = m \langle w_m | \psi \rangle_{L^2(S^1)} = m \cdot \widehat{\psi}(m).$$

Here, we used integration-by-parts and the periodicity of w_m, ψ when regarded as functions on $(-\pi, \pi)$. So the Fourier transform of D is the "position operator" Q on $\ell^2(\mathbb{Z})$, with domain

$$\mathrm{Dom}(Q) = \{\widehat{\psi} \in \ell^2(\mathbb{Z}) \,:\, \sum_{m\in\mathbb{Z}} (1+m^2) |\widehat{\psi}(m)|^2 < \infty\}, \tag{20.8}$$

and action

$$(Q\widehat{\psi})(m) = m \cdot \widehat{\psi}(m), \qquad \widehat{\psi} \in \mathrm{Dom}(Q).$$

Note that Eq. (20.8) is the largest domain on which Q can make sense as an operator mapping into $\ell^2(\mathbb{Z})$. Eq. (20.8) is often used as the *definition* of the Sobolev space $W^{1,2}(S^1)$.

We have found that D is unitarily equivalent to a (self-adjoint, unbounded) multiplication operator. The spectrum of a multiplication operator is the (essential) range, and this is $\mathbb{Z}$ in this case.

20.5 Example: momentum operator on line

The analysis is similar with (a, b) replaced by the unbounded real line $\mathbb{R}$. We say that $\psi : \mathbb{R} \to \mathbb{C}$ is absolutely continuous, denoted $\psi \in \mathrm{AC}(\mathbb{R})$, if it is absolutely continuous on any compact interval. (This is sometimes called "locally absolutely continuous".) In this case, ψ is almost-everywhere differentiable, with $\psi' \in L^1_{\mathrm{loc}}(\mathbb{R})$. If $\psi \in L^1_{\mathrm{loc}}(\mathbb{R})$ has a weak derivative (in $L^1_{\mathrm{loc}}(\mathbb{R})$), then we may consider ψ to be represented by an absolutely continuous function. Integration-by-parts works as in the bounded interval case, with these modifications.

The classical derivative $H = -i\frac{d}{dx}$ is symmetric on the initial domain $C_c^\infty(\mathbb{R})$. The maximal domain is the Sobolev space

$$\mathrm{Dom}(H^*) = \{\psi \in L^2(\mathbb{R}) \,:\, \psi' \in L^2(\mathbb{R})\} =: W^{1,2}(\mathbb{R}).$$

Note that if $\psi \in W^{1,2}(\mathbb{R})$, then integration-by-parts shows that

$$\int_0^x (\overline{\psi}\psi' + \overline{\psi'}\psi) + |\psi(0)|^2 = |\psi(x)|^2,$$

and the left side has a finite limit as $x \to \infty$, by square-integrability of ψ, ψ'; similarly for $x \to -\infty$. So $\lim_{x\to\pm\infty} |\psi(x)|$ exists and it must be zero for square-integrability of ψ to hold.

The fact that $\psi \in \mathrm{Dom}(H^*)$ has $\lim_{x\to\pm\infty} \psi(x) = 0$ implies that H^* is symmetric — the "boundary terms at infinity" vanish. Therefore H^* is self-adjoint, i.e., $H^* = \overline{H}$. In other words, H is essentially self-adjoint on $C_c^\infty(\mathbb{R})$.

20.5.1 Momentum operator on half-line

Now consider the half-line $\mathbb{R}_+ = (0, \infty)$. The maximal domain of $-i\frac{d}{dx}$ is the Sobolev space $W^{1,2}(\mathbb{R}_+)$ (defined in a similar way), but this is no longer a domain of self-adjointness. This is because $\psi \in W^{1,2}(\mathbb{R}_+)$ may have $\psi(0) \neq 0$, and contribute a boundary term violating the symmetry condition.

The deficiency subspace $\mathcal{H}_+$ is spanned by the function $x \mapsto e^{-x}$, while $\mathcal{H}_-$ is trivial (because e^x is not in $L^2(\mathbb{R}_+)$). Since the deficiency indices are unequal, the symmetric operator $H = -i\frac{d}{dx}$ on the initial domain $C_c^\infty(\mathbb{R}_+)$ has no self-adjoint extensions.

Similarly, the deficiency indices of $+i\frac{d}{dx}$ are $(0, 1)$. Thus $\not{D} = -i\frac{d}{dx} \oplus i\frac{d}{dx}$ has deficiency indices $(1, 1)$, and therefore a $\mathrm{U}(1)$-family of self-adjoint extensions.

It is easy to see that the domains of self-adjointness are

$$\mathrm{Dom}(\not{D}_\alpha) = \{(\psi_1, \psi_2) \in W^{1,2}(\mathbb{R}_+; \mathbb{C}^2) \,:\, \psi_2(0) = \alpha\psi_1(0)\}, \qquad \alpha \in \mathrm{U}(1).$$

As a general fact (exercise), if H is self-adjoint on $\mathrm{Dom}(H)$, then for any bounded self-adjoint V, the operator $H + V$ remains self-adjoint on $\mathrm{Dom}(H)$. So, for example, we can add to $\not{D}_\alpha$ an off-diagonal mass term $M = \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix}$, as in Section 3.3.3, obtaining the self-adjoint massive Dirac Hamiltonians on $\mathbb{R}_+$. We had seen that the spectrum $\not{D}_\alpha + M$ depends on the parameter α, and in fact exhibits spectral flow (Fig. 2).

20.5.2 Index of domain-wall Dirac operator

In Section 17.1.3, we mentioned that the massive Dirac Hamiltonian on the real line,

$$\not{D} + M = \begin{pmatrix} -i\frac{d}{dx} & m \\ m & i\frac{d}{dx} \end{pmatrix}, \qquad m \neq 0,$$

has an ambiguity in the sign of m. This operator is self-adjoint on $W^{1,2}(\mathbb{R}; \mathbb{C}^2)$, and the spectrum is easily found by Fourier transform to be

$$\sigma(D_m) = (-\infty, |m|] \cup [|m|, \infty).$$

Let us replace m by the function $m \cdot \mathrm{sgn}(x)$. This means that each half-line has a different sign for the mass term. The corresponding massive Dirac Hamiltonian is called a *domain-wall* Dirac Hamiltonian,

$$D_{\mathrm{dw}} = \begin{pmatrix} -i\frac{d}{dx} & m \cdot \mathrm{sgn} \\ m \cdot \mathrm{sgn} & i\frac{d}{dx} \end{pmatrix}, \qquad \mathrm{Dom}(D_{\mathrm{dw}}) = W^{1,2}(\mathbb{R}; \mathbb{C}^2).$$

Naïvely, it would seem that the spectrum of D_m and D_{dw} should coincide. This is almost true, except for a remarkable 0-eigenfunction (exercise),

$$\psi_{\mathrm{JR}}(x) = \begin{pmatrix} i \\ \mathrm{sgn}(m) \end{pmatrix} e^{-|mx|},$$

called the *Jackiw–Rebbi* solution. Notice that ψ_{JR} has a cusp at $x = 0$, so it does not belong to the initial domain of smooth functions. Nevertheless, it is square-integrable, and has square-integrable weak derivative, so it belongs to $W^{1,2}(\mathbb{R}; \mathbb{C}^2)$, and is admissible as an eigenfunction for D_{dw}.

This is an example of an *index* phenomenon. The operator D_{dw} is odd with respect to the grading operator $i\mathsf{e}_1 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$. In exactly the same way as the Dirac operators on graded spinor bundles, the operator D_{dw} can be rewritten as

$$D_{\mathrm{dw}} = \begin{pmatrix} 0 & (D_{\mathrm{dw}})_- \\ (D_{\mathrm{dw}})_+ & 0 \end{pmatrix}$$

with respect to a basis in which $i\mathsf{e}_1$ is diagonal. The usual spectral supersymmetry argument holds, except for the 0-eigenspace: the kernels of $(D_{\mathrm{dw}})_+$ and $(D_{\mathrm{dw}})_-$ may not coincide, and this asymmetry is measured by

$$\mathrm{Index}(D_{\mathrm{dw}}) = \dim\ker(D_{\mathrm{dw}})_+ - \dim\ker(D_{\mathrm{dw}})_- \in \mathbb{Z}.$$

We have just seen that the mass function $m \cdot \mathrm{sgn}$ leads to an index of ± 1 for the operator D_{dw}.

Of course, the 0-eigenvalue could actually be a spurious spectral phenomenon, which disappears if the mass function is modified slightly. An index theorem says that it is *stable* — the index is actually predicted by the *sign change* in the asymptotic values (at $x = \pm\infty$) of the mass function, independently of the precise form of the function. This is not too hard to prove for the 1D case of D_{dw}. As an exercise, convince yourself that the replacement of $m \cdot \mathrm{sgn}(x)$ with a smooth version, such as $m \cdot \tanh(x)$, still exhibits a ± 1 index.

References

[1] Baez, J., Muniain, J.P.: Gauge Fields, Knots, and Gravity. World Scientific, 1994

[2] Bleecker, D., Booß-Bavnbek, B.: Index theory with applications to mathematics and physics. International Press, 2013

[3] Braverman, M., Milatovic, O., Shubin, M.: Essential self-adjointness of Schrödinger type operators on manifolds. Russian Math. Surveys 57(4) 641 (2002)

[4] Choquet-Bruhat, Y., Dewitt-Morette, C.: Analysis, manifolds, and physics. Part I: Basics. North Holland (1982)

[5] Eguchi, T., Gilkey, P.B., Hanson, A.J.: Gravitation, gauge theories and differential geometry. Physics Reports 66(6) 213–393 (1980)

[6] Guillemin, V., Pollack, A.: Differential Topology. Prentice-Hall, 1974

[7] Hamilton, M.: Mathematical Gauge Theory. With Applications to the Standard Model of Particle Physics. Springer, 2017

[8] Higson, N., Roe, J.: Analytic K-homology. Oxford Univ. Press, 2000

[9] Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry. Vol 1. Wiley, 1963

[10] Lawson, H.B., Michelsohn, M.-L.: Spin Geometry. Princeton Univ. Press, 1989

[11] Lee, J.M.: Introduction to smooth manifolds. Graduate Texts in Math. Vol 218, Springer 2013

[12] Milnor, J.: Topology from the differentiable viewpoint. Univ. Press Virginia, 1965

[13] Naber, G.: Topology, Geometry, and Gauge Fields. Foundations. Texts in Appl. Math., Springer, 2011

[14] Naber, G.: Topology, Geometry, and Gauge Fields. Interactions. Texts in Appl. Math., Springer, 2011

[15] Nash, C.: Differential Topology and Quantum Field Theory. Acad. Press, 1992

[16] Reed, M., Simon, B.: Methods of Modern Mathematical Physics, Vol. 1. Functional Analysis. Acad. Press, 1980

[17] Reed, M., Simon, B.: Methods of Modern Mathematical Physics, Vol. II. Fourier analysis, Self-adjointness. Acad. Press, 1975

[18] Steenrod, N.: The Topology of Fibre Bundles. Princeton Math. Series, 1960