Lecture Notes on Geometry and Physics

Guo Chuan Thiang

August 29, 2024

Assumed background: Basic differential geometry, group theory, and analysis.

These are expanded notes for a set of introductory lectures on the differential geometry of bundles (gauge theory) and quantum mechanics (and some classical mechanics). The aim is to show that the two subjects are inextricable, already at the level of basic mathematical and physical ideas. We will have no occasion to discuss the topology of bundles, or Galilean/special/general relativity, or quantum field theory. So there are no physics prerequisites beyond some curiosity about the natural world.

The content is not readily available from any single source, and details of some of the individual mathematical topics can be found in the references below.

- Lee, J.M.: Introduction to smooth manifolds. Springer (2012)
- Hamilton, M.: Mathematical gauge theory. Springer (2017)
- Kobayashi, S., Nomizu, K.: Foundations of differential geometry Vol 1. Interscience, Wiley (1963)
- Abraham, R., Marsden, J.E.: Foundations of Mechanics: Second Edition. Addison-Wesley (1978)
- Thirring, W.: Classical Mathematical Physics. Springer (2003)
- Arnold, V.I.: Mathematical Methods of Classical Mechanics. GTM 60, Springer (1989)

Somewhat related are lecture notes for a quantum mechanics course with a similar flavour, but focussed on functional analytic aspects:

`http://faculty.bicmr.pku.edu.cn/~guochuanthiang/QT24.html`

1 Manifolds and motion

Classically, motion of a point particle[1] in a d-dimensional *configuration space* M is a time-parametrized curve

$$\gamma : I \to M, \qquad t \mapsto \gamma(t),$$

[1] Or a finite collection of point particles.

where I denotes some open interval of $\mathbb{R}$. There are *equations of motion* (E.O.M.) for predicting γ given some initial conditions. For example, Newton's Laws of motion are usually presented using calculus on $M \cong \mathbb{R}^d$, and involves something called acceleration. More geometrical formulations are Lagrangian and Hamiltonian mechanics. Under mild assumptions, all these formulations lead to the same γ.

1.1 Affine space manifold

We write $\mathbb{R}^d$ for the vector space of d-tuples of real numbers. It comes with a standard topology (induced by the standard inner product).

Let A be the topological space underlying $\mathbb{R}^d$. Any homeomorphism $A \to \mathbb{R}^d$ is called a *chart*, providing d *coordinate functions* for A. We shall also consider *local* charts, which are homeomorphisms between some open subset $U \subseteq A$ and some open subset of $\mathbb{R}^d$. When two charts φ, φ', defined on U, U' respectively, have overlapping domains, there is a change-of-coordinates homeomorphism

$$\varphi' \circ \varphi^{-1} : \varphi(U \cap U') \to \varphi'(U \cap U'). \tag{1}$$

Given a function $f : A \to \mathbb{R}$, we would like to determine its its rate-of-change along motions $\gamma : I \to A$. At time $t \in I$, this would be

$$\frac{d(f \circ \gamma)}{ds}\Big|_{s=t}, \tag{2}$$

provided $f \circ \gamma : I \to \mathbb{R}$ is differentiable.

Pick some chart $\varphi : A \xrightarrow{\cong} \mathbb{R}^d$. Then we can rewrite

$$f \circ \gamma = (\underbrace{f \circ \varphi^{-1}}_{\mathbb{R}^d \to \mathbb{R}}) \circ (\underbrace{\varphi \circ \gamma}_{I \to \mathbb{R}^d}),$$

and declare that f and γ are individually *smooth* if $f \circ \varphi^{-1}$ and $\varphi \circ \gamma$ are smooth in the usual sense of elementary calculus[2]. In this way, we can investigate rates-of-changes of arbitrary smooth functions along arbitrary smooth paths, *with respect to the choice of coordinate chart* φ. Unfortunately, this goes against a basic principle that meaningful physical laws cannot depend on having to pick a particular choice of coordinates.

Thus we consider a maximal *family* of local charts, sufficient to cover all of A, such that the change-of-coordinates maps, (1), are always smooth[3]. Such a family is called a *maximal smooth atlas*, or a *smooth structure*, on A. Now we declare f and γ to be smooth, if $f \circ \varphi^{-1}$ and $\varphi \circ \gamma$ are smooth, for all choices of local chart φ in the maximal atlas. Because smoothness is a local condition, once it holds with respect to a local chart, it will automatically hold on all other charts, due to smoothness of (1). So a smooth structure is a device for doing calculus without preference for any particular chart.

Remark 1.1. On A, the *standard smooth structure* is obtained by starting with any global chart $\varphi : A \to \mathbb{R}^d$, then admitting all possible mutually compatible local charts into the maximal atlas. We then forget the initial choice.

One might worry that we did pick out a preferred initial global chart. As it turns out, apart from the mysterious $d = 4$ case, there is a unique choice of smooth structure on A up to diffeomorphism, namely the one described above.

[2]Partial derivatives of all orders exist for all the component functions.

[3]They map open subsets of $\mathbb{R}^d$ to open subsets of $\mathbb{R}^d$, so we know what "smooth" means.

1.2 General manifold and tangent vectors

The local charts on A allow us to pass to $\mathbb{R}^d$, and invoke the well-developed multivariable calculus of maps $\mathbb{R}^m \to \mathbb{R}^n$. In fact, all the local features of differential and integral calculus, differential equations on $\mathbb{R}^d$, etc., make sense any topological space equipped with local charts.

Definition 1.2. A *topological d-manifold* is a second-countable, Hausdorff topological space M, such that each $x \in M$ is contained in some open neighbourhood homeomorphic to some open subset of $\mathbb{R}^d$. A *smooth d-manifold* is a topological d-manifold equipped with a maximal smooth atlas[4].

So, on a smooth manifold M, there is a consistent coordinate-independent concept of smooth functions $f : M \to \mathbb{R}$ and smooth curves $\gamma : I \to M$. Given a smooth γ, we have the rates-of-change of f along γ, Eq. (2), not just for a single f, but for the whole algebra $C^\infty(M)$ of smooth functions on M.

So γ determines the rate-of-change assignment,

$$\dot\gamma(t) \equiv \frac{d\gamma}{ds}\Big|_{s=t} : C^\infty(M) \to \mathbb{R}$$
$$f \mapsto \frac{d(f\circ\gamma)}{ds}\Big|_{s=t}, \qquad t \in I. \tag{3}$$

Besides being a linear assignment, the key algebraic property of $\dot\gamma(t)$ is that it is a *derivation of* $C^\infty(M)$ *at* $\gamma(t)$, i.e., a *Leibniz* rule holds at $\gamma(t) \in M$:

$$\dot\gamma(t)(f_1 \cdot f_2) = f_1(\gamma(t)) \cdot \dot\gamma(t)(f_2) + \dot\gamma(t)(f_1) \cdot f_2(\gamma(t)), \qquad \forall\, f_1, f_2 \in C^\infty(M).$$

Remark 1.3. If we write (3) in terms of a local coordinate chart $\varphi = (q^1, \ldots, q^d) : U \to \mathbb{R}^d$,

$$\frac{d(f\circ\gamma)}{ds}\Big|_{s=t} = \sum_{i=1}^{d} \frac{\partial(f\circ\varphi^{-1})}{\partial q^i}\Big|_{\varphi\circ\gamma(t)} \frac{d(\varphi\circ\gamma)^i}{ds}\Big|_{s=t}, \qquad \text{(Chain rule)},$$

we see that only the *numerical velocity vector*,

$$\frac{d(\varphi\circ\gamma)}{ds}\Big|_{s=t} \in \mathbb{R}^d,$$

matters for the left-hand-side. So all curves γ possessing the same numerical velocity vector determine the same derivation $\dot\gamma(t)$.

Definition 1.4. The *tangent space* to M at a point $x \in M$, denoted T_xM, is the vector space of derivations of $C^\infty(M)$ at x. For a smooth curve $\gamma : I \to M$, the derivation $\dot\gamma(t) \in T_{\gamma(t)}M$ is called its *velocity* at time t.

[4]Any homeomorphism from an open subset of M to an open subset of $\mathbb{R}^d$ is a *local chart*, providing *local coordinates* for M. A *smooth atlas* is a collection of local charts, sufficient to cover M, such that the change-of-coordinates maps are smooth.

Partial derivatives. Let $x \in M$, and let φ be a local chart with $\varphi(x) = \mathbf{0}$. Associated to φ are the "coordinate axes" centred at x,

$$\gamma_i : t \mapsto \varphi^{-1}(0, \ldots, 0, \underbrace{t}_{i\text{-th}}, 0, \ldots, 0), \qquad i = 1, \ldots, d.$$

These provide the "basic curves" through x, and indeed, a *basis* spanning T_xM. In more detail: the derivation

$$\partial_i|_x := \dot{\gamma}_i(0) : f \mapsto \frac{d(f \circ \gamma_i)}{dt}\Big|_{t=0}, \qquad f \in C^\infty(M),$$

is called the i-th *partial derivative* operator at x. It outputs the rate-of-change of f as the i-th coordinate is increased from 0. Any $v \in T_xM$ is a linear combination of these $\partial_i|_x$,

$$v = \sum_{i=1}^{d} v^i \partial_i|_x. \tag{4}$$

(Exercise.) Thus, in view of Remark 1.3, any derivation $v \in T_xM$ is geometrically realizable as the velocity of some curve through x,

$$v = \dot{\gamma}(0), \qquad \gamma(t) = \varphi^{-1}(tv^1, \ldots, tv^d).$$

To summarize: The tangent space T_xM comprises the velocities of curves passing through x.

1.3 Tangent bundle

The *tangent bundle* of a manifold M is the collection of all its tangent spaces,

$$TM := \bigsqcup_{x \in M} T_xM,$$

equipped with the projection map $\pi : TM \to M$ taking T_xM to x. Since elements of T_xM are velocities of curves, we also refer to TM as the *velocity phase space*.

Let φ be a chart over $U \subseteq M$. For each $x \in U$, we have the basis $\{\partial_i|_x\}_{i=1,\ldots,d}$ for T_xM (Eq. (4)). Thus, the chart induces an identification

$$\begin{aligned} \Phi : TM|_U \equiv \bigsqcup_{x \in U} T_xM &\xrightarrow{\cong} U \times \mathbb{R}^d \\ v = \sum_{i=1}^{d} v^i \partial_i|_{\pi(v)} &\mapsto (\pi(v); v^1, \ldots, v^d), \end{aligned} \tag{5}$$

called a *bundle chart*. The product space $U \times \mathbb{R}^d$, equipped with the projection map to U, is called the *trivial vector bundle* over U. The bundle chart Φ respects the projection maps to U, and is an example of a *local trivialization* of TM. We can further consider

$$\begin{aligned} (\varphi \times \mathrm{id}_{\mathbb{R}^d}) \circ \Phi : TM|_U &\xrightarrow{\cong} U \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d \\ v = \sum_{i=1}^{d} v^i \partial_i|_{\pi(v)} &\mapsto (\varphi(\pi(v)); v^1, \ldots, v^d) = (\underbrace{q^1, \ldots, q^d}_{\mathbf{q}}, \underbrace{v^1, \ldots, v^d}_{\mathbf{v}}). \end{aligned} \tag{6}$$

We refer to $\mathbf{q}, \mathbf{v}$ as *position* and *velocity* coordinates, respectively.

As an exercise, check that if $\varphi_\alpha, \varphi_\beta$ are local charts over $U_\alpha, U_\beta \subseteq M$ respectively, then

$$\Phi_\alpha \circ \Phi_\beta^{-1} : (x, \mathbf{v}) \mapsto (x; g_{\alpha\beta}(x)(\mathbf{v})), \qquad x \in U_\alpha \cap U_\beta, \tag{7}$$

where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(d)$ is smooth. Indeed, if q_α^j and q_β^j are the respective position-coordinate functions, then

$$g_{\alpha\beta}(x) = \left(\frac{\partial q_\alpha^j}{\partial q_\beta^i}\Big|_{\varphi_\beta(x)} \right)_{ji} \tag{8}$$

is the Jacobian of the change-of-position-coordinates map $\varphi_\alpha \circ \varphi_\beta^{-1}$, which depends smoothly on x by assumption.

Thus the local charts φ of M provide local charts identifying the local tangent bundles $TM|_U$ with open subsets of $\mathbb{R}^d \times \mathbb{R}^d$. Then TM itself is a smooth $2d$-manifold (fill in details yourself). Furthermore, TM is a smooth *vector bundle* over M.

Definition 1.5. A (smooth) *vector bundle* over a manifold M is a smooth manifold E with a smooth surjective map $\pi : E \to M$, such that

- Every $x \in M$ lies in an open neighbourhood $U \subseteq M$ such that $E|_U := \pi^{-1}(U)$ is trivializable, i.e., there exists a diffeomorphism, called a *local trivialization*,

  $$\Phi : E|_U \to U \times \mathbb{R}^n,$$

  respecting the projections to U;

- For any pair of local trivializations $(U_\alpha, \Phi_\alpha), (U_\beta, \Phi)$, we have

  $$\begin{aligned} \Phi_\alpha \circ \Phi_\beta^{-1} : (U_\alpha \cap U_\beta) \times \mathbb{R}^n &\to (U_\alpha \cap U_\beta) \times \mathbb{R}^n \\ (x; \xi) &\mapsto (x; g_{\alpha\beta}(x) \cdot \xi) \end{aligned}$$

  for some smooth *transition functions* $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(n)$.

A *section* of a vector bundle E is a smooth map $s : M \to E$ such that $\pi \circ s = \mathrm{id}_M$. The space of sections of E is denoted $\Gamma(E)$.

Some remarks:

- The availability of local trivializations is a defining property, but we do not select any particular choice.

- Each fibre $E|_x = \pi^{-1}\{x\}$ inherits a vector space structure, through its identification with $\mathbb{R}^n$ with respect to any local trivialization. So we can perform scalar-multiplication pointwise,

  $$\begin{aligned} \Gamma(E) \times C^\infty(M) &\to \Gamma(E) \\ (s, f) &\mapsto s \cdot f. \end{aligned}$$

- For TM, the bundle charts provide sufficient local trivializations making $\pi : TM \to M$ a vector bundle. The transition functions are the Jacobians, Eq. (8). Other local trivializations (not necessarily induced by local charts on M) are allowed as well, as long as the transition functions remain smooth.

  We write $\mathfrak{X}(M) = \Gamma(TM)$.

- Some vector bundles, e.g., TS^2, do not admit any nowhere-vanishing global section. Nevertheless, we can always find *local frames* for E — this means a collection of local sections $e_1, \ldots, e_n : U \to E|_U$, such that $\{e_1(x), \ldots, e_n(x)\}$ is a basis for $E|_x$ for every $x \in U$. For example, using a local trivialization, we can map the tautological frame for $U \times \mathbb{R}^n$ to a local frame for E. Conversely, we can use a local frame to construct a local trivialization. Thus E is globally trivializable, $E \cong M \times \mathbb{R}^n$ iff it admits a global frame. If TM is globally trivializable, M is said to be *parallelizable*.

- A function $f \in C^\infty(M)$ is a section of the *trivial*[5] bundle $M \times \mathbb{R}$. The values of f are well-defined numerical *scalars*, since they do not depend on choosing frames. Contrast with a *vector field*, which is a section of a vector bundle: its "value" at $x \in M$ is merely an element of an abstract vector space T_xM, and a frame is needed to achieve $T_xM \cong \mathbb{R}^n$.

1.4 Tangent vector fields and integral curves

Suppose a tangent vector field $X \in \mathfrak{X}(M)$ is given. Then locally in space and time, we can reconstruct a family of uniquely determined *integral curves*, whose velocities precisely match the tangent vector assignment of X,

$$\dot{\gamma}(t) = X|_{\gamma(t)}.$$

To do this, we use local charts to convert the integral curve condition into a set of numerical ODEs with specified first derivatives. Then invoke uniqueness/existence theorems.

We imagine points moving along these integral curves — this is called a local *flow*, and can be thought of as a time-parametrized group of local diffeomorphisms, "generated by X".

1.5 Comparing tangent vectors

Tangent spaces T_xM at different $x \in M$ are not canonically identified. Rather, they are *locally* identifi*able* with the standard vector space $\mathbb{R}^d$, *after* referring to a local trivialization, $TM|_U \cong U \times \mathbb{R}^d$. A choice of local trivialization is called a *local gauge*. With respect to a local gauge, a vector field X becomes locally described as a d-component function $\check{X} : U \to \mathbb{R}^d$. We could differentiate $\check{X}$ component-wise, but this is a *gauge-dependent* procedure!

The problem of differentiating vector fields intrinsically, i.e., gauge-independently, is an old one. In physics, one has differential equations for vector fields, tensors, etc., and demands that they make sense intrinsically. The key insight it that "intrinsic differentiation" requires *extra* geometric data, and this kind of data has profound physical meaning.

Parallelism of affine space. Why is it that, when working with affine space $M = A$, we can speak unambiguously about rate-of-change of velocities[6]? The answer is that we are implicitly using an extra *affine space* structure on A.

[5] Not just trivializ*able*!

[6] Newton's second law, $F = ma$, familiar from high school, is certainly very successful!

More precisely, let V be the vector space underlying $\mathbb{R}^d$, and $(V,+)$ be the underlying additive group. Note that V is a manifold such that $+$ and $-$ are smooth operations[7], so $(V,+)$ is a *Lie group*. Unlike V, the underlying manifold A does not have a preferred origin. But now we do not consider A just as a manifold, we remember that A has a smooth free and transitive action of $(V,+)$,

$$\begin{aligned} A \times (V,+) &\to A \\ (a,v) &\mapsto a+v. \end{aligned}$$

(Jargon: A is a smooth *torsor* over $(V,+)$.) When equipped with this group action, the manifold A is called *affine space*.

The action of $(V,+)$ determines some special curves in A, which we now describe. Consider *1-parameter subgroups* of $(V,+)$, i.e., smooth group homomorphisms

$$\gamma : (\mathbb{R},+) \to (V,+).$$

A little thought shows that such a γ must actually be a linear map. In other words, the space of 1-parameter subgroups of $(V,+)$ is identified with V itself,

$$v \in V \longleftrightarrow \gamma_v : t \mapsto tv.$$

Applying these γ_v to various starting points $a \in A$, we get the following curves

$$\begin{aligned} L_{a,v} : \mathbb{R} &\to A \\ t &\mapsto a + \gamma_v(t) = a + tv. \end{aligned}$$

These $L_{a,v}$ are what we informally call "straight lines", and they are labelled by the starting point $a \in A$ and the direction vector $v \in V$. As an exercise, verify that

$$\begin{aligned} V &\to T_a A \\ v &\mapsto \dot{L}_{a,v}(0) \end{aligned} \tag{9}$$

is a linear isomorphism. (Work with coordinates provided by a linear basis $\beta : \mathbb{R}^d \to V$.)

Altogether, we have a canonical trivialization of TA,

$$\begin{aligned} A \times V &\overset{\cong}{\longleftrightarrow} TA \\ (a,v) &\longleftrightarrow \dot{L}_{a,v}(0). \end{aligned}$$

This is the *global parallelism* of affine space, which allows velocities at distinct points $a, a' \in A$ to be compared, and judged to be parallel or otherwise.

Remark 1.6. If you know about Lie groups, the manifold underlying a Lie group G (understood as a G-torsor) is canonically parallelized in a similar way. The difference is that the vector space V is replaced by the so-called *Lie algebra* $\mathfrak{g}$ of G.

Remark 1.7. The curve $L_{a,v}$ is "straight" in another sense — its velocity vector is always v, so there is no acceleration. On a general manifold M, we need the extra data of a "connection" in order to judge whether the velocity vectors at different points on a curve

[7]Exercise: make this precise.

are changing. With respect to a connection, a "straight curve" with no acceleration is called a *geodesic*.

One way to specify a connection is by putting a Riemannian metric on M, which leads to the so-called *Levi–Civita* connection, and the distance-minimizing geodesics in Riemannian geometry. For example, if we give V an inner product, then the underlying A becomes a Riemannian manifold (usually called *Euclidean space*), and the Levi–Civita connection coincides with the connection coming from the parallelism as an affine space. So straight lines are also shortest paths.

1.6 Derivative of smooth maps between manifolds

A map $f : M_1 \to M_2$ between manifolds is *smooth* if its local representations with respect to (any) charts of M_1 and M_2 are smooth.

Definition 1.8. Let $f : M_1 \to M_2$ be a smooth map. Its *derivative at* $x \in X$ is the map

$$df|_x : T_xM_1 \to T_{f(x)}M_2$$

defined by

$$\underbrace{df|_x(v)}_{\text{derivation at } f(x)}(g) := v(g \circ f), \qquad v \in T_xM_1,\ g \in C^\infty(M_2). \tag{10}$$

Check the following:

- $df|_x(v)$ is indeed a derivation of $C^\infty(M_2)$ at $f(x)$.
- $df|_x$ is linear map.
- The chain rule is satisfied: if $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3$, then $d(g \circ f)|_x = dg_{f(x)} \circ df|_x$.
- $d(\mathrm{id}_M)|_x = \mathrm{id}_{T_xM}$.
- If $f : M_1 \to M_2$ is a diffeomorphism, then $df|_x$ is a linear isomorphism whose inverse is $d(f^{-1})|_{f(x)}$.
- Most importantly: If $v \in T_xM$ is the velocity of a curve γ through x, then $df|_x(v)$ is the velocity of the curve $f \circ \gamma$ through $f(x)$.

Example 1.9. Let $f : M \to \mathbb{R}$ be a smooth function, which we regard as a smooth map $f : M \to A$. Here, A is the 1-dimensional affine space manifold underlying the vector space $\mathbb{R}$. SO we have

$$\begin{aligned} df|_x : T_xM &\to T_{f(x)}A \\ \dot\gamma(0) &\mapsto \frac{d}{dt}\Big|_{t=0}(f \circ \gamma). \end{aligned}$$

Since A is an affine space manifold, we have $T_{f(x)}A = \mathbb{R}$ (see Eq (9)). The map $df|_x : T_xM \to \mathbb{R}$ is thus a linear functional, i.e.,

$$df|_x \in (T_xM)^* =: T_x^*M.$$

The above dual space, T_x^*M, is called the *cotangent space* at x, and $df|_x$ is a *cotangent vector* at x.

Actually, the various $df|_x : T_xM_1 \to T_{f(x)}M_2$ assemble into a single map,

$$df := \bigsqcup_{x\in M} df_x : TM_1 \to TM_2.$$

Using the bundle charts of TM_1, TM_2 induced by local charts on M_1, M_2, one can check that df is a smooth map. Indeed, df is a vector bundle *morphism*, in the sense that the diagram

$$\begin{array}{ccc} TM_1 & \xrightarrow{df} & TM_2 \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

commutes.

1.7 Differential forms

The *cotangent bundle* of M is

$$T^*M = \bigsqcup_{x\in M} T_x^*M.$$

If $\left\{\frac{\partial}{\partial q^i}|_x\right\}_{i=1,\dots,d}$ is a basis for T_xM induced by some local chart $\varphi = (q^1, \dots, q^d)$, then $\{dq^i|_x\}_{i=1,\dots,d}$ provides the dual basis for T_x^*M. So, in the same way as TM, the cotangent bundle T^*M is also a vector bundle,

$$\pi : T^*M \to M.$$

A section of T^*M is called a *differential 1-form*.

Example 1.10. Globalizing Example 1.9, for a smooth function $f : M \to \mathbb{R}$, we can regard

$$df = \bigsqcup_{x\in M} df|_x \tag{11}$$

as a differential 1-form, called the *gradient 1-form* of the scalar function f. The space of 1-forms is denoted

$$\Omega^1(M) := \Gamma(T^*M).$$

For a map $f : M_1 \to M_2$ of manifolds, the dual operation to the derivative map $df : TM_1 \to TM_2$ is the *pullback* of 1-forms,

$$f^* : \Omega^1(M_2) \to \Omega^1(M_1).$$

By definition, if η is a 1-form on M_2, then $f^*\eta$ is the 1-form on M_1 defined by the formula

$$(f^*\eta)|_x(v) = \eta|_{f(x)}\big(df|_x(v)\big), \qquad \forall\, v \in T_xM.$$

Generalizing the idea of 1-forms, we can consider smooth assignments of k-multilinear maps

$$\underbrace{T_xM \times \cdots \times T_xM}_{k \text{ times}} \to \mathbb{R},$$

which are totally antisymmetric in the k arguments; these are called *differential k-forms*. The space of differential k-forms is denoted $\Omega^k(M)$, and like 1-forms, they can be pulled back under smooth maps.

For example, in terms of a local chart, 2-forms can be expanded in terms of the antisymmetrized tensor product of coordinate 1-forms,

$$dq^i \wedge dq^j := dq^i \otimes dq^j - dq^j \otimes dq^i, \qquad 0 \leq i < j \leq d.$$

A 2-form $\omega \in \Omega^2(M)$ can be reduced to a 1-form by substituting a vector field X into the first argument,

$$\iota_X \omega(\,\cdot\,) = \omega(X, \,\cdot\,) \;\in \Omega^1(M).$$

2 Classical mechanics in phase space

A path γ in M automatically determines a path of velocities, $\dot{\gamma}$, in the velocity phase space TM,

$$t \mapsto \dot{\gamma}(t) \in T_{\gamma(t)}M.$$

Note that:

- Each velocity vector $T_{\gamma(t)}M$ is "attached" to the position $\gamma(t) \in M$. Velocities at different positions are not a priori comparable.
- We do not consider arbitrary paths in TM, but focus on those of the form $\dot{\gamma}$.

In *Lagrangian mechanics*, dynamics is specified by a *Lagrangian* function[8] $L : TM \to \mathbb{R}$. The equation of motion is a differential equation, stated in terms of L. This is given in Eq. (31) later.

Rather than try to explain where L comes from, we will start with a reformulation on the cotangent bundle T^*M, called *Hamiltonian mechanics*.

Remark 2.1. In practice, the coordinates of a "point particle" are, for example, the centre-of-mass of some rigid body. Some extra coordinates might be needed to account for the rotational alignment of the body, so in total, M is some abstract "configuration space" rather than literally a "position space". In general, we could think of M as a manifold described by "effective coordinates" which are sufficient to capture the relevant aspects of classical motion of a physical system.

The empirical basis of classical mechanical laws (say, Newton's laws) was celestial body motion, falling balls, etc. It is tempting to think of a macroscopic object as ultimately comprising a large collection of discrete "fundamental point particles" (or even a continuum/field of such particles). Then, in principle, we could be more ambitious and try to capture the full microscopic interactions and motions of the point-like constituents.

However, this is a highly misguided idea, as we know since the discovery of quantum mechanics in the 20th century. The successes of the idea of "point particle" had been in the *modelling* of massive objects that we can typically see. This does not mean that fundamental constituents of matter (whatever they "are"), are literally points moving along well-defined observable[9] paths γ.

At some scale, modelling of "small blobs" of matter by point particles fails. So, although TM seems more intuitive than T^*M, this bias is not that meaningful for post-classical mechanics.

2.1 Hamiltonian mechanics

Let $\varphi = (q^1, \ldots, q^d) : U \to \mathbb{R}^d$ be a local chart over $U \subseteq M$. A cotangent vector $p \in T^*M|_U$ is labelled by the coordinates of the basepoint,

$$q^i(p) = q^i(\pi(p)), \qquad \text{(abuse notation)},$$

together with the coefficients in the expansion

$$p = p_i(p) dq^i|_{\pi(p)}.$$

[8] One can also consider time-dependent Lagrangians, but we will not do so.

[9] And independent of whether an observation is actually made!

So φ induces a bundle chart for $T^*M|_U$, and *position-momentum* coordinates

$$(\mathbf{q}, \mathbf{p}) = (q^i, \dots, q^d, p_1, \dots, p_d) : T^*M|_U \to \mathbb{R}^d \times \mathbb{R}^d. \tag{12}$$

On T^*M, there is a *tautological*[10] 1-*form*, given in position-momentum coordinates by the formula

$$\theta = \sum_{i=1}^{d} p_i dq^i \ \in \Omega^1(T^*M). \tag{13}$$

To clarify, $dq^1, \dots, dq^d$ are regarded as 1-forms on $T^*M|_U$, not 1-forms on U. There are also "vertical" coordinate 1-forms dp_i on $T^*M|_U$, but these do not enter the definition of θ. Importantly, (13) is actually independent of the choice of chart φ (exercise), so (13) is well-defined on all of T^*M, not just on a locally trivialized part $T^*M|_U \cong U \times \mathbb{R}^d$.

Dynamics is specified by a smooth *Hamiltonian* function $H : T^*M \to \mathbb{R}$. The 2-form

$$d\theta = \sum_{i=1}^{d} dp_i \wedge dq^i \ \in \Omega^2(T^*M) \tag{14}$$

is non-degenerate[11] everywhere, so H uniquely defines a *Hamiltonian vector field* $X_H \in \mathfrak{X}(T^*M)$ via the condition

$$\iota_{X_H} d\theta = -dH. \tag{15}$$

Hamilton's equation of motion is the integral curve equation

$$\dot{\mathcal{C}}(t) = X_H|_{\mathcal{C}(t)}, \tag{16}$$

for curves $\mathcal{C} : I \to T^*M$. (Recall Section 1.4.) Once initial position-momentum data (i.e., a point on T^*M) is given, there is a unique integral curve $\mathcal{C}$ solving (16). Then $\pi \circ \mathcal{C}$ is the resulting motion on M.

In position-momentum coordinates, the Hamiltonian vector field is

$$X_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i}, \qquad \text{(Einstein summation convention)}$$

and Hamilton's equations, (16), are the following first-order ODEs in 2d variables,

$$\dot{q}^i(t) = \frac{\partial H}{\partial p_i}\Big|_{(\mathbf{q}(t),\mathbf{p}(t))}, \qquad \dot{p}_i(t) = -\frac{\partial H}{\partial q^i}\Big|_{(\mathbf{q}(t),\mathbf{p}(t))}, \qquad i = 1, \dots, d. \tag{17}$$

Remark 2.2. Let $t \mapsto \mathcal{C}(t)$ be an integral curve of X_H. Then

$$\begin{aligned}
\frac{d}{dt}H(\mathcal{C}(t)) &= dH|_{\mathcal{C}(t)} \cdot \dot{\mathcal{C}}(t) && \text{(Chain rule)} \\
&= dH|_{\mathcal{C}(t)} \cdot X_H|_{\mathcal{C}(t)} && \text{(Integral curve)} \\
&= -d\theta(X_H, X_H)|_{\mathcal{C}(t)} && (X_H \text{ Definition, Eq. (15)}) \\
&= 0. && \text{(Antisymmetry)}
\end{aligned}$$

Thus, the value of H remains constant along $\mathcal{C}$. In fact, the physical meaning of H is *total energy*.

[10] Also "canonical", "Liouville", etc.

[11] Thus $\omega = -d\theta$ gives a canonical symplectic form on T^*M. Hamiltonian mechanics can be formulated on general symplectic manifolds. We are only concerned with T^*M, and the fact that the symplectic form on T^*M comes from a tautological 1-form θ will be important for the relation to Lagrangian mechanics.

2.2 Relating momentum and velocity

The Hamiltonian provides a fibre-preserving way to identify T^*M with TM, as follows. At each fixed position $x \in M$, we have the fibrewise Hamiltonians

$$H^{(x)} := H|_{T^*_x M} : T^*_x M \to \mathbb{R},$$

depending only on the momentum at fixed x. The derivative of $H^{(x)}$ at $p \in T^*_x M$ is the linear map

$$\begin{aligned} dH^{(x)}|_p : T_p(T^*_x M) \cong T^*_x M &\to \mathbb{R} \\ p' &\mapsto \frac{d}{ds}\Big|_{s=0} H(p + sp'). \end{aligned}$$

In other words, $dH^{(x)}|_p$ is an element of $(T^*_x M)^* = T_x M$, i.e., a velocity at x. Thus we can convert momentum at x to velocity at x:

$$\begin{aligned} T^*_x M &\to T_x M \\ p &\mapsto dH^{(x)}|_p, \qquad x \in M. \end{aligned}$$

The above maps assemble globally into the *fibre derivative*,

$$\begin{aligned} \mathcal{F}H : T^*M &\to TM \\ p &\mapsto dH^{(\pi(p))}|_p. \end{aligned}$$

Concretely, in terms of position-velocity coordinates $(\mathbf{q}, \mathbf{v})$ for TM and position-momentum coordinates $(\mathbf{q}, \mathbf{p})$ for T^*M,

$$\mathcal{F}H : (\mathbf{q}, \mathbf{p}) \mapsto (\mathbf{q}, \mathbf{v}), \qquad \mathbf{v} \equiv \mathbf{v}(\mathbf{q}, \mathbf{p}) = \left(\frac{\partial H}{\partial p_1}\Big|_{(\mathbf{q},\mathbf{p})}, \ldots, \frac{\partial H}{\partial p_d}\Big|_{(\mathbf{q},\mathbf{p})} \right). \tag{18}$$

In most examples, $\mathcal{F}H : T^*M \to TM$ is a diffeomorphism, in which case, H is said to be *hyperregular*. We shall always make this assumption. Thus $\mathcal{F}H$ faithfully swaps momenta for velocities, while fixing the position $\mathbf{q}$.

2.3 Hamiltonian to Lagrangian mechanics

If you know about the Lie derivative[12], the Hamiltonian vector field X_H can be equivalently characterized by

$$\begin{aligned} \mathcal{L}_{X_H}\theta &= d\iota_{X_H}\theta + \iota_{X_H} d\theta && \text{(Cartan's formula for Lie derivative } \mathcal{L}) \\ &= d\iota_{X_H}\theta - dH && \text{(Eq. (15))} \\ &= d\big(\theta(H_X) - H\big). && (19) \end{aligned}$$

This prompts us to define the *Lagrangian* associated to H, as the function

$$L_H := \big(\theta(X_H) - H\big) \circ (\mathcal{F}H)^{-1} : TM \to \mathbb{R}. \tag{20}$$

[12]The Lie derivative $\mathcal{L}_X$ along a vector field X refers to the notion of differentiating functions, vector fields, forms, etc., by using the local diffeomorphic flow induced by X to pullback and compare such quantities.

The relationship is summarized by

$$\text{“Hamiltonian + Lagrangian} = \theta(X_H)\text{”}.$$

In coordinates,

$$L_H(\mathbf{q},\mathbf{v}) = p_i \frac{\partial H}{\partial p_i}\Big|_{(\mathbf{q},\mathbf{p})} - H(\mathbf{q},\mathbf{p}) = \mathbf{p}\cdot\mathbf{v} - H(\mathbf{q},\mathbf{p}), \tag{21}$$

where on the right side, $\mathbf{p}$ is determined from $\mathbf{v}$ by inverting Eq. (18)[13].

Remark 2.3. A function $L : TM \to \mathbb{R}$ also has a fibre derivative $\mathcal{F}L$, which swaps velocities for momenta. In coordinates,

$$\begin{aligned} &\mathcal{F}L : TM \to T^*M \\ &\qquad (\mathbf{q},\mathbf{v}) \mapsto (\mathbf{q},\mathbf{p}), \qquad \mathbf{p} = \Big(\frac{\partial L}{\partial v^1}\Big|_{(\mathbf{q},\mathbf{v})}, \dots, \frac{\partial L}{\partial v^d}\Big|_{(\mathbf{q},\mathbf{v})}\Big). \end{aligned} \tag{22}$$

In particular, the fibre derivative of $L = L_H$ is

$$\begin{aligned} \mathcal{F}L_H : (\mathbf{q},\mathbf{v}) &\overset{(22)}{\mapsto} \Big(\mathbf{q}, \frac{\partial L_H}{\partial v^1}, \dots, \frac{\partial L_H}{\partial v^d}\Big) \\ &\overset{(21)}{=} \Big(\mathbf{q}, \frac{\partial p_k}{\partial v^1}v^k + p_1 - \frac{\partial H}{\partial p_k}\frac{\partial p_k}{\partial v^1}, \dots, \frac{\partial p_k}{\partial v^d}v^k + p_d - \frac{\partial H}{\partial p_k}\frac{\partial p_k}{\partial v^d}\Big). \quad \text{(Chain rule)} \end{aligned}$$

If $\mathbf{v}$ is obtained from $\mathbf{p}$ via $\mathcal{F}H$, Eq. (18), substituting into the above calculation gives

$$(\mathbf{q},\mathbf{p}) \overset{\mathcal{F}H}{\mapsto} (\mathbf{q},\mathbf{v}) \overset{\mathcal{F}L_H}{\mapsto} (\mathbf{q},\mathbf{p}),$$

verifying that

$$\mathcal{F}L_H = (\mathcal{F}H)^{-1}. \tag{23}$$

For the rest of this subsection, we write L instead of L_H for convenience. Let us use (23) to work over TM instead of T^*M, by defining

$$\begin{aligned} \tilde{X}_L &:= (\mathcal{F}H)_* X_H = (\mathcal{F}L)_*^{-1} X_H && \in \mathfrak{X}(TM), \\ \tilde{\theta}_L &:= ((\mathcal{F}H)^{-1})^*\theta = (\mathcal{F}L)^*\theta && \in \Omega^1(TM). \end{aligned}$$

The vector field $\tilde{X}_L$ is called the *Lagrangian vector field* associated to L. Just as $X_H \in \mathfrak{X}(T^*M)$ is characterized by H via Eq. (19), $\tilde{X}_L \in \mathfrak{X}(TM)$ is characterized by L via

$$\mathcal{L}_{\tilde{X}_L}\tilde{\theta}_L = dL. \tag{24}$$

So the integral curves $\tilde{\mathcal{C}}$ of $\tilde{X}_L$ correspond to the integral curves $\mathcal{C}$ of X_H, by applying $\mathcal{F}H$ to the latter. By construction, $\mathcal{C}$ and $\tilde{\mathcal{C}} = \mathcal{F}H \circ \mathcal{C}$ project down to the same motions on M.

Eq. (24) is a geometrically concise statement of the E.O.M. in terms of a Lagrangian function on TM. It is not very useful yet, so let us write it out in coordinates. To this end, it is helpful to remember that $L = L_H$ is not an arbitrary function on TM, but is

[13] In other words, L_H is the fibrewise *Legendre transform* of H, in the sense that $L_H^{(x)} := L_H|_{T_xM}$ is the *Legendre transform* of $H^{(x)} := H|_{T_x^*M}$, for each basepoint $x \in M$.

obtained from some H by fibrewise derivative. This ensures that $\tilde{X}_L$, thus its integral curves $\tilde{\mathcal{C}}$, has a certain structure. To be specific, let

$$\mathcal{C}(t) \sim (\mathbf{q}(t), \mathbf{p}(t))$$

be an integral curve of X_H, thus $\mathbf{q}(t), \mathbf{p}(t)$ satisfy Hamilton's equations (Eq. (17)). So

$$\begin{aligned}\tilde{\mathcal{C}}(t) = \mathcal{F}H \circ \mathcal{C}(t) &\sim \Big(q^1(t), \dots, q^d(t), \frac{\partial H}{\partial p_1}\Big|_{(\mathbf{q}(t),\mathbf{p}(t))}, \dots, \frac{\partial H}{\partial p_d}\Big|_{(\mathbf{q}(t),\mathbf{p}(t))}\Big) \\ &\overset{(17)}{=} \big(q^1(t), \dots, q^d(t), \dot{q}^1(t), \dots, \dot{q}^d(t)\big)\end{aligned}$$

is an integral curve of $\tilde{X}_L$. We observe that the velocity coordinates $\mathbf{v}(t)$ of $\tilde{\mathcal{C}}$ are just the time-derivatives of its position coordinates $\mathbf{q}(t)$ (as we would like). Thus $\tilde{X}_L$ always has the form

$$\tilde{X}_L = v^i \frac{\partial}{\partial q^i} + f_L^i \frac{\partial}{\partial v^i}, \tag{25}$$

where $\mathbf{f}_L = f_L^1, \dots, f_L^d$ are some (locally defined) functions of $\mathbf{q}, \mathbf{v}$ implicitly defined by L through Eq. (24). We learn that the integral curves $\tilde{\mathcal{C}}(t) \sim (\mathbf{q}(t), \mathbf{v}(t))$ of $\tilde{X}_L$ satisfy the ODEs

$$\dot{\mathbf{q}}(t) = \mathbf{v}(t) \tag{26}$$

$$\dot{\mathbf{v}}(t) = \mathbf{f}_L(\mathbf{q}(t), \mathbf{v}(t)). \tag{27}$$

In practice, Eq. (26)–(27) is not so useful, because the $\mathbf{f}_L$ are only implicitly known. What we want are E.O.M. formulated directly in terms of the Lagrangian $L = L_H$, which we will now provide.

First, we check the formula (exercise)

$$\tilde{\theta}_L \equiv (\mathcal{F}L)^*\theta = \frac{\partial L}{\partial v^i} dq^i. \tag{28}$$

Then Eq. (24) defining $\tilde{X}_L$ (thus the f_L^i) is, in coordinates,

$$\begin{aligned}0 = \mathcal{L}_{\tilde{X}_L}\tilde{\theta}_L - dL &= \tilde{X}_L\Big(\frac{\partial L}{\partial v^i}\Big)dq^i + \frac{\partial L}{\partial v^i}\mathcal{L}_{\tilde{X}_L}(dq^i) - \frac{\partial L}{\partial q^i}dq^i - \frac{\partial L}{\partial v^i}dv^i \\ &\overset{(25)}{=} v^j \frac{\partial^2 L}{\partial q^j \partial v^i} dq^i + f_L^j \frac{\partial^2 L}{\partial v^j \partial v^i} dq^i + \frac{\partial L}{\partial v^i} d(\underbrace{\iota_{\tilde{X}_L} dq^i}_{v^i \text{ by Eq. (25)}}) - \frac{\partial L}{\partial q^i}dq^i - \frac{\partial L}{\partial v^i}dv^i \\ &= v^j \frac{\partial^2 L}{\partial q^j \partial v^i} dq^i + f_L^j \underbrace{\frac{\partial^2 L}{\partial v^j \partial v^i}}_{\text{Hessian}} dq^i - \frac{\partial L}{\partial q^i}dq^i.\end{aligned} \tag{29}$$

Now substitute the E.O.M. (26)–(27) into (29), to conclude that

$$0 = \frac{dq^j}{dt}\frac{\partial^2 L}{\partial q^j \partial v^i} + \frac{dv^j}{dt}\frac{\partial^2 L}{\partial v^j \partial v^i} - \frac{\partial L}{\partial q^i} \tag{30}$$

$$= \frac{d}{dt}\frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i}. \tag{31}$$

hold along the integral curves of $\tilde{X}_L$. Eq. (31) are the *Euler–Lagrange* equations associated to the Lagrangian L.

Conversely, consider curves in TM of the form

$$(\mathbf{q}(t), \mathbf{v}(t)) = (\mathbf{q}(t), \dot{\mathbf{q}}(t)) \tag{32}$$

which satisfy the Euler–Lagrange equations (31); equivalently, (30) hold. Comparison of (30) with (29) shows that the integral curve condition $\dot{\mathbf{v}} = \mathbf{f}_L$ (Eq. (27)) is satisfied[14].

To summarize, Eq. (31) can be considered as the E.O.M. in velocity phase space TM, to be solved for curves of the form (32).

Remark 2.4. Given $L : TM \to \mathbb{R}$, the associated *total energy* function is

$$\begin{aligned} E_L : TM &\to \mathbb{R} \\ v &\mapsto \mathcal{F}L(v) \cdot v - L(v). \end{aligned}$$

If L is hyperregular, meaning that $\mathcal{F}L : TM \to T^*M$ is a diffeomorphism, then we can define the associated *Hamiltonian* function,

$$H_L := E_L \circ (\mathcal{F}L)^{-1} : T^*M \to \mathbb{R}.$$

It may be checked that H_L is hyperregular, and that

$$L = L_{H_L}, \qquad H = H_{L_H}.$$

Thus we can pass between the Lagrangian and Hamiltonian formulations.

2.4 Riemannian manifolds

So far, we have not said anything about *forces*, which is a prominent concept in Newton's laws of motion. For a *free particle* with *inertial mass*[15] $m > 0$, Newton's First Law says that it undergoes "uniform motion in a straight line[16]". If there is a nett force experienced by the particle, the Second Law states that the motion is accelerated in proportion to the force, with proportionality constant m.

As discussed in Remark 1.7, acceleration — the rate-of-change of velocity along a curve — only makes sense with respect to a connection ∇ on the tangent bundle. Given ∇, the geodesic condition, i.e., zero acceleration condition, reads

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0. \tag{33}$$

Now, even if we work with $M = A$, where a canonical ∇ is available, quite often, we have a *constrained* problem. This means that motion is confined to some submanifold $M' \subseteq A$. This would mean that there are implicit forces conspiring to maintain the motion within M'. For example, the nett force has to be tangential to M'. However, it

[14]To be more precise, under the hyperregularity assumption, the Hessian of L_H is invertible everywhere (see §3.5-3.6 of Abraham–Marsden). So the functions f_L^j satisfying (29) are unique.

[15]Mass is measured as a numerical multiple of, e.g., a "kilogram". In 2019, the "kilogram" reference itself was redefined in terms of Planck's constant from quantum mechanics.

[16]Note the implicit geometric assumption that space is Euclidean. Furthermore, one has to be in an "inertial frames".

is very cumbersome, if at all possible, to work with coordinates on A and write down all these forces explicitly. This is one reason to adopt a formalism intrinsic to M', as the Lagrangian/Hamiltonian does. However, in the intrinsic viewpoint, where would ∇ come from?

Let us consider the affine space A for simplicity. Although velocity is a tangent vector, in practice, measurement devices give the *speed*,

$$||\dot{\gamma}(t)|| = \lim_{\Delta t \to 0} \frac{|\gamma(t+\Delta t) - \gamma(t)|}{|\Delta t|} \in \mathbb{R}, \tag{34}$$

or at least some averaged speed with $\Delta t > 0$ small. Implicit in Eq. (34) is the notion of length of time, $|\Delta t|$, and spatial distance traversed. A measurement device is calibrated against some reference $\frac{\text{yardstick}}{\text{clock tick}}$, so it outputs a numerical multiple of a reference speed.

Today, the standard clock tick is given by some atomic clock (based on quantum mechanics principles!). So a reference speed is obtained if we have a reference yardstick[17]. By requiring rotations to be isometries (circular planetary orbits), we may conclude that the length function on A comes from an inner product on V, i.e., the tangent spaces. When we remember this inner product structure on $V = T_a A$, we are treating A as a Riemannian manifold. This is what "Euclidean space" usually refers to, and one writes $\mathbb{E}^N$ for this Riemannian manifold.

Since the tangent spaces $T_a\mathbb{E}^N$ are inner product spaces, for any submanifold $M \subseteq \mathbb{E}^N$, we can orthogonally project the connection $\nabla^{(A)}$ on $TA = T\mathbb{E}^N$ down to the subbundle $TM \subseteq T\mathbb{E}^N$. The result is precisely the *Levi–Civita connection* on TM,

$$\nabla^{\text{LC}} = \perp \text{ proj} \circ \nabla^{(A)} \circ \text{inclusion}.$$

Intuitively, we perform parallel transport as though we were in the background $\mathbb{E}^N$, then retain only the component tangential to M.

Remark 2.5. Intrinsically, ∇^{LC} is the unique torsion-free connection compatible with the induced Riemannian metric on the submanifold $M \subseteq \mathbb{E}^N$.

2.4.1 Particle motion in a scalar potential

Subsequently, let M be a Riemannian manifold with Riemannian metric[18] g, so we have the squared-norm function

$$||v||^2 = g(v, v), \qquad v \in TM.$$

Similarly, there is a norm/metric on the cotangent spaces,

$$||p||^2 = g(p^\sharp, p^\sharp), \qquad p \in T^*M.$$

Here, we recall that $(\cdot)^\sharp$ is the isomorphism between cotangent and tangent spaces, induced by the metric g through the relation

$$g(p^\sharp, v) = p(v).$$

[17]In relativistic mechanics, there is a reference speed-of-light, therefore there is automatically a reference yardstick.

[18]A smooth assignment of inner products to the tangent spaces.

We also write $(\cdot)^\flat$ for the inverse isomorphism. The cotangent space metric satisfies

$$||v^\flat||^2 = g((v^\flat)^\sharp, (v^\flat)^\sharp)) = g(v, v) = ||v||^2, \qquad \forall v \in TM.$$

For a Riemannian manifold (M, g) and particle mass $m > 0$, the *free particle Lagrangian* and *free particle Hamiltonian* are

$$\begin{aligned} L_{\text{free}} : TM &\to \mathbb{R} & H_{\text{free}} : T^*M &\to \mathbb{R} \\ v &\mapsto \frac{1}{2}m||v||^2, & p &\mapsto \frac{1}{2m}||p||^2. \end{aligned}$$

Notice that

$$\theta(X_{H_{\text{free}}})(p) = p_i dq^i \left(\frac{\partial H_{\text{free}}}{\partial p_j} \frac{\partial}{\partial q^j} \right) = p_i dq^i \left(\frac{g^{jk} p_k}{m} \frac{\partial}{\partial q^j} \right) = \frac{1}{m}||p||^2.$$

The fibre derivative of H_{free} is

$$\begin{aligned} \mathcal{F}H_{\text{free}} : TM &\to T^*M \\ p &\mapsto \frac{1}{2m} d||p||^2 = \frac{1}{m} p^\sharp, \end{aligned}$$

with inverse

$$\mathcal{F}H_{\text{free}}^{-1} : v \mapsto mv^\flat. \tag{35}$$

So the Lagrangian corresponding to H_{free} is

$$L_{H_{\text{free}}} = (\theta(X_{H_{\text{free}}}) - H_{\text{free}}) \circ \mathcal{F}H_{\text{free}}^{-1} : v \mapsto \frac{1}{m}||mv^\flat||^2 - \frac{1}{2m}||mv^\flat||^2 = \frac{1}{2}m||v||^2,$$

which is exactly the free Lagrangian L_{free}.

It is customary to call H_{free} or $L_{\text{free}} = L_{H_{\text{free}}}$ the *kinetic energy* function, and denote either by

$$T = H_{\text{free}} = L_{\text{free}}.$$

A more general Hamiltonian has the form

$$H = T + \pi^* V, \tag{36}$$

where $V \in C^\infty(M)$ is a *scalar potential energy* function depending only on position. Because $\pi^* V \equiv V \circ \pi$ does not depend on fibre coordinates, it follows readily that we still have

$$\theta(X_H) = \theta(X_{H_{\text{free}}}) = \frac{1}{m}||p||^2,$$

as well as $\mathcal{F}H = \mathcal{F}H_{\text{free}}$ (Eq. (35)). Thus, for Riemannian geometry Hamiltonians (Eq. (36)), the relationship between abstract momentum and velocity is the familiar

"momentum = mass × velocity".

(Where one also uses the Riemannian metric to convert momentum cotangent vectors to momentum tangent vectors.)

Finally, the Lagrangian corresponding to H is

$$\begin{aligned} L_H = (\theta(X_H) - H) \circ \mathcal{F}H^{-1} &= \big(\theta(X_{H_{\text{free}}}) - (H_{\text{free}} + V \circ \pi)\big) \circ \mathcal{F}H_{\text{free}}^{-1} \\ &= L_{\text{free}} - V \circ \pi \circ \mathcal{F}H_{\text{free}}^{-1} \\ &= T - \pi^* V. \end{aligned}$$

Let us work out the Euler–Lagrange equation of motion for the Lagrangian L_H above. We have

$$\begin{aligned} \frac{d}{dt}\frac{\partial L_H}{\partial v^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} &= \frac{d}{dt}\frac{\partial T}{\partial v^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} \\ &= m\frac{d}{dt}\left(g_{ij}(\mathbf{q}(t))\dot{q}^j(t)\right) \\ &= m g_{ij}(\mathbf{q}(t))\ddot{q}^j(t) + m\frac{\partial g_{ij}}{\partial q^k}\Big|_{\mathbf{q}(t)}\dot{q}^k(t)\dot{q}^j(t) \end{aligned}$$

and

$$\frac{\partial L_H}{\partial q^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} = \frac{1}{2}m\frac{\partial g_{jk}}{\partial q^i}\Big|_{\mathbf{q}(t)}\dot{q}^j(t)\dot{q}^k(t) - \frac{\partial V}{\partial q^i}\Big|_{\mathbf{q}(t)}.$$

The Euler–Lagrange equation is thus

$$\begin{aligned} 0 &= \frac{d}{dt}\frac{\partial L_H}{\partial v^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} - \frac{\partial L_H}{\partial q^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} \\ &= m\left(g_{ij}\ddot{q}^j + \partial_k g_{ij}\dot{q}^k\dot{q}^j - \frac{1}{2}\partial_i g_{jk}\dot{q}^j\dot{q}^k\right) + \frac{\partial V}{\partial q^i} \\ &= m\left(g_{ij}\ddot{q}^j + \frac{1}{2}(\partial_k g_{ij} + \partial_j g_{ik} - \partial_i g_{jk})\dot{q}^j\dot{q}^k\right) + (dV)_i. \end{aligned}$$

Raising indices with the metric, we arrive at the following equation for the motions $\gamma(t) \sim \mathbf{q}(t)$,

$$0 = m\left(\ddot{q}^l + \Gamma^l_{jk}\dot{q}^j\dot{q}^k\right) + ((dV)^\sharp)^l\Big|_{\mathbf{q}(t)}, \tag{37}$$

where Γ^l_{jk} are the Christoffel symbols[19] associated to g. In coordinate-independent notation,

$$-\operatorname{grad} V\Big|_{\gamma(t)} \equiv -(dV)^\sharp\Big|_{\gamma(t)} = m\underbrace{\nabla_{\dot\gamma(t)}\dot\gamma(t)}_{\text{acceleration}}, \tag{38}$$

where $\nabla = \nabla^{\text{LC}}$ is the Levi–Civita covariant derivative. The left side of (38) is the conservative *force* vector field. So (38) is precisely Newton's second law,

$$\text{Force} = \text{mass} \times \text{acceleration}, \qquad (\text{holds along motions } \gamma)$$

on the Riemannian manifold (M, g).

[19]These are the connection coefficients of the Levi–Civita connection on TM, in terms of a chart-induced local trivialization $(\mathbf{q}, \mathbf{v})$. Explicitly,

$$\Gamma^l_{jk} = \frac{1}{2}g^{li}\big(\partial_j g_{ik} + \partial_k g_{ij} - \partial_i g_{jk}\big).$$

2.4.2 Gauge for metric?

When we fix a reference yardstick and a clock tick, we have a reference speed $\frac{\text{yardstick}}{\text{clock tick}}$. When applied to a velocity vector, the numerical Riemannian metric is understood to be multiplied to this reference speed^2. So if we scale the reference yardstick/clock by a factor λ, the numerical metric would be scaled by λ^{-2}.

H. Weyl was worried about the possibility that there is secretly some non-trivial "parallel transport of "yardstick/clock" it self. This led to his invention of *gauge theory* in 1918. His suggestion was that any yardstick was merely a "local gauge[20]", which could acquire some scale factor $\lambda \neq 1$ after being brought around a closed loop. The criterion for "parallel transport" of yardsticks would be specified by a *gauge field* associated with the non-compact abelian group $(\mathbb{R}_{>0}, \times)$ of scalings[21]. Furthermore, this gauge field was postulated to be the potentials appearing in electromagnetic theory.

Although this particular proposal did not work out, it turned out that the basic concept was correct. It is just that electromangetic potentials are linked to the compact U(1)-phase freedom in quantum mechanics, rather than the noncompact $(\mathbb{R}_{>0}, \times)$-scaling freedom in metrics.

2.5 Charged particle motion in an electromagnetic field

So far, we have discussed motion of a particle with no intrinsic property other than its mass m. In electromagnetism, one learns that there is another particle property called *electric charge*[22] Q, relevant to motion in the presence of an electric field $\mathbf{E}$ and/or a *magnetic flux density* field[23] $\mathbf{B}$ over M, the motion of a charged particle is modified.

Generally, $\mathbf{E}, \mathbf{B}$ are themselves dynamical objects (i.e., time-dependent vector fields), which, together with electric charges and currents, obey equations of motion called *Maxwell's equations*. For example, the following two Maxwell equations hold:

$$\operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{39}$$

$$\operatorname{div} \mathbf{B} = 0. \tag{40}$$

There are two other Maxwell equations[24], but we will not need them. Furthermore, for this section, we make the simplifying assumption that $\mathbf{E}, \mathbf{B}$ are *time-independent* "background" vector fields, and are unaffected by the charged particle's motion[25].

Eq. (39)–(40) apparently require M to be 3-dimensional. Below, we will use an intrinsic formulation in terms of differential forms, which works on any manifold.

[20] In ordinary language, "gauge" refers to some standard of measure, for example; railway track gauges vary from location to location.

[21] Note that the problem of scaling ambiguity is *not* addressed by working with metrics modulo rescaling functions (conformal geometry).

[22] Electric charge is a numerical multiple of the elementary electron charge e.

[23] Often $\mathbf{B}$ is called the "magnetic field". There is a distinct vector field, $\mathbf{H}$, also called the magnetic field, which is closely related to $\mathbf{B}$. To avoid confusion, we use the term *magnetic flux density* for $\mathbf{B}$, which is closer to its geometric meaning.

[24] These involve the $\mathbf{H}$ field and electric displacement field $\mathbf{D}$, and sources (charges/currents).

[25] A moving charge is itself a source of electromagnetic fields, via the remaining Maxwell equations. So we are basically assuming that Q is very small and produces a negligible field compared to $\mathbf{B}, \mathbf{E}$.

2.5.1 Electric 1-form

An electric field $\mathbf{E}$ is a tangent vector field over M. We can use the Riemannian metric to regard $\mathbf{E}$ field as a 1-form

$$\mathcal{E} = \mathbf{E}^\flat \in \Omega^1(M).$$

In physics, an electric field is synonymous with "force-per-unit-charge", meaning that a test particle of charge Q experiences a force 1-form

$$F_{\text{electric}} = Q\mathcal{E}.$$

Now, the differential form version of the Maxwell equation (39) is

$$d\mathcal{E} = 0.$$

So, locally, one can write

$$\mathcal{E} = -d\phi,$$

where $\phi : M \to \mathbb{R}$ is an *electric potential* function, determined up to an overall constant. So

$$F_{\text{electric}} = Q\mathcal{E} = -d(Q\phi).$$

Ignoring global issues, $V = Q\phi \in C^\infty(M)$ is a typical example of a scalar potential energy term that enters the Hamiltonian/Lagrangian $T \pm V \circ \pi$,

$$H^{(\phi)} = T + Q\pi^*\phi, \qquad L^{(\phi)} = T - Q\pi^*\phi.$$

2.5.2 Magnetic 2-form

The story is more subtle for $\mathbf{B}$. The idea of a magnetic flux density might come from inspecting the lines formed by iron filings under the influence of a magnet. The "flux density" refers to the number of such lines piercing through a unit-area surface. Furthermore, in three-dimensional space, it is observed that a moving particle with mass m and charge Q experiences the following acceleration,

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = \frac{Q}{m}\dot\gamma(t) \times \mathbf{B}(\gamma(t)). \tag{41}$$

The left side of (42) multiplied by m, is called the *Lorentz* force. So we have the Lorentz force law[26],

$$F_{\text{magnetic}}(\gamma(t))^\sharp = Q\dot\gamma(t) \times \mathbf{B}(\gamma(t)). \tag{42}$$

Puzzle: On the right side of (45), the cross product of vectors (in three dimensions) depends on a choice of orientation, but the force on the left side of (45) makes no reference to orientation!

Resolution: The right side of (42) is actually an *orientation-dependent* way of spelling out how the acceleration depends on $\mathbf{B}$, and the most common convention is the "right-hand rule[27]" for the cross product. Since the cross product picks up a minus sign when we switch orientation choice, it must be the case that $\mathbf{B}$ also picks up a minus sign. Because of this behaviour, physicists call $\mathbf{B}$ a *pseudovector* field, in contrast to an ordinary (tangent) vector field.

[26]Often, the electric force F_{electric} is also included in "Lorentz force".

[27]The *naming* of this convention as "right-hand rule" is itself a convention!

In modern differential geometric language, a choice of orientation is an example of a *gauge choice* (for the bundle of orientations on M), and $\mathbf{B}$ is a *gauge-dependent* object. One of the main lessons of modern physics is that only quantities which are *gauge-independent*[28] are observable. For example, the lines formed by iron filings are gauge-independent, but any arrows that we decide to draw on them are *conventional*. The Earth's $\mathbf{B}$ points from the geographic South pole to the geographic North pole, by convention. Let us thus formulate the right side of the Lorentz force law, (42), in a manifestly gauge-independent way.

Recall that on an oriented Riemannian manifold (M, g), there is a *Hodge dual* operation $\star : \Omega^k(M) \to \Omega^{d-k}(M)$: the Hodge dual of $\zeta \in \Omega^k(M)$ is defined by the condition

$$\eta \wedge \star\zeta = g(\eta, \zeta)\mathrm{vol}_{M,g}, \qquad \forall\, \eta \in \Omega^k(M).$$

Above, $\mathrm{vol}_{M,g}$ is the Riemannian volume form, which evaluates to 1 on any oriented orthonormal tangent frame. For example,

$$\star 1 = \mathrm{vol}_{M,g}.$$

If it happens that M is 3-dimensional, then the cross product of (tangent) vectors $\mathbf{v}, \mathbf{w}$ can be formulated as

$$(\mathbf{v} \times \mathbf{w})^\flat = \star(\mathbf{v}^\flat \wedge \mathbf{w}^\flat). \tag{43}$$

The orientation-dependent formulation of the Lorentz force law, (42), can be written as

$$\begin{aligned} F_{\text{magnetic}}(\gamma(t)) &= -Q\big(\mathbf{B}(\gamma(t)) \times \dot\gamma(t)\big)^\flat \\ &= -Q \star \big(\mathbf{B}(\gamma(t))^\flat \wedge \dot\gamma(t)^\flat\big) \qquad\qquad \text{(Eq. (43))} \\ &= -Q\iota_{\dot\gamma(t)} \star \mathbf{B}^\flat(\gamma(t)). \end{aligned}$$

For the last equality, we used the identity $\star(a \wedge b) = \iota_{a^\sharp} \star b$, valid for any 1-form b (Exercise). So if we write

$$\mathcal{B} = \star\mathbf{B}^\flat \in \Omega^2(M),$$

the Lorentz 1-form law is

$$F_{\text{magnetic}}(\gamma(t)) = -Q\iota_{\dot\gamma(t)}\mathcal{B}. \tag{44}$$

The 2-form $\mathcal{B} \in \Omega^2(M)$ is called the *magnetic flux density*[29]. Eq. (44) is manifestly orientation-independent, and makes sense in any dimension. In fact, a Riemannian metric is only needed when converting (44) into a statement about acceleration vectors,

$$\nabla_{\dot\gamma(t)}(\gamma(t)) = -\frac{Q}{m}(\iota_{\dot\gamma(t)}\mathcal{B})^\sharp. \tag{45}$$

(Generally, mass m appears when a metric does.)

Remark 2.6. If you have dealt with angular momentum "vectors" in 3D, you might have learned that they are also pseudovectors, thus more properly understood as 2-forms. There are close relationships between magnetic fields, rotations, and "spin angular momentum" from quantum mechanics, to be explored later.

[28] Physicists sometimes say "gauge-invariant" or "gauge-covariant".

[29] It can be integrated over an oriented 2-dimensional surface $S \subseteq M$, to get the *magnetic flux* through S.

Here is another advantage of working with the 2-form $\mathcal{B}$. Suppose $\tilde{\mathcal{B}} \in \Omega^2(\mathbb{R}^3)$ is the magnetic flux density on a background space $\mathbb{R}^3$, and $j : M \hookrightarrow \mathbb{R}^3$ is a submanifold on which motion is confined. Then for a path γ in M, the tangential part of the Lorentz force is readily seen to be

$$j^*F(\gamma(t)) = -Q \cdot j^*\iota_{\dot{\gamma}(t)}\tilde{\mathcal{B}} = -Q \cdot \iota_{\dot{\gamma}(t)}(j^*\tilde{\mathcal{B}}).$$

(Whatever the background Riemannian metric, the normal components of the Lorentz force will, by fiat, be cancelled out by other complicated forces.) The upshot is that the intrinsic magnetic flux density felt by a charged particle confined to M is the 2-form $\mathcal{B} = j^*\tilde{\mathcal{B}}$ on M.

Note that the Lorentz force is only defined along γ, and depend on $\dot{\gamma}$. It does not constitute a "force field/1-form" over M, and it does not make sense to try to write $F = -dV$ for some scalar potential function $V \in C^\infty(M)$. The appropriate way to account for (45) in Lagrangian/Hamiltonian mechanics is as follows. In terms of differential forms, the Maxwell equation (40), is

$$d\mathcal{B} = 0. \tag{46}$$

So locally, we can always choose a 1-form $\mathcal{A}$ such that

$$d\mathcal{A} = \mathcal{B}. \tag{47}$$

Such an $\mathcal{A}$ is called a *magnetic (vector) potential*[30].

For simplicity, *let us assume that $\mathcal{A}$ can be globally chosen*, $\mathcal{A} \in \Omega^1(M)$. Modify the Hamiltonian $H = T + \pi^*V$ to

$$\begin{aligned} H^{(\mathcal{A})} : T^*M &\to \mathbb{R} \\ p &\mapsto \frac{1}{2m}\big|\big|p - Q\mathcal{A}|_{\pi(p)}\big|\big|^2 + V(\pi(p)). \end{aligned} \tag{48}$$

It is straightforward to check that

$$\mathcal{F}H^{(\mathcal{A})} : p \mapsto \frac{1}{m}(p - Q\mathcal{A}|_{\pi(p)})^\sharp,$$

so that

$$(\mathcal{F}H^{(\mathcal{A})})^{-1} : v \mapsto m(v^\flat + Q\mathcal{A}|_{\pi(v)}).$$

We no longer have "momentum = mass × velocity"[31].

We proceed to calculate the corresponding Lagrangian, (exercise)

$$\begin{aligned} L_{H^{(\mathcal{A})}} = \big(\theta(X_{H^{(\mathcal{A})}}) - H^{(\mathcal{A})}\big) \circ (\mathcal{F}H^{(\mathcal{A})})^{-1} : TM &\to \mathbb{R} \\ v \mapsto \cdots &= \frac{1}{2}m||v||^2 + \frac{1}{m}g(Q\mathcal{A}|_{\pi(v)}, mv^\flat) - V(\pi(v)) \\ &= (T - \pi^*V)(v) + Q\mathcal{A}(v), \end{aligned}$$

comprising the usual $T - \pi^*V$ term and an extra $Q\mathcal{A}(v)$ term.

[30] In three dimensions, $\mathcal{A}^\sharp$ is a locally-defined *magnetic vector potential* such that $\operatorname{curl}\mathcal{A}^\sharp = \mathbf{B}$.

[31] In physics, one uses "canonical momentum" to distinguish it from "kinetic momentum = mass × velocity".

In the Euler–Lagrange equation for $L_{H^{(\mathcal{A})}}$, the $T - \pi^* V$ term contributes the usual force term $-dV$. It remains to work out the contribution of the $Q\mathcal{A}(\cdot)$ term. For the latter, we calculate

$$\frac{d}{dt}\frac{\partial \mathcal{A}(v)}{\partial v^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} = \frac{d}{dt}\mathcal{A}_i|_{\mathbf{q}(t)} = \frac{\partial \mathcal{A}_i}{\partial q^j}\Big|_{\mathbf{q}(t)} \dot{q}^j(t),$$

and

$$\frac{\partial \mathcal{A}(v)}{\partial q^i}\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} = \frac{\partial \mathcal{A}_j}{\partial q^i}\Big|_{\mathbf{q}(t)} \dot{q}^j(t).$$

Thus

$$\left(\frac{d}{dt}\frac{\partial \mathcal{A}(v)}{\partial v^i} - \frac{\partial \mathcal{A}(v)}{\partial q^i}\right)\Big|_{(\mathbf{q}(t),\dot{\mathbf{q}}(t))} = \left(\frac{\partial \mathcal{A}_i}{\partial q^j} - \frac{\partial \mathcal{A}_j}{\partial q^i}\right)\Big|_{\mathbf{q}(t)} \dot{q}^j(t),$$

which we recognize as the i-th component of (exercise)

$$\iota_{\dot\gamma(t)}\mathcal{B} = \iota_{\dot\gamma(t)} d\mathcal{A} = \left(\frac{\partial \mathcal{A}_i}{\partial q^j} - \frac{\partial \mathcal{A}_j}{\partial q^i}\right)\Big|_{\mathbf{q}(t)} \dot{q}^j(t) dx^i.$$

We conclude that the Euler–Lagrange E.O.M. associated to $L^{(\mathcal{A})}$ is

$$0 = m\nabla_{\dot\gamma(t)}\dot\gamma(t) + (\operatorname{grad} V|_{\gamma(t)}) + Q(\iota_{\dot\gamma(t)}\mathcal{B})^\sharp.$$

The extra force 1-form due to $\mathcal{B}$ is thus $-Q\iota_{\dot\gamma(t)}\mathcal{B}$, which is precisely the Lorentz force, (44).

2.6 Hamilton–Jacobi equation (sketch)

Let $H : T^*M \to \mathbb{R}$ be a Hamiltonian. Recall that by virtue of the canonical 1-form $\theta \in \Omega^1(T^*M)$, we get the Hamiltonian vector field $X_H \in \mathfrak{X}(T^*M)$ via Eq. (15). Points in T^*M evolve along integral curves of X_H. Let us write the local flow along these curves as a time-parametrized family of local diffeomorphisms,

$$f_t : T^*M \to T^*M, \qquad t \in I.$$

These f_t are actually *symplectomorphisms*, or *canonical transformations*, in the sense that $\omega = -d\theta$ is preserved, i.e., $f_t^*\omega = \omega$ (Exercise). Since

$$0 = \omega - f_t^*\omega = -d\theta + f_t^* d\theta = d(f_t^*\theta - \theta),$$

we can locally write

$$f_t^*\theta - \theta = d\check{S}_t \tag{49}$$

for some locally defined functions $\check{S}_t : T^*M \to \mathbb{R}$. In fact, from (19), we can take $\check{S}_t$ to be the *action integral*

$$\check{S}_t = \int_0^t \Big(\underbrace{\theta(X_H) - H}_{\text{Langrangian } L_H \circ \mathcal{F}H} \Big) \circ f_s \, ds. \tag{50}$$

(See §3.2.9 of Thirring.)

Let us spell out $\check{S}_t$ using a bundle chart for $T^*M|_U$, with position-momentum coordinates denoted $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$. If $p \in T^*M|_U$ has initial position-momentum coordinates $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$, then after time t, it evolves to the point $f_t(p)$ with coordinates

$$(\mathbf{q}, \mathbf{p})(y; t) := (\bar{\mathbf{q}} \circ f_t, \bar{\mathbf{p}} \circ f_t)(y) = (f_t^* \bar{\mathbf{q}}, f_t^* \bar{\mathbf{p}})(y). \tag{51}$$

(At least for small t.) To avoid confusion, in $(\mathbf{q}, \mathbf{p})$, the $\mathbf{p}$ are not necessarily momentum coordinates induced by regarding $\mathbf{q}$ as a local chart on M.

Next, because $\omega = \bar{p}_i dq^i$, we have

$$f_t^* \theta = f_t^*(\bar{p}_i d\bar{q}^i) \overset{(51)}{=} p_i dq^i,$$

at each time t. Eq. (49) becomes the equality of local 1-forms on T^*M,

$$p_i dq^i - \bar{p}_i d\bar{q}^i = d\check{S}_t. \tag{52}$$

Now we make the assumption that the map, $(\bar{\mathbf{q}}, \bar{\mathbf{p}}; t) \mapsto (\mathbf{q}, \mathbf{p})$, satisfies

$$\det \left(\frac{\partial q^i}{\partial \bar{p}^j} \right) \neq 0$$

at every $\bar{\mathbf{q}}$ and t. Then, on some possibly smaller domain in $T^*M|_U$, we can invert the relationship $\mathbf{q} \equiv \mathbf{q}(\bar{\mathbf{q}}, \bar{\mathbf{p}}; t)$ to obtain

$$\bar{\mathbf{p}} \equiv \bar{\mathbf{p}}(\bar{\mathbf{q}}, \mathbf{q}; t).$$

In other words, we can switch to using initial/final-positions $(\bar{\mathbf{q}}, \mathbf{q})$ instead of initial position-momenta $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$ as the local coordinates for T^*M. Taking the $(\bar{\mathbf{q}}, \mathbf{q})$ perspective, we write

$$S \equiv S(\bar{\mathbf{q}}, \mathbf{q}; t) := \check{S}_t(\bar{\mathbf{q}}, \bar{\mathbf{p}}(\bar{\mathbf{q}}, \mathbf{q}; t)), \tag{53}$$

which is called the *characteristic function* determined by H.

Let us investigate the partial derivatives of $S \equiv S(\bar{\mathbf{q}}, \mathbf{q}; t)$. The defining Eq. (52) reads

$$p_i dq^i = \bar{p}_i d\bar{q}^i + \frac{\partial S}{\partial \bar{q}^i} d\bar{q}^i + \frac{\partial S}{\partial q^i} dq^i.$$

Comparing coefficients of $d\bar{q}^i$ and dq^i, we get

$$\frac{\partial S}{\partial \bar{q}^i} = -\bar{p}_i, \tag{54}$$

$$\frac{\partial S}{\partial q^i} = p_i, \qquad i = 1, \ldots, d. \tag{55}$$

The remaining t-partial derivative is

$$\begin{aligned}
\frac{\partial S}{\partial t} &= \frac{\partial \check{S}_t}{\partial t} + \frac{\partial \check{S}_t}{\partial \bar{p}_i}\frac{\partial \bar{p}_i}{\partial t} && \text{(Eq. (53))} \\
&= \frac{\partial \check{S}_t}{\partial t} - \frac{\partial \check{S}_t}{\partial \bar{p}_i}\frac{\partial \bar{p}_i}{\partial q^j}\dot{q}^j && (\frac{d\bar{p}_i}{dt} = 0 = \frac{\partial \bar{p}_i}{\partial t} + \frac{\partial \bar{p}_i}{\partial q^j}\dot{q}^j) \\
&= \frac{\partial \check{S}_t}{\partial t} - \frac{\partial S}{\partial q^j}\dot{q}^j && \text{(Eq. (53))} \\
&= L_H \circ \mathcal{F}H \circ f_t \ - \frac{\partial S}{\partial q^j}\dot{q}^j && \text{(Eq. (50))} \\
&= L_H \circ \mathcal{F}H \circ f_t - \underbrace{p_j \dot{q}^j}_{\theta(X_H)\circ f_t} && \text{(Eq. (55))} \\
&= -H \circ f_t.
\end{aligned}$$

When spelt out in terms of $(\bar{\mathbf{q}}, \mathbf{q})$ coordinates, the last equation for $\frac{\partial S}{\partial t}$ is

$$\frac{\partial}{\partial t} S(\bar{\mathbf{q}}, \mathbf{q}; t) + H\left(\mathbf{q}, \frac{\partial}{\partial q^1} S(\bar{\mathbf{q}}, \mathbf{q}; t), \ldots, \frac{\partial}{\partial q^d} S(\bar{\mathbf{q}}, \mathbf{q}; t)\right) = 0. \tag{56}$$

Remark 2.7. Let a Hamiltonian function $H : T^*M \to \mathbb{R}$ be given, and suppose $S \equiv S(\bar{\mathbf{q}}, \mathbf{q}; t)$ satisfies (56). The time evolution due to H is actually encoded within S (or rather, its partial derivatives), in the following way. Define initial/final momenta $\bar{\mathbf{p}}, \mathbf{p}$ by the partial derivatives of S, as in (54)–(55). Suppose initial data $\bar{\mathbf{q}}, \bar{\mathbf{p}}$ and a final time t are given. Then Eq. (54),

$$\frac{\partial S}{\partial \bar{q}^i}(\bar{\mathbf{q}}, \mathbf{q}; t) = -\bar{p}_i(\bar{\mathbf{q}}, \mathbf{q}; t),$$

specify d equations in the d unknowns $\mathbf{q}$. Solving these gives the future positions $\mathbf{q}$ in terms of $(\bar{\mathbf{q}}, \bar{\mathbf{p}}; t)$. Then Eq. (55) specify $\mathbf{p} \equiv \mathbf{p}(\bar{\mathbf{q}}, \mathbf{q}(\bar{\mathbf{q}}, \bar{\mathbf{p}}; t); t)$ as well. Altogether, we get curves $(\mathbf{q}, \mathbf{p})$ in phase space, labelled by initial $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$ and parametrized by t. One then checks that Hamilton's E.O.M. are satisfied.

Remark 2.7 is meant convey the idea that classical Hamiltonian mechanics for paths $\mathbf{q}(t)$ has a "wave mechanics" reformulation, in terms of a PDE for S. This PDE, Eq. (56), is often stated as the *Hamilton–Jacobi equation*

$$\frac{\partial}{\partial t} S(\mathbf{q}; t) + H\left(\mathbf{q}, \frac{\partial}{\partial \mathbf{q}} S(\mathbf{q}; t)\right) = 0, \tag{57}$$

for *Hamilton's principal ("wave")function* $S \equiv S(\mathbf{q}; t)$.

When studying H with no explicit t-dependent, one usually separates variables, taking

$$S(\mathbf{q}; t) = W(\mathbf{q}) - Et, \qquad E \in \mathbb{R}. \tag{58}$$

Then the Hamilton–Jacobi equation is reduced to

$$H\left(\mathbf{q}, \frac{\partial}{\partial q^j} W(\mathbf{q})\right) = E, \tag{59}$$

2.6.1 Schrödinger equation and wavefunction for a single particle

With a view towards *quantum* wave mechanics, we abandon the premise that there are objective paths deterministically specified by sharp initial values of $\bar{\mathbf{q}}, \bar{\mathbf{p}}$. Instead, we take the wave $S(\mathbf{q}; t)$ as the a classical mechanical analogue to a better model object evolving according to *quantum* mechanics.

It turns out that S itself does not work, and the simplest ansatz which works is that $S/\hbar$ is the $\mathbb{R}$-valued phase function for a quantum mechanical $\mathbb{C}$-valued wavefunction[32],

$$\Psi = R\exp(iS/\hbar), \qquad R = |\Psi| \in C^\infty(M).$$

Here, $\hbar$ is some reference "unit of action[33]" (i.e., having units energy $\times$ time), needed to turn S into a numerical $\mathbb{R}$-valued function that can be exponentiated to a $\mathbb{C}$-valued function.

For now, let consider a Riemannian metric and scalar potential function V, determining the classical Hamiltonian function $H(\mathbf{q}, \mathbf{p}) = \frac{1}{2m}||\mathbf{p}||^2 + V(\mathbf{q})$. This describes a single, isolated particle of mass m in a potential V. The associated Hamilton–Jacobi equation is

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}||dS||^2 + V. \tag{60}$$

In comparison, the *Schrödinger equation* for Ψ (associated to the metric and scalar potential V) is

$$i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\Psi + V\cdot\Psi, \tag{61}$$

where Δ denotes the scalar Laplacian,

$$\Delta u = \frac{1}{\sqrt{|g|}}\partial_j(\sqrt{|g|}g^{jk}\partial_k u).$$

Substitute $\Psi = R\exp(iS/\hbar)$ into the Schrödinger equation. The real part gives (Exercise)

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}g^{jk}\partial_k S\partial_j S + V + \mathcal{O}(\hbar^2) = \frac{1}{2m}||dS||^2 + V + \mathcal{O}(\hbar^2),$$

where $\mathcal{O}(\hbar^2)$ denotes terms with a $\hbar^2$ factor. The imaginary part involves a $\hbar$ factor throughout.

Thus, provided we are studying phenomena where $\hbar$ is much smaller than typical actions[34], the Schrödinger equation for $\Psi = R\exp(iS/\hbar)$ approximately reduces to the classical Hamilton–Jacobi equation (60) for the phase function S.

---

[32] After you learn about the geometric meaning of $\mathcal{A}$, revisit this ansatz, and think about it in terms of compactness of the Lie group $\sim \mathrm{U}(1)$ corresponding to the 1-dimensional abelian Lie algebra.

[33] Nature does provide such a reference, namely, *Planck's constant*.

[34] Actually, $\hbar$ is a universal unit of action. Since 2019, mass is, by definition, measured in terms of Planck's constant (and meters, second), via

$$\text{kilogram} := \frac{1}{6.62607015}\times 10^{34}(\text{Planck's constant})\cdot(\text{metre})^{-2}\cdot(\text{second}).$$

So the classical regime is, more precisely, the regime where

$$\text{mass}\times\text{distance}^2/\text{time} \sim \text{momentum}\times\text{distance}$$

is large compared to $\hbar$ in the above sense. For example, a 1 kilogram ball moving at distance/time scales of meters and seconds, would involve huge (10^{34}) action compared to $\hbar$.

Remark 2.8. We mention that Δ, thus the Schrödinger equation, is independent of the choice of coordinate chart. However, this property, together with the property of reducing to (60) as a classical $\hbar \to 0$ limit, does not fix the form of the Schrödinger uniquely. For example, we could modify the Laplacian Δ by adding some multiple of the scalar curvature function.

The Schrödinger equation, (61), is usually written as

$$i\hbar\frac{\partial\Psi}{\partial t} = H\Psi, \qquad H = -\frac{\hbar^2}{2m}\Delta + V,$$

with H now understood as a linear, *differential operator* acting on wavefunctions Ψ, rather than a function of q^i, p_j. Formally, we are making the replacement

$$p_j \leftrightarrow -i\hbar\frac{\partial}{\partial q^j}, \qquad q^j \leftrightarrow \text{multiplication by coordinate function } q^j,$$

when converting the classical H function to the quantum mechanical H operator.

Furthermore, because H is time-independent, for the factorized S (Eq. (58)), we have

$$\Psi(\mathbf{q}; t) = \psi(\mathbf{q}) \cdot e^{-iEt/\hbar}, \qquad E \in \mathbb{R},$$

where $\psi = R\exp(iW/\hbar)$. The Schrödinger equation reduces to an eigenvalue equation

$$E\psi = H\psi, \qquad \psi \equiv \psi(\mathbf{q}),$$

called the *time-independent Schrödinger equation*. The spectral properties[35] of the Hamiltonian operator H thus play a prominent role in quantum mechanics.

2.6.2 Gauge potential ambiguity

Eq (61) describes a *chargeless* and *spinless* "quantum particle". The ansatz used to "derive" it seems ad-hoc, but let us see that it is actually geometrically very natural. The first physics-related point is that particles exhibiting non-classical behaviour (thus requiring quantum mechanics), usually have electric charge $Q \neq 0$, e.g. electrons.

Recall from Section 2.5.2 that when there is a magnetic flux density $\mathcal{B} \in \Omega^2(M)$, we *chose* a magnetic potential 1-form $\mathcal{A}$ satisfying $d\mathcal{A} = \mathcal{B}$, and declared the classical Hamiltonian $H^{(\mathcal{A})}$ for a charge-Q particle to be

$$H^{(\mathcal{A})} = (f^{(\mathcal{A})})^* H,$$

where

$$\begin{aligned} f^{(\mathcal{A})} : T^*M &\to T^*M \\ p &\mapsto p - Q\mathcal{A}|_{\pi(p)}. \end{aligned}$$

In other words, the momentum coordinate is adjusted in accordance with $\mathcal{A}$, whilst fixing the base position, before evaluating the old Hamiltonian.

[35] One considers the Hilbert space $L^2(M)$ of wavefunctions, and H is formally self-adjoint on the dense subspace $C_c^\infty(M)$.

Importantly, the choice of $\mathcal{A}$ can be freely modified to $\tilde{\mathcal{A}} = \mathcal{A} + d\Lambda$ for any $\Lambda \in C^\infty(M)$, while maintaining the criterion $d\tilde{\mathcal{A}} = \mathcal{B}$. Thus

$$H^{(\mathcal{A}+d\Lambda)} = (f^{(d\Lambda)})^* H^{(\mathcal{A})}$$

is an equally valid Hamiltonian.

The map $f = f^{(d\Lambda)}$ is actually a particular kind of canonical transformation,

$$\begin{aligned} f^*\omega = f^* d\theta = df^*\theta = df^*(p_i dq^i) &= d\big((p_i \circ f) dq^i\big) \\ &= d\big(p_i - Q\partial_i\Lambda) dq^i\big) \\ &= d(p_i dq^i) - Q(\partial_j\partial_i\Lambda \cdot\ dq^j \wedge dq^i) \\ &= d(p_i dq^i) \qquad\qquad \text{((skew)symmetry)} \\ &= \omega. \end{aligned}$$

From this, we deduce that the Hamiltonian vector fields $X_{H^{(\mathcal{A}+d\Lambda)}}$ and $X_{H^{(\mathcal{A})}}$ are related by the diffeomorphism $f^{(d\Lambda)}$, and likewise for their integral curves. Since $f^{(d\Lambda)}$ preserves basepoints, the resulting motions on M do not suffer from the ambiguity in choice of $\mathcal{A}$.

So, there is no canonical choice of $\mathcal{A}$, thus $H^{(\mathcal{A})}$. Similarly, $S \equiv S^{(\mathcal{A})}$ satisfying the Hamilton–Jacobi equation for $H^{(\mathcal{A})}$, (59) is not canonical. Specifically, we know that S comes from an action integral (integrating the Lagrangian, Eq. (50)). When we modify $\mathcal{A}$ by $d\Lambda$, the Lagrangian is modified by $Qd\Lambda$, and the action integral is modified by

$$\Delta\check{S}_t(\bar{\mathbf{q}}, \bar{\mathbf{p}}) = \int_0^t Qd\Lambda|_{\mathbf{q}(s)}\, ds = \Lambda(\mathbf{q}(\bar{\mathbf{q}}; t)) - \Lambda(\bar{\mathbf{q}}).$$

So S gets modified by Λ and possibly an overall constant.

These ambiguities do not matter for determining the classical paths in M. However, the "wave" S, or W in case of time-independent H, is ambiguous, and we cannot interpret it as a literal observable wave. Later on, we will explain how $\mathcal{A}$, or rather $i\mathcal{A}$ is understood as a U(1)-connection 1-form (with experimental justification!). This is tightly related to the ansatz that (ambiguous) W is the ambiguous phase function $\psi = R\exp(iW/\hbar)$. Under $\mathcal{A} \to \mathcal{A} + d\Lambda$, we have

$$\psi \to \exp(iQ\Lambda/\hbar) \cdot \psi. \tag{62}$$

This ambiguity is called the *local* U(1) *gauge ambiguity* in quantum mechanical wavefunctions.

In view of (62), it is *not* correct to regard ψ as a scalar $\mathbb{C}$-valued function on M, with objective measurable pointwise $\mathbb{C}$-values. There is automatically the problem of comparing $\psi(x)$ with $\psi(x')$ for distinct $x, x' \in M$, and therefore what it means to differentiate ψ. The solution to this problem is already apparent from the replacement of p_j with $Q\mathcal{A}_j$ in the classical Hamiltonian. Namely, for the quantum mechanical Hamiltonian corresponding to $H^{(\mathcal{A})}$, we have

$$p_j - Q\mathcal{A}_j \ \leftrightarrow\ -i\hbar\left(\partial_j - i\frac{Q}{\hbar}\mathcal{A}_j\right). \tag{63}$$

replacing $-i\hbar\partial_j$. The operator in brackets is called the j-th *gauge-covariant partial derivative*. Let us explain this terminology. Under the modification $\mathcal{A} \to \tilde{\mathcal{A}} = \mathcal{A} + d\Lambda$,

we transform ψ according to (62). Correspondingly, the conjugated gauge-covariant partial derivative operator is

$$\exp(iQ\Lambda/\hbar)\left(\partial_j - i\frac{Q}{\hbar}\mathcal{A}_j\right)\exp(-iQ\Lambda/\hbar) = \partial_j - i\frac{Q}{\hbar}(\mathcal{A}_j + \partial_j\Lambda) = \partial_j - i\frac{Q}{\hbar}\tilde{\mathcal{A}}_j,$$

which is exactly the gauge-covariant partial derivative with the modified $\tilde{\mathcal{A}}$.

Therefore, despite the apparent $\mathcal{A}$ ambiguity causing apparently different wavefunctions and Schrödinger equations, all the different choices are unitarily related to each other via (62). Physically meaningful quantities must be invariant under such transformations, the spectrum of $H^{(\mathcal{A})}$ being a typical example. The squared absolute value, $|\psi|^2 = R^2$ is unambiguously a well-defined $\mathbb{R}_{\geq 0}$-valued function on M, and has physical meaning is the spatial probability density function for the "quantum state" ψ.

A non-physical quantity is the "U(1)-phase of ψ at $x \in M$". Nevertheless, *relative* U(1)-phases of a pair of wavefunctions ψ, ψ' *are* physical.

The above discussion does not address the issue that $\mathcal{A}$ can actually only be locally defined. Its role in the Schrödinger equation seems to be to ensure a gauge-independent notion of differentiating ψ, and this is a local requirement. In the next section, we will understand what $\mathcal{A}$ really are, in geometric terms.

3 Principal bundles and gauge theory

The mathematical way of summarizing the discussion in Section 2.6.2 is:

- Invariantly, wavefunctions of a (spinless) charged "particle" are sections of a complex Hermitian line bundle $\mathcal{L} \to M$, associated to a principal U(1)-bundle $P \to M$. A choice of (local) trivialization (a "gauge choice") turns such a section into a $\mathbb{C}$-valued function.

- There is a connection on P, which induces a Hermitian *connection* ∇ on $\mathcal{L}$ via the charge Q. This ∇ provides the parallel transport facilitating comparison of $\psi(x), \psi(x')$ at different x, x' (along a curve) in M. This gives a gauge-independent notion of differentiating a section ψ. When a local trivialization is chosen, ∇_j is represented as $\partial_j - i\frac{Q}{\hbar}\mathcal{A}_j$, where $\mathcal{A}_j$ are some local functions. The local 1-form $\mathcal{A}_j dq^j$ is called the *local gauge potential*, or *gauge field* in physics.

- The connection ∇ has a curvature 2-form, which is precisely the magnetic flux density $\mathcal{B}$, and can be computed as $d\mathcal{A}$ (with respect to any local trivialization).

Now, it turns out that in nature fundamental particles with charge, e.g. electrons, also have another non-trivial geometric property, called *spin*. The concept of spin is also gauge-theoretic, and is mathematically understood in the language of principal bundles.

Roughly speaking, for a Lie group G, a principal G-bundle is just a collection of G-torsors, smoothly parametrized by some base space M. And G-torsor refers to the underlying manifold of G equipped with the group-multiplication action of G on the right. For example, affine space is a principal V-bundle over a single point $M = \text{pt}$. We will be interested in G-torsors coming from the idea of *frames*, or simply, vector space bases.

Now, if V is an n-dimensional vector space over $\mathbb{K}$, then a basis is simply a linear isomorphism

$$\beta : \mathbb{K}^n \to V.$$

As we know from linear algebra, we can change a basis by applying an invertible $n \times n$ matrix $g \in \mathrm{GL}(n)$,

$$\beta \circ g : \mathbb{K}^n \to V.$$

So basis-change is nothing but the right action of $\mathrm{GL}(n)$ on the set of bases,

$$\{\text{Bases for } V\} \times \mathrm{GL}(n) \to \{\text{Bases for } V\}.$$

Starting from a reference basis, all other bases are obtained this way by a unique change-of-basis matrix. So $\{\text{Bases for } V\}$ is a $\mathrm{GL}(n)$-torsor.

After making a reference basis choice, any basis is uniquely labelled by $\mathrm{GL}(n)$. Similarly, if V is a complex inner product space, then $\{\text{Orthonormal bases}\}$ is a $\mathrm{U}(n)$-torsor. If V is a real inner product space, then $\{\text{(Oriented) Orthonormal bases}\}$ is a $\mathrm{(S)O}(n)$-torsor.

3.1 Tangent frame bundle as a principal bundle

In the special case of a tangent space T_xM, a basis is usually called a (tangent) *frame at* x. Just as the tangent bundle is the disjoint union of the T_xM, we can consider the *frame bundle*

$$\mathcal{F}M = \bigsqcup_{x\in M} \{\text{Frames at } x\},$$

which is a disjoint union of $\mathrm{GL}(n)$-torsors. Thus $\mathcal{F}M$ possesses a right action of $\mathrm{GL}(n)$, acting fibrewise.

The key thing to remember is that different fibres of $\mathcal{F}(TM)$ are *not* canonically identified with each other, just like tangent spaces T_xM are not identified for different x. Nevertheless, once a local trivialization $TM|_U \cong U \times \mathbb{R}^d$ is chosen, then

$$\mathcal{F}M|_U = \bigsqcup_{x\in U} \{\text{Frames at } x\} \cong U \times \mathrm{GL}(n).$$

Therefore, $\mathcal{F}(TM)$ is locally trivializable. The following definitions make this precise.

Definition 3.1. Let G be a Lie group[36]. The *trivial principal G-bundle* over a manifold U is the product manifold $U \times G$, equipped with the projection map

$$\pi_U : U \times G \to U, \qquad (x, g) \mapsto x,$$

and the fibrewise right G-action,

$$(x, g') \cdot g = (x, g'g), \qquad x \in U,\ g', g \in G.$$

Definition 3.2. A *principal G-bundle* over a manifold M is a manifold P with a smooth surjection $\pi : P \to M$ and a smooth right G-action,

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

such that

- The G-action restricts to fibres,

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

- Each $x \in X$ is contained in some open neighbourhood U such that there exists a diffeomorphism ("local trivialization")

$$\Phi : P|_U :\to U \times G,$$

satisfying

$$\begin{aligned} \pi_U \circ \Phi &= \pi \\ \Phi(p \cdot g) &= \Phi(p) \cdot g, \qquad p \in \pi^{-1}(U), g \in G. \end{aligned}$$

[36] A *Lie group* G is a manifold with a group structure such that

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

is smooth.

So the frame bundle $\pi : \mathcal{F}M \to M$ is a principal $\mathrm{GL}(d)$-bundle in the sense of Definition 3.2. If M is oriented and has a Riemannian metric, so that each tangent space T_xM is an inner product space, then we can restrict to the oriented orthonormal frames, obtaining the *oriented orthonormal frame bundle*

$$\pi : \mathcal{F}^{\mathrm{SO}}M \to M$$

as a principal $\mathrm{SO}(d)$-bundle.

3.2 Gauges and gauge transformations

3.2.1 Local gauge

Definition 3.3. 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 $\pi \circ s = \mathrm{id}_U$.

To get some intuition, consider the tangent bundle TS^2. Over a local patch of S^2, e.g., the upper hemisphere U_N, we can smoothly assign oriented orthonormal frames $\{\searrow, \nearrow\}_x$ to points $x \in U_N$. Such an assignment is precisely a local gauge s of $\mathcal{F}^{\mathrm{SO}}S^2$.

A local gauge s is basically the same thing as a local trivialization of $\mathcal{F}^{\mathrm{SO}}S^2$. This is because any frame β_x for T_xS^2 becomes labelled by the unique $\mathrm{SO}(2)$ rotation which turns the frame $s(x)$ into β_x. So we have an identification

$$\mathcal{F}^{\mathrm{SO}}|_{U_N} \cong U_N \times \mathrm{SO}(2).$$

The purpose of choosing a local gauge is to convert tangent vector fields X over U_N into $\mathbb{R}^2$-valued functions on U_N. Always remember: *the numerical representation of X is non-canonical*, i.e., *local gauge dependent*!

$\mathcal{F}S^2$ is an example of a non-trivializable principal bundle, i.e., no global gauge exists.

3.2.2 Local gauge transformations

Let $\pi : P \to M$ be a principal G-bundle. Let s, s' be two choices of local gauge over $U \subseteq M$. Then we can write

$$s'(x) = s(x) \cdot \lambda(x), \qquad x \in U, \tag{64}$$

for some uniquely determined map $\lambda : U \to G$. The G-valued function λ is said to implement a *local gauge transformation*.

3.2.3 Global gauge group

The *global* notion of a gauge transformation is:

Definition 3.4. A *gauge transformation* of a principal G bundle $\pi : P \to M$ is an automorphism of P inducing the identity map id_M on the base. The group of gauge transformations is called the *gauge group* of P.

The idea of gauge transformation is very foundational. A gauge transformation $F : P \to P$ acts fibrewise, $F : P_x \to P_x$, while respecting the G-torsor structure of P_x,

$$F(p \cdot g) = F(p) \cdot g, \qquad \forall g \in G, p \in P_x, x \in M. \tag{65}$$

In other words, it preserves "relativity of frames".

Remark 3.5. Locally, gauge transformations can be (non-canonically) identified with $\mathrm{Map}(U, G)$, as in (64). In general, because P may not even admit any reference global gauge, we can only characterize the gauge group as the following group of equivariant maps,

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

acting on P by multiplication on the right (exercise).

Example 3.6. If G is Abelian, then

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

Because $\sigma \in \mathrm{Map}(P, G)^G$ is constant on the fibres, it descends to a map on M. Thus, there is a canonical isomorphism

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

We will mostly be concerned with $G = \mathrm{U}(1)$. In this case, the naïve understanding of the gauge group as $\mathrm{Map}(M, \mathrm{U}(1))$ suffices.

3.3 Associated vector bundle

By now, you must have heard:

"A vector is not just a column of numbers!"

A vector $v \in V$ is obtained by pairing of basis $\beta : \mathbb{K}^n \xrightarrow{\cong} V$ with numerical components $\xi \in \mathbb{K}^n$, in many equivalent ways:

$$v = \beta(\xi) = (\beta \circ g)(g^{-1} \cdot \xi), \qquad \forall g \in \mathrm{GL}(n).$$

Therefore, V is the set of equivalence classes,

$$V = \Big(\{\text{Bases}\} \times \mathbb{K}^n\Big)\Big/_{(\beta,\xi) \sim (\beta \cdot g, g^{-1} \cdot \xi)}. \tag{66}$$

Eq. (66) generalizes to the construction of a vector bundle from a principal G-bundle.

Let $\pi : P \to M$ be a principal G-bundle. Let ρ be a linear representation of G on $\mathbb{K}^n$,

$$\begin{aligned}G \times \mathbb{K}^n &\to \mathbb{K}^n \\ (g, \xi) &\mapsto \rho(g) \cdot \xi =: g \cdot \xi.\end{aligned} \tag{67}$$

The set $P \times \mathbb{K}^n$ is equipped with the right G-action,

$$(p, \xi) \cdot g := (p \cdot g, g^{-1} \cdot \xi),$$

and we may pass to the set of equivalence classes,

$$P \times_\rho \mathbb{K}^n := (P \times \mathbb{K}^n)/_{(p,\xi)\sim(p\cdot g, g^{-1}\cdot\xi)}.$$

The projection map

$$\begin{aligned} \pi_\rho : P \times_\rho \mathbb{K}^n &\to M \\ [p,\xi] &\mapsto \pi(p) \end{aligned}$$

is well-defined, so we have constructed a "bundle of equivalence classes". In fact, each fibre can be identified with $\mathbb{K}^n$,

$$\pi_\rho^{-1}(x) = \big\{[p,\xi] \,:\, \xi \in \mathbb{K}^n\big\},$$

where p can be chosen to be any point in $P|_x$ (Exercise).

We call $\pi_\rho : P \times_\rho \mathbb{K}^n \to M$ the *vector bundle*[37] *associated to* P *via the representation* ρ.

Example 3.7. Let $P = \mathcal{F}M$ be the frame bundle of M, and let $\rho = \mathrm{id}$ be the defining representation of $\mathrm{GL}(d)$ on $\mathbb{R}^d$. Then $\mathcal{F}M \times_{\mathrm{id}} \mathbb{R}^d$ recovers the tangent bundle TM as a vector bundle.

Example 3.8. Let ρ_{triv} be the trivial representation of $\mathrm{GL}(d)$ on $\mathbb{K}^n$, meaning that every $g \in \mathrm{GL}(d)$ is represented as the identity $n \times n$ identity matrix . Then in $\mathcal{F}M \times_{\rho_{\mathrm{triv}}} \mathbb{K}^n$, each element is an equivalence class

$$[p,\xi] = [p\cdot g, \xi], \qquad \forall\, g \in \mathrm{GL}(d).$$

So we can unambiguously label the equivalence classes by $(\pi(p), \xi)$. Thus

$$\mathcal{F}M \times_{\rho_{\mathrm{triv}}} \mathbb{K}^n = M \times \mathbb{K}^n$$

is canonically a trivial vector bundle. Sections of this vector bundle are n-component *scalar* fields, or *functions* $M \to \mathbb{K}^n$ — the numerical components ξ have no dependence on the frame at all!

Remark 3.9. Often, we require $\mathbb{K}^n$ to have an inner product and ρ to be an orthogonal or unitary representation of G. In this case, the fibres become inner product spaces. For example, if M is an oriented Riemannian manifold, then $\mathcal{F}^{\mathrm{SO}}M$ is a principal $\mathrm{SO}(d)$-bundle. When we reconstruct TM as

$$\mathcal{F}^{\mathrm{SO}}M \times_{\mathrm{id}} \mathbb{R}^d,$$

the fibres have inner products coinciding with that coming from the Riemannian metric.

3.4 Vertical parallelization of principal bundles

On a Lie group G, we write $L_g : G \to G$ for the diffeomorphism of left-multiplication-by $g \in G$. Similarly, $R_g : G \to G$ denotes the right-multiplication-by-g map.

[37] To make this more precise, we should provide the manifold structure, local trivializations, etc.

3.4.1 Lie algebra of left-invariant vector fields on G

Definition 3.10. A vector field v on a Lie group G is *left-invariant* if

$$(L_g)_* v = v, \qquad \forall g \in G.$$

The *Lie algebra* $\mathfrak{g}$ of G is the vector space of left-invariant vector fields on G, equipped with the commutator *Lie bracket* $[\cdot, \cdot]$.

Certainly, we can evaluate a left-invariant vector field at the identity element $e \in G$, to get a map

$$\mathfrak{g} \to T_e G.$$

It is not too hard to show that this is a linear isomorphism.

Example 3.11. Write $\mathfrak{gl}(n, \mathbb{R})$ for the Lie algebra of $\mathrm{GL}(n, \mathbb{R})$. Regard $\mathrm{GL}(n, \mathbb{R})$ as an open subset of the vector space manifold $\mathrm{M}_n(\mathbb{R}) = \mathbb{R}^{n^2}$. Then

$$\mathfrak{gl}(n, \mathbb{R}) = T_e \mathrm{GL}(n) \cong T_e \mathrm{M}_n(\mathbb{R}) = \mathrm{M}_n(\mathbb{R}).$$

Chasing through the definitions, the Lie bracket on $\mathfrak{gl}(n, \mathbb{R})$ coincides with the matrix commutator on $\mathrm{M}_n(\mathbb{R})$. Similarly for $\mathfrak{gl}(n, \mathbb{C}) \cong \mathrm{M}_n(\mathbb{C})$.

In physics, Lie groups usually arise (non-canonically!) as subgroups $G \subseteq \mathrm{GL}(n)$. Then $\mathfrak{g} \subseteq \mathrm{M}_n(\mathbb{K})$. For example, $\mathfrak{so}(d)$ is the Lie algebra of real antisymmetric $d \times d$ matrices, while $\mathfrak{u}(d)$ is the Lie algebra of $d \times d$ complex skew-Hermitian matrices. In particular, we will use the identification

$$\mathfrak{u}(1) \cong i\mathbb{R}.$$

3.4.2 Exponential map and adjoint action

Starting at the identity element $e \in G$, the integral curves of a left-invariant $v \in \mathfrak{g}$ can be extended indefinitely (use the translation action). We use the suggestive exponential notation for these curves,

$$\begin{aligned} \exp(\,\cdot\, v) : \mathbb{R} &\to G \\ t &\mapsto \exp(tv), \end{aligned}$$

so that the integral curve condition is

$$\frac{d}{dt}\Big|_{t=0} \exp(tv) = v \in T_e G.$$

It is justified by the fact that these curves are also *group homomorphisms* (Exercise),

$$\exp((t + t')v) = \exp(tv) \exp(t'v), \qquad t, t' \in \mathbb{R}.$$

Given $v \in \mathfrak{g}$, we can use $g \in G$ to conjugate the curve $t \mapsto \exp(tv)$ to another curve

$$t \mapsto g \exp(tv) g^{-1}.$$

The resulting curve is itself exponentiated from some other $v' \in \mathfrak{g}$. Thus, we see that conjugation-by-g induces a map $\mathfrak{g} \to \mathfrak{g}$, called the *adjoint action*, and denoted

$$\begin{aligned} \mathrm{Ad}_g : \mathfrak{g} &\to \mathfrak{g} \\ \frac{d}{dt}\Big|_{t=0} \exp(tv) &\mapsto \frac{d}{dt}\Big|_{t=0} g \exp(tv) g^{-1}. \end{aligned}$$

3.4.3 Vertical parallelization

At any $g \in G$, the manifold underlying G "looks the same" as it does at $e \in G$. Indeed, the translated curves

$$t \mapsto g \exp(tv)$$

represent the elements of T_gG via their velocity vectors,

$$\frac{d}{dt}\Big|_{t=0} g \exp(tv) \in T_gG.$$

So all the tangent spaces T_gG are canonically identified with the tangent space at the identity, $T_eG = \mathfrak{g}$. Thus G is a *parallelizable* manifold,

$$TG = \bigsqcup_{g \in G} T_gG \cong G \times \mathfrak{g}.$$

Let us generalize this parallelization to a principal G-bundle $\pi : P \to M$, which is after all, a collection of G-torsors.

At any $p \in P$, there are "vertical" tangent directions represented by curves along the fibre $P|_{\pi(p)}$. Such vertical curves project down to constant curves on the base. So let us define

$$VP = \bigsqcup_{p \in P} V_pP := \ker(d\pi : TP \to TM)$$

to be the *vertical tangent bundle* of P.

It is crucial that VP is canonically parallelizable,

$$VP = \bigsqcup_{p \in P} V_pP \cong P \times \mathfrak{g}.$$

This is due to the fact that at each $p \in P$, there is a canonical collection of vertical curves labelled by $\mathfrak{g}$,

$$p \cdot \exp(tv), \qquad v \in \mathfrak{g}.$$

Thus, each $v \in \mathfrak{g}$ determines a vertical vector field $v^\sharp$ via the assignment

$$v^\sharp|_p := \frac{d}{dt}\Big|_{t=0} p \cdot \exp(tv).$$

We call $v^\sharp \in \mathfrak{X}(P)$ a *fundamental vector field*.

Remark 3.12. We stress that while $V_pP \subset T_pP$ is canonical, there is no canonical complementary "horizontal tangent space" H_pP such that $V_pP \oplus H_pP = T_pP$. One way to specify H_p is to choose a (local) gauge, i.e., a trivialization $P|_U \cong U \times G$, then at $p \sim (x, g)$, we would have

$$T_pP \cong T_{(x,g)}(M \times G) = T_xM \oplus T_gG \cong T_xM \oplus V_pP.$$

But this method is *gauge-dependent*!

3.5 Connections on a principal G-bundle

3.5.1 Moving horizontally between G-torsors

Definition 3.13. A connection on a principal G bundle P is a $\mathfrak{g}$-valued 1-form ω on P which satisfies

$$\omega(v^\sharp) = v, \qquad v \in \mathfrak{g}, \tag{68}$$

$$R_g^*\omega = \mathrm{Ad}_{g^{-1}} \circ \omega, \qquad g \in G. \tag{69}$$

This definition might appear mysterious, so let us spell out its geometric meaning. A 1-form is supposed to do something to tangent vectors. Eq. (68) is just the statement that a connection does the "tautological" thing to vertical vectors. The second condition (69) is automatic when applied to vertical vectors,

$$\begin{aligned} R_g^*\omega|_p(v^\sharp|_p) = \omega|_{p\cdot g}(dR_g(v^\sharp|_p)) &= \omega|_{p\cdot g}\Big(\frac{d}{dt}\Big|_{t=0} p\cdot \exp(tv)g\Big) \\ &= \omega|_{p\cdot g}\Big(\frac{d}{dt}\Big|_{t=0} p\cdot g\cdot g^{-1}\exp(tv)g\Big) \\ &= \omega|_{p\cdot g}\Big((\mathrm{Ad}_{g^{-1}}(v))^\sharp|_{p\cdot g}\Big) \\ &= \mathrm{Ad}_{g^{-1}}(v) && \text{(Eq. (68))} \\ &= \mathrm{Ad}_{g^{-1}}(\omega|_p(v^\sharp|_p)). && \text{(Eq. (68))} \end{aligned}$$

Therefore, the only meaningful data that a connection ω contains is what it does in the *non-vertical* directions[38].

Indeed, a connection supplies the previously missing notion of horizontal tangent spaces, via

$$H_pP := \ker \omega|_p.$$

Furthermore, due to Condition (69), this horizontal subspace assignment is G-invariant,

$$H_{p\cdot g}P = (R_g)_* H_pP.$$

Thus, *a connection is a gauge-independent way to specify "horizontal directions" in P.*

To summarize, a connection ω allows us to write

$$T_pP = \underbrace{H_pP}_{\ker\,\omega|_p} \oplus \underbrace{V_pP}_{\cong\mathfrak{g}}, \tag{70}$$

and measures the vertical component of elements of T_pP.

3.5.2 Curvature

Definition 3.14. 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

$$\Omega(u,v) = d\omega(u^{\mathrm{hor}}, v^{\mathrm{hor}}), \qquad u,v \in \mathfrak{X}(P). \tag{71}$$

[38] In the case $P = G$ which has no non-vertical directions, there is only one connection, called the *Maurer–Cartan form*.

If you know a bit more general differential geometry,

$$d\omega(u,v) = u(\omega(v)) - v(\omega(u)) - \omega([u,v]).$$

Since horizontal vectors are annihilated by ω by definition, an equivalent formula for the curvature is

$$\Omega(u,v) = -\omega[u^{\text{hor}}, v^{\text{hor}}], \tag{72}$$

where $(\cdot)^{\text{hor}}$ means horizontal component (Eq. (70)).

Intuitively: Inside P, imagine moving an small distance horizontally along u^{hor}, then horizontally along v^{hor}. Now do this in the opposite order. Generally, there can be a small vertical mismatch in the result, and this is what (72) measures.

Theorem 3.15 (Cartan structure equation)**.** *The curvature of a connection* $\omega \in \Omega^1(P,\mathfrak{g})$ *is*

$$\Omega = d\omega + \frac{1}{2}[\omega,\omega], \tag{73}$$

where $[\cdot,\cdot]$ *means Lie bracket and wedge product.*

The proof proceeds by direct computation on vertical/horizontal vectors separately (it can be found in standard textbooks).

3.5.3 Local description of connection: gauge potentials

For concrete computations, one describes connections locally, i.e., with respect to some local trivialization/gauge.

Definition 3.16. Let ω be a connection on a principal G-bundle $\pi : P \to M$. Let $s : U \to P$ be a local gauge over an open subset $U \subset M$. Then

- $s^*\omega \in \Omega^1(U,\mathfrak{g})$ is called the *local gauge potential*.
- $s^*\Omega \in \Omega^2(U,\mathfrak{g})$ is called the *local field strength*.

Intuitively, think of a local gauge as a candidate way to specify "horizontal" (Remark 3.12). If this happens to coincide with what ω actually specifies, then the gauge potential $s^*\omega$ will vanish. Otherwise,

$$s^*\omega(u) = \omega(ds(u)) = \omega\big((ds(u))^{\text{vert}}\big), \qquad u \in T_xM,$$

where $(\cdot)^{\text{vert}}$ means vertical component (Eq. (70)).

3.5.4 Local versus global description of connections

A gauge potential is a *local, gauge-dependent* description of a connection. If there is no global gauge available, then we have to use a collection of local sections $s_\alpha : U_\alpha \to P|_{U_\alpha}$ to obtain a collection of local gauge potentials $s_\alpha^*\omega \in \Omega^1(U_\alpha,\mathfrak{g})$. Of course, these gauge potentials have to be consistent with the local gauge transformations $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ relating the various choices of local sections. Explicitly, the consistency condition is

$$s_\beta^*\omega = \text{Ad}_{g_{\alpha\beta}^{-1}} \circ s_\alpha^*\omega + g_{\alpha\beta}^*\Theta, \tag{74}$$

where Θ is the Maurer–Cartan form of G. (The calculation can be found in textbook references.) When G is abelian, this simplifies to

$$s_\beta^*\omega = s_\alpha^*\omega + g_{\alpha\beta}^*\Theta. \tag{75}$$

In physics, it is more usual to think of a connection as a collection $\mathcal{A}_\alpha \in \Omega^1(U_\alpha, \mathfrak{g})$, such that

$$\mathcal{A}_\beta = \mathrm{Ad}_{g_{\alpha\beta}^{-1}} \circ \mathcal{A}_\alpha + g_{\alpha\beta}^*\Theta \quad \text{over } U_\alpha \cap U_\beta, \tag{76}$$

wherever some "local gauge transformation $g_{\alpha\beta}$ is applied". An instance of $\mathcal{A}_\alpha$ is called a *gauge field*, and (76) is understood as the *transformation rule* defining what gauge fields "are", in contrast to "ordinary fields" like tangent vector fields, etc.

Generally, the local field strength $s^*\Omega$ is similarly gauge-dependent. Nevertheless, due to the Cartan structure equation (following (73)),

$$s^*\Omega = d(s^*\omega) + \frac{1}{2}[s^*\omega, s^*\omega], \tag{77}$$

if G is abelian, it follows from (76) and $d\Theta = 0$ that

$$s_\beta^*\Omega = s_\alpha^*\Omega$$

is actually gauge-independent. So the field strength is actually a globally well-defined $\mathfrak{g}$-valued 2-form on M. In physics notation, one writes this 2-form as

$$\mathcal{F} = s^*\Omega = ds^*\omega = d\mathcal{A},$$

where the formula on the right side holds *locally* (because $\mathcal{A}$ is locally defined).

3.6 Connections and gauge-covariant differentiation

We shall now explain how a connection allows us to "parallel transport frames", and consequently, gauge-independent differentiation for *any* associated vector bundle.

3.6.1 Parallel transport of frames

Definition 3.17. Let $\pi : P \to M$ be a principal G-bundle with connection ω. Let $\gamma : I \to M$ be a curve in M. A curve $\tilde{\gamma} : I \to P$ is a *horizontal lift* of γ if $\pi \circ \tilde{\gamma} = \gamma$ and its velocity vectors are horizontal for all $t \in I$.

Theorem 3.18. *Let $\pi : P \to M$ be a principal G-bundle with connection, and let $\gamma : I \to M$. For each $p \in P|_{\gamma(0)}$, there exists a unique horizontal lift $\tilde{\gamma}_p$ starting at p.*

The proof is omitted: It boils down to ODEs when the horizontal lift condition is expressed in terms of a local trivialization.

Definition 3.19. Let $\pi : P \to M$ be a principal G-bundle with connection ω, and let $\gamma : [0,1] \to M$ be a curve. *Parallel transport along* γ, with respect to ω, is the map

$$\begin{aligned} \tau_\gamma^\omega : P|_{\gamma(0)} &\to P|_{\gamma(1)} \\ p &\mapsto \tilde{\gamma}_p(1), \end{aligned}$$

where $\tilde{\gamma}_p$ is the unique horizontal lift of γ starting at $p \in P|_{\gamma(0)}$.

It is crucial, and follows readily from (69), that parallel transport is G-equivariant,

$$\tau_\gamma^\omega(p \cdot g) = (\tau_\gamma^\omega(p)) \cdot g, \qquad p \in P_{\gamma(0)},\ g \in G. \tag{78}$$

In words: the relativity of frames is respected.

3.6.2 Covariant derivatives on associated vector bundles

Recall our perspective that vector bundles are obtained by attaching numerical components $\mathbb{K}^n$ to a principal G-bundle of "frames", via the associated bundle construction.

So let $E = P \times_\rho \mathbb{K}^n$ by any associated vector bundle. Given a connection ω on P, each curve $\gamma : [0,1] \to M$ determines a linear isomorphism

$$\begin{aligned} \tau_\gamma^{E,\omega} : E|_{\gamma(0)} &\to E|_{\gamma(1)} \\ [p, \xi] &\mapsto [\tau_\gamma^\omega(p), \xi]. \end{aligned}$$

This is well-defined:

$$\tau_\gamma^{E,\omega}[p \cdot g, g^{-1} \cdot \xi] = [\tau_\gamma^\omega(p \cdot g), g^{-1} \cdot \xi] \overset{(78)}{=} [\tau_\gamma^\omega(p) \cdot g, g^{-1} \cdot \xi] = [\tau_\gamma^\omega(p), \xi] = \tau_\gamma^{E,\omega}[p, \xi].$$

Definition 3.20. With the above notation, the *induced covariant derivative of a section* $\psi \in \Gamma(E)$ *along* $X \in \mathfrak{X}(M)$ is the section $\nabla_X^\omega \psi$ given by

$$\left(\nabla_X^\omega \psi\right)(x) := \frac{d}{dt}\Big|_{t=0} \Big(\tau^{E,\omega}_{\gamma|_{[0,t]}})^{-1}(\psi(\gamma(t))\Big) \ \in E|_x. \tag{79}$$

Above, γ can be any curve[39] representing the tangent vector $X|_x$.

What is happening is this: we parallel transport $\psi(\gamma(t)) \in E|_{\gamma(t)}$ to the initial fibre $E|_{\gamma(0)} = E|_x$, whence they become comparable and the usual t-derivative makes sense.

As an exercise, one can check the following properties of ∇^ω: for all $X, Y \in \mathfrak{X}(M)$, $f \in C^\infty(M)$, and $\psi \in \Gamma(E)$,

- $\nabla_{X+fY}\psi = \nabla_X \psi + (\nabla_Y \psi) \cdot f$.
- The Leibniz property holds:

$$\nabla_X(\psi \cdot f) = \psi \cdot X(f) + (\nabla_X \psi) \cdot f.$$

3.6.3 Local formula and connection coefficients

Notice that (79) is manifestly gauge-independent!

With respect to a local gauge $s : U \to P$, we would write

$$\psi(x) = [s(x), \xi(x)]$$

with $\xi : U \to \mathbb{K}^n$ being the numerical function representing ψ. Then the local formula for ∇^ω is (exercise)

$$\nabla_X^\omega \psi = [s, d\xi(X) + d\rho_e(s^*\omega(X)) \cdot \xi], \qquad X \in \mathfrak{X}(M),\ \psi \in \Gamma(E). \tag{80}$$

The first term $d\xi(v)$ is the usual directional derivative of ξ along X. The extra term is the "infinitesimal vertical correction".

Let us further choose local coordinates over $U \subseteq M$, so we have partial derivatives $\partial_j, j = 1, \ldots, d$. Then we may expand the local gauge potential as

$$s^*\omega = \mathcal{A}_j dq^j, \qquad \mathcal{A}_j = s^*\omega(\partial_j) \in C^\infty(U, \mathfrak{g}).$$

[39] This is not immediately obvious, but can be checked.

We also have the following basis of sections for $E|_U$,

$$e_a := [s, \eta_a] \equiv [s, (0, \ldots, 0, \underbrace{1}_{a\text{-th}}, 0, \ldots, 0)].$$

Then

$$\nabla^\omega_j e_a := \nabla^\omega_{\partial_j} e_a \overset{(80)}{=} [s, \mathcal{A}_j \cdot \eta_a].$$

Above, $\mathcal{A}_j$ means $(d\rho_e)(\mathcal{A}_j)$, which is an $n \times n$ matrix,

$$(\mathcal{A}_j)^b{}_a \equiv \mathcal{A}^b{}_{ja}.$$

The entries are called *connection coefficients*. In this notation,

$$\nabla^\omega_j e_a = [s, \mathcal{A}^b{}_{ja}\eta_b] = [s, \eta_b]\mathcal{A}^b{}_{ja} = e_b \mathcal{A}^b{}_{ja}.$$

For a general section of E, locally expanded as $\psi = e_a \psi^a$, the Leibniz property of ∇^ω leads to

$$\begin{aligned} \nabla^\omega_j \psi = \nabla^\omega_j (e_a \psi^a) &= e_a(\partial_j \psi^a) + (\nabla^\omega_j e_a)\psi^a \\ &= e_a(\partial_j \psi^a + \mathcal{A}^a{}_{jc}\psi^c). \end{aligned}$$

In physics terminology, the above local formula is the statement that the j-th covariant partial derivative is

$$\text{“ } \nabla_j \ = \ \partial_j + \mathcal{A}_j \text{ ”} \quad \text{acting on } \mathbb{K}^n\text{-valued functions.}$$

3.7 Example: Quantum mechanics of charged particle

For an integer N, the *charge-N* unitary representation of $G = \mathrm{U}(1)$ is

$$\begin{aligned} \rho^{(N)} : G = \mathrm{U}(1) &\to \mathrm{U}(1) \\ e^{i\theta} &\mapsto e^{iN\theta}. \end{aligned}$$

Also,

$$d\rho^{(N)}_e : \mathfrak{u}(1) = i\mathbb{R} \to \mathrm{End}(\mathbb{C}), \qquad i\lambda \mapsto iN\lambda.$$

The local gauge potential $s^*\omega$ has the form

$$d\rho^{(N)}_e(s^*\omega) = d\rho^{(N)}_e(\mathcal{A}_j dq^j) = N\mathcal{A}_j\, dq^j$$

with $\mathcal{A}_j$ being $\mathfrak{u}(1) = i\mathbb{R}$-valued. For physics purposes, it is customary to replace

$$\mathcal{A}_j \rightsquigarrow -i\frac{e}{\hbar}\mathcal{A}_j$$

where e is the electron charge and $\hbar$ is Planck's constant. This ensures that $d\mathcal{A}_j$ is real-valued and has the correct physical dimensions for a magnetic flux density. With these conventions, the j-th covariant partial derivative is represented in the gauge s by the formula

$$\nabla_j = \partial_j - i\frac{Ne}{\hbar}\mathcal{A}_j.$$

With $Q = Ne$ the (quantized!) electric charge, this is precisely what we found in (63).

Furthermore, the abelian transformation formula, (75), says that under a gauge transformation $g = e^{-ie\Lambda/\hbar}$ (which would modify the representation of ψ by g^{-1}), we should replace $s^*\omega$ by

$$s^*\omega + g^{-1}dg = s^*\omega + i\frac{e}{\hbar}d\Lambda.$$

In the physics convention, this amounts to modifying the magnetic potential 1-form to

$$\mathcal{A} \rightsquigarrow \mathcal{A} + d\Lambda.$$

To summarize, *quantum mechanics* naturally provides a consistent home for $\mathcal{A}$:

- Local magnetic vector potentials $\mathcal{A}$ are local gauge potentials for a connection on a principal U(1) bundle P, whose curvature $d\mathcal{A}$ is a (gauge-independent) closed 2-form giving the magnetic flux density $\mathcal{B}$.
- For *integer* charge $Q = Ne$, the "wavefunction" is a section of the associated Hermitian line bundle $\mathcal{L} = P \times_{\rho^{(N)}} \mathbb{C}$. Only in the chargeless $Q = 0$ case, can a wavefunction be unambiguously regarded as a $\mathbb{C}$-valued *function*. (But, can you name any chargeless, spinless quantum particle?)
- The (gauge-invariant) time-independent Schrödinger equation for $\psi \in \Gamma(\mathcal{L})$ is

  $$E\psi = \Big(-\frac{\hbar^2}{2m}\Delta^{\mathcal{L}} + V\Big)\psi,$$

  where $\Delta^{\mathcal{L}}$ is the connection Laplacian associated to the gauge-covariant derivative ∇ on $\mathcal{L}$.

3.8 Aharonov–Bohm effect

Sometimes, one finds the statement that classical electromagnetism is a U(1) gauge theory. Actually, this is not quite true. The transformation property of the potentials $\mathcal{A}$ only narrows down the Lie algebra to the 1-dimensional one (with trivial Lie brackets). In principle, the Lie group could be $(\mathbb{R}_{>0}, \times)$ rather than the compact Lie group U(1). Indeed, the former choice is precisely the group of scalings originally considered by Weyl. The problem with U(1) is that there had been no reason to consider complex vector bundles, until *quantum* mechanics arose. One of the constraints imposed by having a principal U(1)-bundle, is that the charge has to be quantized, and so far, this is consistent with experiment[40].

Despite the geometric unity of electromagnetism and quantum mechanics via gauge theory, one might object that the gauge-theoretic interpretation of $\mathcal{A}$ and ψ is mere formalism, with no experimental consequence. Actually, a connection ω has some further gauge-invariants besides the curvature. The most dramatic example of this is the *holonomy* along a closed loop ℓ in M,

$$\exp\Big(-i\frac{e}{\hbar}\oint_\ell \mathcal{A}\Big) \in \mathrm{U}(1). \tag{81}$$

[40]Caveat: Quarks have 1/3-quantized charge, but are "confined". More interestingly, there are apparently more generally fractionally-quantized charges appearing in 2D materials, and this is a subject of intense modern research.

We can try to detect this holonomy.

Consider the situation where ℓ encloses some thin tube in $M = \mathbb{R}^3$ in which $\mathcal{B}$ is concentrated. Thus $\mathcal{B}$ vanishes on a (non-simply-connected) neighbourhood U of ℓ. So on U, any admissible $\mathcal{A}$ must satisfy $d\mathcal{A} = \mathcal{B}$. Classically, the *local* condition $d\mathcal{A} = \mathcal{B}$ is the only constraint, and we are allowed to choose $\mathcal{A} = 0$ on U — this predicts that the integral (81) vanishes.

However, the *gauge-theoretic* requirement that $\mathcal{A}$ is a local gauge potential for some *global* U(1) connection ω on $P \to M$ constrains which $\mathcal{A}$ can be used in U, namely, the gauge-transformation rule (75). Indeed, let D be a disc with $\partial D = \ell$. There is a $\mathcal{A}'$ defined over D satisfying $d\mathcal{A}' = \mathcal{B}$. Stokes' theorem forces

$$\oint_\ell \mathcal{A}' = \int_D \mathcal{B} \neq 0,$$

so the holonomy (81) does not generally vanish. Any admissible $\mathcal{A}$ on $U \subset D$ must be related to $\mathcal{A}'$ by a gauge transformation over U. If you know some topology, verify that the exponentiated loop integral (81) is unchanged under such a gauge transformation (thus it is gauge-independent).

Quantum mechanics predicts that a charged particle can, in principle, detect this holonomy, basically by picking up the extra phase factor (81) after propagating around ℓ. In 1959 Aharanov–Bohm proposed that this extra (relative) phase can be measured by wavefunction interference experiments, and in the 1960s, this was confirmed.

3.9 Further geometry–physics

The Aharonov–Bohm effect shows indisputably that quantum mechanical wavefunctions should be understood as sections of some vector bundle, *with no absolute numerical meaning to the pointwise values*. This is analogous to:

- Classical tangent vector fields requiring gauges/tangent frames to be numerically described.
- No preferred origin in affine space (or, more generally, G-torsors).
- A vector *not* being a collection of numerical scalars.

It is similar, but not the same thing as

- Points, paths, etc., in a manifold being independent of coordinates used to label them.

Once ψ is understood as a bundle section, it becomes mandatory to provide a connection/parallel transport, to meaningfully differentiate it. This "coupling" of connection to sections models interactions of "quantum particles" mediated by electromagnetic (or more general force) fields. In fact, fundamental "quantum particles" and interactions now understood this way, with *non-Abelian* G-connections — this is the *Standard Model*.

Although we have only studied the Abelian case concretely, the gauge principle is already apparent: only quantities that depend on equivalence classes modulo change-of-gauge are physically measurable[41]. Examples are relative quantum mechanical phases, holonomies, and energy spectra of Hamiltonians.

[41] As numerical multiples of some comparable quantity (i.e., physical units like metres, Joule, etc.)!

So a "quantum mechanical particle" is neither a classical particle (with an objective path), nor a classical wave (with objective values at each time and location). In quantum mechanics, it merely has *observables*, modelled as self-adjoint operators $\mathcal{O}$ on a Hilbert space of wavefunctions ψ. The spectral set of $\mathcal{O}$ comprises the possible measured outcomes. In the normalized state represented by the wavefunction ψ, the probability of observing outcome $\lambda \in \mathrm{Spec}(\mathcal{O})$ is the gauge-invariant number

$$\langle \psi | P_\lambda \psi \rangle \in [0,1],$$

where P_λ is the eigenprojection for λ. General observables are *non-commuting*, e.g., we cannot simultaneously talk about the position and the momentum of a quantum mechanical state, since its wavefunction cannot simultaneously be an eigenfunction of the position and momentum operators. In other words, we precisely abandon the classical paradigm of objective t-parametrized paths in phase space!

3.9.1 Spin geometry and gauge

One of the deepest contributions of quantum theory to geometry is the spin refinement of Riemannian geometry.

From the physics perspective, spin originally referred to a certain mysterious "angular momentum" of electrons. In Euclidean $\mathbb{R}^d$, angular momentum (about the origin) refers to the functions

$$L_{jk} = q^j p_k - q^k p_j, \qquad 1 \leq j < k \leq d.$$

where q^i are Cartesian position coordinates and p_j the corresponding momentum coordinates. They are associated to the symmetry under $\mathrm{SO}(d)$ rotations[42]. A rotating charge Q should have a magnetic moment

$$\mu = \gamma L, \qquad \gamma \overset{\text{classical}}{=} \frac{Q}{2m},$$

where the ratio γ is the *gyromagnetic ratio*. In the presence of a magnetic flux density $\mathcal{B}$, the potential energy is modified by

$$-\sum_{j<k} \mu_{jk} B_{jk}.$$

Since 1915, physicists have experimentally determined that the free charge-e electron has an "intrinsic" contribution to its angular momentum, for which γ is $\frac{e}{m}$ instead of $\frac{e}{2m}$. This discrepancy is inconsistent with a classical picture of the electron orbiting in a small loop. One says that the electron possesses *spin* angular momentum "about itself", in addition to the usual *orbital* angular momentum about some other point (e.g. a nucleus). Mathematically, instead of the classical angular momenta generating $\mathrm{SO}(d)$ rotations around orbits, one considers the *spin group* extension,

$$1 \to \{\pm 1\} \to \mathrm{Spin}(d) \xrightarrow{\chi} \mathrm{SO}(d) \to 1,$$

with corresponding Lie algebra isomorphism

$$d\chi_e : \mathfrak{spin}(d) \to \mathfrak{so}(d).$$

[42]They constitute the $\frac{d(d-1)}{2}$ component functions of the moment map $\mu : T^*\mathbb{R}^d \to \mathfrak{so}(d)^*$.

Spin angular momentum operators come from $\mathfrak{spin}(d)$, and are infinitesimal "internal" rotations of a abstract $\mathrm{Spin}(d)$-torsor of "spin frames". This lifts the usual $\mathrm{SO}(d)$-torsor of oriented orthonormal tangent frames. But spin angular momentum generates only "half as much" rotation, basically due to the $2:1$ mapping χ.

Globally, *spin geometry* is a refinement of Riemannian geometry. Namely, instead of the classical frame bundle $\mathcal{F}^{\mathrm{SO}}M$, one pass to a principal $\mathrm{Spin}(d)$-bundle $\mathcal{F}^{\mathrm{Spin}}M$ which χ-equivariantly lifts of $\mathcal{F}^{\mathrm{SO}}M$; this is called a *spin structure*. Accordingly, the $\mathfrak{so}(d)$-valued Levi–Civita connection is lifted to the $\mathfrak{spin}(d)$-valued *spin connection* ω^{Spin} on $\mathcal{F}^{\mathrm{Spin}}M$. From here, representation theory of $\mathrm{Spin}(d)$ takes over. The simplest irreducible representation is the spin representation of $\mathrm{Spin}(d)$ on $\mathbb{C}^{2^{\lfloor \frac{d}{2} \rfloor}}$, and leads to the associated *spinor bundle*

$$\mathcal{S} = (\mathcal{F}^{\mathrm{Spin}}M) \times_\rho \mathbb{C}^{2^{\lfloor \frac{d}{2} \rfloor}}.$$

Fermionic quantum particles, e.g. electrons, have wavefunctions which are sections of $\mathcal{S}$. Actually, since spin is deduced from electrically charged particles[43] coupled to magnetic fields, the appropriate bundle for charged fermions is the *twisted spinor bundle*,

$$\mathcal{S} \otimes \mathcal{L},$$

whose covariant derivative has an extra electromagnetic contribution.

For example, in $d = 3$ spatial dimensions, fermions have a $\mathbb{C}^2$ spinor degree of freedom at each basepoint. The spin angular momentum operator (about any axis in 3D) acts on $\mathbb{C}^2$ with eigenvalues $\pm\frac{\hbar}{2}$, and contributes an extra 2×2 matrix term to the Hamiltonian (energy) operator for charged fermions in the presence of a magnetic field.

Remarkably, there is a canonical *first-order* gauge-covariant (with respect to $\mathrm{Spin}(d)$ and $\mathrm{U}(1)$) differential operator on sections of $\mathcal{S} \otimes \mathcal{L}$, called the (twisted) *Dirac operator* $\not{D}$. It is popularly described as the "square root of Laplacian". In relativistic physics, it is mandatory to use the $\not{D}$ rather than the Laplacian to achieve a Schrödinger equation compatible with special relativistic spacetime.

First constructed on $\mathbb{R}^4$ by Dirac in 1928, the twisted Dirac operator provides the generating examples for the profound Atiyah–Singer index theory, which counts the zero-eigenvalues of $\not{D}$ through topological invariants of the manifold M. This, in turn, provided a powerful new technique to address various Riemannian geometry questions, e.g., about positive scalar curvature metrics.

[43] The apparent exception of neutrinos is a topic of current research.