2 Gauge and charts

2.1 Vector spaces and gauge

Let V be a vector space of dimension d. A *basis* is a linear isomorphism $\beta : \mathbb{R}^d \to V$. The map β^{-1} provides linear *coordinates* for elements of V,

$$\begin{aligned} V &\to \mathbb{R}^d \\ v &\mapsto \mathbf{v} = \beta^{-1}(v). \end{aligned}$$

Coordinates are always available, but not intrinsic. If β, β' are two choices of bases, and $\mathbf{v}, \mathbf{v}'$ are the respective coordinates of a vector v, then

$$\mathbf{v}' = (\beta')^{-1} \circ \beta(\mathbf{v})$$

where $(\beta')^{-1} \circ \beta \in \mathrm{GL}(d)$ is the invertible matrix implementing the change of coordinates.

In geometry and physics, most of the significant vector spaces do not come with bases[2].

- Elementary plane geometry involves displacement vectors. Starting at some point, the space of displacement vectors is a two-dimensional vector space (we can scale and add/subtract them).
- On a smooth manifold M, the tangent space at a point $p \in M$ is the vector space of *velocities* of curves passing through p.
- The space of possible "values" of a wavefunction at a point.

Consider the wind velocity "function" ψ with domain being the Earth. At each point $p \in$ Earth, $\psi(p)$ is some element of a two-dimensional vector space. But it is wrong to say that ψ is a function Earth $\to \mathbb{R}^2$. There are two problems:

1. At $p \in$ Earth, a basis (e.g., {North, West}) must be chosen in order to turn $\psi(p)$ into a pair of real numbers.
2. As we vary p, there is no globally consistent choice of {North,South}.

[2]Whereas in applications of linear algebra in other disciplines, there are often implicitly preferred bases.

It is true, however, that the wind *speed* function is a genuine function $|\psi| : \text{Earth} \to \mathbb{R}_{\geq 0}$.

Quantum wavefunctions are not functions $\psi : \text{Laboratory} \to \mathbb{C}$ for exactly the same two reasons why the wind velocity is not a function $\text{Earth} \to \mathbb{R}^2$.

1. At each laboratory point, $\psi(p)$ is an element of a 1-dimensional complex vector space (with inner product). Only *after* choosing a basis at p, does $\psi(p)$ becomes a complex number. Such a choice is called a *local gauge* at p.

   In the wind velocity example, we could argue that the Earth's rotation (and a magnetic compass and ruler) distinguishes a {North, West} basis. Then we can measure the wind speed in these directions, and reconstruct the velocities. However, this convention will fail at the two poles.

   There is no physical analogue for quantum wavefunctions — choices of orthonormal bases (at each point) are *always* fictitious, and there is no "reference orthonormal frame" with respect to which the "complex number value" of $\psi(p)$ can be measured! The absolute-value-squared $|\psi(p)|^2$, on the other hand, does not depend on the choice of basis, so it can be measured (at least indirectly, as a probability density[3]).

2. Since do not have a way to choose local gauges, we cannot compare the values of $\psi(p)$ and $\psi(p')$ at distinct points, much less *differentiate* the wavefunction ψ.

   Why, then, does the Schrödinger differential equation make physical sense? The answer can be found in the concept of "gauge-covariant differentiation", which lies at the heart of modern differential geometry and gauge theory. The key is that an extra ingredient, called a "connection" is needed. In quantum mechanics, this connection comes directly from electromagnetism. Indeed, this kind of idea replaces the Newtonian force concept in quantum mechanics! In the basic Schrödinger equation (1.1), this connection data happens to be "trivial" because no magnetic fields are present.

[3]The probabilistic nature of "measuring" a wavefunction also leads to difficult conceptual questions such as: Must two experimentalists agree on the same objective wavefunction for describing the same quantum system, or is the appropriate wavefunction subject to one's knowledge of the system? You may return to this kind of question after the course...

As we will learn, in the presence of magnetic fields, (1.1) has to be modified in a subtle way to be consistent with the principle that there are no preferred local gauges for the wavefunction.

2.2 Torsors

Notice that

$$v = \beta(\beta^{-1}(v)) = (\beta \circ T)(T^{-1}(\beta^{-1}(v)), \qquad \forall\, T \in \mathrm{GL}(d).$$

Think of $\beta \circ T$ as a right action of $T \in \mathrm{GL}(d)$ on the basis β. Under this action, the coordinates are transformed by left multiplication by T^{-1}.

Any basis β' is obtained from a reference choice β by the above action of some (unique) $T \in \mathrm{GL}(d)$. So, fixing β, we obtain a bijection between the set of bases, and $\mathrm{GL}(d)$. However, there is no canonical reference β, so this bijection is not canonical. The more precise statement is: the set of bases for a d-dimensional vector space is a $\mathrm{GL}(d)$-torsor.

Definition 1. Let G be a group. A *G-torsor* is a set X equipped with a free and transitive right G-action,

$$\begin{aligned} X \times G &\to X \\ (x, g) &\mapsto x \cdot g \\ (x, gg') &\mapsto (x \cdot g) \cdot g'. \end{aligned}$$

Free means that every non-identity group element $g \neq 1$ acts on X without fixing any point of X. *Transitive* means that every pair $x, x' \in X$ is related by some g,

$$\exists g \in G \ \text{ such that } \ x \cdot g = x'. \tag{2.1}$$

Free and transitive implies that the g satisfying (2.1) is unique. In particular, if we choose any reference element $x \in X$, then every other $x' \in X$ is labelled by the unique g satisfying (2.1), giving a bijection $X \xrightarrow{\sim} G$.

Example 2.1. The underlying set of a group G is a G-torsor. Here, when we pass from the group to the underlying set, we "forget" the special element 1 (otherwise we would have a pointed set). A torsor is sometimes informally said to be a group which has "forgotten its identity element".

2.3 Coordinates and manifolds

Start from the vector space $\mathbb{R}^d$ with its standard topology.

- Forget the basis. We are left with a d-dimensional vector space V with topology.
- Forget scalar multiplication. We are left with the abelian group $(V, +)$ with translation-invariant topology.
- Forget 0. We are left with a $(V, +)$-torsor A with translation-invariant topology, called *affine space*.
- Forget the $(V, +)$-action. We are left with a topological space $\mathcal{A}$.

We cannot do calculus on $\mathcal{A}$, since there is no notion of $\frac{\partial}{\partial x}$, or "infinitesimally small change in $\mathcal{A}$". Nevertheless, if we pick a point in $\mathcal{A}$, "remember" the $(V, +)$-action and scaling, and pick a basis, we get a homeomorphism

$$\varphi : \mathcal{A} \to A \to V \to \mathbb{R}^d,$$

called a *rectilinear global coordinate chart*. The adjective "rectilinear" refers to the fact that φ maps straight lines in A (what does this mean?) to straight lines in $\mathbb{R}^d$.

Using φ, we can regard a function $f : \mathcal{A} \to W$ (where W is a finite-dimensional vector space) as the function $f \circ \varphi^{-1} : \mathbb{R}^d \to W$, and declare that f is smooth if $f \circ \varphi^{-1}$ is smooth in the usual sense of elementary calculus.

Natural phenomena taking place in $\mathcal{A}$ cannot possibly depend on how we choose to label points in $\mathcal{A}$ by numbers. So we should be uncomfortable at picking out a preferred rectilinear chart φ. Nevertheless, if φ' is another such chart, then the change-of-coordinates map

$$\varphi' \circ \varphi^{-1} : \mathbb{R}^d \to \mathbb{R}^d \tag{2.2}$$

is smooth (why?). Actually, if we are forgetting the $(V, +)$-action, there is no particular reason to only use rectilinear charts[4]. Any homeomorphism $\varphi' : \mathcal{A} \to \mathbb{R}^d$ should be allowed as well, the only constraint being that the

[4]In older literature, one might use the term "curvilinear" to contrast with "rectilinear". Of course this distinction only makes sense if the notion of straight lines in $\mathcal{A}$ is remembered.

change-of-coordinates maps (2.2) are smooth, so that smoothness of functions $f : \mathcal{A} \to W$ becomes an intrinsic, chart-independent notion.

Actually, smoothness of functions $\mathbb{R}^d \to W$ is a *local* condition, and already makes sense when the domain is some open subset of $\mathbb{R}^d$. So we might as well admit local homeomorphisms $\varphi : U \to \mathbb{R}^d$ (with $U \subseteq \mathcal{A}$ open) as *local coordinate charts*. Formally, we allow the use of the *maximal atlas* of local coordinate charts $\varphi : U \to \mathbb{R}^d$ compatible with a fiducial global chart. *Then we forget the fiducial choice.*

Abstracting this example, a general smooth d-dimensional manifold $\mathcal{M}$ is defined as a (second-countable, Hausdorff) topological space equipped with a maximal atlas of local coordinate charts $\varphi : U \to \mathbb{R}^d$ (where $U \subseteq \mathcal{M}$ is open) which are mutually compatible,

$$\varphi' \circ \varphi^{-1} : \varphi(U \cap U') \to \varphi'(U \cap U') \qquad \text{smooth},$$

and which suffices to cover all of $\mathcal{M}$, i.e., $\cup_{U \text{ in atlas}} U = \mathcal{M}$.

In practice, we might specify some convenient fiducial choice of local charts, but we always pass to the maximal compatible atlas, then forget the initial choice. For example, we may use two fiducial local charts to describe the manifold structure on the topological circle S^1. But subsequently, nothing prevents us from using any other compatible local chart.

2.4 Tangent vectors

A function $f : \mathcal{M} \to \mathbb{R}$ is smooth if $f \circ \varphi^{-1}$ is smooth in the elementary calculus sense[5], for any[6] local coordinate chart φ. Smooth functions are meant to be differentiated along smooth curves.

Let $\varphi : U \to \mathbb{R}^d$ be a local coordinate chart. A curve $\gamma : (-\epsilon, \epsilon) \to U$ is *smooth* if

$$\varphi \circ \gamma : (-\epsilon, \epsilon) \to \mathbb{R}^d$$

is smooth. More generally, a curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is *smooth* if the restrictions of the curve to domains of coordinate charts are smooth.

We might attempt to define the "velocity vector" of a curve γ at the point $\gamma(0)$ to be

$$\dot{\gamma}(0) \stackrel{?}{:=} \frac{d\gamma}{dt}\Big|_{t=0} = \lim_{\delta \to 0} \frac{\gamma(\delta) - \gamma(0)}{\delta}. \tag{2.3}$$

[5]Partial derivatives of all orders exist.

[6]At any point $p \in \mathcal{M}$, we only need to check smoothness on one local chart with p in its domain. (Why?)

But the numerator in (2.3) does not even make sense!

Instead, we view γ as specifying an assignment

$$\begin{aligned} \dot\gamma(0) : C^\infty(\mathcal{M}) &\to \mathbb{R} \\ f &\mapsto \frac{d(f\circ\gamma)}{dt}\Big|_{t=0}, \end{aligned} \tag{2.4}$$

where $C^\infty(\mathcal{M})$ denotes the space of smooth functions $\mathcal{M}\to\mathbb{R}$. Eq. (2.4) is nothing but the rate-of-change of f along the curve γ, evaluated at $t=0$.

Clearly, the assignment (2.4) depends linearly on the function f. But it has a crucial extra property with respect to the pointwise multiplication $f_1\cdot f_2$ of functions. Writing $a=\gamma(0)$, we have (Exercise),

$$\dot\gamma(0)(f_1\cdot f_2) = f_1(a)\cdot\dot\gamma(0)(f_2) + \dot\gamma(0)(f_1)\cdot f_2(a), \qquad f_1, f_2\in C^\infty(\mathcal{M}). \tag{2.5}$$

Property (2.5) looks like a "Leibniz rule", and one says that $\dot\gamma(0) : C^\infty(\mathcal{M})\to\mathbb{R}$ is a *derivation* of the algebra $C^\infty(\mathcal{M})$ at $a=\gamma(0)$. In a similar way, for $t\in(-\epsilon,\epsilon)$, we have derivations $\dot\gamma(t) : C^\infty(\mathcal{M})\to\mathbb{R}$ at $\gamma(t)$.

Fix $a\in\mathcal{M}$ and consider curves γ with $\gamma(0)=a$. It is certainly possible that distinct curves give rise to the same derivation at a. Choose a coordinate chart φ with $\varphi(a)=0$, so

$$\varphi\circ\gamma : (-\epsilon,\epsilon)\to\mathbb{R}^d$$

is a curve in $\mathbb{R}^d$ with initial point being the origin. The tangent vector of $\varphi\circ\gamma$ at 0 is defined in the usual way,

$$\frac{d(\varphi\circ\gamma)}{dt}\Big|_{t=0}\in\mathbb{R}^d. \tag{2.6}$$

Let $\hat\gamma$ be any other curve with the same tangent vector,

$$\frac{d(\varphi\circ\hat\gamma)}{dt}\Big|_{t=0} = \frac{d(\varphi\circ\gamma)}{dt}\Big|_{t=0}. \tag{2.7}$$

For any $f \in C^\infty(\mathcal{M})$, we have

$$\begin{aligned}
\dot{\hat{\gamma}}(0)(f) = \frac{d(f\circ\hat{\gamma})}{dt}\Big|_{t=0} &= \frac{d(f\circ\varphi^{-1}\circ\varphi\circ\hat{\gamma})}{dt}\Big|_{t=0} && \text{(Eq. (2.4))}\\
&= \sum_{j=1}^{d} \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}\Big|_{\varphi\circ\hat{\gamma}(0)} \frac{d(\varphi\circ\hat{\gamma})^j}{dt}\Big|_{t=0} && \text{(Chain rule)}\\
&= \sum_{j=1}^{d} \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}\Big|_{\varphi\circ\gamma(0)} \frac{d(\varphi\circ\gamma)^j}{dt}\Big|_{t=0} && \text{(Eq. (2.7))}\\
&= \frac{d(f\circ\varphi^{-1}\circ\varphi\circ\gamma)}{dt}\Big|_{t=0} && \text{(Chain rule)}\\
&= \dot{\gamma}(0)(f). && \text{(Eq. (2.4))}
\end{aligned}$$

This calculation shows that the curves $\gamma, \hat{\gamma}$ define the same derivations of $C^\infty(\mathcal{M})$ at a, iff their tangent vectors (2.6) coincide.

Definition 2. The *tangent space at* $a \in \mathcal{M}$, denoted $T_a\mathcal{M}$, is the vector space of derivations of $C^\infty(\mathcal{M})$ at a. The *velocity* of a curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ at $t = 0$ is the derivation

$$\dot{\gamma}(0) \in T_{\gamma(0)}\mathcal{M}$$

at $\gamma(0)$ defined by (2.4).

Example 2.2. Let φ be a coordinate chart with $\varphi(a) = 0$. For $i = 1, \ldots, d$, consider the i-th coordinate line

$$\ell_i : t \mapsto (0, \ldots, 0, \underbrace{t}_{i\text{-th}}, 0, \ldots, 0).$$

Compose with φ^{-1} to obtain a curve $\gamma_i = \varphi^{-1} \circ \ell_i$ in $\mathcal{M}$. By definition,

$$\dot{\gamma}_i(0)(f) = \frac{d(f\circ\gamma_i)}{dt}\Big|_{t=0} = \sum_{j=1}^{d} \frac{\partial(f\circ\varphi^{-1})}{\partial x^j}\Big|_0 \underbrace{\frac{d(\ell_i)^j}{dt}\Big|_{t=0}}_{=\delta_i^j} = \frac{\partial(f\circ\varphi^{-1})}{\partial x^i}\Big|_0.$$

The right side is the i-th partial derivative of the coordinate representation of f. It is customary to write

$$\partial_i\big|_a := \dot{\gamma}_i(0),$$

and just call it the i-th partial derivative (with respect to the coordinate chart φ).

Example 2.2 shows that we always have the partial derivatives $\partial_i|_a \in T_a\mathcal{M}, i = 1, \ldots, d$, with respect to some coordinate chart. These partial derivatives are easily seen to be linearly independent. As we will learn, tangent spaces have the same dimension as the manifold. This means that we have non-canonical linear isomorphisms $T_a\mathcal{M} \cong \mathbb{R}^d$, by choosing a coordinate chart and taking the partial derivatives $\partial_i|_a$ as a tangent space basis.

The *tangent vector bundle* of $\mathcal{M}$ is the disjoint union

$$T\mathcal{M} = \bigsqcup_{a\in\mathcal{M}} T_a\mathcal{M},$$

equipped with the natural projection map $\pi : T\mathcal{M} \to \mathcal{M}$ taking $T_a\mathcal{M}$ to a. Because of the availability of local coordinate charts on $\mathcal{M}$, we can actually "smoothly relate" $T_a\mathcal{M}$ for nearby a. Indeed, $\pi : T\mathcal{M} \to \mathcal{M}$ has the structure of a smooth vector bundle. We postpone the precise definition and construction to a later lecture.

The manifold structure of $\mathcal{M}$ allows us to make sense of velocities $\dot\gamma(t)$ as tangent vectors in $T_{\gamma(t)}\mathcal{M}$, in a coordinate-independent way. But we still do not have a way to compare velocities at different times (or different points on the curve). This difficulty already applies to the affine space $\mathcal{A}$ when regarded as a bare manifold.

2.4.1 Parallelism of affine space

Let us return to the example of manifold $\mathcal{A}$ underlying a d-dimensional affine space. The $(V, +)$-torsor structure determines the straight lines,

$$\begin{aligned} L_{a,v} : \mathbb{R} &\to \mathcal{A} \\ t &\mapsto a + tv, \qquad\qquad a \in \mathcal{A}, v \in V, \end{aligned}$$

which determine derivations $\dot L_{a,v}(0)$ of $C^\infty(\mathcal{A})$ at a. When applied to $f \in C^\infty(\mathcal{A})$, we get

$$\dot L_{a,v}(0)(f) = \frac{d(f \circ L_{a,v})}{dt}\Big|_{t=0} = \frac{df(a+tv)}{dt}\Big|_{t=0} = \lim_{\delta\to 0} \frac{f(a+\delta v) - f(a)}{\delta} \in \mathbb{R}.$$

This is the *directional derivative* of f in the direction of the vector $v \in V$, evaluated at the point a. The assignment $v \to \dot L_{a,v}(0)$ depends linearly on

v, and gives an injective map $V \to T_a\mathcal{A}$. By dimension reasons, we actually have $V = T_a\mathcal{A}$. Thus the V-action on $\mathcal{A}$ provides *canonical* isomorphisms

$$T_a\mathcal{A} = V \tag{2.8}$$

of all the tangent spaces with the *same* V. This is why, in affine space, it makes sense to talk consistently about a velocity vector without specifying which point we are at. In particular, we are allowed to compare (or "parallel transport") velocity vectors attached to different points on $\mathcal{A}$. We can even differentiate the velocity vectors along a curve in $\mathcal{A}$ to get the *acceleration*. In the modern language of vector bundles, one says that there is a *canonical trivialization*

$$T\mathcal{A} = \mathcal{A} \times V,$$

of the tangent bundle of affine space.

On a general manifold $\mathcal{M}$, the concept of acceleration is much more subtle — one requires an extra structure called a (affine) *connection* on $T\mathcal{M}$. Where might such an extra structure come from? As it turns out, if $\mathcal{M}$ is a Riemannian manifold, there is a canonical connection on $T\mathcal{M}$, called the Levi–Civita connection. In modern physics and mathematics, many other kinds of vector bundles occur, and these often do not come with canonical connections; e.g., the vector bundle where quantum wavefunctions are taken from. Details of these concepts will be provided in a later lecture.