17 Parallel transport and covariant differentiation

17.1 Covariant derivatives on a vector bundle

We can pointwise multiply sections of a vector bundle with smooth functions. It will be convenient to write this as a right-multiplication,

$$\psi \cdot h, \qquad \psi \in \Gamma(E),\, h \in C^\infty(X, \mathbb{K}).$$

Definition 75. A *covariant derivative* on a vector bundle $E \to X$ is a map

$$\begin{aligned} \nabla : \mathfrak{X}(X) \times \Gamma(E) &\to \Gamma(E) \\ (v, \psi) &\mapsto \nabla_v \psi \end{aligned}$$

which is

- $C^\infty(X)$-linear in the argument v;
- $\nabla_v(\psi \cdot h) = \psi \cdot v(h) + (\nabla_v \psi) \cdot h$, for all $\psi \in \Gamma(E)$, $h \in C^\infty(X, \mathbb{K})$.

The *curvature* F^∇ of ∇ is the element of $\Omega^2(X, \mathrm{End}(E))$ defined by

$$F^\nabla(u, v)\psi = \nabla_u \nabla_v \psi - \nabla_v \nabla_u \psi - \nabla_{[u,v]}\psi, \qquad u, v \in \mathfrak{X}(X),\ \psi \in \Gamma(E). \tag{17.1}$$

Example 17.1. Suppose $E \cong X \times \mathbb{K}^n$, so $\Gamma(E) \cong C^\infty(X, \mathbb{K}^n)$. Then there is the "trivial" covariant derivative given by component-wise differentiation,

$$\left(\nabla^{\mathrm{triv}}_v \psi\right)^j = v(\psi^j) \equiv d\psi^j(v), \qquad j = 1, \ldots, n.$$

This covariant derivative is manifestly flat, $F^{\nabla^{\mathrm{triv}}} \equiv 0$.

Definition 75 captures the idea of "intrinsic/covariant differentiation" of vector bundle sections, independent of choosing local trivializations for E. But what is the geometrical mechanism making such a differentiation possible?

17.2 Parallel transport in principal bundles

Definition 76. Let $\pi : P \to X$ be a principal G-bundle with connection ω. Let $\gamma : [0,1] \to X$ be a curve in X. A curve $\tilde{\gamma} : [0,1] \to P$ is a *horizontal lift* of γ if $\pi \circ \tilde{\gamma} = \gamma$ and its velocity vectors are horizontal for all $t \in [0,1]$.

The velocity vectors of γ admit horizontal lifts, but we need to check that these lifts integrate to a curve $\tilde{\gamma}$.

Theorem 17.1. *Let $\pi : P \to X$ be a principal G-bundle with connection, and let $\gamma : [0,1] \to X$. For each $p \in P_{\gamma(0)}$ in the initial fibre, there exists a unique horizontal lift $\tilde{\gamma}$ with starting point $\tilde{\gamma}(0) = p$.*

(Omitted in lectures.) First, consider the case of a trivial bundle $P = U \times G$. A lift of γ starting at $p = (\gamma(0), g_0)$ would have the form

$$\tilde{\gamma}(t) = (\gamma(t), g(t)),$$

where $g : [0,1] \to G$ satisfies $g(0) = g_0$. The connection ω has the form

$$\omega = \omega_{\text{triv}} + \mathcal{A} = \pi_G^* \Theta + \mathcal{A},$$

where $\mathcal{A}$ is some equivariant horizontal $\mathfrak{g}$-valued 1-form. (See Example 16.2 and Exercise 16.3.) The condition for $\tilde{\gamma}$ to be horizontal is

$$\begin{aligned} 0 = \omega_{\tilde{\gamma}(t)}(\dot{\gamma}(t), \dot{g}(t)) &= \mathcal{A}_{(\gamma(t),g(t))}(\dot{\gamma}(t), 0) + \Theta_{g(t)}(\dot{g}(t)) \\ &= (R^*_{g(t)}\mathcal{A})_{(\gamma(t),e)}(\dot{\gamma}(t), 0) + (L_{g(t)^{-1}})_*(\dot{g}(t)) \\ &= \mathrm{Ad}_{g(t)^{-1}} \circ \mathcal{A}_{(\gamma(t),e)}(\dot{\gamma}(t), 0) + (L_{g(t)^{-1}})_*(\dot{g}(t)) \\ &= (L_{g(t)^{-1}})_* \left(R_{g(t)_*} \mathcal{A}_{(\gamma(t),e)}(\dot{\gamma}(t), 0) + \dot{g}(t) \right), \qquad \forall t \in [0,1]. \end{aligned}$$

So we seek a solution $g : t \mapsto g(t) \in G$ to the differential equation

$$\dot{g}(t) = -R_{g(t)_*} \mathcal{A}_{(\gamma(t),e)}(\dot{\gamma}(t), 0), \qquad t \in [0,1], \tag{17.2}$$

with initial condition $g(0) = g_0$. Rephrasing, $t \mapsto (g(t), t)$ is an integral curve of the following vector field on $G \times \mathbb{R}$,

$$v_{(h,t)} = \Big(- (R_h)_* (\mathcal{A}_{(\gamma(t),e)}(\dot{\gamma}(t), 0)), \frac{\partial}{\partial s}\Big|_{s=t} \Big),$$

with starting point $(g_0, 0)$. This integral curve exists at least for small times $t \in [0, \delta)$. (Here, δ is independent of $g_0 \in G$.) We obtain $g(t)$, thus $\tilde{\gamma}(t) =$

$(\gamma(t), g(t))$, for $t \in [0, \delta)$. If we took any other starting time $t' \in [0, 1]$, the same argument will yield a horizontal lift valid for some open time interval containing t'. Compactness of $[0, 1]$ ensures that we only need to patch together finitely many of these partial solutions.

Next, if γ lies within U with $P|_U$ trivializable, then we work in any local trivialization to obtain (the unique) $\tilde{\gamma}$ as above. In the general case, we have to subdivide γ into finitely many subcurves, each of which lies within a single U_i with $P|_{U_i}$ trivializable. □

Theorem 17.1 allows us to make the following definition:

Definition 77. Let $\pi : P \to X$ be a principal G-bundle with connection ω, and let $\gamma : [0, 1] \to X$ be a curve. *Parallel transport along* γ, with respect to ω, is the map

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

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

Proposition 17.2. *Parallel transport along any curve γ is a G-equivariant map,*

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

Proof. Exercise. (The G-invariance of horizontal-ness is relevant here.) □

In words: *Parallel transport respects relativity of frames.*

17.3 Covariant derivatives on associated vector bundles

Definition 78. Let $P \to X$ be a principal G-bundle and E be an associated vector bundle. Given a connection ω on P, each curve $\gamma : [0, 1] \to X$ determines a linear isomorphism

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

called the *parallel transport in E along γ*.

Note that τ_γ^E is well-defined, because for any $g \in G$,

$$\begin{aligned}
\tau_\gamma^{E,\omega}[p \cdot g, g^{-1} \cdot \xi] &= [\tau_\gamma^\omega(p \cdot g), g^{-1} \cdot \xi] \\
&= [\tau_\gamma^\omega(p) \cdot g, g^{-1} \cdot \xi] && \text{(Prop. 17.2)} \\
&= [\tau_\gamma^\omega(p), \xi] && \text{(assoc. bundle)} \\
&= \tau_\gamma^{E,\omega}[p, \xi].
\end{aligned}$$

That $\tau_\gamma^{E,\omega}$ is an isomorphism is easily checked.

Definition 79. Let $P \to X$ be a principal G-bundle with connection ω, and let E be an associated vector bundle. The *induced covariant derivative* on E is defined by

$$\left(\nabla_v^\omega \psi\right)(x) \equiv \left(\nabla_v^{E,\omega} \psi\right)(x) := \frac{d}{dt}\Big|_{t=0} \left(\tau_{\gamma|_{[0,t]}}^{E,\omega}\right)^{-1} \big(\psi(\gamma(t))\big), \quad x \in X,\, v \in \mathfrak{X}(X),\, \psi \in \Gamma(E), \tag{17.3}$$

where γ is any curve representing $v_x \in T_x X$.

Thus, we first parallel transport the $\psi(\gamma(t)) \in E_{\gamma(t)}$ to the initial fibre $E_{\gamma(0)}$, then it becomes meaningful to compare them, and take the usual t-derivative.

Proposition 17.3. *The formula* (17.3) *only depends on the tangent vector v_x, not on the choice of curve γ.*

Proof. (Omitted in lectures.) Pick a local gauge $s : U \to P$ around $x \in U$, so

$$\psi(x') = [s(x'), \xi(x')], \qquad x' \in U,$$

becomes represented by some function $\xi : U \to \mathbb{K}^n$. Also, γ becomes a $s \circ \gamma$ in P starting at $s(x)$. Apply parallel transport to get a curve within the initial fibre $P|_x$, which we parametrize as

$$\left(\tau_{\gamma|_{[0,t]}}^{\omega}\right)^{-1}(s \circ \gamma(t)) =: s(x) \cdot g(t), \qquad g : [0,t] \to G. \tag{17.4}$$

Then

$$\begin{aligned}
\left(\tau_{\gamma|_{[0,t]}}^{E,\omega}\right)^{-1}(\psi(\gamma(t))) &= \left[\left(\tau_{\gamma|_{[0,t]}}^{\omega}\right)^{-1}(s \circ \gamma(t)), \xi(\gamma(t))\right] && \text{(Definition of } \tau_\gamma^{E,\omega}\text{)} \\
&= \left[s(x) \cdot g(t), \xi(\gamma(t))\right] && \text{(Eq. (17.4))} \\
&= \left[s(x), \rho(g(t)) \cdot \xi(\gamma(t))\right], && \text{(associated bundle)}
\end{aligned}$$

so the covariant derivative formula (17.3) is

$$\begin{aligned}(\nabla^\omega_v\psi)(x) &= \frac{d}{dt}\Big|_{t=0}\big[s(x), \rho(g(t))\cdot\xi(\gamma(t))\big] \\ &= \big[s(x), d\xi(v_x) + (d\rho)_e(\dot g(0))\cdot\xi(x)\big],\end{aligned} \tag{17.5}$$

where we used $g(0) = e$ in the last equality, and $(d\rho)_e$ means the induced Lie algebra homomorphism

$$(d\rho)_e : \mathfrak{g} \to \mathfrak{gl}(n, \mathbb{K}). \tag{17.6}$$

In formula (17.5), the term $d\xi(v_x)$ is the directional derivative of the $\mathbb{K}^n$-valued function ξ, as expected. This term does not depend on the choice of γ (or on the connection ω). The extra (ω-dependent) term

$$(d\rho)_e(\dot g(0))\cdot\xi(x)$$

depends on $t \mapsto g(t)$, thus on the path γ (Eq. (17.4)). But actually, the following calculation shows that only $\dot\gamma(0) = v_x$ matters,

$$\begin{aligned}\dot g(0) &= \omega\left(\frac{d}{dt}\Big|_{t=0}(s(x)\cdot g(t))\right) && \text{(connection is vert. tautological)}\\ &= \omega\left(\frac{d}{dt}\Big|_{t=0}\overbrace{\tau^\omega_{\gamma|_{[0,t]}}(s(x))}^{\text{hori. curve}}\cdot g(0) + \frac{d}{dt}\Big|_{t=0}(s(x)\cdot g(t))\right) && \text{(connection kills horizontals)}\\ &= \omega\left(\frac{d}{dt}\Big|_{t=0}(s\circ\gamma)\right) && \text{(Chain rule, (17.4))}\\ &= s^*\omega(v_x).\end{aligned}$$

□

Because of Prop. 17.3, specifically the formula (17.5), the following gives an unambiguous formula for ∇:

Proposition 17.4. *With respect to a local gauge $s : U \to P$, the local formula for $\nabla^\omega \equiv \nabla^{E,\omega}$ is*

$$\nabla^\omega_v\psi = \big[s, d\xi(v) + s^*\omega(v)\cdot\xi\big], \qquad v \in \mathfrak{X}(X),\ \psi \in \Gamma(E), \tag{17.7}$$

where $\xi : U \to \mathbb{K}^n$ is the local function representing ψ with respect to s, and the induced Lie algebra representation (17.6) *is implicit.*

Let Ω^ω be the curvature of ω. Similar to (17.7), the local field strength $s^*\Omega^\omega$ appears in the local formula for the curvature of the induced covariant derivative $\nabla^\omega \equiv \nabla^{E,\omega}$ (Exercise),

$$F^{\nabla^\omega}(u,v)\psi = [s, s^*\Omega^\omega(u,v)\cdot\xi], \qquad u,v\in\mathfrak{X}(X), \psi\equiv[s,\xi]\in\Gamma(E). \quad (17.8)$$

Exercise 17.1. Check that the induced ∇^ω does satisfy the properties of a covariant derivative as axiomatized in Definition 75.

17.4 Connection coefficients and physics notation

Suppose the associated bundle has rank n. Let $s: U\to P$ be a local gauge, and $\partial_j, j=1,\dots,d$ be local coordinate tangent vector fields over $U\subset X$. The local gauge potential may be expanded as

$$s^*\omega = \mathcal{A}_j dx^j,$$

with coefficient functions $\mathcal{A}_j = s^*\omega(\partial_j)\in C^\infty(U,\mathfrak{g})$. More precisely, $\mathcal{A}_j = \left(\mathcal{A}^b{}_{ja}\right)_{a,b=1,\dots,n}$ are $n\times n$ matrix-valued functions obtained after applying the Lie algebra representation $d\rho_e:\mathfrak{g}\to\mathfrak{gl}(n,\mathbb{K})$.

Write η_b for the b-th basis vector of $\mathbb{K}^n$, so $e_b := [s,\eta_b]$ provide a local frame for E. The j-th covariant partial derivative of the section e_b is

$$\begin{aligned}\nabla^E_j e_b := \nabla^E_{\partial_j} e_b &= [s, s^*\omega(\partial_j)\cdot\eta_b] \qquad\qquad \text{(local formula (17.7))}\\ &= [s, \mathcal{A}_j\cdot\eta_b] \\ &= \left[s, \mathcal{A}^a{}_{jb}\eta_a\right] = e_a\mathcal{A}^a{}_{jb}.\end{aligned} \quad (17.9)$$

A general section of E is locally expanded as $\psi = e_b\psi^b$. By the Leibniz rule for covariant derivatives,

$$\begin{aligned}\nabla^E_j\psi = \nabla^E_j(e_b\psi^b) &= e_b(\partial_j\psi^b) + (\nabla^E_j e_b)\psi^b \\ &= e_a(\partial_j\psi^a + \mathcal{A}^a{}_{jb}\psi^b). \qquad\qquad \text{(Eq. (17.9))}\end{aligned}$$

In physics, usually $\mathbb{K}=\mathbb{C}$, and one simply writes the above local formula as

$$\text{“ } \nabla_j \;=\; \partial_j - i\mathcal{A}_j \text{ ”} \quad \text{acting on } \mathbb{C}^n\text{-valued “functions” } \psi,$$

with an extra factor of $-i$ due to differing Lie algebra conventions.

17.4.1 Magnetic Schrödinger operators

Classical electromagnetism In classical electromagnetism, a magnetic field is a 2-form $B \in \Omega^2(X)$ on the spatial manifold X. This is because magnetic fields are measured via their *fluxes* $\int_S B$ across oriented surfaces S. In more elementary presentations, X is an oriented Riemannian 3-manifold, then B is regarded as a tangent vector field via

$$\mathbf{B} = (*B)^\sharp \in \mathfrak{X}(X),$$

where $*$ is the Hodge star operation turning 2-forms into 1-forms. Note that this involves a *choice* of orientation (left/right-hand-rule).

As far as one knows, magnetic fields are divergence free. In the 2-form formalism, means $dB = 0$. Then, locally, we can always find some 1-form A such that $dA = B$ holds. (In tangent vector formalism, this is $\mathbf{B} = \operatorname{curl} \mathbf{A}$ for the mangetic vector potential $\mathbf{A}$). The choice of local A is not unique — one can always modify it by any a local gradient 1-form, $A \rightsquigarrow A + d\Lambda$. This is the *local gauge freedom* in choosing potentials. A *classical* charged particle is only influenced by B (at the particle's position) via the Lorentz force law, and there is no way to observe a preferred (local) potential A.

Minimal coupling Recall that the naïve Schrödinger operator involved a scalar Laplacian over X. Let $B \in \Omega^2(X)$ be a magnetic field. Soon after Schrödinger introduced his wave equation, physicists observed (at least in the Euclidean $\mathbb{R}^d$ case) that if the Cartesian partial derivatives ∂_j in the Schrödinger Hamiltonian operator were replaced by

$$\nabla_j = \partial_j - i\frac{qe}{\hbar}\mathcal{A}_j, \tag{17.10}$$

with $A = \mathcal{A}_j dx^j$ being a magnetic potential for B, then the resulting *magnetic Schrödinger Hamiltonian* would correctly recover the classical Lorentz force law in a suitable classical limit. The prescription (17.10) is called *minimal coupling*.

Apparently, any choice of A satisfying $dA = B$ can be used in (17.10) without affecting observable physics, so that minimal coupling is "gauge-invariant". Many successful applications followed in in atomic and molecular physics.

Actually, the is a very important subtlety in (17.10), which was only uncovered and properly appreciated many decades later. The is related to

the following persistent mystery: why it is essential to introduce A (rather than B directly) into the equations of quantum mechanics, if the potentials A are indeed "fictitious" mathematical devices?

We know that there are U(1) phase ambiguities in the complex values of wavefunctions. Yet we must not simply mod out by U(1) and work with real-valued wavefunctions, because the *relative phases* are observable. Crucially, relative phases are assessed *locally* (e.g., by interference fringes), and this local relativity-of-phases should be respected when we parallel transport wavefunctions.

In modern geometric language:

- There is a principal U(1) bundle over X, with a connection.
- The (gauge-invariant) field strength of the connection is (i times of) the magnetic field.
- The "wavefunctions" for a charge-q particle are sections of an associated Hermitian line bundle $\mathcal{L}$ built from the charge-q unitary representation of U(1),

$$\begin{aligned} \rho^{(q)} : \mathrm{U}(1) &\to \mathrm{U}(1) \subset \mathrm{GL}(1,\mathbb{C}) \\ e^{i\theta} &\mapsto e^{iq\theta}. \end{aligned}$$

- The parallel transport (covariant derivative) on $\mathcal{L}$ is induced from ω.

The local gauge potentials have the form $s^*\omega = iA = i\mathcal{A}_j dx^j$, for some real-valued functions $\mathcal{A}_j$. Because the induced Lie algebra representation is

$$d\rho_e^{(q)} : \mathfrak{u}(1) = i\mathbb{R} \to \mathrm{End}(\mathbb{C}), \qquad i\lambda \mapsto iq\lambda,$$

the induced covariant derivative $\nabla^{\mathcal{L},\omega}$ has the formula given in (17.10). There, the extra $-e/\hbar$ factor is introduced to balance out physical units ($-e$ is the elementary electron charge).

Warning: In Cartesian coordinates for Euclidean space, we can take $-\sum_{j=1}^{d}(\nabla_j^{\mathcal{L},\omega})^2$ as the modified Laplacian in the Schrödinger equation. On a more general space X, the construction is more intricate, and is sketched below.

- We need X to be have a Riemannian metric g. Then there is an associated Riemannian volume form, and an inner product on the smooth sections of $\mathcal{L} \to X$,

$$\langle \varphi | \psi \rangle = \int_X \langle \varphi(x) | \psi(x) \rangle_{\mathcal{L}_x} \operatorname{vol}_g.$$

  The Hilbert space completion can be identified with the square-integrable sections, $\Gamma_{L^2}(X; \mathcal{L})$.

- The covariant derivative can be viewed as a map

$$\nabla^{\mathcal{L}} : \Gamma(X; \mathcal{L}) \to \Omega^1(X; \mathcal{L}),$$

  Here, the 1-form part has an inner product from the Riemannian metric. So the adjoint operator to $\nabla^{\mathcal{L}}$ makes sense, and we may define the *connection Laplacian* as

$$\Delta^{\mathcal{L}} := \nabla^{\mathcal{L}*} \nabla^{\mathcal{L}}, \tag{17.11}$$

  which is a formally self-adjoint operator on the compactly-supported smooth sections of $\mathcal{L}$. Note that $\Delta^{\mathcal{L}}$ is defined in a manifestly gauge-independent way. Locally, we may write $\Delta^{\mathcal{L}} = (d - i\frac{qe}{\hbar}A)^*(d - i\frac{qe}{\hbar}A)$.

- Similarly for connection Laplacians on vector bundles. (There is a more general formulation which makes sense without metrics, but then we would not have self-adjointness concepts.)

The situation described above could be summarized as follows: Classical magnetic potentials A are now interpreted in quantum mechanics as U(1) gauge potentials (after multiplication by i) for a connection on the principal U(1) "phase bundle". In Section 17.6, we will investigate an important subtle consequence.

17.5 Holonomy of parallel transport

For a connection ω on a principal G-bundle, let $\tau = \tau^\omega$ be the associated parallel transport.

Exercise 17.2. Show that τ_γ has the following properties:

- It is independent of the parametrization of the curve γ.

- If γ is the smooth concatenation of two paths $\gamma_1, \gamma_2 : [0,1] \to X$,

$$\gamma(t) = \begin{cases} \gamma_1(2t), & t \in [0, \frac{1}{2}], \\ \gamma_2(2t-1), & t \in [\frac{1}{2}, 1], \end{cases}$$

  then $\tau_\gamma = \tau_{\gamma_2} \circ \tau_{\gamma_1}$.

- For the flipped path $\gamma_{\text{flip}} : t \mapsto \gamma(1-t)$,

$$\tau_{\gamma_{\text{flip}}} = (\tau_\gamma)^{-1}.$$

Parallel transport transforms nicely under gauge transformations:

Proposition 17.5. *Let ω be a connection on a principal G-bundle P. Let $F \in \mathcal{G}(P)$ be a gauge transformation, then*

$$\tau_\gamma^{F^*\omega} = F^{-1} \circ \tau_\gamma^\omega \circ F.$$

Proof. Given p, let $\tilde{\gamma}$ be the ω-horizontal lift starting at $F(p)$. Check that $F^{-1} \circ \tilde{\gamma}$ is the $(F^*\omega)$-horizontal lift starting at p. (Exercise) □

A *loop* on X is a smooth map $\gamma : S^1 \to X$. Fix a basepoint x on the loop, and regard γ as a curve with $\gamma(1) = \gamma(0) = x$. Parallel transport along the loop γ is then a self-map of the fibre P_x,

$$\tau_\gamma^\omega : P_x \to P_x,$$

which is a G-torsor automorphism by Prop. 17.2. It is called the *holonomy of ω along the loop γ*.

The holonomy map can be written as right multiplication,

$$\tau_\gamma(p) = p \cdot h(\gamma; p), \qquad p \in P_x,$$

by a suitable $h(\gamma; p) \in G$. The *conjugacy class* of $h(\gamma; p)$ depends only on γ, and the gauge-equivalence class of ω:

- Changing starting point on fibre:

$$\tau_\gamma(p \cdot g) = \tau_\gamma(p) \cdot g = p \cdot h(\gamma; p) \cdot g = (p \cdot g) \cdot (\underbrace{g^{-1} h(\gamma; p) g}_{h(\gamma; p \cdot g)}),$$

  so $h(\gamma; p \cdot g)$ is conjugate to $h(\gamma; p)$.

- Changing basepoint on loop: Exercise.
- Gauge transformation of connection: Use Prop. 17.5.

Pick a loop γ on X. If G is abelian (e.g. $G = \mathrm{U}(1)$), the holonomy is gauge-invariant by the above discussion. If G is non-abelian, we may pick some group representation $\rho : G \to \mathrm{GL}(n)$, then the trace of the matrix $\rho(h(\gamma; p))$ is a gauge-invariant quantity of the connection, called a *Wilson loop* invariant. Such a ρ usually arises in the associated bundle construction.

17.6 Aharonov–Bohm effect

Consider the plane $X = \mathbb{R}^2$ with unit circle S^1. A magnetic field on X can be written as $B = b\,dx \wedge dy$ for some real-valued function b. Suppose B is supported within the disc D_ϵ of radius $\epsilon \ll 1$, and has total flux $2\pi k$,

$$\int_{D_1} B = \int_{D_\epsilon} b\,dxdy = 2\pi k.$$

Since $\mathbb{R}^2$ is contractible, there exists a globally defined potential A such that $dA = B$. For *classical electromagnetism*, we are free to use any locally-defined A_β with domains $U_\beta \subset \mathbb{R}^2$, as long as

$$dA_\beta = B|_{U_\beta}. \tag{17.12}$$

Let us consider those potentials A_β whose domains U_β contain D_ϵ. Then for any loop ℓ in U_β which winds around D_ϵ once (anticlockwise), Stokes' theorem enforces the constraint

$$\oint_\ell A_\beta = \int_{D_\epsilon} dA_\beta = \int_{D_\epsilon} B = 2\pi k, \qquad \forall\, \beta\, :\, U_\beta \supset D_\epsilon. \tag{17.13}$$

Next, consider potentials defined on the *non-simply-connected* exterior domain $U_{\mathrm{out}} = \{(x, y)\, :\, x^2 + y^2 > \epsilon^2\}$. One obvious choice is

$$A_{\mathrm{zero}}(x, y) = 0, \qquad (x, y) \in U_{\mathrm{out}}.$$

Another possible choice is

$$A_{\mathrm{out}}(x, y) = A_{\mathrm{out}}(r, \theta) = k\,d\theta, \qquad r > \epsilon,$$

since $dA_{\text{out}} = 0 = B|_{U_{\text{out}}}$. (Here (r, θ) are the polar coordinates, and $d\theta$ is defined on all of U_{out}, by the usual abuse of notation.)

However, there seems to be a sense in which A_{out} is "better" than A_{zero}, because

$$\oint_\ell A_{\text{out}} = 2\pi k, \qquad \oint_\ell A_{\text{zero}} \neq 2\pi k.$$

In other words, A_{out} satisfies the loop integral constraint (17.13) but A_{zero} does not.

Classically, A_{zero} is *not* required to satisfy the constraint (17.13), because the Stokes' theorem argument does not apply to it. So there is no problem in using it as the magnetic potential over U_{out}. Indeed, A_{out} and A_{zero} are indistinguishable by classical experiments carried out in U_{out}.

Notice that A_{out} and A_{zero} are *not* gauge equivalent over U_{out} — we cannot write $A_{\text{out}} = A_{\text{zero}} + d\Lambda$ for any $\Lambda \in \Omega^0(U_{\text{out}})$. Therefore, it is reasonable to suspect that A_{out} and A_{zero} are actually descriptions of *distinct, gauge-independent* objects ω and ω', where ω, ω' just happen to encode the same zero-magnetic field information.

We had studied just such an object — a connection admits local $\mathfrak{g}$-valued 1-forms as gauge potentials. It is natural to consider connections on a principal $\mathrm{GL}^+(1, \mathbb{R})$-bundle, since $\mathfrak{gl}(1, \mathbb{R}) = \mathbb{R}$, so the field strength would be a globally defined $\mathbb{R}$-valued 2-form, just like the magnetic field B. In this case, for a loop ℓ encircling D_ϵ, the holonomy for such a connection would be

$$\exp(-\oint_\ell A_\beta) = \exp(-2\pi k) \in \mathbb{R}_{>0}. \tag{17.14}$$

This equality comes from choosing any gauge potential A_β whose domain U_β includes D_ϵ (via Stokes' theorem). Since the holonomy is gauge-independent, it must hold for any gauge potential whose domain contains the loop ℓ (not necessarily D_ϵ as well).

In 1918, Hermann Weyl suggested precisely such an interpretation of the local gauge freedom of (electro)magnetic potentials. Roughly speaking, his idea was that "scale" was only a local concept, so that each spacetime point has an $\mathbb{R}_{>0} = \mathrm{GL}^+(1, \mathbb{R})$-torsor of possible scales. "Parallel transport of scale" would be locally described by gauge potentials identified with electromagnetic potentials, then the electromagnetic field would be the curvature of this parallel transport. This was the first "gauge theory". However, Weyl's proposal predicted possible "holonomies of spacetime scale" such as (17.14), and was rejected by Einstein, Pauli, and others, on empirical grounds.

About a decade later, with the development of *quantum mechanics* as modelled by Schrödinger's "wavefunctions" ψ, one had a physical theory in which U(1) principal bundles occur naturally. Now, all one had to do was to slightly modify Weyl's proposal, and interpret $-i\frac{e}{\hbar}A_\beta$ as the $\mathfrak{u}(1) = i\mathbb{R}$-valued local gauge potentials for a connection on a principal U(1)-bundle. The holonomy around ℓ is now a phase

$$e^{i\frac{e}{\hbar}\oint_\ell \mathcal{A}_\beta} = e^{2\pi ike/\hbar} \in \mathrm{U}(1). \tag{17.15}$$

This phase can be detected through the wavefunctions of a charged quantum particle — if qe is its charge, then it should acquire a *geometric phase* $\exp(2\pi ikqe/\hbar)$ when parallel transported around ℓ. Note that (17.15) gives a constraint on which classical local potentials A_β are admissible as local gauge potentials. In particular, A_{zero} is *not* admissible!

In 1959, Aharonov–Bohm[23] proposed a concrete experiment to measure the geometric phase (17.15). The idea was to pass two beams of charged particles around D_ϵ, one in the upper half-plane and one in the lower half-plane, then observe their interference fringes when then meet again. Because the travel time is the same along each path, the dynamical phases (e.g. from Schrödinger time evolution) cancel, and only the relative geometric phase around the nett loop survives. Theoretically, this phase can be varied by varying the flux $2\pi k$ through D_ϵ, then the interference pattern would shift. This *Aharonov–Bohm effect* was experimentally detected soon after it was proposed, and by the 1980s, had become well-established.

Remark. The AB effect is paradoxical from the classical perspective: B is only varied within D_ϵ, whereas the paths always avoid D_ϵ, yet there is a measurable effect. A classical explanation would require some *non-local* effect of $B|_{D_\epsilon}$ on motion taking place in U_{out}.

The Aharonov–Bohm effect is sometimes said to imply that classical "fictitious" magnetic vector potentials attain "real" physical status in quantum mechanics. Interpreted literally, this is incorrect, and in any case, it misses the important conceptual point. In quantum mechanics, the A_β remain gauge-dependent calculation devices which cannot be measured in any experiment. The AB-effect implies that the *role* of A_β in quantum mechanics is dramatically different from that in classical electromagnetism — its purpose is not (merely) to describe B, to describe the relative-phase-preserving

[23]Historically, W. Franz (1939) as well as W. Ehrenberg and R. Siday (1949) made similar proposals.

parallel transport of wavefunctions of charged quantum particles. Because of this new role, some classically admissible local potentials are no longer valid as local gauge potentials, even if they can reproduce B. Particularly, the ones which predict the wrong holonomies are inadmissible.

Physically speaking, what controls the mechanics of the charged quantum particle in a B-field, is the globally defined connection (whose curvature is iB), not B itself. In particular, the connection enters the Schrödinger equation of motion *locally*.