MATHEMATICAL APPENDICES

Here, for those with a somewhat more substantial mathematical background, I will try to fill in a few of the gaps left open in the lecture. The amount of background required will vary with the topic.

A. Quantum Mechanics

It will help to have at least a general sense of the basic formalism of quantum mechanics. Here, without a word of explanation, are a few rules of the game (this will require a bit of functional analysis).

1. Associated to any quantum mechanical system is a complex Hilbert space $\boldsymbol{H}$ and every *state* of the system is represented by a ray in $\boldsymbol{H}$ (that is, by an equivalence class of vectors that differ by nonzero scalar multiplication). A representative vector Ψ in the ray of norm 1 ($\| \Psi \| = 1$) is called a (*normalized*) *wave function* for the system. Note that Ψ and $e^{i\theta}\,\Psi$ represent the same state of the system.

   For a single particle moving in space, $\boldsymbol{H}$ is generally taken to be the Hilbert space of complex-valued (Lebesgue) square integrable functions on $\mathbb{R}^3$,
$$L^2\left(\mathbb{R}^3, \mathbb{C}\right).$$

2. Any physical quantity associated with the quantum system which can be measured (e.g., a component of the position, or momentum, or angular momentum, the kinetic, or potential, or total energy, etc.) is called an *observable* and is represented by a self-adjoint operator X on $\boldsymbol{H}$ (generally unbounded). There are rules, called the *Born Correspondence Rules*, for writing down the operator associated with any given classical observable. For example, the operator corresponding to the classical kinetic energy of a particle of mass m moving in space is a multiple of the Laplacian
$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) = -\frac{\hbar^2}{2m}\nabla^2$$
   where $\hbar = h/2\pi$ and h is Planck's constant.

3. The possible values for a measurement of an observable X are the elements of its spectrum $\sigma(X)$; these are real numbers because X is self-adjoint.

   According to the Spectral Theorem, every observable X has associated with it a spectral measure ξ_X supported on $\sigma(X)$ and a spectral resolution
$$X = \int_{\sigma(X)} \lambda\, d\xi_X(\lambda).$$

If the spectrum of X happens to be discrete, it will consist of countably many eigenvalues λ_i and the integral reduces to a sum

$$\sum_i \lambda_i P_{\lambda_i}$$

where P_{λ_i} is the orthogonal projection onto the eigenspace of λ_i.

4. If, when a measurement of the observable X is made, the state of the system is Ψ, then the *probability* that the measurement will yield a value in some Borel subset A of $\mathbb{R}$ is

$$\| \xi_X(A)\Psi \|^2 = \langle \Psi, \xi_X(A)\Psi \rangle .$$

Thus, in the case in which X has a discrete spectrum, a measurement will always yield an eigenvalue of X and the probability that the measurement will give the specific eigenvalue λ_i is just the squared norm of the projection of Ψ into the eigenspace of λ_i.

Notice that all of these probabilities are the same for both Ψ and $e^{i\theta}\Psi$.

Quantum mechanics does not predict the outcome of any experiment, but only the probability that a given outcome will occur and these probabilities are all given as the squared norm of a vector in a complex Hilbert space. This complex vector is called the *probability amplitude* of the outcome. If the outcome, viewed classically, could occur in a number of different ways (e.g., an Aharonov-Bohm electron could arrive at a point on the screen by way of slit 1 or by way of slit 2), then each has a probability amplitude associated with it and the probability amplitude of the outcome itself is the sum of these. Probability amplitudes *add* and so probabilities do not because, in general,

$$\| \Psi_1 + \Psi_2 \|^2 \neq \| \Psi_1 \|^2 + \| \Psi_2 \|^2 .$$

This is how interference manifests itself in quantum mechanics.

5. If the state of the system is Ψ when a measurement of the observable X is made and if X has a discrete spectrum and the result of the measurement is the eigenvalue λ_i, then, immediately after the measurement, the state of the system "collapses" to a (normalized) eigenfunction of X.

6. One particular observable for any quantum system is its total classical energy. The corresponding operator is denoted $\mathcal{H}$ and called the *Hamiltonian* of the system. The time evolution of the wave function Ψ is governed by the *Schrödinger equation*

$$i\hbar \frac{\partial \Psi}{\partial t} = \mathcal{H}\Psi .$$

Example: Consider an electromagnetic field with magnetic vector potential $\vec{A}(\vec{x},t)$ and electric scalar potential $V(\vec{x},t)$ so that

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} - \nabla V$$
$$\vec{B} = \nabla \times \vec{A}$$

where c is the speed of light *in vacuo* and $\nabla = \langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \rangle$. Then the Schrödinger equation for the wave function $\Psi(\vec{x},t)$ of a particle of mass m and charge q moving through this field is

$$i\hbar \frac{\partial \Psi}{\partial t} = \left[\frac{1}{2m}\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2 + qV \right]\Psi$$

where the meaning of $(i\hbar\nabla + \frac{q}{c}\vec{A})^2$ as an operator is

$$\begin{aligned}
\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi &= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)\cdot\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)\Psi \\
&= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)\cdot\left(i\hbar\nabla\Psi + \frac{q}{c}\vec{A}\Psi\right) \\
&= -\hbar^2\nabla^2\Psi + \frac{q^2}{c^2}\,|\vec{A}|^2\,\Psi + \frac{i\hbar q}{c}\nabla\left(\Psi\vec{A}\right) + \frac{i\hbar q}{c}\vec{A}\cdot\nabla\Psi \\
&= -\hbar^2\nabla^2\Psi + \frac{q^2}{c^2}\,|\vec{A}|^2\,\Psi + \frac{i\hbar q}{c}\left(\nabla\cdot\vec{A}\right)\Psi + \frac{2i\hbar q}{c}\vec{A}\cdot\nabla\Psi.
\end{aligned}$$

We will return to this example in the next section and show that a gauge transformation of the potentials is equivalent to a phase change in the wave function.

B. The Gauge Principle and Gauge Fields in Electrodynamics

Here we will show explicitly that, for a charged particle in an electromagnetic field, a gauge transformation of the electromagnetic potential corresponds precisely to a phase change in the wave function of the particle (only basic vector calculus is required here).

We consider an electromagnetic field with magnetic vector potential $\vec{A}(\vec{x},t)$ and electric scalar potential $V(\vec{x},t)$ so that

$$\begin{aligned}
\vec{E} &= -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} - \nabla V \\
\vec{B} &= \nabla \times \vec{A}
\end{aligned} \tag{1}$$

where c is the speed of light *in vacuo* and $\nabla = \langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \rangle$.

The wave function $\Psi(\vec{x}, t)$ of a particle of mass m and charge q in this electromagnetic field is assumed to satisfy the Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = \left(\frac{1}{2m} \left(i\hbar\nabla + \frac{q}{c} \vec{A} \right)^2 + qV \right) \Psi \tag{2}$$

where $\hbar$ is Planck's constant divided by 2π and the meaning of $(i\hbar\nabla + \frac{q}{c} \vec{A})^2$ as an operator is

$$\begin{aligned}
\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2 \Psi &= \left(i\hbar\nabla + \frac{q}{c} \vec{A} \right) \cdot \left(i\hbar\nabla + \frac{q}{c} \vec{A} \right) \Psi \\
&= \left(i\hbar\nabla + \frac{q}{c} \vec{A} \right) \cdot \left(i\hbar\nabla\Psi + \frac{q}{c} \vec{A}\Psi \right) \\
&= -\hbar^2\nabla^2\Psi + \frac{i\hbar q}{c} \nabla \cdot \left(\Psi\vec{A}\right) + \frac{i\hbar q}{c} \vec{A} \cdot \nabla\Psi + \frac{q^2}{c^2}\Psi \mid \vec{A} \mid^2 \\
&= -\hbar^2\nabla^2\Psi + \frac{i\hbar q}{c} \left(\Psi\left(\nabla \cdot \vec{A}\right) + \vec{A} \cdot \nabla\Psi\right) + \frac{i\hbar q}{c} \vec{A} \cdot \nabla\Psi + \frac{q^2}{c^2}\Psi \mid \vec{A} \mid^2
\end{aligned}$$

$$\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2 \Psi = -\hbar^2\nabla^2\Psi + \frac{i\hbar q}{c}\Psi\nabla \cdot \vec{A} + \frac{2i\hbar q}{c}\vec{A} \cdot \nabla\Psi + \frac{q^2}{c^2}\Psi \mid \vec{A} \mid^2 \tag{3}$$

Writing the Schrödinger equation (2) out in detail therefore gives

$$i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + \frac{q^2}{2mc^2}\Psi \mid \vec{A} \mid^2 + \frac{i\hbar q}{2mc}\Psi\nabla \cdot \vec{A} + \frac{i\hbar q}{mc}\vec{A} \cdot \nabla\Psi + qV\Psi \, . \tag{4}$$

A gauge transformation of the potentials $\vec{A}(\vec{x}, t)$ and $V(\vec{x}, t)$ is obtained by choosing an arbitrary differentiable, real-valued function $\Lambda(\vec{x}, t)$ and defining

$$\begin{aligned}
\vec{A}'(\vec{x}, t) &= \vec{A}(\vec{x}, t) - \nabla\Lambda(\vec{x}, t) \\
V'(\vec{x}, t) &= V(\vec{x}, t) + \frac{1}{c}\frac{\partial \Lambda}{\partial t} \, .
\end{aligned} \tag{5}$$

These are potentials for the same electromagnetic field because

$$\begin{aligned}
-\frac{1}{c}\frac{\partial \vec{A}'}{\partial t} - \nabla V' &= -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} + \frac{1}{c}\nabla\left(\frac{\partial \Lambda}{\partial t}\right) - \nabla V - \frac{1}{c}\nabla\left(\frac{\partial \Lambda}{\partial t}\right) \\
&= -\frac{1}{c}\frac{\partial \Lambda}{\partial t} - \nabla V \\
&= \vec{E}
\end{aligned}$$

and

$$\begin{aligned}
\nabla \times \vec{A}' &= \nabla \times \left(\vec{A} - \nabla\Lambda\right) = \nabla \times \vec{A} - \nabla \times (\nabla\Lambda) \\
&= \nabla \times \vec{A} - \vec{0} \\
&= \vec{B} \, .
\end{aligned}$$

Now we can state precisely what we intend to prove.

Theorem: $\Psi(\vec{x},t)$ *is a solution to*

$$i\hbar\,\frac{\partial\Psi}{\partial t} \;=\; \left(\,\frac{1}{2m}\left(i\hbar\nabla \,+\, \frac{q}{c}\,\vec{A}\right)^2 \,+\, qV\,\right)\Psi \tag{2}$$

if and only if

$$\Psi'\,(\,\vec{x}\,,\,t\,) \;=\; e^{-(iq/\hbar c)\;\Lambda(\vec{x},t)}\;\Psi\,(\,\vec{x}\,,\,t\,) \tag{6}$$

is a solution to

$$i\hbar\,\frac{\partial\Psi'}{\partial t} \;=\; \left(\,\frac{1}{2m}\left(i\hbar\nabla \,+\, \frac{q}{c}\left(\,\vec{A} - \nabla\Lambda\,\right)\right)^2 \,+\, q\left(\,V \,+\, \frac{1}{c}\,\frac{\partial\Lambda}{\partial t}\,\right)\right)\Psi'\,. \tag{7}$$

Remark: The Schrödinger equation is written in terms of the potentials, but the potentials are determined only up to a gauge transformation. The wave function has an amplitude and a phase, but the physical information about the charged particle is independent of the phase (until it interacts with other charged particles). The Theorem asserts that gauge freedom and phase freedom are really the same thing; choosing a gauge is the same as choosing a phase. From this one can draw a remarkable conclusion. Quantum mechanics decrees that one should be free to choose the phase of $\Psi(\vec{x},t)$ arbitrarily since $\Psi(\vec{x},t)$ and $e^{\theta(\vec{x},t)}\;\Psi(\vec{x},t)$ represent the same state, i.e., the same physics. This is an instance of what is called the *Principle of Local Gauge Invariance*. This is true even when the particle is free, i.e., when the electromagnetic field is zero in the region in which the particle moves so that (2) reduces to

$$i\hbar\,\frac{\partial\Psi}{\partial t} \;=\; \frac{1}{2m}\,(\,i\hbar\nabla\,)^2\;\Psi\,.$$

But then $\Psi'(\vec{x},t) \;=\; e^{-(iq/\hbar c)\;\Lambda(\vec{x},t)}\;\Psi(\vec{x},t)$, which represents the same state, satisfies

$$i\hbar\,\frac{\partial\Psi'}{\partial t} \;=\; \left(\,\frac{1}{2m}\left(i\hbar\nabla \,-\, \frac{q}{c}\,\nabla\Lambda\right)^2 \,+\, \frac{q}{c}\,\frac{\partial\Lambda}{\partial t}\,\right)\Psi'$$

and this is the Schrödinger equation for a particle moving in the potential field $\vec{A} = -\frac{q}{c}\nabla\Lambda,\;\; V \,=\, \frac{q}{c}\,\frac{\partial\Lambda}{\partial t}$. The Principle of Local Gauge Invariance *implies* the existence of Gauge Fields.

Proof: Assume Ψ satisfies (2), i.e., (4), and consider $\Psi' \;=\; e^{-(iq/\hbar c)\;\Lambda}\;\Psi$. Showing that Ψ' satisfies (7) is just a routine, but rather tedious calculation, but since no one seems to bother doing it we will carry out the computations in detail. First, the product rule gives

$$\begin{aligned}\frac{\partial\Psi'}{\partial t} &= \frac{\partial}{\partial t}\,\left(\,e^{-(iq/\hbar c)\;\Lambda}\;\Psi\,\right)\\ &= e^{-(iq/\hbar c)\;\Lambda}\,\frac{\partial\Psi}{\partial t} \,-\, \frac{iq}{\hbar c}\,\frac{\partial\Lambda}{\partial t}\,e^{-(iq/\hbar c)\;\Lambda}\;\Psi\end{aligned}$$

so

$$i\hbar \frac{\partial \Psi'}{\partial t} \;=\; e^{-(iq/\hbar c)\,\Lambda} \left(i\hbar \frac{\partial \Psi}{\partial t} \;+\; \frac{q}{c}\frac{\partial \Lambda}{\partial t}\,\Psi \right) . \tag{8}$$

Moreover,

$$q\left(V \;+\; \frac{1}{c}\frac{\partial \Lambda}{\partial t} \right) \Psi' \;=\; e^{-(iq/\hbar c)\,\Lambda} \left(qV\Psi \;+\; \frac{q}{c}\frac{\partial \Lambda}{\partial t}\,\Psi \right) . \tag{9}$$

Next we need to expand $(\,i\hbar\nabla + \frac{q}{c}\,(\,\vec{A} - \nabla\Lambda\,)\,)^2\,\Psi'$ just as we did for $(\,i\hbar\nabla + \frac{q}{c}\,\vec{A}\,)^2\,\Psi$ earlier. Here it is.

$$\begin{aligned}
&\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 \Psi' = \left(\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right) - \frac{q}{c}\nabla\Lambda\right)\cdot\left(\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right) - \frac{q}{c}\nabla\Lambda\right)\Psi' \\
&= \left(\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right) - \frac{q}{c}\nabla\Lambda\right)\cdot\left(\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)\Psi' - \frac{q}{c}\Psi'\nabla\Lambda\right) \\
&= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi' - \frac{q}{c}\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)(\Psi'\nabla\Lambda) - \frac{q}{c}\nabla\Lambda\cdot\left(i\hbar\nabla\Psi' + \frac{q}{c}\Psi'\vec{A}\right) + \frac{q^2}{c^2}\Psi'\mid\nabla\Lambda\mid^2 \\
&= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi' - \frac{i\hbar q}{c}\nabla\cdot(\Psi'\nabla\Lambda) - \frac{q^2}{c^2}\Psi'\left(\vec{A}\cdot\nabla\Lambda\right) - \frac{i\hbar q}{c}\nabla\Lambda\cdot\nabla\Psi' \\
&\qquad\qquad - \frac{q^2}{c^2}\Psi'\nabla\Lambda\cdot\vec{A} + \frac{q^2}{c^2}\Psi'\mid\nabla\Lambda\mid^2 \\
&= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi' - \frac{i\hbar q}{c}\Psi'\nabla^2\Lambda - \frac{i\hbar q}{c}\nabla\Psi'\cdot\nabla\Lambda - \frac{q^2}{c^2}\Psi'\left(\vec{A}\cdot\nabla\Lambda\right) - \frac{i\hbar q}{c}\nabla\Lambda\cdot\nabla\Psi' \\
&\qquad\qquad - \frac{q^2}{c^2}\Psi'\nabla\Lambda\cdot\vec{A} + \frac{q^2}{c^2}\Psi'\mid\nabla\Lambda\mid^2 \\
&= \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi' - \frac{2i\hbar q}{c}\nabla\Psi'\cdot\nabla\Lambda + e^{-(iq/\hbar c)\,\Lambda}\left[-\frac{i\hbar q}{c}\Psi\nabla^2\Lambda - \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda + \frac{q^2}{c^2}\Psi\mid\nabla\Lambda\mid^2\right]
\end{aligned}$$

In the second term we substitute

$$\nabla\Psi' \;=\; \nabla\left(e^{-(iq/\hbar c)\,\Lambda}\,\Psi\right) \;=\; e^{-(iq/\hbar c)\,\Lambda}\left(\nabla\Psi \;-\; \frac{iq}{\hbar c}\,\Psi\nabla\Lambda\right) \tag{10}$$

to obtain

$$\begin{aligned}
\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2\Psi' = \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi' + e^{-(iq/\hbar c)\,\Lambda}\Bigg[&-\frac{2i\hbar q}{c}\nabla\Psi\cdot\nabla\Lambda \\
&- \frac{2q^2}{c^2}\Psi\mid\nabla\Lambda\mid^2 - \frac{i\hbar q}{c}\Psi\nabla^2\Lambda - \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda \\
&+ \frac{q^2}{c^2}\Psi\mid\nabla\Lambda\mid^2\Bigg]
\end{aligned}$$

$$\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 \Psi' = \left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2 \Psi' + e^{-(iq/\hbar c)\,\Lambda}\,\left[-\frac{2i\hbar q}{c}\nabla\Psi\cdot\nabla\Lambda \right. \tag{11}$$

$$\left. -\frac{q^2}{c^2}\Psi\mid\nabla\Lambda\mid^2 - \frac{i\hbar q}{c}\Psi\nabla^2\Lambda - \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda\right].$$

Finally, we need to use (3) applied to Ψ' rather that Ψ to expand the first term on the right-hand side of (11). For this we will need $\nabla\Psi'$, which we have in (10), but also

$$\begin{aligned}
\nabla^2\Psi' &= \nabla\cdot(\nabla\Psi') = \nabla\cdot\left(e^{-(iq/\hbar c)\,\Lambda}\left(\nabla\Psi - \frac{iq}{\hbar c}\Psi\nabla\Lambda\right)\right) \\
&= \nabla\left(e^{-(iq/\hbar c)\,\Lambda}\right)\cdot\left(\nabla\Psi - \frac{iq}{\hbar c}\Psi\nabla\Lambda\right) + e^{-(iq/\hbar c)\,\Lambda}\,\nabla\cdot\left(\nabla\Psi - \frac{iq}{\hbar c}\Psi\nabla\Lambda\right) \\
&= -\frac{iq}{\hbar c}e^{-(iq/\hbar c)\,\Lambda}\,\nabla\Lambda\cdot\left(\nabla\Psi - \frac{iq}{\hbar c}\Psi\nabla\Lambda\right) + e^{-(iq/\hbar c)\,\Lambda}\left(\nabla^2\Psi - \frac{iq}{\hbar c}\nabla\cdot(\Psi\nabla\Lambda)\right) \\
&= e^{-(iq/\hbar c)\,\Lambda}\left(-\frac{iq}{\hbar c}\nabla\Lambda\cdot\nabla\Psi - \frac{q^2}{\hbar^2 c^2}\Psi\mid\nabla\Lambda\mid^2 + \nabla^2\Psi - \frac{iq}{\hbar c}\left(\Psi\nabla^2\Lambda + \nabla\Psi\cdot\nabla\Lambda\right)\right)
\end{aligned}$$

$$\nabla^2\Psi' = e^{-(iq/\hbar c)\,\Lambda}\left(\nabla^2\Psi - \frac{2iq}{\hbar c}\,\nabla\Psi\cdot\nabla\Lambda - \frac{q^2}{\hbar^2 c^2}\Psi\mid\nabla\Lambda\mid^2 + \nabla^2 - \frac{iq}{\hbar c}\Psi\nabla^2\Lambda\right). \tag{12}$$

Now notice that

$$\begin{aligned}
\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi &= (i\hbar)^2\left(\nabla - \frac{iq}{\hbar c}\vec{A}\right)^2\Psi \\
&= -\hbar^2\left(\nabla - \frac{iq}{\hbar c}\vec{A}\right)^2\Psi
\end{aligned}$$

so

$$\left(\nabla - \frac{iq}{\hbar c}\vec{A}\right)^2\Psi \;=\; -\frac{1}{\hbar^2}\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2\Psi\,.$$

Apply (3) with $\vec{A} = \nabla\Lambda$ to obtain

$$\left(\nabla - \frac{iq}{\hbar c}\nabla\Lambda\right)^2 = \nabla^2\Psi - \frac{iq}{\hbar c}\Psi\nabla^2\Lambda - \frac{2iq}{\hbar c}\nabla\Lambda\cdot\nabla\Psi - \frac{q^2}{\hbar^2 c^2}\Psi\mid\nabla\Lambda\mid^2 \tag{13}$$

so that (12) becomes

$$\nabla^2\Psi' = e^{-(iq/\hbar c)\,\Lambda}\left(\nabla - \frac{iq}{\hbar c}\,\nabla\Lambda\right)^2\Psi\,. \tag{14}$$

Now we can compute, from (3) applied to Ψ' with the aid of (14) and (10),

$$\left(i\hbar\nabla + \frac{q}{c}\vec{A}\right)^2 \Psi' = -\hbar^2\nabla^2\Psi' + \frac{i\hbar q}{c}\Psi'\nabla\cdot\vec{A} + \frac{2i\hbar q}{c}\vec{A}\cdot\nabla\Psi' + \frac{q^2}{c^2}\Psi' \mid \vec{A}\mid^2$$

$$= e^{-(iq/\hbar c)\,\Lambda}\left[-\hbar^2\left(\nabla - \frac{iq}{\hbar c}\nabla\Lambda\right)^2\Psi + \frac{i\hbar q}{c}\Psi\nabla\cdot\vec{A} + \frac{2i\hbar q}{c}\vec{A}\cdot\left(\nabla\Psi - \frac{iq}{\hbar c}\Psi\nabla\Lambda\right) + \frac{q^2}{c^2}\Psi \mid \vec{A}\mid^2\right]$$

$$= e^{-(iq/\hbar c)\,\Lambda}\left[-\hbar^2\left(\nabla - \frac{iq}{\hbar c}\nabla\Lambda\right)^2\Psi + \frac{i\hbar q}{c}\Psi\nabla\cdot\vec{A} + \frac{2i\hbar q}{c}\vec{A}\cdot\nabla\Psi + \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda + \frac{q^2}{c^2}\Psi \mid \vec{A}\mid^2\right]$$

and substitute this into (11) to obtain

$$\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 \Psi' = e^{-(iq/\hbar c)\,\Lambda}\left[-\hbar^2\left(\nabla - \frac{iq}{\hbar c}\nabla\Lambda\right)^2\Psi + \frac{i\hbar q}{c}\Psi\nabla\cdot\vec{A} + \frac{2i\hbar q}{c}\vec{A}\cdot\nabla\Psi + \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda + \frac{q^2}{c^2}\Psi \mid \vec{A}\mid^2 - \frac{2i\hbar q}{c}\nabla\Psi\cdot\nabla\Lambda - \frac{q^2}{c^2}\Psi \mid \nabla\Lambda\mid^2 - \frac{i\hbar q}{c}\Psi\nabla^2\Lambda - \frac{2q^2}{c^2}\Psi\vec{A}\cdot\nabla\Lambda\right].$$

But, by (13),

$$-\hbar^2\left(\nabla - \frac{iq}{\hbar c}\nabla\Lambda\right)^2\Psi = -\hbar^2\nabla^2\Psi + \frac{i\hbar q}{c}\Psi\nabla^2\Lambda + \frac{2i\hbar q}{c}\nabla\Psi\cdot\nabla\Lambda + \frac{q^2}{c^2}\Psi \mid \nabla\Lambda\mid^2$$

so this reduces to

$$\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 \Psi' = e^{-(iq/\hbar c)\,\Lambda}\left[-\hbar^2\nabla^2\Psi + \frac{i\hbar q}{c}\Psi\nabla\cdot\vec{A} + \frac{2i\hbar q}{c}\vec{A}\cdot\nabla\Psi + \frac{q^2}{c^2}\Psi \mid \vec{A}\mid^2\right]. \tag{15}$$

Finally, we use this to compute the right-hand side of (7).

$$\left(\frac{1}{2m} \left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 + q\left(V + \frac{1}{c}\frac{\partial\Lambda}{\partial t}\right)\right)^2 \Psi'$$

$$= \frac{1}{2m}\left(i\hbar\nabla + \frac{q}{c}\left(\vec{A} - \nabla\Lambda\right)\right)^2 \Psi' + qV\Psi' + \frac{q}{c}\frac{\partial\Lambda}{\partial t}\Psi'$$

$$= e^{-(iq/\hbar c)\,\Lambda}\left[-\frac{\hbar^2}{2m}\nabla^2\Psi + \frac{i\hbar q}{2mc}\Psi\nabla\cdot\vec{A} + \frac{i\hbar q}{mc}\vec{A}\cdot\nabla\Psi \right.$$

$$\left. + \frac{q^2}{2mc^2}\Psi \mid \vec{A}\mid^2 + qV\Psi + \frac{q}{c}\frac{\partial\Lambda}{\partial t}\Psi \right]$$

$$= e^{-(iq/\hbar c)\,\Lambda}\left[i\hbar\frac{\partial\Psi}{\partial t} + \frac{q}{c}\frac{\partial\Lambda}{\partial t}\Psi\right] \qquad (\text{ by } (4)\,)$$

$$= i\hbar\frac{\partial\Psi'}{\partial t} \qquad (\text{by } (8)\,)$$

and this is (7). To show that if Ψ' satisfies (7), then $\Psi = e^{(iq/\hbar c)\,\Lambda}\,\Psi'$ satisfies (2) one need only apply what we have just proved to the gauge transformation $\vec{A} - \nabla\Lambda \to (\vec{A} - \nabla\Lambda) + \nabla\Lambda,\ V + \frac{1}{c}\frac{\partial\Lambda}{\partial t} \to (V + \frac{1}{c}\frac{\partial\Lambda}{\partial t}) - \frac{1}{c}\frac{\partial\Lambda}{\partial t}$. ■

C. Phase Shift for Aharonov-Bohm

In this section we will calculate explicitly the phase shift of the electrons moving on opposite sides of the solenoid in the Aharonov-Bohm experiment (only basic vector calculus is required here, but we will also appeal to the theorem proved in Appendix B).

We consider a solenoid of radius R along the z-axis and the magnetic field

$$\vec{B}\,(\,x,\,y,\,z\,) \;=\; \begin{cases} \langle\,0,\,0,\,B\,\rangle & , \quad x^2 + y^2 \le R^2 \\ \vec{0} & , \quad x^2 + y^2 > R^2 \end{cases} \tag{1}$$

it determines on $\mathbb{R}^3$ (here B is a constant). There is no electric field, i.e., $\vec{E}(x,y,z) = \vec{0}$ for all (x,y,z). Notice that the flux of the magnetic field through the solenoid (in the positive z-direction) is

$$\Phi_0 \;=\; \iint\limits_{\substack{x^2+y^2 \le R^2 \\ z=0}} \vec{B}\cdot\vec{k}\,dS \;=\; B\left(\pi R^2\right)\,. \tag{2}$$

We will construct a magnetic vector potential $\vec{A}(x,y,z)$ for $\vec{B}(x,y,z)$ by piecing together potentials on $x^2+y^2 \le R^2$ and $x^2+y^2 > R^2$. Notice first that the vector field

$$\vec{F}\,(\,x,\,y,\,z\,)\;=\;\frac{B}{2}\,\langle\,-\,y,\,x,\,0\,\rangle$$

is smooth (infinitely differentiable component functions) on all of $\mathbb{R}^3$ and

$$\nabla\,\times\,\vec{F}\;=\;\langle\,0,\,0,\,B\,\rangle$$

on $\mathbb{R}^3$. We will let $\vec{A}_1(x,y,z)$ be the restriction of $\vec{F}$ to $x^2+y^2 \le R^2$.

On $x^2+y^2 > R^2$ the magnetic field is zero. However, we are not free to take the potential to be zero on this region since, in order to be continuous on $\mathbb{R}^3$, $\vec{A}(x,y,z)$ must approach $\vec{A}_1(x,y,z)$ as $x^2+y^2 \to R^2$ and $\vec{A}_1$ is not zero on $x^2+y^2 = R^2$. We consider the vector field

$$\vec{G}\,(\,x,\,y,\,z\,)\;=\;\frac{BR^2}{2}\left\langle\,-\,\frac{y}{x^2+y^2}\,,\,\frac{x}{x^2+y^2}\,,\,0\right\rangle\,.$$

This is smooth on all of $\mathbb{R}^3$ except the z-axis ($x^2+y^2 = 0$) and a quick calculation shows that, on this region,

$$\nabla\,\times\,\vec{G}\;=\;\langle\,0\,,\,0\,,\,0\,\rangle\,.$$

Furthermore, $\vec{G}(x,y,z) \to \vec{F}(x,y,z)$ as $x^2+y^2 \to R^2$ so we can take $\vec{A}_2(x,y,z)$ to be the restriction of $\vec{G}(x,y,z)$ to $x^2+y^2 > R^2$. Thus, we arrive at

$$\begin{aligned}\vec{A}\,(\,x,\,y,\,z\,) &= \begin{cases}\vec{A}_1\,(x,y,z)\;,\;x^2+y^2 \le R^2\\ \vec{A}_2\,(x,y,z)\;,\;x^2+y^2 > R^2\end{cases}\\ &= \begin{cases}\dfrac{B}{2}\,\langle -y,x,0\rangle & ,\quad x^2+y^2 \le R^2\\ \dfrac{BR^2}{2}\left\langle -\dfrac{y}{x^2+y^2}\,,\,\dfrac{x}{x^2+y^2}\,,\,0\right\rangle & ,\quad x^2+y^2 > R^2\,.\end{cases}\end{aligned} \tag{3}$$

Remark: $\vec{A}(x,y,z)$ is continuous on all of $\mathbb{R}^3$, but it is not smooth there. Indeed, the partial derivatives of its component functions have jump discontinuities on $x^2+y^2 = R^2$.

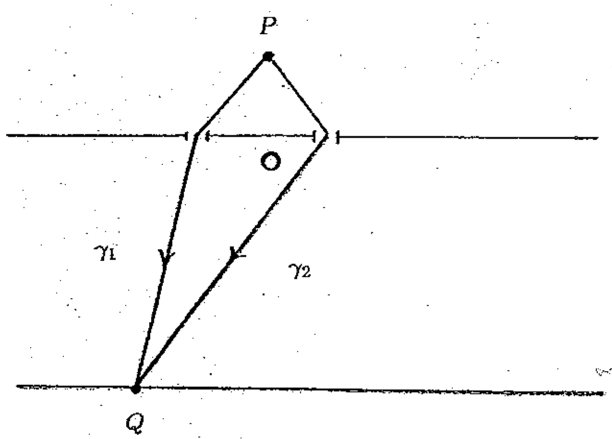
We will be interested in line integrals of $\vec{A}$ over various curves in the exterior of the solenoid (classical paths for the Aharonov-Bohm electrons). Notice that, on $x^2+y^2 > R^2$, $\vec{A}$ agrees with $\vec{G}$ and $\vec{G}$ is actually smooth with zero curl on all of $x^2+y^2 > 0$. Consequently, Stokes' Theorem implies that the integrals of $\vec{G}$ over all piecewise smooth, simple closed curves with the same orientation and surrounding the z-axis are the same. Thus, it suffices to integrate $\vec{G}$ over $x^2 + y^2 = R^2,\ z = 0$, so we take $\gamma(t) = (R\cos t,\ R\sin t,\ 0),\ 0 \le t \le 2\pi$. Then

$$\begin{aligned}\oint_\gamma \vec{G}\cdot d\vec{r} &= \int_0^{2\pi} \frac{BR^2}{2}\left\langle -\frac{R\sin t}{R^2},\ \frac{R\cos t}{R^2},\ 0\right\rangle \cdot \langle -R\sin t,\ R\cos t,\ 0\rangle\, dt \\ &= \frac{BR^2}{2}\int_0^{2\pi} dt \\ &= B\left(\pi R^2\right) = \Phi_0\ .\end{aligned}$$

Consequently, if γ is any closed curve outside, but surrounding the solenoid and oriented counterclockwise as seen from above, then

$$\oint_\gamma \vec{A}\cdot d\vec{r} = B\left(\pi R^2\right) = \Phi_0\ . \tag{4}$$

In particular, if we have two curves γ_1 and γ_2 as shown below on opposite sides of the solenoid joining a point P (say, at the source of the Aharonov-Bohm electrons) and a point Q (say, at the screen)



then

$$\oint_{\gamma_1+(-\gamma_2)} \vec{A}\cdot d\vec{r} = \Phi_0$$

so that

$$\int_{\gamma_1} \vec{A}\cdot d\vec{r} = \int_{\gamma_2} \vec{A}\cdot d\vec{r} + \Phi_0\ . \tag{5}$$

To relate all of this to phase shifts of wave functions we will introduce cylindrical coordinates (r, θ, z) in space. We will need to be careful about the polar angle

θ, however, since this cannot be defined globally on $\mathbb{R}^3$. For any (x, y, z) in space we define $r = \sqrt{x^2 + y^2}$ and, of course, z is the same in cylindrical coordinates. Now define

$$\theta_1(x, y, z) = \begin{cases} \arctan\left(\frac{y}{x}\right) & , \quad x > 0,\ y > 0 \\ \dfrac{\pi}{2} & , \quad x = 0,\ y > 0 \\ \pi + \arctan\left(\frac{y}{x}\right) & , \quad x < 0 \\ \dfrac{3\pi}{2} & , \quad x = 0,\ y < 0 \\ 2\pi + \arctan\left(\frac{y}{x}\right) & , \quad x > 0,\ y < 0 \end{cases} .$$

Then $0 < \theta_1 < 2\pi$ and $\theta_1(x, y, z)$ is discontinuous on the half-plane $x \geq 0$, $y = 0$, but smooth on its complement U_1. Moreover, on U_1, $\nabla\theta_1(x, y, z) = \vec{G}(x, y, z)$. Similarly, we define

$$\theta_2(x, y, z) = \begin{cases} \arctan\left(\frac{y}{x}\right) & , \quad x > 0 \\ \dfrac{\pi}{2} & , \quad x = 0,\ y > 0 \\ \pi + \arctan\left(\frac{y}{x}\right) & , \quad x < 0,\ y > 0 \\ -\pi + \arctan\left(\frac{y}{x}\right) & , \quad x < 0,\ y < 0 \\ -\dfrac{\pi}{2} & , \quad x = 0,\ y < 0 \end{cases} .$$

Then $-\pi < \theta_2 < \pi$ and $\theta_2(x, y, z)$ is discontinuous on the half-plane $x \leq 0$, $y = 0$, but smooth on its complement U_2. Moreover, on U_2, $\nabla\theta_2(x, y, z) = \vec{G}(x, y, z)$. Thus, on the intersection of U_i with $x^2 + y^2 > R^2$,

$$\nabla\left(\frac{BR^2}{2}\,\theta_i\,(\,x,\ y,\ z\,)\right) = \vec{A}\,(\,x,\ y,\ z\,) \tag{6}$$

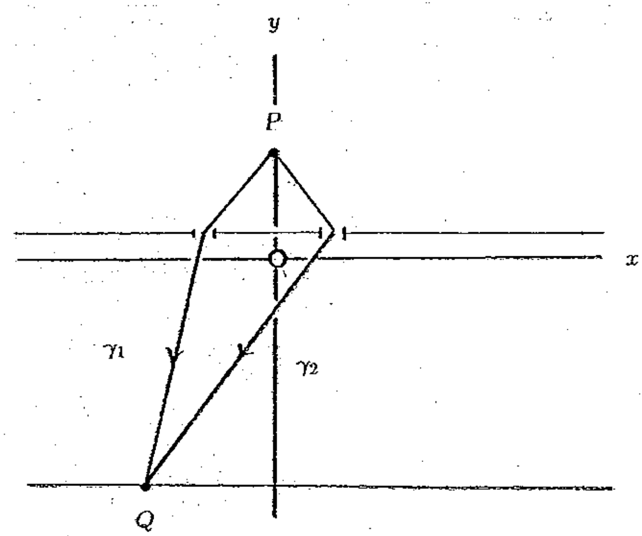
for $i = 1, 2$. Consequently, if γ_i is a curve in U_i,

$$\begin{aligned} \int_{\gamma_i} \vec{A} \cdot d\vec{r} &= \int_{\gamma_i} \nabla\left(\frac{BR^2}{2}\,\theta_i\right) \cdot d\vec{r} \\ &= \text{change in } \frac{BR^2}{2}\,\theta_i \text{ over } \gamma_i\ . \end{aligned}$$

For future use we insert a few extra constants and write “change over γ_i” as Δ_{γ_i} .

$$\frac{q}{\hbar c}\int_{\gamma_i} \vec{A} \cdot d\vec{r} = \Delta_{\gamma_i}\left(\frac{q}{\hbar c}\,\frac{BR^2}{2}\,\theta_i\right) = \Delta_{\gamma_i}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_i\right) \tag{7}$$

Notice that, if $y > 0$, $\theta_2 = \theta_1$ and, if $y < 0$, $\theta_2 = \theta_1 - 2\pi$, Now suppose γ_1 and γ_2 are two curves from P (source) to Q (screen) on opposite sides of the solenoid as shown below.



Then γ_1 remains in U_1 and γ_2 remains in U_2 and P and Q are both in $U_1 \cap U_2$. Thus,

$$\begin{aligned}\Delta_{\gamma_2}(\theta_2) &= \theta_2(Q) - \theta_2(P) = \theta_1(Q) - 2\pi - \theta_1(P) \\ &= \Delta_{\gamma_1}(\theta_1) - 2\pi\end{aligned}$$

so

$$\Delta_{\gamma_2}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_2\right) = \Delta_{\gamma_1}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_1\right) - \left(\frac{q\Phi_0}{2\pi\hbar c}\right)(2\pi)$$

$$\Delta_{\gamma_1}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_1\right) - \Delta_{\gamma_2}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_2\right) = \frac{q\Phi_0}{\hbar c}\,. \tag{8}$$

Now we return to the Aharonov-Bohm experiment and show that (8) describes the shift in the interference pattern at the screen due to the flux Φ_0 in the solenoid. We will be interested in the probability amplitude ψ for an electron emitted at P to be detected at Q on the screen. Classically, an electron can reach Q from P along a route through either Slit 1 or Slit 2 and each of these has an associated probability amplitude; if we call these ψ_1 and ψ_2, then $\psi = \psi_1 + \psi_2$ (see Appendix A).

Remark: We should admit at the outset that the argument we are about to give is only an approximation to an honest quantum mechanical calculation. Electrons do not move on classical paths (remember that they "travel like waves" between the source and the screen). The rules of the game require that one take into account *all possible classical paths* from P through the slits to Q and one does this by computing a *Feynman path integral*. However, if you would like to see this the best I can do for you is to provide a reference (Kobe, D.H., *The Aharonov-Bohm Effect Revisited*, Annals of Physics 123, 1979, 381-410).

Now let us suppose first that the current in the solenoid is turned off. Then $\vec{B} = \vec{0}$ everywhere so we can take $\vec{A} = \vec{0}$ everywhere ($V = 0$ also since there is no electric field). The solution to the corresponding Schrödinger equation is called the free particle wave function and is denoted $\psi_0(\vec{x}, t)$.

Now turn the current on. Outside the solenoid $\vec{B}$ is still $\vec{0}$ so $\vec{A}$ must be a gauge transformation of $\vec{0}$. Specifically, if we let θ denote θ_1 on U_1 or θ_2 on U_2, then, by (6), the gauge transformation must be

$$\vec{0} \longrightarrow \vec{0} - \nabla\left(-\frac{BR^2}{2}\,\theta \right) \quad = \nabla\left(\frac{BR^2}{2}\,\theta \right)$$

$$0 \longrightarrow 0 + \frac{1}{c}\frac{\partial}{\partial t}\left(-\frac{BR^2}{2}\,\theta \right) = 0\,.$$

Now, according to our general result on gauge transformations of the Schrödinger equation in Appendix B the new wave functions on U_1 and U_2 are

$$\psi_i\,(\,\vec{x},\,t\,) = e^{-(iq/\hbar c)\left(-\frac{BR^2}{2}\,\theta_i \right)}\,\psi_0\,(\,\vec{x},\,t\,)\ \ ,\ \ i = 1, 2\,,$$

$$\psi_i\,(\,\vec{x},\,t\,) = e^{i\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_i \right)}\,\psi_0\,(\,\vec{x},\,t\,)\ \ ,\ \ i = 1, 2\,. \tag{9}$$

Now consider an electron emitted at P and detected at Q. We will show that the difference in the phase associated to the electron at Q by ψ_1 and ψ_2 depends on the flux Φ_0 of the magnetic field through the solenoid and, indeed, is given by (8). This not only accounts for the Aharonov-Bohm effect, but, since the phase shift $q\Phi_0\,/\,\hbar c$ is independent of Q, shows that the entire interference pattern at the screen will be shifted by the same amount $q\Phi_0\,/\,\hbar c$.

To do the arithmetic we will write $\omega_i(P)$ and $\omega_i(Q)$, $i = 0, 1, 2$, to denote the phase associated to P at the emission and to Q at the detection by ψ_i, $i = 0, 1, 2$. We want to compute $\omega_1(Q) - \omega_2(Q)$. By (9),

$$\omega_i\,(\,P\,) \;=\; \omega_0\,(\,P\,) + \frac{q\Phi_0}{2\pi\hbar c}\,\theta_i\,(\,P\,)\ \ ,\ \ i = 1, 2$$

and

$$\omega_i\,(\,Q\,) \;=\; \omega_0\,(\,Q\,) + \frac{q\Phi_0}{2\pi\hbar c}\,\theta_i\,(\,Q\,)\ \ ,\ \ i = 1, 2\,.$$

Thus,

$$\omega_1(Q) - \omega_2(Q) = \frac{q\Phi_0}{2\pi\hbar c}\,(\theta_1(Q) - \theta_2(Q))$$
$$= \frac{q\Phi_0}{2\pi\hbar c}\Big(\,(\theta_1(Q) - \theta_1(P)) - (\theta_2(Q) - \theta_2(P))\Big)$$

since $\theta_2(P) = \theta_1(P)$. Since γ_1 and γ_2 both begin at P and end at Q, this can be written

$$\omega_1(Q) - \omega_2(Q) = \Delta_{\gamma_1}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_1\right) - \Delta_{\gamma_2}\left(\frac{q\Phi_0}{2\pi\hbar c}\,\theta_2\right)$$

which is the left-hand side of (8). We conclude that a magnetic flux of Φ_0 through the solenoid shifts the free particle interference pattern on the screen by $q\Phi_0 / \hbar c$, as required.

D. Fiber Bundle Interpretation of Aharonov-Bohm

In this final section we will describe the geometrical view of the Aharonov-Bohm effect as the holonomy of a certain flat connection on a principal $U(1)$-bundle. We will have to assume a basic familiarity with manifolds, differential forms, principal bundles and connections, but the particularly simple context in which we find ourselves will greatly simplify much of the general theory and will permit us to do explicit and relatively straightforward calculations. Let me first describe precisely why the context in which we find ourselves is particularly simple.

1. The structure group is $U(1)$. This is an Abelian Lie group and its Lie algebra $u(1)$ is identified with the pure imaginary complex numbers $i\mathbb{R}$. As a result, all brackets in $u(1)$ are zero, the adjoint action of $U(1)$ on $u(1)$ is trivial and $u(1)$-valued differential forms are just ordinary real-valued differential forms with an extra factor of i.

2. The base of our principal bundle is $\mathbb{R}^3$ minus a closed, solid cylinder which is homotopically equivalent to S^1 and the classification, up to equivalence, of principal bundles over a space depends only on the homotopy type of that space. Moreover, there is, up to equivalence, only one principal $U(1)$-bundle over S^1, namely, the trivial (product) bundle. Consequently, our bundle is just the trivial $U(1)$-bundle over $\mathbb{R}^3$ minus a cylinder.

We will, in fact, allow ourselves two more physically motivated and mathematically innocuous simplifications. We will assume that, in comparison with the dimensions of the Aharonov-Bohm experiment, the solenoid is "very long" so that, by symmetry, we can restrict our attention to classical paths in the plane $\mathbb{R}^2$ minus a disc rather than $\mathbb{R}^3$ minus a cylinder (note that $\mathbb{R}^2$ minus a disc is still homotopically equivalent to S^1). Furthermore, in our discussion of the phase shift in Appendix C, the only role played by the radius R of the cylinder was to tell us the flux $\Phi_0 = B(\pi R^2)$ of the magnetic field through the solenoid. We will therefore assume that our solenoid is also "very thin" and can be identified with a line carrying a magnetic flux of Φ_0. Together these two assumptions amount to assuming that the Aharonov-Bohm electrons move in $\mathbb{R}^2$ minus a point (still homotopically equivalent to S^1).

With this we can begin building our geometrical model; first, the principal bundle itself. The *base space* is

$$M = \mathbb{R}^2 - \{ (0, 0) \} .$$

This is where the electrons move. The *structure group* is

$$G = U(1) .$$

This is the "notebook" in which we record phases. The *bundle space* is

$$P = M \times G$$

and the *projection*

$$\pi_1 : P \longrightarrow M$$

is just the projection onto the first factor, i.e., if $p = (m, g) \in P$, then $\pi_1(p) = m$. The *right action* of G on P is just multiplication in the second coordinate, i.e.,

$$\sigma : P \times G \longrightarrow P$$

is defined by

$$\sigma(p, h) = p \cdot h = (m, g) \cdot h = (m, gh) .$$

What we have just described is the structure of the *trivial* (*product*) $U(1)$-*bundle over* $\mathbb{R}^2 - \{ (0,0) \}$.

In order to introduce connections on $\pi_1 : P \to G$ we will need to briefly review some basic tools and terminology.

> **WARNING:** I will not hesitate to use the very special structure of our bundle to simplify matters without further comment so what follows *cannot* be regarded as an introduction to the general theory of connections on principal bundles.

For each $h \in G$ we have a diffeomorphism

$$\sigma_h : P \longrightarrow P$$

of P onto itself defined by

$$\sigma_h(p) = p \cdot h = (m, g) \cdot h = (m, gh) .$$

Thus, σ_h rotates each fiber $\pi_1^{-1}(m)$ by h. Note that we can write

$$\sigma_h(p) = \sigma_h(m, g) = (m, gh) = \big(m, R_h(g)\big), \tag{1}$$

where $R_h : G \to G$ is right multiplication by h. Similarly, for each $p = (m, g) \in P$ we have a map

$$\sigma_p : G \longrightarrow P$$

defined by

$$\sigma_p(h) = p \cdot h = (m, g) \cdot h = (m, gh) = \big(m, L_g(h)\big), \tag{2}$$

where $L_g : G \to G$ is left multiplication by g.

A (*global*) *section* (or *gauge*) for $\pi_1 : P \to G$ is a smooth selection of a point in each fiber $\pi_1^{-1}(m)$, $m \in M$, i.e., it is a smooth map

$$s : M \longrightarrow P$$

satisfying

$$\pi_1 \circ s = id_M .$$

Every such section has the form

$$s(m) = \big(m, g(m)\big) \tag{3}$$

for some smooth map $g : M \to G$ and every such g can be written as $g(m) = e^{i\theta(m)}$ for some smooth map $\theta : M \to \mathbb{R}$. The *canonical section* s_0 is just $s_0(m) = (m, 1) = (m, e^{i0})$. Any section s gives rise to a diffeomorphism

$$\varphi_s : P \longrightarrow P$$

defined by

$$\varphi_s(p) = \varphi_s(m, g) = s(m) \cdot g = \big(m, g(m)g\big) \tag{4}$$

which preserves the action of G on P ($\varphi_s(p \cdot g_0) = \varphi_s(p) \cdot g_0$) and restricts to an automorphism on each fiber $\pi_1^{-1}(m)$, i.e., φ_s is a *bundle isomorphism*. Notice that if $s_0 : M \to P$ is the canonical section, then

$$\varphi_s \circ s_0 = s . \tag{5}$$

Notice also that if $s_1(m) = (m, g_1(m))$ and $s_2(m) = (m, g_2(m))$ are two sections and if we define

$$g_{ij} : M \longrightarrow G$$

for $i, j = 1, 2$ by

$$g_{ij}(m) = \big(g_i(m)\big)^{-1} g_j(m) \tag{6}$$

then

$$s_j(m) = s_i(m) \cdot g_{ij}(m) \ , \ i,j = 1,2\,. \tag{7}$$

The maps g_{ij} are called the *transition functions* relating the two sections and we also have

$$\begin{aligned}
\varphi_{s_j}(p) = \varphi_{s_j}(m, g) &= s_j(m) \cdot g = \big(s_i(m) \cdot g_{ij}(m) \big) \cdot g \\
&= s_i(m) \cdot \big(g_{ij}(m)\, g \big) \\
&= s_i(m) \cdot \big(g\, g_{ij}(m) \big) \qquad (\text{because } G \text{ is Abelian}) \\
&= \big(s_i(m) \cdot g \big) \cdot g_{ij}(m)
\end{aligned}$$

$$\varphi_{s_j}(p) = \varphi_{s_i}(p) \cdot g_{ij}(m)\,. \tag{8}$$

We will need derivatives for the maps we have introduced so we will first describe convenient identifications of the tangent spaces. For $M = \mathbb{R}^2 - \{ (0,0) \}$ we fix an $m = (x,y) \in M$. Any element of $T_m(M)$ is the tangent vector at $t = 0$ to a curve

$$\alpha : I \longrightarrow M$$

$$\alpha(t) = \big(At + x\,,\, Bt + y \big)\,, \tag{9}$$

where $I \subseteq \mathbb{R}$ is an interval and $A, B \in \mathbb{R}$. We will write such tangent vectors

$$v_m = \alpha'(0) = (A\,,\,B)_m\,. \tag{10}$$

The elements of $U(1)$ are 1×1 unitary matrices and so can be identified with complex numbers $g = e^{i\theta}$ of modulus one, i.e., points on the unit circle S^1. Fixing a $g = e^{i\theta}$ in $U(1)$, any element of $T_g(G)$ is the tangent vector at $t = 0$ to a curve

$$\beta : I \longrightarrow G$$

$$\beta(t) = e^{i(Ct + \theta)}\,, \tag{11}$$

where $I \subseteq \mathbb{R}$ is an interval and $C \in \mathbb{R}$. Thus, the elements of $T_g(G)$ can be identified with the complex numbers

$$v_g = \beta'(0) = iCe^{i\theta} = (iC)\,g\,. \tag{12}$$

In particular, if $g = 1$ the tangent space to G at the identity $1 \in G$ is just the pure imaginary complex numbers and this is the Lie algebra $\mathcal{G}$ of G.

$$\mathcal{G} = i\,\mathbb{R} \tag{13}$$

The tangent space to P at $p = (m,g)$ is then identified with

$$T_p(P) = T_{(m,g)}(M \times G) = T_m(M) \times T_g(G)\,. \tag{14}$$

With this we can now compute derivatives of the various maps introduced above. I will record the results we need and prove one of them to remind you how it goes. We will write m, g, and $p=(m,g)$ for points in M, G, and P, respectively. For any $h \in G$,

$$\sigma_h \,:\, P \longrightarrow P$$

$$\sigma_h(m,g) = (m, gh) = \big(m, R_h(g)\big)$$

$$(\sigma_h)_{*p} \,:\, T_m(M) \times T_g(G) \longrightarrow T_m(M) \times T_{gh}(G)$$

$$(\sigma_h)_{*p}\big(v_m, v_g\big) = \Big(v_m, (R_h)_{*g}\big(v_g\big)\Big) = \big(v_m, v_g h\big) \qquad (15)$$

To see this we note that $\sigma_h = \mathrm{id}_M \times R_h$, $(\mathrm{id}_M)_{*m} = \mathrm{id}_{T_m(M)}$ and

$$\begin{aligned}(R_h)_{*g}\big(v_g\big) &= (R_h)_{*g}\big(\beta'(0)\big)\\ &= (R_h \circ \beta)'(0)\,.\end{aligned}$$

But

$$(R_h \circ \beta)(t) = R_h\big(\beta(t)\big) = \beta(t)\,h$$

so

$$(R_h \circ \beta)'(0) = \beta'(0)\,h = v_g\,h\,.$$

Similarly, one obtains the following derivatives.

$$\sigma_p \,:\, G \longrightarrow P$$

$$\sigma_p(h) = p \cdot h = (m,g)\cdot h = (m, gh) = \big(m, L_g(h)\big)$$

$$\big(\sigma_p\big)_{*g_0} \,:\, T_{g_0}(G) \longrightarrow T_m(M) \times T_{gg_0}(G)$$

$$\big(\sigma_p\big)_{*g_0}\big(v_{g_0}\big) = \Big(0, \big(L_g\big)_{*g_0}\big(v_{g_0}\big)\Big) = \big(0, g\,v_{g_0}\big)\,. \qquad (16)$$

$$s \,:\, M \longrightarrow P \qquad (\pi_1 \circ s = \mathrm{id}_M)$$

$$s(m) = \big(m, g(m)\big) \qquad (g : M \longrightarrow G)$$

$$s_{*m} \,:\, T_m(M) \longrightarrow T_{s(m)}(P) = T_m(M) \times T_{g(m)}(G)$$

$$s_{*m}\,(v_m) = \big(v_m\,,\, g_{*m}(v_m)\big)$$

$$= \left(v_m\,,\, ie^{i\theta(m)} \left(A\,\frac{\partial\theta}{\partial x}\,(m) + B\,\frac{\partial\theta}{\partial y}\,(m) \right) \right)\,, \qquad (17)$$

where $g(m) = e^{i\theta(m)}$ and $v_m = (A\,,\,B)_m$.

In particular, for the canonical section $s_0(m) = (m, 1)$,

$$(s_0)_{*m}\,(v_m) = (v_m\,,\,0)\ . \qquad (18)$$

$$\pi_1 : P \longrightarrow M$$

$$\pi_1(p) = \pi_1(m, g) = m$$

$$(\pi_1)_{*p} : T_m(M) \times T_g(G) \longrightarrow T_m(M)$$

$$(\pi_1)_{*p}\,\big(v_p\big) = (\pi_1)_{*(m,g)}\,\big(v_m\,,\,v_g\big) = v_m \qquad (19)$$

$$\pi_2 : P \longrightarrow G$$

$$\pi_2(p) = \pi_2(m, g) = g$$

$$(\pi_2)_{*p} : T_m(M) \times T_g(G) \longrightarrow T_g(G)$$

$$(\pi_2)_{*p}\,\big(v_p\big) = (\pi_2)_{*(m,g)}\,\big(v_m\,,\,v_g\big) = v_g \qquad (20)$$

We have seen that the Lie algebra $\mathcal{G}$ of G can be identified with the pure imaginary complex numbers $i\mathbb{R}$. Consequently, the ordinary exponential map

$$\exp : \mathcal{G} \longrightarrow G$$

$$\exp(i\theta) = e^{i\theta}$$

is surjective. For any fixed $\xi = i\theta$ in $\mathcal{G}$,

$$t \longrightarrow \exp(t\xi) = e^{it\theta}$$

is a curve in G that passes through 1 at $t = 0$ with tangent vector ξ. Thus, for any $p = (m, g) \in P$,

$$t \longrightarrow p \cdot \exp(t\xi) = \big(m\,,\,ge^{it\theta}\big)$$

is a curve in the fiber $\pi_1^{-1}(m)$ through p in P. Its tangent vector is therefore an element of $T_p(\,\pi_1^{-1}(m)\,) \subseteq T_p(P)$.

> **Note:** $T_p(\,\pi_1^{-1}(m)\,)$ is called the *vertical space* at $p = (m, g)$ and denoted $\mathrm{Vert}_p\,(P)$.

Doing this for each $p \in P$ we obtain a vector field $\xi^{\#}$ on P defined by

$$\xi^{\#}\,(\,p\,) \;=\; \frac{d}{dt}\,\big(\,p \cdot \exp(\,t\xi\,)\,\big)\Big|_{t=0} \;=\; \big(\,\sigma_p\,\big)_{*1}\,(\,\xi\,) \tag{21}$$

and called the *fundamental vector field* on P determined by $\xi \in \mathcal{G}$. For each fixed p, $\xi \to \xi^{\#}(p)$ is an isomorphism of $\mathcal{G}$ onto $\mathrm{Vert}_p\,(P)$.

A *Lie algebra-valued* 1*-form* $\boldsymbol{\omega}$ on P is a 1-form on P with values in $\mathcal{G} = i\mathbb{R}$ and each of these can be written uniquely as

$$\boldsymbol{\omega} \;=\; i\,\omega\;,$$

where ω is an ordinary real-valued 1-form on P. Given a section $s : M \to P$, $s(m) = (\,m,\, g(m)\,)$, we can pull $\boldsymbol{\omega}$ back to a Lie algebra-valued 1-form $s^*\boldsymbol{\omega}$ on M defined by

$$(\,s^*\boldsymbol{\omega}\,)_m\,(\,v_m\,) \;=\; \boldsymbol{\omega}_{s(m)}\,\big(\,s_{*m}\,(\,v_m\,)\,\big)\;. \tag{22}$$

For any $h \in G$ we can also pull $\boldsymbol{\omega}$ back by the diffeomorphism $\sigma_h : P \to P$ to obtain another Lie algebra-valued 1-form $\sigma_h^*\boldsymbol{\omega}$ on P defined by

$$(\,\sigma_h^*\boldsymbol{\omega}\,)_p\,\big(\,v_p\,\big) \;=\; (\,\sigma_h^*\boldsymbol{\omega}\,)_{(m,g)}\,\big(\,v_m\,,\,v_g\,\big) \;=\; \boldsymbol{\omega}_{(m,gh)}\,\big(\,v_m\,,\,v_g\,h\,\big)\;. \tag{23}$$

With all of this behind us we can now at last define the objects of interest to us. A *connection* (or *gauge field*) on the trivial $U(1)$-bundle $\pi_1 : P \to M$ over $M = \mathbb{R}^2 - \{\,(0,0)\,\}$ is a Lie algebra-valued 1-form $\boldsymbol{\omega}$ on P which satisfies the following two conditions.

(a) $\sigma_h^*\boldsymbol{\omega} = \boldsymbol{\omega}$ for every $h \in G$

and

(b) $\boldsymbol{\omega}(\,\xi^{\#}\,) = \xi$ for every $\xi \in \mathcal{G}$.

Let's write these two conditions out in more detail.

(a) For all $h \in G$, $(m, g) \in P$ and $(v_m, v_g) \in T_m(M) \times T_g(G)$, $(\sigma_h^*\boldsymbol{\omega})_{(m,g)}\,(v_m, v_g) = \boldsymbol{\omega}_{(m,g)}\,(v_m, v_g)$, i.e., by (23),

$$\boldsymbol{\omega}_{(m,gh)}\,\big(\,v_m\,,\,v_g\,h\,\big) \;=\; \boldsymbol{\omega}_{(m,g)}\,\big(\,v_m\,,\,v_g\,\big)\;. \tag{24}$$

(b) For all $\xi \in \mathcal{G}$ and $p = (m, g) \in P$,

$$\boldsymbol{\omega}_p \left(\xi^{\#} (p) \right) = \boldsymbol{\omega}_p \Big((\sigma_p)_{*1} (\xi) \Big) = \xi ,$$

i.e., by (16),

$$\boldsymbol{\omega}_{(m,g)} (0 , g \xi) = \xi . \tag{25}$$

Example: We begin by introducing a Lie algebra-valued 1-form $\boldsymbol{\omega}_0$ on G called the *Cartan* 1-*form*. Fix a $g \in G$. Notice that $L_{g^{-1}}$ carries g to $L_{g^{-1}}(g) = g^{-1} g = 1$ so $(L_{g^{-1}})_{*g} : T_g(G) \to T_1(G) = \mathcal{G}$. We define $\boldsymbol{\omega}_0$ at each $g \in G$ by

$$(\boldsymbol{\omega}_0)_g \left(v_g \right) = \Big(L_{g^{-1}} \Big)_{*g} \left(v_g \right) = g^{-1} v_g . \tag{26}$$

Now pull this back to P by $\pi_2 : P = M \times G \to G$ to define

$$\boldsymbol{\omega}_{MC} = \pi_2^* \boldsymbol{\omega}_0 .$$

$\boldsymbol{\omega}_{MC}$ is called the *Maurer-Cartan* (or *canonical flat*) *connection* on the trivial bundle $\pi_1 : M \times G \to M$. Thus, for each $p = (m, g) \in P$ and $v_p = (v_m, v_g) \in T_p(P)$,

$$\begin{aligned} (\boldsymbol{\omega}_{MC})_p \left(v_p \right) &= (\pi_2^* \boldsymbol{\omega}_0)_p \left(v_p \right) = (\pi_2^* \boldsymbol{\omega}_0)_{(m,g)} \left(v_m , v_g \right) \\ &= (\boldsymbol{\omega}_0)_{\pi_2 (m,g)} \Big((\pi_2)_{*(m,g)} \left(v_m , v_g \right) \Big) \\ &= (\boldsymbol{\omega}_0)_g \left(v_g \right) = g^{-1} v_g . \end{aligned}$$

$$(\boldsymbol{\omega}_{MC})_p \left(v_p \right) = (\boldsymbol{\omega}_{MC})_{(m,g)} \left(v_m , v_g \right) = g^{-1} v_g . \tag{27}$$

We will check that this is a connection on $\pi_1 : P \to G$. For part (a) of the definition ($\sigma_h^* \boldsymbol{\omega} = \boldsymbol{\omega}$) we fix $h \in G$, $(m, g) \in P$ and $(v_m, v_g) \in T_m(M) \times T_g(G)$ and compute

$$\begin{aligned} (\sigma_h^* \boldsymbol{\omega}_{MC})_{(m,g)} \left(v_m , v_g \right) &= (\boldsymbol{\omega}_{MC})_{(m,gh)} \Big((\sigma_h)_{(m,g)} \left(v_m , v_g \right) \Big) \\ &= (\boldsymbol{\omega}_{MC})_{(m,gh)} \left(v_m , v_g h \right) \qquad \text{by (15)} \\ &= (gh)^{-1} \left(v_g h \right) \\ &= h^{-1} \left(g^{-1} v_g \right) h \\ &= g^{-1} v_g \qquad \textit{because } G \textit{ is Abelian} \\ &= (\boldsymbol{\omega}_{MC})_{(m,g)} \left(v_m , v_g \right) \end{aligned}$$

as required. For part (b) of the definition ($\boldsymbol{\omega}(\xi^{\#}) = \xi$) we fix $\xi \in \mathcal{G}$ and $p = (m, g) \in P$ and compute

$$
\begin{aligned}
(\boldsymbol{\omega}_{MC})_p\left(\xi^{\#}(p)\right) &= (\boldsymbol{\omega}_{MC})_p\Big(\left(\sigma_p\right)_{*1}(\xi)\Big) && \text{by (21)} \\
&= (\boldsymbol{\omega}_{MC})_{(m,g)}(0\,,\,g\,\xi) && \text{by (16)} \\
&= g^{-1}(g\,\xi) \\
&= \xi
\end{aligned}
$$

as required. Thus, we have an example of a connection on $\pi_1 :\ P \to G$. Shortly, we will discuss the reason it is referred to as "flat", but first we would like to take a somewhat different view of it.

Notice that $\boldsymbol{\omega}_{MC}$ is just the Cartan 1-form $\boldsymbol{\omega}_0$ on G "lifted" to $P = M \times G$ and that its value at any $v_p = (v_m, v_g)$ is independent of v_m. We begin, therefore, by looking at $\boldsymbol{\omega}_0$. This is an $i\mathbb{R}$-valued 1-form on $U(1)$ and $U(1)$ can be identified with the unit circle S^1 in $\mathbb{R}^2$. We would like to view $\boldsymbol{\omega}_0$ as the restriction to S^1 of an $i\mathbb{R}$-valued 1-form on $M = \mathbb{R}^2 - \{(0,0)\}$. Consider the 1-form $\boldsymbol{\gamma}$ on $\mathbb{R}^2 - \{(0,0)\}$ defined by

$$
\boldsymbol{\gamma} \;=\; i\left(\frac{-y}{x^2+y^2}\,dx \;+\; \frac{x}{x^2+y^2}\,dy\right)\,.
$$

At any $g = (g_1, g_2) \in S^1 \subseteq \mathbb{R}^2 - \{(0,0)\}$ we have seen ((12)) that every tangent vector v_g to S^1 at g can be written as

$$
\begin{aligned}
v_g = v_{(g_1,g_2)} &= iCg \;=\; iC\,(\,g_1 \;+\; g_2\,i\,) \;=\; -Cg_2 \;+\; Cg_1\,i \\
&= -Cg_2\,\frac{\partial}{\partial x}\bigg|_g \;+\; Cg_1\,\frac{\partial}{\partial y}\bigg|_g
\end{aligned}
$$

for some real number C. Thus, regarding $g = (g_1, g_2)$ as an element of $\mathbb{R}^2 - \{(0,0)\}$,

$$
\boldsymbol{\gamma}\,(\,v_g\,) \;=\; i\left(\frac{-g_2}{g_1^2+g_2^2}\,(\,-Cg_2\,) \;+\; \frac{g_1}{g_1^2+g_2^2}\,(\,Cg_1\,)\right) \;=\; iC\,.
$$

On the other hand,

$$
(\,\boldsymbol{\omega}_0\,)_g\,(\,v_g\,) \;=\; g^{-1}\,v_g \;=\; (\,g_1 \;-\; g_2\,i\,)\,(\,-Cg_2 \;+\; Cg_1\,i\,) \;=\; iC
$$

so $\boldsymbol{\omega}_0$ is just the restriction to S^1 of $\boldsymbol{\gamma}$. More precisely,

$$
\boldsymbol{\omega}_0 \;=\; \iota^*\,\boldsymbol{\gamma}
$$

where $\iota : S^1 \hookrightarrow M$ is the inclusion map. In particular, we can compute

$$d\boldsymbol{\omega}_0 = d(\iota^* \boldsymbol{\gamma}) = \iota^*(d\boldsymbol{\gamma})$$

and a simple coordinate calculation shows that $d\boldsymbol{\gamma} = 0$ so

$$d\boldsymbol{\omega}_0 = 0 .$$

As a result,

$$d\boldsymbol{\omega}_{MC} = d(\pi_2^* \boldsymbol{\omega}_0) = \pi_2^*(d\boldsymbol{\omega}_0) = 0 .$$

Now, any connection $\boldsymbol{\omega}$ on the trivial $U(1)$-bundle over M has associated with it a Lie algebra-valued 2-form on P called the *curvature* of $\boldsymbol{\omega}$, denoted $\boldsymbol{\Omega}$ and defined by

$$\boldsymbol{\Omega} = d\boldsymbol{\omega}$$

> **WARNING:** This is the correct definition of curvature *only* because $U(1)$ is Abelian.

Thus, $\boldsymbol{\omega}_{MC}$ is "flat" because its curvature is zero.

Note: The 1-form $\boldsymbol{\gamma}$ should look familiar. Except for a constant, it is precisely the 1-form on $\mathbb{R}^2 - \{(0,0)\}$ corresponding to the vector field $\vec{G}$ with which we represented (in Appendix C) the magnetic vector potential outside a solenoid. The fact that the corresponding connection $\boldsymbol{\omega}_{MC} = \pi_2^*(\iota^* \boldsymbol{\gamma}) = (\iota \circ \pi_2)^* \boldsymbol{\gamma}$ on P is flat is just another way of saying that this vector potential gives rise to a magnetic field that is $\vec{0}$ everywhere on its domain.

Notice that, if $\boldsymbol{\omega}$ is any connection on $\pi_1 : P \to M$, then we can pull back $\boldsymbol{\omega}$ by the canonical section $s_0 : M \to P$ (which is just the inclusion $s_0(m) = (m, 1)$) to obtain a Lie algebra-valued 1-form

$$\boldsymbol{A}_0 = s_0^* \boldsymbol{\omega}$$

on M (for $\boldsymbol{\omega}_{MC}$, the corresponding 1-form on M is zero by (18) and (27)).

We would like to show next that s_0^* and $\boldsymbol{\omega}_{MC}$ can be used to establish a bijection between the connections on P and the Lie algebra-valued 1-forms on M. Specifically, for any Lie algebra-valued 1-form $\boldsymbol{A}$ on M we define a Lie algebra-valued 1-form $\boldsymbol{\omega}_{\boldsymbol{A}}$ on P by

$$\boldsymbol{\omega}_{\boldsymbol{A}} = \pi_1^* \boldsymbol{A} + \boldsymbol{\omega}_{MC} . \tag{28}$$

Thus, for any $p = (m, g) \in P$ and any $v_p = (v_m, v_g) \in T_p(P)$,

$$\begin{aligned} (\boldsymbol{\omega}_{\boldsymbol{A}})_p \left(v_p \right) &= (\pi_1^* \boldsymbol{A})_{(m,g)} \left(v_m , v_g \right) + (\boldsymbol{\omega}_{MC})_{(m,g)} \left(v_m , v_g \right) \\ &= \boldsymbol{A}_m (v_m) + (\boldsymbol{\omega}_0)_g \left(v_g \right) . \end{aligned} \tag{29}$$

Note that $\boldsymbol{\omega}_{\boldsymbol{A}}$ is, in fact, a connection on P because

$$\begin{aligned}
\sigma_h^* \boldsymbol{\omega}_{\boldsymbol{A}} &= \sigma_h^* \left(\pi_1^* \boldsymbol{A} + \boldsymbol{\omega}_{MC} \right) \\
&= \left(\pi_1 \circ \sigma_h \right)^* \boldsymbol{A} + \sigma_h^* \boldsymbol{\omega}_{MC} \\
&= \pi_1^* \boldsymbol{A} + \boldsymbol{\omega}_{MC} \\
&= \boldsymbol{\omega}_{\boldsymbol{A}}
\end{aligned}$$

and

$$\begin{aligned}
\left(\boldsymbol{\omega}_{\boldsymbol{A}} \right)_p \Big(\xi^{\#} (p) \Big) &= \left(\boldsymbol{\omega}_{\boldsymbol{A}} \right)_p \Big(\left(\sigma_p \right)_{*1} (\xi) \Big) \\
&= \left(\pi_1^* \boldsymbol{A} \right)_p \Big(\left(\sigma_p \right)_{*1} (\xi) \Big) + \left(\boldsymbol{\omega}_{MC} \right)_p \Big(\left(\sigma_p \right)_{*1} (\xi) \Big) \\
&= \boldsymbol{A}_m \Big(\left(\pi_1 \circ \sigma_p \right)_{*1} (\xi) \Big) + \xi \\
&= 0 + \xi \\
&= \xi \, .
\end{aligned}$$

Moreover,

$$s_0^* \boldsymbol{\omega}_{\boldsymbol{A}} = \boldsymbol{A}$$

since

$$\begin{aligned}
s_0^* \boldsymbol{\omega}_{\boldsymbol{A}} &= s_0^* \left(\pi_1^* \boldsymbol{A} \right) + s_0^* \boldsymbol{\omega}_{MC} \\
&= \left(\pi_1 \circ s_0 \right)^* \boldsymbol{A} + 0 \\
&= \boldsymbol{A} \, .
\end{aligned}$$

Finally, observe that if $\boldsymbol{\omega}$ is any connection on P and if we take $\boldsymbol{A}_0 = s_0^* \boldsymbol{\omega}$, then $\boldsymbol{\omega} = \boldsymbol{\omega}_{\boldsymbol{A}_0}$ so every connection on P arises from some $\boldsymbol{A}$ in this way. Indeed,

$$\begin{aligned}
\boldsymbol{\omega}_{\boldsymbol{A}_0} &= \pi_1^* \boldsymbol{A}_0 + \boldsymbol{\omega}_{MC} = \pi_1^* \left(s_0^* \boldsymbol{\omega} \right) + \boldsymbol{\omega}_{MC} \\
&= \left(s_0 \circ \pi_1 \right)^* \boldsymbol{\omega} + \boldsymbol{\omega}_{MC}
\end{aligned}$$

so

$$
\begin{aligned}
\left(\boldsymbol{\omega}_{\boldsymbol{A}_0}\right)_{(m,g)}\left(v_m\,,\,v_g\right) &= \left(\,(\,s_0\,\circ\,\pi_1\,)^*\,\boldsymbol{\omega}\,\right)_{(m,g)}\left(v_m\,,\,v_g\right)\,+\,(\,\boldsymbol{\omega}_{MC}\,)_{(m,g)}\left(v_m\,,\,v_g\right)\\
&= \boldsymbol{\omega}_{(m,1)}\left(\,(\,s_0\,)_{*m}\left(\,(\,\pi_1\,)_{*(m,g)}\left(v_m\,,\,v_g\right)\right)\right)\\
&\qquad\qquad +\,(\,\boldsymbol{\omega}_{MC}\,)_{(m,g)}\left(v_m\,,\,v_g\right)\\
&= \boldsymbol{\omega}_{(m,1)}\left(\,(\,s_0\,)_{*m}\,(\,v_m\,)\,\right)\,+\,g^{-1}\,v_g\\
&= \boldsymbol{\omega}_{(m,1)}\,(\,v_m\,,\,0\,)\,+\,\boldsymbol{\omega}_{(m,g)}\left(\,0\,,\,v_g\,\right)\\
&\qquad\qquad\qquad\qquad \text{by (25) with } \xi = g^{-1}\,v_g\\
&= \boldsymbol{\omega}_{(m,g)}\,(\,v_m\,,\,0\,)\,+\,\boldsymbol{\omega}_{(m,g)}\left(\,0\,,\,v_g\,\right)\\
&\qquad\qquad\qquad\qquad \text{by (24)}\\
&= \boldsymbol{\omega}_{(m,g)}\left(\,v_m\,,\,v_g\,\right)
\end{aligned}
$$

as required.

We conclude then that any connection $\boldsymbol{\omega}$ on P can be written as

$$\boldsymbol{\omega} \,=\, \pi_1^*\,\boldsymbol{A}_0 \,+\, \boldsymbol{\omega}_{MC}\;.$$

where

$$\boldsymbol{A}_0 \,=\, s_0^*\,\boldsymbol{\omega}$$

and so $\boldsymbol{\omega}$ is completely determined by its pullback to M by the canonical section s_0.

Now suppose $s:\; M \to P,\; s(m) = (m, g(m)\,)$ is some other section of P and consider

$$\boldsymbol{A} \,=\, s^*\,\boldsymbol{\omega}\;.$$

Each such $\boldsymbol{A}$ is called a *gauge potential* for the connection (gauge field) $\boldsymbol{\omega}$. We have seen ((4)) that s gives rise to a diffeomorphism $\varphi_s:\; P \to P$ that preserves the action of G on P and carries each fiber $\pi_1^{-1}(m) \cong G$ isomorphically onto itself. Such a diffeomorphism is called a (*global*) *gauge transformation* and these are all of the form φ_s for some (global) section s. For each of these,

$$\varphi_s^*\,\boldsymbol{\omega}$$

is also a connection on P and it is said to be *gauge equivalent* to $\boldsymbol{\omega}$. By (5) we have $\varphi_s \circ s_0 = s$ so

$$\boldsymbol{A} \,=\, s^*\,\boldsymbol{\omega} \,=\, s_0^*\,(\,\varphi_s^*\,\boldsymbol{\omega}\,)$$

and we can write

$$\varphi_s^*\,\boldsymbol{\omega} \,=\, \pi_1^*\,\boldsymbol{A} \,+\, \boldsymbol{\omega}_{MC}\;. \tag{30}$$

Consequently, the connections on $\pi_1 : P \to G$ that are gauge equivalent to $\boldsymbol{\omega}$ are precisely those that correspond (via (28)) to pullbacks of $\boldsymbol{\omega}$ by different sections $s : M \to P$. A gauge transformation can therefore also be thought of as a change of section. If $s_i : M \to P,\ i = 1, 2,$ are two sections, related by the transition functions $g_{ij} : M \to G,\ i = 1, 2$ (see (6) and (7)), and $\boldsymbol{A}_i = s_i^* \boldsymbol{\omega}$, $i = 1, 2,$ then

$$\boldsymbol{A}_j = \boldsymbol{A}_i + g_{ij}^* \boldsymbol{\omega}_0 , \quad i = 1, 2 . \tag{31}$$

Notice that any connection that is gauge equivalent to a flat connection is also flat (because $d(\varphi_s^* \boldsymbol{\omega}) = \varphi_s^*((d\boldsymbol{\omega}))$. However, it is *not* true that all flat connections on $\pi_1 : P \to G$ are gauge equivalent (in fact, there is an entire "moduli space" of gauge equivalence classes of flat connections on P which one can show is just S^1).

Let's compute an example. We know that, for the Maurer-Cartan connection $\boldsymbol{\omega}_{MC}$ on $\pi_1 : P \to M,\ s_0^* \boldsymbol{\omega}_{MC}$ is identically zero. To compute a gauge transformed potential we choose the section

$$s : M \longrightarrow P$$
$$s(m) = \big(m , g(m) \big) .$$

where

$$g : M = \mathbb{R}^2 - \{ (0,0) \} \longrightarrow U(1) \subseteq \mathbb{R}^2$$

is defined by

$$g(m) = g(x, y) = \left(\frac{x}{\sqrt{x^2+y^2}} , \frac{y}{\sqrt{x^2+y^2}} \right) .$$

Since

$$(s^* \boldsymbol{\omega}_{MC})_m (v_m) = \big((\pi_2 \circ s)^* \boldsymbol{\omega}_0 \big)_m (v_m) = (g^* \boldsymbol{\omega}_0)_m (v_m) = (\boldsymbol{\omega}_0)_{g(m)} \big(g_{*m}(v_m) \big)$$
$$= \big(g(m) \big)^{-1} g_{*m}(v_m) \qquad \big(\text{by } (26)\big)$$

we will need to compute the derivative of g. For this we will regard g as a map into $\mathbb{R}^2$ and compute in standard coordinates. Then $g_{*(x,y)}$ is just the Jacobian of g so

$$g_{*(x,y)} = \frac{1}{(x^2+y^2)^{3/2}} \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix} .$$

Write $v_{(x,y)} \in T_{(x,y)}(M)$ as $v_{(x,y)} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Then

$$g_{*(x,y)}\Big(v_{(x,y)}\Big) \;=\; \frac{1}{(\,x^2+y^2\,)^{3/2}}\begin{pmatrix} y^2v_1 - xyv_2 \\ -xyv_1 + x^2v_2\end{pmatrix}$$

which we write as

$$g_{*(x,y)}\Big(v_{(x,y)}\Big) = g_{*(x,y)}(\,v_1\,,\,v_2\,) = \big(\,x^2+y^2\,\big)^{-3/2}\,(\,y^2v_1 - xy\,v_2\,,\,-xy\,v_1 + x^2v_2\,)\,.$$

Thus,

$$\begin{aligned}(\,s^*\boldsymbol{\omega}_{MC}\,)_{(x,y)}\Big(v_{(x,y)}\Big) \;&=\; \big(\,g\,(\,x,y\,)\,\big)^{-1}\,g_{*(x,y)}(\,v_2\,,\,v_2\,)\\ &=\;\big(\,x^2+y^2\,\big)^{-\frac{1}{2}}\,(\,x,\,-y\,)\,\big(\,x^2+y^2\,\big)^{-3/2}\,(\,y^2v_1 - xy\,v_2\,,\,-xy\,v_1 + x^2v_2\,)\\ &=\;\big(\,x^2+y^2\,\big)^{-2}\,\Big(\,0\,,\,(\,x^2+y^2\,)\;(\,-\,y\,v_1 + x\,v_2\,)\Big)\\ &=\;\Big(0\,,\,\frac{-y\,v_1 + x\,v_2}{x^2+y^2}\Big)\;=\;i\,\Big(\frac{-y^2v_1 + x\,v_2}{x^2+y^2}\Big)\\ &=\;i\Big(\frac{-y}{x^2+y^2}\,dx \;+\;\frac{x}{x^2+y^2}\,dy\Big)\,\Big(v_{(x,y)}\Big)\end{aligned}$$

and we conclude that

$$(\,s^*\boldsymbol{\omega}_{MC}\,)_{(x,y)} \;=\; i\,\Big(\,\frac{-y}{x^2+y^2}\,dx \;+\;\frac{x}{x^2+y^2}\,dy\Big)\;. \tag{32}$$

This is, of course, just the Lie algebra-valued 1-form on M that we earlier called $\boldsymbol{\gamma}$. It is, except for a constant, the 1-form on $\mathbb{R}^2 - \{\,(0,0)\,\}$ corresponding to the vector field representing the magnetic vector potential outside a solenoid. We see it now as a gauge potential for the Maurer-Cartan connection on the trivial $U(1)$-bundle over $\mathbb{R}^2 - \{\,(0,0)\,\}$ or perhaps better, as a gauge transformation of the potential $s_0^*\,\boldsymbol{\omega}_{MC} = 0$. According to (30), the section s gives rise to a gauge equivalent connection

$$\varphi_s^*\,\boldsymbol{\omega}_{MC} \;=\; \pi_1^*\,(\,s^*\,\boldsymbol{\omega}_{MC}\,) \;+\; \boldsymbol{\omega}_{MC}\;. \tag{33}$$

Our final objective is to show how the Aharonov-Bohm phase shift can be viewed as a path lifting, or holonomy effect associated with this connection (up to a constant which we will insert shortly).

Let $\boldsymbol{\omega}$ be any connection on our trivial $U(1)$-bundle $\pi_1 :\; P \to M$. At each $p = (m, g) \in P$ we define the *horizontal space* of $\boldsymbol{\omega}$ at p by

$$\mathrm{Hor}_p^{\boldsymbol{\omega}}\,(\,P\,) \;=\; \mathrm{Ker}\,\boldsymbol{\omega}_p \;=\; \Big\{\;(\,v_m\,,\,v_g\,) \in T_p\,(\,P\,):\;\boldsymbol{\omega}_{(m,g)}(\,v_m\,,\,v_g\,) \;=\; 0\Big\}\;.$$

For example, if $\boldsymbol{\omega} = \boldsymbol{\omega}_{MC}$, then $(\,\boldsymbol{\omega}_{MC}\,)_{(m,g)}(\,v_m\,,\,v_g\,) = g^{-1}\,v_g$, which is zero only if $v_g = 0$ so

$$\mathrm{Hor}_p^{\boldsymbol{\omega}_{MC}}\,(\,P\,) = \Big\{\;(\,v_m,\,0\,):\;v_m \in T_m\,(\,M\,)\Big\} \;\cong\; T_m\,(\,M\,) \subseteq T_m\,(\,M\,)\,\times\,T_g\,(\,G\,)$$

which actually looks horizontal. The significance of the horizonal subspaces of a connection resides in the following two properties.

$$T_p(P) \cong \mathrm{Hor}_p^{\boldsymbol{\omega}}(P) \oplus \mathrm{Vert}_p(P) \qquad \text{for each } p \in P\,, \tag{34}$$

and

$$(\sigma_h)_{*p}\big(\mathrm{Hor}_p^{\boldsymbol{\omega}}(P)\big) = \mathrm{Hor}_{p\cdot h}^{\boldsymbol{\omega}}(P) \qquad \text{for each } p \in P \text{ and } h \in G\,. \tag{35}$$

These are clear for $\boldsymbol{\omega}_{MC}$ and easy to prove in general, but we will not need the result so I won't bother to do so.

We have seen that any connection $\boldsymbol{\omega}$ on P can be written as $\boldsymbol{\omega} = \pi_1^{\,*}\boldsymbol{A} + \boldsymbol{\omega}_{MC}$, where $\boldsymbol{A} = s_0^{\,*}\boldsymbol{\omega}$. Thus,

$$\begin{aligned}\boldsymbol{\omega}_{(m,g)}(v_m, v_g) &= (\boldsymbol{\omega}_A)_{(m,g)}(v_m, v_g)\\ &= (\pi_1^{\,*}\boldsymbol{A})_{(m,g)}\big(v_m, v_g\big) + (\boldsymbol{\omega}_{MC})_{(m,g)}\big(v_m, v_g\big)\\ &= \boldsymbol{A}_m(v_m) + g^{-1}v_g\end{aligned}$$

and this is zero if and only if

$$v_g = -g\,\boldsymbol{A}_m(v_m) \tag{36}$$

so

$$\mathrm{Hor}_p^{\boldsymbol{\omega}}(P) = \big\{ \big(v_m, v_g\big) \in T_p(P) :\ v_g = -g\,\boldsymbol{A}_m(v_m) \big\}\,.$$

It is a general fact about connections that the distribution of horizontal spaces determines lifts to P of arbitrary smooth curves in M and that these are unique once an initial point is chosen. More precisely, given a smooth curve $\gamma(t)$, $a \le t \le b$, in M and a point $\tilde{\gamma}(a)$ in $\pi_1^{-1}(\gamma(a))$ there is a unique smooth curve $\tilde{\gamma}(t)$, $a \le t \le b$, in P satisfying $\pi_1 \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}'(t) \in \mathrm{Hor}_{\gamma(t)}^{\boldsymbol{\omega}}(P)$ for each t. Let us see why.

Let $\gamma(t) = (x(t), y(t))$, $a \le t \le b$, be a smooth curve in $M = \mathbb{R}^2 - \{(0,0)\}$. Consider the canonical section $s_0 : M \to P$, $s_0(x,y) = ((x,y), 1)$. Then any curve $\tilde{\gamma}(t)$ in P satisfying $\pi_1 \circ \tilde{\gamma}(t) = \gamma(t)$ for each t must be of the form

$$\tilde{\gamma}(t) = s_0\big(\gamma(t)\big)\cdot g(t) = \big(\gamma(t), g(t)\big) = \Big(\big(x(t), y(t)\big), g(t)\Big)$$

for some smooth function $g : [a,b] \to G$. Then

$$\tilde{\gamma}'(t) = \big(\gamma'(t), g'(t)\big)$$

and this is horizontal if and only if

$$g'(t) = -g(t)\,\boldsymbol{A}_{(x(t),y(t))}\big(x'(t), y'(t)\big) \tag{37}$$

(by (36)). Since $\boldsymbol{A}_{(\,x(t),y(t)\,)}\,(\,x'(\,t\,),\,y'(\,t\,)\,)$ is just some known function of t, (37) is a simple first order linear ordinary differential equation for $g(t)$ and so is uniquely solvable once $g(a)$ is specified. Thus, given $\gamma(t)$ and $(\,\gamma(a),\,g(a)\,)$ there is a unique $\tilde{\gamma}(t) = (\,\gamma(t),\,g(t)\,)$ in P satisfying $\pi_1 \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}'\,(t)$ is horizontal (with respect to $\boldsymbol{\omega}$) for each t.

Now we specialize to a particular potential $\boldsymbol{A}$. Specifically, we consider $\boldsymbol{A} = s^*\,\boldsymbol{\omega}_{MC} \;=\; s_0^*(\,\varphi_s^*\,\boldsymbol{\omega}_{MC}\,)$, where the section s is given by

$$s\,(\,x\,,\,y\,)\;=\;\left(\;(\,x\,,\,y\,)\;,\;\left(\frac{x}{\sqrt{x^2+y^2}}\;,\;\frac{y}{\sqrt{x^2+y^2}}\right)\right)\;.$$

We have already calculated $\boldsymbol{A}$ ((32)), the result being

$$\boldsymbol{A}_{(\,x,y\,)}\;=\;i\left(\frac{-y}{x^2+y^2}dx\;+\;\frac{x}{x^2+y^2}dy\right)\;.$$

Thus,

$$\boldsymbol{A}_{(\,x(t),y(t)\,)}\;\left(\,x'\,(\,t\,)\;,\;y'\,(\,t\,)\,\right)\;=\;i\left(\frac{-y(t)x'(t)+x(t)y'(t)}{(\,x(t)\,)^2+(\,y(t)\,)^2}\right)$$

and (37) becomes

$$g'\,(\,t\,)\;=\;ig\,(\,t\,)\;\left(\frac{y(t)x'(t)-x(t)y'(t)}{(\,x(t)\,)^2+(\,y(t)\,)^2}\right)\;.$$

We'll compute the horizontal lifts of two curves in M from (1,0) to (-1,0). Let $\gamma_1 = (\,\cos t,\,\sin t\,),\;0 \le t \le \pi$. Then

$$\frac{y(t)x'(t)-x(t)y'(t)}{(\,x(t)\,)^2+(\,y(t)\,)^2}\;=\;-\,1\,.$$

so (38) becomes

$$g'\,(\,t\,)\;=\;-ig\,(\,t\,)$$

and so

$$g\,(\,t\,)\;=\;g\,(\,0\,)\,e^{-it}$$

and

$$\tilde{\gamma}_1\,(\,t\,)\;=\;\left(\,\gamma_1\,(\,t\,)\;,\,g\,(\,0\,)\,e^{-it}\,\right)\,,\qquad 0 \le t \le \pi\;.$$

In exactly the same way, if $\gamma_2(t) = (\,\cos{(2\pi - t)},\,\sin{(2\pi - t)}\,),\;0 \le t \le \pi$, one obtains

$$\tilde{\gamma}_2\,(\,t\,)\;=\;\left(\,\gamma_2\,(\,t\,)\;,\,g\,(\,0\,)\,e^{it}\,\right)\,,\qquad 0 \le t \le \pi\;.$$

Note that

$$\tilde{\gamma}_1\,(\,\pi\,)\;=\;\left(\,(\,-\,1,0\,)\;,\,-g\,(\,0\,)\,\right)\;=\;\tilde{\gamma}_2\,(\,\pi\,)$$

so there is no "phase difference" for γ_1 and γ_2 .

Now we return to the Aharonov-Bohm experiment and apply exactly the same ideas to the "physical" gauge potential

$$\boldsymbol{A}_{AB} = i\left(\frac{q\Phi_0}{2\pi\hbar c}\right)\left(\frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy\right).$$

(which corresponds to the connection $\boldsymbol{\omega}_{AB} = \pi_1^*\boldsymbol{A}_{AB} + \boldsymbol{\omega}_{MC}$. Thus, (38) becomes

$$g'(t) = i\left(\frac{q\Phi_0}{2\pi\hbar c}\right) g(t)\left(\frac{y(t)x'(t) - x(t)y'(t)}{(x(t))^2 + (y(t))^2}\right).$$

The solutions for $\gamma_1(t)$ and $\gamma_2(t)$ are, of course, obtained in exactly the same way and the results are

$$\tilde{\gamma}_1(t) = \left(\gamma_1(t)\,,\, g(0)\,e^{-i(\frac{q\Phi_0}{2\pi\hbar c})t}\right)$$

and

$$\tilde{\gamma}_2(t) = \left(\gamma_2(t)\,,\, g(0)\,e^{i(\frac{q\Phi_0}{2\pi\hbar c})t}\right).$$

In particular,

$$\tilde{\gamma}_1(\pi) = \left((-1,0)\,,\, e^{-i(\frac{q\Phi_0}{2\hbar c})}\right)$$

and

$$\tilde{\gamma}_1(\pi) = \left((-1,0)\,,\, e^{i(\frac{q\Phi_0}{2\hbar c})}\right)$$

which gives a phase difference of

$$\frac{q\Phi_0}{2\hbar c} - \left(-\frac{q\Phi_0}{2\hbar c}\right) = \frac{q\Phi_0}{\hbar c}$$

exactly as in (8) of Appendix C. As promised, the connection $\boldsymbol{\omega}_{AB}$ is doing its job of keeping track of the phase shifts of Aharonov-Bohm electrons as they traverse their classical paths.