Chapter 19
Gauge invariance and quantum mechanics

December 23, 2020

The deep relationship between gauge transformations in EM and local phase transformations in quantum mechanics is introduced. The vector potential emerges as a geometric concept: a connection 1-form that accounts for the misalignment of basis vectors. The field emerges as the curvature 2-form. A very brief introduction to non-abelian gauge theories is given.

Contents

1 Introduction

Relating two concepts

This Chapter presents the deep and perhaps unexpected relationship between two ideas, one in EM, the other in quantum mechanics (QM).

- EM allows gauge transformations

$$\boxed{A^\mu(x) \mapsto A^\mu(x) + \partial^\mu \Lambda(x)} \tag{1}$$

- The phase of QM wavefunctions can be changed:

$$\Psi(x) \mapsto \exp i\Theta \cdot \Psi(x) \tag{2}$$

It turns out that the change in phase can be spacetime-dependent:

$$\boxed{\Psi(x) \mapsto \exp i\Theta(x) \cdot \Psi(x)} \tag{3}$$

Importantly, the gauge transformation in EM and the phase transformation in QM (for a point charge q) must be related:

$$\boxed{\Theta(x) = \frac{q}{\hbar}\Lambda(x)} \tag{4}$$

Geometric point of view

The transformation (2) rotates the real and imaginary axes of the complex plane backwards by Θ, while (3) means that the rotation can vary with x; thus the basis vectors can be *misaligned*, for which a standard framework exists. The key *geometric*

idea is the *connection 1-form*, which will turn out to be just the 4-vector potential.

Considering simple cases
Although the formalism is valid in 4D (x refers to spacetime), the ideas can be explained in a simpler context: time-independent situations in 3D space.

Study guide
Students should first focus on Section 2. The rest of the Chapter introduces advanced and general concepts that have grown out of gauge invariance in EM.

This Chapter is slightly repetitive, but for good reasons. (a) The whole idea is to unite two points of view (EM and QM), so the same thing has to be said twice from these perspectives. (b) The idea is described both for the abelian case of EM and more generally to include non-abelian (e.g., Yang–Mills) gauge fields.

2 Dynamics of a charged particle

2.1 Ways to describe dynamics

There are three ways to describe the dynamics of a point charge q.

Newtonian mechanics
In Newtonian mechanics and in obvious notation,

$$m\frac{d\mathbf{v}}{dt} \quad = \quad \mathbf{F} = q\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right) \tag{5}$$

Potentials do not appear.

Classical Hamiltonian mechanics
From the classical Hamiltonian for a free particle

$$H \quad = \quad \frac{\mathbf{p}^2}{2m}$$

the replacements (minimal coupling)

$$\begin{aligned} \mathbf{p} &\mapsto \mathbf{p} - q\mathbf{A} \\ H &\mapsto H - q\Phi \end{aligned} \tag{6}$$

lead to

$$H - q\Phi \quad = \quad \frac{(\mathbf{p}-q\mathbf{A})^2}{2m}$$

or

$$\boxed{H = \frac{(\mathbf{p}-q\mathbf{A})^2}{2m} + q\Phi} \tag{7}$$

Incidentally, (6) can be combined into

$$\boxed{p_\mu \mapsto p_\mu - qA_\mu} \tag{8}$$

since H is the energy $E = p^0/c$ and $\Phi = A^0/c$.

The equations of motion are then obtained from

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i} \quad , \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i} \tag{9}$$

Although the potentials appear explicitly, (9) is equivalent to (5), so there is no problem.

Quantum mechanics
To go from a classical Hamiltonian to QM ("canonical quantization"), make the replacement[1] in H:

$$\mathbf{p} \quad \mapsto \quad -i\hbar\boldsymbol{\nabla} \tag{10}$$

Time-independent problems are to be solved with[2] $H\,\Psi = E\,\Psi$. More explicitly

$$\left[\frac{1}{2m}\left(-i\hbar\boldsymbol{\nabla} - q\mathbf{A}\right)^2 + q\Phi\right]\Psi = E\,\Psi \tag{11}$$

Do the answers (e.g., the energy E and the probability density $|\Psi|^2$, both measurable) remain unchanged under a gauge transformation?

2.2 Covariant derivative

Definition
From the kinetic energy term in (11), introduce the covariant derivative $\mathbf{D}$:

$$-i\hbar\boldsymbol{\nabla} - q\mathbf{A} \quad = \quad -i\hbar\mathbf{D}$$

or

$$\mathbf{D} \quad = \quad \boldsymbol{\nabla} - i(q/\hbar)\mathbf{A} \tag{12}$$

Generalizing to spacetime,

$$\boxed{D_\mu = \partial_\mu - i(q/\hbar)A_\mu} \tag{13}$$

Properties
D_μ is the generalization of ∂_μ when there is a 4-vector potential. It is called the *covariant derivative* (for a reason explained in Section 3.6).

The ordinary derivative commutes:

$$[\partial_\mu, \partial_\nu] \quad = \quad 0$$

meaning that $\partial_\mu\partial_\nu\,\Psi(x) = \partial_\nu\partial_\mu\,\Psi(x)$ for any Ψ. For the covariant derivative,

$$\begin{aligned} &[D_\mu, D_\nu] = -i(q/\hbar)\left\{[\partial_\mu, A_\nu] - [\partial_\nu, A_\mu]\right\} \\ = \quad &-i(q/\hbar)\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) \end{aligned}$$

[1] Acting on a plane wave, $\boldsymbol{\nabla} \mapsto i\mathbf{k}$, where $\mathbf{k}$ is the wave number. So this replacement is just a mathematical expression of de Broglie's relation.

[2] The operator character of H will not be indicated explicitly. A label n on the energy E and the wavefunction Ψ is omitted for simplicity.

In other words

$$\boxed{[D_\mu, D_\nu] = -i(q/\hbar)F_{\mu\nu}} \tag{14}$$

where $F_{\mu\nu}$ is the EM field tensor. Refer to Appendix A for commutators.

The Hamiltonian
In terms of the covariant derivative,

$$H = -\frac{\hbar^2}{2m}\mathbf{D}^2 + q\Phi \tag{15}$$

2.3 Consistency?

The problem
Alice and Bob respectively solve for the energy and wavefunction:

$$H\,\Psi = E\,\Psi \quad \text{(Alice)} \tag{16}$$
$$H'\,\Psi' = E'\,\Psi' \quad \text{(Bob)} \tag{17}$$

with H given by (15) and $\mathbf{D}$ by (12), while H' is given in terms of

$$\mathbf{D}' = \boldsymbol{\nabla} - i(q/\hbar)\mathbf{A}'$$
$$\mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\Lambda \tag{18}$$

Since Alice and Bob refer to the same physical situation (differing only by a gauge transformation), they should get the same energy $E' = E$ and the same probability density:

$$|\Psi'(\mathbf{x})|^2 = |\Psi(\mathbf{x})|^2 \tag{19}$$

This implies

$$\Psi'(\mathbf{x}) = e^{i\Theta(\mathbf{x})}\,\Psi(\mathbf{x}) \tag{20}$$

Since (19) holds for each $\mathbf{x}$ separately, $\Theta(\mathbf{x})$ can depend on $\mathbf{x}$.

How can we be sure that Ψ' as given by (20) will satisfy (17) with $E' = E$?

The condition
It is claimed that the required condition is

$$\boxed{H'e^{i\Theta} = e^{i\Theta}H} \tag{21}$$

Problem 1
Show that if (21) is satisfied, then (16) implies (17) with $E' = E$. §

To check whether (21) holds, first note that the $q\Phi$ term is the same in H and H', and does not cause any trouble. It remains to check

$$(\mathbf{D}')^2\,e^{i\Theta} = e^{i\Theta}\,\mathbf{D}^2$$

which in turn will be guaranteed if

$$\mathbf{D}'\,e^{i\Theta} = e^{i\Theta}\,\mathbf{D}$$

or, putting it in the 4D form

$$\boxed{D'_\mu\,e^{i\Theta} = e^{i\Theta}\,D_\mu} \tag{22}$$

This can also be written as

$$\boxed{D'_\mu = e^{i\Theta}\,D_\mu\,e^{-i\Theta}} \tag{23}$$

Interpretation
Operator products should be read from right to left (because, when acting on a wavefunction, the rightmost factor acts first), so the RHS of (23) tells Bob to do the following (**Figure 4**): (a) Rotate the phase back to that used by Alice ($e^{-i\Theta}$). (b) Do the covariant derivative according to Alice (D_μ). (c) Rotate the phase forward to that used by Bob ($e^{i\Theta}$).

Problem 2
Consider (23). (a) Show that LHS $= D_\mu - i(q/\hbar)\partial_\mu\Lambda$. (b) Show that RHS $= D_\mu - i\partial_\mu\Theta$. You need to make use of the rule for commutators $[\partial_\mu, f(x)] = \partial_\mu f(x)$. (c) Hence show that consistency between Alice and Bob would be guaranteed if (4) in Section 1 holds. §

Using these, we also find that under a phase rotation, $D_\mu \mapsto D'_\mu$:

$$\boxed{D'_\mu = D_\mu - i\partial_\mu\Theta} \tag{24}$$

This is all we need for ensuring that QM is consistent with the gauge degree of freedom in EM. Appendix B provides insight through one realistic example. In this context it might be appropriate to quote Dyson [1], who, apart from Tomonaga, Schwinger and Feynman who shared the 1965 Nobel Prize in Physics [3], was regarded as the fourth architect of quantum electrodynamics:

> The progress of science requires the growth of understanding in both directions, downward from the whole [general theory] to the parts [specific examples and applications] and upwards from the parts to the whole.
> (Added remarks in parentheses)

3 Geometric point of view

3.1 Local reference frame

Phase is universal
Importantly, $\Theta(\mathbf{x})$ must be the same for all wavefunctions, so that inner products such as

$$\langle\Psi_b|\Psi_a\rangle = \int \Psi_b^*\Psi_a\,d^3x$$

remain invariant. So the local phase convention should be associated not with the wavefunction, but with the complex plane itself.

Complex plane as 2D real plane
For our purpose, regard the complex wavefunction Ψ as a vector with two real components:[3]

$$\vec{\Psi} \quad = \quad \Psi^j \, \vec{e}_j = \Psi^1 \, \vec{e}_1 + \Psi^2 \, \vec{e}_2 = \begin{pmatrix} \Psi^1 \\ \Psi^2 \end{pmatrix}$$

with $\vec{e}_1 = 1$, $\vec{e}_2 = i$ as the two basis vectors. The phase convention determines the orientation of the real and imaginary axes.

We record here, for later use, a correspondence between the complex representation Ψ and the vector representation $\vec{\Psi}$:

$$\begin{aligned} \Psi_b &= (\alpha + \beta i) \, \Psi_a \\ \vec{\Psi}_b &= (\alpha + \beta \Sigma) \, \vec{\Psi}_a \end{aligned}$$

where

$$\Sigma \quad = \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad (25)$$

Problem 3
(a) Prove the above equivalence. (b) Check that $\Sigma^2 = -I$ where I is the identity operator. §

Thus Σ plays the role of $i = \sqrt{-1}$ in the vector representation.

Misaligned basis vectors
The gauge degree of freedom in EM ($\Lambda(\mathbf{x})$) implies freedom to add a phase ($\Theta(\mathbf{x})$), i.e., rotating the real and imaginary axes backwards by Θ (**Figure 1**). Since Θ can depend on $\mathbf{x}$, the basis vectors may be *misaligned*: there is a change of the basis vector $d\mathbf{e}$ upon a displacement dx (**Figure 2**).[4]

3.2 Generalizations

The language and the tools for dealing with misaligned basis vectors should be introduced in a broader context for a vector field $\vec{\Psi}(x)$.[5]

The domain
The domain $\mathcal{M}$ is the set of points $x = (x^1, \ldots, x^M)$. Do not assume $M = 3$ or 4, or that these are Cartesian coordinates. Do not identify x with the displacement from the origin (a vector); indeed there need not be an origin. For example, $x = (\theta, \phi)$, where these are the polar coordinates on a unit sphere.

On the *manifold* $\mathcal{M}$ there is the concept of being *nearby* (x^μ changing by an infinitesimal amount dx^μ), but there need not be a concept of distance, much less that it is given by the Pythagorean formula.

The range
The dependent variable $\vec{\Psi}$ is a vector with (real or complex) components Ψ^j, $j = 1, \ldots, N$. Thus the mapping is

$$\mathcal{M} \quad \mapsto \quad V$$

represented as in **Figure 3a**, which deliberately suggests that $\mathcal{M}$ need not be "flat".

Dimensions
Importantly, M and N can be different. In our example, $M = 3$ (position labelled by three coordinates) and $N = 2$ (wavefunction described by two real numbers). As another example, we can have $M = 4$ (spacetime, which can be curved or at least characterized by curvilinear coordinates), and at each x, there is a vector with components Ψ^j, $j = 1, 2, 3$, being the complex amplitudes of finding a red, green or blue quark.

Misaligned basis vectors
Define basis vectors $\vec{e}_j$, $j = 1, \ldots, N$ at each point.[6] For simplicity, assume the basis vectors are orthonormal:[7]

$$\vec{e}_j \cdot \vec{e}_k \quad = \quad \delta_{jk}$$

The basis vectors can be different at different points (see **Figure 2**) — that is the meaning of "misalignment".[8] So we think of a *different* or "private" vector space $V(x)$ associated with each point x (**Figure 3b**) and the mapping

$$\mathcal{M} \quad \mapsto \quad V(x)$$

3.3 Differentiating a vector

Consider the differential of a vector, i.e., the infinitesimal version of vector subtraction:

$$d\vec{\Psi} \quad = \quad \vec{\Psi}(x + dx) - \vec{\Psi}(x)$$

[3] As usual, whenever $i = \sqrt{-1}$ may appear, we avoid using the index i to denote a component; indices $jk\ell$ are used instead.

[4] First, this statement is schematic, omitting the necessary labels on $\mathbf{e}$ (which basis vector) and on x (which coordinate); see below for details. Second, we are careful to refer only to $d\mathbf{e}$ and not $\mathbf{e}$; see the next Section.

[5] The notation $\mathbf{x}$ is restricted to 3D vectors.

[6] For the general discussion, we use the notation $\vec{e}_j$ rather than $\mathbf{e}_j$; bold letters are reserved for 3D or 2D Euclidean space.

[7] In that case j, k etc. can be freely written as upper or lower indices. For students who have learnt general relativity: in Riemann geometry, the "natural" basis vectors for the tangent plane are not orthonormal.

[8] This discussion is heuristic. As shown in the next Section, we may not be able to speak of $\vec{e}_j(x)$ but only about its change $d\vec{e}_j$ upon a displacement dx^μ.

Expressing in terms of basis vectors:

$$\begin{aligned} d\vec{\Psi} &= d\left(\Psi^j\,\vec{e}_j\right) \\ &= \left(d\,\Psi^i\right)\vec{e}_j + \boxed{\Psi^k\,d\vec{e}_k} \end{aligned} \qquad (26)$$

The first term alone corresponds to differentiating component-by-component; all the new features come from the second term. A different dummy variable k has been used for later convenience.

The connection 1-form
The differential $d\vec{e}_k$ (a) is linearly related to the infinitesimal displacement dx^μ, and (b) is a vector, therefore expressible as a sum over basis vectors $\vec{e}_j$. The *connection 1-form* Γ is defined as the coefficient in the proportionality:[9]

$$\boxed{d\vec{e}_k = \Gamma^j_{\mu,k}\,dx^\mu\,\vec{e}_j} \qquad (27)$$

It connects the basis vectors at neighboring points; hence the name. There are 3 indices: two (j, k) associated with directions in V, and one (μ) with a an infinitesimal displacement on $\mathcal{M}$. In (27), k is a free index, to be chosen consistently on both sides; j and μ are dummy indices to be summed over.

Combining (26) and (27) leads to

$$\boxed{d\vec{\Psi} = \left(d\Psi^j + \Psi^k\Gamma^j_{\mu,k}dx^\mu\right)\,\vec{e}_j} \qquad (28)$$

Antisymmetry
The 1-form $\Gamma^j_{\mu,k}$ is antisymmetric in its indices j, k.

Problem 4
Prove the above statement starting from the fact that $\vec{e}_j \cdot \vec{e}_k$ is the same at every point. §

Analogy
The connection 1-form relates the *phase* convention between two points x and $x + dx$. There is an interesting analogy (apart from a factor of $i = \sqrt{-1}$) with the *scaling* convention of different monetary units [2].

3.4 An elementary example

The coordinate system
Let $\mathcal{M}$ be the 2D plane described in polar coordinates: $x = (x^1, x^2) = (r, \phi)$. Let V be the same as $\mathcal{M}$, with the usual orthonormal basis vectors in the radial and tangential directions:[10]

$$\begin{aligned} \vec{e}_1 &= \vec{e}_r = \cos\phi\,\hat{\mathbf{i}} + \sin\phi\,\hat{\mathbf{j}} \\ \vec{e}_2 &= \vec{e}_\phi = -\sin\phi\,\hat{\mathbf{i}} + \cos\phi\,\hat{\mathbf{j}} \end{aligned} \qquad (29)$$

which are of course not constants.

Problem 5
(a) Show that

$$d\vec{e}_r = \vec{e}_\phi\,d\phi \quad , \quad d\vec{e}_\phi = -\vec{e}_r\,d\phi$$

and hence (b) the only non-zero elements of the connection 1-form are

$$\Gamma^\phi_{\phi,r} = 1 \quad , \quad \Gamma^r_{\phi,\phi} = -1$$

This is all about geometry, i.e., the basis vectors. §

Uniform circular motion
A particle undergoes uniform circular motion: $r(t) = R$, $\phi(t) = \omega t$. Its velocity $\vec{v} = \omega R\,\vec{e}_\phi$ has radial and tangential components

$$v^r = 0 \quad , \quad v^\phi = \omega R$$

The acceleration $\vec{a}$ is *not* given by $a^i = dv^i/dt$; in particular, even a high school student knows that $a^r \neq 0$ even though v^r is constant. The correct calculation is to use (28):

$$\vec{a}\,dt = d\vec{v} \;=\; \left(dv^j + v^k\Gamma^j_{\mu,k}dx^\mu\right)\vec{e}_j$$

To simplify, note the following.

- The velocity components do not change: $dv^j = 0$.
- For the factor v^k, the only contribution is $k = \phi$, with $v^k = v^\phi = \omega R$.
- The only non-zero dx^μ is $\mu = \phi$, for which $dx^\mu = d\phi = \omega dt$.

Then we get, upon cancelling dt,

$$\begin{aligned} \vec{a} &= (\omega R)\,\Gamma^j_{\phi,\phi}\,\omega\,\vec{e}_j \\ &= \omega^2 R\left(\Gamma^r_{\phi,\phi}\vec{e}_r + \Gamma^\phi_{\phi,\phi}\vec{e}_\phi\right) \\ &= -\omega^2 R\,\vec{e}_r \end{aligned}$$

by using the values of $\Gamma^j_{\mu,k}$ evaluated in Problem 5. The familiar centripetal acceleration is due to the connection 1-form!

Comments about this example
Some special features of this example should be mentioned. The vector space V is the same as the (infinitesimal)[11] displacement on $\mathcal{M}$; $V(x)$ is called the *tangent plane* at x. Therefore the two types of indices j, k and μ are the same, and they are usually denoted all with Greek letters. It is also better to use "natural" basis vectors which are not necessarily orthonormal; the connection 1-form would have nicer properties. (Students who have studied Riemann geometry should be more familiar with this version.)

[9] When Γ is written without indices, it is a shorthand to mean the whole object $\Gamma^j_{\mu,k}$.

[10] These are not the standard basis vectors used in Riemann geometry.

[11] In general, if $\mathcal{M}$ is not flat, only *infinitesimal* displacements are "straight arrows" and hence vectors.

3.5 Expressing in terms of components

Go back to the general case and (26). Since $d\Psi^j = \partial_\mu \Psi^j \, dx^\mu$ (the rule for the differential of any function), we get

$$\begin{aligned} d\vec{\Psi} &= \partial_\mu \Psi^j \, dx^\mu \, \vec{e}_j + \Psi^k \left(\Gamma^j_{\mu,k} \, dx^\mu \vec{e}_j \right) \\ &= \left(\delta^j{}_k \, \partial_\mu + \Gamma^j_{\mu,k} \right) \Psi^k \, dx^\mu \, \vec{e}_j \\ &\equiv (D_\mu)^j{}_k \Psi^k dx^\mu \, \vec{e}_j \\ (d\Psi)^j &= (D_\mu)^j{}_k \, \Psi^k \, dx^\mu \end{aligned} \tag{30}$$

The LHS means differentiate the vector and then take the j component (not the other way round); on the RHS, the factor $\vec{e}_j$ is simply peeled off in order to pick out the j component.

Matrix notation
Use matrix notation to hide the (j,k) indices:

$$\begin{aligned} d\vec{\Psi} &= [D_\mu] \, \vec{\Psi} \, dx^\mu \\ [D_\mu]^j{}_k &= \delta^j{}_k \, \partial_\mu + \Gamma^j_{\mu,k} \\ [D_\mu] &= [I] \, \partial_\mu + [\Gamma_\mu] \end{aligned}$$

The matrix notation for Γ_μ should be understood with the upper index on the left (a row index) and the lower index on the right (a column index).

Often we adopt a more compact notation, without writing $[\ldots]$ for matrices or an arrow for vectors; moreover the identity is omitted:

$$\boxed{D_\mu = \partial_\mu + \Gamma_\mu} \tag{31}$$

Problem 6
Write the results of Problem 5 in terms of the matrices $[\Gamma_r]$ and $[\Gamma_\phi]$. §

Whenever the basis vectors (in this case equivalent to the choice of real and imaginary axes, i.e., the phase convention) are allowed to vary, the natural way to differentiate a vector — called the *covariant derivative* — is to replace ∂_μ with D_μ; the latter is a matrix because it mixes components.

Separating two properties
The 2×2 matrix Γ_μ has two properties.

- It is an antisymmetric matrix.
- It depends on position: $\Gamma_\mu = \Gamma_\mu(x)$.

To separate these two properties, note that any 2×2 antisymmetric matrix must be proportional to Σ defined in (25). Thus define a function $A_\mu(x)$ by

$$\Gamma_\mu(x) = -\frac{q}{\hbar} A_\mu(x) \, \Sigma \tag{32}$$

The *matrix* property is contained in Σ (which is not position-dependent) and the position-dependence is contained in A_μ (which is not a matrix). The factor $-(q/\hbar)$ is a matter of convention; it will turn out that A_μ so defined is exactly the 4-vector potential in EM.

Back to complex form
Thus we have

$$\begin{aligned} D_\mu \vec{\Psi} &= \partial_\mu \vec{\Psi} - \frac{q}{\hbar} A_\mu(\Sigma \vec{\Psi}) \\ D_\mu \Psi &= \partial_\mu \Psi - i \frac{q}{\hbar} A_\mu \Psi \end{aligned}$$

where the second line is obtained by converting the vector expression to the corresponding complex representaiton, with $\Sigma \mapsto i$. Thus we recover (13).

Therefore D_μ defined in (13), coming from EM, is the same as D_μ defined in (31) and (32) from the concept of non-constant basis vectors — thus giving a geometric picture to the vector potential.

Transformation of the 1-form
Now that the two notions are identified, (24) and (31) lead to the transformation rule under a rotation of the real and imaginary axes: $\Gamma_\mu \mapsto \Gamma'_\mu$,

$$\boxed{\Gamma'_\mu = \Gamma_\mu - i \partial_\mu \Theta} \tag{33}$$

This is written in complex representation; in vector representation, $i \mapsto \Sigma$.

Note that $\Gamma_\mu = 0$ does not guarantee that $\Gamma'_\mu = 0$. That is why it is not regarded as a tensor.

3.6 Meaning of "covariance"

The concept of covariance is analogous to that in Lorentz transformations.

Lorentz transformations
Consider the Lorentz force law

$$\frac{dp^\mu}{d\tau} = q F^{\mu\nu} u_\nu$$

or schematically $X^\mu = Y^\mu$ where X^μ and Y^μ are shorthands for the LHS and the RHS. Equivalently, $Z^\mu \equiv X^\mu - Y^\mu = 0$. The law is said to be covariant because, under a Lorentz transformation (a general "rotation" of the 4 spacetime coordinates), the quantities transform in the same way,[12] e.g.

$$Z^\mu \mapsto Z'^\mu = L^\mu{}_\nu Z^\nu$$

Therefore the law $Z^\mu = 0$ guarantees $Z'^\mu = 0$: the same law is valid in any reference frame.

Phase transformations
Consider rotations not of the spacetime axes but of the axes in V, equivalent to adding a phase, i.e.,

$$\Psi \mapsto \Psi' = e^{i\Theta} \Psi$$

[12] That is the meaning of "co" or together.

Repeating some earlier formulas, we have $D_\mu \Psi \mapsto D'_\mu \Psi'$,

$$D'_\mu \Psi' \quad = \quad D'_\mu (e^{i\Theta}\Psi) = e^{i\Theta}(D_\mu \Psi)$$

i.e., transforming in the same way ("together", or covariantly) as Ψ, with the same multiplicative factor on the left. Therefore equations constructed with D_μ will remain valid after phase rotations — this is the essence of Section 2.3.

Recall that $D_\mu = \partial_\mu + \Gamma_\mu$, but each of the two terms is *not* covariant.

4 Curvature

4.1 Key idea

Are there basis vectors $\vec{e}_j(x)$, defined in a path-independent way, which change with position according to the following rule?

$$d\vec{e}_j(x) \quad = \quad \left(\Gamma^k_{\mu,j}\, \vec{e}_k\right)\, dx^\mu \equiv C_\mu\, dx^\mu \qquad (34)$$

If so, all the complications can be removed if we "undo" the misalignment. (The variable C_μ suppresses some indices to make the discussion below clearer.)

To say that $d\vec{e}_j$ exists but $\vec{e}_j$ may not exist seems mathematically perplexing; yet it is something familiar and central in physics (Appendix C).

Suppose there is such a function of state $\vec{e}_j(x)$. If the changes in each small step dx^μ are added up around a closed loop, the result should be zero:

$$\oint C_\mu\, dx^\mu \quad \stackrel{?}{=} \quad 0$$

$$\partial_\nu C_\mu - \partial_\mu C_\nu \quad \stackrel{?}{=} \quad 0$$

Inserting the definition of C_μ, we get

$$\begin{aligned} \partial_\nu \left(\Gamma^k_{\mu,j}\, \vec{e}_k\right) - (\nu \leftrightarrow \mu) \quad &\stackrel{?}{=} \quad 0 \\ (\partial_\nu \Gamma^k_{\mu,j})\, \vec{e}_k + \Gamma^k_{\mu,j} (\partial_\nu \vec{e}_k) & \\ - (\nu \leftrightarrow \mu) \quad &\stackrel{?}{=} \quad 0 \end{aligned}$$

In the second term, use

$$\partial_\nu \vec{e}_k \quad = \quad \Gamma^\ell_{\nu,k} \vec{e}_\ell$$

which is just another way of expressing (27). In the first term, change the dummy variable $k \mapsto \ell$, so that everything multiplies $\vec{e}_\ell$, which can be peeled off, leading to the condition

$$\partial_\nu \Gamma^\ell_{\mu,j} + \Gamma^\ell_{\nu,k} \Gamma^k_{\mu,j} - (\nu \leftrightarrow \mu) \quad \stackrel{?}{=} \quad 0$$

The sum over k can be understood as matrix multiplication:

$$\partial_\nu [\Gamma_\mu]^\ell{}_j + \{[\Gamma_\mu][\Gamma_\nu]\}^\ell{}_j - (\nu \leftrightarrow \mu) \quad \stackrel{?}{=} \quad 0$$

which is just the $^\ell{}_j$ element of the matrix equation

$$\partial_\nu \Gamma_\nu + \Gamma_\nu \Gamma_\mu - (\nu \leftrightarrow \mu) \quad \stackrel{?}{=} \quad 0 \qquad (35)$$

where Γ_ν etc., written with one index, are matrices.

So there will be something non-trivial if this object is not zero; this is how the concept of curvature is defined.

4.2 Curvature 2-form

We are led to define the *curvature 2-form*

$$R_{\nu\mu} \quad = \quad \partial_\nu \Gamma_\nu + \Gamma_\mu \Gamma_\nu - (\nu \leftrightarrow \mu)$$

or, more compactly

$$\boxed{R_{\nu\mu} = \partial_\nu \Gamma_\mu - \partial_\mu \Gamma_\nu + [\Gamma_\nu, \Gamma_\mu]} \qquad (36)$$

All the genuine physics depends on this object.

Putting back the indices j, k etc.,

$$\begin{aligned} R^\ell_{\nu\mu,j} \quad = \quad & \partial_\nu \Gamma^\ell_{\mu,j} - \partial_\mu \Gamma^\ell_{\nu,j} \\ & + \Gamma^\ell_{\nu,k} \Gamma^k_{\mu,j} - \Gamma^\ell_{\mu,k} \Gamma^k_{\nu,j} \end{aligned} \qquad (37)$$

Those who have studied general relativity should find this formula familiar. (There may be some differences in sign depending on convention.)

The four indices
Two indices (say ℓ, j) refer to the vector space V. The elements of R are antisymmetric between these indices. Two indices (say ν, μ) refer to infinitesimal displacements along $\mathcal{M}$. The elements of R are antisymmetric between these two indices.

Problem 7
Evaluate the curvature 2-form for the example in Problem 6. §

Another expression
We also have

$$\begin{aligned} & [D_\nu, D_\mu] = [\partial_\nu + \Gamma_\nu\, ,\, \partial_\mu + \Gamma_\mu] \\ = \quad & \partial_\nu \Gamma_\mu - \partial_\mu \Gamma_\nu + [\Gamma_\nu, \Gamma_\mu] \\ = \quad & R_{\mu\nu} \end{aligned} \qquad (38)$$

This is probably the best way to remember $R_{\nu\mu}$ in general. Also see (14).

4.3 The case of EM

The general formalism can be simplified in the case of EM.

Abelian
From (32), Γ_μ involves only one matrix Σ, so

$$[\Gamma_\nu, \Gamma_\mu] \quad \propto \quad A_\nu \, A_\mu \, [\Sigma, \Sigma] = 0$$

After all, phase rotations constitute a commutative or *abelian* group, i.e., the order of two rotations (2D real pace) does not matter.[13]

For any abelian group (36) simplifies to

$$R_{\nu\mu} \quad = \quad \partial_\nu \Gamma_\mu - \partial_\mu \Gamma_\nu \tag{39}$$

The connection 1-form appears linearly.

Factor out the matrix
Next, the matrix property of $R_{\nu\mu}$ can be factored out:

$$R_{\nu\mu}(x) \quad = \quad -\frac{q}{\hbar} \, F_{\nu\mu}(x) \, \Sigma \tag{40}$$

where from (39)

$$F_{\nu\mu} \quad = \quad \partial_\nu A_\mu - \partial_\mu A_\nu$$

is just the usual field tensor in EM — now understood to be just the curvature 2-form.

Transformation property
The covariant derivative is designed to satisfy the transformation property (23). Thus, from (38) it follows that under a phase transformation,[14]

$$\boxed{R'_{\mu\nu} = e^{i\Theta} R_{\mu\nu} e^{-i\Theta}} \tag{41}$$

Thus a statement such as $R_{\mu\nu} = 0$ is preserved under phase transformations. In contrast, a statement such sa $\Gamma_\mu = 0$ would not be preserved; see (33).

Consider[15]

$$R^2 \quad \equiv \quad \text{tr } R_{\mu\nu} R^{\mu\nu} \propto F_{\mu\nu} F^{\mu\nu}$$

This is invariant under Lorentz transformations (because all the $\mu\nu$ indices are contracted), and also under phase rotations $\Theta(x)$, on account of (**??**) and all the j, k indices being contracted in the trace. Thus it is a suitable candidate (indeed the only viable one up to a multiple) for the Lagrangian density. These considerations are in fact general, as discussed in Section 5.

[13] This is true for rotations in 2D, but not in higher dimensions.

[14] This is written in the complex representation. In terms of two real directions, i would be replaced by the matrix Σ.

[15] In the complex representation, we would be dealing with numbers, and there is no need to take trace. In terms of two real directions, everything is expressed in terms of 2×2 matrices, hence we take trace.

5 General gauge theories*

**This Section goes beyond EM, and can be skipped.*

This Section presents the key ideas in general gauge theories, as the first step towards the strong interactions (quantum chromodynamics, QCD) as well as electroweak interactions.[16] This Section repeats much of the earlier material, and gathers in one place those formulas that remain valid in the general case.

5.1 Symmetry for the matter field

Consider a matter field Ψ with components Ψ^j, $j = 1, \ldots, N$, corresponding to N directions in internal space, e.g., $N = 3$ colors of the quark field. Assume the internal degree of freedom (e.g., color) is irrelevant; then we are free to make transformations

$$\Psi(x) \mapsto \Psi'(x) = \exp iT \cdot \Psi(x)$$

Next, assume that T can be made local, $T = T(x)$; the symmetry is said to be *gauged*.

$$\Psi(x) \mapsto \Psi'(x) = \exp iT(x) \cdot \Psi(x) \tag{42}$$

Here T or $T(x)$ is an $N \times N$ matrix. The class of matrices considered depends on the symmetry group, briefly explained in Appendix D.

The local symmetry means an arbitrary convention or basis vectors for the N directions in internal space at each spacetime point.

5.2 Connection 1-form emerges

Covariant derivative
Since the basis vectors may vary from place to place, differentiation should be defined in terms of

$$D_\mu \quad = \quad \partial_\mu + \Gamma_\mu \tag{43}$$

which is the same as (31) except that Γ_μ is an $N \times N$ matrix.

Transformation of covariant derivative
As before,

$$D_\mu \Psi \quad \mapsto \quad D'_\mu \Psi' = D'_\mu \left(e^{iT} \cdot \Psi \right)$$

and we need to ensure that this gives $e^{iT} \cdot D_\mu \Psi$. In operator terms

$$\boxed{D'_\mu e^{iT} = e^{iT} D_\mu} \tag{44}$$

[16] In the latter case, the additional concept of spontaneous symmetry breaking and the Higgs mechanism is required.

This allows a direct expression for D'_μ

$$\boxed{D'_\mu = e^{iT} D_\mu e^{-iT}} \tag{45}$$

Compare (22) and (23).

Transformation of 1-form
The 1-form transforms as $\Gamma_\mu \mapsto \Gamma'_\mu$,

$$\boxed{\Gamma'_\mu = e^{iT}\Gamma_\mu e^{-iT} + e^{iT}\left(\partial_\mu e^{-iT}\right)} \tag{46}$$

Note that this equation reduces to (33) when the matrices commute, and $T(x) \mapsto \Theta(x)$.

Problem 8
Prove (46) from (43) and (45). §

Problem 9
Note that in general the last term in (46) is not equal to $-i\partial_\mu T$. To understand this, take a simple example:

$$T(x) = \lambda[A + f(x)B]$$

where $\lambda \ll 1$ and A, B are matrices. Expand in powers of λ and evaluate the difference

$$\Delta \;=\; e^{iT}\left(\partial_\mu e^{-iT}\right) + i\partial_\mu T$$

Show that there is a leading term proportional to

$$\lambda^2\,(\partial_\mu f)\,[A, B]$$

Thus when matrices do not commute, care is needed in working with matrix formulas. §

5.3 Curvature 2-form

In exactly the same way as before, non-trivial physics comes about through the *curvature 2-form.*

$$\boxed{R_{\nu\mu} = \partial_\nu\Gamma_\mu - \partial_\mu\Gamma_\nu + [\Gamma_\nu, \Gamma_\mu]} \tag{47}$$

which again can be expressed as

$$\boxed{R_{\nu\mu} = [D_\nu, D_\mu]} \tag{48}$$

Compare (36) and (38). Importantly, the commutator term cannot be omitted in general.

Transformation property
The 2-form transforms in a simple way:

$$\boxed{R'_{\nu\mu} = e^{iT} R_{\nu\mu} e^{-iT}} \tag{49}$$

Compare (41). In this case, we cannot change the order of matrices, and the two phase factors do not cancel. Nevertheless the following quantity is invariant under gauge transformations

$$R^2 \;\equiv\; \text{tr } R^{\nu\mu}R_{\nu\mu}$$

as well as Lorentz transformations. To prove the gauge invariance,

$$\begin{aligned}(R')^2 &= \text{tr } e^{iT}R^{\nu\mu}e^{-iT}\cdot e^{iT}R_{\nu\mu}e^{-iT}\\ &= \text{tr } e^{iT}R^{\nu\mu}R_{\nu\mu}e^{-iT}\end{aligned}$$

The two exponentials on the "outside" also cancel because the factors in a trace can be cyclically permuted: tr $ABC =$ tr BCA. Thus the invariant R^2 is a candidate for the Lagrangian density: $\mathcal{L} \sim R^2$.

5.4 Another representation

The above formulation is presented in terms of a spacetime-dependent matrix $T(x)$. It is convenient to separate the spacetime dependence from the matrix properties. (Compare (32) for separating out a matrix factor Σ.) Let T_a, $a = 1, \ldots, N_g$ be a set of $N \times N$ matrices that serve as a basis for the class under consideration. In general $N_g \le N^2$ expresses the degrees of freedom for the group of transformations. Introduce functions $\Theta^a(x)$ and $A^a_\mu(x)$ through

$$\begin{aligned}T(x) &= \Theta^a(x)\,T_a\\ \Gamma_\mu(x) &= -i\frac{g}{\hbar}A^a_\mu(x)\,T_a\\ D_\mu &= \partial_\mu - i\frac{g}{\hbar}A^a_\mu(x)\,T_a\\ (D_\mu\Psi)^j &= \partial_\mu\Psi^j - i\frac{g}{\hbar}A^a_\mu(x)\;(T_a)^j{}_k\;\Psi^k\end{aligned}$$

where a coupling constant g has been introduced in analogy to q.

The 2-form involves commutators. Consider $[T^a, T^b]$. It should be expressible as a sum over the other basis matrices, through a set of coefficients $f_{ab}{}^c$ (see Problem 12 and the related discussion in Appendix D):

$$[T_a, T_b] \;=\; if_{ab}{}^c\,T_c \tag{50}$$

Then

$$\begin{aligned}R_{\nu\mu}(x) \;=\; &\left(-i\frac{g}{\hbar}\right)\left[\partial_\nu A^a_\mu(x) - \partial_\mu A^a_\nu(x)\right]T_a\\ &+\left(-i\frac{g}{\hbar}\right)^2 A^b_\nu(x)A^c_\mu(x)\cdot(if_{bc}{}^aT_a)\end{aligned}$$

This is (a) position-dependent, and (b) a matrix. We separate out these two aspects by writing

$$R_{\nu\mu}(x) \;=\; R^a_{\nu\mu}(x)\,T_a$$

so that all the matrix aspects are carried by the constant factors T_a. If we further identify

$$R^a_{\nu\mu}(x) \quad = \quad -i\frac{g}{\hbar}F^a_{\nu\mu}(x)$$

then we have

$$\boxed{F^a_{\nu\mu} = \partial_\nu A^a_\mu - \partial_\mu A^a_\nu + \frac{g}{\hbar} f_{bc}{}^a A^b_\nu A^c_\mu} \qquad (51)$$

There are now $a = 1, \ldots, N_g$ types of field tensors, each like an EM field. But they are related to the potentials in a way that is (a) nonlinear, and (b) involving other components b, c. In other words, the Lagrangian density $\mathcal{L} \sim R^2$ represents N_g types of *interacting* vector fields.

A note on convention
First, in elementary particle physics, it is conventional to take $\hbar = 1$. Secondly, unlike EM, we do not have a ready and natural definition of the unit for measuring field and potential. So we may as well take

$$g\, A^a_\mu \quad \mapsto \quad A^a_\mu$$

so that the covariant derivative then looks simple

$$D_\mu \quad = \quad \partial_\mu - iA^a_\mu(x)T_a$$

But then

$$R^2 \quad \mapsto \quad g^{-2}\, R^2$$

Gauge theory is often written in this form.

5.5 Qualitative insights

While a proper study of the resultant theory would take us very far afield, we can at least make some qualitative remarks.

If Ψ has an internal degree of freedom that is irrelevant, then transformations $\exp iT$ can be made. In general, these matrices T do not commute, and the theory is said to be *non-abelian*. If the theory is gauged, i.e., $T \mapsto T(x)$, that would require the introduction of a covariant derivative $D_\mu = \partial_\mu + \Gamma_\mu$. Separating out the matrix character leads to N_g 4-vector potentials A^a_μ — which correspond to N_g types of photon-like particles, i.e., vector bosons. The relationship of field tensor to potential is nonlinear. This means that $\mathcal{L} \sim R^2$ has terms that go schematically as A^3 and A^4 — these photon-like particles have self-interaction and do not satisfy the law of superposition.

This set of ideas was first introduced in 1954 by Yang and Mills [4] in the context of isospin — which actually would not work because isopsin is not an exact symmetry. In the early 1970s, it was applied to the newly invented color degree of freedom for quarks; many people contributed different ingredients to this work, but the name[17] quantum chromodyanamics (QCD) in analogy to quantum electrodynamics (QED) was due to Gell-Mann. Students may like to read about the history [5] and the key ideas [6]. Students should now be able to understand the equations written on the "blackboard" in the latter article. The application to the weak interactions is another story. In short, the idea of gauge theories — as a beautiful extension of the ideas of electrodynamics as outlined in this Chapter — is now at the core of fundamental physics.

A Commutators

Operators
A linear operator $\mathcal{A}$ turns one wavefunction Ψ into another one $\mathcal{A}\Psi$. All operators considered are assumed to be *local*: if $\Psi_b = \mathcal{A}\Psi_a$, the value of $\Psi_b(x_0)$ depends only on $\Psi_a(x)$ in an infinitesimal neighborhood of x_0. Such local and linear operators are built up from two elementary operations.

- Multiplication by a function of position:

$$\Psi(x) \quad \mapsto \quad f(x)\,\Psi(x)$$

  This operator is simply denoted as $f(x)$.
- Derivative (along some direction μ):

$$\Psi(x) \quad \mapsto \quad \partial_\mu \Psi(x)$$

Product of operators
A product operator $\mathcal{C} = \mathcal{A}\mathcal{B}$ means the following

$$\mathcal{C}\Psi \quad = \quad \mathcal{A}\mathcal{B}\Psi = \mathcal{A}(\mathcal{B}\Psi) \qquad (52)$$

In other words, $\mathcal{B}$ acts first, and then $\mathcal{A}$ acts on the result. Thus, products should be read from right to left.

Commutator
Define

$$[\mathcal{A}, \mathcal{B}] \quad = \quad \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}$$

Obviously

$$\begin{aligned} [f(x), g(x)] &= 0 \\ [\partial_\mu, \partial_\nu] &= 0 \end{aligned}$$

the latter because the order of mixed derivatives does not matter.

However

$$\begin{aligned} & [\partial_\mu, f(x)]\Psi(x) \\ = \; & \partial_\mu(f(x)\Psi(x)) - f(x)\partial_\mu\Psi(x) \\ = \; & (\partial_\mu f(x))\Psi(x) \end{aligned}$$

[17] The word "chromo" means color.

In the second line we have inserted brackets according to (52). Thus

$$[\partial_\mu, f(x)] \quad = \quad \partial_\mu f(x) \tag{53}$$

B Dynamics of a point charge: an example

The main body of this Chapter focuses on general and abstract issues. In order not to lose sight of realistic problems, this Appendix studies one simple example from three points of view: Newtonian, classical Hamiltonian, quantum.

A point charge q moves in the following uniform field:

$$\mathbf{B} = B\,\mathbf{e}_z \quad , \quad \mathbf{E} = -E\,\mathbf{e}_x$$

where $E \ll Bc$. This simple problem is surprisingly rich, being related to (a) the (classical) Hall effect discovered in 1879 [9, 10] and (b) the quantum Hall effect (QHE) which has led to several Nobel Prizes: von Klitzing in 1985 for the integer QHE [11]; Laughlin, Stormer and Tsui in 1998 for the fractional QHE [12]; Thouless, Haldane and Kosterlitz in 2016 [13] for the underlying topological concepts. We shall not get to the intricacies, but everything starts with the Hamiltonian described below.

B.1 Newtonian mechanics

General motion
The equations of motion are

$$\begin{aligned} m\frac{dv_x}{dt} &= q(-E + v_y B) \\ m\frac{dv_y}{dt} &= -qv_x B \\ m\frac{dv_z}{dt} &= 0 \end{aligned}$$

The last of these allows motion in the x–y plane with $v_z = 0$. Make the substitution

$$\begin{aligned} v_x(t) &= v'_x(t) \\ v_y(t) &= U + v'_y(t) \\ U &= E/B \end{aligned}$$

The choice of U removes the constant term on the RHS of the v_x equation, giving

$$\begin{aligned} \frac{dv'_x}{dt} &= \omega v'_y \\ \frac{dv'_y}{dt} &= -\omega v'_x \\ \omega &= \frac{qB}{m} \end{aligned} \tag{54}$$

with solution (up to displacing the origin of time)

$$v'_x = v_0 \sin\omega t \quad , \quad v'_y = v_0 \cos\omega t$$

Integration gives

$$\begin{aligned} x' &= x_c - R\cos\omega t \\ y' &= y_c + R\sin\omega t \\ R &= \frac{v_0}{\omega} = \frac{mv_0}{qB} \end{aligned}$$

Thus a uniform circular motion is superimposed on a uniform linear motion of velocity U.[18] The circular motion can be ignored if averaged over time (as for the Hall effect below).

The Hall effect
A thin conducting film (dimensions $0 \le z \le b$, $0 \le x \le a$, $0 \le y \le L$, typically with $b \ll a \ll L$) carries a current $I = Jab$ in the $+y$ direction (either positive charges flowing along $+y$ or negative charges flowing along $-y$, in both cases $qv_y > 0$).

Impose the magnetic field as described, initially without any electric field. Then for either sign of q, the force in the x direction is $qv_yB > 0$ and the charge carriers drift to and accumulate at the edge $x = a$, leading to an electric field opposing further charge accumulation. Equilibrium is obtained when $E = UB$ where U is the mean v_y. Thus the (small) E is not imposed, but self-generated. See diagram in Ref. [9] for the schematic experimental configuration.[19]

The Hall voltage across the width of the film is

$$V_{\rm H} = Ea \quad = \quad UBa$$

Sign of Hall voltage
Measurement of $V_{\rm H}$ determines the sign of U and hence of q.

Magnitude of Hall voltage and the Hall resistance
From this point onwards, $V_{\rm H}$ will mean the absolute value. The current is

$$I = Jab \quad = \quad NqUab$$

Thus define a Hall resistance

$$\boxed{R_{\rm H} \equiv \frac{V_{\rm H}}{I} = \frac{B}{Nqb}} \tag{55}$$

It is often convenient (e.g., for the 2D interface between two materials) to refer to the number of charge carriers per unit *surface area*

$$N_s = Nb$$

[18] The assumption that $E \ll Bc$ ensures that this speed is not relativistic and it is correct to use Newtonian equations.

[19] The notation may be slightly different.

in term so which (55) becomes

$$\boxed{R_{\rm H} \equiv \frac{V_{\rm H}}{I} = \frac{B}{N_s q}} \tag{56}$$

$R_{\rm H}$ has the dimension of resistance. It is an unusual kind of "resistance": the voltage is across x but the current is along y. Its value is independent of the width a — which would not be the case with a "normal" resistance. In the form (56), all reference to the dimensions of the system disappears.

The following problem gives the typical orders of magnitude involved.

Problem 10
(a) Find N for silver (density 10.5 g cm^{-3}, atomic weight 107.8, 1 conduction electron per atom). (b) A silver thin film of thickness 0.1 mm is placed in a magnetic field of 2.0 T. What is $R_{\rm H}$? (c) If a current is $I = 5$ A is passed through the film, what is $V_{\rm H}$ §

Applications
The Hall effect is especially useful for semiconductors.

- Because the density of carriers is much smaller (by a factor of 1000 or more), $V_{\rm H}$ is easier to measure.
- In semiconductors, neither the sign of the carriers (electrons or holes) nor their density (dependent on doping) is known a priori, and both can be determined from $R_{\rm H}$.
- For a fixed thin film and fixed current, $V_{\rm H} \propto B$, so magnetic fields can be measured via *Hall probes*.

B.2 Using classical Hamiltonian

The potentials are

$$\Phi = Ex \quad , \quad \mathbf{A} = Bx\,\hat{\mathbf{e}}_y \tag{57}$$

the latter being *one possible* choice. The Hamiltonian is (ignoring the z variable)

$$\begin{aligned} & H = H(x, y, p_x, p_y) \\ = \;& \frac{1}{2m}\left\{(p_x - qA_x)^2 + (p_y - qA_y)^2\right\} + q\Phi \\ = \;& \frac{1}{2m}\left\{p_x^2 + (p_y - qBx)^2\right\} + qEx \end{aligned} \tag{58}$$

Since y does not appear in H,

$$p_y(t) \;=\; P = \text{ const}$$

Putting this into H yields an effective 1D problem:

$$\begin{aligned} \tilde{H} \;&=\; \tilde{H}(x, p_x) = \frac{1}{2m}p_x^2 + V(x) \\ V(x) \;&=\; \frac{q^2B^2}{2m}x^2 + q(E - PB/m)x \\ &=\; \frac{q^2B^2}{2m}(x - x_c)^2 + \text{ const} \end{aligned} \tag{59}$$

with x_c chosen to reproduce the x term in the previous line. A term $p_y^2/(2m)$ with nothing to do with the x sector is dropped.[20]

Thus we have a harmonic oscillator centered about x_c, with "force constant"

$$K \;=\; \frac{q^2B^2}{m}$$

The frequency $\omega = \sqrt{K/m}$ agreees with (54). The motion is, for example,

$$x(t) - x_c \;=\; -R\cos\omega t$$

For the y degree of freedom, go back to (58). Then

$$\begin{aligned} \frac{dy}{dt} = \frac{\partial H}{\partial p_y} &= \frac{1}{m}(P - qBx) \\ &= (P/m - \omega x_c) - \omega(x - x_c) \\ &\equiv U + \omega R\cos\omega t \\ y \;&=\; Ut + y_c + R\sin\omega t \end{aligned}$$

agreeing with the Newtonian result.

You might like to try solving the problem in another gauge.

B.3 Using quantum mechanics*

**This Section goes beyond classical EM, and can be skipped.*

Only a brief sketch will be given for the main features of the quantum version. There are many references, and we largely follow Ref. [14].

The object is to solve for the eigenfunctions from H as given in (58), but with $p_i \mapsto -i\hbar\partial_i$. We begin by reviewing an auxiliary problem.

Harmonic oscillator
For a harmonic oscillator with Hamiltonian

$$H_0 \;=\; \frac{1}{2m}p_x^2 + \frac{K}{2}x^2$$

the eigenvalues are

$$E_n = (n + \tfrac{1}{2})\hbar\omega \quad , \quad \omega^2 = K/m$$

Students can look up the eigenfunctions ψ_n; here we only cite the widths in terms of $\langle x^2\rangle$, easily obtained by equating the mean potential energy to $(1/2)E_n$, giving

$$\langle x^2\rangle \;=\; (n + \tfrac{1}{2})\hbar/\sqrt{mK} \tag{60}$$

[20] If you keep this term, then everywhere in the calculation for x and p_x, it would be just a constant.

which is small, i.e. $O(\hbar)$, for finite n.

Factorized solution
From Section B.2, the classical p_y is conserved; thus the eigenfunctions can be labelled by eigenvalues of $p_y = -i\hbar\partial_y$:

$$-i\hbar\partial_y \Psi(x,y) \quad = \quad P\,\Psi(x,y)$$

which means the y-dependence is a plane wave $\propto \exp(iky)$, $P = \hbar k$, and the wavefunction can be factorized:

$$\Psi(x,y) \quad = \quad e^{iky}\,\psi(x)$$

where ψ is described by the 1D Hamiltonian $\tilde{H}$ in (59). Therefore (a) the energies are given by

$$E_{n,k} \quad = \quad (n+\tfrac{1}{2})\hbar\omega + \frac{\hbar^2 k^2}{2m}$$

with ω having the same meaning as in the rest of this Appendix, and (b) the wavefunction ψ are centered on x_c with a narrow width given by (60). The eigenfunction problem is in principle solved; refer to Ref. [14] for details as well as the interpretation.

The QM problem is solved in a convenient gauge, and Section 2 ensures that the physical results are independent of this choice. Moreover, the QM formalism closely parallels the treatment using the classical Hamiltonian, e.g., the conservation of p_y leads to a plane-wave dependence on y.

C Exact differentials

C.1 Conservative forces

The infinitesimal work done by a force $\mathbf{F}$ in a displacement $d\mathbf{x}$ is

$$dW \quad = \quad F_j\,dx^j \tag{61}$$

The implied sum over j corresponds to the dot product. Is it possible to define $W(\mathbf{x})$, i.e., the work done (from a reference point) to $\mathbf{x}$ that is a function of $\mathbf{x}$ alone, in a manner that is *independent of path*? That is possible only if the work done over any closed loop is zero: $\oint \mathbf{F}\cdot d\mathbf{x} = 0$. By Stokes' theorem, the above condition can be converted to

$$\boldsymbol{\nabla}\times\mathbf{F} \quad = \quad 0$$

Generalizing the idea of a curl
This statement can be broken up into two steps:

First, we can say that all antisymmetric derivatives vanish, e.g.,

$$\partial_1 F_2 - \partial_2 F_1 \quad = \quad 0 \tag{62}$$

This statement about the 1–2 plane does not mention what other directions there may be.

Secondly, invent a concept $\mathbf{C} = \text{curl}\,\mathbf{F}$, which replaces the 1-2 indices by the "remaining" index 3: $C_3 \equiv \partial_1 F_2 - \partial_2 F_1$, and (62) is equivalent to $C_\ell = 0$ for all ℓ, i.e., $\mathbf{C} = \text{curl}\,\mathbf{F} = \boldsymbol{\nabla}\times\mathbf{F} = 0$.

The second half works only in 3D (otherwise what is the "remaining" direction?) and is best *unlearnt*. We should think of the condition as the generalization of (62), namely

$$\partial_j F_k - \partial_k F_j \quad = \quad 0 \tag{63}$$

for all pairs j,k. This condition works for any number of variables $x^1,\ldots,x^N$, and does not even require these to be position coordinates. This last generalization will be illustrated in Section C.3.

Terminology
An expression such as dW in (61) is called a *differential*. If $dW = F_j dx^j$ satisfies (63), it is called an *exact differential*. In that case one can define a *function of state* $W = W(x^1,\ldots,x^N)$.

C.2 Solving PDEs

Suppose you are asked to solve the PDE

$$\frac{\partial W(x^1,x^2,\ldots,x^N)}{\partial x^j} \quad = \quad F_j(x^1,x^2,\ldots,x^N) = \text{ given}$$

A solution exists only if (63) is satisfied. Phrasing the problem this way makes it obvious that N is arbitrary, and the variables x^j can be anything.

C.3 Thermodynamics

The First Law of Thermodynamics states, in obvious notation,[21]

$$dU \quad = \quad \delta Q - \delta W = \delta Q - p\,dV$$

Beginning students may not appreciate that given the *state* of the system, e.g. the temperature T and the volume V (or any reasonable pair of variables, such as T and p), one cannot say how much heat there is in the system; a function $Q = Q(T,V)$ does not exist. The concept "heat" refers only to a *process*, i.e., a change from one state to another (say infinitesimally close by). Compare the idea of work in Section C.1.

The Second Law of Thermodynamics tells us that

$$dS \quad = \quad \frac{\delta Q}{T}$$

[21] The sign in front of δW is a matter of convention: it is minus if δW is the work done *by* the system.

(defined for reversible processes) is an exact differential and defines a function of state $S(T,V)$.

The abstract statement about a function of state leads to testable consequences, and we just give one example. From the First Law and using (T,V) as the independent variables, we have

$$\begin{aligned} dS = \frac{1}{T}\delta Q &= \frac{1}{T}(dU + p\,dV) \\ &= \frac{1}{T}\left(\frac{\partial U}{\partial T}dT + \frac{\partial U}{\partial V}dV\right) + \frac{p}{T}dV \\ &= \left(\frac{1}{T}\frac{\partial U}{\partial T}\right)dT + \left(\frac{1}{T}\frac{\partial U}{\partial V} + \frac{p}{T}\right)dV \\ &\equiv F_j\,dx^j \end{aligned}$$

where, to connect to the general discussion, we define the variables $(x^1,x^2) = (T,V)$ and F_1, F_2 are the brackets in the previous line. The condition (63) then leads to

$$\frac{\partial}{\partial V}\left(\frac{1}{T}\frac{\partial U}{\partial T}\right) = \frac{\partial}{\partial T}\left(\frac{1}{T}\frac{\partial U}{\partial V} + \frac{p}{T}\right) \quad (64)$$

A little bit of algebra gives

$$\frac{\partial}{\partial T}\left(\frac{p}{T}\right) = \frac{1}{T^2}\frac{\partial U}{\partial V} \quad (65)$$

Problem 11
(a) Derive (65) from (64).
(b) Consider one mole of ideal gas, so that $pV = RT$ (R = gas constant). Show that the LHS of (65) is zero.
(c) Hence show that the internal energy of an ideal gas depends only on T. This result, well known from the kinetic theory of gases, is here derived without invoking any microscopic model. §

Notation
It is conventional to denote exact differentials (e.g., dV, dT, dU, dS) with the symbol d and other differentials (e.g., $\delta Q, \delta W$) with the symbol δ.

D Symmetry groups

Symmetry groups
Let there be a set of field

$$\Psi = (\Psi^1, \ldots, \Psi^N)$$

regarded as a vector. Suppose there are linear transformations that leave the physics invariant. A linear transformation can always be expressed as multiplication by a matrix g. Invariance means that the mapping

$$\Psi \mapsto g\,\Psi$$

takes a solution to another solution. At the level of action, the condition is

$$\mathcal{L}(\Psi) = \mathcal{L}(g\Psi)$$

The set of all such g's form a *group*, called the symmetry group of the theory.

Unitary groups
In physics, a reasonable condition is that the length of the vector Ψ must be preserved. Then g must be *unitary* in the sense

$$g^\dagger g = 1$$

This group is called the *unitary group* in N dimensions, denoted as $U(N)$.

A unitary matrix can be written as

$$g = \exp iT$$

where T is hermitian: $T^\dagger = T$. This is the matrix generalization of the theorem that a complex number z satisfying $z^*z = 1$ can be expressed as $z = \exp it$ where t is real, i.e., $t^* = t$.

The corresponding algebra
Thus the possible group elements g can be enumerated by finding all the possible T's — which form a linear space over the real numbers. A convenient basis is provided by the following T_a, $a = 0, \ldots, N^2-1$:

- The identity, conventionally labelled as T_0 (up to a normalization).
- Another $N-1$ matrices which are diagonal and traceless.
- $N(N-1)/2$ real symmetric matrices, with $+1$ in the jk and kj elements, with $j \neq k$.
- $N(N-1)/2$ imaginary antisymmetric matrices, with $+i$ in the jk element and $-i$ in the kj element, with $j \neq k$.

It is conventional to choose them such that

$$\mathrm{tr}\; T_a T_b = \delta_{ab}$$

so that they are orthonormal when the trace acts as the inner product in this space.

Take the case $N = 2$. There is only one matrix in each category, and they are respectively $1/2$ times $\sigma_0 = I$, σ_z, σ_x and σ_y, where σ_j are the Pauli matrices. For $N = 3$, they are conventionally chosen to be $1/2$ times the Gell-Mann matrices λ_j [8], with $\lambda_0 = I$.

Thus the most general T is

$$T = \Theta^a T_a$$

where Θ^a are real. If T is spacetime-dependent,

$$\Theta^a \mapsto \Theta^a(x)$$

With this, the general representation for g is

$$g \quad = \quad \exp i\Theta^a T_a$$

It turns out to be enough to study those group elements close to the identity

$$g \quad \approx \quad 1 + i\Theta^a T_a$$

Thus the linear space of all T's defines the infinitesimal transformations, and the T's are called the *generators* of the group. The set of generators has a richer structure than just a linear space; this set is closed under commutation in the following sense, and is said to constitute a *Lie algebra*.

By the way, for the group of rotations in 3D space, the corresponding Lie algebra is that of the angular momentum operators.

Problem 12
Show that if T and T' are hermitian matrices, then so is $-i[T, T']$. §

Structure constants
Because of the above result, the commutator (with the factor of $-i$) is also a member of the linear space, and can therefore be written as a linear combination over the basis

$$[T_a, T_b] \quad = \quad i f_{ab}{}^c T_c \tag{66}$$

where the real numbers $f_{ab}{}^c$ are called *structure constants* of the Lie algebra.

Henceforth assume the basis to be orthonormal, in which case it is conventional to write all three indices as subscripts, and

$$f_{abc} \quad = \quad -i \,\mathrm{tr}\, \{[T_a, T_b]\, T_c\} \tag{67}$$

Problem 13
Take the case $N = 2$ and show that the structure constants are simply the totally antisymmetric symbol. §

Problem 14
It turns out that, in general (i.e., not restricted to $N = 2$) the structure constants share many of the properties of the totally antisymmetric symbol: antisymmetry, invariance under cyclic permutations. Prove these properties using (67). §

The unitary group in 1 dimension
The most important example is $U(1)$ — a unitary 1×1 matrix is just a complex number of modulus unit: $e^{i\Theta}$. There is only one generator $T = 1$. Such phase transformations lead to EM. In other words:

EM is the gauge theory for $U(1)$.

Special unitary groups
Consider any $N > 1$. For any group element $g = \exp iT$, we can write

$$T \quad = \quad \alpha I + T'$$

where T' is traceless. Then there is the decomposition

$$g = e^{i\alpha} \cdot g' \quad , \quad g' = e^{iT'}$$

The two factors have nothing to do with each other, in the sense that if

$$g_j \quad = \quad e^{i\alpha_j} \cdot g'_j$$

and $g = \prod_j g_j$, then

$$\alpha = \sum_j \alpha_j \quad , \quad e^{iT'} = \prod_j e^{iT'_j}$$

Thus we can consider the two factors separately; in particular each subset of transformations forms a group by itself. In physics, this means the two factors need not both be symmetries (local or not), and even if they are both symmetries and moreover gauged, need not have the same coupling constant. The first factor is a $U(1)$ group.

The second factor has the property that

$$\det\, g' \quad = \quad \exp(i \,\mathrm{tr}\, T') = 1$$

The group of such unitary matrices with determinant unity is called the *special unitary group* in N dimensions, denoted as $SU(N)$.

Applications
The most important examples are the following.

Isospin is an approximate $SU(2)$ symmetry. Roughly speaking, it corresponds to interchanging two flavors of quarks, u and d. Interchange of three flavors of light quarks u, d, s leads to an approximate $SU(3)$ symmetry that was important in the classification of elementary particles. Now that we know there are six flavors of quarks (the additional three being quite heavy), one can think about an $SU(6)$ group that interchanges them. But this symmetry is very badly broken, because the quark masses are so different. These symmetries are not exact, and cannot be gauged.

In our present context, the most important example is the interchange of three quark colors. This leads to an exact $SU(3)$ of color. Moreover, this group is gauged (possible only because it is an exact symmetry), and the concepts introduced in this Chapter is the foundation of QCD. In QCD, there are $3^2 - 1 = 8$ vector fields, and the corresponding photon-like particles are called gluons.

Particles (or more precisely their left-handed components) form doublets, whose members are interchanged in the weak interactions. An example of such a doublet is (ν_e, e). This leads to an $SU(2)$, whose gauge theory describes the weak interactions.[22] There are $2^2 - 1 = 3$ gauge fields that mediate the weak interactions. They are $W^{\pm}$ and Z^0, which are massive because of spontaneous symmetry breaking and the Higgs mechanism.

To summarize many of the above points,

The Standard Model of fundamental interactions is the gauge theory corresponding to $SU(3) \times SU(2) \times U(1)$.

The first factor refers to color; the second refers to the weak interactions and the last refers to EM. All the interactions come from the gauge principle.

References

[1] FJ Dyson, *The scientist as rebel*, The New York Review of Books, (2006), p. 13

[2] K Young, "Foreign exchange market as a lattice gauge theory", Am. J. Phys., **67**, 864 (1999).

[3] The Nobel Prize in Physics, 1965.
https://www.nobelprize.org/prizes/physics/1965/summary/

[4] CN Yang and R Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance", Phys. Rev. **96**, 19 (1954).
DOI: 10.1103/PhysRev.96.191

[5] H Fritzsch, "The history of QCD", CERN COURIER Sep 27, 2012.
http://cerncourier.com/cws/article/cern/50796

[6] F Wilczek, "QCD made simple", Physics Today, August 2000.
http://frankwilczek.com/Wilczek_Easy_Pieces/298_QCD_Made_Simple.pdf

[7] M Schwartz, "Lecture Notes for 253a: QFT1 Lecture 9: Scalar QED", Harvard University (2009). Read on 1 May 2016.
http://isites.harvard.edu/fs/docs/icb.topic624321.files/09-scalarQED.pdf

[8] Wikipedia, "Gell-Mann matrices".
https://en.wikipedia.org/wiki/Gell-Mann_matrices
Read on 1 May 2016. (Wikipedia is not recommended as a reference unless the content is checked to be correct, which it is in this case.)

[9] Hall Effect.
http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/Hall.html

[10] The Hall Effect.
https://www.tf.uni-kiel.de/matwis/amat/mw2_ge/kap_2/backbone/r2_1_3.html

[11] The Nobel Prize in Physics, 1985.
https://www.nobelprize.org/prizes/physics/1985/press-release/

[12] The Nobel Prize in Physics, 1998.
https://www.nobelprize.org/prizes/physics/1998/press-release/

[13] The Nobel Prize in Physics, 2016.
https://www.nobelprize.org/prizes/physics/2016/press-release/

[14] D Tong, "The quantum hall effeect", TIFR Infosys Lectures (2016).
http://www.damtp.cam.ac.uk/user/tong/qhe.html

[22] But here, one needs to deal with the further complication of spontaneous symmetry breaking.

$$\Psi(x) \mapsto \exp i\Theta \quad \cdot \Psi(x)$$

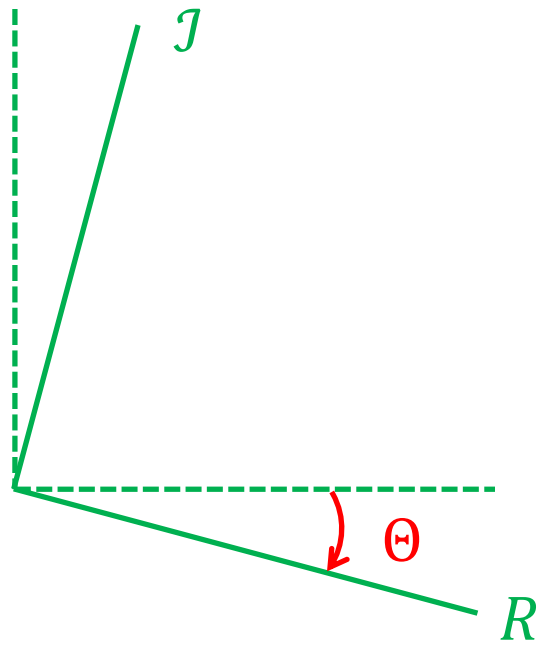


Figure 1. Rotate axes backwards

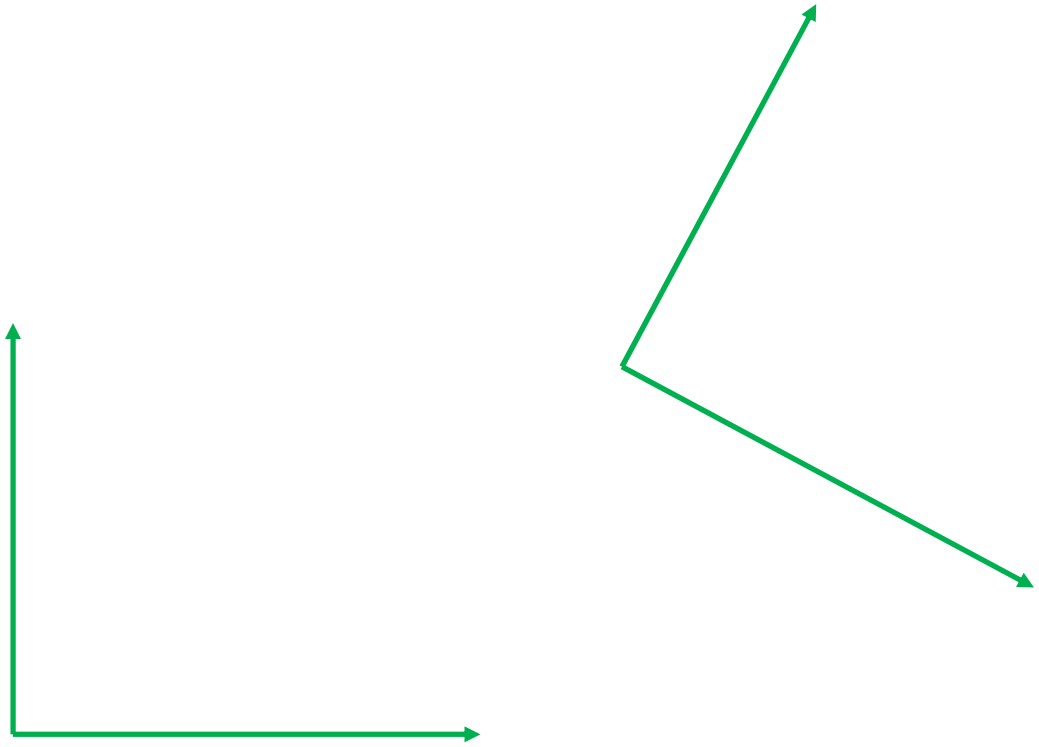


Figure 2. Basis vectors may vary with position

Figure 3a

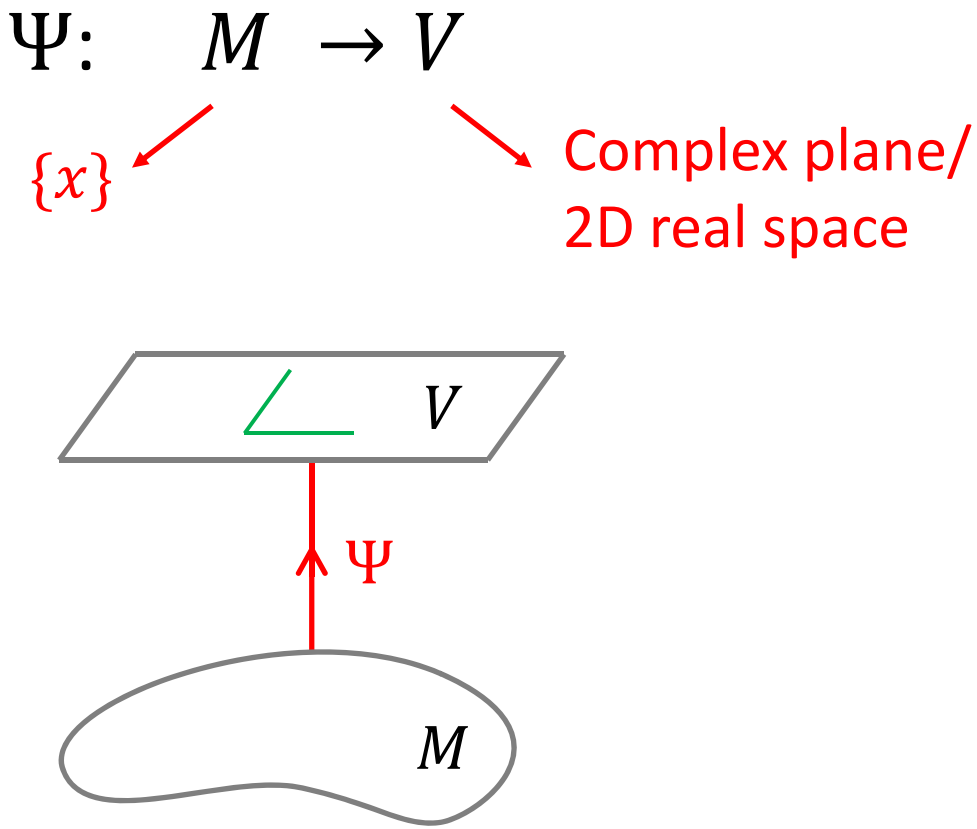
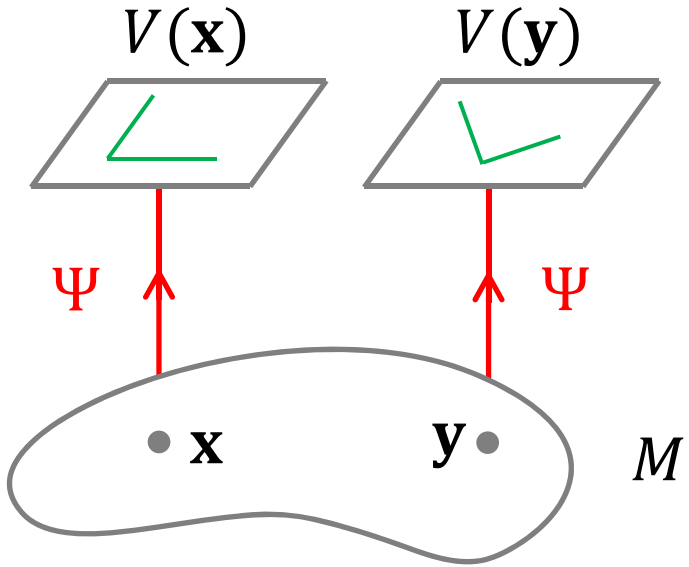


Figure 3b

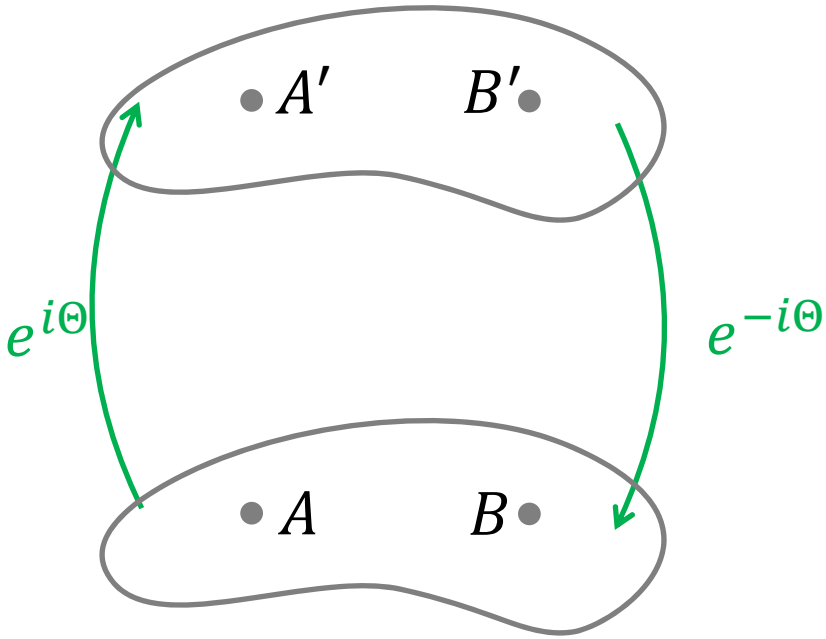
$$\Psi: \quad M \rightarrow V(\mathbf{x})$$



Γ connects $V(\mathbf{x})$ and $V(\mathbf{x} + d\mathbf{x})$

Figure 4

$$D'_\mu = e^{i\Theta}\, D_\mu\, e^{-i\Theta}$$



$D'_\mu \sim B' - A'$

$D_\mu \sim B - A$