8 Self-adjointness

Definition 26. A densely-defined operator H is said to be *self-adjoint*, if $H = H^*$ (i.e. $\mathrm{Dom}(H) = \mathrm{Dom}(H^*)$).

A self-adjoint operator is symmetric. But a symmetric H may fail to be self-adjoint due to the domain of H^* being too large, thereby failing the symmetry condition, (7.6). Then H^* can have imaginary eigenvalues, and this has undesirable consequences for spectrum of H, e.g., Exercise 7.8.

So we really want to find an intermediate symmetric extension $\tilde{H}$ of H, leading to a corresponding restriction $\tilde{H}^*$ of H^*, satisfying

$$H \subset \tilde{H} = \tilde{H}^* \subset H^*.$$

Such a $\tilde{H}$ is a *self-adjoint extension* of H. The basic question is *existence* and *uniqueness* of self-adjoint extensions.

The following is a basic criterion for self-adjointness.

Proposition 8.1. *A symmetric operator H on $\mathcal{H}$ is self-adjoint iff H is closed and* $\ker(H^* \mp i) = 0$ *iff* $\mathrm{Ran}(H \pm i) = \mathcal{H}$.

Proof. A self-adjoint $H = H^*$ is closed (Prop. 7.3); it cannot have $\pm i$ eigenvalues, since

$$H^*\psi = H\psi = \pm i\psi \;\Rightarrow\; \pm i\langle\psi|\psi\rangle = \langle\psi|H\psi\rangle = \langle H\psi|\psi\rangle = \mp i\langle\psi|\psi\rangle \;\Rightarrow \psi = 0.$$

Now suppose H is closed and H^* has no $\pm i$-eigenvalue. If $0 \neq \psi \in \mathrm{Range}(H \pm i)^\perp$, then for all $\varphi \in \mathrm{Dom}(H)$,

$$\langle\psi|(H \pm i)\varphi\rangle = 0 \;\Rightarrow\; \psi \in \mathrm{Dom}(H^*) \text{ with } 0 = (H^* \mp i)\psi,$$

contradicting the assumption. Thus $\mathrm{Ran}(H \pm i)$ must be dense in $\mathcal{H}$. We also have

$$\begin{aligned} ||(H \pm i)\psi||^2 &= ||H\psi||^2 + \langle H\psi| \pm i\psi\rangle + \langle \pm i\psi|H\psi\rangle + ||\psi||^2 \\ &= ||H\psi||^2 + ||\psi||^2, \qquad\qquad \forall\, \psi \in \mathrm{Dom}(H). \end{aligned}$$

So if $\xi = \lim(H \pm i)\psi_n$ for some sequence $\psi_n \in \mathrm{Dom}(H)$, then ψ_n and $H\psi_n$ converge. Because H is closed, $\lim \psi_n \in \mathrm{Dom}(H)$ and $H(\lim \psi_n) = \lim H\psi_n$,

thus $(H \pm i)(\lim \psi_n) = \lim(H \pm i)\psi_n = \xi$. This shows that $\mathrm{Ran}(H \pm i)$ is closed.

Finally, assume $\mathrm{Range}(H \pm i) = \mathcal{H}$. So given any $\varphi \in \mathrm{Dom}(H^*)$, we can write $(H^* \mp i)\varphi = (H \mp i)\psi_\pm$ for some $\psi_\pm \in \mathrm{Dom}(H)$. Since $\mathrm{Dom}(H) \subset \mathrm{Dom}(H^*)$, we actually have

$$\begin{aligned}(H^* \mp i)(\varphi - \psi_\mp) = 0 &\Leftrightarrow \varphi - \psi_\mp \in \ker(H^* \mp i) \\ &\Leftrightarrow \varphi - \psi_\mp \in \mathrm{Range}(H \pm i)^\perp = 0 \quad \text{(Lemma 7.2)}.\end{aligned}$$

Thus $\varphi = \psi_\mp \in \mathrm{Dom}(H)$, i.e., $\mathrm{Dom}(H^*) \subset \mathrm{Dom}(H)$. So H and H^* actually have the same domains, meaning that H is self-adjoint. □

Since any self-adjoint extension is closed, it is enough to investigate the closed symmetric extensions, and check which ones are self-adjoint. An obvious candidate is the closure.

Lemma 8.2. *A symmetric operator is closable, with symmetric closure.*

Proof. Let H be symmetric. Note that $\mathrm{Dom}(H^*)$ is dense, since it contains the dense domain of H. Prop. 7.3 says that H is closable. The second statement is left as an exercise. □

Definition 27. A symmetric operator H is *essentially self-adjoint*, if $\overline{H}$ is self-adjoint.

Lemma 8.3. *A symmetric operator H is essentially self-adjoint iff H^* is self-adjoint iff $H^* = \overline{H}$. In this case, H has a unique self-adjoint extension, given by its closure.*

Proof. Let H be symmetric. So it is closable by Lemma 8.2, and $\overline{H} = H^{**}$ holds (see Prop. 7.3). The essential self-adjointness condition is $\overline{H} = \overline{H}^*$, and this occurs iff

$$H^{**} = \overline{H} = \overline{H}^* = H^{***} = \overline{H^*} = H^*,$$

i.e., iff H^* is self-adjoint iff $\overline{H} = H^*$.

Let H be essentially self-adjoint, and let T be any self-adjoint extension of H. In particular, T is a closed extension of H, so

$$T \supset \overline{H}.$$

But T is also self-adjoint, so (recalling (7.2))

$$T = T^* \subset \overline{H}^* = \overline{H}$$

as well. Thus we must have equality,

$$T = \overline{H} \quad (= H^*).$$

□

Usually, H is initially defined on a nice domain $\mathrm{Dom}(H)$ for which the symmetric condition is easily checked. Then, provided H is known to be essentially self-adjoint, we can simply refer to Lemma 8.3 and work with $\overline{H}$ when we want to utilize all the structure theory of self-adjoint operators. Unfortunately, essential self-adjointness is generally difficult to prove, and often false. A similar argument to Prop. 8.1 gives criteria for essential self-adjointness:

Proposition 8.4. *A symmetric operator H is essentially self-adjoint iff* $\ker(H^* \mp i) = 0$ *iff* $\mathrm{Ran}(H \pm i)$ *is dense.*

Proof. Exercise. □

8.1 Self-adjoint position and momentum operators

For an open subset Ω of $\mathbb{R}^d$, we write $C_c^\infty(\Omega)$ for the space of smooth functions $\Omega \to \mathbb{C}$ with compact support. We have

$$C_c^\infty(\mathbb{R}^d) \subset \mathscr{S}(\mathbb{R}^d) \subset L^p(\mathbb{R}^d),$$

and a standard fact is that $C_c^\infty(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$, $1 \le p \le \infty$.

Definition 28. On $L^2(\mathbb{R})$ (with Lebesgue measure), the *position operator* is

$$\begin{aligned} \mathrm{Dom}(\hat{x}) &= \left\{ \psi \in L^2(\mathbb{R}) \,:\, \int_{\mathbb{R}} |x\psi(x)|^2 < \infty \right\} \\ (\hat{x}\psi)(x) &= x\psi(x). \end{aligned}$$

The *momentum operator* is

$$\begin{aligned} \mathrm{Dom}(\hat{p}) &= H^1(\mathbb{R}) \equiv \left\{ \psi \in L^2(\mathbb{R}) \,:\, \psi' \in L^2(\mathbb{R}) \right\} \\ \hat{p}\psi &= -i\psi', \end{aligned}$$

where ψ' means the *weak derivative*.

Remark. Let Ω be an open subset of $\mathbb{R}$. We recall that a locally integrable function $f \in L^1_{\text{loc}}(\Omega)$ has *weak derivative* $f' \in L^1_{\text{loc}}(\Omega)$ if the following holds for all $g \in C_c^\infty(\Omega)$,

$$\int_{\mathbb{R}} f'g = -\int_{\mathbb{R}} f\frac{dg}{dx}.$$

Of course, if f is smooth, then f' is the usual $\frac{df}{dx}$. But, for example, the function $x \mapsto |x|$ has weak derivative being the (a.e. defined) sign function.

Theorem 8.5. *The position operator $\hat{x}$ and the momentum operator $\hat{p}$ are self-adjoint. Both operators are essentially self-adjoint when restricted to either $C_c^\infty(\mathbb{R})$ or $\mathscr{S}(\mathbb{R})$.*

Proof. Certainly $\hat{x}$ is densely-defined (its domain contains $\mathscr{S}(\mathbb{R})$), and symmetric. By definition,

$$\varphi \in \text{Dom}(\hat{x}^*) \;\Leftrightarrow\; \exists \eta \in L^2(\mathbb{R}) \,:\, \int_{\mathbb{R}} \big(\overline{\varphi(x)}x - \overline{\eta(x)}\big)\psi(x)\,dx = 0 \quad \forall \psi \in \text{Dom}(\hat{x}).$$

As $\text{Dom}(\hat{x})$ is dense, this forces $x\varphi(x) = \eta(x)$ for almost every $x \in \mathbb{R}$. So

$$\text{Dom}(\hat{x}^*) = \left\{\varphi \in L^2 \,:\, \int_{\mathbb{R}} |x\varphi(x)|^2 < \infty\right\} = \text{Dom}(\hat{x}).$$

As for $\hat{x}|_{C_c^\infty(\mathbb{R})}$ or $\hat{x}|_{\mathscr{S}(\mathbb{R})}$, they are symmetric, and the above arguments show that in each case, the adjoint is the self-adjoint operator $\hat{x}$. By Lemma 8.3, $\hat{x}|_{C_c^\infty(\mathbb{R})}$ and $\hat{x}|_{\mathscr{S}(\mathbb{R})}$ are essentially self-adjoint.

As for $\hat{p}$, let us first consider

$$\tilde{p} := \hat{p}|_{C_c^\infty(\mathbb{R})},$$

which is easily seen to be symmetric using integration-by-parts. By definition, $\varphi \in \text{Dom}(\tilde{p}^*)$ iff there exists $\eta \in L^2(\mathbb{R})$ such that

$$\int_{\mathbb{R}} \overline{\eta}\psi = \int_{\mathbb{R}} \overline{\varphi}\tilde{p}(\psi) = -i\int_{\mathbb{R}} \overline{\varphi}\frac{d\psi}{dx}, \qquad \forall \psi \in C_c^\infty(\mathbb{R}).$$

This is the condition that φ admits an L^2 weak derivative (namely, η). Thus

$$\tilde{p}^* = \hat{p}.$$

Now,

$$\ker(\tilde{p}^* \pm i) = \ker(\hat{p} \pm i) = \{\varphi \in \mathrm{Dom}(\hat{p}) \,:\, -i\varphi' \pm i\varphi = 0\}\,.$$

The equation $-\varphi' \pm \varphi = 0$ would be solved by $\varphi(x) = e^{\pm x}\cdot$constant, but this solution is inadmissible as it is not in $L^2(\mathbb{R})$. So $\ker(\tilde{p}^* \pm i) = 0$. By Prop. 8.4, $\tilde{p}$ is essentially self-adjoint. Then $\hat{p} = \tilde{p}^*$ is the (unique) self-adjoint extension of $\tilde{p}$.

Similarly,

$$\check{p} := \hat{p}|_{\mathscr{S}(\mathbb{R})}$$

is symmetric, and it follows from $\tilde{p} \subset \check{p} \subset \hat{p} = \tilde{p}^*$ that $\check{p}^* = \hat{p}$. So $\check{p}$ is essentially self-adjoint by Lemma 8.3. □

8.2 Reality of spectrum

Why must we be fussy about self-adjointness?

First, we check that self-adjoint operators have real spectrum.

Proposition 8.6. *Let H be a (possibly unbounded) self-adjoint operator. Then $\sigma(H) \subset \mathbb{R}$.*

Proof. Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$, and write $\lambda = a + ib$, with $b \neq 0$. We first show that $H - \bar{\lambda}$ has closed range.

Since H is symmetric,

$$\begin{aligned} ||(H-\bar{\lambda})\psi||^2 &= ||(H-a)\psi||^2 - ib\langle\psi|(H-a)\psi\rangle + ib\langle(H-a)\psi|\psi\rangle + b^2||\psi||^2 \\ &\geq b^2||\psi||^2 \end{aligned} \tag{8.1}$$

holds for all $\psi \in \mathrm{Dom}(H)$. Suppose $(H - \bar{\lambda})\psi_n \to \Psi$ converges (with $\psi_n \in \mathrm{Dom}(H)$). Eq. (8.1) shows that ψ_n is Cauchy, thus convergent to some $\psi \in \mathcal{H}$. Since H is closed, $\psi \in \mathrm{Dom}(H)$ and $(H - \bar{\lambda})\psi = \Psi$. We conclude that $\mathrm{Ran}(H - \bar{\lambda})$ is a closed subspace of $\mathcal{H}$.

Together with Lemma 7.2, we have

$$H - \bar{\lambda} \text{ surjective iff } \mathrm{Ran}(H-\bar{\lambda})^\perp = \ker(H^* - \lambda) = \{0\}. \qquad (\lambda \in \mathbb{C}\setminus\mathbb{R}) \tag{8.2}$$

Injectivity of $H - \bar{\lambda}$ is automatic, since H is symmetric and can only have real eigenvalues. We deduce that

$$\bar{\lambda} \in \rho(H) \;\text{ iff }\; \dim\ker(H^* - \lambda) = 0. \qquad (\lambda \in \mathbb{C}\setminus\mathbb{R}) \tag{8.3}$$

Finally, because of self-adjointness, $H^* = H$, the right side of (8.3) is verified for all $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Thus $\mathbb{C} \setminus \mathbb{R} \subset \rho(H)$, which is equivalent to $\sigma(H) \subset \mathbb{R}$. $\square$

Does Prop. 8.6 hold if H is a closed symmetric operator, which is not self-adjoint? The proof works only up to (8.3). We already know from Prop. 8.1 that H^* has imaginary eigenvalues. This has a catastrophic consequence for the spectrum of H.

Theorem 8.7. *Let H be a closed symmetric operator. Then, $\sigma(H)$ is either the closed upper-half plane, closed lower-half plane, all of $\mathbb{C}$, or a subset of $\mathbb{R}$. The last case occurs iff H is self-adjoint.*

Proof. (Optional.) Let $\lambda = a + ib \in \mathbb{C} \setminus \mathbb{R}$. We will prove that $\dim \ker(H^* - \lambda)$ is a stable quantity.

Consider a perturbation $\lambda + \eta$, with $|\eta| \ll |b|$ to ensure that $\lambda + \eta$ remains in the same upper/lower half-plane as λ. Take any unit vector $\varphi \in \ker(H^* - (\lambda + \eta))$, and suppose $\varphi \in \ker(H^* - \lambda)^\perp$ as well. By Eq. (8.2), we may write $\varphi = (H - \overline{\lambda})\psi$ for some $\psi \in \mathrm{Dom}(H)$, and we have

$$0 = \langle \underbrace{(H^* - (\lambda + \eta))\varphi}_{0} | \psi \rangle = \langle \varphi | \underbrace{(H - \overline{\lambda})\psi}_{\varphi} \rangle - \overline{\eta} \langle \varphi | \psi \rangle = \underbrace{||\varphi||^2}_{1} - \overline{\eta} \langle \varphi | \psi \rangle. \quad (8.4)$$

By Eq. (8.1), we also have

$$||\psi|| \leq ||(H - \overline{\lambda})\psi||/|b| = ||\varphi||/|b| = 1/|b| \ll 1/|\eta|,$$

so ψ is too small for the right-side of Eq. (8.4) to vanish (Cauchy–Schwarz). This contradiction shows that

$$\ker(H^* - (\lambda + \eta)) \cap \ker(H^* - \lambda)^\perp = 0, \qquad \forall\, |\eta| \ll |b|.$$

It follows that

$$\dim \ker(H^* - (\lambda + \eta)) \leq \dim \ker(H^* - \lambda).$$

By swapping the roles of $\lambda + \eta$ and λ, we see that $\dim \ker(H^* - \lambda)$ is locally constant for λ in the upper/lower half-plane. By connectedness, $\dim \ker(H^* - \lambda)$ is constant on the upper/lower half-planes.

Putting this last result into (8.3), we learn that the whole upper/lower half plane lies in either $\rho(H)$ or $\sigma(H)$, according to whether $\dim \ker(H^* \mp i) = 0$

or not. Since $\sigma(H)$ is closed, it can only be one of the four given options. Furthermore, the fourth option, $\sigma(H) \subset \mathbb{R}$, occurs iff $\dim\ker(H^* \mp i) = 0$, iff H is self-adjoint, by Prop. 8.1. $\square$

Theorem 8.7 highlights why symmetric operators are not good enough for spectral theory. The next result shows that self-adjoint operators have well-behaved (real) spectrum.

Theorem 8.8. *Let $H = H^*$ be a (possibly unbounded) self-adjoint operator. Then $\sigma(H) = \sigma_{\mathrm{ap}}(H)$.*

Proof. Let $\lambda \notin \sigma_{\mathrm{ap}}(H)$. In particular, λ is not an eigenvalue of H, so $\lambda - H$ is injective. We just need to show that $\lambda - H$ is surjective, then $(\lambda - H)^{-1} : \mathcal{H} \to \mathrm{Dom}(H)$ exists and $\lambda \notin \sigma(H)$ is deduced.

We first check the density of $\mathrm{Ran}(\lambda - H)$. Suppose for a contradiction, that there exists a non-zero $\varphi \in \mathrm{Ran}(\lambda - H)^{\perp}$, i.e.,

$$0 = \langle\varphi|(\lambda - H)\psi\rangle = \langle\bar{\lambda}\varphi|\psi\rangle - \langle\varphi|H\psi\rangle, \qquad \forall\psi \in \mathrm{Dom}(H).$$

This means that $\varphi \in \mathrm{Dom}(H^*) = \mathrm{Dom}(H)$, and $\bar{\lambda}\varphi = H^*\varphi = H\varphi$. Thus $\bar{\lambda} \in \sigma_{\mathrm{p}}(H)$. As H is symmetric, we have $\lambda = \bar{\lambda} \in \sigma_{\mathrm{p}}(H) \subset \sigma_{\mathrm{ap}}(H)$, which is a contradiction.

(Compare Prop. 6.3 proof.) Since $\lambda - H$ has dense range, for any $\psi \in \mathcal{H}$, we can find a sequence ψ_n in $\mathrm{Dom}(H) = \mathrm{Dom}(\lambda - H)$ such that $(\lambda - H)\psi_n \to \psi$. Note that $\lambda - H$ is bounded below by some $\alpha > 0$ (by definition of $\lambda \notin \sigma_{\mathrm{ap}}(H)$). Thus

$$||\psi_n - \psi_m|| \leq \alpha^{-1}||(\lambda - H)\psi_n - (\lambda - H)\psi_m|| \to 0.$$

So $\{\psi_n\}$ is Cauchy, and converges to some $\tilde{\psi} \in \mathcal{H}$. Now, $(\lambda - H)$ is a closed operator, so $\tilde{\psi} \in \mathrm{Dom}(\lambda - H)$ and $(\lambda - H)\tilde{\psi} = \psi$. Thus $\lambda - H$ is surjective as claimed. $\square$

Exercise 8.1. Does the self-adjoint position operator $\hat{x}$ (Definition 28) have any eigenvalues? Compute its spectrum, and check that each spectral value is an approximate eigenvalue.

Exercise 8.2. Give an example of an operator on some $\mathbb{C}^N$ which has eigenvalues lying within $\mathbb{R}$, but whose eigenspaces do not sum up to the identity operator.

8.3 Self-adjoint extensions

Let us give names to the obstructions to (essential) self-adjointness provided in Prop. 8.1–8.4.

Definition 29. Let H be a symmetric operator on $\mathcal{H}$. Its *deficiency subspaces* and (possibly infinite) *deficiency indices* are

$$\mathcal{H}_\pm := \ker(H^* \mp i), \qquad n_\pm := \dim \mathcal{H}_\pm.$$

Example 8.1. [Anomalous momentum on half-line] Let us attempt to define a momentum operator P on the half-line $\mathbb{R}_+ = [0,\infty)$. As usual, the initial domain is

$$\mathrm{Dom}(P) = C_c^\infty(0,\infty).$$

The Dirichlet condition at $x = 0$ ensures that $P = -i\frac{d}{dx}$ is a symmetric operator. Let us compute P^*. We have

$$\begin{aligned}\mathrm{Dom}(P^*) &= \left\{\varphi \in L^2(\mathbb{R}_+) \,:\, \exists \eta \in L^2(\mathbb{R}_+) \text{ with } \int_0^\infty \overline{\eta}\psi = \int_0^\infty \overline{\varphi}\left(-i\tfrac{d\psi}{dx}\right) \;\; \forall \psi \in C_c^\infty(0,\infty)\right\} \\ &= \left\{\varphi \in L^2(\mathbb{R}_+) \,:\, \varphi' \in L^2(\mathbb{R}_+)\right\}, \end{aligned} \tag{8.5}$$

with $P^*\varphi = -i\varphi'$ the weak derivative (times $-i$). The deficiency subspaces are

$$\mathcal{H}_+ = \mathrm{span}_{\mathbb{C}}\{e^{-x}\}, \qquad \mathcal{H}_- = \{0\}.$$

Since $n_+ \neq 0$, we know that P is not essentially self-adjoint.

For the closure $\overline{P} = P^{**}$, we have

$$\mathrm{Dom}(\overline{P}) = \left\{\xi \in L^2(\mathbb{R}_+) \,:\, \exists \eta \in L^2(\mathbb{R}_+) \text{ with } \int_0^\infty \overline{\eta}\varphi = \int_0^\infty \overline{\xi}(-i\varphi') \;\; \forall \varphi \in \mathrm{Dom}(P^*)\right\}.$$

Consideration of $\varphi \in C_c^\infty(0,\infty)$ shows that $\xi \in \mathrm{Dom}(\overline{P})$ has L^2 weak derivative and $\overline{P}\xi = -i\xi'$. But we also need to consider general $\varphi \in \mathrm{Dom}(P^*)$, which also has L^2 weak derivative by (8.5). So ξ, φ are locally absolutely continuous (see, e.g., 10.9 of Ziemer's Modern Real Analysis). We may therefore integrate-by-parts,

$$\begin{aligned}\langle \xi | P^*\varphi\rangle &= \int_0^\infty \overline{\xi}(-i\varphi') \\ &= i\overline{\xi}(0)\varphi(0) + \underbrace{i\int_0^\infty \overline{\xi'}\varphi}_{\langle \overline{P}\xi|\varphi\rangle} \qquad \xi \in \mathrm{Dom}(\overline{P}), \varphi \in \mathrm{Dom}(P^*). \end{aligned} \tag{8.6}$$

Since $\varphi(0)$ could be arbitrary, it follows that $\xi(0) = 0$. To summarize,

$$\mathrm{Dom}(\overline{P}) = \left\{\xi \in L^2(\mathbb{R}_+) \,:\, \xi' \in L^2(\mathbb{R}_+),\ \xi(0) = 0\right\}.$$

Let us consider any domain $\mathrm{Dom}(\tilde{P})$ lying between $\mathrm{Dom}(\overline{P})$ and $\mathrm{Dom}(P^*)$. The integration-by-parts calculation will lead, in particular, to

$$\langle \xi | \tilde{P}\xi \rangle = i|\xi(0)|^2 + \langle \tilde{P}\xi | \xi \rangle, \qquad \xi \in \mathrm{Dom}(\tilde{P}).$$

In order for $\tilde{P}$ to remain symmetric, we have to enforce $\xi(0) = 0$, yielding $\mathrm{Dom}(\tilde{P}) = \mathrm{Dom}(\overline{P})$. But we already know that $\mathrm{Dom}(\overline{P})$ is not self-adjoint.

We conclude that there does not exist any self-adjoint extension of P.

Exercise 8.3. Let us return to the problem of defining a self-adjoint momentum operator on the interval $[0, 2\pi]$. As before, the initial domain is $\mathrm{Dom}(P) = C_c^\infty(0, 2\pi)$, on which $P = -i\frac{d}{dx}$ is symmetric. The adjoint P^* is the weak (or almost-everywhere) derivative operator on the domain

$$\mathrm{Dom}(P^*) = \mathrm{AC}[0, 2\pi].$$

- What are the eigenfunctions of P? Do they span all of $L^2[0, 2\pi]$?
- Compute the deficiency indices of P, and deduce that P is not essentially self-adjoint.
- What is the domain of $\overline{P}$?
- Check that

$$\begin{aligned} \mathrm{Dom}(P_\alpha) &= \{\psi \in \mathrm{AC}[0, 2\pi] \,:\, \psi(2\pi) = \alpha\psi(0)\}, \\ P_\alpha &= P^*|_{\mathrm{Dom}(P_\alpha)}, \qquad\qquad \alpha \in \mathrm{U}(1), \end{aligned} \tag{8.7}$$

  define self-adjoint extensions of P.
- Are P_α and $P_{\alpha'}$ unitarily equivalent for distinct α, α'?

There is a systematic classification and construction of self-adjoint extensions of closed symmetric operators H, due to von Neumann. It is summarized below, and we refer to Reed–Simon Theorem X.2 for a proof.

1. H admits self-adjoint extensions iff its deficiency indices n_+, n_- are equal.

2. In case H has $n_+ = n_-$, its self-adjoint extensions are parametrized by unitary maps $U : \mathcal{H}_+ \to \mathcal{H}_-$,

$$\begin{aligned} \mathrm{Dom}(H_U) &:= \{\psi + \psi_+ + U\psi_+ \,:\, \psi \in \mathrm{Dom}(H),\ \psi_+ \in \mathcal{H}_+\}\,, \\ H_U(\psi + \psi_+ + U\psi_+) &:= H\psi + i\psi_+ - iU\psi_+. \end{aligned}$$

Definition 30. An map $C : \mathcal{H} \to \mathcal{H}$ is *antilinear* if $C(\psi + \alpha\varphi) = C\psi + \bar{\alpha}C\varphi$. An antilinear map C is called a *complex conjugation* if it is norm-preserving and $C^2 = 1$.

There is an easy sufficient criterion for the *existence* of self-adjoint extensions.

Exercise 8.4. Let H be a symmetric operator. Suppose there exists a complex conjugation C such that $C : \mathrm{Dom}(H) \to \mathrm{Dom}(H)$ and $CH = HC$. Deduce that H has equal deficiency indices (thus it admits self-adjoint extensions).

Example 8.2. For a Hamiltonian of the form $-\nabla^2 + V$, where V is a *real*-valued polynomial, it is easy to see that it is symmetric on the domain of Schwartz functions. These are the non-magnetic Schrödinger operators. Clearly, the operation of pointwise complex-conjugation provides a C such that Exercise 8.4 applies.

Example 8.3. For the operator $\tilde{p} = -i\frac{d}{dx}$ on $C_c^\infty(\mathbb{R})$, we can consider the conjugation

$$C\psi(x) = \overline{\psi(-x)}, \qquad \psi \in C_c^\infty(\mathbb{R}),$$

and apply Exercise 8.4 to deduce that it admits self-adjoint extensions. However, this conjugation does not work on $C_c^\infty(\mathbb{R}_+)$.

9 Preview of gauge theory and Aharonov–Bohm effect

9.1 Momentum on a circle

Exercise 8.3 shows that there is no unique notion of "momentum operator" on a compact interval. The different possibilities are usually interpreted as different boundary conditions. However, this is a rather naïve view. The issue is that the boundary condition $\psi(2\pi) = \alpha\psi(0)$ for P_α is *non-local*. How could such a constraint be physical, if the interval is very large?

One way out of the non-local condition is to compactify $[0, 2\pi]$ into a unit circle S^1 by identifying $x = 0$ with $x = 2\pi$. In other words, we are considering momentum operators on a circle, and we use the (local) angular coordinate θ instead of x. What we called boundary conditions are now discontinuities in ψ.

Write $\alpha = e^{-2\pi ik}$, where $k \in \mathbb{R}$ is determined up to an integer. Observe that the α discontinuity can be fixed by the following modification of ψ,

$$\tilde{\psi}(x) = (u_k\psi)(x) = e^{ikx}\psi(x). \tag{9.1}$$

The map u_k above is a unitary multiplication operator. We also have

$$u_k P_\alpha u_k^{-1} = u_k \left(-i\frac{d}{d\theta}\right) u_k^{-1} = -i\left(\frac{d}{d\theta} - ik\right) = -i\frac{d}{d\theta} - k.$$

So P_α can be unitarily transformed to a shifted momentum operator acting on a domain of *continuous* functions.

The rest of this section may not be comprehensible until after we learn about gauge theory. But it gives some idea of what lies ahead, as well as some motivation from quantum physics.

From the gauge theory perspective, what is happening is the following. There is a connection ∇ on a Hermitian line bundle $\mathcal{L}$ over S^1, telling us what parallel transport means. Starting from $x = 0$, we can obtain a normalized section s_0 defined over $(0, 2\pi)$ which is parallel, or "constant", with respect to ∇. (Normalized means $|s_0(x)| = 1$). Generally, $s_0(0)$ and $s_0(2\pi)$ do not coincide. Rather,

$$s_0(2\pi) = \alpha^{-1}s_0(0), \tag{9.2}$$

with $\alpha^{-1} = e^{2\pi ik} \in \mathrm{U}(1)$ the *holonomy* of ∇.

It is natural to use s_0 as a reference local section. Any other section s is obtained from s_0 by multiplying by some $\mathbb{C}$-valued "local wavefunction" $\psi_s : (0, 2\pi) \to \mathrm{U}(1)$,

$$s = \psi_s \cdot s_0.$$

This is completely analogous to describing a vector (field) as a combination, [basis (field), component (function)]. It is important to remember that the wavefunction ψ_s representing s depends on the choice of reference section s_0.

The notion of parallel transport provided by ∇ allows us to make sense of differentiation of sections s, independently of how we choose to describe it locally as a wavefunction ψ_s. Let us make this concrete. The Leibniz rule for ∇ is

$$\nabla s \equiv \nabla(\psi_s \cdot s_0) = \psi_s \cdot \nabla s_0 + \tfrac{d\psi_s}{d\theta} \cdot s_0.$$

Since we have chosen s_0 to be ∇-parallel, i.e. $\nabla s_0 = 0$, the above reduces to

$$\tfrac{d\psi_s}{d\theta} \cdot s_0.$$

In other words, $-i\nabla$ is implemented, at the level of representative wavefunctions, as ordinary differentiation, $-i\frac{d}{d\theta}$.

Now, we really want to restrict $-i\nabla$ to a domain of *globally continuous* L^2-sections s. With respect to s_0, the wavefunctions representing such continuous sections must be *dis*continuous in the following way,

$$\psi_s(2\pi) = \alpha\psi_s(0),$$

to compensate for (9.2).

Therefore, the *description* of $-i\nabla$ with respect to the reference local section s_0 is precisely the operator P_α we encountered in (8.7). The discontinuity of wavefunctions $\psi_s \in \mathrm{Dom}(P_\alpha)$ is only *apparent* — it is an artifact of the failure of the ∇-parallel reference section s_0 to extend to a globally continuous section.

Let us switch the reference section to

$$s_k = u_k^{-1} \cdot s_0, \qquad u_k(x) = e^{ikx},$$

and note that s_k is no longer ∇-parallel. Since

$$s_k(2\pi) = u_k^{-1}(2\pi)s_0(2\pi) = e^{-2\pi ik}e^{2\pi ik}s_0(0) = s_k(0),$$

s_k is a globally smooth section over S^1. With respect to s_k, a general globally smooth section s is

$$s = s_k \cdot \tilde{\psi}_s, \qquad \tilde{\psi}_s \in C^\infty(S^1).$$

Note that

$$\psi_s \cdot s_0 = s = \tilde{\psi}_s \cdot s_k = \tilde{\psi}_s u_k^{-1} \cdot s_0,$$

so

$$\tilde{\psi}_s = u_k \psi_s.$$

This is precisely what (9.1) was about. The unitary u_k conjugates P_α to the operator

$$u_k P_\alpha u_k^{-1} = -i\frac{d}{d\theta} - k = -i\left(\frac{d}{d\theta} - ik\right) \tag{9.3}$$
$$\mathrm{Dom}(u_k P_\alpha u_k^{-1}) = u_k \cdot \mathrm{Dom}(P_\alpha) = \{\tilde{\psi} \in \mathrm{AC}[0, 2\pi] \,:\, \tilde{\psi}(2\pi) = \tilde{\psi}(0)\}.$$

With regards to the $\mathbb{Z}$ ambiguity in the choice of k, let us mention that for each nonzero integer n,

$$u_{k+n} P_\alpha u_{k+n}^{-1} = u_k P_\alpha u_k^{-1} - n,$$

acting on the same domain of periodic absolutely continuous functions.

From such arguments, we deduce that for each $n \in \mathbb{Z}$, the operators P_α and $P_\alpha - n$ are unitarily equivalent, via u_n. This implies that the spectrum of P_α is invariant under integer shifts. The operators u_n are called *large gauge transformations*. Whereas u_k is sometimes called a *singular* or *local* gauge transformation.

9.2 Local U(1) phase freedom in quantum mechanics

Although one often calls normalized elements ψ of some Hilbert space $L^2(X, \mu)$ "quantum states", only real-valued *transition probabilities*,

$$|\langle\psi|\varphi\rangle|^2 = |\langle\psi|\varphi\rangle_{L^2(X)}|^2 = \Big|\int_X \overline{\psi(x)}\varphi(x)\, d\mu(x)\Big|^2 \tag{9.4}$$

are observable. The probabilistic interpretation of $|\langle\psi|\varphi\rangle|^2$ is called the *Born rule*. The number $\langle\psi|\varphi\rangle$ is called a *transition amplitude*.

The absolute value in Eq. (9.4) tells us that we may independently multiply ψ and φ by some U(1) phases,

$$\psi \mapsto \alpha \cdot \psi, \quad \varphi \mapsto \beta \cdot \varphi, \qquad \alpha, \beta \in \mathrm{U}(1),$$

without affecting the transition probability. So the transition probabilities depend only on the equivalence classes in the *projective* Hilbert space. Note that when we take a *superposition*,

$$\frac{1}{\sqrt{2}}(\psi + \varphi),$$

a relative phase change in the linear combination,

$$\frac{1}{\sqrt{2}}(\alpha\psi + \beta\varphi), \qquad \beta/\alpha \neq 1$$

changes the projective Hilbert space element.

The situation is more interesting once we take the spatial information of X into account. Observe that at each point $x \in X$, we are actually free to multiply the $\psi(x)$ and $\varphi(x)$ by a *common* U(1)-valued phase *function* $u(x)$, without changing the transition amplitude in Eq. (9.4). This follows from the pointwise sesquilinearity of the L^2-inner product. The pointwise phase information, i.e., $\frac{\psi}{|\psi|} : X \to \mathrm{U}(1)$, is not directly observable.

How, then, should we understand ψ modulo the above "phase redundancy"? If we make the phase information totally redundant, and just work with $|\psi|$, what remains is the real-valued *probability density* of ψ,

$$\rho_\psi : X \to \mathbb{R}_{\geq 0}, \qquad \rho_\psi(x) = |\psi(x)|^2.$$

Note that ρ_ψ integrates to 1,

$$\int_X \rho_\psi = ||\psi||^2_{L^2(X)} = 1,$$

as required for a probabilistic interpretation of ψ. Indeed, for any measurable $Z \subset X$, the integral $\int_Z \rho_\psi$ is the probability of finding the particle in the set Z, when its quantum state is ψ. The phase information is indeed redundant for this purpose.

However, general transition probabilities will generally change if we simply replace ψ, φ by $|\psi|, |\varphi|$,

$$|\langle\psi|\varphi\rangle| \neq \big|\big\langle |\psi| \,\big|\, |\varphi| \big\rangle\big|.$$

The "phase redundancy" is, more precisely, in a *simultaneous* redefinition of all wavefunctions ψ by any phase function $u : X \to \mathrm{U}(1)$,

$$\psi \rightsquigarrow u \cdot \psi, \qquad \forall \psi \in L^2(X).$$

This operation is an abelian unitary *gauge transformation*. We saw some examples in Section 9.1.

9.3 Aharonov–Bohm effect

With Section 9.2 in mind, let us proceed to explain how the momentum operators on S^1 studied in Section 9.1 arises in quantum mechanics, via the *Aharonov–Bohm* effect.

Let us imagine S^1 as the unit circle in the horizontal plane $\mathbb{R}^2$. A magnetic field with total flux $2\pi k \in \mathbb{R}$ is introduced through the "hole" in S^1, perpendicularly to the plane. For concreteness, let the magnetic field vector field $\mathbf{B} = B(x, y)\, \partial_z$ be supported on the disc D_ϵ of radius $\epsilon < 1$. We have the flux integral

$$\int_{D_1} \mathbf{B} \cdot d\mathbf{S} = \int_{D_\epsilon} B\, dxdy = 2\pi k.$$

Magnetic fields are divergence-free (Gauss' Law). Since $\mathbb{R}^2$ is contractible, there exists a globally defined *magnetic vector potential* $\mathbf{A}$ with

$$\operatorname{curl} \mathbf{A} = \mathbf{B}.$$

The form of such an $\mathbf{A}$ could be very complicated, particularly in the region inside S^1.

A classical charged particle is only influenced by $\mathbf{B}$ via the Lorentz force law, not by $\mathbf{A}$. This makes sense given the above indeterminacy of $\mathbf{A}$. Furthermore, only the value of $\mathbf{B}$ at the particle's position matters. We are free to use locally-defined $\mathbf{A}_\beta$ with domains $U_\beta \subset \mathbb{R}^2$, as long as all the U_β cover $\mathbb{R}^2$, and

$$\operatorname{curl} \mathbf{A}_\beta(\mathbf{x}) = \mathbf{B}(\mathbf{x}), \qquad \forall \mathbf{x} \in U_\beta,\ \forall\, \beta. \tag{9.5}$$

There is also no canonical choice of $\mathbf{A}_\beta$ for any given domain U_β. Indeed, we can add to $\mathbf{A}_\beta$ the gradient of any function $\Lambda : U_\beta \to \mathbb{R}$,

$$\mathbf{A}_\beta \rightsquigarrow \mathbf{A}_\beta + \nabla\Lambda, \tag{9.6}$$

whilst maintaining (9.5). This is called a *local gauge freedom* in classical electromagnetism.

Suppose the domain U_β of $\mathbf{A}_\beta$ contains D_ϵ. Then for any loop ℓ winding around D_ϵ once (anticlockwise), Stokes' theorem enforces the following value for the loop integral,

$$\oint_\ell \mathbf{A}_\beta = \int_{D_\epsilon} \operatorname{curl} \mathbf{A}_\beta = \int_{D_\epsilon} \mathbf{B} \cdot d\mathbf{S} = 2\pi k, \qquad \forall\, \beta \,:\, U_\beta \supset D_\epsilon. \tag{9.7}$$

For domains not containing D_ϵ, the above constraint need not apply.

For example, on the region excluding D_ϵ, we are allowed to use

$$\mathbf{A}_{\text{zero}}(x, y) = 0, \qquad x^2 + y^2 > \epsilon^2$$

to reproduce $\operatorname{curl} \mathbf{A}_{\text{zero}} = 0 = \mathbf{B}$ there. Note that $\mathbf{A}_{\text{zero}}$ *does not* satisfy (9.7),

$$\oint_\ell \mathbf{A}_{\text{zero}} \neq 2\pi k.$$

Another example on the same domain, is

$$\mathbf{A}_{\text{out}}(x, y) = \left(-\frac{ky}{x^2 + y^2}, \frac{kx}{x^2 + y^2} \right), \qquad x^2 + y^2 > \epsilon^2.$$

It satisfies $\operatorname{curl} \mathbf{A}_{\text{out}} = \mathbf{B}$ (Exercise), and *does* satisfy (9.7),

$$\oint_\ell \mathbf{A}_{\text{out}} = 2\pi k.$$

Classically, whether or not $\mathbf{A}_\beta$ satisfies (9.7) is irrelevant; their only role is to reproduce the magnetic field on their domains of definition. So both $\mathbf{A}_{\text{out}}$ and $\mathbf{A}_{\text{zero}}$ are admissible as magnetic vector potentials.

However, it seems that $\mathbf{A}_{\text{zero}}$ is, in some sense, "incompatible" with all those $\mathbf{A}_\beta$ with $U_\beta \supset D_\epsilon$, because $\mathbf{A}_{\text{zero}}$ does not share the constraint (9.7). Whereas $\mathbf{A}_{\text{out}}$ is "compatible". The "compatible" local vector potentials $\mathbf{A}_\beta$ must be describing some object which encodes more information than just the curl, $\mathbf{B}$. Let us tentatively call this object ∇.

Before explaining what ∇ is, let us discuss investigate the difference between $\mathbf{A}_{\text{out}}$ and $\mathbf{A}_{\text{zero}}$ more closely. We can certainly write

$$\mathbf{A}_{\text{out}} = \operatorname{grad}\big(k \tan^{-1}(y/x)\big), \qquad x^2 + y^2 > \epsilon^2,\ x \neq 0, \tag{9.8}$$

away from a *branch cut*. Without performing the cut, $\mathbf{A}_{\text{out}}$ and $\mathbf{A}_{\text{zero}}$ are gauge *in*equivalent over any subdomain containing a full loop. Such gauge inequivalence of curl-free vector potentials is possible when the domain is not simply-connected. This indicates that $\mathbf{A}_{\text{zero}}$ and $\mathbf{A}_{\text{out}}$ are actually describing *distinct* gauge-independent objects, ∇^0 and ∇.

Let us now reveal what ∇ is. First, one passes between the local vector field $\mathbf{A}_\beta$ and the corresponding locally differential 1-form $\mathcal{A}_\beta$ using the Riemannian metric on $\mathbb{R}^2$. The equivalence under local gauge transformations, (9.6), is understood as the following statement. The 1-forms $\mathcal{A}_\beta$ or $-i\mathcal{A}_\beta$ are local connection 1-forms describing a *globally*-defined connection ∇ on a principal fibre bundle with abelian gauge group $(\mathbb{R}_{>0}, \times)$ or U(1). Each $\mathcal{A}_\beta$ describes ∇ numerically, with respect to a *local gauge choice* over U_β. The magnetic field $\mathbf{B}$ is recovered as the curvature of ∇, which can be calculated as $\mathbf{B} = \operatorname{curl} \mathbf{A}_\beta$ for any choice of $\mathcal{A}_\beta$. The key point is that a connection ∇ provides strictly more information than just its curvature $\mathbf{B}$. Namely, it provides the notion of *parallel transport* for the fibre bundle. In particular, the exponentiated loop integral (9.7),

$$e^{-\oint_\ell \mathcal{A}_\beta} \in \mathbb{R}_{>0} \qquad \text{or} \qquad e^{i\oint_\ell \mathcal{A}_\beta} \in \mathrm{U}(1) \tag{9.9}$$

computes the *holonomy* of the ∇-parallel transport around the loop ℓ, and is independent of the choice of $\mathcal{A}_\beta$ used to describe ∇. A word of caution: if we want to break up the loop integral into contributions from different $\mathcal{A}_\beta$, we must take into account the fact that the different $\mathcal{A}_\beta$ are generally referring to different local gauge choices on the overlaps.

Now, for classical electromagnetism on its own, the fibre bundle and connection formalism is superfluous. Only the curvature $\mathbf{B}$ matters, and further properties of the connection ∇ are irrelevant. Indeed, no physical candidate for the principal bundle or line bundle (thus ∇) is specified. Whether the gauge group is $(\mathbb{R}_{>0}, \times)$ or U(1) is also indeterminate, and there is no observable corresponding to the holonomies in (9.9).

In 1918, Hermann Weyl introduced gauge theory ideas, for the $(\mathbb{R}_{>0}, \times)$ case, with the proposal that there is a principal $(\mathbb{R}_{>0}, \times)$-bundle over spacetime, encoding local metric *rescalings*. The electromagnetic potentials $\mathcal{A}_\beta$ would now find a role here, as the local connection 1-forms for a connection on this gauge bundle. However, Weyl's original idea was rejected by Einstein, Pauli, and others, on empirical grounds, and remained merely a piece of beautiful mathematics in search of physical meaning.

As it turns out, Weyl was far ahead of his time. As quantum mechanics developed in the 1920s, it became apparent that the *quantum* charged particle's wave function had a local U(1) gauge freedom in its phases (see Section 9.2). Quantum mechanics automatically comes with a principal U(1)-bundle of gauge choices for the wavefunctions!

So Weyl's idea, now with $-i\mathcal{A}$ rather than $\mathcal{A}$, fits perfectly with quantum mechanics. Thus, $-i\mathcal{A}_\beta$ would be the local connection 1-forms describing a U(1)-connection ∇ with respect to local gauge choices. The curvature of ∇ gives the electromagnetic fields as usual. Furthermore, ∇ tells us how the U(1) phases should be parallel transported, *independently of local gauge choices*. In turn, we obtain a gauge-independent way to differentiate the quantum wavesections, circumventing the apparent issue of phase ambiguities.

> ... In my opinion the origin and necessity for the electromagnetic field is in the following... From the arbitrariness of the gauge-factor in ψ appears the necessity of introducing the electromagnetic potential... H. Weyl, Elektron und Gravitation. I. Z. Phys. (1929)

Remark. $\mathbf{A}_{\text{out}}$ corresponds to the 1-form

$$\mathcal{A}_{\text{out}} = k\, d\theta, \qquad x^2 + y^2 > \epsilon^2.$$

The coefficient $-ik$ of $-i\mathcal{A}_{\text{out}}$ is the same as that which appears in the "twisted" momentum operator $-i(\frac{d}{d\theta} - ik)$ on S^1, (9.3). This differential operator provides a description of the U(1)-connection ∇.

In 1959, Aharonov–Bohm[17] proposed an experiment to test the role of $\mathbf{A}_\beta$ in quantum mechanics. The basic claim was that charged quantum particles would pick up an extra phase factor $e^{iq \oint_{S^1} \mathbf{A}_\beta/\hbar} = e^{iq\Phi_B/\hbar}$, when moved around a circle S^1 with magnetic flux Φ_B piercing its centre. Here, q is the electric charge of the quantum particle being used. This was a dramatic claim:

> Suppose the AB effect is true. If we insist that vector potentials $\mathbf{A}_\beta$ are fictitious and only the magnetic field $\mathbf{B}$ is real, then motion is affected non-locally by $\mathbf{B}$, since the motion only takes place in a region with no magnetic field. This would invalidate a basic reason for introducing electromagnetic fields into physics in the first place!

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

Early experiments in the 1960s were able to detect the AB effect. By the 1980s, the effect had become well-established.

The Aharonov–Bohm effect is sometimes said to imply that magnetic vector potentials attain a "real physical status" in quantum mechanics. This is not really correct. In fact, the $\mathbf{A}_\beta$ remain gauge-dependent calculation devices, whose values are still unmeasurable in any experiment. What is fundamentally different is the *role* of $\mathbf{A}_\beta$ in quantum mechanics. Its purpose is not (merely) to describe $\mathbf{B}$, but another *geometric* object — the connection ∇. In this new capacity, $\mathbf{A}_\beta$ is sometimes called a local *gauge potential*, to be contrasted with its previous classical role as a *magnetic vector potential*. For example, $\mathbf{A}_{\text{out}}$ is a valid gauge potential for ∇, as is $\mathbf{A}_{\text{zero}}$ *with a branch cut*, in view of (9.8). However, $\mathbf{A}_{\text{zero}}$ is not a valid gauge potential for ∇.

It is the connection ∇ which affects the motion of the *quantum* charged particle *locally*, in the sense of specifying what "parallel", or "geodesic" motion means. There is no mysterious non-local action of the magnetic field! The connection entails the *availability* of local potentials $\mathbf{A}$ for its numerical description, but does not distinguish any particular choice.

The connection ∇ is a gauge-independent object. According to the *gauge principle*, only gauge-independent numerical quantities can be measured. One such quantity is the pointwise curvatures, i.e., the pointwise magnetic field $\mathbf{B}(\mathbf{x})$. There are other gauge-invariant numerical properties of ∇, e.g., the holonomies around closed loops ℓ. Concretely, a holonomy is calculated as (9.9) using a(ny) representative gauge potential $\mathbf{A}$. Even if the connection is flat over the whole loop ($\mathbf{B} = 0$ along the loop), the holonomy does not need to vanish!

A holonomy of ∇ is "seen" by a charged quantum particle in the following way. The charge q is the strength of coupling of the quantum wavefunction ψ to the connection ∇. Including physical units, the actual resulting phase picked up by ψ as it moves around a loop ℓ in a ∇-parallel manner is $e^{iq \oint_\ell \mathbf{A}/\hbar}$. This phase shift is what gets measured via interference experiments. Physicists call such holonomies *Wilson loops*.

Finally, we briefly discuss a subtle question: which connection ∇ is present in the first place? There is an infinite-dimensional space of possible connections on any given principal bundle. Ultimately, only the *gauge equivalence class* of the connection matters, as far as measurable quantities in quantum mechanics are concerned. This must be the case because the principal U(1) bundle itself only makes sense up to automorphisms (i.e. gauge equivalence).

In many situations (e.g. over a contractible space), the curvature completely determines the gauge-equivalence class of a connection with that curvature. So very often, if we know what the electromagnetic field is, the answer is simply: use any connection whose curvature is that field[18] — they are all equally valid and gauge-equivalent. However, over other topologically non-trivial spaces, the answer is more subtle, and sometimes, one resorts to considering the moduli space of *all* possible gauge-equivalence classes of compatible connections, and doing some kind of averaging procedure. Furthermore, the topology of the principal bundle may be "non-trivial" — this is the subject of *monopoles*. Finally, the connection itself can be studied as a dynamical object obeying certain equations of motion, e.g. Yang–Mills equation. This is particular interesting in the case of *non-abelian* gauge theory.

[18] A connection comes first and determines the field in a local manner. If we state this in reverse order, we would run into the non-local causality issue highlighted by the Aharonov–Bohm effect.