The Intrinsic Anomalous Hall Effect from a Geometric Viewpoint

Lewerissa, Tobias

The Intrinsic Anomalous Hall Effect from a Geometric Viewpoint

THESIS

submitted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE
in
MATHEMATICS AND PHYSICS

Author :	T.R. Lewerissa
Student ID :	s2124777
Supervisor mathematics :	Dr. P.J. Bruin
Supervisor physics :	Prof.dr. K.E. Schalm

Leiden, The Netherlands, June 14, 2024

The Intrinsic Anomalous Hall Effect from a Geometric Viewpoint

T.R. Lewerissa

Huygens-Kamerlingh Onnes Laboratory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

June 14, 2024

Abstract

We investigate the geometric Berry phase interpretation of the intrinsic anomalous Hall effect in a d-dimensional crystalline solid. It has three parts: (1) By the TKNN formula, the contribution $\sigma_{xy}^{(n)}$ of the n-th band to the interband Hall conductivity is the integral over the first Brillouin zone of $f_{\mathrm{FD}}(\epsilon(\mathbf{k}))f^{(n)}(\mathbf{k})$, where f_{FD} is the Fermi-Dirac distribution and $f^{(n)}$ can be interpreted as a local coordinate representation of the curvature of a connection on a principal $U(1)$-bundle over the d-dimensional torus T^d. (2) This connection gives rise to a notion of parallel transport on the bundle which reflects the physical time evolution of the system in the adiabatic limit. (3) In two dimensions at low temperatures, it follows that $\sigma_{xy}^{(n)}$ is a universal constant multiplied by a geometric phase acquired by the system when the wave vector traverses a loop in the Brillouin zone. A classification theorem for connections on principal G-bundles in terms of their pullbacks along local sections is proved, in order to be able to construct the connection on the $U(1)$-bundle over T^d. Finally, the Rashba model for ferromagnetism is discussed to demonstrate the necessity of time reversal symmetry breaking and spin-orbit coupling in the anomalous Hall effect.

Contents

Chapter 1

Introduction

In 1879, American physicist Edwin Hall discovered that when a current-carrying conductor is placed in a perpendicular magnetic field, its electrons are pressed to one side by the Lorentz force. This induces a voltage perpendicular to both the charge current and the magnetic field, a phenomenon now known as the *Hall effect* [1]. It can be summarized by saying that the material's *conductivity* matrix σ, which measures the linear response to an applied electric field $\mathbf{E}$ of the current density $\mathbf{j}$, is not diagonal when $\mathbf{B}$ is nonzero. An elementary computation in the Drude model shows that if $\mathbf{B} = B\hat{\mathbf{z}}$, the inverse conductivity $\rho = \sigma^{-1}$ (also called the *resistivity*) satisfies

$$\rho_{xy} = \frac{B}{ne} = R_0 B,$$

with n the charge carrier density and $R_0 \coloneqq 1/(ne)$ the *Hall coefficient* [2, Section 3.1.2]. It is difficult to overstate the importance of the discovery of the Hall effect. To this day, efforts are being made by both experimental and theoretical physicists to precisely characterize the effect in different materials and to better understand the theory behind it. It has practical applications as well: Hall sensors use the effect to measure magnetic field strengths and charge carrier densities, and are common in many industrial and consumer devices.

Only one year after his initial discovery, Edwin Hall also stumbled upon what came to be known as the *anomalous Hall effect*: in ferromagnetic materials, induced voltages can get up to ten times as large as in non-magnetic conductors [3]. In addition, the linear relation between Hall resistivity ρ_{xy} and magnetic field strength B breaks down: ρ_{xy} initially increases steeply with B, but the curve flattens out and "saturates" after a while. It was quickly discovered that both observations can be accounted for by factoring in the magnetization $\mathbf{M}$, i.e. the density of magnetic dipole moments in the material. Around 1930, the empirical relation

$$\rho_{xy} = R_0 B + R_s M$$

was established, with a second term representing the anomalous Hall contribution [4]. The precise physical origin of this second term and the nature of R_s, the *anomalous Hall coefficient*, were studied extensively for the next 70 years or so. Besides an *intrinsic contribution* depending only on the electronic band structure, two distinct disorder related microscopic scattering mechanisms were identified as potential influencing factors: *side jumps* and *skew-scattering*. For a long time, it remained unclear which should dominate [5].

The discussion shifted with the discovery of the *integer quantum Hall effect* in the '80s: in a two-dimensional conductor at low temperatures in the presence of a strong magnetic field, the Hall conductivity σ_{xy} is always (approximately) an integer multiple of the universal constant e^2/h [6]. This

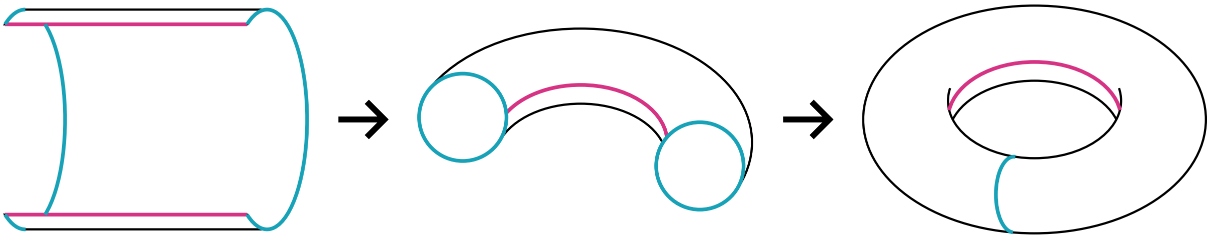


Figure 1.1: In topology, a torus T^2 is just a rectangle with opposite edges identified.

phenomenon turned out to have deep geometric and topological roots. To understand why, recall that the potential energy function of an electron in a crystalline solid is periodic with the lattice Λ describing the crystal structure. Bloch's theorem then tells us that the eigenstates of the one-electron Hamiltonian H take the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r})$$

with N the number of unit cells making up the system and $u_{n,\mathbf{k}}$ a lattice periodic *Bloch function*. The integer n is known as the *band index*, and the *wave vector* $\mathbf{k}$ can always be chosen within the *first Brillouin zone*, a unit cell of the reciprocal lattice Λ^*. For any $\mathbf{k}$ in the first Brillouin zone, the corresponding Bloch functions $u_{n,\mathbf{k}}$ are precisely the normalized eigenstates of the operator $H(\mathbf{k}) := e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}}$ acting on the space of lattice periodic functions, so the wave vector can be interpreted as a sort of parameter of the system. Opposite edges of the Brillouin zone can be identified, meaning that for a two-dimensional conductor, the parameter space takes the shape of a torus T^2.

Fixing n and assuming the Hamiltonian to be (sufficiently) nondegenerate, we thus have for each point $\mathbf{k}$ on T^2 a one-dimensional complex eigenspace of $H(\mathbf{k})$, namely the one spanned by $u_{n,\mathbf{k}}$. The physically relevant eigenstates are those of norm 1, giving us a copy of the circle $U(1) := \{z \in \mathbb{C} : |z| = 1\}$. Gluing to the torus at each of its points the corresponding copy of $U(1)$, we get what is known as a *principal* $U(1)$*-bundle* over T^2: a bundle of fibers, one for each point in T^2, which are all circles. Locally, it looks like the Cartesian product $T^2 \times U(1)$, but globally, it might be "twisted" to some extent. A lower-dimensional analog of this is the Möbius strip, shown in Figure 1.2, which is just the circle $U(1)$ with a line segment glued to every point. If we consider only the endpoints of each such line segment, we get a principal C_2-bundle over $U(1)$, because each fiber looks like the discrete group of two elements, C_2. Pick a point x in the fiber of $1 \in U(1)$, then the loop $t \mapsto \exp(2\pi it)$, $t \in [0,1]$ in $U(1)$ has a unique "lift" to the Möbius strip boundary which starts at x. The *endpoint* of that lift is precisely the other point in the fiber of 1, because the strip is twisted.

Something similar holds for our $U(1)$-bundle over the Brillouin zone: when you prepare an electron in some state $\psi_{n,\mathbf{k}}$ and have its wave vector $\mathbf{k}$ traverse a closed loop on the torus, the electron ends up in the same copy of $U(1)$ as it started in but perhaps not the same state; it might have acquired a nonzero

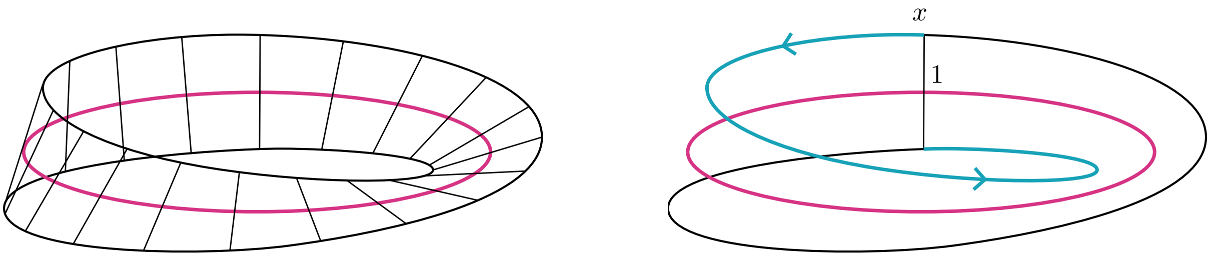


Figure 1.2: (Left) The Möbius strip can be viewed as a circle $U(1)$ with a line segment glued to every point. (Right) The lift to the boundary of the Möbius strip starting at x of the loop $t \mapsto \exp(2\pi it)$ in $U(1)$ has a different endpoint.

phase. This phase shift is measured by the *Berry connection* on the bundle, which gives rise to notions of parallel transport and path lifting. When "pulled back" along a *local section* of the bundle, which is just a smooth choice of a phase for $u_{n,\mathbf{k}}$ for all $\mathbf{k}$ in some open subset of T^2, this connection takes the form of a vector valued function $\mathbf{a}^{(n)}$ with components

$$a_\mu^{(n)}(\mathbf{k}) := -i \int \mathrm{d}\mathbf{r}\ \overline{u_{n,\mathbf{k}}(\mathbf{r})} \partial_{k_\mu} u_{n,\mathbf{k}}(\mathbf{r}),$$

where the integral runs over a single unit cell of Λ. In general, $\mathbf{a}^{(n)}$ depends on the chosen section, but

$$f^{(n)}(\mathbf{k}) := \partial_{k_x} a_y^{(n)}(\mathbf{k}) - \partial_{k_y} a_x^{(n)}(\mathbf{k})$$

does not, so f is a well-defined function on the whole T^2. It is also the pullback along any section of the *curvature* of the Berry connection, the so-called *Berry curvature*.

Now, Thouless et al. showed that when exactly M energy bands are filled, the Hall conductivity of a two-dimensional conductor at zero temperature can be written as

$$\sigma_{xy} = \frac{e^2}{2\pi h} \sum_{n=1}^{M} \int_{T^2} \mathrm{d}\mathbf{k}\ f^{(n)}(\mathbf{k}).$$

So, in words, σ_{xy} is a sum of integrals of local forms of the Berry curvature on principal $U(1)$-bundles over the first Brillouin zone [7]. Algebraic topology dictates that any such integral must in fact be an integer. Moreover, it is a *topological invariant* of the bundle, its so-called *top Chern number*, meaning that slight perturbations of the Hamiltonian leave the conductivity invariant. This expression for σ_{xy} was coined the *TKNN formula*, and it is applicable to the intrinsic anomalous Hall effect as well: the two phenomena share the same topological roots.

The main purpose of this thesis is to explore this connection between topology and the anomalous Hall effect. Most of what will be said can also be read elsewhere in the literature, but not in one place. Texts which highlight the physical side of the story often leave out a lot of the mathematics, and vice versa. This thesis is an attempt to gather in one place the essential ingredients – both physical and mathematical – needed to understand the geometric interpretation of the intrinsic anomalous Hall effect, and perhaps to provide a new viewpoint for some of them or fill in some gaps.

Chapter 2 starts with the discussion of Lie groups, of which the circle $U(1)$ and the cyclic groups C_n are examples. The reader is assumed to be familiar with basic concepts from differential geometry, including but not limited to smooth manifolds, smooth maps, vector bundles, smooth flows, differential forms and de Rham cohomology. Most prerequisite knowledge is covered in Leiden University's *Differentiable Manifolds* course and can be found in [8]; that which isn't, is listed in Appendix A. In Chapter 3, we move on to the definition of fiber bundles. Principal bundles, such as the $U(1)$-bundle over the first Brillouin zone and the Möbius strip boundary discussed above, have some particularly nice properties and deserve special attention. The theory of vector valued differential forms is developed in Chapter 4 and immediately applied in Chapter 5, where we rigorously define connections and curvatures on principal bundles and discuss a way to classify them by their pullbacks along smooth local sections. In Chapter 6, we derive the TKNN formula along with the Kubo formula for conductivity that it is based on. For this, we assume basic knowledge of quantum statistical mechanics and solid state physics; terms such such as statistical ensembles, density operators, Bloch's theorem and Brillouin zones should all sound at least somewhat familiar. Some important concepts from second quantization that the derivation relies on are listed in Appendix B. Finally, everything comes together in Chapter 7, where we examine the geometric Berry phase interpretation of the TKNN formula and discuss its implications in the Rashba model for ferromagnetism.

Conventions

Listed here are some of the conventions that are used throughout this thesis, so as to prevent confusion and make sure all symbols are interpreted the same way by both mathematicians and physicists.

- Inner products are denoted by $\langle \cdot | \cdot \rangle$ and are assumed to be conjugate linear in the *first* coordinate.
- The symbol $:=$ means "is by definition equal to", whereas $\equiv$ means "is identically equal to".
- Quantum mechanical operators such as observables and creation and annihilation operators are given a hat ($\hat{}$) to distinguish them from other symbols.
- The complex conjugate of a complex number $z \in \mathbb{C}$ is written as $\overline{z}$ and the adjoint (also known as Hermitian conjugate) of a linear operator $\hat{A}$ as $\hat{A}^*$.
- Whenever an index variable appears twice in an expression, once as a subscript and once as a superscript, summation over that index is implied. This is known as the *Einstein summation convention.*
- If ρ is a function from a set X to a (multiplicatively written) group G, we will often use the notation ρ^{-1} for the function $X \to G$, $x \mapsto \rho(x)^{-1}$. This should not be confused with the inverse function of ρ, which of course only exists when ρ is bijective.

Chapter 2

Lie groups

In this chapter we introduce Lie groups, which stand at the basis of the geometric interpretation of the TKNN formula. A *Lie group* is a topological group endowed with a smooth structure making the multiplication and inversion into smooth maps. A Lie group G can be studied via its tangent bundle TG as any other smooth manifold, but TG is special for multiple reasons. One such reason is that the tangent space T_eG to G at the identity element e has a natural *Lie algebra* structure. Furthermore, information about the Lie algebra T_eG can be transferred back to G using the canonically defined *exponential map*, which is quite a powerful tool.

Smooth group actions of Lie groups on other manifolds are crucial in many of their applications, including principal bundles, which will be discussed in Chapter 3. If a Lie group G acts smoothly on a manifold M, the exponential map induces a mapping from the Lie algebra of G to the space $\mathfrak{X}(M)$ of smooth vector fields on M, known as the *infinitesimal generator* of the action. Infinitesimal generators will play a crucial role in the treatment of connections on principal bundles in Chapter 5.

2.1 Lie groups and Lie algebras

We start with the main definitions.

Definition 2.1. A *Lie group* is a group G (usually written multiplicatively) that is also a smooth manifold, with the property that the multiplication map $G \times G \to G$, $(g,h) \mapsto gh$ and the inversion map $G \to G$, $g \mapsto g^{-1}$ are smooth.

Lie groups are examples of *topological groups*, i.e. groups with a topology with respect to which the multiplication and inversion are continuous.

Example 2.2. The relevant Lie groups for us are the following; see [8, Example 7.3] for details.

(1) The *general linear group* $\mathrm{GL}(n,\mathbb{R})$ of invertible $n \times n$ matrices over $\mathbb{R}$ is an open submanifold of the $\mathbb{R}$-vector space $\mathrm{M}(n,\mathbb{R})$ of all $n \times n$ matrices over $\mathbb{R}$, and an n^2-dimensional Lie group under matrix multiplication.

(2) Similarly, the *complex general linear group* $\mathrm{GL}(n,\mathbb{C})$ is an open submanifold of $\mathrm{M}(n,\mathbb{C})$ and a $2n^2$-dimensional Lie group under matrix multiplication.

(3) The circle $U(1) \coloneqq \{z \in \mathbb{C} : |z| = 1\}$ is a 1-dimensional Lie group under complex multiplication.

(4) Any countable group with the discrete topology is a 0-dimensional Lie group. △

The natural structure-preserving maps between Lie groups are the smooth group homomorphisms.

Definition 2.3. A *Lie group homomorphism* is a smooth group homomorphism between Lie groups.

Any element $g \in G$ of a Lie group G defines smooth left and right multiplication maps $L_g \colon G \to G$, $h \mapsto gh$ and $R_g \colon G \to G$, $h \mapsto hg$. These are diffeomorphisms, their inverses are $L_{g^{-1}}$ and $R_{g^{-1}}$. The pushforwards $\mathrm{d}(L_g), \mathrm{d}(R_g) \colon TG \to TG$ provide a way to "translate" tangent spaces to G along g. Of particular importance are those vector fields on G which are invariant under any such translation.

Definition 2.4. Let G be a Lie group. A smooth vector field $X \in \mathfrak{X}(G)$ is called *left-invariant* if it satisfies $\mathrm{d}(L_g)_h(X_h) = X_{gh}$ for all $g, h \in G$. The *Lie algebra of* G, denoted $\mathrm{Lie}(G)$, is the set of all smooth left-invariant vector fields on G.

The differential $\mathrm{d}(L_g)_h$ of L_g at h is linear for all $g, h \in G$, which implies that $\mathrm{Lie}(G)$ is a linear subspace of $\mathfrak{X}(G)$. The following lemma shows that more is true. Recall that the *Lie bracket* of two smooth vector fields $X, Y \in \mathfrak{X}(M)$ on a smooth manifold M is the vector field $[X, Y] \in \mathfrak{X}(M)$ defined by

$$[X, Y]f = X(Yf) - Y(Xf) \in C^\infty(M)$$

for any $f \in C^\infty(M)$.

Lemma 2.5. *Let G be a Lie group and $X, Y \in \mathfrak{X}(G)$ two left-invariant vector fields, then their Lie bracket $[X, Y] \in \mathfrak{X}(G)$ is also left-invariant.*

See [8, Proposition 8.33] for a proof. It follows that $\mathrm{Lie}(G)$ is a (non-associative) algebra over $\mathbb{R}$, since the Lie bracket is bilinear. The following definition is inspired by its other properties.

Definition 2.6. A *Lie algebra* is a real vector space $\mathfrak{g}$ together with a binary operation $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, $(X, Y) \mapsto [X, Y]$ satisfying the following properties for all $X, Y, Z \in \mathfrak{g}$:

(i) *bilinearity*: for any $a, b \in \mathbb{R}$,

$$[aX + bY, Z] = a[X, Z] + b[Y, Z],$$
$$[X, aY + bZ] = a[X, Y] + b[X, Z];$$

(ii) *antisymmetry*:

$$[X, Y] = -[Y, X];$$

(iii) the *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

Example 2.7. Let G be a Lie group, then $\mathrm{Lie}(G)$ is a Lie algebra under the Lie bracket. △

As a vector space, $\mathrm{Lie}(G)$ is nothing more than the tangent space to G at the identity element $e \in G$.

Lemma 2.8. *Let G be a Lie group, then the evaluation map*

$$\begin{aligned} \epsilon \colon \mathrm{Lie}(G) &\longrightarrow T_eG \\ X &\longmapsto X_e \end{aligned}$$

is a vector space isomorphism.

A proof can be found in [8, Theorem 8.37]. The result of Lemma 2.8 still holds if we replace e by an arbitrary group element $g \in G$; in fact, the composition of ϵ and the isomorphism $\mathrm{d}(L_g)_e$ is the evaluation map $\mathrm{Lie}(G) \to T_gG$ which sends $X \in \mathrm{Lie}(G)$ to X_g.

Example 2.9. Set $U \coloneqq U(1) \setminus \{-1\}$ and define $\theta \colon U \to (-\pi, \pi)$ by $e^{i\theta(z)} = z$ for all $z \in U$, then (U, θ) is a smooth chart for $U(1)$. The corresponding coordinate vector $\mathrm{d}/\mathrm{d}\theta\,|_1 \in T_1U(1)$ induces a left-invariant vector field $\mathrm{d}/\mathrm{d}\theta \coloneqq \epsilon^{-1}(\,\mathrm{d}/\mathrm{d}\theta|_1)$ on $U(1)$. It constitutes a basis for $\mathrm{Lie}(U(1))$ and is equal to the coordinate vector field on any angle coordinate chart for $U(1)$. △

An important consequence of Lemma 2.8 is that any Lie group G admits a smooth global frame of left-invariant vector fields, which implies that the space $\mathfrak{X}(G)$ of all smooth vector fields on G is generated as a $C^\infty(G)$-module by $\mathrm{Lie}(G)$. We refer to [8, Corollary 8.39] for a proof.

Lemma 2.10. *Let G be a Lie group and $X \in \mathfrak{X}(G)$ a smooth vector field on G, then X can be written as a $C^\infty(G)$-linear combination of smooth left-invariant vector fields on G.*

The relevant maps between Lie algebras are those which respect both the vector space structure and the bracket operation.

Definition 2.11. A *Lie algebra homomorphism* is a linear map $A\colon \mathfrak{g} \to \mathfrak{h}$ between Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$ such that for all $X, Y \in \mathfrak{g}$, we have $A[X,Y] = [AX, AY]$.

If $F\colon G \to H$ is a Lie group homomorphism, then via the canonical identifications $T_eG \cong \mathrm{Lie}(G)$ and $T_eH \cong \mathrm{Lie}(H)$ from Lemma 2.8, the differential $\mathrm{d}F_e$ can be interpreted as a linear map $\mathrm{Lie}(G) \to \mathrm{Lie}(H)$. The following lemma, proved in [8, Proposition 8.44], states that it is even a Lie group homomorphism.

Lemma 2.12. *Let G and H be Lie groups and $F\colon G \to H$ a Lie group homomorphism. For any smooth left-invariant vector field $X \in \mathrm{Lie}(G)$, there exists a unique $Y \in \mathrm{Lie}(H)$ such that $\mathrm{d}F_g(X_g) = Y_{F(g)}$ for all $g \in G$. Writing $Y = F_*X$, the map $F_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ so defined is a Lie algebra homomorphism.*

We call F_* the *induced Lie algebra homomorphism*.

Lemma 2.13. *Let G, H and K be Lie groups.*

(1) $(\mathrm{id}_G)_*$ *is the identity on* $\mathrm{Lie}(G)$.

(2) *Let $F_1\colon G \to H$ and $F_2\colon H \to K$ be Lie group homomorphisms, then $(F_2 \circ F_1)_* = (F_2)_* \circ (F_1)_*$*

Lemma 2.13 follows directly from the properties of the differential and can be summarized by saying that the assignments $G \mapsto \mathrm{Lie}(G)$, $F \mapsto F_*$ define a covariant functor from the category of Lie groups to the category of Lie algebras. As a consequence, if $F\colon G \to H$ is a Lie group isomorphism, the induced map $F_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ is an isomorphism of Lie algebras.

2.2 Lie group actions

Many important applications of Lie groups involve actions of Lie groups on other smooth manifolds.

Definition 2.14. Let M be a smooth manifold and G a Lie group. A *smooth left action* of G on M is a left action in the group theoretical sense that is smooth as a map $G \times M \to M$. A *smooth right action* of G on M is defined analogously.

Example 2.15. Let M be a smooth manifold, then a smooth global flow on M is the same as a smooth left action of $\mathbb{R}$ on M, where $\mathbb{R}$ is considered as a Lie group under addition. △

A smooth manifold M together with a smooth left action $\theta\colon G \times M \to M$ of a Lie group G on M is known as a *left G-manifold*. We have:

- for any $g \in G$ a smooth left multiplication map $\theta_g\colon M \to M$, $p \mapsto gp$;
- for any $p \in M$ a smooth orbit map $\theta^{(p)}\colon G \to M$, $g \mapsto gp$.

For all $g \in G$, the map θ_g is a diffeomorphism with inverse $\theta_{g^{-1}}$. *Right G-manifolds* and the corresponding orbit and right multiplication maps are defined analogously.

Definition 2.16. Let M and N be left G-manifolds and $F\colon M \to N$ a smooth map. F is said to be *G-equivariant* if for all $p \in M$ and $g \in G$, we have $F(gp) = gF(p)$. G-equivariant maps between right G-manifolds are defined analogously.

An important example is the action of a Lie group G on itself by conjugation: the map

$$\begin{aligned} \mathrm{ad}\colon G \times G &\longrightarrow G \\ (g,h) &\longmapsto ghg^{-1} \end{aligned}$$

defines a smooth left action of G on itself. For all $g \in G$, ad_g is a Lie group automorphism of G, which means that it induces a Lie algebra isomorphism

$$\begin{aligned} \mathrm{Ad}_g \coloneqq (\mathrm{ad}_g)_*\colon \mathrm{Lie}(G) &\longrightarrow \mathrm{Lie}(G) \\ X &\longmapsto \epsilon^{-1}(\mathrm{d}(\mathrm{ad}_g)_e(X_e)), \end{aligned} \tag{2.1}$$

see Lemmas 2.12 and 2.13. The map

$$\begin{aligned} \mathrm{Ad}\colon G &\longrightarrow \mathrm{GL}(\mathrm{Lie}(G)) \\ g &\longmapsto \mathrm{Ad}_g \end{aligned}$$

thus defined is often referred to as the *adjoint representation*. It is in fact a representation of G in the group theoretical sense, and a smooth one if we consider $\mathrm{GL}(\mathrm{Lie}(G))$ as an open submanifold of the vector space $\mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$ of all linear maps from $\mathrm{Lie}(G)$ to itself.

Lemma 2.17. *Let G be a Lie group, then $\mathrm{Ad}\colon G \to \mathrm{GL}(\mathrm{Lie}(G))$ is a Lie group homomorphism.*

Proof. Let $g, h \in G$, then $\mathrm{ad}_{gh} = \mathrm{ad}_g \circ \mathrm{ad}_h$, so $\mathrm{Ad}_{gh} = \mathrm{Ad}_g \circ \mathrm{Ad}_h$ by functoriality, i.e. Lemma 2.13. For smoothness, we refer to [8, Proposition 20.24]. □

Ad being a Lie group homomorphism means that it, too, induces a Lie algebra homomorphism. Under some canonical identifications, this map $\mathrm{Ad}_*\colon \mathrm{Lie}(G) \to \mathrm{Lie}(\mathrm{GL}(\mathrm{Lie}(G)))$ is essentially just the Lie bracket.

Lemma 2.18. *Let G be a Lie group, $g \in G$ and $D_{\mathrm{Ad}_g}\colon \mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G)) \to T_{\mathrm{Ad}_g} \mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$ the isomorphism from Lemma A.3, then for all $X, Y \in \mathrm{Lie}(G)$,*

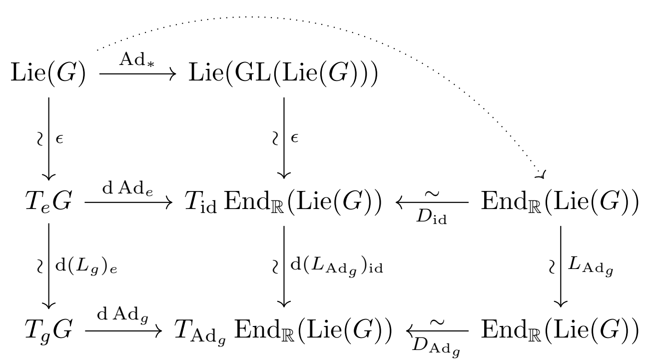
$$(D^{-1}_{\mathrm{Ad}_g} \circ \mathrm{d}\,\mathrm{Ad}_g)(X_g)Y = \mathrm{Ad}_g([X,Y]) \in \mathrm{Lie}(G),$$

where we interpret Ad as a map $G \to \mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$.

Proof. First, note that if $g = e$, the claim follows from [8, Proposition 20.25]. For the general case, observe that Lemma 2.17 implies $\mathrm{Ad} \circ L_g = L_{\mathrm{Ad}_g} \circ \mathrm{Ad}$, where L_{Ad_g} is the endomorphism of $\mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$ which sends any B to its composition $\mathrm{Ad}_g \circ B$ with Ad_g. With Lemma A.3, we now find

$$\begin{aligned} (D^{-1}_{\mathrm{Ad}_g} \circ \mathrm{d}\,\mathrm{Ad}_g)(X_g)Y &= (D^{-1}_{\mathrm{Ad}_g} \circ \mathrm{d}\,\mathrm{Ad}_g)(\mathrm{d}(L_g)_e(X_e))Y = (D^{-1}_{\mathrm{Ad}_g} \circ \mathrm{d}(\mathrm{Ad} \circ L_g)_e)(X_e)Y \\ &= (D^{-1}_{\mathrm{Ad}_g} \circ \mathrm{d}(L_{\mathrm{Ad}_g} \circ \mathrm{Ad})_e)(X_e)Y = (L_{\mathrm{Ad}_g} \circ D^{-1}_{\mathrm{id}} \circ \mathrm{d}\,\mathrm{Ad}_e)(X_e)Y \\ &= (\mathrm{Ad}_g \circ (D^{-1}_{\mathrm{id}} \circ \mathrm{d}\,\mathrm{Ad}_e)(X_e))Y = \mathrm{Ad}_g([X,Y]) \end{aligned}$$

for any $X, Y \in \mathrm{Lie}(G)$. □



The proof is summarized in the above commutative diagram, in which we have identified the tangent spaces of $\mathrm{GL}(\mathrm{Lie}(G))$ and $\mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$. The content of Proposition 20.25 in [8] is that the dotted arrow sends $X \in \mathrm{Lie}(G)$ to $[X, \cdot\,] = (Y \mapsto [X, Y]) \in \mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$.

We now introduce some notation which will be useful later. Let M and N be smooth manifolds and θ a smooth right action of a Lie group G on N. Let $F\colon M \to N$ and $\rho\colon M \to G$ be smooth maps and let $q \in N$, $g \in G$, then we define

$$\begin{aligned} F \cdot \rho &\coloneqq \theta \circ (F \times \rho)\colon M \longrightarrow N,\ p \longmapsto F(p)\rho(p), \\ F \cdot g &\coloneqq \theta_g \circ F\colon M \longrightarrow N,\ p \longmapsto F(p)g, \\ q \cdot \rho &\coloneqq \theta^{(q)} \circ \rho\colon M \longrightarrow N,\ p \longmapsto q\rho(p). \end{aligned} \tag{2.2}$$

With this notation, we get the following product rule for the differential.

Lemma 2.19. *Let $p \in M$, then*

$$\mathrm{d}(F \cdot \rho)_p = \mathrm{d}(F \cdot \rho(p))_p + \mathrm{d}(F(p) \cdot \rho)_p.$$

Proof. Define the injections $j_1\colon N \to N \times G,\ q \mapsto (q, \rho(p))$ and $j_2\colon G \to N \times G,\ g \mapsto (F(p), g)$ and let $\pi_1\colon N \times G \to N$, $\pi_2\colon N \times G \to G$ be the projections, then

$$\begin{aligned} \mathrm{d}(F \times \rho)_p &= (\mathrm{d}(j_1 \circ \pi_1)_{(F(p),\rho(p))} + \mathrm{d}(j_2 \circ \pi_2)_{(F(p),\rho(p))}) \circ \mathrm{d}(F \times \rho)_p \\ &= \mathrm{d}(j_1 \circ \pi_1 \circ (F \times \rho))_p + \mathrm{d}(j_2 \circ \pi_2 \circ (F \times \rho))_p \\ &= \mathrm{d}(j_1 \circ F)_p + \mathrm{d}(j_2 \circ \rho)_p \end{aligned}$$

by Lemma A.1, so

$$\begin{aligned} \mathrm{d}(F \cdot \rho)_p &= \mathrm{d}\theta_{(F(p),\rho(p))} \circ (\mathrm{d}(j_1 \circ F)_p + \mathrm{d}(j_2 \circ \rho)_p) \\ &= \mathrm{d}(\theta \circ j_1 \circ F)_p + \mathrm{d}(\theta \circ j_2 \circ \rho)_p \\ &= \mathrm{d}(\theta_{\rho(p)} \circ F)_p + \mathrm{d}(\theta^{(F(p))} \circ \rho)_p \\ &= \mathrm{d}(F \cdot \rho(p))_p + \mathrm{d}(F(p) \cdot \rho)_p. \end{aligned}$$

□

2.3 The exponential map

By Lemma 2.8, the Lie algebra of the circle $U(1) \subseteq \mathbb{C}^*$ is isomorphic as an $\mathbb{R}$-vector space to the tangent space $T_1U(1)$ at 1. Heuristically, $T_1U(1)$ is just a vertical line, i.e. a shifted copy of $i\mathbb{R}$. The complex exponential function $\mathbb{C} \to \mathbb{C}^*$ maps $i\mathbb{R}$ to the circle, so it can be interpreted as a map $\mathrm{Lie}(U(1)) \to U(1)$. It provides a way to go from the Lie algebra, which is a “linear approximation” of $U(1)$, back to $U(1)$.

A similar map can be constructed for general Lie groups. The construction is based on the following observation. For any $s \in \mathbb{R}$, is corresponds to $sX \in \mathrm{Lie}(U(1))$ with $X \coloneqq \mathrm{d}/\mathrm{d}\theta$ the angle coordinate vector field from Example 2.9. The map $\gamma_s\colon \mathbb{R} \to U(1),\ t \to e^{ist}$ is an integral curve of sX starting at $1 \in U(1)$, i.e. its velocity vector $\gamma_s'(t) \in T_{\gamma_s(t)}U(1)$ at $\gamma(t)$ is equal to $sX|_{\gamma(t)}$ for all $t \in \mathbb{R}$, and it is maximal since it is defined on the whole real line. We know that such maximal integral curves are unique, so $is \mapsto \gamma_s(1) = e^{is}$ gives us an alternative, intrinsic definition of the exponential map on $i\mathbb{R}$. This one we can generalize, provided that left-invariant vector fields on arbitrary Lie groups generate large enough flows.

Lemma 2.20. *Let G be a Lie group and $X \in \mathrm{Lie}(G)$, then X generates a global flow.*

A proof can be found in [8, Theorem 9.18]. Now, let G be a Lie group and denote the flow of any left-invariant vector field $X \in \mathrm{Lie}(G)$ by $\theta_{(X)}$. Lemma 2.20 states that the domain of $\theta_{(X)}$ is $\mathbb{R} \times G$ for all $X \in \operatorname{Lie}(G)$, so we can define the *exponential map* of G as

$$\begin{aligned} \exp\colon \operatorname{Lie}(G) &\longrightarrow G \\ X &\longmapsto \theta^{(e)}_{(X)}(1) = \theta_{(X)}(1, e). \end{aligned}$$

In words, exp sends $X \in \operatorname{Lie}(G)$ to the value at 1 of the maximal integral curve $\theta^{(e)}_{(X)}$ of X starting at e. The exponential map has a number of useful properties, as the next lemma shows; see [8, Proposition 20.8] for a proof.

Lemma 2.21. *Let G be a Lie group.*

(1) *The exponential map* $\exp\colon \operatorname{Lie}(G) \to G$ *is smooth.*

(2) *Let $X \in \operatorname{Lie}(G)$ and $s, t \in \mathbb{R}$, then* $\exp((s+t)X) = \exp(sX)\exp(tX)$.

(3) *Let $X \in \operatorname{Lie}(G)$, then* $\exp(-X) = \exp(X)^{-1}$.

(4) *Let $X \in \operatorname{Lie}(G)$ and $n \in \mathbb{Z}$, then* $\exp(nX) = \exp(X)^n$.

(5) *Let $X, Y \in \operatorname{Lie}(G)$ with $[X, Y] = 0$, then* $\exp(X + Y) = \exp(X)\exp(Y)$.

(6) *The differential of* exp *at* 0 *is* $\mathrm{d}\exp_0 = \epsilon \circ D_0^{-1}$ *with* $\epsilon\colon \operatorname{Lie}(G) \to T_eG$ *and* $D_0\colon \operatorname{Lie}(G) \to T_0 \operatorname{Lie}(G)$ *the isomorphisms from Lemmas 2.8 and A.3, respectively. That is,* $\mathrm{d}\exp_0$ *is the identity under the canonical identifications* $T_0 \operatorname{Lie}(G) \cong \operatorname{Lie}(G) \cong T_eG$.

(7) *Let $X \in \operatorname{Lie}(G)$, then $\theta_{(X)}(t, g) = g\exp(tX)$ for all $t \in \mathbb{R}$ and $g \in G$.*

Example 2.22.

(1) Under the identification of $\operatorname{Lie}(U(1))$ with $i\mathbb{R}$ (which maps $\mathrm{d}/\mathrm{d}\theta$ to i), the exponential map of $U(1)$ is just the complex exponential map.

(2) We know $\mathrm{GL}(n, \mathbb{R})$ is an open submanifold of $\mathrm{M}(n, \mathbb{R})$, so there are canonical $\mathbb{R}$-vector space isomorphisms $\operatorname{Lie}(\mathrm{GL}(n, \mathbb{R})) \cong T_{I_n} \mathrm{GL}(n, \mathbb{R}) \cong \mathrm{M}(n, \mathbb{R})$. Under this identification, the exponential map of $\mathrm{GL}(n, \mathbb{R})$ is the matrix exponential $A \mapsto e^A$ (see [8, Example 20.6]). △

2.4 Infinitesimal generators

Now let G be a Lie group and M a right G-manifold, and denote the right action by $\theta\colon M \times G \to M$. Following [8, pp. 525–527], any element $X \in \operatorname{Lie}(G)$ induces a map

$$\begin{aligned} \theta_{(X)}\colon \mathbb{R} \times M &\longrightarrow M \\ (t, p) &\longmapsto p\exp(tX) \end{aligned}$$

via the exponential map of G, which is a smooth global flow on M by Lemma 2.21. Denoting by $\underline{X}$ the infinitesimal generator of this flow, i.e. the smooth vector field on M whose value at $p \in M$ is the velocity at $t = 0$ of the curve $\theta^{(p)}_{(X)}$, we obtain a map $\underline{\theta}\colon \operatorname{Lie}(G) \to \mathfrak{X}(M)$ given by $\underline{\theta}(X) = \underline{X}$. We can characterize $\underline{X}$ in two different ways.

First of all, for any $p \in M$ and $f \in C^\infty(M)$,

$$\underline{X}_p f = \theta^{(p)\prime}_{(X)}(0) f = \mathrm{d}\big(\theta^{(p)}_{(X)}\big)_0 \left(\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_0 \right) f = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_0 f(p\exp(tX))$$

by definition of $\underline{X}$. Second, we have $\theta^{(p)}_{(X)} = \theta^{(p)} \circ \gamma$ with $\gamma := \exp \circ c$, where $c\colon \mathbb{R} \to \operatorname{Lie}(G)$ is defined as $c(t) := tX$. Under the identification $T_0 \operatorname{Lie}(G) \cong \operatorname{Lie}(G)$, $c'(0)$ corresponds to X, so by Lemma 2.21,

$$\gamma'(0) = \mathrm{d}\exp_0(c'(0)) = X_e.$$

We also know $\gamma(0) = p$, so

$$\underline{X}_p = \theta_{(X)}^{(p)\prime}(0) = \mathrm{d}(\theta^{(p)})_e(\gamma'(0)) = \mathrm{d}(\theta^{(p)})_e(X_e). \tag{2.3}$$

From this, it is easy to see that $\underline{\theta}$ is $\mathbb{R}$-linear. In fact, more is true.

Proposition 2.23. *The map $\underline{\theta}\colon \mathrm{Lie}(G) \to \mathfrak{X}(M)$ defined above is a Lie algebra homomorphism.*

See [8, Theorem 20.15] for a proof. We call $\underline{\theta}$ the *infinitesimal generator* of θ and for all $X \in \mathrm{Lie}(G)$, $\underline{X}$ is the *fundamental vector field* on M associated to X. The next example shows that, in some sense, fundamental vector fields are a generalization of left-invariant vector fields on Lie groups.

Example 2.24. Suppose $M = G$ and $\theta\colon G \times G \to G$ is just the group operation, so that $\theta^{(g)} = L_g$ and $\theta_g = R_g$ for any $g \in G$. Let $X \in \mathrm{Lie}(G)$, then

$$\underline{X}_g = \mathrm{d}(\theta^{(g)})_e(X_e) = \mathrm{d}(L_g)_e(X_e) = X_g$$

for any $g \in G$ since X is left-invariant, so $\underline{X} = X$. In other words, the infinitesimal generator of θ is just the inclusion $\mathrm{Lie}(G) \hookrightarrow \mathfrak{X}(G)$ in this case. △

We know from Lemma 2.8 that evaluation at any point $g \in G$ defines a vector space isomorphism $\mathrm{Lie}(G) \to T_gG$. The following lemma, adapted from [9, Corollary 27.16], generalizes this fact.

Lemma 2.25. *Let $p \in M$ and suppose p has a trivial stabilizer, then the map $\mathrm{Lie}(G) \to T_pM$, $X \mapsto \underline{X}_p$ is injective.*

Proof. Let $X \in \mathrm{Lie}(G)$ with $0 = \underline{X}_p = \mathrm{d}(\theta^{(p)})_e(X_e)$, then the constant map $\gamma\colon \mathbb{R} \to M$, $t \mapsto p$ is an integral curve of $\underline{X}$ starting at p. Since $\underline{X}$ is by definition the infinitesimal generator of the global flow $\theta_{(X)}$ on M, $\theta_{(X)}^{(p)}$ is too, so $p = \theta_{(X)}^{(p)}(t) = p\exp(tX)$ and thus $\exp(tX) = e$ for all $t \in \mathbb{R}$. We know $t \mapsto \exp(tX)$ is the maximal integral curve of X starting at e, so $X_e = 0$ and $X = 0$. □

Finally, it is worth noting that infinitesimal generators behave well under pushforwards of G-equivariant maps.

Lemma 2.26. *Let N be another smooth manifold, $\vartheta\colon N \times G \to N$ a smooth right action of G on N and $\underline{\vartheta}\colon \mathrm{Lie}(G) \to \mathfrak{X}(N)$ the corresponding infinitesimal generator. If $F\colon M \to N$ is a smooth G-equivariant map and $X \in \mathrm{Lie}(G)$, then $\underline{\theta}(X)$ and $\underline{\vartheta}(X)$ are F-related. That is, $\mathrm{d}F_p(\underline{\theta}(X)_p) = \underline{\vartheta}(X)_{F(p)}$ for any point $p \in M$.*

Proof. Let $p \in M$, then

$$\vartheta^{(F(p))}(g) = F(p)g = F(pg) = (F \circ \theta^{(p)})(g)$$

for any $g \in G$ by the G-equivariance of F, so $\vartheta^{(F(p))} = F \circ \theta^{(p)}$. Using (2.3), we find

$$\underline{\vartheta}(X)_{F(p)} = \mathrm{d}(\vartheta^{(F(p))})_e(X_e) = \mathrm{d}F_p(\mathrm{d}(\theta^{(p)})_e(X_e)) = \mathrm{d}F_p(\underline{\theta}(X)_p),$$

as required. □

Chapter 3

Fiber bundles

This chapter kicks off the discussion of fiber bundles, which play a central role in the geometric interpretation of the anomalous Hall effect. Heuristically, if M and F are smooth manifolds, a *fiber bundle* over M with *typical fiber* F is a smooth manifold E obtained by gluing to M at every point x a copy E_x of F, in such a way that E locally looks like the Cartesian product $M \times F$. That is to say, we can cover M with open sets U_α and find for each α a diffeomorphism from $\bigcup_{x \in U_\alpha} E_x$ to $U_\alpha \times F$ which "preserves the fibers". Two such fiber-preserving local diffeomorphisms differ on any copy of F by an element of $\mathrm{Diff}(F)$. If each of these elements is contained in a (finite-dimensional) Lie group $G \subseteq \mathrm{Diff}(F)$, we say that E is a *G-bundle.* Of particular importance are those G-bundles whose typical fiber F is precisely G; these so-called *principal G-bundles* are the main object of study in this chapter. Given a Lie group G and a smooth manifold M, the principal G-bundles over M can be classified via so-called *universal bundles.* This classification scheme has an intuitive physical interpretation in the context of the Hall effect, which we will return to in Chapter 6. Finally, the fiber structure of a principal G-bundle gives rise to a notion of *vertical tangent vectors*, but there is no natural horizontal analog. This leads to the definition of a connection, which will be discussed in Chapter 5.

3.1 Principal bundles

Definition 3.1. Let G be a Lie group. A *principal G-bundle* is a smooth surjection $\pi\colon P \to M$ between smooth manifolds P and M together with a smooth right action of G on P such that there exist an open cover $\{U_\alpha\}_{\alpha \in A}$ of M and for every $\alpha \in A$ a diffeomorphism $\phi_\alpha\colon \pi^{-1}U_\alpha \to U_\alpha \times G$ with the properties:

(i) ϕ_α is G-equivariant, where $U_\alpha \times G$ is considered as a right G-manifold under right multiplication in the second coordinate;

(ii) $\pi_1 \circ \phi_\alpha = \pi$, with $\pi_1\colon U_\alpha \times G \to U_\alpha$ the projection onto the first coordinate.

Such a pair (U, ϕ) of an open subset $U \subseteq M$ and a G-equivariant fiber-preserving diffeomorphism $\phi\colon \pi^{-1}U \to U \times G$ is known as a *local trivialization* for the bundle, and a collection $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ of local trivializations such that the U_α cover M is a *bundle atlas* or *G-atlas.* We call P the *total space* of the bundle, M its *base space* and π the *projection.* The projection of any principal G-bundle is a smooth submersion.

Lemma 3.2. *Let $\pi\colon P \to M$ be a principal G-bundle,*[1] *then π is a submersion.*

[1] Strictly speaking, the right G-action on P is part of the data, but we will usually leave it implicit.

Proof. Let $p \in P$ and choose a local trivialization (U, ϕ) for π with $\pi(p) \in U$, then $\pi_1 \circ \phi = \pi$ implies $\mathrm{d}\pi_p = \mathrm{d}(\pi_1)_{\phi(p)} \circ \mathrm{d}\phi_p$. We know know $\mathrm{d}\phi_p$ is an isomorphism since ϕ is a difeomorphism, and π_1 is a smooth submersion, which implies $\mathrm{d}(\pi_1)_{\phi(p)}$ is surjective. It follows that $\mathrm{d}\pi_p$ is surjective, as required. □

As a consequence, the fiber $P_x \coloneqq \pi^{-1}\{x\}$ of π over any point $x \in M$ is an embedded submanifold of P. If (U, ϕ) is a local trivialization for π, then the requirement $\pi_1 \circ \phi = \pi$ essentially means that for all $x \in U$, the restriction $\phi|_{P_x}$ of ϕ to P_x is a diffeomorphism onto $\{x\} \times G \cong G$. We can therefore think of the total space P as the base space M with a copy of G glued to every point x, namely P_x. In other words, P is a "bundle of fibers", one for each point in M and all diffeomorphic to G via local trivializations. The G-equivariance of the local trivializations now implies the following, see also [9, Proposition 27.6].

Lemma 3.3. *Let $\pi\colon P \to M$ be a principal G-bundle. The right action of G on P is free, and transitive on each fiber of π.*

Proof. Let $p \in P$ and choose a local trivialization (U, ϕ) for π with $x \coloneqq \pi(p) \in U$, then $\phi(p) = (x, g)$ for some $g \in G$. Let $h \in G$ and suppose $ph = p$, then

$$(x, g) = \phi(p) = \phi(ph) = \phi(p)h = (x, g)h = (x, gh)$$

since ϕ is G-equivariant, so $g = gh$ and $h = e$. To show that G acts transitively on the fibers, let q be any point in the fiber $P_x = \pi^{-1}\{x\}$ of π over x, so that $\phi(q) = (x, h)$ for some $h \in G$. It follows that

$$pg^{-1}h = \phi^{-1}(x, g)g^{-1}h = \phi^{-1}(x, h) = q,$$

which proves the claim. □

The simplest example of a principal G-bundle over a smooth manifold M is the projection $\pi\colon M \times G \to M$.

Example 3.4. Let M be a smooth manifold and G a Lie group, then the projection $\pi\colon P \coloneqq M \times G \to M$ onto the first coordinate together with the smooth right action of G on P given by right multiplication in the second coordinate is a principal G-bundle. The pair (M, id_P) is a global trivialization for π. △

One could say that the definition of a principal G-bundle was conjured up precisely to describe spaces that look like the Cartesian product $M \times G$ *locally*, but have some sort of twist *globally*. This is reflected in the concept of transition functions. Let $\pi\colon P \to M$ be a principal G-bundle, choose a bundle atlas $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ and let $\alpha, \beta \in A$. For any $x \in U_\alpha \cap U_\beta$, there exists $\rho_{\alpha\beta}(x) \in G$ such that the composition

$$\phi_\alpha \circ \phi_\beta^{-1}\colon (U_\alpha \cap U_\beta) \times G \longrightarrow (U_\alpha \cap U_\beta) \times G$$

maps (x, e) to $(x, \rho_{\alpha\beta}(x))$, since

$$\pi_1((\phi_\alpha \circ \phi_\beta^{-1})(x, e)) = (\pi \circ \phi_\beta^{-1})(x, e) = \pi_1(x, e) = x.$$

We also know $\phi_\alpha \circ \phi_\beta^{-1}$ is G-equivariant, so

$$(\phi_\alpha \circ \phi_\beta^{-1})(x, g) = (\phi_\alpha \circ \phi_\beta^{-1})(x, e)g = (x, \rho_{\alpha\beta}(x))g = (x, \rho_{\alpha\beta}(x)g) \tag{3.1}$$

for all $g \in G$. The maps $\rho_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$ thus defined are known as the *transition functions* of the bundle relative to the chosen bundle atlas, and it is straightforward to show that they are smooth.

Lemma 3.5. *Let $\pi\colon P \to M$ be a principal G-bundle, then the transition functions relative to any bundle atlas are smooth.*

Proof. Let $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ be a G-atlas for π and denote the corresponding transition functions by $\rho_{\alpha\beta}$. Let $\alpha, \beta \in A$, then by definition, $\rho_{\alpha\beta}$ is the composition of the map $U_\alpha \cap U_\beta \to (U_\alpha \cap U_\beta) \times G$ sending x to (x, e), $\phi_\alpha \circ \phi_\beta^{-1}$ and the projection $(U_\alpha \cap U_\beta) \times G \to G$ onto the second coordinate. All three of these are smooth, so $\rho_{\alpha\beta}$ is too. □

They also satisfy the following important property: for all $\alpha, \beta, \gamma \in A$ and $x \in U_\alpha \cap U_\beta \cap U_\gamma$, we have

$$\begin{aligned}(x, \rho_{\alpha\gamma}(x)) &= (\phi_\alpha \circ \phi_\gamma^{-1})(x, e) = (\phi_\alpha \circ \phi_\beta^{-1} \circ \phi_\beta \circ \phi_\gamma^{-1})(x, e) \\ &= (\phi_\alpha \circ \phi_\beta^{-1})(x, \rho_{\beta\gamma}(x)) = (x, \rho_{\alpha\beta}(x)\rho_{\beta\gamma}(x))\end{aligned}$$

by Equation (3.1), so

$$\rho_{\alpha\gamma}(x) = \rho_{\alpha\beta}(x)\rho_{\beta\gamma}(x). \tag{3.2}$$

This is known as the *cocycle condition.* The following properties are immediate consequences of it.

Lemma 3.6. *Let $\pi \colon P \to M$ be a principal G-bundle, then the transition functions $\rho_{\alpha\beta}$ of π relative to a bundle atlas $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ satisfy the following properties for all $\alpha, \beta \in A$ and $x \in U_\alpha \cap U_\beta$:*

(i) $\rho_{\alpha\alpha}(x) = e$;

(ii) $\rho_{\alpha\beta}(x) = \rho_{\beta\alpha}(x)^{-1}$.

The transition functions $\rho_{\alpha\beta}$ describe how the locally trivial pieces $U_\alpha \times G$ glue together to form the total space P of the bundle. As the next example illustrates, that means they often indicate the extent to which the bundle is twisted.

Example 3.7. Let $k \in \mathbb{Z}_{\geq 1}$ and denote by $C_k = \langle g \rangle$ the cyclic group of order k. Define $\zeta := e^{\pi i/k} \in \mathbb{C}^*$, then the right action of C_k on the circle $S^1 \subseteq \mathbb{C}^*$ given by $z \cdot g = z\zeta^2$ is smooth. With it, the k-sheeted covering $\pi \colon S^1 \to S^1$, $z \to z^k$ of S^1 is a principal C_k-bundle. A bundle atlas can be constructed as follows. Write $U := S^1 \setminus \{1\}$ and $V := S^1 \setminus \{-1\}$, then $U \cap V = S^1 \setminus \{\pm 1\}$ is the disjoint union of W_+ and W_-, where $W_\pm := \{z \in S^1 : \pm \operatorname{Im}(z) > 0\}$. Set $W_j := \{e^{2\pi i(j+t)/k} : t \in (0,1)\}$ for all $j \in \mathbb{Z}$, then $W_j = W_{j+k}$ for all $j \in \mathbb{Z}$ and

$$\pi^{-1}U = \bigcup_{j=0}^{k-1} W_j, \qquad \pi^{-1}V = \bigcup_{j=0}^{k-1} \zeta W_j.$$

Note that the restrictions $\pi|_{W_j}$ are all diffeomorphisms onto U. If $\phi_U \colon \pi^{-1}U \to U \times C_k$ is defined on W_j by $\phi_U(z) = (\pi(z), g^j)$, it follows that (U, ϕ_U) is a local trivialization for π. Similarly, there is a local trivialization (V, ϕ_V) with ϕ_V defined on ζW_j by $\phi_V(z) = (\pi(z), g^j)$; together, these two constitute a bundle atlas for π. It is straightforward to check that the composition $\phi_U \circ \phi_V^{-1}$ is given for all $w \in U \cap V$ and $g^j \in C_k$ by

$$(\phi_U \circ \phi_V^{-1})(w, g^j) = \begin{cases} (w, g \cdot g^j) & \text{if } w \in W_+ \\ (w, e \cdot g^j) & \text{if } w \in W_-, \end{cases}$$

so the transition function ρ_{UV} is g on W_+ and e on W_-. See Figure 3.1 for an illustration. △

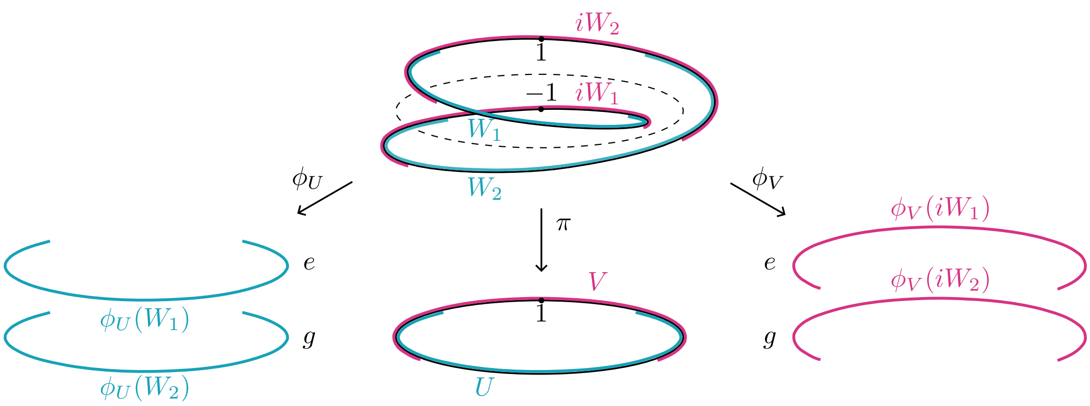


Figure 3.1: Illustration of the bundle atlas for the 2-sheeted covering $S^1 \to S^1$, $z \mapsto z^2$, as constructed in Example 3.7. The open subsets U and V of S^1 have been drawn slightly smaller for clarity.

One useful property of principal bundles is that the right G-action induces a natural one-to-one correspondence between local trivializations and local sections of the bundle.

Definition 3.8. Let $\pi\colon P \to M$ be a principal G-bundle and $U \subseteq M$ an open subset. A *smooth local section* of π over U is a smooth map $\sigma\colon U \to \pi^{-1}U$ such that $\pi \circ \sigma = \mathrm{id}_U$.

Lemma 3.9. *Let $\pi\colon P \to M$ be a principal G-bundle and $U \subseteq M$ an open subset. There is a natural one-to-one correspondence between smooth local sections and local trivializations of π over U.*

Proof. Given a smooth local section $\sigma\colon U \to \pi^{-1}U$, we define a map

$$\begin{aligned}\psi_\sigma\colon U \times G &\longrightarrow \pi^{-1}U\\ (x,g) &\longmapsto \sigma(x)g.\end{aligned}$$

This is clearly a G-equivariant diffeomorphism. Conversely, given a diffeomorphism $\psi\colon U \times G \to \pi^{-1}U$ which respects the G-action, we can define a local section

$$\begin{aligned}\sigma_\psi\colon U &\longrightarrow \pi^{-1}U\\ x &\longmapsto \psi(x,e).\end{aligned}$$

These two operations are inverses of each other; details are left to the reader. □

This proof is based on [10, Proposition 1.1.6]. Given a local trivialization of a principal bundle, we will often speak of the associated or corresponding local section, and vice versa.

The structure-preserving maps between principal G-bundles are defined as follows.

Definition 3.10. Let $\pi_1\colon P_1 \to M_1$ and $\pi_2\colon P_2 \to M_2$ be principal G-bundles. A smooth map $F\colon P_1 \to P_2$ is a *principal G-bundle morphism* if it is G-equivariant and there exists a smooth map $f\colon M_1 \to M_2$ such that $\pi_2 \circ F = f \circ \pi_1$. We say F *covers* f, and call F

(i) a *principal G-bundle isomorphism* if F is a diffeomorphism, and

(ii) a *principal G-bundle morphism over M* if $M \coloneqq M_1 = M_2$ and $f = \mathrm{id}_M$.

Note that the map f covered by F is uniquely determined by F. If F is a principal G-bundle morphism over a smooth manifold M, it is automatically an isomorphism.

Lemma 3.11. *Let $\pi_1\colon P_1 \to M$ and $\pi_2\colon P_2 \to M$ be principal G-bundles and $F\colon P_1 \to P_2$ a principal G-bundle morphism over M, then F is a diffeomorphism.*

Proof. Let $p \in P_1$, set $x \coloneqq \pi_1(p)$ and choose an open neighborhood $U \subseteq M$ of x admitting local trivializations $\phi_1\colon \pi_1^{-1}U \to U \times G$ and $\phi_2\colon \pi_2^{-1}U \to U \times G$. Note that this is always possible, since there exist such neighborhoods for both bundles separately and we can take U to be their intersection.

Note that $\pi_2(F(q)) = \pi_1(q) \in U$ for any $q \in \pi_1^{-1}U$ since $\pi_2 \circ F = \pi_1$ by definition of a principal G-bundle morphism over M, so F restricts to a map $\pi_1^{-1}U \to \pi_2^{-1}U$. Hence, $\hat{F} \coloneqq \phi_2 \circ F \circ \phi_1^{-1}\colon U \times G \to U \times G$ is smooth, and so too is the map $\rho\colon U \to G$ defined by $\rho(x) = \pi_G(\hat{F}(x,e))$, with $\pi_G\colon U \times G \to G$ the projection. Now, note that $\pi_2 \circ F = \pi_1$ implies $\pi_U \circ \hat{F} = \pi_U$ with $\pi_U\colon U \times G \to U$ the other projection, so $\hat{F}(x,e) = (x,\rho(x))$ for all $x \in U$. We also know $\hat{F}$ is G-equivariant as a composition of G-equivariant maps, so $\hat{F}(x,g) = (x,\rho(x)g)$ for any $x \in U$, $g \in G$. It follows that $\hat{F}$ is bijective with inverse $(x,g) \mapsto (x,\rho(x)^{-1}g)$, which is smooth by the smoothness of ρ, so $\hat{F}$ is a diffeomorphism. It follows that $\phi_2^{-1} \circ \hat{F} \circ \phi_1 = F|_{\pi^{-1}U}$ is a diffeomorphism too, and we are done. □

3.2 Pullback bundles

An important way to generate new principal G-bundles out of old ones, along with principal G-bundle morphisms between them, is through the following construction. Let $\pi\colon P \to N$ be a principal G-bundle, M a smooth manifold and $F\colon M \to N$ a smooth map, then we define the *pullback* of P by F as

$$F^*P \coloneqq \{(x,p) \in M \times P : F(x) = \pi(p)\} = \bigcup_{x\in M} \{x\} \times \pi^{-1}\{F(x)\} \subseteq M \times P. \tag{3.3}$$

It is the inverse image of

$$\Delta_N \coloneqq \{(y,y) : y \in N\} \subseteq N \times N$$

under the smooth map $F \times \pi\colon M \times P \to N \times N$. The diagonal Δ_N is just the graph of the identity id_N and thus an embedded submanifold of $N \times N$. Together with the next lemma, this implies that F^*P is an embedded submanifold of $M \times P$.

Lemma 3.12. *For any $x \in M$ and $p \in P$ such that $y \coloneqq F(x) = \pi(p) \in N$,*

$$\mathrm{d}(F \times \pi)_{(x,p)}(T_{(x,p)}(M \times P)) + \mathrm{d}j_{(y,y)}(T_{(y,y)}\Delta_N) = T_{(y,y)}(N \times N),$$

where $j\colon \Delta_N \hookrightarrow N \times N$ is the inclusion.

Proof. Define inclusions $j_1\colon N \hookrightarrow N \times N$, $y' \mapsto (y', y)$ and $j_2 \coloneqq N \hookrightarrow N \times N$, $y' \mapsto (y, y')$ and set $T_i \coloneqq \mathrm{d}(j_i)_y(T_yN)$ for $i \in \{1, 2\}$, then $T_{(y,y)}(N \times N) = T_1 \oplus T_2$ by Theorem A.2. By that same theorem, $T_1 = \ker \mathrm{d}(\pi_2)_{(y,y)}$ and $T_2 = \ker \mathrm{d}(\pi_1)_{(y,y)}$ with $\pi_i\colon N \times N \to N$ the projections.

Define $i\colon P \hookrightarrow M \times P$, $p' \mapsto (x, p')$ and let $v \in T_pP$, then

$$\mathrm{d}(F \times \pi)_{(x,p)}(\mathrm{d}i_p(v)) = \mathrm{d}((F \times \pi) \circ i)_p(v) = \mathrm{d}(j_2 \circ \pi)_p(v) = \mathrm{d}(j_2)_y(\mathrm{d}\pi_p(v)).$$

By Lemma 3.2, π is a submersion, so $\mathrm{d}\pi_p(T_pP) = T_yN$ and $T_2 \subseteq \mathrm{d}(F \times \pi)_{(x,p)}(T_{(x,p)}(M \times M))$.

Now, let $v \in \mathrm{d}j_{(y,y)}(T_{(y,y)}\Delta_N)$ with $v \in T_2 = \ker \mathrm{d}(\pi_1)_{(y,y)}$, then also $v \in \ker \mathrm{d}(\pi_2)_{(y,y)} = T_1$ since $\pi_1 \circ j = \pi_2 \circ j$. We know T_1 and T_2 intersect trivially, so $v = 0$ and $T_2 \cap \mathrm{d}j_{(y,y)}(T_{(y,y)}\Delta_N) = \{0\}$. Both these subspaces of $T_{(y,y)}(N \times N)$ have dimension $\dim N$, so together, they span the whole tangent space and we are done. □

Lemma 3.12 can be summarized by saying that the smooth map $F \times \pi$ is *transverse* to the diagonal Δ_N, so that $F^*P = (F \times \pi)^{-1}\Delta_N$ is an embedded submanifold of $M \times P$ by [8, Theorem 6.30]. Equation (3.3) shows that it is a disjoint union of fibers of π and of course, each such fiber is diffeomorphic to G. Therefore, the following result is perhaps not entirely unexpected.

Proposition 3.13. *The projection $\pi_1\colon F^*P \to M$ onto the first coordinate together with the right action of G on F^*P given by right multiplication in the second coordinate is a principal G-bundle, and the second projection $\pi_2\colon F^*P \to P$ is a principal G-bundle morphism covering F.*

Proof. It is clear that π_1 is a smooth surjection, so it remains to prove the existence of a G-atlas. To this end, let $\{(V_\alpha, \psi_\alpha)\}_{\alpha\in A}$ be a G-atlas for π and set $U_\alpha \coloneqq F^{-1}V_\alpha$ for all $\alpha \in A$, then $\{U_\alpha\}_{\alpha\in A}$ is an open cover of M. Write $g_\alpha \coloneqq \pi_G \circ \psi_\alpha\colon \pi^{-1}V_\alpha \to G$ for all $\alpha \in A$, with $\pi_G\colon V_\alpha \times G \to G$ the projection. Note that $\pi(p) = F(x) \in V_\alpha$ for all $(x,p) \in \pi_1^{-1}U_\alpha$ since $x = \pi_1(x,p) \in U_\alpha = F^{-1}V_\alpha$, so $\pi_1^{-1}U_\alpha \subseteq U_\alpha \times \pi^{-1}V_\alpha$. It follows that we can define a map

$$\begin{aligned} \phi_\alpha\colon \pi_1^{-1}U_\alpha &\longrightarrow U_\alpha \times G \\ (x,p) &\longmapsto (x, g_\alpha(p)), \end{aligned}$$

which is smooth since g_α is smooth. We know g_α is G-equivariant, so ϕ_α is G-equivariant too, by definition of the G-action on F^*P. It also clearly preserves the fibers and it has a smooth inverse

$$\begin{aligned} U_\alpha \times G &\longrightarrow \pi_1^{-1} U_\alpha \\ (x, g) &\longmapsto (x, \psi_\alpha^{-1}(F(x), g)), \end{aligned}$$

so $\{(U_\alpha, \phi_\alpha)\}$ is a bundle atlas and π_1 is a principal G-bundle. The projection π_2 is G-equivariant by definition of the G-action on F^*P and it covers F by definition of the space F^*P, so the second claim is immediate. ☐

$$\begin{array}{ccc} F^*P & \xrightarrow{\pi_2} & P \\ \big\downarrow{\scriptstyle \pi_1} & & \big\downarrow{\scriptstyle \pi} \\ M & \xrightarrow{F} & N \end{array}$$

The above diagram shows an overview of the situation. Denoting the transition functions of π relative to the bundle atlas $\{(V_\alpha, \psi_\alpha)\}_{\alpha \in A}$ by $\rho_{\alpha\beta}$, it is easy to see that the transition functions of π_1 relative to the induced bundle atlas $\{(U_\alpha, \phi_\alpha)\}$ are precisely the compositions $\rho_{\alpha\beta} \circ F$.

3.3 The tautological bundle

Using the pullback construction, the principal G-bundles over any smooth manifold M can be classified in a natural way. For $G = U(1)$, this classification scheme will prove particularly useful in the context of the anomalous Hall effect later on in Chapter 7.

For any $\mathbb{C}$-vector space V, the *complex projective space* $\mathbb{P}(V)$ is defined as the quotient of $V \setminus \{0\}$ by the equivalence relation $\sim$ defined by

$$z \sim w \iff \text{there exists } \lambda \in \mathbb{C}^* \text{ such that } z = \lambda w.$$

The equivalence class of any point $z \in V \setminus \{0\}$ is denoted by $[z]$, and can be identified with the 1-dimensional (complex) subspace of V spanned by z. Now let $n \in \mathbb{Z}_{\geq 0}$ and set $\mathcal{H} := \mathbb{C}^{n+1}$. Supply $\mathcal{H}$ with the Euclidean norm $\|\cdot\|$, then $\mathbb{P}(\mathcal{H})$ with the quotient topology and the unit sphere

$$S := \{z \in \mathcal{H} : \|z\| = 1\}$$

in $\mathcal{H}$ with the subspace topology are compact topological manifolds of dimension $2n+1$ and $2n$, respectively. Additionally, they admit smooth structures with respect to which the inclusion $j \colon S \hookrightarrow \mathcal{H}$ is a smooth embedding and the quotient map $q \colon \mathcal{H} \setminus \{0\} \to \mathbb{P}(\mathcal{H})$ a smooth submersion. For details, we refer to [8, Example 5.15, Exercise 1-9, Exercise 4-5]. It follows that the composition

$$\begin{aligned} \gamma_n := q \circ j \colon S &\longrightarrow \mathbb{P}(\mathcal{H}) \\ z &\longmapsto [z] \end{aligned}$$

is smooth, and it is clearly surjective. Note that for any $z \in S$, the fiber $\pi^{-1}\{[z]\}$ of π is diffeomorphic to the circle $U(1)$ since the elements of S being mapped to $[z]$ are precisely those of the form λz for some $\lambda \in U(1)$, so one might suspect that we are dealing with a principal $U(1)$-bundle here.

Proposition 3.14. *The map $\gamma_n \colon S \to \mathbb{P}(\mathcal{H})$ together with the right action of $U(1)$ on S by scalar multiplication is a principal $U(1)$-bundle.*

Proof. Let $e_0, \dots, e_n$ be the standard basis for $\mathcal{H}$ and set $V_k := \mathrm{Sp}\{e_0, \dots, \hat{e}_k, \dots, e_n\}$ for all $0 \leq k \leq n$, where the hat denotes that we omit that particular basis vector. Define $U_k := \mathbb{P}(\mathcal{H}) \setminus \mathbb{P}(V_k)$, then

$$q^{-1} U_k = \{z \in \mathcal{H} : [z] \notin \mathbb{P}(V_k)\} = \{z \in \mathcal{H} : z \notin V_k\} = \mathcal{H} \setminus V_k$$

since V_k is a linear subspace of $\mathcal{H}$. We know V_k is closed in $\mathcal{H}$ since it is finite-dimensional, so $q^{-1}U_k$ is open in $\mathcal{H}$ and U_k open in $\mathbb{P}(\mathcal{H})$. Clearly, the U_k cover $\mathbb{P}(\mathcal{H})$: for any $z = (z^0, \dots, z^n) \in \mathcal{H} \setminus \{0\}$, there is a k with $z^k \neq 0$, which implies $z \notin V_k$ and thus $[z] \in U_k$.

Now let $0 \leq k \leq n$, then

$$\gamma_n^{-1}U_k = \{z \in S : z \notin V_k\} = \{z \in S : z^k \neq 0\},$$

so we can define maps

$$\phi_k \colon \gamma_n^{-1}U_k \longrightarrow U_k \times U(1) \qquad z \longmapsto \left([z], \frac{z^k}{|z^k|}\right), \qquad\qquad \psi_k \colon U_k \times U(1) \longrightarrow \gamma_n^{-1}U_k \qquad ([z], \lambda) \longmapsto \frac{|z^k| z}{z^k \, \|z\|}\lambda.$$

Note that the definition of ψ_k does not depend on the choice of representative. Both maps are smooth and by a straightforward calculation, they are inverses of each other, so $\phi_k = \psi_k^{-1}$ is a diffeomorphism. Clearly, $\pi_1 \circ \phi_k = \gamma_n$ with $\pi_1 \colon U_k \times U(1) \to U_k$ the projection since ϕ_k is just $\gamma_n \colon z \mapsto [z]$ in the first coordinate. Also,

$$\phi_k(z\lambda) = \left([z\lambda], \frac{z^k \lambda}{|z^k \lambda|}\right) = \left([z], \frac{z^k}{|z^k|}\lambda\right) = \phi_k(z)\lambda$$

for all $z \in \gamma_n^{-1}U_k$ and $\lambda \in U(1)$, so (U_k, ϕ_k) is a local trivialization for γ_n and we are done. □

We call γ_n the *tautological bundle* over $\mathbb{P}(\mathcal{H})$. It has a special universal property that allows us to classify the principal $U(1)$-bundles over any sufficiently low-dimensional manifold, see [10, Theorem 3.4.10] and [10, Theorem 3.6.7] for a proof. By an isomorphism class of principal $U(1)$-bundles over a smooth manifold M, we mean an equivalence class under the equivalence relation on the set of all principal $U(1)$-bundles defined by isomorphism *over* M. That is, two bundles belong to the same class if and only if there exists a principal G-bundle (iso)morphism between them which covers id_M.

Theorem 3.15. *Let M be a smooth manifold of dimension $m \leq 2n$, then the assignment $F \mapsto F^*S$ induces a bijection from the set of homotopy classes of smooth maps $M \to \mathbb{P}(\mathcal{H})$ to the set of isomorphism classes of principal $U(1)$-bundles over M.*

But what if we want to construct a bundle over a smooth manifold M using a map F from M to the projectivization $\mathbb{P}(\mathcal{H})$ of an *infinite-dimensional* Hilbert space $\mathcal{H}$, instead? Does there also exist a tautological bundle in this case, which we can pull back along F? This question opens the door to the realm of topological and smooth manifolds modeled not on $\mathbb{R}^n$, but on general Banach spaces. There still exists a notion of differentiability for maps between such spaces, which allows for the generalization of a large portion of the theory of finite-dimensional manifolds to this more general setting; see for instance [11]. However, we will not attempt to do that here. In order to still be able to make sense of bundles over infinite-dimensional projective spaces, we take a different approach.

The definition of a principal bundle we gave in Section 3.1 can be generalized to arbitrary topological spaces. By replacing every occurrence of

- "Lie group" with "topological group",
- "smooth manifold" with "topological space",
- "smooth" with "continuous" and
- "diffeomorphism" with "homeomorphism"

in the definitions in Section 3.1, one obtains the definitions for *topological principal G-bundles* and *topological principal G-bundle morphisms.* Most of what we have discussed in this chapter so far has a topological analog. In particular, a topological principal G-bundle $\pi \colon P \to N$ can be pulled back along a continuous map $F \colon M \to N$ to obtain a topological principal G-bundle over M in exactly the same way as for the smooth category, as was outlined in Section 3.2.

Now, the crucial fact is that for any smooth manifold M and Lie group G, the topological and smooth principal G-bundles over M turn out to be essentially the same thing, up to isomorphism.

Theorem 3.16. *Let M be a smooth manifold and G a Lie group. Forgetting about the smooth structure defines a bijection from the set of isomorphism classes of (smooth) principal G-bundles over M to the set of isomorphism classes of topological principal G-bundles over M.*

The interested reader may consult [10, Theorem 3.6.6] for a proof. Again, by isomorphism classes we mean isomorphism classes *over M*.

Theorem 3.16 states that for any topological principal G-bundle $\pi_1 : P_1 \to M$ over a smooth manifold M, there exist a *smooth* principal G-bundle $\pi_2 : P_2 \to M$ and a G-equivariant homeomorphism $F : P_1 \to P_2$ with $\pi_2 \circ F = \pi_1$. Via F, we can supply P_1 with a smooth structure with respect to which π_1 is also a smooth principal G-bundle. Therefore, in order to be able to construct smooth bundles over M using continuous maps from M to the infinite-dimensional projective Hilbert space $\mathbb{P}(\mathcal{H})$, all we need now is a topological bundle over $\mathbb{P}(\mathcal{H})$. This bundle is constructed in much the same way as in the finite-dimensional case. Provide the unit sphere

$$S \coloneqq \{z \in \mathcal{H} : \|z\| = 1\}$$

in $\mathcal{H}$ with the subspace topology and $\mathbb{P}(\mathcal{H})$ with the quotient topology, then the inclusion $j : S \hookrightarrow \mathcal{H}$ and the quotient map $q : \mathcal{H} \setminus \{0\} \to \mathbb{P}(\mathcal{H})$ are continuous, so their composition

$$\begin{aligned} \gamma \coloneqq q \circ j : S &\longrightarrow \mathbb{P}(\mathcal{H}) \\ z &\longmapsto [z]. \end{aligned}$$

is too. We have now the following infinite-dimensional version of Proposition 3.14.

Proposition 3.17. *The map $\gamma : S \to \mathbb{P}(\mathcal{H})$ together with the right action of $U(1)$ on S by scalar multiplication is a topological principal $U(1)$-bundle.*

Proof. The proof is virtually identical to that of Proposition 3.14. Choose an orthonormal basis $\{e_k\}_{k\in I}$ for $\mathcal{H}$ and for any $k \in I$, let $V_k \coloneqq \overline{\mathrm{Sp}}\{e_\ell : \ell \neq k\}$ be the closed linear span of all basis vectors except the k-th. It follows that $U_k \coloneqq \mathbb{P}(\mathcal{H}) \setminus \mathbb{P}(V_k)$ is open for all $k \in I$, and the U_k clearly cover $\mathbb{P}(\mathcal{H})$. For any $k \in I$, we have continuous maps

$$\begin{aligned} \phi_k : \gamma^{-1}U_k &\longrightarrow U_k \times U(1) \\ z &\longmapsto \left([z], \frac{\langle e_k|z\rangle}{|\langle e_k|z\rangle|}\right) \end{aligned} \qquad \begin{aligned} \psi_k : U_k \times U(1) &\longrightarrow \gamma^{-1}U_k \\ ([z], \lambda) &\longmapsto \frac{|\langle e_k|z\rangle|}{\langle e_k|z\rangle} \cdot \frac{z}{\|z\|} \cdot \lambda \end{aligned}$$

which can be shown to be inverses of each other, so $\phi_k = \psi_k^{-1}$ is a homeomorphism. It is also clearly $U(1)$-equivariant and fiber-preserving, so (U_k, ϕ_k) is a local trivialization. □

Again, we call γ the *tautological bundle* over $\mathbb{P}(\mathcal{H})$, but it is important to keep in mind that it is only a topological one.

3.4 The vertical subbundle

At this point, one might wonder why we need smooth structures on our bundles in the first place. One important reason is that we want to be able to look at their tangent bundles and define certain objects on them. These objects will be the focal point of Chapters 4 and 5, and we kick off the discussion here.

If $\pi : P \to M$ is any (smooth) principal G-bundle, there is a well-defined notion of "verticality" on the tangent spaces of P.

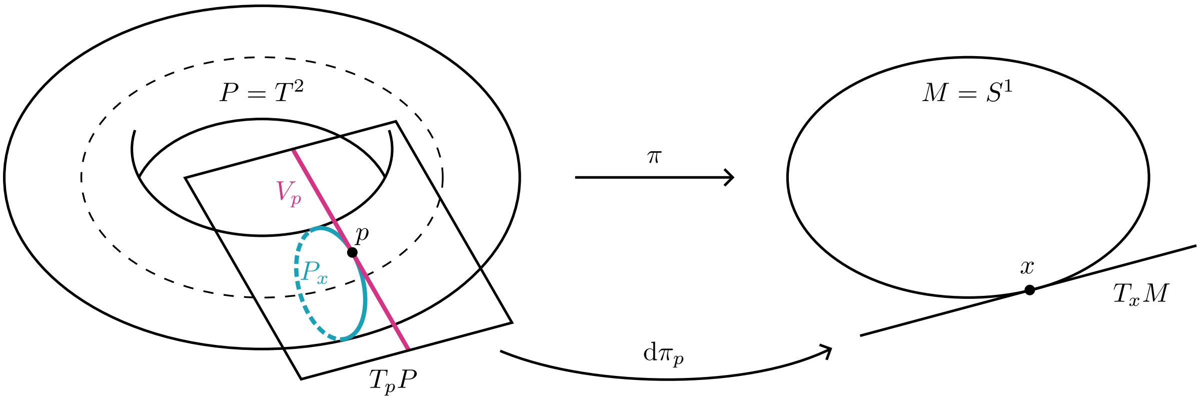


Figure 3.2: Illustration of Lemma 3.19 for the trivial $U(1)$-bundle $\pi\colon T^2 = S^1 \times U(1) \to S^1$. The vertical tangent space V_p to P at a point $p \in P$ is equal to the tangent space at p to the fiber $P_x := \pi^{-1}\{x\}$, where $x := \pi(p)$.

Definition 3.18. For all $p \in P$, the *vertical tangent subspace* of T_pP is defined as $V_p := \ker \mathrm{d}\pi_p \subseteq T_pP$. The union $V := \bigcup_{p\in P} V_p = \ker \mathrm{d}\pi$ is known as the *vertical subbundle* of TP.

V_p can be shown to vary smoothly with p. More precisely, V is a so-called *smooth subbundle* of the tangent bundle TP of P since π has constant rank by Lemma 3.2; see [8, Theorem 10.34] for details. The vertical tangent subspaces have an intuitive geometric interpretation.

Lemma 3.19. *Let $p \in P$ and set $x := \pi(p)$. Let $P_x := \pi^{-1}\{x\}$ be the fiber of π containing p and $i\colon P_x \hookrightarrow P$ the inclusion, then $V_p = \mathrm{d}i_p(T_pP_x)$.*

Proof. Let $w \in T_pP_x$, then $\mathrm{d}\pi_p(\mathrm{d}i_p(w)) = \mathrm{d}(\pi \circ i)_p(w) = 0$ since $\pi \circ i$ is the constant map $q \mapsto x$, so $\mathrm{d}i_p(w) \in \ker \mathrm{d}\pi_p = V_p$ and $\mathrm{d}i_p(T_pP_x) \subseteq V_p$.

Now, note that we have a short exact sequence of finite-dimensional vector spaces

$$0 \longrightarrow V_p \hookrightarrow T_pP \xrightarrow{\mathrm{d}\pi_p} T_xM \longrightarrow 0$$

where the surjectivity of $\mathrm{d}\pi_p$ follows from Lemma 3.2, so comparing dimensions gives

$$\dim V_p = \dim T_pP - \dim T_xM = \dim P - \dim M = \dim G.$$

Since P_x is diffeomorphic to G, it follows that $\mathrm{d}i_p(T_pP_x)$ also has dimension $\dim G$, so $V_p = \mathrm{d}i_p(T_pP_x)$. □

In words, V_p is just the subspace of T_pP consisting of those vectors which are tangent to the fiber P_x of π containing p, as illustrated in Figure 3.2. Since any local trivialization of π around p induces a diffeomorphism between P_x and the Lie group G, Lemma 2.8 implies $\mathrm{Lie}(G) \cong V_p$ as vector spaces. As it turns out, there is a *canonical* identification of $\mathrm{Lie}(G)$ and V_p, independent of the chosen trivialization. It is given by the infinitesimal generator $\underline{\theta}\colon \mathrm{Lie}(G) \to \mathfrak{X}(P)$, $X \mapsto \underline{X}$ of the right action θ of G on P, see also [9, Proposition 27.18].

Lemma 3.20. *Let $p \in P$, then the linear map $\eta_p\colon \mathrm{Lie}(G) \to T_pP$, $X \mapsto \underline{X}_p$ is an isomorphism onto V_p.*

Proof. Note that $\eta_p = \mathrm{d}(\theta^{(p)})_e \circ \epsilon$ by Equation (2.3), with $\epsilon\colon \mathrm{Lie}(G) \to T_eG$ the isomorphism from Lemma 2.8. It follows that

$$\mathrm{d}\pi_p(\underline{X}_p) = \mathrm{d}(\pi \circ \theta^{(p)})_e(X_e) = 0$$

for any $X \in \mathrm{Lie}(G)$ since $\pi \circ \theta^{(p)}$ is the constant map $g \mapsto \pi(pg) = \pi(p)$, so $\underline{X}_p \in \ker \mathrm{d}\pi_p = V_p$. Also, $\mathrm{d}(\theta^{(p)})_e$ is injective by Lemma 2.25 since G acts freely on P. Finally, note that V_p and $\mathrm{Lie}(G) \cong T_eG$ both have dimension $\dim G$ by Lemma 3.19, so we are done. □

Lemma 3.20 essentially says that the vertical tangent subspace consists precisely of the fundamental vectors in T_pP associated to elements of the Lie algebra $\mathrm{Lie}(G)$.

The important thing to realize at this point is that in general, there is no *canonically defined* "horizontal" counterpart of the vertical subbundle $V \subseteq TP$. This is because π is inherently asymmetric: if (U, ϕ) is a local trivialization for π, then on $\pi^{-1}U \cong U \times G$, the bundle map π is just the projection on the first coordinate. In general, though, there is no analogous global projection on the second coordinate. This asymmetry gives rise to the notion of a *connection* on π, which is essentially just a smooth choice of a complementary subspace H_p of V_p in T_pP for all $p \in P$. Giving such a subspace H_p amounts to choosing a projection $T_pP \to V_p$ with kernel H_p, which we can interpret as a linear map $T_pP \to \mathrm{Lie}(G)$ through Lemma 3.20. We would therefore like to be able to talk about connections on the principal bundle π as differential forms on P taking values in the vector space $\mathrm{Lie}(G)$ instead of $\mathbb{R}$.

Chapter 4

Lie algebra valued forms

The main subject of this chapter is the theory of vector valued differential forms. If M is a smooth manifold and V a finite-dimensional $\mathbb{R}$-vector space, then a *V-valued k-form* ω on M is an assignment of an alternating multilinear map $\omega_p\colon (T_pM)^k \to V$ to every point $p \in M$. If $e_1, \dots, e_n$ is a basis for V, ω can also be interpreted as an n-tuple of ordinary $\mathbb{R}$-valued differential forms of degree k on M, which allows us to generalize the wedge product, exterior derivative and pullback operations. The case $V = \mathrm{Lie}(G)$ will be particularly important for the discussion of connections and curvature on principal bundles in Chapter 5. The *Maurer-Cartan form* is a distinguished $\mathrm{Lie}(G)$-valued one-form ω_G on the Lie group G itself with many useful properties. Pulling ω_G back along a smooth map $\rho\colon M \to G$ yields a $\mathrm{Lie}(G)$-valued one-form on M which we call the *logarithmic derivative* of ρ, because it generalizes the logarithmic derivative of real-valued functions.

4.1 Vector valued forms

In this section and the next, we largely follow [9, §21]. Let T, V, W and Z be arbitrary finite-dimensional vector spaces over $\mathbb{R}$ and write

$$A_k(T, V) \coloneqq \{f\colon T^k \to V : f \text{ is multilinear and alternating}\}$$

for all $k \in \mathbb{Z}_{\geq 1}$, then $A_k(T, V)$ has a natural $\mathbb{R}$-vector space structure with addition and scalar multiplication defined pointwise. Taking $A_0(T, V)$ to just be V by convention (any map taking 0 arguments is automatically multilinear and alternating), the universal mapping property of the exterior power induces an isomorphism

$$A_k(T, V) \cong \mathrm{Hom}_{\mathbb{R}}\Big(\bigwedge^k T, V\Big)$$

for all $k \in \mathbb{Z}_{\geq 0}$. Since all spaces are finite-dimensional, standard linear algebra also gives us identifications

$$\mathrm{Hom}_{\mathbb{R}}\Big(\bigwedge^k T, V\Big) \cong \Big(\bigwedge^k T\Big)^* \otimes V \cong \Big(\bigwedge^k T^*\Big) \otimes V.$$

In particular, if $V = \mathbb{R}$, this yields an identification of $\bigwedge^k T^*$ with the space of alternating multilinear forms $T^k \to \mathbb{R}$.

Now, recall that a differential form $\omega \in \Omega^k(M)$ of degree $k \geq 0$ on a smooth manifold M is a section of the vector bundle $\bigwedge^k T^*M$ over M. The fiber at $p \in M$ of this bundle is the k-th exterior power $\bigwedge^k T_p^*M$ of the cotangent space T_p^*M, so we can interpret ω_p as an alternating multilinear form $(T_pM)^k \to \mathbb{R}$ by the above discussion. More generally, we have the following.

Definition 4.1. Let M be a smooth manifold and $k \in \mathbb{Z}_{\geq 0}$. A *$V$-valued k-form* on M is a section of the tensor product bundle $(\bigwedge^k T^*M) \otimes (M \times V)$ over M.

If ω is a V-valued k-form on M, then for any $p \in M$, we can view $\omega_p \in (\bigwedge^k T_p^*M) \otimes V$ as an alternating multilinear map $(T_pM)^k \to V$. Choose a basis $e_1, \ldots, e_n$ for V, then for any $p \in M$, it follows that there exist unique multilinear forms $\omega_p^1, \ldots, \omega_p^n \colon (T_pM)^k \to \mathbb{R}$ such that

$$\omega_p(v_1, \ldots, v_k) = \omega_p^i(v_1, \ldots, v_k)e_i$$

for all $v_1, \ldots, v_k \in T_pM$, where we have adopted the Einstein summation convention. The multilinear forms ω_p^i define ($\mathbb{R}$-valued) k-forms $\omega^1, \ldots, \omega^n$ on M satisfying $\omega = \omega^i \otimes e_i$, which is also often written without the tensor product sign as $\omega = \omega^i e_i$. We say ω is *smooth* if ω^i is smooth for all i. It is easy to see that this definition is independent of the chosen basis: if $f_1, \ldots, f_n$ is another basis for V and $\omega^i e_i = \eta^j f_j$, then the ω^i are smooth if and only if the η^j are since there exist $c_j^i, d_i^j \in \mathbb{R}$ such that

- $f_j = c_j^i e_i$ for all j and thus $\omega^i = \eta^j c_j^i$ for all i, and
- $e_i = d_i^j f_j$ for all i and thus $\eta^j = \omega^i d_i^j$ for all j.

We denote the space of smooth V-valued k-forms on M by $\Omega^k(M, V)$.

There is another useful characterization of smoothness of vector valued forms, for which we first need to introduce some notation. Let ω be a V-valued k-form on a smooth manifold M, then for any vector fields $X_1, \ldots, X_k$ on M, we define

$$\begin{aligned} \omega(X_1, \ldots, X_k) \colon M &\longrightarrow V \\ p &\longmapsto \omega_p(X_1|_p, \ldots, X_k|_p). \end{aligned}$$

In terms of a basis $e_1, \ldots, e_n$ for V, we can write $\omega = \omega^i e_i$ for k-forms $\omega^1, \ldots, \omega^n$ on M, so that

$$\omega(X_1, \ldots, X_k) = \omega^i(X_1, \ldots, X_k)e_i. \tag{4.1}$$

Now, recall that in the case $V = \mathbb{R}$, ω is smooth if and only if $\omega(X_1, \ldots, X_k)$ is smooth for all smooth vector fields $X_1, \ldots, X_k \in \mathfrak{X}(M)$; see for instance [8, Proposition 12.19]. It follows that ω induces a map

$$\begin{aligned} \mathfrak{X}(M)^k &\longrightarrow C^\infty(M) \\ (X_1, \ldots, X_k) &\longmapsto \omega(X_1, \ldots, X_k) \end{aligned}$$

which can be shown to be linear over $C^\infty(M)$. Something similar holds for the general case.

Lemma 4.2. *Let M be a smooth manifold and ω a V-valued k-form on M, then ω is smooth if and only if $\omega(X_1, \ldots, X_k)$ is smooth for all $X_1, \ldots, X_k \in \mathfrak{X}(M)$.*

Proof. Choose a basis $e_1, \ldots, e_n$ for V, then $\omega = \omega^i e_i$ for some k-forms $\omega^1, \ldots, \omega^n$ on M, so by (4.1), $\omega(X_1, \ldots, X_k) = \omega^i(X_1, \ldots, X_k)e_i$ for all $X_1, \ldots, X_k \in \mathfrak{X}(M)$. It follows that

$$\begin{aligned} \omega \text{ is smooth} &\iff \omega^i \text{ is smooth for all } i \\ &\iff \omega^i(X_1, \ldots, X_k) \in C^\infty(M) \text{ for all } i \text{ and } X_1, \ldots, X_k \in \mathfrak{X}(M) \\ &\iff \omega(X_1, \ldots, X_k) \in C^\infty(M, V) \text{ for all } X_1, \ldots, X_k \in \mathfrak{X}(M). \end{aligned}$$

□

Thus, any smooth V-valued k-form $\omega \in \Omega^k(M, V)$ induces a mapping $\mathfrak{X}(M)^k \to C^\infty(M, V)$, which can again be shown to be $C^\infty(M)$-linear.

4.2 Fundamental operations

Now, there are three operations on $\mathbb{R}$-valued differential forms which can be extended in a natural way to vector valued forms: the wedge product, the exterior derivative and the pullback along a smooth map.

The wedge product. Let $\mu\colon V\times W \to Z$ be a bilinear map. The *product* of two alternating multilinear maps $\omega \in A_k(T,V)$ and $\eta \in A_\ell(T,W)$ is then defined as

$$\begin{aligned}\mu(\omega,\eta)\colon T^{k+\ell} &\longrightarrow Z\\ (v_1,\dots,v_{k+\ell}) &\longmapsto \frac{1}{k!\ell!}\sum_{\sigma\in S_{k+\ell}}(\operatorname{sgn}\sigma)\mu(\omega(v_{\sigma(1)},\dots,v_{\sigma(k)}),\eta(v_{\sigma(k+1)},\dots,v_{\sigma(k+\ell)})).\end{aligned}$$

Clearly, $\mu(\omega,\eta)$ is multilinear and alternating, so we have a map

$$\begin{aligned}A_k(T,V)\times A_\ell(T,W) &\longrightarrow A_{k+\ell}(T,Z)\\ (\omega,\eta) &\longmapsto \mu(\omega,\eta).\end{aligned}$$

If M is a smooth manifold, ω a V-valued k-form on M and η a W-valued ℓ-form, then their *product* is the Z-valued $(k+\ell)$-form $\mu(\omega,\eta)$ on M defined as $\mu(\omega,\eta)_p = \mu(\omega_p,\eta_p)$ for all $p\in M$. The following lemma, proved in [9, Proposition 21.1], gives an alternative characterization of this product.

Lemma 4.3. *Let M be a smooth manifold and $\mu\colon V\times W\to Z$ a bilinear map. Let $e_1,\dots,e_n\in V$ and $f_1,\dots,f_m\in W$ be vectors, $\omega^1,\dots,\omega^n$ forms of degree k on M and $\eta^1,\dots,\eta^m$ forms of degree ℓ, then*

$$\mu(\omega^i e_i,\eta^j f_j) = (\omega^i\wedge\eta^j)\mu(e_i,f_j).$$

In particular, this shows that $\mu(\omega,\eta)$ is smooth whenever ω and η are smooth forms, so the product defines a pairing

$$\begin{aligned}\Omega^k(M,V)\times\Omega^\ell(M,W) &\longrightarrow \Omega^{k+\ell}(M,Z)\\ (\omega,\eta) &\longmapsto \mu(\omega,\eta).\end{aligned}$$

Example 4.4. Let M be a smooth manifold.

(1) If $V=W=Z=\mathbb{R}$ and μ is the multiplication on $\mathbb{R}$, the corresponding product is the regular wedge product of $\mathbb{R}$-valued differential forms on M.

(2) If $V=\operatorname{Hom}_{\mathbb{R}}(W,Z)$ and μ is the evaluation map $(L,y)\mapsto Ly$, then the product of a V-valued 0-form $F\colon M\to V$ and a W-valued k-form ω on M is just the "pointwise composition", i.e. $(F\omega)_p = F(p)\circ\omega_p$ for all $p\in M$.

(3) If $V=W$ is a finite-dimensional vector space over $\mathbb{C}$ considered as an $\mathbb{R}$-vector space through restriction of scalars, then any Hermitian inner product $\langle\,\cdot\,|\,\cdot\,\rangle$ on V is an $\mathbb{R}$-bilinear map $V\times V\to\mathbb{C}$. The product of V-valued forms ω and η on a smooth manifold M with respect to this bilinear map is written as $\langle\omega|\eta\rangle$. △

The exterior derivative. Let M be a smooth manifold and $e_1,\dots,e_n$ a basis for V, then we define

$$\begin{aligned}\mathrm{d}\colon\Omega^k(M,V) &\longrightarrow \Omega^{k+1}(M,V)\\ \omega^i e_i &\longmapsto (\mathrm{d}\omega^i)e_i\end{aligned}$$

for all $k\in\mathbb{Z}_{\geq 0}$. If $f_1,\dots,f_n$ is another basis for V and $\omega^i e_i = \eta^j f_j$, then there exist $c^i_j\in\mathbb{R}$ such that $f_j = c^i_j e_i$ for all j, so $\omega^i = \eta^j c^i_j$ and thus $\mathrm{d}\omega^i = (\mathrm{d}\eta^j)c^i_j$ for all i by the linearity of the exterior derivative on $\mathbb{R}$-valued forms on M. We conclude that $(\mathrm{d}\omega^i)e_i = (\mathrm{d}\eta^j)c^i_j e_i = (\mathrm{d}\eta^j)f_j$, so the definition of d is independent of the chosen basis for V.

As in the case $V=\mathbb{R}$, the exterior derivative on V-valued forms is an antiderivation. A proof of this fact can be found in [9, Proposition 21.3].

Lemma 4.5. *Let M be a smooth manifold and $\mu\colon V\times W\to Z$ a bilinear map. Let $\omega\in\Omega^k(M,V)$ and $\eta\in\Omega^\ell(M,W)$, then*

$$\mathrm{d}\mu(\omega,\eta)=\mu(\mathrm{d}\omega,\eta)+(-1)^k\mu(\omega,\mathrm{d}\eta).$$

The coordinate-free expressions for the exterior derivatives of $\mathbb{R}$-valued 0- and 1-forms on a smooth manifold M have natural generalizations. Recall that the exterior derivative of a smooth function (i.e. a 0-form) $f\in C^\infty(M)$ is defined by $\mathrm{d}f(X)=Xf$ for $X\in\mathfrak{X}(M)$, where $Xf\in C^\infty(M)$ is given by $(Xf)_p=X_pf$ for all $p\in M$. This makes sense because the tangent vectors to M are defined as linear maps $C^\infty(M)\to\mathbb{R}$ satisfying a Leibniz rule. We can extend this to the general case as follows: choose a basis $e_1,\dots,e_n$ for V, then for any $X\in\mathfrak{X}(M)$ and $F=F^ie_i\in C^\infty(M,V)$, we define

$$\begin{aligned}XF\coloneqq(XF^i)e_i\colon M&\longrightarrow V\\ p&\longmapsto X_pF=(X_pF^i)e_i.\end{aligned}\tag{4.2}$$

XF is smooth by definition of the smooth structure on V, and it is easy to see that its definition is independent of the chosen basis. By Equation (4.1), the exterior derivative of a smooth map (i.e. a 0-form) $F=F^ie_i\colon M\to V$ is now given by

$$\mathrm{d}F(X)=\mathrm{d}F^i(X)e_i=(XF^i)e_i=XF$$

for all $X\in\mathfrak{X}(M)$. For 1-forms, we have the following lemma, based on [9, Exercise 21.8].

Lemma 4.6. *Let M be a smooth manifold, $\omega\in\Omega^1(M,V)$ and $X,Y\in\mathfrak{X}(M)$, then*

$$\mathrm{d}\omega(X,Y)=X(\omega(Y))-Y(\omega(X))-\omega([X,Y]).$$

Proof. Choose a basis $e_1,\dots,e_n$ for V, then $\omega=\omega^ie_i$ for some $\omega^1,\dots,\omega^n\in\Omega^1(M)$, so $\mathrm{d}\omega=(\mathrm{d}\omega^i)e_i$. Using the invariant formula for the exterior derivative of $\mathbb{R}$-valued 1-forms (cf. [8, Proposition 14.29]) together with Equations (4.1) and (4.2), we find

$$\begin{aligned}\mathrm{d}\omega(X,Y)=\mathrm{d}\omega^i(X,Y)e_i&=(X(\omega^i(Y))-Y(\omega^i(X))-\omega^i([X,Y]))e_i\\&=X(\omega^i(Y)e_i)-Y(\omega^i(X)e_i)-\omega^i([X,Y])e_i\\&=X(\omega(Y))-Y(\omega(X))-\omega([X,Y]).\end{aligned}$$ □

It is important to note at this point that if M is a smooth manifold and $F\colon M\to V$ a smooth map, we can interpret $\mathrm{d}F$ as either the differential of F as a smooth map, or its exterior derivative as a 0-form. The first interpretation gives for all $p\in M$ a linear map $\mathrm{d}F_p\colon T_pM\to T_{F(p)}V$, whereas in the second, $\mathrm{d}F_p$ is a map $T_pM\to V$. Under the canonical identification of V with its tangent spaces, however, these objects are one and the same.

Lemma 4.7. *Let M be a smooth manifold and $F\colon M\to V$ a smooth map. Denoting the exterior derivative by d_e, we have $D_{F(p)}\circ\mathrm{d}_eF_p=\mathrm{d}F_p$ for all $p\in M$, where $D_{F(p)}\colon V\to T_{F(p)}V$ is the canonical isomorphism as in Lemma A.3.*

Proof. Choose a basis $e_1,\dots,e_n$ for V and let $\phi=(x^1,\dots,x^n)\colon V\to\mathbb{R}^n$ be the associated smooth coordinate chart for V as in (A.1), then $F=(x^i\circ F)e_i$. The isomorphism $D_{F(p)}$ maps the e_i to the coordinate vectors associated to the chart ϕ, so

$$(D_{F(p)}\circ\mathrm{d}_eF_p)(v)=D_{F(p)}(v(x^i\circ F)e_i)=v(x^i\circ F)D_{F(p)}(e_i)=v(x^i\circ F)\left.\frac{\partial}{\partial x^i}\right|_{F(p)}$$

for all $v\in T_pM$ by linearity. On the other hand, the differential $\mathrm{d}F_p$ of F at p satisfies

$$\mathrm{d}F_p(v)=\mathrm{d}_ex^i\big|_{F(p)}(\mathrm{d}F_p(v))\left.\frac{\partial}{\partial x^i}\right|_{F(p)}=\mathrm{d}_e(x^i\circ F)_p(v)\left.\frac{\partial}{\partial x^i}\right|_{F(p)}=v(x^i\circ F)\left.\frac{\partial}{\partial x^i}\right|_{F(p)}$$

for all $v \in T_pM$ since the $\mathrm{d}_e x^i\big|_{F(p)}$ form a basis of T_p^*M dual to the coordinate vector basis of T_pM. To be able to write $\mathrm{d}_e x^i\big|_{F(p)} \circ \mathrm{d}F_p = \mathrm{d}_e(x^i \circ F)_p$, we used that the claim is known to hold for $V = \mathbb{R}$. □

As a consequence, we have the following chain rules for the exterior derivative.

Lemma 4.8. *Let M and N be smooth manifolds, $F\colon M \to N$ and $G\colon N \to V$ smooth maps and $L\colon V \to W$ a linear map. Denoting the exterior derivative by d_e, we have:*

(i) $\mathrm{d}_e(G \circ F)_p = \mathrm{d}_e G_{F(p)} \circ \mathrm{d}F_p$ *for all* $p \in M$*;*

(ii) $\mathrm{d}_e(L \circ G)_q = L \circ \mathrm{d}_e G_q$ *for all* $q \in N$*.*

Proof.

(i) This follows directly from Lemma 4.7 together with the chain rule for the differential.

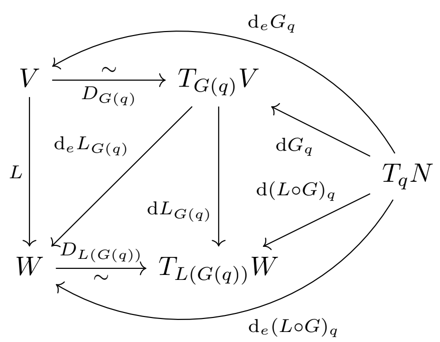
(ii) Lemmas 4.7 and A.3 together imply

$$\mathrm{d}_e L_{G(q)} = D^{-1}_{L(G(q))} \circ \mathrm{d}L_{G(q)} = L \circ D^{-1}_{G(q)},$$

so with (i), we find

$$\mathrm{d}_e(L \circ G)_q = \mathrm{d}_e L_{G(q)} \circ \mathrm{d}G_q = L \circ D^{-1}_{G(q)} \circ \mathrm{d}G_q = L \circ \mathrm{d}_e G_q.$$ □

The proof of (ii) boils down to the commutativity of the following diagram:



Pullbacks. Let M and N be smooth manifolds, $F\colon M \to N$ a smooth map and ω a V-valued k-form on N, then the *pullback* of ω along F is the V-valued k-form $F^*\omega$ on M defined by

$$(F^*\omega)_p(v_1, \ldots, v_k) = \omega_{F(p)}(\mathrm{d}F_p(v_1), \ldots, \mathrm{d}F_p(v_k))$$

for all $p \in M$ and $v_1, \ldots, v_k \in T_pM$. If ω is smooth, then so is $F^*\omega$, as shown by the following lemma.

Lemma 4.9. *Let M and N be smooth manifolds, $F\colon M \to N$ a smooth map, V a finite-dimensional $\mathbb{R}$-vector space and $k \in \mathbb{Z}_{\geq 0}$. Let $e_1, \ldots, e_n \in V$ be vectors and $\omega^1, \ldots, \omega^n$ forms of degree k on N, then $F^*(\omega^i e_i) = (F^*\omega^i)e_i$.*

Proof. Set $\omega \coloneqq \omega^i e_i$, then for all $p \in M$ and $v_1, \ldots, v_k \in T_pM$,

$$\begin{aligned}(F^*\omega)_p(v_1, \ldots, v_k) &= \omega_{F(p)}(\mathrm{d}F_p(v_1), \ldots, \mathrm{d}F_p(v_k)) \\ &= \omega^i_{F(p)}(\mathrm{d}F_p(v_1), \ldots, \mathrm{d}F_p(v_k))e_i \\ &= (F^*\omega^i)_p(v_1, \ldots, v_k)e_i.\end{aligned}$$ □

Many important properties of pullbacks of $\mathbb{R}$-valued forms transfer over to the more general setting. For a proof, we refer to [9, Proposition 21.8].

Lemma 4.10. *Let M, N and P be smooth manifolds, $G\colon M \to N$ and $F\colon N \to P$ smooth maps and $\mu\colon V \times W \to Z$ a bilinear map. Let $\omega \in \Omega^k(P, V)$ and $\eta \in \Omega^\ell(P, W)$, then:*

(i) $(F \circ G)^*\omega = G^*(F^*\omega)$;

(ii) $F^*\mu(\omega, \eta) = \mu(F^*\omega, F^*\eta)$;

(iii) $F^*(\mathrm{d}\omega) = \mathrm{d}(F^*\omega)$.

4.3 Lie algebra valued forms

Let G be a Lie group. From now on, we focus on differential forms with values in the Lie algebra $\mathrm{Lie}(G)$ of G. The Lie bracket is a bilinear map $\mathrm{Lie}(G) \times \mathrm{Lie}(G) \to \mathrm{Lie}(G)$; the product of $\mathrm{Lie}(G)$-valued forms ω and η on a smooth manifold M with respect to this bilinear map is called their *Lie bracket* and written as $[\omega, \eta]$. It satisfies the following "graded anticommutativity" property, proved in [9, Proposition 21.5].

Lemma 4.11. *Let M be a smooth manifold, ω a $\mathrm{Lie}(G)$-valued k-form on M and η a $\mathrm{Lie}(G)$-valued ℓ-form, then*

$$[\omega, \eta] = (-1)^{k\ell+1}[\eta, \omega].$$

In particular, if ω and η both have degree 1, we get $[\omega, \eta] = [\eta, \omega]$ and $[\omega, \omega]$ need not be the zero form.

Also important in this context is the following special case of Example 4.4 (2). Recall from Lemma 2.17 that the adjoint representation of G is a smooth group homomorphism $\mathrm{Ad}\colon G \to \mathrm{GL}(\mathrm{Lie}(G))$, so we can interpret it as a smooth 0-form on G with values in the vector space $\mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$ of linear maps $\mathrm{Lie}(G) \to \mathrm{Lie}(G)$. Its product $\mathrm{Ad}\,\omega$ with a $\mathrm{Lie}(G)$-valued k-form ω on G is the pointwise composition, i.e. $(\mathrm{Ad}\,\omega)_g = \mathrm{Ad}_g \circ \omega_g$ for all $g \in G$. Its exterior derivative $\mathrm{d}\,\mathrm{Ad} \in \Omega^1(G, \mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G)))$ is strongly related to the Lie bracket.

Lemma 4.12. *Let $g \in G$ and $X, Y \in \mathrm{Lie}(G)$, then $\mathrm{d}\,\mathrm{Ad}_g(X_g)Y = \mathrm{Ad}_g([X, Y])$.*

Proof. This follows immediately from Lemmas 2.18 and 4.7. □

If M is a smooth manifold and $\rho\colon M \to G$ any map, we now have two product operations on $\mathrm{Lie}(G)$-valued forms on M: the Lie bracket and the pointwise composition with the smooth $\mathrm{End}_{\mathbb{R}}(\mathrm{Lie}(G))$-valued 0-form $\mathrm{Ad}_\rho \coloneqq \mathrm{Ad} \circ \rho$. These two operations behave well together.

Lemma 4.13. *Let M be a smooth manifold and $\rho\colon M \to G$ any map. Let ω and η be $\mathrm{Lie}(G)$-valued forms on M of degree k and ℓ, respectively, then*

$$\mathrm{Ad}_\rho[\omega, \eta] = [\mathrm{Ad}_\rho\,\omega, \mathrm{Ad}_\rho\,\eta].$$

Proof. Let $p \in M$ and $v_1, \ldots, v_{k+\ell} \in T_pM$, then since $\mathrm{Ad}_{\rho(p)}$ is a Lie algebra homomorphism,

$$\begin{aligned}
&\mathrm{Ad}_{\rho(p)} \circ [\omega_p, \eta_p](v_1, \ldots, v_{k+\ell}) \\
&\quad = \mathrm{Ad}_{\rho(p)}\left(\frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn}\sigma)[\omega_p(v_{\sigma(1)}, \ldots, v_{\sigma(k)}), \eta_p(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)})]\right) \\
&\quad = \frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn}\sigma)\,\mathrm{Ad}_{\rho(p)}([\omega_p(v_{\sigma(1)}, \ldots, v_{\sigma(k)}), \eta_p(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)})]) \\
&\quad = \frac{1}{k!\ell!}\sum_{\sigma \in S_{k+\ell}} (\operatorname{sgn}\sigma)[\mathrm{Ad}_{\rho(p)} \circ \omega_p(v_{\sigma(1)}, \ldots, v_{\sigma(k)}), \mathrm{Ad}_{\rho(p)} \circ \eta_p(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+\ell)})] \\
&\quad = [\mathrm{Ad}_{\rho(p)} \circ \omega_p, \mathrm{Ad}_{\rho(p)} \circ \eta_p](v_1, \ldots, v_{k+\ell}).
\end{aligned}$$

□

4.4 The Maurer-Cartan form

On any Lie group G, there is a distinguished $\mathrm{Lie}(G)$-valued form which naturally occurs in many situations.

Definition 4.14. The *Maurer-Cartan form* on G is the $\mathrm{Lie}(G)$-valued 1-form ω_G on G defined by

$$(\omega_G)_g := \epsilon^{-1} \circ \mathrm{d}(L_{g^{-1}})_g \colon T_g G \to \mathrm{Lie}(G)$$

for any $g \in G$, where $\epsilon\colon \mathrm{Lie}(G) \to T_e G$ is the vector space isomorphism from Lemma 2.8.

The Maurer-Cartan form has a number of very useful properties, see also [9, Exercise 21.10].

Lemma 4.15. *The Maurer-Cartan form ω_G on G satisfies the following properties:*

(i) *$\omega_G(X)$ is the constant map $g \mapsto X$ for all $X \in \mathrm{Lie}(G)$;*

(ii) *$L_g^* \omega_G = \omega_G$ for all $g \in G$;*

(iii) *$R_g^* \omega_G = \mathrm{Ad}_{g^{-1}} \omega_G$ for all $g \in G$.*

Proof.

(i) Let $X \in \mathrm{Lie}(G)$, then

$$\omega_G(X)(g) = (\omega_G)_g(X_g) = \epsilon^{-1}(\mathrm{d}(L_{g^{-1}})_g(X_g)) = \epsilon^{-1}(X_e) = X$$

for all $g \in G$ since X is left-invariant.

(ii) Let $g, h \in G$, then

$$(L_g^* \omega_G)_h = (\omega_G)_{gh} \circ \mathrm{d}(L_g)_h = \epsilon^{-1} \circ \mathrm{d}(L_{h^{-1}g^{-1}} \circ L_g)_h = \epsilon^{-1} \circ \mathrm{d}(L_{h^{-1}})_h = (\omega_G)_h.$$

(iii) Let $g, h \in G$, then

$$(R_g^* \omega_G)_h = (\omega_G)_{hg} \circ \mathrm{d}(R_g)_h = \epsilon^{-1} \circ \mathrm{d}(L_{g^{-1}h^{-1}} \circ R_g)_h = \epsilon^{-1} \circ \mathrm{d}(\mathrm{ad}_{g^{-1}} \circ L_{h^{-1}})_h$$

since $(L_{g^{-1}h^{-1}} \circ R_g)(k) = g^{-1}h^{-1}kg = (\mathrm{ad}_{g^{-1}} \circ L_{h^{-1}})(k)$ for all $k \in G$. Using the definition of the adjoint representation (see Equation (2.1)), we find

$$(R_g^* \omega_G)_h = \mathrm{Ad}_{g^{-1}} \circ \epsilon^{-1} \circ \mathrm{d}(L_{h^{-1}})_h = \mathrm{Ad}_{g^{-1}} \circ (\omega_G)_h.$$ □

We can use the first of these three properties to show that ω_G is smooth.

Lemma 4.16. *The Maurer-Cartan form ω_G on G is smooth.*

Proof. By Lemma 4.2, it suffices to show that $\omega_G(X) \in C^\infty(M, \mathrm{Lie}(G))$ for all $X \in \mathfrak{X}(G)$, so let $X \in \mathfrak{X}(G)$ be a smooth vector field on G. From Lemma 2.10, we know $X = f^i X_i$ for some $f^1, \ldots, f^n \in C^\infty(G)$ and $X_1, \ldots, X_n \in \mathrm{Lie}(G)$, so $\omega_G(X) = f^i \omega_G(X_i)$. Now, by Lemma 4.15, $\omega_G(X_i)$ is constant for all i, so $\omega_G(X)$ is smooth and we are done. □

Our next lemma is inspired by [9, Exercise 21.9].

Lemma 4.17. *ω_G satisfies the* Maurer-Cartan equation

$$\mathrm{d}\omega_G + \frac{1}{2}[\omega_G, \omega_G] = 0.$$

Proof. Let $X, Y \in \mathrm{Lie}(G)$, then Lemma 4.6 gives

$$\mathrm{d}\omega_G(X, Y) = X(\omega_G(Y)) - Y(\omega_G(X)) - \omega_G([X, Y]).$$

By Lemma 4.15, $\omega_G(X)$, $\omega_G(Y)$ and $\omega_G([X, Y])$ are the constant functions X, Y and $[X, Y]$, so

$$\mathrm{d}\omega_G(X, Y)(g) = -\omega_G([X, Y])(g) = -[X, Y]$$

for all $g \in G$. On the other hand,

$$\begin{aligned}\frac{1}{2}[\omega_G, \omega_G](X, Y)(g) &= \frac{1}{2}\big([(\omega_G)_g(X_g), (\omega_G)_g(Y_g)] - [(\omega_G)_g(Y_g), (\omega_G)_g(X)]\big)\\ &= [\omega_G(X)(g), \omega_G(Y)(g)] = [X, Y],\end{aligned}$$

for all $g \in G$ by definition of the Lie bracket of forms, so

$$\mathrm{d}\omega_G(X, Y) + \frac{1}{2}[\omega_G, \omega_G](X, Y) = 0.$$

The claim now follows from $C^\infty(G)$-linearity together with Lemma 2.10. □

4.5 The logarithmic derivative

Now if G is a Lie group, its Maurer-Cartan form ω_G can be used to extend the notion of the logarithmic derivative of a real-valued function to that of a map $M \to G$, for any smooth manifold M. Suppose that $f\colon \mathbb{R} \to \mathbb{R}_{>0}$ is smooth, then we can view $\log f\colon \mathbb{R} \to \mathbb{R}$ as a smooth 0-form on $\mathbb{R}$. Its exterior derivative at any point $t_0 \in \mathbb{R}$ is

$$\mathrm{d}(\log f)_{t_0} = \frac{\mathrm{d}\log f}{\mathrm{d}t}(t_0)\,\mathrm{d}t|_{t_0} = \frac{f'(t_0)\,\mathrm{d}t|_{t_0}}{f(t_0)} = \frac{\mathrm{d}f_{t_0}}{f(t_0)} = \mathrm{d}(f(t_0)^{-1}f)_{t_0}$$

by the linearity of d. This inspires the following definition.

Definition 4.18. Let M be a smooth manifold and $\rho\colon M \to G$ a smooth map. The *logarithmic derivative* of ρ is the $\mathrm{Lie}(G)$-valued 1-form $\mathrm{d}\log\rho$ on M defined for any $p \in M$ by

$$\mathrm{d}\log\rho_p \coloneqq \epsilon^{-1} \circ \mathrm{d}(\rho(p)^{-1}\rho)_p = \epsilon^{-1} \circ \mathrm{d}(L_{\rho(p)^{-1}} \circ \rho)_p\colon T_pM \to \mathrm{Lie}(G),$$

where ϵ is the isomorphism from Lemma 2.8 and we have used the product notation (2.2).

The link between Definition 4.18 and the logarithmic derivative of real-valued functions is further solidified by the next lemma.

Lemma 4.19. *Suppose G is abelian, then $\mathrm{d}\log\exp_X = D_X^{-1}$ for any $X \in \mathrm{Lie}(G)$, where D_X is the canonical isomorphism $\mathrm{Lie}(G) \to T_X\,\mathrm{Lie}(G)$ from Lemma A.3.*

Proof. Let $X \in \mathrm{Lie}(G)$ and let $T\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$, $Y \mapsto -X + Y$ be the translation by $-X$, then

$$\exp(X)^{-1}\exp(Y) = \exp(-X + Y) = (\exp \circ T)(Y)$$

for all $Y \in \mathrm{Lie}(G)$ by Lemma 2.21 (3) and (5) since G is abelian, so $\exp(X)^{-1}\exp = \exp \circ T$. Using Lemma 2.21 (6) and Lemma A.4, we find

$$\mathrm{d}\log\exp_X \circ D_X = \epsilon^{-1} \circ \mathrm{d}(\exp \circ T)_X \circ D_X = \epsilon^{-1} \circ \mathrm{d}\exp_0 \circ D_{X-X} = \mathrm{id}_{\mathrm{Lie}(G)}$$

and thus $\mathrm{d}\log\exp = D_X^{-1}$, as required. □

Definition 4.18 is really nothing new: the logarithmic derivative of $\rho\colon M \to G$ is just the pullback of the Maurer-Cartan form ω_G on G along ρ.

Lemma 4.20. *Let M be a smooth manifold and $\rho\colon M \to G$ a smooth map, then $\mathrm{d}\log\rho = \rho^*\omega_G$.*

Proof. This follows directly from the definitions: for any $p \in M$,

$$\mathrm{d}\log\rho_p = \epsilon^{-1} \circ \mathrm{d}(L_{\rho(p)^{-1}})_{\rho(p)} \circ \mathrm{d}\rho_p = (\omega_G)_{\rho(p)} \circ \mathrm{d}\rho_p = (\rho^*\omega_G)_p. \qquad \square$$

In particular, this implies that $\mathrm{d}\log\rho$ is smooth, since ω_G is smooth by Lemma 4.16. Lemma 4.20 can also be used to translate many of the useful properties of the Maurer-Cartan form to statements about logarithmic derivatives.

Lemma 4.21. *Let M and N be smooth manifolds and $F\colon N \to M$, $\rho, g\colon M \to G$ smooth maps, then:*

(i) $F^*\mathrm{d}\log\rho = \mathrm{d}\log(\rho \circ F)$;

(ii) $\mathrm{d}\log(\rho \cdot g) = \mathrm{Ad}_{g^{-1}}\mathrm{d}\log\rho + \mathrm{d}\log g$;

(iii) $\mathrm{d}(\mathrm{d}\log\rho) = -\frac{1}{2}[\mathrm{d}\log\rho, \mathrm{d}\log\rho]$.

Proof.

(i) This follows directly from Lemmas 4.20 and 4.10.

(ii) Let $p \in M$, then Lemmas 4.20 and 2.19 give

$$\begin{aligned}
\mathrm{d}\log(\rho \cdot g)_p &= (\omega_G)_{\rho(p)g(p)} \circ \mathrm{d}(\rho \cdot g)_p \\
&= (\omega_G)_{\rho(p)g(p)} \circ (\mathrm{d}(\rho g(p))_p + \mathrm{d}(\rho(p)g)_p) \\
&= (R^*_{g(p)}\omega_G)_{\rho(p)} \circ \mathrm{d}\rho_p + (L^*_{\rho(p)}\omega_G)_{g(p)} \circ \mathrm{d}g_p.
\end{aligned}$$

Using Lemma 4.15, we now find

$$\begin{aligned}
\mathrm{d}\log(\rho \cdot g)_p &= \mathrm{Ad}_{g(p)^{-1}} \circ (\omega_G)_{\rho(p)} \circ \mathrm{d}\rho_p + (\omega_G)_{g(p)} \circ \mathrm{d}g_p \\
&= \mathrm{Ad}_{g(p)^{-1}} \circ (\rho^*\omega_G)_p + (g^*\omega_G)_p \\
&= \mathrm{Ad}_{g(p)^{-1}} \circ \mathrm{d}\log\rho_p + \mathrm{d}\log g_p.
\end{aligned}$$

(iii) This is just an application of the Maurer-Cartan equation (Lemma 4.17), combined with the fact that pullbacks commute with both exterior derivatives and products (Lemma 4.10):

$$\begin{aligned}
\mathrm{d}(\mathrm{d}\log\rho) &= \mathrm{d}(\rho^*\omega_G) = \rho^*(\mathrm{d}\omega_G) = \rho^*\left(-\frac{1}{2}[\omega_G, \omega_G]\right) \\
&= -\frac{1}{2}[\rho^*\omega_G, \rho^*\omega_G] = -\frac{1}{2}[\mathrm{d}\log\rho, \mathrm{d}\log\rho]. \qquad \square
\end{aligned}$$

Logarithmic derivatives of G-equivariant maps have some additional properties which will be useful in the discussion of connections on principal bundles later on.

Lemma 4.22. *Let M be a smooth manifold, θ a smooth right action of G on M and $\underline{\theta}\colon \mathrm{Lie}(G) \to \mathfrak{X}(M)$, $X \mapsto \underline{X}$ the corresponding infinitesimal generator. Let $\rho\colon M \to G$ be a smooth G-equivariant map, where G is considered as a right G-manifold under multiplication, and $p \in M$ a point, then:*

(i) $\mathrm{d}\log\rho_p(\underline{X}_p) = X$ *for all* $X \in \mathrm{Lie}(G)$;

(ii) $(\theta_g^*\mathrm{d}\log\rho)_p = \mathrm{Ad}_{g^{-1}} \circ \mathrm{d}\log\rho_p$ *for all* $g \in G$.

Proof.

(i) Let $X \in \mathrm{Lie}(G)$, then $\underline{X}$ and X are ρ-related by Example 2.24, Lemma 2.26 and the G-equivariance of ρ, so $\mathrm{d}\rho_p(\underline{X}_p) = X_{\rho(p)}$ for all $p \in M$. Property (i) in Lemma 4.15 now gives

$$\mathrm{d}\log\rho_p(\underline{X}_p) = (\rho^*\omega_G)_p(\underline{X}_p) = (\omega_G)_{\rho(p)}(X_{\rho(p)}) = X.$$

(ii) The G-equivariance of ρ implies $\rho \circ \theta_g = \rho \cdot g$, so by Lemma 4.21,

$$(\theta_g^* \,\mathrm{d}\log\rho)_p = \mathrm{d}\log(\rho\circ\theta_g)_p = \mathrm{d}\log(\rho\cdot g)_p = \mathrm{Ad}_{g^{-1}} \circ \mathrm{d}\log\rho_p.$$

In the last equality, we used that the logarithmic derivative of a constant function vanishes. □

The logarithmic derivative also pops up when computing the exterior derivative of the pointwise composition $\mathrm{Ad}_\rho\,\omega$ of a smooth $\mathrm{Lie}(G)$-valued form ω on a smooth manifold M with $\mathrm{Ad}\circ\rho$, for some smooth map $\rho\colon M \to G$.

Lemma 4.23. *Let M be a smooth manifold, $\rho\colon M \to G$ a smooth map and $\omega \in \Omega^k(M, \mathrm{Lie}(G))$, then*

$$\mathrm{d}(\mathrm{Ad}_\rho\,\omega) = \mathrm{Ad}_\rho\,\mathrm{d}\omega + (-1)^k[\mathrm{Ad}_\rho\,\omega, \mathrm{d}\log(\rho^{-1})].$$

Proof. By Lemma 4.5, it suffices to show

$$\mathrm{d}(\mathrm{Ad}_\rho)\omega = (-1)^k[\mathrm{Ad}_\rho\,\omega, \mathrm{d}\log(\rho^{-1})].$$

Let $p \in M$ and $v \in T_pM$, then $\mathrm{d}\log\rho_p(v) \in \mathrm{Lie}(G)$ satisfies

$$\mathrm{d}\log\rho_p(v)\big|_{\rho(p)} = \mathrm{d}(L_{\rho(p)})_e \circ \mathrm{d}(\rho(p)^{-1}\rho)_p(v) = \mathrm{d}\rho_p(v)$$

by left invariance, so Lemma 4.12 gives

$$\mathrm{d}(\mathrm{Ad}_\rho)_p(v)Y = \mathrm{d}(\mathrm{Ad}\circ\rho)_p(v)Y = \mathrm{d}\,\mathrm{Ad}_{\rho(p)}(\mathrm{d}\log\rho_p(v)\big|_{\rho(p)})Y = \mathrm{Ad}_{\rho(p)}([\mathrm{d}\log\rho_p(v), Y])$$

for all $Y \in \mathrm{Lie}(G)$. It follows that

$$\begin{aligned}
(\mathrm{d}(\mathrm{Ad}_\rho)\omega)_p(v_1,\dots,v_{k+1}) &= \frac{1}{k!}\sum_{\sigma\in S_{k+1}} (\operatorname{sgn}\sigma)\,\mathrm{d}(\mathrm{Ad}_\rho)_p(v_{\sigma(1)})\omega_p(v_{\sigma(2)},\dots,v_{\sigma(k+1)}) \\
&= \frac{1}{k!}\sum_{\sigma\in S_{k+1}} (\operatorname{sgn}\sigma)\,\mathrm{Ad}_{\rho(p)}([\mathrm{d}\log\rho_p(v_{\sigma(1)}), \omega_p(v_{\sigma(2)},\dots,v_{\sigma(k+1)})]) \\
&= \frac{1}{k!}\sum_{\sigma\in S_{k+1}} (\operatorname{sgn}\sigma)[\mathrm{Ad}_{\rho(p)}\circ\mathrm{d}\log\rho_p(v_{\sigma(1)}), \mathrm{Ad}_{\rho(p)}\circ\omega_p(v_{\sigma(2)},\dots,v_{\sigma(k+1)})] \\
&= [\mathrm{Ad}_\rho\,\mathrm{d}\log\rho, \mathrm{Ad}_\rho\,\omega]_p(v_1,\dots,v_{k+1})
\end{aligned}$$

for any $v_1,\dots,v_{k+1} \in T_pM$, so

$$\mathrm{d}(\mathrm{Ad}_\rho)\omega = [\mathrm{Ad}_\rho\,\mathrm{d}\log\rho, \mathrm{Ad}_\rho\,\omega] = (-1)^{k+1}[\mathrm{Ad}_\rho\,\omega, \mathrm{Ad}_\rho\,\mathrm{d}\log\rho]$$

by Lemma 4.11. To conclude the proof, note that

$$\begin{aligned}
(\mathrm{Ad}_\rho\,\mathrm{d}\log\rho)_p(v) &= \mathrm{Ad}_{\rho(p)}\circ\epsilon^{-1}\circ\mathrm{d}(\rho(p)^{-1}\rho)_p(v) \\
&= \epsilon^{-1}\circ\mathrm{d}(\mathrm{ad}_{\rho(p)}\circ\rho(p)^{-1}\rho)_p(v) \\
&= \epsilon^{-1}\circ\mathrm{d}(\rho\rho(p)^{-1})_p(v),
\end{aligned}$$

where we have used the definition of the adjoint representation (see Equation (2.1)). Lemma 2.19 now gives $\mathrm{d}(\rho(p)\rho^{-1})_p + \mathrm{d}(\rho\rho(p)^{-1})_p = 0$ since $\rho\cdot\rho^{-1}$ is constant, so $\mathrm{Ad}_\rho\,\mathrm{d}\log\rho = -\mathrm{d}\log(\rho^{-1})$ and we are done. □

Finally, consider the special case that G is the unit circle $U(1) \subseteq \mathbb{C}$. The inclusion $j_0 \colon U(1) \hookrightarrow \mathbb{C}$ is a smooth $\mathbb{C}$-valued 0-form on $U(1)$, and its exterior derivative $\mathrm{d}(j_0)_1$ gives a way of identifying $T_1U(1)$ with a subspace of $\mathbb{C}$.

Lemma 4.24. *The exterior derivative* $\mathrm{d}(j_0)_1 \colon T_1U(1) \to \mathbb{C}$ *of* j_0 *at* 1 *is an isomorphism onto* $i\mathbb{R}$ *mapping the angle coordinate vector* $\mathrm{d}/\mathrm{d}\theta\,|_1$ *from Example 2.9 to* i*.*

Proof. Recall from Lemma 4.7 that the exterior derivative of j_0 at 1 is the composition of its differential $\mathrm{d}(j_0)_1 \colon T_1U(1) \to T_1\mathbb{C}$ with the inverse of the vector space isomorphism $D_1 \colon \mathbb{C} \to T_1\mathbb{C}$ from Lemma A.3. $U(1)$ is an embedded submanifold of $\mathbb{C}$, so j_0 is an immersion and $\mathrm{d}(j_0)_1$ is injective. Let $f \in C^\infty(\mathbb{C})$, then

$$\mathrm{d}(j_0)_1\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\bigg|_1\right) f = \frac{\mathrm{d}}{\mathrm{d}\theta}\bigg|_1 (f \circ j_0) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 (f \circ j_0 \circ \theta^{-1}) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(\exp(it))$$

by definition of the angle coordinate vector. On the other hand,

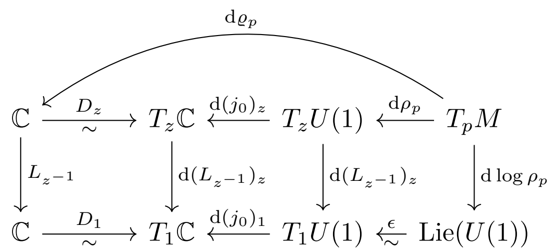
$$D_1(i)f = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(1 + it) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(\exp(it)),$$

so $\mathrm{d}(j_0)_1(\,\mathrm{d}/\mathrm{d}\theta\,|_1) = D_1(i)$ and we are done. ☐

By composing with the evaluation map $\epsilon\colon \operatorname{Lie}(U(1)) \to T_1U(1)$ from Lemma 2.8, we thus obtain an isomorphism $\mathrm{d}(j_0)_1 \circ \epsilon\colon \operatorname{Lie}(U(1)) \to i\mathbb{R}$. Under this identification, the logarithmic derivative has the following alternative characterization.

Lemma 4.25. *Let* M *be a smooth manifold,* $\rho\colon M \to U(1)$ *a smooth map and set* $\varrho \coloneqq j_0 \circ \rho\colon M \to \mathbb{C}$*. Considering* ϱ *as a smooth* $\mathbb{C}$*-valued* 0*-form on* M*, we have* $\mathrm{d}\log\rho = \varrho^{-1}\,\mathrm{d}\varrho$ *under the canonical identification* $\operatorname{Lie}(U(1)) \cong i\mathbb{R}$*, where the product of forms is the one induced by complex multiplication.*

Proof. The claim follows from the fact that the diagram



is commutative for all $p \in M$, $z \coloneqq \rho(p) \in U(1)$. The left square commutes by Lemma A.3, the right one by definition of the logarithmic derivative and the middle one by the fact that $L_{z^{-1}} \circ j_0 = j_0 \circ L_{z^{-1}}$. For the map $\mathrm{d}\varrho_p$ on top, we used Lemma 4.7. ☐

Chapter 5

Connections on principal bundles

With a solid foundation in the theory of Lie groups, fiber bundles and vector valued differential forms, we are finally in a position to start talking about connections on principal bundles. Consider the k-sheeted covering $\pi\colon S^1 \to S^1$, $z \mapsto z^k$ of the circle $S^1 \subseteq \mathbb{C}^*$. Let $z \in S^1$ be a point and $\gamma\colon [0,1] \to S^1$ a path starting at z. It is a well-known fact in topology that for any $w \in \pi^{-1}\{z\}$, there is a unique lift of γ starting at w, i.e. a path $\tilde{\gamma}\colon [0,1] \to S^1$ with $\pi \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = w$. This can also be interpreted as there being a canonical way to "parallel transport" the point w in the fiber of π over z along the path γ.

In Example 3.7, it was shown that π is a principal C_k-bundle with C_k the cyclic group of order k, which begs the question: does every principal G-bundle $\pi\colon P \to M$ have this path lifting property? The answer is a resounding no. Instead, there are natural objects which provide the bundle with precisely such a notion of parallel transport. These objects are called *connections*, and form the subject of this chapter. As alluded to at the end of Chapter 3, they can be defined either as subbundles of the tangent bundle of P, or as 1-forms on P with values in the Lie algebra $\mathrm{Lie}(G)$ of G. One way to construct such a 1-form is to cover the base space M with open subsets U_α and define a $\mathrm{Lie}(G)$-valued 1-form $\mathcal{A}_\alpha$ on every U_α, in such a way that a certain *compatibility condition* is satisfied. This construction will be important later in the geometric interpretation of the TKNN formula, as will the definition of the *curvature* of a connection.

In this entire chapter, $\pi\colon P \to M$ is a principal G-bundle, with G a Lie group. We denote the smooth right action of G on P by θ and the corresponding infinitesimal generator by $\underline{\theta}\colon \mathrm{Lie}(G) \to \mathfrak{X}(P)$, $X \mapsto \underline{X}$.

5.1 Connections

Geometrically, a connection on the principal bundle π is a smooth choice of a *horizontal tangent subspace* H_p of T_pP for all $p \in P$ such that $T_pP = H_p \oplus V_p$, where V_p is the vertical tangent subspace as defined in Section 3.4. Choosing such a complementary subspace of V_p in T_pP is equivalent to choosing a projection $v_p\colon T_pP \to V_p$, under the correspondence $v_p \mapsto \ker v_p$. Via the isomorphism $\eta_p\colon \mathrm{Lie}(G) \to V_p$ from Lemma 3.20, we can interpret v_p as a linear map $T_pP \to \mathrm{Lie}(G)$, which leads to the following definition.

Definition 5.1. A *connection (form)* on π is a smooth $\mathrm{Lie}(G)$-valued one-form $\omega \in \Omega^1(P, \mathrm{Lie}(G))$ on P satisfying the following properties for all $p \in P$:

(i) $\omega_p(\underline{X}_p) = X$ for all $X \in \mathrm{Lie}(G)$;

(ii) $(\theta_g^*\omega)_p = \mathrm{Ad}_{g^{-1}} \circ \omega_p$ for all $g \in G$.

For all $p \in P$, we call $\ker \omega_p \subseteq T_pP$ the *horizontal tangent subspace* of T_pP.

The first property in Definition 5.1 means that ω_p is required to be a projection onto V_p for all $p \in P$, under the identification $\mathrm{Lie}(G) \cong V_p$. The second implies that once a horizontal tangent subspace is chosen at some point $p \in P$, it is immediately fixed for the whole fiber $P_{\pi(p)}$ of π containing p, since

$$\ker \omega_{pg} = \ker(\mathrm{Ad}_g \circ \omega_{pg}) = \ker(\theta^*_{g^{-1}}\omega)_{pg} = \ker(\omega_p \circ \mathrm{d}(\theta_g)_p^{-1}) = \mathrm{d}(\theta_g)_p(\ker \omega_p)$$

for any $g \in G$. That is, the horizontal tangent space at pg is the pushforward of the one at p along the right multiplication map θ_g. We have already seen a few examples of connection forms.

Example 5.2.

(1) By Lemma 4.15, the Maurer-Cartan form $\omega_G \in \Omega^1(G, \mathrm{Lie}(G))$ on the Lie group G is a connection form on the principal G-bundle $G \to \{*\}$ over the one-point space $\{*\}$.

(2) Let $\rho\colon P \to G$ be any G-equivariant map where we consider G as a right G-manifold under multiplication, then $\mathrm{d}\log \rho \in \Omega^1(P, \mathrm{Lie}(G))$ is a connection form on P by Lemma 4.22. △

It is always possible to pull back connection forms along morphisms of principal G-bundles.

Proposition 5.3. *Let $\pi'\colon Q \to N$ be another principal G-bundle and $F\colon P \to Q$ a principal G-bundle morphism. If ω is a connection form on π', then $F^*\omega$ is a connection form on π.*

Proof. We know $F^*\omega$ is a smooth $\mathrm{Lie}(G)$-valued one-form on P, so it remains to check the two properties in Definition 5.1. Denote by ϑ the right action of G on Q, by $\underline{\vartheta}$ the corresponding infinitesimal generator and let $p \in P$.

(i) Let $X \in \mathrm{Lie}(G)$. The vector fields $\underline{\theta}(X) \in \mathfrak{X}(P)$ and $\underline{\vartheta}(X) \in \mathfrak{X}(Q)$ are F-related by Lemma 2.26 since F is G-equivariant, so

$$(F^*\omega)_p(\underline{\theta}(X)_p) = \omega_{F(p)}(\mathrm{d}F_p(\underline{\theta}(X)_p)) = \omega_{F(p)}(\underline{\vartheta}(X)_{F(p)}) = X$$

since ω is a connection form on π'.

(ii) Let $g \in G$, then $F \circ \theta_g = \vartheta_g \circ F$ by the G-equivariance of F, so

$$\begin{aligned}(\theta_g^* F^*\omega)_p &= ((F \circ \theta_g)^*\omega)_p = ((\vartheta_g \circ F)^*\omega)_p = (F^*\vartheta_g^*\omega)_p \\ &= (\vartheta_g^*\omega)_{F(p)} \circ \mathrm{d}F_p = \mathrm{Ad}_{g^{-1}} \circ \omega_{F(p)} \circ \mathrm{d}F_p = \mathrm{Ad}_{g^{-1}} \circ (F^*\omega)_p.\end{aligned}$$ □

For trivial bundles, there is a natural choice because they have a globally defined projection on the *second* coordinate.

Lemma 5.4. *Suppose $\pi\colon P = M \times G \to M$ is a product bundle and let $\pi_2\colon P \to G$ be the projection onto the second coordinate, then $\mathrm{d}\log \pi_2$ is a connection form on P with horizontal subspace $\ker \mathrm{d}(\pi_2)_p$ at any point $p \in P$.*

Proof. The projection π_2 is smooth and G-equivariant by definition of the G-action on P, so this is a special case of Example 5.2 (2). To see why $\ker \mathrm{d}(\pi_2)_p$ is the horizontal tangent space at $p \in P$, note that $\mathrm{d}\log(\pi_2)_p$ is the composition of $\mathrm{d}(\pi_2)_p$ with the vector space isomorphisms $\mathrm{d}(L_{\pi_2(p)^{-1}})_{\pi_2(p)}$ and ϵ^{-1} for all $p \in P$, so its kernel is $\ker \mathrm{d}(\pi_2)_p$. □

5.2 The tautological bundle revisited

Another important case in which there exists a canonical choice of connection is when π is the tautological bundle γ_n constructed in Section 3.3. Recall its definition $\gamma_n\colon S \to \mathbb{P}(\mathcal{H})$, $z \mapsto [z]$ where $\mathcal{H} := \mathbb{C}^{n+1}$ is supplied with the Euclidean norm, $\mathbb{P}(\mathcal{H})$ is its projectivization and S is the unit sphere in $\mathcal{H}$. Remember also that the right action θ of $U(1)$ on S is just scalar multiplication.

Now, the norm on $\mathcal{H}$ is induced by the Euclidean inner product $\langle\cdot|\cdot\rangle$ on $\mathcal{H}$ and the inclusion $j\colon S\hookrightarrow\mathcal{H}$ is a smooth $\mathcal{H}$-valued 0-form on S, so using the product defined in Example 4.4 (3), we get a smooth $\mathbb{C}$-valued 1-form $\langle j|\mathrm{d}j\rangle\in\Omega^1(S,\mathbb{C})$ on S. The exterior derivative is an antiderivation and $\langle j|j\rangle\colon S\to\mathbb{C}$ is the constant function $z\mapsto 1$ by definition of S, so $\langle \mathrm{d}j|j\rangle+\langle j|\mathrm{d}j\rangle=0$. It follows that

$$\langle j|\mathrm{d}j\rangle_z(v)=-\langle \mathrm{d}j|j\rangle_z(v)=-\langle \mathrm{d}j_z(v)|z\rangle=-\overline{\langle z|\mathrm{d}j_z(v)\rangle}=-\overline{\langle j|\mathrm{d}j\rangle_z(v)}$$

for all $z\in S$ and $v\in T_zS$ by the conjugate symmetry of $\langle\cdot|\cdot\rangle$, so $\langle j|\mathrm{d}j\rangle$ actually takes values in $i\mathbb{R}$. We know from Lemma 4.24 that the exterior derivative $\mathrm{d}(j_0)_1\colon T_1U(1)\to\mathbb{C}$ at $1\in U(1)$ of the inclusion $j_0\colon U(1)\hookrightarrow\mathbb{C}$ is an isomorphism onto $i\mathbb{R}$ mapping the angle coordinate vector $\mathrm{d}/\mathrm{d}\theta\,|_1$ to i, so under the fixed identification $\mathrm{d}(j_0)_1\circ\epsilon\colon \mathrm{Lie}(U(1))\to i\mathbb{R}$, we can consider $\langle j|\mathrm{d}j\rangle$ to be a smooth $\mathrm{Lie}(U(1))$-valued 1-form on S.

Proposition 5.5. *$\langle j|\mathrm{d}j\rangle$ is a connection form on γ_n.*

Proof. We check the two defining properties separately. Let $\vartheta\colon\mathcal{H}\times U(1)\to\mathcal{H}$, $(z,\lambda)\mapsto z\lambda$ be the scalar multiplication map of $U(1)$ on $\mathcal{H}$, so that $j\circ\theta_\lambda=\vartheta_\lambda\circ j$ and $j\circ\theta^{(z)}=\vartheta^{(z)}$ for all $\lambda\in U(1)$ and $z\in S$.

(i) Let $X\in\mathrm{Lie}(U(1))$, then $X=s\,\mathrm{d}/\mathrm{d}\theta$ for some $s\in\mathbb{R}$. The curve $\gamma\colon\mathbb{R}\to U(1)$, $t\mapsto e^{ist}$ in $U(1)$ has initial velocity $\gamma'(0)=s\,\mathrm{d}/\mathrm{d}\theta\,|_1=X_1\in T_1U(1)$, so for any $z\in S$,

$$\omega_z(\underline{X}_z)=\omega_z(\mathrm{d}(\theta^{(z)})_1(X_1))=\langle z|\mathrm{d}(j\circ\theta^{(z)})_1(\gamma'(0))\rangle=\langle z|(\vartheta^{(z)}\circ\gamma)'(0)\rangle=\langle z|isz\rangle=is$$

by Equation (2.3). Under the identification $\mathrm{Lie}(U(1))\cong i\mathbb{R}$, is corresponds to $s\,\mathrm{d}/\mathrm{d}\theta=X$.

(ii) Let $\lambda\in U(1)$, then

$$\theta_\lambda^*\langle j|\mathrm{d}j\rangle=\langle\theta_\lambda^*j|\mathrm{d}(\theta_\lambda^*j)\rangle=\langle j\circ\theta_\lambda|\mathrm{d}(j\circ\theta_\lambda)\rangle=\langle\vartheta_\lambda\circ j|\mathrm{d}(\vartheta_\lambda\circ j)\rangle$$

since pullbacks commute with products and exterior derivatives (Lemma 4.10). By Lemma 4.8, we have $\mathrm{d}(\vartheta_\lambda\circ j)_z=\vartheta_\lambda\circ\mathrm{d}j_z$ for $z\in S$, so for any $v\in T_zS$,

$$(\theta_\lambda^*\langle j|\mathrm{d}j\rangle)_z(v)=\langle(\vartheta_\lambda\circ j)(z)|(\vartheta_\lambda\circ\mathrm{d}j_z)(v)\rangle=\langle z\lambda|\mathrm{d}j_z(v)\lambda\rangle=\langle z|\mathrm{d}j_z(v)\rangle|\lambda|^2=\langle j|\mathrm{d}j\rangle_z(v).\quad\square$$

One reason this connection form is a natural one, is that the corresponding horizontal tangent vectors to S at z are precisely those which are *orthogonal* to the vertical tangent subspace $V_z\subseteq T_zS$. To see why, let $z\in S$ and note that by the proof of (i), the fundamental vector $\underline{X}_z\in T_zS$ associated to the angle coordinate vector field $X:=\mathrm{d}/\mathrm{d}\theta\in\mathrm{Lie}(U(1))$ corresponds to iz under the identification $\mathrm{d}j_z\colon T_zS\to\mathcal{H}$ of T_zS with a subspace of $\mathcal{H}$. That means $v\in T_zS$ is horizontal if and only if

$$\begin{aligned}\omega_z(v)=0 &\iff \langle z|\mathrm{d}j_z(v)\rangle=0\iff\langle\mathrm{d}j_z(\underline{X}_z)|\mathrm{d}j_z(v)\rangle=0\\ &\iff \mathrm{d}j_z(v)\in\{\mathrm{d}j_z(\underline{X}_z)\}^\perp\iff\mathrm{d}j_z(v)\in\mathrm{d}j_z(V_z)^\perp,\end{aligned}$$

since V_z is generated by $\underline{X}_z$. Let us now return to the general setting.

5.3 Local connection forms

In this section,[1] we set up a one-to-one correspondence between the connections on the principal G-bundle π and the families of so-called *local connection forms* on open subsets of the base space M. The proofs are based on [12, Section 10.1.3]. First, choose a G-atlas $\{(U_\alpha,\phi_\alpha)\}_{\alpha\in A}$ for π, then we can construct for each $\alpha\in A$ a smooth local section

$$\begin{aligned}\sigma_\alpha\colon U_\alpha&\longrightarrow\pi^{-1}U_\alpha\\ x&\longmapsto\phi_\alpha^{-1}(x,e)\end{aligned}$$

[1] We leave it up to the reader to decide if it is a smooth one.

of π over U_α as in Lemma 3.9. Recall that the transition functions $\rho_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$ of π relative to the chosen G-atlas are defined by

$$\phi_\beta^{-1}(x, g) = \phi_\alpha^{-1}(x, \rho_{\alpha\beta}(x)g) \tag{5.1}$$

for all $x \in U_\alpha \cap U_\beta$ and $g \in G$. Plugging in $g = e$ gives $\sigma_\beta(x) = \sigma_\alpha(x)\rho_{\alpha\beta}(x)$, i.e. $\sigma_\beta = \sigma_\alpha \cdot \rho_{\alpha\beta}$ in the product notation (2.2). With this, we find that the differentials of the local sections are related on intersections as follows.

Lemma 5.6. *Let $\alpha, \beta \in A$, $x \in U_\alpha \cap U_\beta$ and $v \in T_xM$, then*

$$\mathrm{d}(\sigma_\beta)_x(v) = \mathrm{d}(\sigma_\alpha \cdot \rho_{\alpha\beta}(x))_x(v) + \underline{\mathrm{d}\log(\rho_{\alpha\beta})_x(v)}_{\sigma_\beta(x)}.$$

Proof. Applying Lemma 2.19 to $\sigma_\beta = \sigma_\alpha \cdot \rho_{\alpha\beta}$ yields

$$\mathrm{d}(\sigma_\beta)_x = \mathrm{d}(\sigma_\alpha \cdot \rho_{\alpha\beta}(x))_x + \mathrm{d}(\sigma_\alpha(x) \cdot \rho_{\alpha\beta})_x.$$

Now, by Equation (2.3) and the definition of the logarithmic derivative,

$$\begin{aligned}
\underline{\mathrm{d}\log(\rho_{\alpha\beta})_x(v))}_{\sigma_\beta(x)} &= \mathrm{d}(\theta^{(\sigma_\beta(x))})_e(\mathrm{d}(\rho_{\alpha\beta}(x)^{-1}\rho_{\alpha\beta})_x(v)) \\
&= \mathrm{d}(\theta^{(\sigma_\beta(x))} \circ (\rho_{\alpha\beta}(x)^{-1}\rho_{\alpha\beta}))_x(v) \\
&= \mathrm{d}(\theta^{(\sigma_\beta(x)\rho_{\alpha\beta}(x)^{-1})} \circ \rho_{\alpha\beta})_x(v) \\
&= \mathrm{d}(\theta^{(\sigma_\alpha(x))} \circ \rho_{\alpha\beta})_x(v) = \mathrm{d}(\sigma_\alpha(x) \cdot \rho_{\alpha\beta})_x(v)
\end{aligned}$$

and we are done. □

Now, let $\omega \in \Omega^1(P, \mathrm{Lie}(G))$ be a connection form on π. For all $\alpha \in A$, pulling ω back along σ_α gives a smooth $\mathrm{Lie}(G)$-valued one-form $\mathcal{A}_\alpha := \sigma_\alpha^*\omega \in \Omega^1(U_\alpha, \mathrm{Lie}(G))$ on U_α. The $\mathcal{A}_\alpha$ are known as the *local connection forms* of ω relative to the chosen bundle atlas and on intersections, they satisfy the following *compatibility condition.*

Proposition 5.7. *Let $\alpha, \beta \in A$, then on $U_\alpha \cap U_\beta$,*

$$\boxed{\mathcal{A}_\beta = \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha + \mathrm{d}\log(\rho_{\alpha\beta}).} \tag{5.2}$$

Proof. Let $x \in U_\alpha \cap U_\beta$ and $v \in T_xM$. By Lemma 5.6,

$$\begin{aligned}
(\mathcal{A}_\beta)_x(v) = (\sigma_\beta^*\omega)_x(v) &= \omega_{\sigma_\beta(x)}(\mathrm{d}(\sigma_\beta)_x(v)) \\
&= \omega_{\sigma_\beta(x)}(\mathrm{d}(\sigma_\alpha \cdot \rho_{\alpha\beta}(x))_x(v)) + \omega_{\sigma_\beta(x)}\big(\underline{\mathrm{d}\log(\rho_{\alpha\beta})_x(v)}_{\sigma_\beta(x)}\big).
\end{aligned}$$

We examine the two terms separately. For the first, note that $\sigma_\beta(x) = \sigma_\alpha(x)\rho_{\alpha\beta}(x) = \theta_{\rho_{\alpha\beta}(x)}(\sigma_\alpha(x))$, so

$$\begin{aligned}
\omega_{\sigma_\beta(x)}(\mathrm{d}(\sigma_\alpha \cdot \rho_{\alpha\beta}(x))_x(v)) &= \omega_{\sigma_\alpha(x)\rho_{\alpha\beta}(x)}(\mathrm{d}(\theta_{\rho_{\alpha\beta}(x)} \circ \sigma_\alpha)_x(v)) \\
&= (\theta_{\rho_{\alpha\beta}(x)}^*\omega)_{\sigma_\alpha(x)}(\mathrm{d}(\sigma_\alpha)_x(v)) \\
&= \mathrm{Ad}_{\rho_{\alpha\beta}(x)^{-1}}(\omega_{\sigma_\alpha(x)}(\mathrm{d}(\sigma_\alpha)_x(v))) \\
&= (\mathrm{Ad}_{\rho_{\alpha\beta}(x)^{-1}} \circ (\sigma_\alpha^*\omega)_x)(v) \\
&= (\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha)_x(v)
\end{aligned}$$

by property (ii) in Definition 5.1. For the second term, property (i) implies

$$\omega_{\sigma_\beta(x)}\big(\underline{\mathrm{d}\log(\rho_{\alpha\beta})_x(v)}_{\sigma_\beta(x)}\big) = \mathrm{d}\log(\rho_{\alpha\beta})_x(v),$$

so we are done. □

Suppose now that instead of a connection one-form on π, we are given for all $\alpha \in A$ a $\mathrm{Lie}(G)$-valued one-form $\mathcal{A}_\alpha \in \Omega^1(U_\alpha, \mathrm{Lie}(G))$ on U_α. Then Proposition 5.7 gives a *necessary* condition on the $\mathcal{A}_\alpha$ for the existence of a connection form ω on π satisfying $\mathcal{A}_\alpha = \sigma_\alpha^*\omega$ for all $\alpha \in A$. This compatibility condition turns out to be *sufficient*, too. We prove this in two steps: we first show that each $\mathcal{A}_\alpha$ can be "lifted" to a connection on $\pi^{-1}U_\alpha$, and then that these lifts agree on intersections if the compatibility condition is satisfied.

Lemma 5.8. *Let $U \subseteq M$ be an open subset, $\sigma\colon U \to \pi^{-1}U$ a smooth local section of π and $\mathcal{A} \in \Omega^1(U, \mathrm{Lie}(G))$ a $\mathrm{Lie}(G)$-valued one-form on U, then there exists a connection form ω on $\pi^{-1}U$ such that $\mathcal{A} = \sigma^*\omega$.*

Proof. Let $\phi\colon \pi^{-1}U \to U\times G$ be the local trivialization induced by σ, i.e. $\phi^{-1}(x,h) = \sigma(x)h$ for all $x \in U$ and $h \in G$. Set $g := \pi_2 \circ \phi\colon \pi^{-1}U \to G$ with $\pi_2\colon U \times G \to G$ the projection, so that $\phi(p) = (\pi(p), g(p))$ for all $p \in \pi^{-1}U$. The map g is smooth and G-equivariant when considering G as a right G-manifold under multiplication, since both π_2 and ϕ are smooth and G-equivariant. Lemma 4.22 now implies that $\chi := \mathrm{d}\log g$ is a connection form on $\pi^{-1}U$. By Lemma 4.20, it is precisely the pullback along the principal G-bundle morphism ϕ of the canonical connection form $\mathrm{d}\log(\pi_2)$ on $U \times G$.

Now, note that $\zeta := \mathrm{Ad}_{g^{-1}}(\pi^*\mathcal{A})$ is a smooth $\mathrm{Lie}(G)$-valued one-form on $\pi^{-1}U$, given by

$$\zeta_p = \mathrm{Ad}_{g(p)^{-1}} \circ (\pi^*\mathcal{A})_p = \mathrm{Ad}_{g(p)^{-1}} \circ \mathcal{A}_{\pi(p)} \circ \mathrm{d}\pi_p$$

for any $p \in \pi^{-1}U$. It clearly satisfies $\zeta_p(v) = 0$ for all $v \in V_p = \ker \mathrm{d}\pi_p$, and for any $h \in G$, we have

$$\begin{aligned}(\theta_h^*\zeta)_p &= \zeta_{ph} \circ \mathrm{d}(\theta_h)_p = \mathrm{Ad}_{g(ph)^{-1}} \circ \mathcal{A}_{\pi(ph)} \circ \mathrm{d}(\pi \circ \theta_h)_p \\ &= \mathrm{Ad}_{h^{-1}g(p)^{-1}} \circ \mathcal{A}_{\pi(p)} \circ \mathrm{d}\pi_p = \mathrm{Ad}_{h^{-1}} \circ (\mathrm{Ad}_{g(p)^{-1}} \circ \mathcal{A}_{\pi(p)} \circ \mathrm{d}\pi_p) = \mathrm{Ad}_{h^{-1}} \circ \zeta_p,\end{aligned}$$

where we used Lemma 2.17. From this, it follows that $\omega := \zeta + \chi$ is a connection form on $\pi^{-1}U$. We now show $\sigma^*\omega = \mathcal{A}$. Note that $g \circ \sigma$ is the constant map $x \mapsto e$ while $\pi \circ \sigma$ is the identity on U, so

$$\begin{aligned}(\sigma^*\zeta)_x &= \zeta_{\sigma(x)} \circ \mathrm{d}\sigma_x = \mathrm{Ad}_{g(\sigma(x))^{-1}} \circ \mathcal{A}_{\pi(\sigma(x))} \circ \mathrm{d}(\pi \circ \sigma)_x = \mathrm{Ad}_e \circ \mathcal{A}_x = \mathcal{A}_x,\\ (\sigma^*\chi)_x &= \chi_{\sigma(x)} \circ \mathrm{d}\sigma_x = \epsilon^{-1} \circ \mathrm{d}((g(\sigma(x))^{-1} \cdot g) \circ \sigma)_x = \epsilon^{-1} \circ \mathrm{d}(g \circ \sigma)_x = 0\end{aligned}$$

and thus $(\sigma^*\omega)_x = \mathcal{A}_x$ for all $x \in U$. □

Proposition 5.9. *Suppose the $\mathcal{A}_\alpha$ satisfy the compatibility condition* (5.2)*, then there exists a connection form $\omega \in \Omega^1(P, \mathrm{Lie}(G))$ on π such that $\mathcal{A}_\alpha = \sigma_\alpha^*\omega$ for all $\alpha \in A$.*

Proof. As in the proof of Lemma 5.8, we write $g_\alpha := \pi_2 \circ \phi_\alpha\colon \pi^{-1}U_\alpha \to G$ for all $\alpha \in A$ with π_2 the projection $U_\alpha \times G \to G$, so that $\phi_\alpha(p) = (\pi(p), g_\alpha(p))$ for $p \in \pi^{-1}U_\alpha$. With this, the transition function equation (5.1) becomes $g_\alpha(p) = \rho_{\alpha\beta}(\pi(p))g_\beta(p)$ for $p \in \pi^{-1}(U_\alpha \cap U_\beta)$, i.e. $g_\alpha = (\rho_{\alpha\beta} \circ \pi) \cdot g_\beta$.

For all $\alpha \in A$, we have a connection form $\omega_\alpha = \zeta_\alpha + \chi_\alpha \in \Omega^1(\pi^{-1}U_\alpha, \mathrm{Lie}(G))$ on $\pi^{-1}U_\alpha$ defined as

$$\begin{aligned}\zeta_\alpha &= \mathrm{Ad}_{g_\alpha^{-1}}(\pi^*\mathcal{A}_\alpha),\\ \chi_\alpha &= \mathrm{d}\log g_\alpha\end{aligned}$$

and satisfying $\sigma_\alpha^*\omega_\alpha = \mathcal{A}_\alpha$ by Lemma 5.8. It suffices to show that the ω_α agree on overlaps, so let $\alpha, \beta \in A$, $p \in \pi^{-1}(U_\alpha \cap U_\beta)$ and write $x := \pi(p)$, then by the compatibility condition (5.2),

$$\begin{aligned}(\zeta_\beta)_p &= \mathrm{Ad}_{g_\beta(p)^{-1}} \circ (\pi^*\mathcal{A}_\beta)_p\\ &= \mathrm{Ad}_{g_\beta(p)^{-1}} \circ (\mathcal{A}_\beta)_x \circ \mathrm{d}\pi_p\\ &= \mathrm{Ad}_{g_\beta(p)^{-1}} \circ (\mathrm{Ad}_{\rho_{\alpha\beta}(x)^{-1}} \circ (\mathcal{A}_\alpha)_x + \mathrm{d}\log(\rho_{\alpha\beta})_x) \circ \mathrm{d}\pi_p.\end{aligned}$$

For the first term, note that

$$\begin{aligned}\mathrm{Ad}_{g_\beta(p)^{-1}} \circ \mathrm{Ad}_{\rho_{\alpha\beta}(x)^{-1}} \circ (\mathcal{A}_\alpha)_x \circ \mathrm{d}\pi_p &= \mathrm{Ad}_{(\rho_{\alpha\beta}(x)g_\beta(p))^{-1}} \circ (\mathcal{A}_\alpha)_x \circ \mathrm{d}\pi_p \\ &= \mathrm{Ad}_{g_\alpha(p)^{-1}} \circ (\pi^*\mathcal{A}_\alpha)_p = (\zeta_\alpha)_p\end{aligned}$$

since Ad is a group homomorphism (Lemma 2.17). For the second, Lemma 4.21 gives

$$\begin{aligned}\mathrm{Ad}_{g_\beta(p)^{-1}} \circ \mathrm{d}\log(\rho_{\alpha\beta})_x \circ \mathrm{d}\pi_p &= \mathrm{Ad}_{g_\beta(p)^{-1}} \circ \mathrm{d}\log(\rho_{\alpha\beta} \circ \pi)_p \\ &= \mathrm{d}\log((\rho_{\alpha\beta} \circ \pi) \cdot g_\beta)_p - \mathrm{d}\log(g_\beta)_p \\ &= \mathrm{d}\log(g_\alpha)_p - \mathrm{d}\log(g_\beta)_p = (\chi_\alpha)_p - (\chi_\beta)_p.\end{aligned}$$

We conclude $\zeta_\beta - \zeta_\alpha = \chi_\alpha - \chi_\beta$ and thus $\omega_\alpha = \omega_\beta$ on $\pi^{-1}(U_\alpha \cap U_\beta)$. ☐

A very natural question to ask at this point is if the connection ω on π as constructed in the proof of Proposition 5.9 is the only one which has the given $\mathcal{A}_\alpha$ as its local connection forms. Our next proposition answers this question.

Proposition 5.10. *Let $\omega, \eta \in \Omega^1(P, \mathrm{Lie}(G))$ be connection forms on π such that $\sigma_\alpha^*\omega = \sigma_\alpha^*\eta$ for all $\alpha \in A$, then $\omega = \eta$.*

Proof. Let $\alpha \in A$, $x \in U_\alpha$ and set $p \coloneqq \sigma_\alpha(x)$, we show $\omega_p = \eta_p$. Define $\sigma \colon U_\alpha \to U_\alpha \times G$, $y \mapsto (y, e)$ and let $\pi_1 \colon U_\alpha \times G \to U_\alpha$ be the projection onto the first coordinate, then $\sigma = \phi_\alpha \circ \sigma_\alpha$ by definition of the local section σ_α and $\pi_1 \circ \phi_\alpha = \pi$ by definition of a local trivialization. Theorem A.2 now tells us

$$T_{(x,e)}(U_\alpha \times G) = \mathrm{d}\sigma_x(T_xU_\alpha) \oplus \ker \mathrm{d}(\pi_1)_{(x,e)},$$

so the tangent space T_pP to P at $p = \sigma_\alpha(x) = \phi_\alpha^{-1}(x, e)$ is the direct sum of

$$\mathrm{d}(\phi_\alpha^{-1})_{(x,e)}(\mathrm{d}\sigma_x(T_xU_\alpha)) = \mathrm{d}(\phi_\alpha^{-1} \circ \sigma)_x(T_xU_\alpha) = \mathrm{d}(\sigma_\alpha)_x(T_xU_\alpha)$$

and

$$\mathrm{d}(\phi_\alpha^{-1})_{(x,e)}(\ker \mathrm{d}(\pi_1)_{(x,e)}) = \ker \mathrm{d}(\pi_1 \circ \phi_\alpha)_p = \ker \mathrm{d}\pi_p = V_p,$$

the vertical tangent subspace at p. By linearity, it therefore suffices to prove $\omega_p(v) = \eta_p(v)$ for v in these two subspaces separately. If $v \in V_p$, this holds by property (i) in Definition 5.1 since $v = \underline{X}_p$ for some $X \in \mathrm{Lie}(G)$ (see Lemma 3.20). If $v \in \mathrm{d}(\sigma_\alpha)_x(T_xU_\alpha)$, then $v = \mathrm{d}(\sigma_\alpha)_x(w)$ for some $w \in T_xU_\alpha$ and thus

$$\omega_p(v) = (\sigma_\alpha^*\omega)_x(w) = (\sigma_\alpha^*\eta)_x(w) = \eta_p(v).$$

We conclude $\omega_p = \eta_p$, so also

$$\omega_{pg} = \mathrm{Ad}_{g^{-1}} \circ \omega_p \circ \mathrm{d}(\theta_{g^{-1}})_{pg} = \mathrm{Ad}_{g^{-1}} \circ \eta_p \circ \mathrm{d}(\theta_{g^{-1}})_{pg} = \eta_{pg}$$

for all $g \in G$ by property (ii) in Definition 5.1. Since α and x were arbitrary and G acts transitively on the fibers, we are done. ☐

Combining all the previous results yields the following classification theorem.

Theorem 5.11. *The assignment $\omega \mapsto \{\sigma_\alpha^*\omega\}_{\alpha \in A}$ defines a one-to-one correspondence between the connections on π and the families $\{\mathcal{A}_\alpha\}_{\alpha \in A}$ of smooth* Lie(G)*-valued forms $\mathcal{A}_\alpha \in \Omega^1(U_\alpha, \mathrm{Lie}(G))$ satisfying the compatibility condition* (5.2).

Proof. The given map is well-defined by Proposition 5.7, surjective by Proposition 5.9 and injective by Proposition 5.10. ☐

5.4 Curvature

Associated to any connection on π is a $\mathrm{Lie}(G)$-valued 2-form on P.

Definition 5.12. Let $\omega \in \Omega^1(P, \mathrm{Lie}(G))$ be a connection form on π. The *curvature* of ω is defined as the $\mathrm{Lie}(G)$-valued 2-form

$$\Omega \coloneqq \mathrm{d}\omega + \frac{1}{2}[\omega, \omega] \in \Omega^2(P, \mathrm{Lie}(G)).$$

Let ω be a connection form on π and Ω its curvature. We are mainly interested in the corresponding *local curvature forms*, i.e. pullbacks of Ω along local sections of π, so let $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ be a G-atlas for π with associated local sections σ_α and transition functions $\rho_{\alpha\beta}$ as in the previous section. Set $\mathcal{A}_\alpha \coloneqq \sigma_\alpha^* \omega$ and $\mathcal{F}_\alpha \coloneqq \sigma_\alpha^* \Omega$, then

$$\mathcal{F}_\alpha = \sigma_\alpha^* \, \mathrm{d}\omega + \frac{1}{2}\sigma_\alpha^*[\omega, \omega] = \mathrm{d}\sigma_\alpha^* \omega + \frac{1}{2}[\sigma_\alpha^* \omega, \sigma_\alpha^* \omega] = \mathrm{d}\mathcal{A}_\alpha + \frac{1}{2}[\mathcal{A}_\alpha, \mathcal{A}_\alpha] \tag{5.3}$$

because pullbacks commute with exterior derivatives and products. For these local curvature forms, the compatibility condition (5.2) on the $\mathcal{A}_\alpha$ now implies the following.

Proposition 5.13. *Let $\alpha, \beta \in A$, then on $U_\alpha \cap U_\beta$,*

$$\boxed{\mathcal{F}_\beta = \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{F}_\alpha.} \tag{5.4}$$

Proof. Plugging (5.2) into (5.3) gives

$$\begin{aligned}
\mathcal{F}_\beta &= \mathrm{d}\mathcal{A}_\beta + \frac{1}{2}[\mathcal{A}_\beta, \mathcal{A}_\beta] \\
&= \mathrm{d}(\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha + \mathrm{d}\log(\rho_{\alpha\beta})) + \frac{1}{2}[\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha + \mathrm{d}\log(\rho_{\alpha\beta}), \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha + \mathrm{d}\log(\rho_{\alpha\beta})] \\
&= \mathrm{d}(\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha) + \mathrm{d}(\mathrm{d}\log(\rho_{\alpha\beta})) + \frac{1}{2}[\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha, \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha] \\
&\qquad + [\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha, \mathrm{d}\log(\rho_{\alpha\beta})] + \frac{1}{2}[\mathrm{d}\log(\rho_{\alpha\beta}), \mathrm{d}\log(\rho_{\alpha\beta})],
\end{aligned}$$

where we used that the Lie bracket is commutative for 1-forms (cf. Lemma 4.11) in the last equation. By Lemmas 4.13, 4.21 and 4.23,

$$\begin{aligned}
\mathrm{d}(\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha) &= \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathrm{d}\mathcal{A}_\alpha - [\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha, \mathrm{d}\log(\rho_{\alpha\beta})], \\
\mathrm{d}(\mathrm{d}\log(\rho_{\alpha\beta})) &= -\frac{1}{2}[\mathrm{d}\log(\rho_{\alpha\beta}), \mathrm{d}\log(\rho_{\alpha\beta})], \\
[\mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha, \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}} \mathcal{A}_\alpha] &= \mathrm{Ad}_{\rho_{\alpha\beta}^{-1}}[\mathcal{A}_\alpha, \mathcal{A}_\alpha],
\end{aligned}$$

from which the claim follows. □

5.5 Parallel transport

Now that we have seen what connections and curvatures on the principal bundle π are and how they can be constructed from Lie algebra valued forms on open subsets of M, it is time to talk about their *raison d'être*: path lifting and parallel transport. Let ω be a connection form on π, then we can define what it means for a path in P to be "horizontal".

Definition 5.14. Let $\gamma\colon [0,1] \to M$ and $\tilde{\gamma}\colon [0,1] \to P$ be smooth paths in M and P, respectively. We say $\tilde{\gamma}$ is a *lift* of γ if $\pi \circ \tilde{\gamma} = \gamma$, and a *horizontal lift* if in addition $\omega_{\tilde{\gamma}(t)}(\tilde{\gamma}'(t)) = 0$ for all $t \in [0,1]$.

In words, a horizontal lift of a smooth path γ in M is a smooth path in P which projects onto γ and whose velocity at any point $\tilde{\gamma}(t)$ lies in the horizontal tangent subspace $\ker \omega_{\tilde{\gamma}(t)}$ defined by the connection ω. The main theorem on parallel transport in principal bundles states that γ has a *unique* horizontal lift with a given starting point $p \in \pi^{-1}\{\gamma(0)\}$.

Theorem 5.15. *Let $\gamma\colon [0,1] \to M$ be a smooth path and choose $p \in \pi^{-1}\{\gamma(0)\}$, then there exists a unique horizontal lift $\tilde{\gamma}\colon [0,1] \to P$ of γ with $\tilde{\gamma}(0) = p$.*

The reader is referred to [10, Proposition 17.2] for a proof, which essentially boils down to the existence and uniqueness of solutions to ordinary first order differential equations. We give here an explicit construction of the horizontal lift in case G is abelian and $\gamma([0,1])$ is contained within an open subset U of M whose inverse image under π can be trivialized.

Proposition 5.16. *Suppose G is abelian. Let (U, ϕ) be a local trivialization for π, $\sigma\colon U \to \pi^{-1}U$ the associated local section of π under the correspondence of Lemma 3.9 and set $\mathcal{A} := \sigma^*\omega \in \Omega^1(U, \mathrm{Lie}(G))$. Let $\gamma\colon [0,1] \to U$ be a smooth path in U and define*

$$\begin{aligned} g\colon [0,1] &\longrightarrow G \\ t &\longmapsto \exp\Big(-\int_0^t \gamma^*\mathcal{A}\Big), \end{aligned}$$

then $\tilde{\gamma} := (\sigma \circ \gamma) \cdot g\colon [0,1] \to \pi^{-1}U$ is the unique horizontal lift of γ starting at $\sigma(\gamma(0))$.

Proof. The right action of G on P is fiberwise, so $\pi(\tilde{\gamma}(t)) = \pi((\sigma \circ \gamma)(t))$ for all $t \in [0,1]$. It follows that $\pi(\tilde{\gamma}(t)) = \gamma(t)$ since $\pi \circ \sigma = \mathrm{id}_U$, which means $\tilde{\gamma}$ is a lift of γ.

To prove horizontality, let $t_0 \in [0,1]$ and note that

$$\tilde{\gamma}'(t_0) = \mathrm{d}((\sigma \circ \gamma) \cdot g)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) = \mathrm{d}((\sigma \circ \gamma) \cdot g(t_0))_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) + \mathrm{d}((\sigma \circ \gamma)(t_0) \cdot g)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big)$$

by Lemma 2.19. We can plug both terms into $\omega_{\tilde{\gamma}(t_0)}$ and show that they sum up to zero. For the first term, observe that $(\sigma \circ \gamma) \cdot g(t_0) = \theta_{g(t_0)} \circ (\sigma \circ \gamma)$ and thus

$$\begin{aligned} \omega_{\tilde{\gamma}(t_0)} \circ \mathrm{d}((\sigma \circ \gamma) \cdot g(t_0))_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) &= \omega_{(\sigma\circ\gamma)(t_0)g(t_0)} \circ \mathrm{d}(\theta_{g(t_0)})_{(\sigma\circ\gamma)(t_0)} \circ \mathrm{d}(\sigma \circ \gamma)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) \\ &= (\theta_{g(t_0)}^*\omega)_{(\sigma\circ\gamma)(t_0)} \circ \mathrm{d}(\sigma \circ \gamma)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) \\ &= \omega_{\sigma(\gamma(t_0))} \circ \mathrm{d}\sigma_{\gamma(t_0)}(\gamma'(t_0)) \\ &= (\sigma^*\omega)_{\gamma(t_0)}(\gamma'(t_0)) = \mathcal{A}_{\gamma(t_0)}(\gamma'(t_0)) \end{aligned}$$

by property (ii) in Definition 5.1 (the adjoint representation is trivial since G is assumed to be abelian).

For the second term, $(\sigma \circ \gamma)(t_0) \cdot g = \tilde{\gamma}(t_0) \cdot g(t_0)^{-1}g = \theta^{(\tilde{\gamma}(t_0))} \circ g(t_0)^{-1}g$ yields

$$\begin{aligned} \omega_{\tilde{\gamma}(t_0)} \circ \mathrm{d}((\sigma \circ \gamma)(t_0) \cdot g)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) &= \omega_{\tilde{\gamma}(t_0)} \circ \mathrm{d}(\theta^{(\tilde{\gamma}(t_0))})_e \circ \mathrm{d}(g(t_0)^{-1}g)_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big) \\ &= \omega_{\tilde{\gamma}(t_0)}\Big(\underline{\mathrm{d}\log g_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big)}_{\tilde{\gamma}(t_0)}\Big) = \mathrm{d}\log g_{t_0}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t_0}\Big). \end{aligned}$$

We used Equation (2.3) in the second equality and property (i) from Definition 5.1 in the third. Define

$$\begin{aligned} I\colon [0,1] &\longrightarrow \mathrm{Lie}(G) \\ t &\longmapsto -\int_0^t \gamma^*\mathcal{A}, \end{aligned}$$

so that $g = \exp \circ I$ and thus $\mathrm{d}\log g_{t_0} = \mathrm{d}\log \exp_{I(t_0)} \circ \mathrm{d}I_{t_0}$ by Lemma 4.21. From Lemma 4.19, we know $\mathrm{d}\log \exp_{I(t_0)}$ is the inverse of the vector space isomorphism $D_{I(t_0)} \colon \mathrm{Lie}(G) \to T_{I(t_0)} \mathrm{Lie}(G)$. By Lemma 4.7, that means $\mathrm{d}\log g_{t_0} = D^{-1}_{I(t_0)} \circ \mathrm{d}I_{t_0}$ is just the exterior derivative of I at t_0 when we consider I as a $\mathrm{Lie}(G)$-valued 0-form on $[0,1]$, so

$$\mathrm{d}\log g_{t_0}\left(\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t_0}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t_0} I = -\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t_0} \int_0^t (\gamma^*\mathcal{A})\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\,\mathrm{d}t = -(\gamma^*\mathcal{A})_{t_0}\left(\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t_0}\right) = -\mathcal{A}_{\gamma(t_0)}(\gamma'(t_0)).$$

Putting everything together, we find $\omega_{\tilde{\gamma}(t_0)}(\tilde{\gamma}'(t_0)) = 0$, as required. □

As a special case, suppose γ is a *loop* in U, so that $\tilde{\gamma}(1) = \tilde{\gamma}(0)g(1)$. Let Ω be the curvature of ω and $\mathcal{F} := \sigma^*\Omega$ the corresponding local curvature form on U, then $\Omega = \mathrm{d}\omega$ and $\mathcal{F} = \mathrm{d}\mathcal{A}$ since G is abelian. If $\Sigma \subseteq M$ is an appropriately oriented and compact embedded 2-submanifold of U with the image of the curve γ as its boundary, Stokes' theorem implies

$$\begin{aligned} g(1) &= \exp\left(-\int_0^1 \gamma^*\mathcal{A}\right) = \exp\left(-\int_{\partial\Sigma} \mathcal{A}\right) = \exp\left(-\int_{\partial\Sigma} \mathrm{d}\mathcal{A}\right) \\ &= \exp\left(-\int_{\Sigma} \mathcal{F}\right) = \exp\left(-\int_{\Sigma} \sigma^*\Omega\right) = \exp\left(-\int_{\sigma(\Sigma)} \Omega\right). \end{aligned}$$

In words, the starting point $\tilde{\gamma}(0)$ and the endpoint $\tilde{\gamma}(1)$ of the horizontal lift $\tilde{\gamma}$ of the loop γ in U lie in the same fiber of the bundle map π, but they differ by a factor $g(1)$ which is measured by the curvature Ω of the connection. In the case $G = U(1)$, we can interpret this as a sort of "phase shift", a concept we will return to in Chapter 7.

Chapter 6

The TKNN formula

With all this mathematical machinery in place, we now turn to the derivation of the TKNN formula for σ_{xy}. The starting point is the *Kubo formula* for the linear response of an observable to a weak perturbation, which can then be applied to the current density $\hat{\mathbf{j}}$ to obtain a very general expression for the components of the conductivity matrix $\boldsymbol{\sigma}$. This expression reduces to a much simpler and more manageable form in the *independent electron approximation*, i.e. when electron-electron interactions in the system are ignored. Finally, the *TKNN formula* can be derived by assuming the potential energy function has the periodicity of a lattice – which is approximately the case for electrons in a crystalline solid – and using Bloch's theorem to characterize the energy eigenstates. The derivations are based largely on [13], [14] and [15] and take place in the framework of *second quantization*. For an overview of the important definitions and results, we refer to Appendix B.1.

6.1 Linear response theory

Consider a system of fermionic particles described by a time-independent Hamiltonian $\hat{H}_0$ acting on the antisymmetric Fock space $\mathcal{F}(\mathcal{H})$ of a one-particle Hilbert space $\mathcal{H}$, i.e. the completion of the direct sum of the N-particle spaces $\mathcal{H}_a^{(N)}$ for all $N \geq 0$. Assuming the grand canonical ensemble, the equilibrium density matrix of the system is

$$\hat{\rho}_0 = \frac{1}{Z_0} e^{-\beta(\hat{H}_0 - \mu\hat{N})} \tag{6.1}$$

with $\beta = 1/k_B T$ the inverse temperature, μ the chemical potential, $\hat{N} = \mathrm{d}\Gamma(I)$ the particle number operator on $\mathcal{F}(\mathcal{H})$ and $Z_0 = \mathrm{Tr}\, e^{-\beta(\hat{H}_0 - \mu\hat{N})}$ the partition function. In the presence of a (weak) time-dependent perturbation $\hat{W}(t)$, the total Hamiltonian becomes

$$\hat{H}(t) = \hat{H}_0 + \hat{W}(t)$$

and the time evolution of the new density matrix $\hat{\rho}(t)$ is dictated by the von Neumann equation

$$i\hbar \frac{\partial \hat{\rho}(t)}{\partial t} = [\hat{H}(t), \hat{\rho}(t)]. \tag{6.2}$$

Since the Hamiltonian is split into a time-independent part and a perturbation, it is useful to switch to the *interaction picture*, in which part of the time evolution of the density matrix is transferred to the observables. We do this by setting $\hat{A}_I(t) := e^{i\hat{H}_0 t/\hbar} \hat{A}(t) e^{-i\hat{H}_0 t/\hbar}$ for any (possibly time-dependent) operator $\hat{A}(t)$ on $\mathcal{F}(\mathcal{H})$. For the interaction picture density matrix, we then find

$$\begin{aligned}
i\hbar\frac{\partial\hat{\rho}_I(t)}{\partial t} &= i\hbar\Big(\frac{\partial}{\partial t}e^{i\hat{H}_0 t/\hbar}\Big)\hat{\rho}(t)e^{-i\hat{H}_0 t/\hbar} + i\hbar e^{i\hat{H}_0 t/\hbar}\frac{\partial\hat{\rho}(t)}{\partial t}e^{-i\hat{H}_0 t/\hbar} + i\hbar e^{i\hat{H}_0 t/\hbar}\hat{\rho}(t)\Big(\frac{\partial}{\partial t}e^{-i\hat{H}_0 t/\hbar}\Big) \\
&= -\hat{H}_0\hat{\rho}_I(t) + e^{i\hat{H}_0 t/\hbar}[\hat{H}(t),\hat{\rho}(t)]e^{-i\hat{H}_0 t/\hbar} + \hat{\rho}_I(t)\hat{H}_0 \\
&= -[\hat{H}_0,\hat{\rho}_I(t)] + e^{i\hat{H}_0 t/\hbar}[\hat{H}_0 + \hat{W}(t),\hat{\rho}(t)]e^{-i\hat{H}_0 t/\hbar}.
\end{aligned} \tag{6.3}$$

Note that $e^{i\hat{H}_0 t/\hbar}$ and $\hat{H}_0$ commute, so

$$e^{i\hat{H}_0 t/\hbar}[\hat{H}_0,\hat{\rho}(t)]e^{-i\hat{H}_0 t/\hbar} = [\hat{H}_0, e^{i\hat{H}_0 t/\hbar}\hat{\rho}(t)e^{-i\hat{H}_0 t/\hbar}] = [\hat{H}_0,\hat{\rho}_I(t)].$$

Also,

$$e^{i\hat{H}_0 t/\hbar}[\hat{W}(t),\hat{\rho}(t)]e^{-i\hat{H}_0 t/\hbar} = \hat{W}_I(t)\hat{\rho}_I(t) - \hat{\rho}_I(t)\hat{W}_I(t) = [\hat{W}_I(t),\hat{\rho}_I(t)],$$

so putting everything together, we find

$$i\hbar\frac{\partial\hat{\rho}_I(t)}{\partial t} = [\hat{W}_I(t),\hat{\rho}_I(t)], \tag{6.4}$$

which is just the von Neumann equation in the interaction picture. Its integral form is

$$\hat{\rho}_I(t) = \hat{\rho}_0 + \frac{1}{i\hbar}\int_{-\infty}^{t} \mathrm{d}t_1[\hat{W}_I(t_1),\hat{\rho}_I(t_1)],$$

where we have made the assumption that $\hat{W}(t)$ goes to zero in the limit $t \to -\infty$. Iteratively plugging this equation back into itself yields the solution

$$\hat{\rho}_I(t) = \hat{\rho}_0 + \sum_{n=1}^{\infty}\frac{1}{(i\hbar)^n}\int_{-\infty}^{t}\mathrm{d}t_1\int_{-\infty}^{t_1}\mathrm{d}t_2\cdots\int_{-\infty}^{t_{n-1}}\mathrm{d}t_n[\hat{W}_I(t_1),[\hat{W}_I(t_2),[\ldots[\hat{W}_I(t_n),\hat{\rho}_0]\ldots]]],$$

so to first order in the perturbation $\hat{W}(t)$,

$$\hat{\rho}_I(t) = \hat{\rho}_0 + \frac{1}{i\hbar}\int_{-\infty}^{t}\mathrm{d}t_1[\hat{W}_I(t_1),\hat{\rho}_0]. \tag{6.5}$$

Note that $[\hat{H}_0,\hat{\rho}_0] = 0$ by Equation (6.1), so Equation (6.5) implies

$$\hat{\rho}(t) = e^{-i\hat{H}_0 t/\hbar}\hat{\rho}_I(t)e^{i\hat{H}_0 t/\hbar} = \hat{\rho}_0 + \frac{1}{i\hbar}\int_{-\infty}^{t}\mathrm{d}t_1[e^{-i\hat{H}_0 t/\hbar}\hat{W}_I(t_1)e^{i\hat{H}_0 t/\hbar},\hat{\rho}_0].$$

Now, assume the perturbation can be written as a product $\hat{W}(t) = \hat{Q}F(t)$ of a real-valued function of time F and an operator $\hat{Q}$ with no explicit time dependence, then

$$e^{-i\hat{H}_0 t/\hbar}\hat{W}_I(t_1)e^{i\hat{H}_0 t/\hbar} = e^{i\hat{H}_0(t_1-t)/\hbar}\hat{Q}e^{-i\hat{H}_0(t_1-t)/\hbar}F(t_1) = \hat{Q}_I(t_1-t)F(t_1)$$

and thus

$$\hat{\rho}(t) = \hat{\rho}_0 + \frac{1}{i\hbar}\int_{-\infty}^{0}\mathrm{d}t_1[\hat{Q}_I(t_1),\hat{\rho}_0]F(t_1+t). \tag{6.6}$$

With this approximate solution to (6.2), we can now compute the linear response to the perturbation $\hat{W}$ of any observable $\hat{A}$. The expectation value of $\hat{A}$ at time t is given in terms of the density operator $\hat{\rho}(t)$ by

$$\langle\hat{A}\rangle(t) = \operatorname{Tr}\hat{\rho}(t)\hat{A},$$

so using Equation (6.6), we find

$$\boxed{\langle\hat{A}\rangle(t) = \langle\hat{A}\rangle_0 + \frac{1}{i\hbar}\int_{-\infty}^{0}\mathrm{d}t_1\operatorname{Tr}\big([\hat{Q}_I(t_1),\hat{\rho}_0]\hat{A}\big)F(t_1+t),} \tag{6.7}$$

where $\langle\hat{A}\rangle_0 \coloneqq \operatorname{Tr}\hat{\rho}_0\hat{A}$ is the expectation value of $\hat{A}$ in the absence of any perturbation.

6.2 The Kubo formula for conductivity

This expression for the linear response of an observable $\hat{A}$ can now be applied to the current density $\hat{\mathbf{j}}$, defined as

$$\hat{\mathbf{j}}(t) \coloneqq \frac{q}{i\hbar V}[\mathrm{d}\Gamma(\hat{\mathbf{r}}^{(1)}), \hat{H}(t)] \tag{6.8}$$

with q the charge of the particles in the system, V its volume and $\hat{\mathbf{r}}^{(1)}$ the position operator on $\mathcal{H}$. On the N-particle subspace $\mathcal{H}_a^{(N)} \subseteq \mathcal{F}(\mathcal{H})$, Equation (B.1) gives

$$\hat{\mathbf{j}}(t)|_{\mathcal{H}_a^{(N)}} = \frac{q}{i\hbar V}[\mathrm{d}\Gamma^{(N)}(\hat{\mathbf{r}}^{(1)}), \hat{H}^{(N)}(t)] = \frac{q}{i\hbar V}\sum_{n=1}^{N}[\hat{\mathbf{r}}_n, \hat{H}^{(N)}(t)] = \frac{1}{V}\sum_{n=1}^{N} q\hat{\mathbf{v}}_n(t), \tag{6.9}$$

where $\hat{\mathbf{r}}_n \coloneqq \mathrm{d}\Gamma_n^{(N)}(\hat{\mathbf{r}}^{(1)})$ is the position of particle n,

$$\hat{\mathbf{v}}_n(t) \coloneqq \frac{1}{i\hbar}[\hat{\mathbf{r}}_n, \hat{H}^{(N)}(t)]$$

its velocity and $\hat{H}^{(N)}(t)$ the Hamiltonian on $\mathcal{H}_a^{(N)}$. Equation (6.9) is just the classical expression for current density. In order to be able to compute the conductivity of the system, we need to know the linear response of $\mathbf{j}$ in the presence of an external electric field. A field of the form $\mathbf{E}(t) = \mathbf{E}_0 e^{-i(\omega+i\delta)t}$ (with $\delta > 0$ so that $\mathbf{E}(t) \to \mathbf{0}$ for $t \to -\infty$) gives rise to a perturbation energy

$$\hat{W}(t) = -\hat{\mathbf{P}} \cdot \mathbf{E}(t) = \hat{Q}F(t),$$

with $F(t) \coloneqq e^{-i(\omega+i\delta)t}$ and

$$\hat{Q} \coloneqq -\hat{\mathbf{P}} \cdot \mathbf{E}_0 = -\sum_{\nu} \hat{P}_\nu E_{0\nu}.$$

Here, $\hat{\mathbf{P}} \coloneqq \mathrm{d}\Gamma(q\hat{\mathbf{r}}^{(1)})$ is the polarization operator, given on $\mathcal{H}_a^{(N)}$ by the classical expression

$$\hat{\mathbf{P}}|_{\mathcal{H}_a^{(N)}} = \sum_{n=1}^{N} q\hat{\mathbf{r}}_n. \tag{6.10}$$

Under the assumption $\langle\hat{\mathbf{j}}\rangle_0 = \mathbf{0}$, Equation (6.7) now gives

$$\begin{aligned}\langle\hat{j}_\mu\rangle(t) &= \frac{1}{i\hbar}\int_{-\infty}^{0} \mathrm{d}t_1 \operatorname{Tr}\Big(\big[-\sum_{\nu}\hat{P}_{\nu I}(t_1)E_{0\nu}, \hat{\rho}_0\big]\hat{j}_\mu\Big)e^{-i(\omega+i\delta)(t_1+t)} \\ &= \sum_{\nu}\frac{i}{\hbar}\int_{-\infty}^{0} \mathrm{d}t_1 \operatorname{Tr}\big([\hat{P}_{\nu I}(t_1), \hat{\rho}_0]\hat{j}_\mu\big)e^{-i(\omega+i\delta)t_1}E_\nu(t),\end{aligned}$$

so using $\mathbf{j} = \boldsymbol{\sigma}\mathbf{E}$, we get

$$\sigma_{\mu\nu}(\omega) = \frac{i}{\hbar}\int_{-\infty}^{0} \mathrm{d}t \operatorname{Tr}\big([\hat{P}_{\nu I}(t), \hat{\rho}_0]\hat{j}_\mu\big)e^{-i(\omega+i\delta)t}. \tag{6.11}$$

It is important to note that whereas Equation (6.7) gives only an *approximation* of the true expectation value of an observable in the presence of a time-dependent perturbation, Equation (6.11) is *exact*, essentially by definition of the conductivity. In many materials, the relation $\mathbf{j} = \boldsymbol{\sigma}\mathbf{E}$ is only approximately true; $\boldsymbol{\sigma}$ is just the "coefficient" of the linear term in the power series expansion of $\mathbf{j}$ in $\mathbf{E}$, which is precisely what we have calculated here.

Proceeding with the derivation, note that for any time-independent operator $\hat{A}$ commuting with $\hat{N}$,

$$\begin{aligned}\hat{\rho}_0\hat{A}_I(t - i\hbar\beta) &= \frac{1}{Z_0}e^{-\beta(\hat{H}_0-\mu\hat{N})}e^{i\hat{H}_0(t-i\hbar\beta)/\hbar}\hat{A}e^{-i\hat{H}_0(t-i\hbar\beta)/\hbar} \\ &= \frac{1}{Z_0}e^{i\hat{H}_0 t/\hbar}\hat{A}e^{-i\hat{H}_0 t/\hbar}e^{-\beta(\hat{H}_0-\mu\hat{N})} = \hat{A}_I(t)\hat{\rho}_0\end{aligned}$$

and thus

$$\hat{\rho}_0 \int_0^\beta \mathrm{d}\lambda \frac{\partial}{\partial t} \hat{A}_I(t - i\hbar\lambda) = \hat{\rho}_0 \int_t^{t-i\hbar\beta} \Big(-\frac{\mathrm{d}t'}{i\hbar}\Big) \frac{\partial}{\partial t} \hat{A}_I(t') = \frac{i}{\hbar}\hat{\rho}_0(\hat{A}_I(t - i\hbar\beta) - \hat{A}_I(t))$$
$$= \frac{i}{\hbar}(\hat{A}_I(t)\hat{\rho}_0 - \hat{\rho}_0\hat{A}_I(t)) = \frac{i}{\hbar}[\hat{A}_I(t), \hat{\rho}_0].$$

With this, Equation (6.11) becomes

$$\sigma_{\mu\nu}(\omega) = \int_{-\infty}^0 \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \, \mathrm{Tr}\Big(\hat{\rho}_0 \frac{\partial}{\partial t}\hat{P}_{\nu I}(t - i\hbar\lambda)\hat{j}_\mu\Big) e^{-i(\omega+i\delta)t}$$
$$= \int_{-\infty}^0 \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \Big\langle \frac{\partial}{\partial t}\hat{P}_{\nu I}(t - i\hbar\lambda)\hat{j}_\mu \Big\rangle_0 e^{-i(\omega+i\delta)t}.$$

One might wonder why we went through so much effort only to introduce an extra integral as well as a derivative into the expression for $\sigma_{\mu\nu}$. The reason is that we can relate the time derivative of the (interaction picture) polarization operator to the current density, as follows. First note that the linearity of $\mathrm{d}\Gamma$ (Lemma B.1) implies

$$\hat{W}(t) = -\,\mathrm{d}\Gamma(q\hat{\mathbf{r}}^{(1)}) \cdot \mathbf{E}(t) = -\,\mathrm{d}\Gamma(q\hat{\mathbf{r}}^{(1)} \cdot \mathbf{E}(t)),$$

so by that same lemma,

$$[\mathrm{d}\Gamma(\hat{\mathbf{r}}^{(1)}), \hat{W}(t)] = -[\mathrm{d}\Gamma(\hat{\mathbf{r}}^{(1)}), \mathrm{d}\Gamma(q\hat{\mathbf{r}}^{(1)} \cdot \mathbf{E}(t))] = -\,\mathrm{d}\Gamma([\hat{\mathbf{r}}^{(1)}, q\hat{\mathbf{r}}^{(1)} \cdot \mathbf{E}(t)]) = \mathbf{0}. \tag{6.12}$$

It follows that we can replace $\hat{H}(t)$ by $\hat{H}_0$ in Equation (6.8), so

$$V\hat{\mathbf{j}}_I(t) = e^{i\hat{H}_0 t/\hbar} \frac{1}{i\hbar}[\mathrm{d}\Gamma(q\hat{\mathbf{r}}^{(1)}), \hat{H}_0] e^{-i\hat{H}_0 t/\hbar}$$
$$= \frac{1}{i\hbar}[e^{i\hat{H}_0 t/\hbar}\hat{\mathbf{P}}e^{-i\hat{H}_0 t/\hbar}, \hat{H}_0]$$
$$= \frac{1}{i\hbar}[\hat{\mathbf{P}}_I(t), \hat{H}_0] = \frac{\partial \hat{\mathbf{P}}_I(t)}{\partial t}$$

by a similar argument as in Equation (6.3) and thus

$$\sigma_{\mu\nu}(\omega) = V \int_{-\infty}^0 \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \langle \hat{j}_{\nu I}(t - i\hbar\lambda)\hat{j}_\mu \rangle_0 e^{-i(\omega+i\delta)t}$$
$$= V \int_0^\infty \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \langle \hat{j}_\nu \hat{j}_{\mu I}(t + i\hbar\lambda) \rangle_0 e^{i(\omega+i\delta)t}.$$

In the last equality, we used that

$$\mathrm{Tr}(\hat{\rho}_0\hat{j}_{\nu I}(-t - i\hbar\lambda)\hat{j}_\mu) = \mathrm{Tr}(\hat{\rho}_0 e^{-i\hat{H}_0(t+i\hbar\lambda)/\hbar}\hat{j}_\nu e^{i\hat{H}_0(t+i\hbar\lambda)/\hbar}\hat{j}_\mu)$$
$$= \mathrm{Tr}(\hat{\rho}_0\hat{j}_\nu e^{i\hat{H}_0(t+i\hbar\lambda)/\hbar}\hat{j}_\mu e^{-i\hat{H}_0(t+i\hbar\lambda)/\hbar})$$
$$= \mathrm{Tr}(\hat{\rho}_0\hat{j}_\nu\hat{j}_{\mu I}(t + i\hbar\lambda)),$$

since $\hat{\rho}_0$ commutes with $e^{-i\hat{H}_0(t+i\hbar\lambda)/\hbar}$ and the trace is invariant under cyclic permutations. We conclude that the DC conductivity reads

$$\boxed{\sigma_{\mu\nu} = V \int_0^\infty \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \langle \hat{j}_\nu \hat{j}_{\mu I}(t + i\hbar\lambda) \rangle_0 e^{-\delta t}.} \tag{6.13}$$

This is the *Kubo formula* for the conductivity, first derived in 1957 by Ryogo Kubo [16]. In order to actually be able to use it in practice, we need to make some simplifying assumptions.

6.3 The independent electron approximation

In the *independent electron approximation*, electron-electron interactions in the system are ignored. This implies that the Hamiltonian can be written as $\hat{H}_0 = \mathrm{d}\Gamma(\hat{H}_0^{(1)})$, where $\hat{H}_0^{(1)}$ is the one-particle Hamiltonian operator on $\mathcal{H}$ (see Appendix B.1). We denote its eigenstates by $|k\rangle$ and the corresponding energy eigenvalues by ϵ_k. Substitution into Equation (6.8) yields

$$\hat{\mathbf{j}} = \frac{q}{i\hbar V}[\mathrm{d}\Gamma(\hat{\mathbf{r}}^{(1)}), \mathrm{d}\Gamma(\hat{H}_0^{(1)})] = \frac{q}{i\hbar V}\,\mathrm{d}\Gamma([\hat{\mathbf{r}}^{(1)}, \hat{H}_0^{(1)}]) = \mathrm{d}\Gamma(\hat{\mathbf{j}}^{(1)})$$

with

$$\hat{\mathbf{j}}^{(1)} \coloneqq \frac{1}{V} q \hat{\mathbf{v}}^{(1)} \qquad \text{and} \qquad \hat{\mathbf{v}}^{(1)} \coloneqq \frac{1}{i\hbar}[\hat{\mathbf{r}}^{(1)}, \hat{H}_0^{(1)}] \tag{6.14}$$

the current density and velocity operators on $\mathcal{H}$, respectively. We again used Equation (6.12) to replace $\hat{H}(t)$ by $\hat{H}_0$ in the definition of $\hat{\mathbf{j}}$ and Lemma B.1 to pull $\mathrm{d}\Gamma$ out of the commutator. Lemma B.1 also gives us

$$\hat{\mathbf{j}}_I(t) = e^{i\hat{H}_0 t/\hbar}\hat{\mathbf{j}}e^{-i\hat{H}_0 t/\hbar} = \mathrm{d}\Gamma(e^{i\hat{H}_0^{(1)} t/\hbar}\hat{\mathbf{j}}^{(1)} e^{-i\hat{H}_0^{(1)} t/\hbar}) = \mathrm{d}\Gamma(\hat{\mathbf{j}}_I^{(1)}(t)),$$

so by Lemma B.2,

$$\begin{aligned}
\hat{H}_0 &= \sum_{k,\ell} \langle k|\hat{H}_0^{(1)}|\ell\rangle \hat{c}_k^* \hat{c}_\ell = \sum_k \epsilon_k \hat{c}_k^* \hat{c}_k = \sum_k \epsilon_k \hat{n}_k, \\
\hat{j}_\nu &= \sum_{k,\ell} \langle k|\hat{j}_\nu^{(1)}|\ell\rangle \hat{c}_k^* \hat{c}_\ell, \\
\hat{j}_{\mu I}(t) &= \sum_{k,\ell} \langle k|\hat{j}_{\mu I}^{(1)}(t)|\ell\rangle \hat{c}_k^* \hat{c}_\ell
\end{aligned}$$

where $\hat{c}_k^*$ and $\hat{c}_k$ are the fermionic creation and annihilation operators. The corresponding occupation number operator $\hat{n}_k = \hat{c}_k^* \hat{c}_k$ is known to have expectation value $\langle \hat{n}_k \rangle = f(\epsilon_k)$, where

$$f(\epsilon) \coloneqq \frac{1}{e^{\beta(\epsilon-\mu)} + 1}$$

is the Fermi-Dirac distribution.

These expansions of the Hamiltonian and current density operators allow us to compute the trace in the Kubo formula (6.13). Noting that

$$\begin{aligned}
\langle \ell|\hat{j}_{\mu I}^{(1)}(t + i\hbar\lambda)|k\rangle &= \langle \ell|e^{i\hat{H}_0^{(1)} t/\hbar} e^{-\lambda \hat{H}_0^{(1)}} \hat{j}_\mu^{(1)} e^{-i\hat{H}_0^{(1)} t/\hbar} e^{\lambda \hat{H}_0^{(1)}}|k\rangle \\
&= e^{-i\epsilon_k t/\hbar} e^{\lambda\epsilon_k} \overline{\langle k|(e^{i\hat{H}_0^{(1)} t/\hbar} e^{-\lambda \hat{H}_0^{(1)}} \hat{j}_\mu^{(1)})^*|\ell\rangle} \\
&= e^{-i\epsilon_k t/\hbar} e^{\lambda\epsilon_k} \overline{\langle k|\hat{j}_\mu^{(1)} e^{-i\hat{H}_0^{(1)} t/\hbar} e^{-\lambda \hat{H}_0^{(1)}}|\ell\rangle} \\
&= e^{-i\epsilon_k t/\hbar} e^{\lambda\epsilon_k} e^{i\epsilon_\ell t/\hbar} e^{-\lambda\epsilon_\ell} \overline{\langle k|\hat{j}_\mu^{(1)}|\ell\rangle} \\
&= e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\lambda(\epsilon_\ell - \epsilon_k)} \langle \ell|\hat{j}_\mu^{(1)}|k\rangle
\end{aligned}$$

for any k, ℓ, we get

$$\langle \hat{j}_\nu \hat{j}_{\mu I}(t + i\hbar\lambda)\rangle_0 = \sum_{k_1,k_2,k_3,k_4} \langle k_1|\hat{j}_\nu^{(1)}|k_2\rangle \langle k_3|\hat{j}_\mu^{(1)}|k_4\rangle e^{i(\epsilon_{k_3} - \epsilon_{k_4})t/\hbar} e^{-\lambda(\epsilon_{k_3} - \epsilon_{k_4})} \langle \hat{c}_{k_1}^* \hat{c}_{k_2} \hat{c}_{k_3}^* \hat{c}_{k_4}\rangle_0. \tag{6.15}$$

Taking the Fock states $|(n_k)\rangle$ as the basis states for the trace, where $n_k \in \{0, 1\}$ is the number of particles in state $|k\rangle$, it becomes clear that there are only three cases in which the trace does not vanish:

(1) *Case 1*: $k_1 = k_2 = k_3 = k_4 =: k$. We have

$$\hat{c}_k^* \hat{c}_k \hat{c}_k^* \hat{c}_k = \hat{c}_k^*(1 - \hat{c}_k^* \hat{c}_k)\hat{c}_k = \hat{c}_k^* \hat{c}_k - (\hat{c}_k^*)^2 (\hat{c}_k)^2 = \hat{n}_k$$

since $\{\hat{c}_k, \hat{c}_k^*\} = 1$ by Equation (B.2), so in this case,

$$\langle \hat{c}_{k_1}^* \hat{c}_{k_2} \hat{c}_{k_3}^* \hat{c}_{k_4} \rangle_0 = \langle \hat{n}_k \rangle = f(\epsilon_k).$$

(2) *Case 2*: $k_1 = k_2 =: k \neq \ell := k_3 = k_4$. Heuristically, the occupation number operators $\hat{n}_k$ and $\hat{n}_\ell$ are independent as random variables, so

$$\langle \hat{c}_{k_1}^* \hat{c}_{k_2} \hat{c}_{k_3}^* \hat{c}_{k_4} \rangle_0 = \langle \hat{n}_k \hat{n}_\ell \rangle_0 = \langle \hat{n}_k \rangle_0 \langle \hat{n}_\ell \rangle_0 = f(\epsilon_k) f(\epsilon_\ell).$$

This can also be derived by simply evaluating the trace

$$\langle \hat{c}_k^* \hat{c}_k \hat{c}_\ell^* \hat{c}_\ell \rangle_0 = \sum_{(n_{k'})} \langle (n_{k'}) | \hat{\rho}_0 \hat{n}_k \hat{n}_\ell | (n_{k'}) \rangle = \sum_{\substack{(n_{k'}): \\ n_k = n_\ell = 1}} \langle (n_{k'}) | \hat{\rho}_0 | (n_{k'}) \rangle.$$

We know $\hat{\rho}_0 = e^{-\beta(\hat{H}_0 - \mu \hat{N})}/Z_0$ from Equation (6.1) and $\hat{H}_0 - \mu \hat{N} = \sum_{k'} (\epsilon_{k'} - \mu) \hat{n}_{k'}$, so

$$\begin{aligned}\langle \hat{c}_k^* \hat{c}_k \hat{c}_\ell^* \hat{c}_\ell \rangle_0 = \frac{1}{Z_0} \sum_{\substack{(n_{k'}): \\ n_k = n_\ell = 1}} e^{-\beta \sum_{k'} (\epsilon_{k'} - \mu) n_{k'}} &= \frac{1}{Z_0} e^{-\beta(\epsilon_k - \mu)} e^{-\beta(\epsilon_\ell - \mu)} \sum_{\substack{(n_{k'}): \\ k' \neq k, \ell}} \prod_{k' \neq k, \ell} e^{-\beta(\epsilon_{k'} - \mu) n_{k'}} \\ &= \frac{1}{Z_0} e^{-\beta(\epsilon_k - \mu)} e^{-\beta(\epsilon_\ell - \mu)} \prod_{k' \neq k, \ell} (1 + e^{-\beta(\epsilon_{k'} - \mu)}).\end{aligned}$$

Applying the same trick of swapping sum and product to Z_0 gives

$$Z_0 = \prod_{k'} (1 + e^{-\beta(\epsilon_{k'} - \mu)})$$

and thus

$$\langle \hat{c}_k^* \hat{c}_k \hat{c}_\ell^* \hat{c}_\ell \rangle_0 = \frac{e^{-\beta(\epsilon_k - \mu)} e^{-\beta(\epsilon_\ell - \mu)}}{(1 + e^{-\beta(\epsilon_k - \mu)})(1 + e^{-\beta(\epsilon_\ell - \mu)})} = f(\epsilon_k) f(\epsilon_\ell).$$

(3) *Case 3*: $k_1 = k_4 =: k \neq \ell := k_2 = k_3$. Note that

$$\hat{c}_k^* \hat{c}_\ell \hat{c}_\ell^* \hat{c}_k = -\hat{c}_k^* \hat{c}_\ell \hat{c}_k \hat{c}_\ell^* = \hat{c}_k^* \hat{c}_k \hat{c}_\ell \hat{c}_\ell^* = \hat{c}_k^* \hat{c}_k (1 - \hat{c}_\ell^* \hat{c}_\ell) = \hat{n}_k - \hat{n}_k \hat{n}_\ell$$

by the anticommutation relations $\{\hat{c}_k^*, \hat{c}_\ell^*\} = 0$, $\{\hat{c}_k, \hat{c}_\ell\} = 0$ and $\{\hat{c}_\ell, \hat{c}_\ell^*\} = 1$, see (B.2). Using case 2, we find

$$\langle \hat{c}_{k_1}^* \hat{c}_{k_2} \hat{c}_{k_3}^* \hat{c}_{k_4} \rangle_0 = \langle \hat{n}_k \rangle_0 - \langle \hat{n}_k \hat{n}_\ell \rangle_0 = f(\epsilon_k)(1 - f(\epsilon_\ell)).$$

Combining this with Equation (6.15) gives

$$\begin{aligned}\langle \hat{j}_\nu \hat{j}_{\mu I}(t + i\hbar\lambda) \rangle_0 = &\sum_k \langle k | \hat{j}_\nu^{(1)} | k \rangle \langle k | \hat{j}_\mu^{(1)} | k \rangle f(\epsilon_k) + \\ &\sum_{k \neq \ell} \langle k | \hat{j}_\nu^{(1)} | k \rangle \langle \ell | \hat{j}_\mu^{(1)} | \ell \rangle f(\epsilon_k) f(\epsilon_\ell) + \\ &\quad \sum_{k \neq \ell} \langle k | \hat{j}_\nu^{(1)} | \ell \rangle \langle \ell | \hat{j}_\mu^{(1)} | k \rangle e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\lambda(\epsilon_\ell - \epsilon_k)} f(\epsilon_k)(1 - f(\epsilon_\ell)).\end{aligned}$$

Now, $\langle \hat{\mathbf{j}} \rangle_0 = \mathbf{0}$ by assumption, so

$$0 = \langle \hat{j}_\lambda \rangle_0 = \sum_{k,\ell} \langle k | \hat{j}_\lambda^{(1)} | \ell \rangle \langle \hat{c}_k^* \hat{c}_\ell \rangle_0 = \sum_k \langle k | \hat{j}_\lambda^{(1)} | k \rangle \langle \hat{n}_k \rangle_0 = \sum_k \langle k | \hat{j}_\lambda^{(1)} | k \rangle f(\epsilon_k).$$

for any λ. It follows that

$$\begin{aligned}0 = \langle \hat{j}_\nu \rangle_0 \langle \hat{j}_\mu \rangle_0 &= \left(\sum_k \langle k | \hat{j}_\nu^{(1)} | k \rangle f(\epsilon_k) \right) \left(\sum_\ell \langle \ell | \hat{j}_\mu^{(1)} | \ell \rangle f(\epsilon_\ell) \right) \\ &= \sum_{k \neq \ell} \langle k | \hat{j}_\nu^{(1)} | k \rangle \langle \ell | \hat{j}_\mu^{(1)} | \ell \rangle f(\epsilon_k) f(\epsilon_\ell) + \sum_k \langle k | \hat{j}_\nu^{(1)} | k \rangle \langle k | \hat{j}_\mu^{(1)} | k \rangle f(\epsilon_k)^2\end{aligned}$$

and thus

$$\begin{aligned}\langle \hat{j}_\nu \hat{j}_{\mu I}(t + i\hbar\lambda) \rangle_0 &= \sum_k \langle k | \hat{j}_\nu^{(1)} | k \rangle \langle k | \hat{j}_\mu^{(1)} | k \rangle f(\epsilon_k)(1 - f(\epsilon_k)) + \\ &\quad \sum_{k \neq \ell} \langle k | \hat{j}_\nu^{(1)} | \ell \rangle \langle \ell | \hat{j}_\mu^{(1)} | k \rangle e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\lambda(\epsilon_\ell - \epsilon_k)} f(\epsilon_k)(1 - f(\epsilon_\ell)) \\ &= \sum_{k,\ell} \langle k | \hat{j}_\nu^{(1)} | \ell \rangle \langle \ell | \hat{j}_\mu^{(1)} | k \rangle e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\lambda(\epsilon_\ell - \epsilon_k)} f(\epsilon_k)(1 - f(\epsilon_\ell)).\end{aligned}$$

Substitution into Equation (6.13) yields

$$\sigma_{\mu\nu} = V \sum_{k,\ell} \int_0^\infty \mathrm{d}t \; e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\delta t} \int_0^\beta \mathrm{d}\lambda \; e^{-\lambda(\epsilon_\ell - \epsilon_k)} f(\epsilon_k)(1 - f(\epsilon_\ell)) \langle k | \hat{j}_\nu^{(1)} | \ell \rangle \langle \ell | \hat{j}_\mu^{(1)} | k \rangle,$$

which we can simplify further. We compute

$$\begin{aligned}\int_0^\infty \mathrm{d}t \; e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\delta t} &= \frac{\hbar}{i(\epsilon_\ell - \epsilon_k) - \hbar\delta} e^{i(\epsilon_\ell - \epsilon_k)t/\hbar} e^{-\delta t} \Big|_0^\infty = -\frac{i\hbar}{\epsilon_k - \epsilon_\ell - i\hbar\delta}, \\ \int_0^\beta \mathrm{d}\lambda \; e^{\lambda(\epsilon_k - \epsilon_\ell)} &= \frac{1}{\epsilon_k - \epsilon_\ell} e^{\lambda(\epsilon_k - \epsilon_\ell)} \Big|_0^\beta = \frac{e^{\beta(\epsilon_k - \epsilon_\ell)} - 1}{\epsilon_k - \epsilon_\ell},\end{aligned}$$

interpreting the second integral as β if $\epsilon_k = \epsilon_\ell$. Noting that

$$\begin{aligned}(e^{\beta(\epsilon_k - \epsilon_\ell)} - 1) f(\epsilon_k)(1 - f(\epsilon_\ell)) &= (e^{\beta(\epsilon_k - \epsilon_\ell)} - 1) \cdot \frac{1}{e^{\beta(\epsilon_k - \mu)} + 1} \cdot \frac{e^{\beta(\epsilon_\ell - \mu)}}{e^{\beta(\epsilon_\ell - \mu)} + 1} \\ &= \frac{e^{\beta(\epsilon_k - \mu)} - e^{\beta(\epsilon_\ell - \mu)}}{(e^{\beta(\epsilon_k - \mu)} + 1)(e^{\beta(\epsilon_\ell - \mu)} + 1)} = f(\epsilon_\ell) - f(\epsilon_k),\end{aligned}$$

we arrive at the final expression

$$\boxed{\sigma_{\mu\nu} = i\hbar V \sum_{k,\ell} \frac{f(\epsilon_k) - f(\epsilon_\ell)}{(\epsilon_k - \epsilon_\ell - i\hbar\delta)(\epsilon_k - \epsilon_\ell)} \langle k | \hat{j}_\nu^{(1)} | \ell \rangle \langle \ell | \hat{j}_\mu^{(1)} | k \rangle.} \tag{6.16}$$

This expression is consistent with the literature, see for instance Equation (1.2) in [5]. It may seem like the $\epsilon_k = \epsilon_\ell$ terms blow it up, but this is not the case since the λ integral is just β if $\epsilon_k = \epsilon_\ell$.

6.4 The TKNN formula

At last, we can reduce (6.16) to a form that has a geometric interpretation in the context of connections on principal $U(1)$-bundles. This was first done by Thouless, Kohmoto, Nightingale and Den Nijs in 1982 [7]; we follow [15, Section 4.2]. All operators now act on the one-particle Hilbert space $\mathcal{H}$, so henceforth, any superscripts $^{(1)}$ will be dropped.

This is the point where we assume $\hat{H}_0$ is a *Bloch Hamiltonian*, i.e. that it commutes with all translations by elements of a full rank lattice Λ in $\mathbb{R}^d$, with d the dimension of the system (usually 2 or 3). That is, if we define for all $\mathbf{R} \in \Lambda$ an operator $\hat{T}_{\mathbf{R}}$ on $\mathcal{H}$ by $\hat{T}_{\mathbf{R}}\psi(\mathbf{r}) \coloneqq \psi(\mathbf{r}+\mathbf{R})$, then we require $[\hat{H}_0, \hat{T}_{\mathbf{R}}] = 0$ for all $\mathbf{R} \in \Lambda$. The simplest Bloch Hamiltonians are those of the form

$$\hat{H}_0 = -\frac{\hbar^2}{2m}\nabla_{\mathbf{r}}^2 + U(\mathbf{r}) \tag{6.17}$$

for some Λ-periodic potential energy function $U(\mathbf{r})$, but in general, $\hat{H}_0$ can take many different shapes. For instance, magnetic fields can give rise to terms *linear* in the momentum $\hat{\mathbf{p}}$ and if spin is included in the picture, terms involving the Pauli matrices can pop up too.

We now attach some symbols to the crystal lattice Λ. Choose a basis matrix $A \in \mathrm{GL}(d, \mathbb{R})$ for Λ, then

$$\Lambda = \{A\mathbf{m} : \mathbf{m} \in \mathbb{Z}^d\}$$

and $\Gamma \coloneqq A[0,1)^d$ is a primitive unit cell. For simplicity, assume the system is made up of finitely many such unit cells, arranged in a d-dimensional cube. In other words, there exists $L \in \mathbb{Z}_{>0}$ such that the particles are confined to

$$\Gamma_L \coloneqq A[0,L)^d = \bigcup_{\mathbf{R}\in\Lambda_L} (\mathbf{R} + \Gamma),$$

where

$$\Lambda_L \coloneqq \{A\mathbf{m} : \mathbf{m} \in \{0, 1, \ldots, L-1\}^d\} \subseteq \Lambda.$$

The number of unit cells is $N \coloneqq |\Lambda_L| = L^d$ and the system volume is $V = \mathrm{vol}(\Gamma_L) = NV_0$ with $V_0 \coloneqq \mathrm{vol}(\Gamma) = |\det(A)|$ the determinant of Λ. The *reciprocal lattice* of Λ is defined as

$$\Lambda^* \coloneqq \{B\mathbf{n} : \mathbf{n} \in \mathbb{Z}^d\}$$

with $B \coloneqq 2\pi A^{-T} \in \mathrm{GL}(d, \mathbb{R})$. Its has determinant $|\det(B)| = (2\pi)^d/V_0$ and can also be characterized as follows.

Lemma 6.1. *Let $\mathbf{K} \in \mathbb{R}^d$, then $\mathbf{K} \in \Lambda^*$ if and only if $e^{i\mathbf{K}\cdot\mathbf{R}} = 1$ for all $\mathbf{R} \in \Lambda$.*

Proof. First suppose $\mathbf{K} \in \Lambda^*$, then $\mathbf{K} = B\mathbf{n}$ for some $\mathbf{n} \in \mathbb{Z}^d$. Note that $B^T A = 2\pi A^{-1}A = 2\pi I_d$, so for any $\mathbf{R} = A\mathbf{m} \in \Lambda$,

$$e^{i\mathbf{K}\cdot\mathbf{R}} = e^{i(B\mathbf{n})^T(A\mathbf{m})} = e^{i\mathbf{n}^T B^T A\mathbf{m}} = e^{2\pi i\mathbf{n}\cdot\mathbf{m}} = 1.$$

For the converse, assume $e^{i\mathbf{K}\cdot\mathbf{R}} = 1$ holds for all $\mathbf{R} \in \Lambda$, and let $\mathbf{n} \in \mathbb{R}^d$ be such that $\mathbf{K} = B\mathbf{n}$; we need to show $\mathbf{n} \in \mathbb{Z}^d$. Let $1 \le j \le d$ and denote by $\mathbf{e}_j$ the j-th standard basis vector of $\mathbb{R}^d$. Set $\mathbf{R} \coloneqq A\mathbf{e}_j \in \Lambda$, then by a similar computation as before,

$$1 = e^{i\mathbf{K}\cdot\mathbf{R}} = e^{2\pi i\mathbf{n}\cdot\mathbf{e}_j} = e^{2\pi i n_j}$$

and thus $n_j \in \mathbb{Z}$. □

With this language of lattices and unit cells in place, the eigenstates of $\hat{H}_0$ can be characterized quite easily. Bloch's theorem guarantees the existence of an orthonormal basis for $\mathcal{H}$ consisting of $\hat{H}_0$-eigenstates $|n, \mathbf{k}\rangle$ of the form

$$\langle \mathbf{r}|n, \mathbf{k}\rangle = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}) \tag{6.18}$$

for some Λ-periodic *Bloch function* $u_{n,\mathbf{k}}(\mathbf{r})$. Here, n is an integer known as the *band index* and the *wave vector* $\mathbf{k} \in \mathbb{R}^d$ can be chosen inside the *(first) Brillouin zone*, a primitive unit cell of the reciprocal lattice Λ^*. Any unit cell would suffice, but in the literature, the term Brillouin zone is usually reserved for the *Wigner-Seitz cell* of Λ^*, which is the set of points for which $\mathbf{0}$ is among the closest reciprocal lattice vectors.[1] Furthermore, if we impose periodic boundary conditions on the eigenstates so that $\mathcal{H}$

[1] Strictly speaking, this is not a fundamental domain of Λ^* since it is symmetric around $\mathbf{0}$. To fix this, part of the boundary should be omitted, but this is usually ignored.

can be regarded as a space of $L\Lambda$-periodic functions on $\mathbb{R}^d$, it follows that $\mathbf{k}$ always lies in the discretized reciprocal space

$$\frac{\Lambda^*}{L} := \Big\{\frac{B\mathbf{n}}{L} : \mathbf{n} \in \mathbb{Z}^d\Big\}.$$

Hence, the total number of points in reciprocal space representing a physical state is N, and each such point has volume $(2\pi)^d/(NV_0) = (2\pi)^d/V$. The factor $1/\sqrt{N}$ in Equation (6.18) ensures that the Bloch functions are normalized as elements of the space $\mathcal{H}'$ of Λ-periodic functions with inner product

$$\langle u|v\rangle := \int_\Gamma \mathrm{d}\mathbf{r}\ \overline{u(\mathbf{r})}v(\mathbf{r}). \tag{6.19}$$

Now, define for all $\mathbf{k}$ in the Brillouin zone a new operator

$$\hat{H}(\mathbf{k}) := e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{H}_0 e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.20}$$

on this smaller space $\mathcal{H}'$ and note that $\hat{H}(\mathbf{k})$ is Hermitian.[2] By definition, then, $u \in \mathcal{H}'$ is an eigenstate of $\hat{H}(\mathbf{k})$ with eigenvalue $\epsilon \in \mathbb{R}$ if and only if $e^{i\mathbf{k}\cdot\mathbf{r}}u \in \mathcal{H}$ is an eigenstate of the Hamiltonian $\hat{H}_0$ with that same eigenvalue, so diagonalizing $\hat{H}(\mathbf{k})$ yields precisely the Bloch functions $u_{n,\mathbf{k}}$ with wave vector $\mathbf{k}$. One consequence of this is that *for any* $\mathbf{k}$, the $|u_{n,\mathbf{k}}\rangle$ form an orthonormal basis of $\mathcal{H}'$ since $\hat{H}(\mathbf{k})$ is Hermitian, a fact which we will need later. The definition of this new $\mathbf{k}$-dependent Hamiltonian $\hat{H}(\mathbf{k})$ signifies an important conceptual shift: instead of solving for all $\mathbf{k}$ in the first Brillouin zone simultaneously, we can fix $\mathbf{k}$ and diagonalize $\hat{H}(\mathbf{k})$, regarding the wave vector as a sort of parameter of the system.

Having characterized the eigenstates of the one-particle Hamiltonian, we can now reduce Equation (6.16) further by computing the current density operator's matrix elements. Recall from (6.14) that $\hat{\mathbf{j}} = q\hat{\mathbf{v}}/V$, so Equation (6.16) can be written

$$\sigma_{\mu\nu} = \frac{i\hbar q^2}{V}\sum_{n,n'}\sum_{\mathbf{k},\mathbf{k}'}\frac{f(\epsilon_{n,\mathbf{k}}) - f(\epsilon_{n',\mathbf{k}'})}{(\epsilon_{n,\mathbf{k}} - \epsilon_{n',\mathbf{k}'} - i\hbar\delta)(\epsilon_{n,\mathbf{k}} - \epsilon_{n',\mathbf{k}'})}\langle n,\mathbf{k}|\hat{v}_\nu|n',\mathbf{k}'\rangle\langle n',\mathbf{k}'|\hat{v}_\mu|n,\mathbf{k}\rangle. \tag{6.21}$$

It turns out that the matrix elements of $\hat{\mathbf{v}}$ can be related to those of the $\mathbf{k}$-dependent velocity operator

$$\hat{\mathbf{v}}(\mathbf{k}) := e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{\mathbf{v}}e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.22}$$

on $\mathcal{H}'$, whose definition is of course inspired by (6.20). To see this, first note that for all band indices n, n' and wave vectors $\mathbf{k}, \mathbf{k}' \in \Lambda^*/L$, we have

$$\begin{aligned}\langle n,\mathbf{k}|\hat{\mathbf{v}}|n',\mathbf{k}'\rangle &= \frac{1}{N}\int_{\Gamma_L}\mathrm{d}\mathbf{r}\ e^{-i\mathbf{k}\cdot\mathbf{r}}\overline{u_{n,\mathbf{k}}(\mathbf{r})}\hat{\mathbf{v}}e^{i\mathbf{k}'\cdot\mathbf{r}}u_{n',\mathbf{k}'}(\mathbf{r})\\ &= \frac{1}{N}\int_{\Gamma_L}\mathrm{d}\mathbf{r}\ e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}\overline{u_{n,\mathbf{k}}(\mathbf{r})}\hat{\mathbf{v}}(\mathbf{k}')u_{n',\mathbf{k}'}(\mathbf{r}).\end{aligned}$$

Splitting the domain of integration into N shifted copies of the unit cell Γ yields

$$\begin{aligned}\langle n,\mathbf{k}|\hat{\mathbf{v}}|n',\mathbf{k}'\rangle &= \frac{1}{N}\sum_{\mathbf{R}\in\Lambda_L}\int_{\mathbf{R}+\Gamma}\mathrm{d}\mathbf{r}\ e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}\overline{u_{n,\mathbf{k}}(\mathbf{r})}\hat{\mathbf{v}}(\mathbf{k}')u_{n',\mathbf{k}'}(\mathbf{r})\\ &= \frac{1}{N}\sum_{\mathbf{R}\in\Lambda_L}\int_{\Gamma}\mathrm{d}\mathbf{r}\ e^{i(\mathbf{k}'-\mathbf{k})\cdot(\mathbf{R}+\mathbf{r})}\overline{u_{n,\mathbf{k}}(\mathbf{R}+\mathbf{r})}\hat{\mathbf{v}}(\mathbf{k}')u_{n',\mathbf{k}'}(\mathbf{R}+\mathbf{r})\\ &= \frac{1}{N}\Big(\sum_{\mathbf{R}\in\Lambda_L}e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}}\Big)\int_{\Gamma}\mathrm{d}\mathbf{r}\ e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}\overline{u_{n,\mathbf{k}}(\mathbf{r})}\hat{\mathbf{v}}(\mathbf{k}')u_{n',\mathbf{k}'}(\mathbf{r}),\end{aligned} \tag{6.23}$$

[2] If $\hat{H}_0$ is of the form (6.17), it is easy to see that

$$\hat{H}(\mathbf{k}) = -\frac{\hbar^2}{2m}(\boldsymbol{\nabla}_\mathbf{r} + i\mathbf{k})^2 + U(\mathbf{r}),$$

which is just $\hat{H}_0$ with the momentum operator $\hat{\mathbf{p}} = -i\hbar\boldsymbol{\nabla}_\mathbf{r}$ replaced by $\hat{\mathbf{p}} + \hbar\mathbf{k}$.

where we made use of the Λ-periodicity of the Bloch functions. This can be reduced further using the following lemma.

Lemma 6.2. *If* $\mathbf{k}, \mathbf{k}' \in \Lambda^*/L$, *then*

$$\sum_{\mathbf{R}\in\Lambda_L} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}} = \begin{cases} N & \text{if } \mathbf{k}'-\mathbf{k}\in\Lambda^*, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The case $\mathbf{k}'-\mathbf{k}\in\Lambda^*$ holds by Lemma 6.1 (since $|\Lambda_L| = L^d = N$), so assume $\mathbf{k}'-\mathbf{k}\notin\Lambda^*$ and choose $\mathbf{n}\in\mathbb{Z}^d$ such that $\mathbf{k}'-\mathbf{k} = B\mathbf{n}/L$. Writing $\mathbb{Z}_L \coloneqq \{0,1,\dots,L-1\}$, it follows that

$$\begin{aligned}\sum_{\mathbf{R}\in\Lambda_L} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}} &= \sum_{\mathbf{m}\in\mathbb{Z}_L^d} e^{i(B\mathbf{n})\cdot(A\mathbf{m})/L} = \sum_{\mathbf{m}\in\mathbb{Z}_L^d} e^{2\pi i\mathbf{n}\cdot\mathbf{m}/L} \\ &= \prod_{j=1}^{d}\sum_{m=0}^{L-1} e^{2\pi i n_j m/L} = \prod_{j=1}^{d}\sum_{m=0}^{L-1}\left(e^{2\pi i n_j/L}\right)^m,\end{aligned}$$

where we used $B^TA = 2\pi A^{-1}A = 2\pi I_d$. By assumption, there exists a j such that n_j is not a multiple of L, which means $e^{2\pi i n_j/L}\neq 1$ and thus

$$\sum_{m=0}^{L-1}\left(e^{2\pi i n_j/L}\right)^m = \frac{1-e^{2\pi i n_j}}{1-e^{2\pi i n_j/L}} = 0. \qquad \square$$

Finally, if $\mathbf{k}$ and $\mathbf{k}'$ live inside the first Brillouin zone, $\mathbf{0}$ is the only reciprocal lattice vector that they can differ by. Combining this with Equations (6.19) and (6.23) gives

$$\begin{aligned}\langle n,\mathbf{k}|\hat{\mathbf{v}}|n',\mathbf{k}'\rangle &= \delta_{\mathbf{k}\mathbf{k}'}\int_\Gamma \mathrm{d}\mathbf{r}\, \overline{u_{n,\mathbf{k}}(\mathbf{r})}\hat{\mathbf{v}}(\mathbf{k})u_{n',\mathbf{k}}(\mathbf{r}) \\ &= \delta_{\mathbf{k}\mathbf{k}'}\langle u_{n,\mathbf{k}}|\hat{\mathbf{v}}(\mathbf{k})|u_{n',\mathbf{k}}\rangle,\end{aligned}$$

so Equation (6.21) reduces to

$$\sigma_{\mu\nu} = \frac{i\hbar q^2}{V}\sum_{n,n'}\sum_{\mathbf{k}}\frac{f(\epsilon_{n,\mathbf{k}})-f(\epsilon_{n',\mathbf{k}})}{(\epsilon_{n,\mathbf{k}}-\epsilon_{n',\mathbf{k}}-i\hbar\delta)(\epsilon_{n,\mathbf{k}}-\epsilon_{n',\mathbf{k}})}\langle u_{n,\mathbf{k}}|\hat{v}_\nu(\mathbf{k})|u_{n',\mathbf{k}}\rangle\langle u_{n',\mathbf{k}}|\hat{v}_\mu(\mathbf{k})|u_{n,\mathbf{k}}\rangle. \qquad (6.24)$$

As explained at the end of the previous section, the $n = n'$ terms remain finite due to some cancellations, but this breaks down when we take the limit $\delta\downarrow 0$. A more careful analysis which also takes disorder and impurity concentrations into account shows that the $n = n'$ terms are proportional to the mean scattering time; see for instance [15, Section 4.2]. Their sum is known as the *Drude conductivity* and it blows up in the clean limit – which we are working in here – essentially because a perfect metal has zero resistance. We therefore choose to ignore this "very stupid infinity" and focus instead on the $n\neq n'$ terms, which constitute the *interband conductivity*. In the continuum limit $L\to\infty$ with $\delta\downarrow 0$, it reads

$$\sigma_{\mu\nu} = \frac{i\hbar q^2}{(2\pi)^d}\sum_{n\neq n'}\int \mathrm{d}\mathbf{k}\,\frac{f(\epsilon_{n,\mathbf{k}})-f(\epsilon_{n',\mathbf{k}})}{(\epsilon_{n,\mathbf{k}}-\epsilon_{n',\mathbf{k}})^2}\langle u_{n,\mathbf{k}}|\hat{v}_\nu(\mathbf{k})|u_{n',\mathbf{k}}\rangle\langle u_{n',\mathbf{k}}|\hat{v}_\mu(\mathbf{k})|u_{n,\mathbf{k}}\rangle. \qquad (6.25)$$

In converting the sum over $\mathbf{k}$ to an integral, we divided by the volume per $\mathbf{k}$-point $(2\pi)^d/V$.

Now, note that for any $u\in\mathcal{H}'$ and wave vector $\mathbf{k}$,

$$\begin{aligned}(\boldsymbol{\nabla}_{\mathbf{k}}\hat{H}(\mathbf{k}))u = \boldsymbol{\nabla}_{\mathbf{k}}(e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{H}_0e^{i\mathbf{k}\cdot\mathbf{r}}u) &= -i\mathbf{r}e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{H}_0e^{i\mathbf{k}\cdot\mathbf{r}}u + e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{H}_0i\mathbf{r}e^{i\mathbf{k}\cdot\mathbf{r}}u \\ &= -ie^{-i\mathbf{k}\cdot\mathbf{r}}[\hat{\mathbf{r}},\hat{H}_0]e^{i\mathbf{k}\cdot\mathbf{r}}u = \hbar\hat{\mathbf{v}}(\mathbf{k})u\end{aligned}$$

by the product rule together with Equations (6.14) and (6.22), so

$$\langle u_{n,\mathbf{k}}|\hat{\mathbf{v}}(\mathbf{k})|u_{n',\mathbf{k}}\rangle = \frac{1}{\hbar}\langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}\hat{H}(\mathbf{k})|u_{n',\mathbf{k}}\rangle$$

for any pair of band indices n, n'. We want to use the product rule for differentiation to transfer the gradient to other parts of the inner product, but for that, we need the map assigning $|u_{n,\mathbf{k}}\rangle$ to every point $\mathbf{k}$ in the first Brillouin zone to be *smooth* for all band indices n. However, quantum mechanical eigenstates are defined up to phase, so we can *guarantee* that $|u_{n,\mathbf{k}}\rangle$ depends smoothly on $\mathbf{k}$ by choosing an appropriate phase for every Bloch function. There is an important caveat, though: for this to be possible, we need to assume that the bands are *isolated*, i.e. that $\epsilon_{n,\mathbf{k}} \neq \epsilon_{n+1,\mathbf{k}}$ for all n and $\mathbf{k}$. If two distinct bands cross or touch, which happens quite regularly at high symmetry points in the first Brillouin zone, singularities can occur. We will see an example of this in Chapter 7.

The Bloch states $|u_{n,\mathbf{k}}\rangle$ and $|u_{n',\mathbf{k}}\rangle$ have inner product $\delta_{nn'}$, so

$$\mathbf{0} = \boldsymbol{\nabla}_{\mathbf{k}}\langle u_{n,\mathbf{k}}|u_{n',\mathbf{k}}\rangle = \langle \boldsymbol{\nabla}_{\mathbf{k}}u_{n,\mathbf{k}}|u_{n',\mathbf{k}}\rangle + \langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}u_{n',\mathbf{k}}\rangle \tag{6.26}$$

and if $n \neq n'$,

$$\begin{aligned}\mathbf{0} &= \boldsymbol{\nabla}_{\mathbf{k}}\langle u_{n,\mathbf{k}}|\hat{H}(\mathbf{k})|u_{n',\mathbf{k}}\rangle \\ &= \langle \boldsymbol{\nabla}_{\mathbf{k}}u_{n,\mathbf{k}}|\hat{H}(\mathbf{k})|u_{n',\mathbf{k}}\rangle + \langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}\hat{H}(\mathbf{k})|u_{n',\mathbf{k}}\rangle + \langle u_{n,\mathbf{k}}|\hat{H}(\mathbf{k})|\boldsymbol{\nabla}_{\mathbf{k}}u_{n',\mathbf{k}}\rangle \\ &= (-\epsilon_{n',\mathbf{k}} + \epsilon_{n,\mathbf{k}})\langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}u_{n',\mathbf{k}}\rangle + \langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}\hat{H}(\mathbf{k})|u_{n',\mathbf{k}}\rangle,\end{aligned}$$

which means

$$\frac{\langle u_{n,\mathbf{k}}|\hat{\mathbf{v}}(\mathbf{k})|u_{n',\mathbf{k}}\rangle}{\epsilon_{n,\mathbf{k}} - \epsilon_{n',\mathbf{k}}} = -\frac{1}{\hbar}\langle u_{n,\mathbf{k}}|\boldsymbol{\nabla}_{\mathbf{k}}u_{n',\mathbf{k}}\rangle.$$

Substitution into Equation (6.25) now gives

$$\sigma_{\mu\nu} = -\frac{iq^2}{(2\pi)^d\hbar}\sum_{n,n'}\int \mathrm{d}\mathbf{k}\ \big(f(\epsilon_{n,\mathbf{k}}) - f(\epsilon_{n',\mathbf{k}})\big)\langle u_{n,\mathbf{k}}|\partial_\nu u_{n',\mathbf{k}}\rangle\langle u_{n',\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle,$$

where ∂_μ and ∂_ν denote partial derivatives with respect to k^μ and k^ν. We have also reintroduced the $n = n'$ terms because they now contribute nothing. Splitting the sum into two and using Equation (6.26) to again transfer $\mathbf{k}$-derivatives to the other side of inner products at the cost of a minus sign yields

$$\begin{aligned}\sigma_{\mu\nu} &= \frac{iq^2}{(2\pi)^d\hbar}\int \mathrm{d}\mathbf{k}\ \Bigg(\sum_{n,n'} f(\epsilon_{n,\mathbf{k}})\langle \partial_\nu u_{n,\mathbf{k}}|u_{n',\mathbf{k}}\rangle\langle u_{n',\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle - \\ &\qquad \sum_{n,n'} f(\epsilon_{n',\mathbf{k}})\langle \partial_\mu u_{n',\mathbf{k}}|u_{n,\mathbf{k}}\rangle\langle u_{n,\mathbf{k}}|\partial_\nu u_{n',\mathbf{k}}\rangle\Bigg) \\ &= \frac{iq^2}{(2\pi)^d\hbar}\sum_n\int \mathrm{d}\mathbf{k}\ f(\epsilon_{n,\mathbf{k}})\big(\langle \partial_\nu u_{n,\mathbf{k}}|\hat{O}(\mathbf{k})|\partial_\mu u_{n,\mathbf{k}}\rangle - \langle \partial_\mu u_{n,\mathbf{k}}|\hat{O}(\mathbf{k})|\partial_\nu u_{n,\mathbf{k}}\rangle\big),\end{aligned}$$

where we have introduced the projection operator

$$\hat{O}(\mathbf{k}) \coloneqq \sum_n |u_{n,\mathbf{k}}\rangle\langle u_{n,\mathbf{k}}|$$

on $\mathcal{H}'$ for all $\mathbf{k}$. Now, recall that for any fixed $\mathbf{k}$, the $|u_{n,\mathbf{k}}\rangle$ constitute an orthonormal basis of $\mathcal{H}'$. That means $\hat{O}(\mathbf{k})$ is the identity, so

$$\sigma_{\mu\nu} = -\frac{iq^2}{(2\pi)^d\hbar}\sum_n\int \mathrm{d}\mathbf{k}\ f(\epsilon_{n,\mathbf{k}})\big(\langle \partial_\mu u_{n,\mathbf{k}}|\partial_\nu u_{n,\mathbf{k}}\rangle - \langle \partial_\nu u_{n,\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle\big).$$

For all n, define the n-th *Berry connection* $\mathbf{a}^{(n)}$ and *Berry curvature* $\mathbf{f}^{(n)}$ as

$$a_\mu^{(n)}(\mathbf{k}) \coloneqq -i\langle u_{n,\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle, \tag{6.27}$$

$$f_{\mu\nu}^{(n)}(\mathbf{k}) \coloneqq \partial_\mu a_\nu^{(n)}(\mathbf{k}) - \partial_\nu a_\mu^{(n)}(\mathbf{k}). \tag{6.28}$$

Recall that earlier on, we had to choose a phase for each Bloch function to be able to take derivatives with respect to $\mathbf{k}$; a priori, $\mathbf{a}^{(n)}$ and $\mathbf{f}^{(n)}$ depend on that choice of phase. By the equality of mixed partials,

$$\begin{aligned} f_{\mu\nu}^{(n)}(\mathbf{k}) &= -i\big(\langle\partial_\mu u_{n,\mathbf{k}}|\partial_\nu u_{n,\mathbf{k}}\rangle + \langle u_{n,\mathbf{k}}|\partial_\mu\partial_\nu u_{n,\mathbf{k}}\rangle - \\ &\qquad \langle\partial_\nu u_{n,\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle - \langle u_{n,\mathbf{k}}|\partial_\nu\partial_\mu u_{n,\mathbf{k}}\rangle\big) \\ &= -i\big(\langle\partial_\mu u_{n,\mathbf{k}}|\partial_\nu u_{n,\mathbf{k}}\rangle - \langle\partial_\nu u_{n,\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle\big), \end{aligned}$$

with which we finally arrive at the *TKNN formula*

$$\boxed{\sigma_{\mu\nu} = \frac{q^2}{(2\pi)^d\hbar}\sum_n \int \mathrm{d}\mathbf{k}\; f(\epsilon_{n,\mathbf{k}}) f_{\mu\nu}^{(n)}(\mathbf{k}).} \tag{6.29}$$

Taking a step back, recall that $\sigma_{\mu\nu}$ by definition measures the strength of the current in the μ-direction induced by an electric field in the ν-direction. Thus, the *off-diagonal* elements $\sigma_{\mu\nu}$, $\mu \neq \nu$ of the conductivity matrix quantify the anomalous Hall effect. Equation (6.29) then states that there is a contribution to the AHE by each (partially) occupied band n, given by the integral over the Brillouin zone of the n-th Berry curvature $f_{\mu\nu}^{(n)}$ weighted by the Fermi-Dirac distribution. As we will see in the next chapter, the TKNN formula explains why the anomalous Hall effect can occur only in magnetic materials, and it provides a fascinating geometric viewpoint.

Chapter 7

The anomalous Hall effect

To those readers who have gotten lost in the sea of symbols and equations: do not fret, this is where the fun begins. In this final chapter, we discuss the geometric interpretation of the anomalous Hall effect, the puzzling ferromagnetic cousin of the ordinary Hall effect discovered in 1881 by Edwin Hall. One could say that it has three parts. Firstly, the functions $\mathbf{a}^{(n)}$ and $\mathbf{f}^{(n)}$ on the Brillouin zone which make an appearance in the TKNN formula (6.29) can be interpreted as the coordinate representations of the local forms of a connection ω and its curvature Ω on a principal $U(1)$-bundle $\pi\colon P \to T^d$ over the d-dimensional torus T^d. Interpreting T^d as the "folded up" Brillouin zone, the total space P of the bundle can be constructed by gluing to each wave vector $\mathbf{k} \in T^d$ the $U(1)$ of normalized $\hat{H}(\mathbf{k})$ eigenstates. Second, the notion of parallel transport on P that ω gives rise to has real physical significance: the horizontal lift to P of a smooth path γ in T^d describes the actual time evolution of the quantum system (as dictated by the Schrödinger equation) when the wave vector $\mathbf{k}$ (which can be interpreted as a parameter of the system) traverses the path γ. If γ is a loop in T^d, then the phase difference between the initial and final states of the system is known as a *Berry phase*. Finally, in two dimensions and in the low temperature limit, the contribution of a band to the TKNN conductivity σ_{xy} is a fundamental constant multiplied by precisely the Berry phase acquired by the system when walking around the edge of the occupied Brillouin zone for that particular band.

Interesting as it may be, however, this geometric interpretation still does not answer the question of why the anomalous Hall effect is measured only in ferromagnets; for that, we need to look at symmetries. It turns out that σ_{xy} can be nonvanishing only if *time reversal symmetry* is broken. This happens spontaneously in a ferromagnet: if it is magnetized, it can retain its nonzero magnetization even after any external fields have been turned off. Time reversal symmetry is usually broken in the spin sector, and *spin-orbit coupling* is needed to somehow "transfer" this information to the momentum $\mathbf{k}$. Finally, to conclude the chapter, we discuss the *Rashba model* for itinerant ferromagnetism, which nicely illustrates many of the concepts.

7.1 The bundle and its connection

The following ideas are based on [12, Section 10.6]. Consider a system of non-interacting electrons described by a Bloch Hamiltonian $\hat{H}_0$ in a crystal lattice $\Lambda \subseteq \mathbb{R}^d$. Recall from Equation (6.20) that we can then define a $\mathbf{k}$-dependent Hamiltonian $\hat{H}(\mathbf{k})$ acting on the space of Λ-periodic functions (which we now denote by $\mathcal{H}$) for all $\mathbf{k}$ in the first Brillouin zone, i.e. a primitive unit cell of the reciprocal lattice Λ^* of Λ. Diagonalizing $\hat{H}(\mathbf{k})$ yields a collection of eigenspaces, namely those spanned by the Bloch functions $u_{n,\mathbf{k}}$ with wave vector $\mathbf{k}$, along with the corresponding energy eigenvalues $\epsilon_{n,\mathbf{k}}$.

As before, we have to make the crucial assumption that all these eigenspaces have (complex) dimension 1, meaning that all energy bands in the system are *isolated*.[1] There may still very well be degeneracy within any band, just not between them: states with the same n but a different $\mathbf{k}$ can have the same energy eigenvalue, while states with the same $\mathbf{k}$ but a different n cannot. Under this assumption, we end up with a collection of functions – one for each band – from the first Brillouin zone to the set of 1-dimensional subspaces of $\mathcal{H}$, which can be canonically identified with the projective Hilbert space $\mathbb{P}(\mathcal{H})$. These descend to maps $F^{(n)}\colon T^d \to \mathbb{P}(\mathcal{H})$ with T^d the d-dimensional torus, since points on opposite edges of the Brillouin zone have identical spectra. Mathematically, $T^d = \mathbb{R}^d/\Lambda^*$ is the quotient of reciprocal space by the reciprocal lattice Λ^*, and the functions $\mathrm{BZ} \to \mathbb{P}(\mathcal{H})$ factor through the quotient map $q\colon \mathbb{R}^d \to T^d$, $\mathbf{k} \mapsto [\mathbf{k}]$. In some sense, the $F^{(n)}$ form the quantum mechanical description of the system, and they are precisely what we need to give the TKNN formula a geometric interpretation.

The finite-dimensional case. To see why, fix n and assume for the moment that the image of $F^{(n)}$ is contained in the projectivization of some finite-dimensional subspace of $\mathcal{H}$. For all intents and purposes, then, we can just assume $\mathcal{H}$ itself to be finite-dimensional, at least while we are looking at only the n-th band. Recall from Section 3.3 that the natural map $\gamma_n\colon S \to \mathbb{P}(\mathcal{H})$, $z \mapsto [z]$ with $S \subseteq \mathcal{H}$ the unit sphere is then a principal $U(1)$-bundle, and from Section 5.2 that there is a canonical connection form $\langle j|\mathrm{d}j\rangle \in \Omega^1(S, i\mathbb{R})$ on γ_n, where $j\colon S \hookrightarrow \mathcal{H}$ is the inclusion. Hence, under the assumption that $F^{(n)}$ is smooth (which is a well-defined notion now), we can do two things:

(i) pull back γ_n along $F^{(n)}$ to obtain a principal $U(1)$-bundle $\pi\colon P \coloneqq (F^{(n)})^*S \to T^d$ and a principal $U(1)$-bundle morphism $\zeta\colon P \to S$, which is just the projection onto the second coordinate (see Proposition 3.13);

(ii) pull back $\langle j|\mathrm{d}j\rangle$ along ζ to obtain a connection form $\omega \coloneqq \zeta^*\langle j|\mathrm{d}j\rangle$ on π (see Proposition 5.3).

One way to summarize this is to say that we are "pulling back the mathematics along the physics".

By definition of the pullback bundle, the fiber of π over $[\mathbf{k}] \in T^d$ consists precisely of those points $([\mathbf{k}], |u\rangle)$ with $|u\rangle \in S$ an eigenstate of $\hat{H}(\mathbf{k})$ with eigenvalue $\epsilon_{n,\mathbf{k}}$. Intuitively, P is just the space we obtain by gluing to each point $[\mathbf{k}] \in T^d$ on the folded up Brillouin zone the $U(1)$ of normalized Bloch functions with wave vector $\mathbf{k}$ in band n. As for the pullback connection ω, let us try to compute its local forms in coordinates. Let $V \subseteq \mathbb{R}^d$ be the Brillouin zone without its boundary and set $U \coloneqq q(V) \subseteq T^d$, then U is open in T^d. The restriction $q|_V$ is a diffeomorphism onto U and its inverse $\psi \coloneqq (q|_V)^{-1}\colon U \to V$ is a smooth chart for T^d. Assume U admits a smooth local section $\sigma\colon U \to \pi^{-1}U$ of π,[2] then the composition $\tau \coloneqq j \circ \zeta \circ \sigma\colon U \to \mathcal{H}$ sends $[\mathbf{k}] \in U$ to one such normalized $\hat{H}(\mathbf{k})$ eigenstate, which we can denote by $|u_{n,\mathbf{k}}\rangle \coloneqq \tau([\mathbf{k}]) = (\tau \circ \psi^{-1})(\mathbf{k})$ for all $\mathbf{k} \in V$. The function τ is essentially what was referred to earlier as a "smooth choice of phase" on the open Brillouin zone. The local form of ω on U can now be written as

$$\mathcal{A} \coloneqq \sigma^*\omega = \langle(\zeta \circ \sigma)^* j|\mathrm{d}((\zeta \circ \sigma)^* j)\rangle = \langle \tau|\mathrm{d}\tau\rangle \in \Omega^1(U, i\mathbb{R}), \tag{7.1}$$

since pullbacks commute with products and exterior derivatives (see Lemma 4.10). Similarly, pulling $\mathcal{A}$ back along $\psi^{-1} = q|_V\colon V \to U$ yields a smooth $i\mathbb{R}$-valued 1-form

$$A \coloneqq (\psi^{-1})^*\mathcal{A} = \langle(\psi^{-1})^*\tau|\mathrm{d}((\psi^{-1})^*\tau)\rangle = \langle \tau \circ \psi^{-1}|\mathrm{d}(\tau \circ \psi^{-1})\rangle \in \Omega^1(V, i\mathbb{R})$$

on V. Now, since $\tau \circ \psi^{-1}$ is just a map from an open subset of $\mathbb{R}^d$ to $\mathcal{H}$,

$$\mathrm{d}(\tau \circ \psi^{-1}) = \partial_\mu(\tau \circ \psi^{-1})\,\mathrm{d}k^\mu = |\partial_\mu u_{n,\mathbf{k}}\rangle\,\mathrm{d}k^\mu$$

[1] This assumption is not strictly necessary for the construction to work. For example, if the system is invariant under both time-reversal and spatial inversion, each band is doubly degenerate due to the spin of the electrons. However, the spin-up and spin-down populations can be regarded as their own separate subsystems, with isolated and non-singular bands. We still make the assumption for the sake of clarity and ease of language.

[2] This assumption is actually redundant, since U is contractible and any principal bundle over a contractible base space admits a global section.

with $k^1, \dots, k^d \colon \mathbb{R}^d \to \mathbb{R}$ the coordinate projections, so

$$\boxed{A = \langle u_{n,\mathbf{k}} | \partial_\mu u_{n,\mathbf{k}} \rangle \, \mathrm{d}k^\mu.} \tag{7.2}$$

By Equation (6.27), the component functions $A_\mu \colon V \to i\mathbb{R}$ of A satisfy $A_\mu = a_\mu^{(n)} i$, with $\mathbf{a}^{(n)}$ the Berry connection of the n-th band. Now, let $\Omega := \mathrm{d}\omega \in \Omega^2(P, i\mathbb{R})$ be the curvature of ω (the Lie bracket term vanishes because $U(1)$ is abelian) and $\mathcal{F} := \sigma^*\Omega$ its corresponding local form, then $\mathcal{F} = \mathrm{d}\mathcal{A}$ since pullbacks commute with exterior derivatives. The coordinate representation of $\mathcal{F}$ reads

$$F := (\psi^{-1})^*\mathcal{F} = \mathrm{d}A = \partial_\mu A_\nu \, \mathrm{d}k^\mu \wedge \mathrm{d}k^\nu = \frac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu) \, \mathrm{d}k^\mu \wedge \mathrm{d}k^\nu$$

where the factor $1/2$ compensates for the fact that the implicit summations runs over each unordered pair of distinct indices twice, so for the component functions $F_{\mu\nu} \colon V \to i\mathbb{R}$, we have $F_{\mu\nu} = f_{\mu\nu}^{(n)} i$ with $\mathbf{f}^{(n)}$ the n-th Berry curvature. This is the essence of the geometric interpretation of the TKNN formula:

> *The Berry connection $\mathbf{a}^{(n)}$ and curvature $\mathbf{f}^{(n)}$ are the coordinate representations of the local forms of a connection ω and its curvature $\Omega = \mathrm{d}\omega$ on a principal $U(1)$-bundle π over the folded up Brillouin zone T^d.*

The infinite-dimensional case. Things get a little more messy if the image of $F^{(n)}$ is *not* contained in $\mathbb{P}(\mathcal{H}')$ for some finite-dimensional subspace $\mathcal{H}' \subseteq \mathcal{H}$. In this case, $\mathbb{P}(\mathcal{H})$ is no longer locally Euclidean and a more general theory of infinite-dimensional differentiable manifolds is needed to be able to talk about smoothness of maps and bundles. What we *can* do, however, is assume that $F^{(n)}$ is at least continuous and pull back the topological bundle $\gamma \colon S \to \mathbb{P}(\mathcal{H})$ from Proposition 3.17 along $F^{(n)}$. This yields a topological principal $U(1)$-bundle $\pi \colon P \to T^d$ along with a topological principal $U(1)$-bundle morphism $\zeta \colon P \to S$. By virtue of Theorem 3.16, P can then be given a compatible smooth structure such that π becomes a smooth bundle.

To construct a connection on π, we now argue backwards. Let $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$ be a bundle atlas for π with associated smooth local sections $\sigma_\alpha \colon U_\alpha \to \pi^{-1}U_\alpha$. Again, denote by $j \colon S \hookrightarrow \mathcal{H}$ the inclusion of the unit sphere $S \subseteq \mathcal{H}$ in $\mathcal{H}$ and set $\tau_\alpha := j \circ \zeta \circ \sigma_\alpha \colon U_\alpha \to \mathcal{H}$ for all $\alpha \in A$. Choose an orthonormal basis $\{e_\ell\}_{\ell \in I}$ for $\mathcal{H}$, then for all $\alpha \in A$, there exist functions $\tau_\alpha^\ell \colon U_\alpha \to \mathbb{C}$ such that $\tau_\alpha = \tau_\alpha^\ell e_\ell$. Assuming these to be smooth, we can use the product of $\mathbb{C}$-valued forms induced by complex multiplication to define

$$\mathcal{A}_\alpha^\ell := \overline{\tau_\alpha^\ell} \, \mathrm{d}(\tau_\alpha^\ell) \in \Omega^1(U_\alpha, \mathbb{C})$$

for all $\alpha \in A$ and $\ell \in I$. Finally, assume the sum

$$\mathcal{A}_\alpha := \sum_{\ell \in I} \mathcal{A}_\alpha^\ell$$

to be a well-defined smooth $\mathbb{C}$-valued 1-form on U_α for all $\alpha \in A$. These are our local connection form candidates, inspired by Equation (7.1). It is easy to see that they are in fact $i\mathbb{R}$-valued: for all $[\mathbf{k}] \in U_\alpha$ and $v \in T_{[\mathbf{k}]}T^d$,

$$\begin{aligned}
(\mathcal{A}_\alpha)_{[\mathbf{k}]}(v) &= \sum_{\ell \in I} (\mathrm{d}(\overline{\tau_\alpha^\ell}\tau_\alpha^\ell)_{[\mathbf{k}]}(v) - \mathrm{d}(\overline{\tau_\alpha^\ell})_{[\mathbf{k}]}(v)\tau_\alpha^\ell([\mathbf{k}])) \\
&= \mathrm{d}\left(\sum_{\ell \in I} |\tau_\alpha^\ell|^2\right)_{[\mathbf{k}]}(v) - \sum_{\ell \in I} \overline{\overline{\tau_\alpha^\ell([\mathbf{k}])} \, \mathrm{d}(\tau_\alpha^\ell)_{[\mathbf{k}]}(v)} \\
&= \mathrm{d}(|\tau_\alpha|^2)_{[\mathbf{k}]}(v) - \overline{(\mathcal{A}_\alpha)_{[\mathbf{k}]}(v)} = -\overline{(\mathcal{A}_\alpha)_{[\mathbf{k}]}(v)}
\end{aligned}$$

since the exterior derivative is an antiderivation and τ_α by definition takes values in the unit sphere S, so that $|\tau_\alpha|^2$ is a constant map. Now, denote by $\rho_{\alpha\beta}$ the transition functions relative to the chosen bundle atlas and let $\alpha, \beta \in A$. The topological fiber bundle morphism $\zeta\colon P \to S$ is $U(1)$-equivariant and we know $\sigma_\beta = \sigma_\alpha \cdot \rho_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, so also $\tau_\beta = \tau_\alpha \cdot \rho_{\alpha\beta}$ and thus $\tau_\beta^\ell = \tau_\alpha^\ell \cdot \rho_{\alpha\beta}$ for all $\ell \in I$. Again using that d is an antiderivation gives

$$\begin{aligned}\mathcal{A}_\beta &= \sum_{\ell\in I} \overline{\tau_\alpha^\ell \rho_{\alpha\beta}}(\mathrm{d}(\tau_\alpha^\ell)\rho_{\alpha\beta} + \tau_\alpha^\ell \,\mathrm{d}(\rho_{\alpha\beta}))\\ &= \sum_{\ell\in I} \overline{\tau_\alpha^\ell}\,\mathrm{d}(\tau_\alpha^\ell)|\rho_{\alpha\beta}|^2 + \Bigl(\sum_{\ell\in I}|\tau_\alpha^\ell|^2\Bigr)\overline{\rho_{\alpha\beta}}\,\mathrm{d}(\rho_{\alpha\beta})\\ &= \mathcal{A}_\alpha + \rho_{\alpha\beta}^{-1}\,\mathrm{d}(\rho_{\alpha\beta}) = \mathcal{A}_\alpha + \mathrm{d}\log(\rho_{\alpha\beta}),\end{aligned}$$

where Lemma 4.25 was used in the last equation. Hence, by Proposition 5.9, there exists a connection form $\omega \in \Omega^1(P, i\mathbb{R})$ on π such that $\mathcal{A}_\alpha = \sigma_\alpha^*\omega$ for all $\alpha \in A$. Note that the $\mathcal{A}_\alpha$ were defined in such a way that pushing them forward along smooth coordinate chart maps $\psi_\alpha\colon U_\alpha \to \mathbb{R}^d$ gives the same results as in the finite-dimensional case, so from this point on, the two cases can again be treated as one.

7.2 Berry's phase

With the connection form ω on the principal $U(1)$-bundle $\pi\colon P \to T^d$ comes a notion of parallel transport on P, as was discussed in Section 5.5. Remarkably, the horizontal lifts of smooth paths in the folded up Brillouin zone T^d have real physical significance: they represent the actual time evolution of quantum states as dictated by the Schrödinger equation. We follow [17, Section 3.3].

As above, let V be the open Brillouin zone and $U \subseteq T^d$ its image under the quotient map $q\colon \mathbb{R}^d \to T^d$, so that $\psi \coloneqq (q|_V)^{-1}\colon U \to V$ is a smooth coordinate map. Choose a smooth local section $\sigma\colon U \to \pi^{-1}U$ of π and write $|u_{n,\mathbf{k}}\rangle \coloneqq \zeta(\sigma([\mathbf{k}])) \in S$ for all $\mathbf{k} \in V$; σ can be interpreted as a smooth choice of phase on U. Again, we denote by $\mathcal{A} \coloneqq \sigma^*\omega$ and $\mathcal{F} \coloneqq \sigma^*\Omega$ the corresponding local forms of the connection ω and its curvature $\Omega = \mathrm{d}\omega$, and by $A \coloneqq (\psi^{-1})^*\mathcal{A}$ and $F \coloneqq (\psi^{-1})^*\mathcal{F}$ the coordinate representations of $\mathcal{A}$ and $\mathcal{F}$. Now comes a bit of a conceptual leap: we have to start viewing the wave vector $\mathbf{k}$ as a parameter of the system that can be tuned by turning an imaginary knob. The system is described by the $\mathbf{k}$-dependent Hamiltonian $\hat{H}(\mathbf{k})$, and its parameter space is T^d. If $\gamma\colon t \mapsto \mathbf{k}(t)$ is a smooth curve in V, then its composition $\psi^{-1} \circ \gamma\colon t \mapsto [\mathbf{k}(t)]$ with ψ^{-1} describes a possible evolution of the parameter through time. Suppose the system is prepared at time $t = 0$ in the state $|\psi(0)\rangle = \bigl|u_{n,\mathbf{k}(0)}\bigr\rangle$, then its state $|\psi(t)\rangle$ at time $t > 0$ can be found by solving the Schrödinger equation

$$\bigl(i\hbar\partial_t - \hat{H}(\mathbf{k}(t))\bigr)|\psi(t)\rangle = 0. \tag{7.3}$$

If the evolution of the parameter $\mathbf{k}(t)$ is "slow enough",[3] then the adiabatic theorem tells us that to good approximation, $|\psi(t)\rangle = e^{i\theta(t)}\bigl|u_{n,\mathbf{k}(t)}\bigr\rangle$ for some phase function θ. That is, if the system changes in time only very gradually, it will stay in the n-th band and barely mix with other eigenstates. If γ were the *constant* map $\mathbf{k}(t) \equiv \mathbf{k}$, then the system would evolve as $|\psi(t)\rangle = e^{-i\epsilon_{n,\mathbf{k}}t/\hbar}|u_{n,\mathbf{k}}\rangle$ by standard quantum mechanics: $|u_{n,\mathbf{k}}\rangle$ is an eigenstate of $\hat{H}(\mathbf{k})$ with eigenvalue $\epsilon_{n,\mathbf{k}}$, so all we have to do is tack on its characteristic wiggle factor. For general γ, this motivates the ansatz $|\psi(t)\rangle = c(t)e^{i\theta(t)}\bigl|u_{n,\mathbf{k}(t)}\bigr\rangle$ with

$$\theta(t) \coloneqq -\frac{1}{\hbar}\int_0^t \epsilon_{n,\mathbf{k}(t')}\,\mathrm{d}t'.$$

Substitution into the Schrodinger equation (7.3) yields

$$\begin{aligned}0 &= c'(t)e^{i\theta(t)}\bigl|u_{n,\mathbf{k}(t)}\bigr\rangle + i\theta'(t)|\psi(t)\rangle + c(t)e^{i\theta(t)}\bigl|\partial_t u_{n,\mathbf{k}(t)}\bigr\rangle - \frac{1}{i\hbar}\hat{H}(\mathbf{k}(t))|\psi(t)\rangle\\ &= c'(t)e^{i\theta(t)}\bigl|u_{n,\mathbf{k}(t)}\bigr\rangle + \frac{1}{i\hbar}\epsilon_{n,\mathbf{k}(t)}|\psi(t)\rangle + c(t)e^{i\theta(t)}\bigl|\partial_t u_{n,\mathbf{k}(t)}\bigr\rangle - \frac{1}{i\hbar}\epsilon_{n,\mathbf{k}(t)}|\psi(t)\rangle\\ &= \bigl(c'(t)\bigl|u_{n,\mathbf{k}(t)}\bigr\rangle + c(t)\bigl|\partial_t u_{n,\mathbf{k}(t)}\bigr\rangle\bigr)e^{i\theta(t)},\end{aligned}$$

[3] The parameter must evolve slowly relative to $\Delta E/\hbar$ with ΔE a typical band gap size. It is possible to precisely quantify this, see [17, Section 3.3] and [18, Section 2.1].

which implies

$$c'(t) = -c(t)\langle u_{n,\mathbf{k}(t)}|\partial_t u_{n,\mathbf{k}(t)}\rangle.$$

With the initial condition $c(0) = 1$, we thus find the solution

$$c(t) = \exp\biggl(-\int_0^t \langle u_{n,\mathbf{k}(t')}|\partial_t u_{n,\mathbf{k}(t')}\rangle \,\mathrm{d}t'\biggr).$$

This can be plugged back into the ansatz to obtain

$$\begin{aligned}|\psi(t)\rangle &= e^{i\theta(t)}\big|u_{n,\mathbf{k}(t)}\big\rangle \exp\biggl(-\int_0^t \langle u_{n,\mathbf{k}(t')}|\partial_\mu u_{n,\mathbf{k}(t')}\rangle \frac{\mathrm{d}k^\mu}{\mathrm{d}t}(t')\,\mathrm{d}t'\biggr)\\ &= e^{i\theta(t)}\big|u_{n,\mathbf{k}(t)}\big\rangle \exp\biggl(-\int_0^t \langle u_{n,\mathbf{k}(t')}|\partial_\mu u_{n,\mathbf{k}(t')}\rangle \,\mathrm{d}(k^\mu\circ\gamma)\biggr)\\ &= e^{i\theta(t)}\big|u_{n,\mathbf{k}(t)}\big\rangle \exp\biggl(-\int_0^t \gamma^*(\langle u_{n,\mathbf{k}}|\partial_\mu u_{n,\mathbf{k}}\rangle\,\mathrm{d}k^\mu)\biggr)\\ &= e^{i\theta(t)}\big|u_{n,\mathbf{k}(t)}\big\rangle \exp\biggl(-\int_0^t \gamma^* A\biggr),\end{aligned}$$

where we have used the expression (7.2) for $A = (\psi^{-1})^*\mathcal{A}$. By writing $|\psi(t)\rangle$ in this form, the connection (pun intended) with horizontal path lifting becomes clear. Theorem 5.15 states that the smooth path $\psi^{-1}\circ\gamma\colon t\mapsto[\mathbf{k}(t)]$ in T^d has a unique lift $\tilde\gamma$ to P with initial point $\sigma([\mathbf{k}(0)])$ which is horizontal with respect to the connection ω on π. By Proposition 5.16, it is given by

$$(\zeta\circ\tilde\gamma)(t) = \zeta(\sigma([\mathbf{k}(t)]))\exp\biggl(-\int_0^t(\psi^{-1}\circ\gamma)^*\mathcal{A}\biggr) = \big|u_{n,\mathbf{k}(t)}\big\rangle\exp\biggl(-\int_0^t\gamma^*A\biggr),$$

which is precisely $|\psi(t)\rangle$ if the dynamical phase $\theta(t)$ is left out. This brings us to the following conclusion:

> *The connection ω on the principal $U(1)$-bundle π over T^d describes a physical geometric phase acquired by the system's wave function when traversing a path in parameter space.*

In this context, such a geometric phase is also known as a *Berry phase*. In $d = 2$ dimensions at low temperatures, it allows for taking the geometric interpretation of the TKNN formula a step further. In the limit $T\downarrow 0$, the Fermi-Dirac distribution $f(\epsilon)$ converges to a step function which cuts off from 1 to 0 at the chemical potential $\epsilon = \mu$. The occupied states in the n-th band are thus precisely those with a wave vector lying in the subset

$$\Sigma \coloneqq \{\mathbf{k} : \epsilon_{n,\mathbf{k}} < \mu\} \subseteq \mathbb{R}^2$$

of the first Brillouin zone. By Equation (6.29) then, the contribution of the n-th band to the anomalous Hall conductivity σ_{xy} is

$$\sigma_{xy}^{(n)} = \frac{e^2}{2\pi h}\int_\Sigma f_{xy}^{(n)}(\mathbf{k})\,\mathrm{d}k^x\,\mathrm{d}k^y = -\frac{ie^2}{2\pi h}\int_\Sigma F_{xy}(\mathbf{k})\,\mathrm{d}k^x\,\mathrm{d}k^y = -\frac{ie^2}{2\pi h}\int_\Sigma F,$$

where we used that $F = F_{xy}\,\mathrm{d}k^x\wedge\mathrm{d}k^y$ since we are working in 2D. Assume that the closure $\overline{\Sigma}$ of Σ is contained in the open Brillouin zone V, and that $\gamma\colon[0,1]\to V$ is a smooth closed curve traversing the boundary of Σ once in the positive direction, then by Stokes' theorem,

$$\sigma_{xy}^{(n)} = -\frac{ie^2}{2\pi h}\int_\Sigma \mathrm{d}A = -\frac{ie^2}{2\pi h}\int_{\partial\Sigma}A = -\frac{ie^2}{2\pi h}\int_0^1\gamma^*A.$$

From the above discussion, we also know that when walking around the closed loop $\psi^{-1}\circ\gamma$ in parameter space, the initial and final states of the system differ by a geometric phase factor $e^{i\phi}$ with

$$\phi = i\int_0^1(\psi^{-1}\circ\gamma)^*\mathcal{A} = i\int_0^1\gamma^*A,$$

which implies

$$\boxed{\sigma_{xy}^{(n)} = -\frac{e^2}{h} \cdot \frac{\phi}{2\pi}.}$$

In words:

> *In two dimensions in the low temperature limit, the contribution of the n-th band to the anomalous Hall conductivity σ_{xy} is the universal constant $-e^2/h$ multiplied the Berry phase in units of 2π acquired by the wave function when traversing the boundary of the occupied part of the Brillouin zone in reciprocal space.*

Something special happens when the entire n-th band is filled, so that

$$\sigma_{xy}^{(n)} = -\frac{ie^2}{2\pi h}\int_V F = -\frac{ie^2}{2\pi h}\int_{T^2} \mathcal{F}.$$

It is important to note here that $\mathcal{F} = \sigma^*\Omega$ is really only defined on the dense subset U of T^2, but by Proposition 5.13, it agrees with other local forms of Ω on intersections since $U(1)$ is abelian. Therefore, it can be extended to a smooth $i\mathbb{R}$-valued 2-form on the whole space T^2, which by abuse of notation is also called F. Now, by Shiing-Shen Chern's generalization of the famous Gauss-Bonnet theorem, the integral of $i\mathcal{F}/2\pi \in \Omega^2(T^2, \mathbb{R})$ over the torus T^2 is an *integer*, known as the *Chern number*, which implies that the Hall conductivity σ_{xy} is quantized [19].

This fits into the broader theory of characteristic classes as follows. Since $\mathcal{F}$ is just a bunch of exterior derivatives of local connection forms patched together, it is closed, i.e. $\mathrm{d}\mathcal{F} = 0$. As a consequence, the $\mathbb{R}$-valued 2-form $i\mathcal{F}/2\pi$ on T^2 defines a de Rham cohomology class $[i\mathcal{F}/2\pi] \in H^2(T^2)$, which can be shown to be independent of the connection ω. That is, if you follow the same procedure starting with a different connection on the principal $U(1)$-bundle π, you end up with something which differs from $i\mathcal{F}/2\pi$ by an exact form. We call $[i\mathcal{F}/2\pi]$ the *Chern class* of the bundle and under the identification of de Rham and singular cohomology, it can be shown to have integer coefficients. Furthermore, homotopic maps $F^{(n)} \colon T^2 \to \mathbb{P}(\mathcal{H})$ give isomorphic $U(1)$-bundles by Theorem 3.15 (at least in the finite-dimensional case), which means that the Chern classes and numbers are invariant under small perturbations of the Hamiltonian: the quantization of σ_{xy} is *topologically protected.* For details, see [9, §32], [20, Chapter C].

7.3 Symmetry considerations

One question that remains unanswered is why the anomalous Hall effect is measured only in ferromagnetic materials. It can be answered at least to some extent by considering the symmetries of the system. In quantum mechanics, a symmetry is represented by a (conjugate) linear operator $\hat{S}$ acting on the system's Hilbert space. The system is said to possess the symmetry if its Hamiltonian $\hat{H}$ commutes with $\hat{S}$, i.e. $[\hat{H}, \hat{S}] = 0$. For example, the operator $\hat{T}$ associated to translational symmetry over some vector $\mathbf{a}$ shifts any wave function $\psi(\mathbf{r})$ by $\mathbf{a}$, so $\hat{T}\psi(\mathbf{r}) = \psi(\mathbf{r} - \mathbf{a})$. For a Hamiltonian of the form

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}),$$

having translational symmetry then means $U(\mathbf{r} - \mathbf{a}) = U(\mathbf{r})$ for all $\mathbf{r}$.

For us, two symmetry operators are relevant. The first is the *parity* or *inversion* operator $\hat{\mathcal{I}}$, defined in the position basis by $\hat{\mathcal{I}}\psi(\mathbf{r}) = \psi(-\mathbf{r})$. It is linear, Hermitian ($\hat{\mathcal{I}}^* = \hat{\mathcal{I}}$) and unitary ($\hat{\mathcal{I}}^* = \hat{\mathcal{I}}^{-1}$), which implies that its only possible eigenvalues are $+1$ and -1. The corresponding eigenstates are precisely the wave functions which are even and odd in $\mathbf{r}$. The second is the *time reversal* operator $\hat{\mathcal{T}}$, defined for spin 1/2 particles as $\hat{\mathcal{T}} = i\hat{\sigma}_y\hat{K}$ with $\hat{\sigma}_y$ the second Pauli spin matrix and $\hat{K}$ the position basis complex conjugation operator. It is conjugate linear and antiunitary ($\hat{\mathcal{T}}^* = \hat{\mathcal{T}}^{-1}$),[4] which means all its eigenvalues have norm 1, and it satisfies $\hat{\mathcal{T}}^2 = -1$; see [17, Section 2.1.6].

[4] For a conjugate linear operator $\hat{A}$ on a Hilbert space $\mathcal{H}$, the adjoint $\hat{A}^*$ is defined by $\langle \hat{A}x|y\rangle = \overline{\langle x|\hat{A}^*y\rangle}$ for all $x, y \in \mathcal{H}$.

The associated symmetries have important consequences, also explained in [17, Section 2.1.6]. If a system has *inversion symmetry*, then we always can find an orthonormal basis of energy eigenstates which are all either even or odd in $\mathbf{r}$. *Time reversal symmetry*, on the other hand, implies that every energy eigenspace is at least twofold degenerate. To see why, let $|\psi\rangle$ be a normalized eigenstate of the Hamiltonian $\hat{H}$ with eigenvalue ϵ, then also $\hat{H}\hat{\mathcal{T}}|\psi\rangle = \epsilon|\psi\rangle$ since $\hat{H}$ and $\hat{\mathcal{T}}$ commute and ϵ is real. Suppose now that this does *not* imply degeneracy, i.e. that $\hat{\mathcal{T}}|\psi\rangle = \lambda|\psi\rangle$ for some $\lambda \in \mathbb{C}$, then

$$|\lambda|^2 = |\lambda|^2\langle\psi|\psi\rangle = \langle\hat{\mathcal{T}}\psi|\hat{\mathcal{T}}\psi\rangle = \overline{\langle\psi|\hat{\mathcal{T}}^*\hat{\mathcal{T}}|\psi\rangle} = \overline{\langle\psi|\psi\rangle} = 1$$

since $\hat{\mathcal{T}}$ is antiunitary, so

$$-|\psi\rangle = \hat{\mathcal{T}}^2|\psi\rangle = \hat{\mathcal{T}}\lambda|\psi\rangle = \overline{\lambda}\hat{\mathcal{T}}|\psi\rangle = |\lambda|^2|\psi\rangle = |\psi\rangle,$$

which is clearly a contradiction. It follows that $|\psi\rangle$ and $\hat{\mathcal{T}}|\psi\rangle$ are linearly independent eigenstates of $\hat{H}$ with the same eigenvalue E, and $|\psi\rangle$ was arbitrary. This is known as *Kramers degeneracy*.

Let us now return to our system of non-interacting electrons in a crystalline solid described by a Bloch Hamiltonian $\hat{H}_0$. Under inversion symmetry,

$$\hat{\mathcal{I}}\psi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}}e^{-i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(-\mathbf{r}) = \frac{1}{\sqrt{N}}e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{\mathcal{I}}u_{n,\mathbf{k}}(\mathbf{r})$$

is again an eigenstate of $\hat{H}_0$ with eigenvalue $\epsilon_{n,\mathbf{k}}$, for any wave vector $\mathbf{k}$ and band index n. It has wave vector $-\mathbf{k}$, so $\hat{\mathcal{I}}|u_{n,\mathbf{k}}\rangle = \lambda|u_{n,-\mathbf{k}}\rangle$ for some ($\mathbf{k}$-dependent) $\lambda \in U(1)$ and $\epsilon_{n,\mathbf{k}} = \epsilon_{n,-\mathbf{k}}$. Similarly, under time reversal symmetry,

$$\hat{\mathcal{T}}\psi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}}e^{-i\mathbf{k}\cdot\mathbf{r}}\hat{\mathcal{T}}u_{n,\mathbf{k}}(\mathbf{r})$$

implies $\hat{\mathcal{T}}|u_{n,\mathbf{k}}\rangle = \lambda|u_{n,-\mathbf{k}}\rangle$ and $\epsilon_{n,\mathbf{k}} = \epsilon_{n,-\mathbf{k}}$ for all $\mathbf{k}$ and n. These two are linearly independent by the Kramers argument, so if $\mathbf{k}$ and $-\mathbf{k}$ differ by a reciprocal lattice vector (i.e. if $2\mathbf{k} \in \Lambda^*$), we find that all eigenspaces of $\hat{H}(\mathbf{k})$ are at least doubly degenerate. To reiterate, if time reversal symmetry is present, all $\hat{H}_0$ eigenstates are degenerated, but this need not be *fixed* $\mathbf{k}$ degeneracy: the partner of any state could lie at a different wave vector in the same band. However, for all $\mathbf{k} \in \Lambda^*/2$, the degenerated partner must have the same wave vector, which means that two bands cross or touch. These special $\mathbf{k}$ are known as the *time-reversal invariant momenta* or TRIM for short, and there are precisely 2^d of them in d dimensions; see Figure 7.1 for an illustration. If the system is invariant under both inversion *and* time reversal, then the product $\hat{\mathcal{I}}\hat{\mathcal{T}}$ is a conjugate linear and antiunitary operator mapping $|u_{n,\mathbf{k}}\rangle$ to itself, up to phase. Thus, by the Kramers argument, the eigenspaces of $\hat{H}(\mathbf{k})$ are at least doubly degenerate for *all* $\mathbf{k}$, meaning that each band has "multiplicity" ≥ 2.

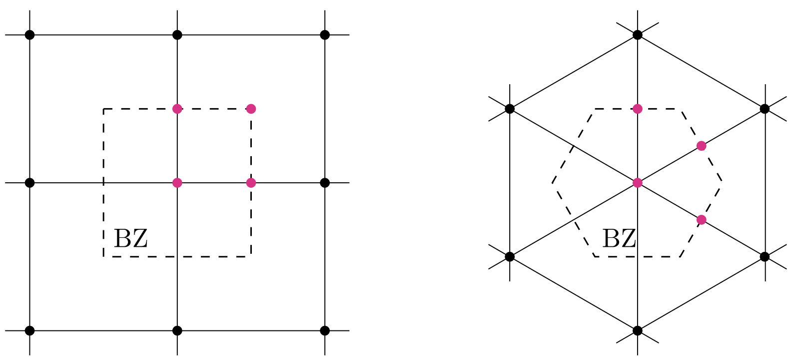


Figure 7.1: Illustration of the time-reversal invariant momenta (marked with pink dots) in the Brillouin zone of a square lattice (left) and triangular lattice (right)

Now, recall that the components of the n-th Berry curvature are given by

$$f_{\mu\nu}^{(n)}(\mathbf{k}) = -i\big(\langle\partial_\mu u_{n,\mathbf{k}}|\partial_\nu u_{n,\mathbf{k}}\rangle - \overline{\langle\partial_\mu u_{n,\mathbf{k}}|\partial_\nu u_{n,\mathbf{k}}\rangle}\big),$$

where we now explicitly write $|u_{n,\mathbf{k}}\rangle$ as a 2-component spinor

$$|u_{n,\mathbf{k}}\rangle = \begin{pmatrix} |u_{n,\mathbf{k}}^{\uparrow}\rangle \\ |u_{n,\mathbf{k}}^{\downarrow}\rangle \end{pmatrix}.$$

Applying the time-reversal operator $\hat{\mathcal{T}}$ gives

$$\hat{\mathcal{T}}|u_{n,\mathbf{k}}\rangle = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \overline{|u_{n,\mathbf{k}}^{\uparrow}\rangle} \\ \overline{|u_{n,\mathbf{k}}^{\downarrow}\rangle} \end{pmatrix} = \begin{pmatrix} \overline{|u_{n,\mathbf{k}}^{\downarrow}\rangle} \\ -\overline{|u_{n,\mathbf{k}}^{\uparrow}\rangle} \end{pmatrix},$$

so under time-reversal symmetry,

$$\begin{aligned} if_{\mu\nu}^{(n)}(-\mathbf{k}) &= \langle\partial_\mu\hat{\mathcal{T}}u_{n,\mathbf{k}}|\partial_\nu\hat{\mathcal{T}}u_{n,\mathbf{k}}\rangle - \overline{\langle\partial_\mu\hat{\mathcal{T}}u_{n,\mathbf{k}}|\partial_\nu\hat{\mathcal{T}}u_{n,\mathbf{k}}\rangle} \\ &= \overline{\langle\partial_\mu u_{n,\mathbf{k}}^{\downarrow}|\partial_\nu u_{n,\mathbf{k}}^{\downarrow}\rangle} + \overline{\langle\partial_\mu u_{n,\mathbf{k}}^{\uparrow}|\partial_\nu u_{n,\mathbf{k}}^{\uparrow}\rangle} - \langle\partial_\mu u_{n,\mathbf{k}}^{\downarrow}|\partial_\nu u_{n,\mathbf{k}}^{\downarrow}\rangle - \langle\partial_\mu u_{n,\mathbf{k}}^{\uparrow}|\partial_\nu u_{n,\mathbf{k}}^{\uparrow}\rangle \\ &= -if_{\mu\nu}^{(n)}(\mathbf{k}). \end{aligned}$$

We also know $\epsilon_{n,\mathbf{k}} = \epsilon_{n,-\mathbf{k}}$ for all n and $\mathbf{k}$, so $f(\epsilon_{n,\mathbf{k}})f_{\mu\nu}^{(n)}(\mathbf{k})$ is odd in $\mathbf{k}$ and the TKNN conductivity

$$\sigma_{xy} = \frac{e^2}{(2\pi)^d\hbar}\sum_n \int \mathrm{d}\mathbf{k} f(\epsilon_{n,\mathbf{k}})f_{\mu\nu}^{(n)}(\mathbf{k})$$

vanishes. In conclusion, time reversal symmetry *must be broken* in order for the anomalous Hall effect to occur. This always happens in ferromagnets when the material spontaneously magnetizes. The symmetry breaking usually happens in the spin sector, and experience has shown that spin-orbit coupling is needed to "transfer" the information to the orbital sector and thus develop a non-zero Berry curvature. We now examine a simple model of itinerant ferromagnetism which demonstrates this.

7.4 The Rashba model

This section is based on [21]. Consider a generic 2D $\mathbf{k}$-dependent Bloch Hamiltonian of the form

$$\hat{H}(\mathbf{k}) = \frac{1}{2m}(\hat{\mathbf{p}} + \hbar\mathbf{k})^2 + U(\mathbf{r}) = \hat{H}(\mathbf{0}) + \frac{\hbar^2 k^2}{2m} + \frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m}$$

for some potential $U(\mathbf{r})$, acting on the space $\mathcal{H}$ of lattice periodic functions. In order to be able to compute the Berry curvatures of the system, we essentially need to diagonalize $\hat{H}(\mathbf{k})$ for all $\mathbf{k}$ in the first Brillouin zone. To this end, it is often useful to take $\hat{H}(\mathbf{0})$ with eigenstates $|n,\mathbf{0}\rangle$ and energies $\epsilon_{n,\mathbf{0}}$ as a starting point and treat the remaining terms perturbatively. To second order,

$$\epsilon_{n,\mathbf{k}} \approx \epsilon_{n,\mathbf{0}} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\frac{|\langle u_{n,\mathbf{0}}|\mathbf{k}\cdot\hat{\mathbf{p}}|u_{n',\mathbf{0}}\rangle|^2}{\epsilon_{n',\mathbf{0}} - \epsilon_{n,\mathbf{0}}} = \epsilon_{n,\mathbf{0}} + \gamma k^2,$$

assuming the dispersion has an extremum at $\mathbf{k} = \mathbf{0}$ so that the linear term vanishes. In the last equality, we introduced $\gamma \coloneqq \hbar^2/2m^*$ with

$$m^* = \left(\frac{1}{m} + \frac{2}{m^2k^2}\sum_{n'\neq n}\frac{|\langle u_{n,\mathbf{0}}|\mathbf{k}\cdot\hat{\mathbf{p}}|u_{n',\mathbf{0}}\rangle|^2}{\epsilon_{n',\mathbf{0}} - \epsilon_{n,\mathbf{0}}}\right)^{-1}$$

the *effective mass*. Note that m^* depends only on the direction of $\mathbf{k}$ and not on its magnitude. In the two-dimensional subspace of $\mathcal{H}$ spanned by the unperturbed eigenstates $|n, \mathbf{0}, \uparrow\rangle$ and $|n, \mathbf{0}, \downarrow\rangle$, the Hamiltonian now has the approximate form $\hat{H}(\mathbf{k}) \approx (\epsilon_{n,\mathbf{0}} + \gamma k^2) I_2$, with I_2 the 2×2 identity matrix. For details about this so-called $\mathbf{k} \cdot \mathbf{p}$ method, we refer to [22, Section 2.6].

Now, in the *Rashba model*, we add two more terms to $\hat{H}(\mathbf{k})$. As we will see later, one breaks inversion symmetry while the other breaks time reversal invariance; when put together, all spin degeneracy is lifted. The first term is the *Rashba spin-orbit coupling*

$$\hat{V}_{\text{SO}} = \alpha f(k)(\hat{\boldsymbol{\sigma}} \times \mathbf{k}) \cdot \mathbf{e}_z = \alpha f(k)(\hat{\sigma}_x k_y - \hat{\sigma}_y k_x) \tag{7.4}$$

with α a non-zero real number, $f(k)$ an arbitrary function and $\hat{\boldsymbol{\sigma}}$ the vector of Pauli spin matrices. To understand where it comes from, note that an electron moving through an electric field $\mathbf{E}$ experiences an effective magnetic field $\mathbf{B}_{\text{eff}} = -(\mathbf{p} \times \mathbf{E})/mc^2$ in its rest frame, with c the speed of light. The resulting Zeeman energy contribution $\hat{V}_{\text{SO}} \sim (\hat{\mathbf{p}} \times \mathbf{E}) \cdot \hat{\boldsymbol{\sigma}}$ then takes the form (7.4) if $\mathbf{E}$ is assumed to point in the z-direction (see [23]).[5] In the given basis,

$$\hat{V}_{\text{SO}} = \alpha f(k) \begin{pmatrix} 0 & k_y + i k_x \\ k_y - i k_x & 0 \end{pmatrix} = i\alpha f(k) \begin{pmatrix} 0 & k_- \\ -k_+ & 0 \end{pmatrix},$$

where $k_\pm \coloneqq k_x \pm i k_y$. Note that $\overline{k_+} = k_-$ and $|k_+| = |k_-| = k$. The second term reads

$$\hat{V}_{\text{EF}} = h_0 \hat{\sigma}_z = \begin{pmatrix} h_0 & 0 \\ 0 & -h_0 \end{pmatrix}$$

and it represents the coupling of spins to a uniform *exchange field* $h_0 \mathbf{e}_z$. Such an effective exchange field is present in ferromagnets due to the spontaneous magnetization, but it can also be artificially realized by introducing magnetic impurities through *doping*. Adding these two terms together gives the total perturbation

$$\hat{H}' = \hat{V}_{\text{SO}} + \hat{V}_{\text{EF}} = \begin{pmatrix} h_0 & i\alpha f(k) k_- \\ -i\alpha f(k) k_+ & -h_0 \end{pmatrix}.$$

The corrections to the energy eigenvalues as well as the associated "good" linear combinations of basis states can be found by diagonalizing this matrix. Its eigenvalues are quickly found to be $\pm\Delta$ with

$$\Delta = \sqrt{h_0^2 + \alpha^2 f(k)^2 k^2},$$

and corresponding eigenvectors are

$$\begin{pmatrix} i\alpha f(k) k_- \\ -(h_0 \mp \Delta) \end{pmatrix}, \text{ with eigenvalue } \pm\Delta.$$

These have squared norm

$$\begin{aligned} N_\pm^2 &= |i\alpha f(k) k_-|^2 + (h_0 \mp \Delta)^2 \\ &= \alpha^2 f(k)^2 k^2 + h_0^2 + \Delta^2 \mp 2h_0\Delta \\ &= 2\Delta^2 \mp 2h_0\Delta \\ &= 2\Delta(\Delta \mp h_0), \end{aligned}$$

so we obtain the following (approximate) spin-split subband eigenstates:

$$|n, \mathbf{k}, \pm\rangle = \frac{i\alpha f(k) k_-}{N_\pm} |n, \mathbf{0}, \uparrow\rangle - \frac{h_0 \mp \Delta}{N_\pm} |n, \mathbf{0}, \downarrow\rangle. \tag{7.5}$$

[5] If $\mathbf{E}$ were radial, we would instead get the usual $\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}$ spin-orbit coupling.

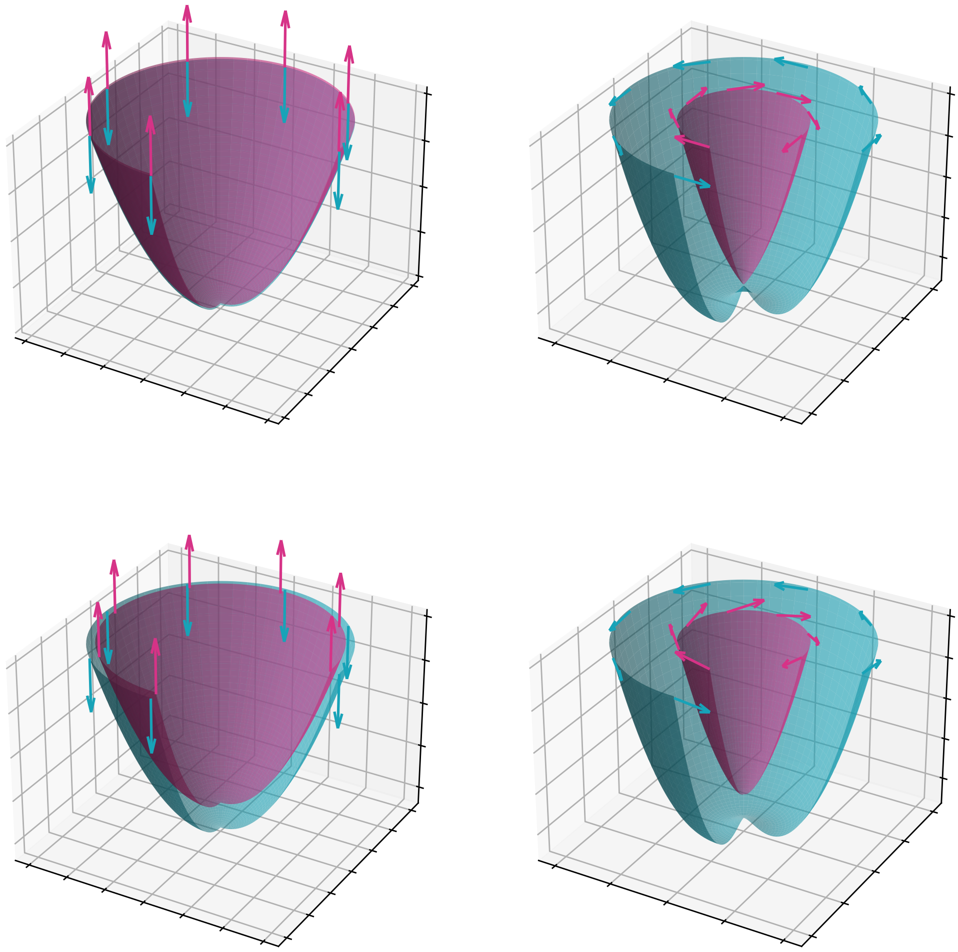


Figure 7.2: Sketch of the dispersion and spin quantization direction for the upper (pink) and lower (blue) branches of any band in the Rashba model, for $\gamma \equiv 1$, $f(\mathbf{k}) \equiv 1$ and different values of the parameters α and h_0. If both are zero (top left), the system is invariant under both spatial inversion and time reversal and the two subbands coincide. If only $\alpha \neq 0$ (top right), only inversion symmetry is broken and the bands cross at $\mathbf{k} = \mathbf{0}$. If only $h_0 \neq 0$ (bottom left), we get two vertically shifted copies of the unperturbed dispersion, with a gap of size $2h_0$. If both are nonzero (bottom right), the band crossing at $\mathbf{k} = \mathbf{0}$ is narrowly avoided.

By computing the expected value of the spin operator $\hat{\mathbf{S}} = \hbar\hat{\boldsymbol{\sigma}}/2$ in the states $|n, \mathbf{k}, \pm\rangle$, we can see how the spin-orbit coupling and exchange field terms in the Rashba Hamiltonian give rise to ($\mathbf{k}$-dependent) spin quantization directions. A straightforward albeit rather tedious computation gives

$$\begin{aligned}
\langle \hat{\mathbf{S}}_\parallel \rangle_{n,\mathbf{k},\pm} &= \pm\frac{\hbar}{2} \cdot \frac{\alpha f(k) k}{\sqrt{h_0^2 + \alpha^2 f(k)^2 k^2}} \cdot \frac{1}{k} \begin{pmatrix} k_y \\ -k_x \end{pmatrix}, \\
\langle \hat{S}_z \rangle_{n,\mathbf{k},\pm} &= \pm\frac{\hbar}{2} \cdot \frac{h_0}{\sqrt{h_0^2 + \alpha^2 f(k)^2 k^2}}.
\end{aligned}$$

From this, we see that the spin-orbit coupling $\hat{V}_{\mathrm{SO}}$ aligns spins in the xy-plane and perpendicular to the wave vector $\mathbf{k}$ ("coupling" spin to momentum), while the exchange field $\hat{V}_{\mathrm{EF}}$ tries to tilt the spins in the z-direction. This is shown in Figure 7.2, together with some sketches of the dispersion.

Observe:

- if $\alpha = 0$ and $h_0 = 0$, both spatial inversion and time reversal symmetry are present, which means the band is doubly degenerate everywhere;
- if $\alpha \neq 0$ and $h_0 = 0$, only inversion symmetry is broken and the spin-split subbands cross at the time-reversal invariant momentum $\mathbf{k} = \mathbf{0}$;
- if $\alpha = 0$ and $h_0 \neq 0$, only time reversal symmetry is broken;
- if $\alpha \neq 0$ and $h_0 \neq 0$, both symmetries are broken and the bands are isolated.

With Equation (7.5), we can approximate the Berry curvatures

$$\begin{aligned} f_{xy}^{(n,\pm)} &= 2\,\mathrm{Im}\langle \partial_{k_x}(n,\mathbf{k},\pm)|\partial_{k_y}(n,\mathbf{k},\pm)\rangle \\ &= 2\,\mathrm{Im}\bigg(\frac{\partial}{\partial k_x}\overline{\left[\frac{i\alpha f(k)k_-}{N_\pm}\right]}\frac{\partial}{\partial k_y}\left[\frac{i\alpha f(k)k_-}{N_\pm}\right] + \frac{\partial}{\partial k_x}\overline{\left[\frac{h_0 \mp \Delta}{N_\pm}\right]}\frac{\partial}{\partial k_y}\left[\frac{h_0 \mp \Delta}{N_\pm}\right]\bigg) \\ &= 2\,\mathrm{Im}\bigg(-\frac{\partial}{\partial k_x}\left[\frac{i\alpha f(k)k_+}{N_\pm}\right]\frac{\partial}{\partial k_y}\left[\frac{i\alpha f(k)k_-}{N_\pm}\right] + \frac{\partial}{\partial k_x}\left[\frac{h_0 \mp \Delta}{N_\pm}\right]\frac{\partial}{\partial k_y}\left[\frac{h_0 \mp \Delta}{N_\pm}\right]\bigg) \end{aligned}$$

of the two subbands. First, note that the second term within the parentheses is real, so

$$f_{xy}^{(n,\pm)} = 2\alpha^2\,\mathrm{Im}\bigg(\frac{\partial}{\partial k_x}\left[\frac{f(k)k_+}{N_\pm}\right]\frac{\partial}{\partial k_y}\left[\frac{f(k)k_-}{N_\pm}\right]\bigg).$$

After some algebra, it follows that

$$\begin{aligned} \frac{\partial}{\partial k_x}\left[\frac{f(k)k_+}{N_\pm}\right] &= \frac{1}{N_\pm}\bigg(\frac{1}{k}f'(k)k_x(k_x+ik_y)+f(k)\bigg) - \frac{1}{N_\pm^3}\alpha^2 f(k)^2 k_x(k_x+ik_y)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k]\bigg(2 \mp \frac{h_0}{\Delta}\bigg), \\ \frac{\partial}{\partial k_y}\left[\frac{f(k)k_-}{N_\pm}\right] &= \frac{1}{N_\pm}\bigg(\frac{1}{k}f'(k)k_y(k_x-ik_y)-if(k)\bigg) - \frac{1}{N_\pm^3}\alpha^2 f(k)^2 k_y(k_x-ik_y)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k]\bigg(2 \mp \frac{h_0}{\Delta}\bigg). \end{aligned}$$

The product of these two derivatives has four terms. The $N_\pm^{-6}$ term has no imaginary part since $k_+k_- = k^2$ is real, which means it will not contribute to the Berry curvature. Writing out the other three, we eventually obtain

$$\begin{aligned} f_{xy}^{(n,\pm)} &= 2\alpha^2\bigg(-\frac{1}{N_\pm^2}f(k)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k] + \frac{1}{N_\pm^4}\alpha^2 f(k)^3(k_x^2+k_y^2)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k]\bigg(2 \mp \frac{h_0}{\Delta}\bigg)\bigg) \\ &= 2\alpha^2 f(k)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k]\frac{1}{N_\pm^2}\bigg(\frac{1}{N_\pm^2}(\Delta^2-h_0^2)\bigg(2 \mp \frac{h_0}{\Delta}\bigg) - 1\bigg). \end{aligned}$$

Now,

$$\begin{aligned} \frac{1}{N_\pm^2}(\Delta^2-h_0^2)\bigg(2 \mp \frac{h_0}{\Delta}\bigg) - 1 &= \frac{\Delta^2-h_0^2}{2\Delta(\Delta \mp h_0)}\cdot\frac{2\Delta \mp h_0}{\Delta} - 1 = \frac{\Delta \pm h_0}{2\Delta}\cdot\frac{2\Delta \mp h_0}{\Delta} - 1 \\ &= \frac{2\Delta^2 \pm h_0\Delta - h_0^2}{2\Delta^2} - 1 = \pm\frac{h_0(\Delta \mp h_0)}{2\Delta^2} = \pm\frac{h_0 N_\mp^2}{4\Delta^3} \end{aligned}$$

and thus

$$\boxed{f_{xy}^{(n,\pm)} = \pm\frac{\alpha^2 h_0 f(k)\frac{\mathrm{d}}{\mathrm{d}k}[f(k)k]}{2\Delta^3} = \mp\frac{h_0}{2k}\cdot\frac{\mathrm{d}}{\mathrm{d}k}\left[\frac{1}{\Delta}\right].}$$

See Figure 7.3 for a plot. The Berry curvature is heavily spiked around $\mathbf{k} = \mathbf{0}$, the point at which a band crossing is narrowly avoided.

Now, the TKNN formula (6.29) for $d = 2$ says that the contribution of any subband to the Hall conductivity σ_{xy} is just the integral of its Berry curvature over the occupied part of the first Brillouin zone, multiplied by $e^2/2\pi h$. We make two further assumptions:

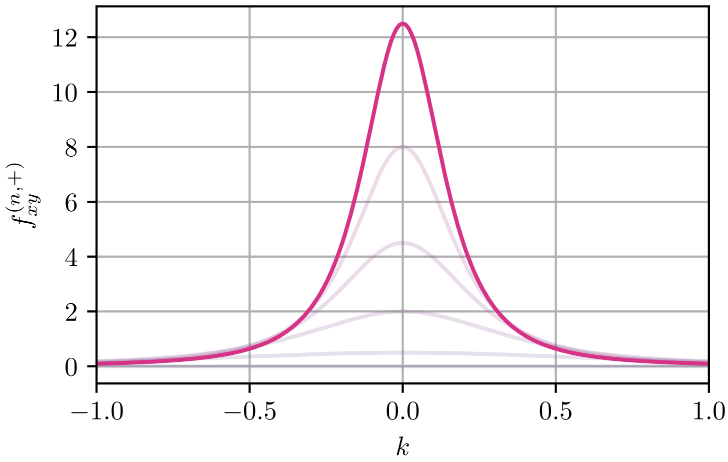


Figure 7.3: Sketch of the Berry curvature $f_{xy}^{(n,+)}$ around $\mathbf{k} = \mathbf{0}$ of any upper branch in the Rashba model, for $\gamma \equiv 1$, $f(k) \equiv 1$, $h_0 = 0.4$ and 6 distinct values of α ranging from 0 to 2. The height of the peak increases with α. The Berry curvature of the corresponding lower branch is just $-f_{xy}^{(n,+)}$.

(i) the effective mass m^* is isotropic, so that the (approximate) dispersion is rotationally symmetric;

(ii) the Fermi energy lies between the local maximum at $\mathbf{k} = \mathbf{0}$ of the lower branch of some band (the n-th, say) and the local minimum at $\mathbf{k} = \mathbf{0}$ of its upper branch, see Figure 7.2.

In particular, the second assumption implies the presence of an exchange field, since there is no gap at the origin when $h_0 = 0$. The contributions to σ_{xy} of completely filled upper and lower subbands cancel each other out since $F_{xy}^{(n',-)} = -F_{xy}^{(n',+)}$ for any band index n', so the only surviving term is the one corresponding to the lower branch of band n. By rotational symmetry of the dispersion relation, the occupied states within that subband form a circle of radius k_F in the first Brillouin zone, so

$$\begin{aligned}\sigma_{xy} &= \frac{e^2}{2\pi h} \iint_{k<k_F} F_{xy}^{(n,-)} \,\mathrm{d}\mathbf{k} = \frac{e^2}{2\pi h} \int_0^{2\pi} \int_0^{k_F} F_{xy}^{(n,-)} k \,\mathrm{d}k \,\mathrm{d}\theta \\ &= \frac{e^2}{h} \cdot \frac{h_0}{2} \left[\frac{1}{\Delta} \right]_0^{k_F} = \frac{e^2}{2h} \left(\frac{h_0}{(h_0^2 + \alpha^2 f(k_F)^2 k_F^2)^{3/2}} - 1 \right).\end{aligned}$$

The geometric interpretation of the TKNN formula tells us the following: if the system is prepared in a state $|n, \mathbf{k}, -\rangle$ with $|\mathbf{k}| = k_F$ and we then let $\mathbf{k}$ walk counterclockwise along the circle of radius k_F in reciprocal space, a geometric Berry phase is acquired which is given precisely by the term in parentheses, multiplied by π.

Chapter 8

Conclusion and outlook

In this thesis, we have investigated the geometric Berry phase interpretation of the intrinsic anomalous Hall effect. In the finite-dimensional case, the tautological $U(1)$-bundle and its canonical connection form can be pulled back along the Bloch eigenspace map $T^d \to \mathbb{P}(\mathcal{H})$ of any band to obtain a principal $U(1)$-bundle π over the folded up Brillouin zone T^d, as well as a connection on π. In the infinite-dimensional case, the theory of local connection forms is required to lift smooth $i\mathbb{R}$-valued forms on open subsets of T^d to the total space of the bundle and patch them together. Either way, the connection on π gives rise to a notion of horizontal lifting which reflects the physical time evolution of the system according to the Schrödinger equation whenever the wave vector traverses a path in $\mathbf{k}$-space. If that path is a loop, the initial and final states differ by a geometric phase factor known as a Berry phase, and in two-dimensional systems at low temperatures, the interband anomalous Hall conductivity can be interpreted as a sum of such Berry phases through the TKNN formula. For this sum to be nonvanishing, a combination of time reversal symmetry breaking and spin-orbit coupling is required, as demonstrated by the Rashba model.

There are still some loose ends on both the mathematics and physics side, which might be worth looking into. Recall from Theorem 3.15 that principal $U(1)$-bundles over a smooth manifold M are classified by homotopy classes of smooth maps $M \to \mathbb{P}^n(\mathbb{C})$ for n large enough. However, there also exists an alternative classification scheme if M is the orbit space for the smooth left action of $H = \mathbb{Z}^d$ on $X = \mathbb{R}^d$ by translation, i.e. if $M = H\backslash X = T^d$. In short, any principal G-bundle $\pi\colon P \to M$ can be pulled back along the quotient map $q\colon X \to M$ to obtain a principal G-bundle $\tilde{\pi}\colon q^*P \to X$, which is necessarily trivial since X is contractible. That means $\tilde{\pi}$ admits a global section $\tilde{\sigma}\colon X \to q^*P$, which we can compose with the projection $q^*P \to P$ on the second coordinate to obtain a map $\sigma\colon X \to P$ satisfying $\pi(\sigma(h+x)) = \pi(\sigma(x))$ for all $h \in H$ and $x \in X$. This uniquely defines a map $\gamma\colon H \times X \to G$ with $\sigma(h+x)\gamma(h,x) = \sigma(x)$, which in turn implies that γ obeys the *cocycle condition* $\gamma(h'+h,x) = \gamma(h', h+x)\gamma(h,x)$. Such a map $H \times X \to G$ satisfying the cocycle condition is known as a *cocycle*. Two cocycles γ and γ' are said to be equivalent if there exists a function $f\colon X \to G$ such that $\gamma'(h,x) = f(h+x)\gamma(h,x)f(x)^{-1}$ for all $h \in H$ and $x \in X$, and it can be shown that choosing a different global section for $\tilde{\pi}$ or starting with an isomorphic bundle results in an equivalent cocycle. The details have not yet been checked, but it seems like under some smoothness conditions on the γ and f, this sets up a one-to-one correspondence between isomorphism classes of principal G-bundles over M and equivalence classes of cocycles. This of course begs the question how the two classifications are related: given a smooth map $M \to \mathbb{P}^n(\mathbb{C})$, which cocycle yields the same principal G-bundle isomorphism class and vice versa?

On the same note, it might be interesting to examine how much of the theory of principal G-bundles, connections and their classification generalizes to the case of infinite-dimensional manifolds modeled on Banach spaces, as defined in for example [11]. Our construction of the connection on the bundle π over T^d in the infinite-dimensional case (see Section 7.1) was rather messy because we could not do much more than simply assume that everything works. It would be great if both cases could instead be treated on an equal footing.

As for the derivation of the TKNN formula in Chapter 6, (at least) two things deserve attention. The first is the continuum limit $L \to \infty$ with which the sum over the allowed wave vectors in discretized reciprocal space was converted to an integral over the Brillouin zone: why is it even necessary to confine the system to a finite volume in the first place? We chose to do it because it resolves the issue of the Bloch waves not being normalizable in an unbounded system and makes it easier to show that the off-$\mathbf{k}$-diagonal matrix elements of the velocity operator vanish (see Lemma 6.2), but it might be possible to circumvent these problems in a different way. Second, we ignored the Drude conductivity (the $n = n'$ terms in (6.24)) without real justification, outside of the fact that it blows up in the clean limit. According to [24], one of the two terms in the so-called Středa formula for conductivity vanishes whenever the Fermi energy of the system lies in a band gap. Since the classical Drude conductivity applies mainly to metals, it seems plausible that this holds for the $n = n'$ terms in the Kubo formula, too. Perhaps there is a link between the distinction of the Drude and interband conductivity in the Kubo formula and the splitting of the Středa conductivity which has not yet been explored.

Another important thing to note is that the TKNN formula only describes one contribution to the anomalous Hall effect, an *intrinsic* one which depends only on the material's band structure. As explained in the introduction, other contributions have also been identified, the most important being the side jump and skew-scattering mechanisms. Both rely heavily on disorder in the system. Attempts have been made to combine all three contributions into a unified theory. For an overview of three such attempts – the semiclassical Boltzmann approach and the Kubo and Keldysh formalisms – we refer to [5, Section IV]. In certain temperature and longitudinal conductivity regimes, the side-jump and skew scattering dominate (see [5]), so the intrinsic anomalous Hall effect we have discussed is only part of the story.

Finally, it is worth noting that the Rashba model calculations from Section 7.4 seem to disagree qualitatively with principles from the field of *topological matter* that the intrinsic anomalous Hall effect is a part of. As was explained at the end of Section 7.2, the TKNN conductivity of a filled band is quantized: it labels a "topological phase". For a topological phase transition to occur, there must be some structural change in the system's band structure, such as the opening or closing of a gap. In the Rashba model, the Berry curvature (and thus the conductivity) of a band vanishes in the absence of spin-orbit coupling. When spin-orbit coupling is turned on, the Berry curvature starts to spike around $\mathbf{k} = \mathbf{0}$, as shown in Figure 7.3, which should imply that a band gap opens or closes somewhere along the way. However, this is not visible in the dispersion shown in Figure 7.2. Perhaps this has to do with the fact that the approximation breaks down further away from $\mathbf{k} = \mathbf{0}$, which is also why we can't directly see quantization in σ_{xy}: for that, we would need to be able to integrate over the entire Brillouin zone. Either way, it would be interesting to see if things change when higher order terms in the approximation are included, and to compare it to other models, such as the tight binding honeycomb lattice model introduced by Haldane in 1988 ([25], see also [17, Section 5.1.1]).

Appendices

Appendix A

Preliminaries: differential geometry

A.1 Product manifolds

In this section, we show that the tangent spaces to a product manifold split as a direct sum in a natural way, a result which is used on numerous occasions throughout the rest of this thesis. Let M_1 and M_2 be smooth manifolds of dimension n_1 and n_2, respectively, then $M \coloneqq M_1 \times M_2$ can be made into a smooth manifold of dimension $n \coloneqq n_1 + n_2$. Let $p = (p_1, p_2) \in M$ and define injections $j_1 \colon M_1 \to M$, $q_1 \mapsto (q_1, p_2)$ and $j_2 \colon M_2 \to M$, $q_2 \mapsto (p_1, q_2)$. Also let $\pi_i \colon M \to M_i$ be the projection for $i \in \{1, 2\}$.

Lemma A.1. *The linear map* $\mathrm{d}(j_1 \circ \pi_1)_p + \mathrm{d}(j_2 \circ \pi_2)_p$ *is the identity on* T_pM*.*

Proof. Choose coordinate neighborhoods (U_1, ϕ_1) and (U_2, ϕ_2) of p_1 and p_2 in M_1 and M_2, respectively, then (U, ϕ) with $U \coloneqq U_1 \times U_2 \subseteq M$ and

$$\begin{aligned} \phi \colon U = U_1 \times U_2 &\longrightarrow \mathbb{R}^n = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \\ (q_1, q_2) &\longmapsto (\phi_1(q_1), \phi_2(q_2)) \end{aligned}$$

is a chart on M containing p. This gives a basis $\partial_1|_p, \ldots, \partial_n|_p$ for T_pM, where

$$\partial_i|_p f \coloneqq \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ \phi^{-1})$$

for all $i \in \{1, \ldots, n\}$ and $f \in C^\infty(M)$. It suffices to show that $\mathrm{d}(j_1 \circ \pi_1)_p + \mathrm{d}(j_2 \circ \pi_2)_p$ sends each of these basis vectors to itself. Let $f \in C^\infty(M)$ and $x = (x^1, \ldots, x^n) \in \phi(U) = \phi_1(U_1) \times \phi_2(U_2)$, then

$$\begin{aligned} (f \circ j_1 \circ \pi_1 \circ \phi^{-1})(x) &= f(\phi_1^{-1}(x^1, \ldots, x^{n_1}), p_2) = (f \circ \phi^{-1})(x^1, \ldots, x^{n_1}, \phi_2(p_2)), \\ (f \circ j_2 \circ \pi_2 \circ \phi^{-1})(x) &= f(p_1, \phi_2^{-1}(x^{n_1+1}, \ldots, x^n)) = (f \circ \phi^{-1})(\phi_1(p_1), x^{n_1+1}, \ldots, x^n). \end{aligned}$$

It follows that

$$\mathrm{d}(j_1 \circ \pi_1)_p(\partial_i|_p) f = \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ j_1 \circ \pi_1 \circ \phi^{-1}) = \begin{cases} \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ \phi^{-1}) & \text{if } 1 \le i \le n_1 \\ 0 & \text{if } n_1 < i \le n, \end{cases}$$

$$\mathrm{d}(j_2 \circ \pi_2)_p(\partial_i|_p) f = \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ j_2 \circ \pi_2 \circ \phi^{-1}) = \begin{cases} 0 & \text{if } 1 \le i \le n_1 \\ \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ \phi^{-1}) & \text{if } n_1 < i \le n \end{cases}$$

for all $i \in \{1, \ldots, n\}$, so $\mathrm{d}(j_1 \circ \pi_1)_p(\partial_i|_p) + \mathrm{d}(j_2 \circ \pi_2)_p(\partial_i|_p) = \partial_i|_p$ and we are done. □

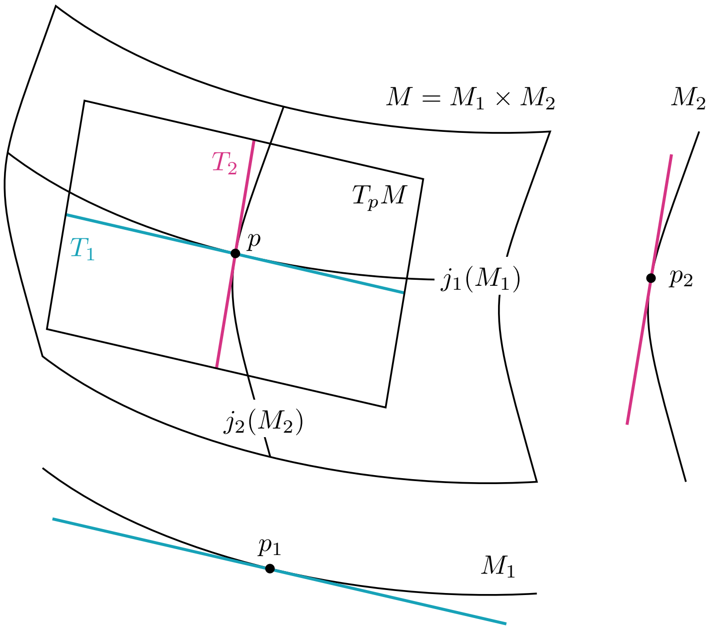


Figure A.1: Illustration of Theorem A.2.

Using this technical lemma, we can show that T_pM splits as a direct sum, and that this direct sum decomposition has two different characterizations.

Theorem A.2. *Set* $T_i \coloneqq \mathrm{d}(j_i)_{p_i}(T_{p_i}M_i) \subseteq T_pM$ *for* $i \in \{1, 2\}$*, then:*

(i) $T_pM = T_1 \oplus T_2$*;*

(ii) $T_1 = \ker \mathrm{d}(\pi_2)_p$ *and* $T_2 = \ker \mathrm{d}(\pi_1)_p$*.*

Proof. We prove the two claims separately.

(i) Let $v \in T_1 \cap T_2$, then there exist $w_1 \in T_{p_1}M_1$ and $w_2 \in T_{p_2}M_2$ such that $\mathrm{d}(j_1)_{p_1}(w_1) = v = \mathrm{d}(j_2)_{p_2}(w_2)$. Applying $\mathrm{d}(\pi_1)_p$ on both sides gives $w_1 = 0$ since $\pi_1 \circ j_1 = \mathrm{id}_{M_1}$ while $\pi_1 \circ j_2$ is the constant map $q_2 \mapsto p_1$. It follows that $v = 0$, so $T_1 \cap T_2 = \{0\}$.

The fact that $T_pM = T_1 + T_2$ holds, follows directly from Lemma A.1: for any $v \in T_pM$,

$$v = \mathrm{id}_{T_pM}(v) = \mathrm{d}(j_1)_{p_1}(\mathrm{d}(\pi_1)_p(v)) + \mathrm{d}(j_2)_{p_2}(\mathrm{d}(\pi_2)_p(v)) \in T_1 + T_2.$$

(ii) We prove $T_2 = \ker \mathrm{d}(\pi_1)_p$, the other equality follows from symmetry. Let $w \in T_{p_2}M_2$, then

$$\mathrm{d}(\pi_1)_p(\mathrm{d}(j_2)_{p_2}(w)) = \mathrm{d}(\pi_1 \circ j_2)(w) = 0$$

by the same argument as above, so $\mathrm{d}(j_2)_{p_2}(w) \in \ker \mathrm{d}(\pi_1)_p$ and $T_2 \subseteq \ker \mathrm{d}(\pi_1)_p$. For the other inclusion, let $v \in \ker \mathrm{d}(\pi_1)_p$, then

$$v = \mathrm{id}_{T_pM}(v) = \mathrm{d}(j_1)_{p_1}(\mathrm{d}(\pi_1)_p(v)) + \mathrm{d}(j_2)_{p_2}(\mathrm{d}(\pi_2)_p(v)) = \mathrm{d}(j_2)_{p_2}(\mathrm{d}(\pi_2)_p(v)) \in T_2$$

by Lemma A.1. ☐

A.2 Vector spaces

The tangent spaces to any finite-dimensional $\mathbb{R}$-vector space V are also easily classified: they are canonically isomorphic to the space V itself. More precisely, we have the following. We refer to [8, Proposition 3.13] for a proof.

Lemma A.3. *For all $x \in V$, the map*

$$\begin{aligned} D_x \colon V &\longrightarrow T_xV \\ y &\longmapsto \left(f \mapsto \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(x+ty) \right), \end{aligned}$$

is a linear isomorphism such that for any $\mathbb{R}$-linear map $L \colon V \to W$, the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow[D_x]{\sim} & T_xV \\ \big\downarrow L & & \big\downarrow \mathrm{d}L_x \\ W & \xrightarrow[D_{Lx}]{\sim} & T_{Lx}W \end{array}$$

Recall that the smooth structure on V is defined such that for any basis $e_1, \dots, e_n$ for V, the map

$$\begin{aligned} \phi \colon V &\longrightarrow \mathbb{R}^n \\ x^i e_i &\longmapsto (x^1, \dots, x^n) \end{aligned} \tag{A.1}$$

is a smooth global coordinate chart, where we used the Einstein summation convention. Now let $x = x^i e_i \in V$, then for all i, the isomorphism D_x of Lemma A.3 maps e_i to the coordinate vector

$$\frac{\partial}{\partial x^i}\bigg|_x := \mathrm{d}(\phi^{-1})_{\phi(x)} \left(\frac{\partial}{\partial x^i}\bigg|_{\phi(x)} \right) \in T_xV$$

associated to the chart ϕ, since

$$\begin{aligned} D_x(e_i)f = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(x+te_i) &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 (f \circ \phi^{-1})(x^1, \dots, x^i + t, \dots, x^n) \\ &= \frac{\partial}{\partial x^i}\bigg|_{\phi(x)} (f \circ \phi^{-1}) = \mathrm{d}(\phi^{-1})_{\phi(x)} \left(\frac{\partial}{\partial x^i}\bigg|_{\phi(x)} \right) f = \frac{\partial}{\partial x^i}\bigg|_x f \end{aligned}$$

for all $f \in C^\infty(V)$ by the chain rule. In other words, $e_1, \dots, e_n$ and $\partial/\partial x^1\,|_x, \dots, \partial/\partial x^n\,|_x$ correspond under the canonical isomorphism D_x.

If we use Lemma A.3 to identify V with two different tangent spaces, it turns out that these identifications differ by the pushforward of a translation.

Lemma A.4. *Let $x, y \in V$ and define $T \colon V \to V$, $x \mapsto x + y$, then $\mathrm{d}T_x \circ D_x = D_{x+y}$.*

Proof. Let $z \in V$ and $f \in C^\infty(V)$, then

$$(\mathrm{d}T_x \circ D_x)(z)f = D_x(z)(f \circ T) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 (f \circ T)(x+tz) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f((x+y)+tz) = D_{x+y}(z)f. \qquad \square$$

Appendix B

Preliminaries: quantum mechanics

B.1 Second quantization

The state of a quantum mechanical system is described at any time t by an element of a Hilbert space $\mathcal{H}$, i.e. a complex inner product space which is complete with respect to the metric induced by the inner product, while observables are represented by Hermitian operators on $\mathcal{H}$. *Second quantization* provides a description of the state space and observables of a system made up of arbitrarily many identical particles, starting from the state space and observables of the system describing one such particle. A mathematically rigorous discussion of this formalism requires quite a lot of machinery from linear analysis and lies outside the scope of this thesis. At the same time, many texts discussing the subject from a physicist's point of view leave out almost all of the mathematics. Here, we take an intermediate stance: the antisymmetric Fock space is defined somewhat rigorously, but in the results concerning linear operators, all subtleties regarding for instance domains of definition and boundedness are swept under the rug. This level of precision suffices for our derivation of the TKNN formula in Chapter 6, which is already inherently imprecise.

Fock space. To start, let $\mathcal{H}$ be any complex Hilbert space, which we interpret as the state space of the one-particle system. For all $N \in \mathbb{Z}_{\geq 0}$, it can be shown that the bilinear form on the N-fold tensor product $\mathcal{H}^{\otimes N} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$ defined for elementary tensors by

$$\langle x_1 \otimes \cdots \otimes x_N | y_1 \otimes \cdots \otimes y_N \rangle = \prod_{i=1}^{N} \langle x_i | y_i \rangle$$

is an inner product, see for example [26, Proposition II.4.1]. The completion of $\mathcal{H}^{\otimes N}$ with respect to this inner product is denoted by $\mathcal{H}^{(N)}$. The inner product on $\mathcal{H}^{\otimes N}$ extends naturally to $\mathcal{H}^{(N)}$, making it into a Hilbert space. Any σ in the permutation group S_N on N elements induces a bounded linear operator of norm 1 on $\mathcal{H}^{\otimes N}$ defined on elementary tensors by

$$\sigma(x_1 \otimes \cdots \otimes x_N) = x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(N)},$$

which then extends to a bounded linear operator of the same norm on $\mathcal{H}^{(N)}$ since $\mathcal{H}^{(N)}$ contains $\mathcal{H}^{\otimes N}$ as a dense subspace. By [26, Problem II.23], the *alternation operator*

$$A_N := \frac{1}{N!} \sum_{\sigma \in S_N} (\operatorname{sgn} \sigma)\sigma$$

on $\mathcal{H}^{(N)}$ is an orthogonal projection, i.e. $A_N^2 = A_N$ and $A_N^* = A_N$. Its range $\mathcal{H}_a^{(N)} := A_N \mathcal{H}^{(N)}$ is a closed subspace of $\mathcal{H}^{(N)}$ and thus a Hilbert space. Finally, we can define the *fermionic Fock space* of $\mathcal{H}$ as the $\mathbb{C}$-vector space

$$\mathcal{F}(\mathcal{H}) := \left\{ (\psi_N)_{N\geq 0} : \psi_N \in \mathcal{H}_a^{(N)}, \sum_{N=0}^{\infty} \|\psi_N\|^2 < \infty \right\},$$

which naturally contains each $\mathcal{H}_a^{(N)}$ as a subspace. As shown in [27, Proposition 6.2], it follows that

$$\langle (\psi_N)_{N\geq 0} | (\phi_N)_{N\geq 0} \rangle = \sum_{N=0}^{\infty} \langle \psi_N | \phi_N \rangle$$

defines an inner product on $\mathcal{F}(\mathcal{H})$ which extends the inner products on the $\mathcal{H}_a^{(N)}$ and with respect to which $\mathcal{F}(\mathcal{H})$ is complete. With it, the vector space direct sum of the $\mathcal{H}_a^{(N)}$ becomes a dense subspace of $\mathcal{F}(\mathcal{H})$, which means we can view the fermionic Fock space as its completion:

$$\mathcal{F}(\mathcal{H}) = \overline{\bigoplus_{N=0}^{\infty} \mathcal{H}_a^{(N)}}.$$

If the particle described by $\mathcal{H}$ is a *fermion* (e.g. an electron) and we put together arbitrarily many indistinguishable copies of it in one system, the resulting state space is precisely $\mathcal{F}(\mathcal{H})$. The elements describing states with N particles live in $\mathcal{H}_a^{(N)}$, which is therefore known as the *N-particle subspace*. For example, $A_N(x_1 \otimes \cdots \otimes x_N) \in \mathcal{H}_a^{(N)}$ is interpreted (after normalization) as the state with one x_1 particle, one x_2 particle, etc. Only elements in the range of the alternation operator are allowed since fermionic wave functions must be odd under particle interchange.

Operators on Fock space. Given a Hermitian operator $\hat{A}$ on $\mathcal{H}$ corresponding to some observable, we can construct for all $N \geq \mathbb{Z}_{\geq 0}$ and $1 \leq n \leq N$ a corresponding operator

$$\mathrm{d}\Gamma_n^{(N)}(\hat{A}) := \hat{I} \otimes \cdots \otimes \hat{A} \otimes \cdots \otimes \hat{I}$$

on $\mathcal{H}^{(N)}$, where $\hat{I}$ is the identity on $\mathcal{H}$ and the $\hat{A}$ is in the n-th spot. Note that $\mathrm{d}\Gamma_n^{(N)}$ need not have the range $\mathcal{H}_a^{(N)}$ of A_N as an invariant subspace, but

$$\mathrm{d}\Gamma^{(N)}(\hat{A}) := \sum_{n=1}^{N} \mathrm{d}\Gamma_n^{(N)}(\hat{A}) = \hat{A} \otimes \hat{I} \otimes \cdots \otimes \hat{I} + \hat{I} \otimes \hat{A} \otimes \cdots \otimes \hat{I} + \cdots + \hat{I} \otimes \hat{I} \otimes \cdots \otimes \hat{A} \tag{B.1}$$

does, so we obtain an operator $\mathrm{d}\Gamma(\hat{A})$ on the fermionic Fock space $\mathcal{F}(\mathcal{H})$ of $\mathcal{H}$ defined on the N-particle subspace $\mathcal{H}_a^{(N)}$ as $\mathrm{d}\Gamma(\hat{A})|_{\mathcal{H}_a^{(N)}} = \mathrm{d}\Gamma^{(N)}(\hat{A})$. Explicitly,

$$\begin{aligned} &\mathrm{d}\Gamma(\hat{A})(A_N(x_1 \otimes x_2 \otimes \cdots \otimes x_N)) \\ &\quad = A_N(\hat{A}x_1 \otimes x_2 \otimes \cdots \otimes x_N + x_1 \otimes \hat{A}x_2 \otimes \cdots \otimes x_N + \cdots + x_1 \otimes x_2 \otimes \cdots \otimes \hat{A}x_N) \in \mathcal{H}_a^{(N)} \end{aligned}$$

for any $x_1, \ldots, x_N \in \mathcal{H}$. $\mathrm{d}\Gamma(\hat{A})$ is known as the *second quantization* of $\hat{A}$, and it is also Hermitian; see for instance [28, Section 6.3.2]. It represents the same observable as $\hat{A}$, under the assumption that the system's particles are *independent*. For example, if $\hat{A} = \hat{H}$ is the Hamiltonian of the one-particle system, then $\mathrm{d}\Gamma(\hat{H})$ represents the sum of the energies that the individual particles would have on their own, which is precisely the total energy of the system provided that there are no extra interactions. Another important special case is $\hat{A} = \hat{I}$, for which we write $\hat{N} := \mathrm{d}\Gamma(\hat{I})$. The second quantized operator $\hat{N}$ has each N-particle subspace $\mathcal{H}_a^{(N)}$ of $\mathcal{F}(\mathcal{H})$ as an eigenspace, with corresponding eigenvalue N. For this reason, we call $\hat{N}$ the *particle number operator* on $\mathcal{F}(\mathcal{H})$. As the next lemma shows, the second quantization operator $\mathrm{d}\Gamma$ has a number of useful properties. For details, we refer to [29].

Lemma B.1. *The second quantization map* $\mathrm{d}\Gamma$ *satisfies the following properties:*

(i) $\mathrm{d}\Gamma$ *is* $\mathbb{C}$*-linear;*

(ii) *let* $\hat{A}$ *and* $\hat{B}$ *be Hermitian operators on* $\mathcal{H}$*, then* $\mathrm{d}\Gamma(e^{i\hat{A}}\hat{B}e^{-i\hat{A}}) = e^{i\,\mathrm{d}\Gamma(\hat{A})}\,\mathrm{d}\Gamma(\hat{B})e^{-i\,\mathrm{d}\Gamma(\hat{A})}$*;*

(iii) *let* $\hat{A}$ *and* $\hat{B}$ *be Hermitian operators on* $\mathcal{H}$*, then* $[\mathrm{d}\Gamma(\hat{A}), \mathrm{d}\Gamma(\hat{B})] = \mathrm{d}\Gamma([\hat{A}, \hat{B}])$.

The number representation. Suppose now that $\mathcal{H}$ is an infinite-dimensional separable Hilbert space and let $\{e_k\}_{k\geq 1}$ be an orthonormal basis for $\mathcal{H}$. We can view the e_k as the possible states for any individual particle to be in. For any $N \in \mathbb{Z}_{\geq 0}$, the set

$$\{e_{k_1} \otimes \cdots \otimes e_{k_N} : k_1, \ldots, k_N \geq 1\}$$

is an orthonormal basis for $\mathcal{H}^{(N)}$ by [26, Proposition II.4.2]. As a result,

$$\{\sqrt{N!}A_N(e_{k_1} \otimes \cdots \otimes e_{k_N}) : 1 \leq k_1 < \cdots < k_N\}$$

is an orthonormal basis for $\mathcal{H}_a^{(N)}$. Its elements have their own alternative notation in the physics literature. Let $(n_k)_{k\geq 1}$ be a sequence of elements of $\{0, 1\}$ with $n_k = 0$ for all but finitely many k and let $1 \leq k_1 < \cdots < k_N$ be the indices of the nonzero terms, we then write

$$|(n_k)\rangle \coloneqq \sqrt{N!}A_N(e_{k_1} \otimes \cdots \otimes e_{k_N}).$$

In words, $|(n_k)\rangle$ describes the system when precisely the states $k_1, \ldots, k_N$ are occupied and n_k represents the occupation of state k for all $k \geq 1$. This notation is therefore often referred to as the *occupation number representation*, and the $|(n_k)\rangle$ are known as *Fock states*. For a sequence $(n_k)_{k\geq 1}$ in $\mathbb{Z}$ with only finitely many nonzero terms, we understand $|(n_k)\rangle$ to be 0 whenever $n_k > 1$ or $n_k < 0$ for some k; no two fermions can occupy the same state by the Pauli exclusion principle, nor can any state be occupied by a negative number of particles.

Now, it can be shown (see for example [30, Section 1.3.3]) that there exist operators $\hat{c}_k$ on $\mathcal{F}(\mathcal{H})$ for all $k \geq 1$ with the following properties:

- $\hat{c}_\ell|(n_k)\rangle = \pm|(n_k - \delta_{k\ell})\rangle$ for any $(n_k)_{k\geq 1}$;
- $\hat{c}_\ell^*|(n_k)\rangle = \pm|(n_k + \delta_{k\ell})\rangle$ for any $(n_k)_{k\geq 1}$;
- for all $k, \ell \geq 1$, we have the anticommutation[1] relations

$$\{\hat{c}_k, \hat{c}_\ell\} = 0, \quad \{\hat{c}_k^*, \hat{c}_\ell^*\} = 0, \quad \{\hat{c}_k, \hat{c}_\ell^*\} = \delta_{k\ell}. \tag{B.2}$$

$\hat{c}_k$ and its adjoint $\hat{c}_k^*$ are known as the *fermionic annihilation* and *creation operators* at state k, respectively, since $\hat{c}_k$ lowers the occupation of state k by one while $\hat{c}_k^*$ raises it. An immediate consequence of the above properties is that $\hat{c}_\ell^*\hat{c}_\ell|(n_k)\rangle = n_\ell|(n_k)\rangle$, which is why $\hat{n}_\ell \coloneqq \hat{c}_\ell^*\hat{c}_\ell$ is known as the *occupation number operator* at state ℓ.

Any second quantized operator on $\mathcal{F}(\mathcal{H})$ can be expanded in terms of the creation and annihilation operators as follows. For a derivation, we refer to [30, Equation (1.60)]

Lemma B.2. *Let* $\hat{A}$ *be a Hermitian operator on* $\mathcal{H}$*, then*

$$\mathrm{d}\Gamma(\hat{A}) = \sum_{k,\ell} \langle e_k|\hat{A}e_\ell\rangle \hat{c}_k^*\hat{c}_\ell.$$

[1] The anticommutator of two operators $\hat{A}$ and $\hat{B}$ is defined as the operator $\{\hat{A}, \hat{B}\} \coloneqq \hat{A}\hat{B} + \hat{B}\hat{A}$.

Bibliography

[1] E. H. Hall. On a New Action of the Magnet on Electric Currents. *American Journal of Mathematics*, 2(3):287, September 1879. ISSN 0002-9327. doi: 10.2307/2369245. URL http://dx.doi.org/10.2307/2369245.

[2] Steven H. Simon. *The Oxford Solid State Basics*. Oxford University Press, 2013.

[3] E.H. Hall. On the “Rotational Coefficient” in nickel and cobalt. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 12(74):157–172, September 1881. ISSN 1941-5990. doi: 10.1080/14786448108627086. URL http://dx.doi.org/10.1080/14786448108627086.

[4] Emerson M. Pugh. Hall Effect and the Magnetic Properties of Some Ferromagnetic Materials. *Phys. Rev.*, 36:1503–1511, Nov 1930. doi: 10.1103/PhysRev.36.1503. URL https://link.aps.org/doi/10.1103/PhysRev.36.1503.

[5] Naoto Nagaosa, Jairo Sinova, Shigeki Onoda, A. H. MacDonald, and N. P. Ong. Anomalous Hall effect. *Rev. Mod. Phys.*, 82:1539–1592, May 2010. doi: 10.1103/RevModPhys.82.1539. URL https://link.aps.org/doi/10.1103/RevModPhys.82.1539.

[6] K. v. Klitzing, G. Dorda, and M. Pepper. New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance. *Phys. Rev. Lett.*, 45:494–497, Aug 1980. doi: 10.1103/PhysRevLett.45.494. URL https://link.aps.org/doi/10.1103/PhysRevLett.45.494.

[7] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. Quantized Hall Conductance in a Two-Dimensional Periodic Potential. *Phys. Rev. Lett.*, 49:405–408, Aug 1982. doi: 10.1103/PhysRevLett.49.405. URL https://link.aps.org/doi/10.1103/PhysRevLett.49.405.

[8] John M. Lee. *Introduction to Smooth Manifolds*. Springer New York, 2012. ISBN 9781441999825. doi: 10.1007/978-1-4419-9982-5. URL http://dx.doi.org/10.1007/978-1-4419-9982-5.

[9] Loring W. Tu. *Differential Geometry*. Springer International Publishing, 2017. ISBN 9783319550848. doi: 10.1007/978-3-319-55084-8. URL http://dx.doi.org/10.1007/978-3-319-55084-8.

[10] Gerd Rudolph and Matthias Schmidt. *Differential Geometry and Mathematical Physics: Part I. Manifolds, Lie Groups and Hamiltonian Systems*. Springer Netherlands, 2013. ISBN 9789400753457. doi: 10.1007/978-94-007-5345-7. URL http://dx.doi.org/10.1007/978-94-007-5345-7.

[11] Serge Lang. *Differential and Riemannian Manifolds*. Springer New York, 1995. ISBN 9781461241829. doi: 10.1007/978-1-4612-4182-9. URL http://dx.doi.org/10.1007/978-1-4612-4182-9.

[12] Michiko Nakahara. *Geometry, Topology and Physics*. Graduate student series in physics. Adam Hilger, Bristol [etc, 1990. ISBN 0852740948.

[13] Kristina Chadova. Electronic transport within the Kubo-Bastin Formalism, December 2017. URL http://nbn-resolving.de/urn:nbn:de:bvb:19-216095.

[14] Jan Mrozek. Study of the effect of spin-orbit interaction in solids, 2017. URL http://hdl.handle.net/20.500.11956/90983.

[15] Alexander B. Watson, Dionisios Margetis, and Mitchell Luskin. Mathematical aspects of the Kubo formula for electrical conductivity with dissipation. *Japan Journal of Industrial and Applied Mathematics*, 40(3):1765–1795, September 2023. ISSN 1868-937X. doi: 10.1007/s13160-023-00613-7. URL http://dx.doi.org/10.1007/s13160-023-00613-7.

[16] Ryogo Kubo. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *Journal of the Physical Society of Japan*, 12 (6):570–586, June 1957. ISSN 1347-4073. doi: 10.1143/jpsj.12.570. URL http://dx.doi.org/10.1143/JPSJ.12.570.

[17] David Vanderbilt. *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators*. Cambridge University Press, 2018.

[18] Dariusz Chruściński and Andrzej Jamiołkowski. *Geometric Phases in Classical and Quantum Mechanics*. Birkhäuser Boston, 2004. ISBN 9780817681760. doi: 10.1007/978-0-8176-8176-0. URL http://dx.doi.org/10.1007/978-0-8176-8176-0.

[19] Hongming Weng, Rui Yu, Xiao Hu, Xi Dai, and Zhong Fang. Quantum anomalous Hall effect and related topological electronic states. *Advances in Physics*, 64(3):227–282, 2015. doi: 10.1080/00018732.2015.1068524. URL https://doi.org/10.1080/00018732.2015.1068524.

[20] J.W. Milnor and J.D. Stasheff. *Characteristic Classes*. Annals of mathematics studies. Princeton University Press, 1974. ISBN 9780691081229. URL https://books.google.nl/books?id=5zQ9AFk1i4EC.

[21] Dimitrie Culcer, Allan MacDonald, and Qian Niu. Anomalous Hall effect in paramagnetic two-dimensional systems. *Phys. Rev. B*, 68:045327, Jul 2003. doi: 10.1103/PhysRevB.68.045327. URL https://link.aps.org/doi/10.1103/PhysRevB.68.045327.

[22] Peter Y. Yu and Manuel Cardona. *Fundamentals of Semiconductors: Physics and Materials Properties*. Springer Berlin Heidelberg, 2010. ISBN 9783642007101. doi: 10.1007/978-3-642-00710-1. URL http://dx.doi.org/10.1007/978-3-642-00710-1.

[23] A. Manchon, H. C. Koo, J. Nitta, S. M. Frolov, and R. A. Duine. New perspectives for Rashba spin–orbit coupling. *Nature Materials*, 14(9):871–882, August 2015. ISSN 1476-4660. doi: 10.1038/nmat4360. URL http://dx.doi.org/10.1038/nmat4360.

[24] Shigeki Onoda, Naoyuki Sugimoto, and Naoto Nagaosa. Intrinsic Versus Extrinsic Anomalous Hall Effect in Ferromagnets. *Physical Review Letters*, 97(12), September 2006. ISSN 1079-7114. doi: 10.1103/physrevlett.97.126602. URL http://dx.doi.org/10.1103/PhysRevLett.97.126602.

[25] F. D. M. Haldane. Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the "Parity Anomaly". *Phys. Rev. Lett.*, 61:2015–2018, Oct 1988. doi: 10.1103/PhysRevLett.61.2015. URL https://link.aps.org/doi/10.1103/PhysRevLett.61.2015.

[26] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional analysis*. Academic Press, 1980.

[27] John B. Conway. *A Course in Functional Analysis*. Springer New York, 2007. ISBN 9781475743838. doi: 10.1007/978-1-4757-4383-8. URL http://dx.doi.org/10.1007/978-1-4757-4383-8.

[28] Edson de Faria and Welington de Melo. *Mathematical Aspects of Quantum Field Theory*. Cambridge University Press, August 2010. ISBN 9780511760532. doi: 10.1017/cbo9780511760532. URL http://dx.doi.org/10.1017/CBO9780511760532.

[29] J. M. Cook. The mathematics of second quantization. *Transactions of the American Mathematical Society*, 74(2):222–245, 1953. ISSN 1088-6850. doi: 10.1090/s0002-9947-1953-0053784-4. URL http://dx.doi.org/10.1090/S0002-9947-1953-0053784-4.

[30] Henrik Bruus and Karsten Flensberg. *Many-body quantum theory in condensed matter physics - an introduction*. Oxford University Press, 2002.